UNIVERSITÉ DE STRASBOURG

Ecole Doctorale des Sciences de la Vie et de la Santé

THÈSE

présentée pour obtenir le grade de

Docteur de l'Université de Strasbourg

Discipline : Sciences du Vivant

Domaine : Biophysique

par

Kamal kant SHARMA

Mécanisme moléculaire des propriétés chaperonnes de la protéine Core du virus de l'hépatite C : étude physicochimique par fluorescence et résonance plasmonique de surface.

Soutenue le 02 mai 2011 devant la commission d'examen :

Pr. Vincent PEYROT	**Rapporteur externe**
Dr. Philippe FOSSE	**Rapporteur externe**
Dr. Catherine ROYER	**Examinateur**
Dr. Catherine SCHUSTER	**Rapporteur interne**
Pr. Jean-Marc LESSINGER	**Co-Directeur de thèse**
Pr. Yves MÉLY	**Directeur de thèse**

UMR CNRS 7213, Faculté de Pharmacie, ILLKIRCH

UNIVERSITÉ DE STRASBOURG

Ecole Doctorale des Sciences de la Vie et de la Santé

THÈSE

présentée pour obtenir le grade de

Docteur de l'Université de Strasbourg

Discipline : Sciences du Vivant

Domaine : Biophysics

par

Kamal kant SHARMA

Molecular mechanism of the Hepatitis C Virus core protein chaperone properties: physicochemical investigation by Fluorescence and Surface Plasmon Resonance

Soutenue le 02 mai 2011 devant la commission d'examen :

Pr. Vincent PEYROT	**Rapporteur externe**
Dr. Philippe FOSSE	**Rapporteur externe**
Dr. Catherine ROYER	**Examinateur**
Dr. Catherine SCHUSTER	**Rapporteur interne**
Pr. Jean-Marc LESSINGER	**Co-Directeur de thèse**
Pr. Yves MÉLY	**Directeur de thèse**

UMR CNRS 7213, Faculté de Pharmacie, ILLKIRCH

My thesis project was performed in the Laboratoire de Biophotonique et Pharmacologie, directed by Prof. Yves MELY. I am grateful to him for providing me the opportunity to work in his lab.

I would like to express my heartfelt gratitude to Director of my thesis Prof. Yves MELY for his supervision, inspiration and encouragement during my research. He has been generous with his time and knowledge. It has been a privilege for me, to be a student of him. I would like to express my deepest gratitude and sincere appreciation to my co-Director Prof. Jean-Marc LESSINGER for help and guidance for monitoring my research.

I am greatly honored by the kind acceptance of Prof. Vincent PEYROT, Dr. Philippe FOSSE, Dr. Catherine ROYER, and Pr. Thomas F. BAUMERT as members of jury for my thesis and would like to thanks them for serving as my advisory committee.

Special thanks to Dr. Guy DUPORTAIL for his help, guidance and moral support during my research, and Dr. Hugues de ROCQUIGNY for his help with HPLC, peptide synthesis and of course for some good French translation. I would like to extend my gratitude to Dr. Andrey KLYMCHENKO for his valuable discussion and motivation, Dr. Christian BOUDIER for valuable discussion and help in stop flow and calorimetry experiments, Dr. Ludovic RICHERT, Dr Pascal DIDIER and Dr. Youri ARNTZ for fluorescence imaging and lifetime studies and for analytical model simulations. I would like give special thanks to my friend Dr. Julien GODET for his help and great discussions in experiments, especially in single molecule experiments and of course their analysis. I thank to the group of Dr. D. Altschuh and Dr. M. Baltzinger from ESBS, France, for providing data and informations related to SPR expeiments.

I would like to thank all my friends form the lab Dr. Namrata JAIN, Dr. Nicolas HUMBERT, Viktoria POSTUPALENKO, Oleksandr KUCHERAK, Beata BASTA, Zeinab DARWICH, Vanille GRENIER, Armelle JOUONANG, Hussein FTOUNI, Avisek GHOSH, and Nada KRAYEM for their constant help, encouragement and our great discussions. I wish them all the best of luck in their future endeavours.

On a personal note, I would like to thank my family and friends who have supported me along the way in my pursuit for my goals.

Table of Contents

Abbreviations

ASGP-R	asialoglycoprotein receptor
ALT	alanine aminotransferase
BVDV	Bovine viral diarrhoea virus
CSF	Classical swine fever virus
CKII	Casein kinase II
CLDN	claudins
DNA	De-oxyribonucleic Acid
DLS	Dimer linkage sequence
dTAR	DNA copy of the Transactivation response element of HIV-1
DC	dendritic cell
HCV	Hepatitis C Virus
HCC	hepatocellular carcinoma
HDL	High density lipoproteins
HVR	Hypervariable region
HCVpp	HCV pseudo particles
hnRNPC	heterogenous nuclear ribonucleoprotein C
HCVcc	cell culture model of HCV
IRES	Internal ribosome entry site
IFN	alpha-interferon
IFN-α/β	Type I interferons
ISGs	Interferon simulating genes
JFH-1	Japanese Fulminant hepatitis 1
LVP	Lipo-viroparticules
LTR	long terminal repeat
LPS	lipopolysacharrides
LDLR	lipoprotein receptor
MLV	murine leukemia virus
MHC	Histocompatibility Complex
NANBH	Non A non B hepatitis
NF-κB	nuclear factor κB
ORF	Open reading frame
ODN	Oligonucleotide
OCLN	occluding
PTB	Polypyrimidine tract-binding protein
PBS	primer binding site
PPT	polypurine track
PK	Proteinase K
pDCs	plasmacytoid dendritic cells
PAMP	pathogen-associated molecular pattern
RNA	Ribonucleic Acid
RdRp	Rna dependent Rna polymerase
SL2	Stem loop 2
SPR	Surface Plasmon resonance
SR-BI	scavenger receptor BI
SEL and LEL	small and large extracellular loop
siRNA	small interfering RNAs
SPP	signal peptide peptidase
SP	signal peptidase

TMD	Transmembrane domain
TJs	Tight junctions
TLR	Toll-like receptor
UTR	Non-translated region
VSL	Variable Stem loop
VLDL or LDL	Very low and low density lipoproteins
VLPs	Virus like particles
VAP	Vesicle-associated membrane protein
VRE	virus-responsive element

General Introduction

Hepatitis C virus (HCV) is an important human pathogen [1, 2], and major cause of chronic hepatitis and liver cirrhosis, leading to hepatocellular carcinoma [3]. It infects approximately 170 million people worldwide and its prevalence varies from country to country [4, 3]. The transmission of virus is done through blood contacts during transfusion or exchanges of syringes among drug addicts. Transmission from mother to child and sexual transmission also were deferred but seemed not very effective [5, 6], except during co-infection with HIV [7, 8]. Although combination of interferon and Ribavirin has been used as a therapeutic strategy, this treatment is expensive with an average efficiency and considerable side-effects [9]. It is therefore important to study the mechanism of HCV replication cycle in order to develop new antiviral strategies.

HCV, an enveloped virus member of the *Flaviviridae* family, has a positive-sense and single stranded RNA genome that encodes a polyprotein of 3010 amino acids [2]. The 9.6 kb genomic RNA plays complex, temporally and spatially regulated roles throughout the virus life cycle, by serving both as a template for minus-strand RNA synthesis and as an mRNA directing the translation of viral proteins, and lastly, by specifically packaging itself into newly made progeny virions. The coding part of the genomic RNA is flanked by two non-translated, but highly preserved and structured regions (5' UTR and 3' UTR) of 341 and 230 nucleotides, respectively. The 5' UTR folds into a complex, four-domain structure that contains an internal ribosome entry site (IRES), which directs the cap-independent translation of ~3011-codon-long open reading frame, ORF, as well as upstream signals required for viral RNA synthesis [10]. The 3'UTR is a tripartite structure consisting of: (I) an upstream variable part with VSL1 and VSL2 stem-loops, (II) a poly (U/UC) tract, and (III) a highly conserved 98 nucleotide long 3' X-tail, that putatively forms three stem-loop structures (designated as SL1, SL2 and SL3, in the 3'-to-5' direction) [11, 12]. The X-tail is highly conserved even between the most divergent HCV isolates, suggesting its important role in HCV replication. In addition, it plays a crucial role, in controlling the amount of genomic RNA product in its protein translation and encapsidation, during replication of (-)RNA and (+)RNA in the cytoplasm of the host cell.

The polyprotein precursor is co- and post-translationally processed by cellular and viral proteases at the level of the endoplasmic reticulum (ER) membrane to give ten proteins, four of whom have structural functions (Core, E1, E2 and p7). The N-terminus of the polyprotein

is cleaved by a cellular peptidase, generating the core of 191 residues that binds to the membrane of the endoplasmic reticulum (ER) [13]. This protein is then matured by a SPP (signal peptide peptidase) [14, 15], giving a protein of 179 residues. This protein of 179 amino acids consists of a 117 residue long hydrophilic and poorly folded N-terminal domain (Domain D1) with three basic sub-domains (BD1-BD3) responsible for the interactions with RNA and a hydrophobic C-terminal domain (Domain D2) involved in targeting the protein to lipid droplets, membranes of the nucleus and mitochondria [16-19].

The protein has multiple functions. By interacting with several cellular proteins [4, 20], the core alters the intracellular trafficking and consequently, proliferation and cell differentiation. In addition, transgenic mouse models indicate that this protein also alters the fat metabolism in the cells and induces liver carcinomas [16, 21]. The core also opposed to apoptosis, but its role in this phenomenon remains controversial [17]. Moreover, HCV infection is characterized by chronic hepatitis and many authors have highlighted the role of core in maintaining the virus in the host cell. Recently it was shown that this protein acts in *trans* on IRES (Internal Ribosome Entry Site) [22], by specifically modulating its dose dependent translational activity. Finally, the chaperone activities of core [23, 24] were shown to direct nucleic acid structures to their most stable conformations [25] and stimulate hybridization of complementary DNA or RNA sequences [26]. These activities enable the core to dimerize the 3' UTR end of HCV genomic RNA as well as its transcription into a copy of negative polarity, as (-)HCV RNA. This dimerization is based on the palindromic sequence (residues 23-38), named DLS (Dimer Linkage Structure), present in the stem-loop SL2 of the 3'UTR X-Tail region. The chaperone activity of the core is carried by the D1 domain and more specifically, the basic areas BD1 to BD3 [24].

Apart from encapsidation of single-stranded (+) RNA, the chaperone activity of core in the X-region may be needed at several stages of the viral life cycle. It could for instance play a role in the synthesis and translation of (-) RNA strands [27]. Moreover, dimerization of HCV RNA could facilitate the release of the (+) RNA from the translation machinery and allow its encapsidation in the viral particle. Moreover, as the nucleocapsid protein of HIV, the chaperone properties of the core could play a role in packaging and compaction of the genomic RNA during virus assembly. Finally, by homology with the crucial role of dimers in RNA recombination in retroviruses [28], dimerization of the X-region may promote the hybridization of genomic subtypes. Thus, it would be one of the factors that, concomitantly with the lack of fidelity of the RNA dependent RNA polymerase activity, promote the

production of a heterogeneous population of virus responsible for the impaired immune response during infection [29, 27, 30].

The objective of this thesis was to characterize and quantify the molecular chaperone activities of the HCV core protein by fluorescence techniques and surface plasmon resonance (SPR). For this, we used oligonucleotides (ODNs) corresponding to the 3 'UTR X-Tail region (native and mutated forms) and a model system based on the annealing of the DNA copy of the Transactivation response element (dTAR) of HIV-1 and its complementary sequence cTAR with synthetic peptides corresponding to fragments of the N-terminal D1 domain of the core. We characterize and quantify the binding of core peptides to HCV ODNs, destabilization of intramolecular secondary structures of hairpin-loop-structured ODNs and the dimerization of self-complementary sequences in the presence and absence of core. We also characterized the molecular mechanism of the RNA chaperone activity of HCV core protein on model sequences and HCV oligonucleotides from 3' X-tail region. We further proved that the structural determinants for the core activity mainly reside in its 3 basic domains of N-terminus and core can bind specifically to the structured ODNs that are in abundance in the 3' X-Tail region of the HCV genome.

This study has uncovered important details regarding the chaperone mechanisms and interaction of HCV core protein with the 3'X-tail region and model sequences. This understanding of the molecular mechanism of the core could help to propose new strategies to identify, by screening or rational design, inhibitors targeting the interaction between core protein and its nucleic targets.

The project was further prolonged by revealing molecular mechanisms and structural determinants of other chaperone proteins like the Human prion protein, Tat and NCp7 from HIV-1 virus and Nucleocapsid protein from Bunyamwera Orthobunyavirus.

Bibliographic Review

1: Background

Liver is the largest organ of the body. It is composed of several types of cells with well-defined functionalities (Figure 1).

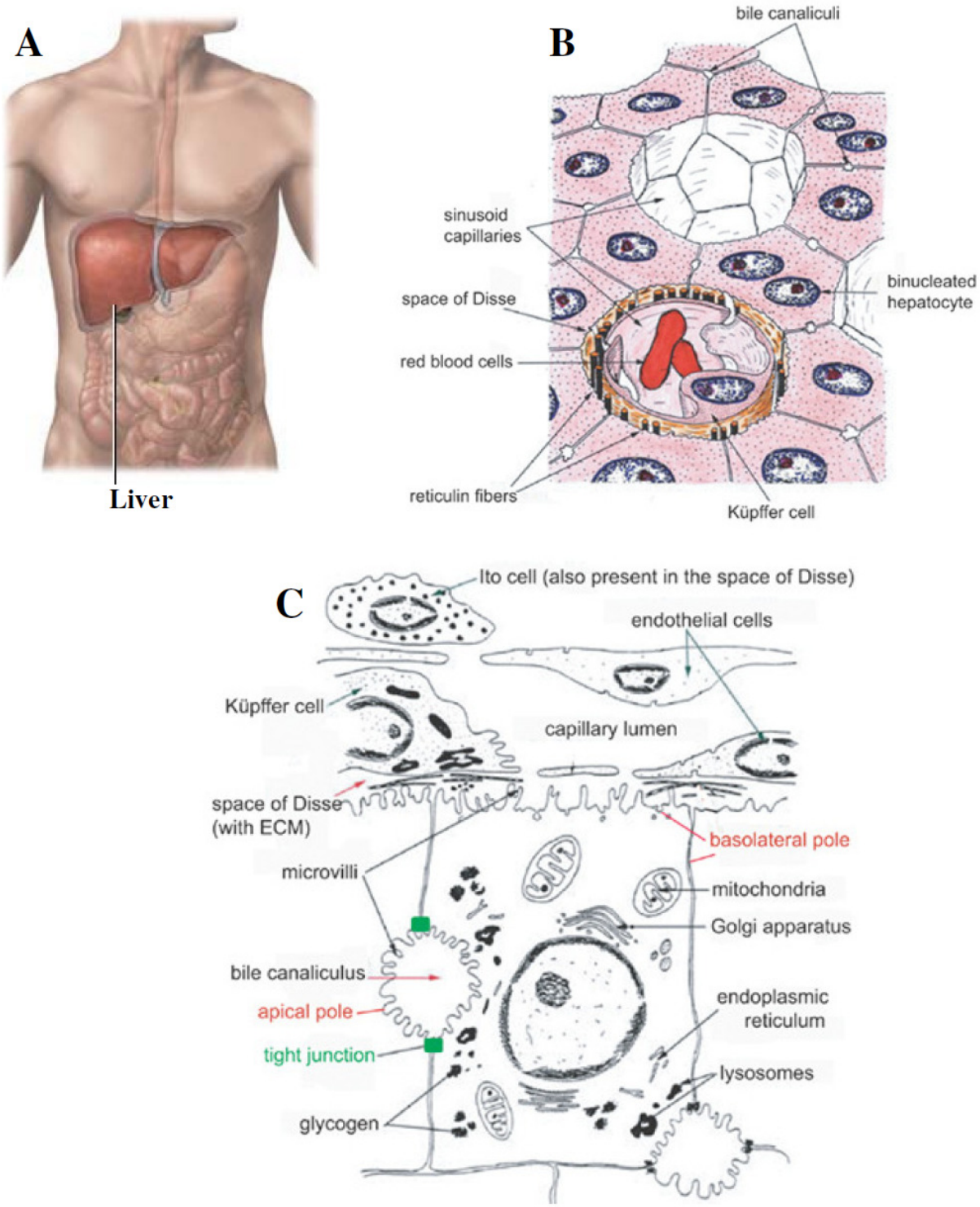


Figure 1: Architecture of Liver parenchyma. Location of human liver (A) and Architecture of the hepatic trabecula (B) and hepatocyte (C) (based on [31])

The liver plays essential roles in metabolism, including carbohydrate metabolism (glycogenogenesis, glyogenolysis, neoglucogenesis in order to maintain glucose homeostasis), lipid metabolism (beta-oxidation of fatty acids, cholesterol synthesis, lipoprotein synthesis) and protein metabolism (synthesis of most plasma proteins, protein catabolism and removal of ammonia by synthesis of urea).

Since liver has vascular, metabolic, secretory and excretory essencial functions, effect of any liver diseases will hamper the proper functioning of human body. There are several causes for these deseases like drugs, alcohol, toxins, viral infections, autoimmune diseases, or congenital diseases. There are several specific signatures of liver diseases like cholestatis (reduced or stopped bile flow), liver inflammation, hepatomegaly, portal hypertension (increased blood pressure in the portal blood vessels which supplies blood to liver from intestine), ascites (fluid retention in the abdominal cavity due to leakage from the liver and intestine) and hepatic encephalopathy (deterioration of brain functions due to non-removal of toxins and ammonia by liver from the blood). Jaundice, caused by increased levels of bilirubin, is the major clinical sign of liver diseases.

Hepatitis or inflammation of the liver, first described in the 1900s, caused by viral infection is one of the major causes of concern. The associated groups of viruses were defined as hepatitis viruses. These viruses are taxonomically different but they are related by their primary disease site i.e. liver. These viruses cause liver inflammation by targeting host cell hepatocytes which represent 80% of the total liver cell count (Figure 1). Among them, one of the major and deadly causative agents of the inflammation is Hepatitis C Virus (HCV).

2: Discovery of Hepatitis C Virus

In the 1970s, while developing serological tests for detection of hepatitis A or B (HAV or HBV) in patients suffering from hepatitis, Feinstone and colleagues observed that some patients were negative for virus hepatitis A and B. On the basis of these tests, they deduced the existence of an unknown agent which was capable of causing hepatitis non A non B (NANBH) [32]. In fact, in 1978, Alter and his team succeeded in infecting chimpanzees from sera of patients with hepatitis NANBH [33]. But no conventional technique did permit the isolation of antigens, or infectious particles.

However, ten years later after many attempts, Choo et al. produced a cDNA expression library from the serum of infected chimpanzees by NANBH [1]. The modern techniques of molecular cloning and phage displayed the molecular characteristics of the causative agent of NANBH,

a new RNA virus, termed Hepatitis C Virus (HCV). HCV causes a persistent infection in the majority of infected people and can lead to cirrhosis of the liver and subsequently hepatocellular carcinoma (HCC) [34-36]. For this reason, and the high prevalence of its infection worldwide, HCV is rightly classed as a major human pathogen.

3: Epidemiology

The World Health Organization estimates that approximately 170 million people in the world, representing a prevalence of 3%, are infected with HCV. Among them, 130 million are chronic carriers and may develop cirrhosis or hepatocellular carcinoma (HCC). Although HCV is endemic worldwide, there is great geographic variability of its distribution (Figure 2). Africa and Asia have the highest reported prevalence rates while the lowest are found in industrialised countries such as North America, Australia and those in Northern and Western Europe [37]. The highest reported seroprevalence rate is in Egypt where approximately 22% of the population are HCV seropositive, probably due to contaminated glass syringes used in nationwide schistosomiasis treatment campaigns from 1960 to 1987 [38]. In industrialized countries, the disease prevalence is low (<1%) while its high (> 2%) in many countries of Africa, America Latin and Central and South-East Asia. In these developing countries, the prevalence varies between 5% and 10% and sometimes more. In France, it is estimated that 550,000-600,000 people are infected with HCV which counts to 1 to 1.2% of the total population.

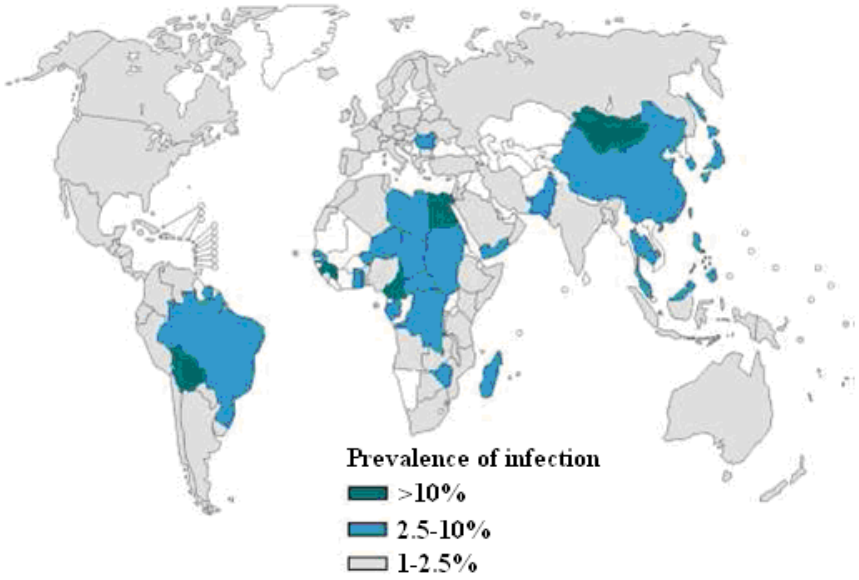


Figure 2: HCV prevalence worldwide (source OMS 2004)

4: Classification of HCV

HCV has been categorized as a member of *Hepacivirus* genus within *Flaviviridae* family by genome analogy with other members of this family [1, 2, 38, 39]. This family also includes other flaviviruses [dengue virus and Japanese encephalitis virus [40], the pestiviruses [such as bovine viral diarrhoea virus and classic swine fever virus [40], hepatitis G virus [41] and GBV-A, B and C viruses [42, 43]. GBV-A,B and C viruses were named from initials of a patient (GB), whose inoculum with acute sporadic hepatitis was used to successfully infect tamarins. The viral genomes subsequently cloned from serum of an infected animal.

There are many similarities between HCV and other members of family.

- 3011 amino acids [1, 2] long viral-genome-encoded ploy protein is comparable in size to other members of the *Flaviviridae* such as the flavivirus yellow fever virus (YFV, ~3460 aa), and the pestivirus bovine viral diarrhoea virus (BVDV, ~3960 aa).
- The structural proteins of both the flavi- and pestiviruses are located at the N termini of their polyproteins also beginning with a small, basic nucleocapsid protein [44, 45].
- HCV, flavi- and pestiviral polyproteins, all share similar kind of hydropathy index. Alignment of the HCV genome with other genomes of the *Flaviviridae* shows regions of sequences homology as well as comparable genomic organisation [46].
- There exist similarities (like internal ribosome entry site, IRES) and conserved stem-loop structures between nucleotide sequencing of the HCV 5' un-translated region (UTR) compared to those of pestiviruses. Several tracts identical to 5'UTR sequence of HCV have been identified in BVDV and classical swine fever virus (CSF) [47, 48].

Despite many similarities between HCV and both flaviriruses and pestiviruses, significant differences exist that lead to the proposal of a third *Flaviviridae* genus, the *hepaciviruses* [42].

In this new genus, HCV is divided into 6 genotypes (1-6) and several subtypes. The genotypes differ from each other by 30% to 35% in their nucleotide sequencing [49]. In the same genotype there are different subtypes (designated by letters) with a difference of 20% on average in sequencing [49]. This leads to great genetic heterogeneity within the same individual, resulting in the existence of quasi-species [50].

HCV genotypes are distributed according to geographical areas of world (Figure 3).

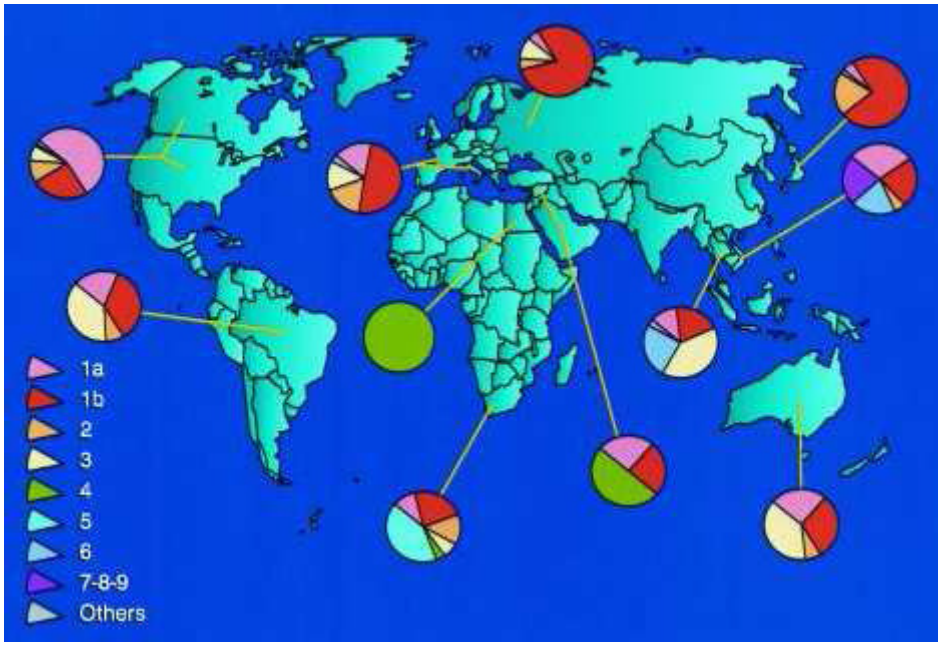


Figure 3: Geographical distribution of different genotypes of HCV (based on [51]).

Although HCV genotypes 1, 2 and 3 seem to have a worldwide distribution, their relative prevalence varies from one geographic region to another (Figure 3). Subtypes 1a and 1b are the most common genotypes in the United States and Europe, subtypes 2a and 2b are relatively common in North America, Europe and Japan while subtype 2c is commonly found in northern Italy [51]. The genotype 4 is widespread in Egypt and the Middle East. Genotypes 5 and 6 appear to be confined to South Africa and Southeast Asia. In addition, there is a genotype 7, identified in some Asian populations [51].

These genotypes also differ in their response to treatment and disease progression. Patients infected with genotype 1 respond less well to alpha-interferon (IFN) treatment than compared to those infected with genotype 2 or 3.

So, Infection with HCV is a very dynamic process with hepatocytes infected with a half-life time of few hours and 10^{12} virions produced per day within single individual. This high replicative capacity as well as high frequency error (10^{-3} to 10^{-4} per nucleotide position) made it a highly variable genetic virus.

5: Biophysical properties of the virus particle

HCV is a spherically enveloped virus of 55-65 nm in diameter [52]. It consists of a nucleocapsid of 30-35 nm in diameter [53], surrounded by an envelope in which glycoproteins E1 and E2 are anchored (Figure 4).

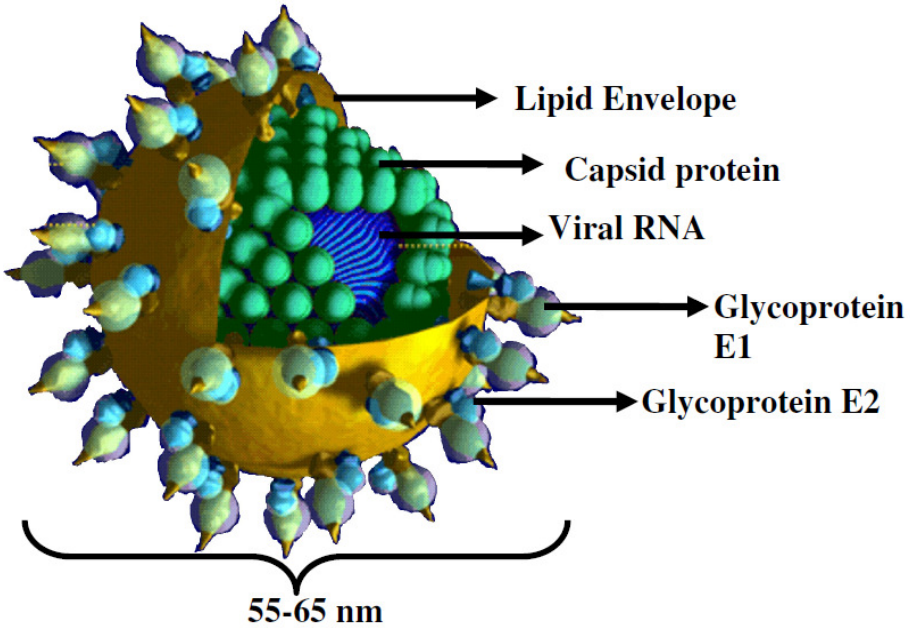


Figure 4: Schematic representation of HCV

Studies have shown that HCV circulates in different forms in the blood of infected patients. It has a heterogeneous distribution in its density gradient and is associated with lipoproteins of both low density, (VLDL or LDL) (1.03 to 1.08 g / ml), and high density (1.17 to 1.25 g / ml) [53]. It has been found that the particles of low density (associated with VLDL or LDL) are more infectious than those of higher density (associated with HDL and IgG) [53-56], and thus called low density lipo-viroparticules (LVP) [53]. In addition to HCV RNA, they also contain capsid proteins and envelope glycoproteins E1 and E2 of apoB and apoE. Furthermore, the production of HCV in hepatocytes has been shown to be dependent on assembly and secretion of VLDL, suggesting that HCV particles are associated with VLDL [57, 58].

6: HCV Genome

6.a: Genome Organization

HCV has a single stranded, positive sense genomic RNA of ~9.6 kb. The coding part of the genomic RNA is flanked by two non-translated, but highly preserved and structured regions (5' UTR and 3' UTR) of 341 and 230 nucleotides, respectively (Figure 5) [59, 48, 60]. In between these flanked regions the genome contains one long open reading frame (ORF) encoding a polyprotein of ~3011 amino acids [2]. Similarities in structure between the HCV 5'UTR and that of picornaviruses led to the conclusion that HCV had an internal ribosome entry site (IRES), required for translation initiation [60] (Figure 8).

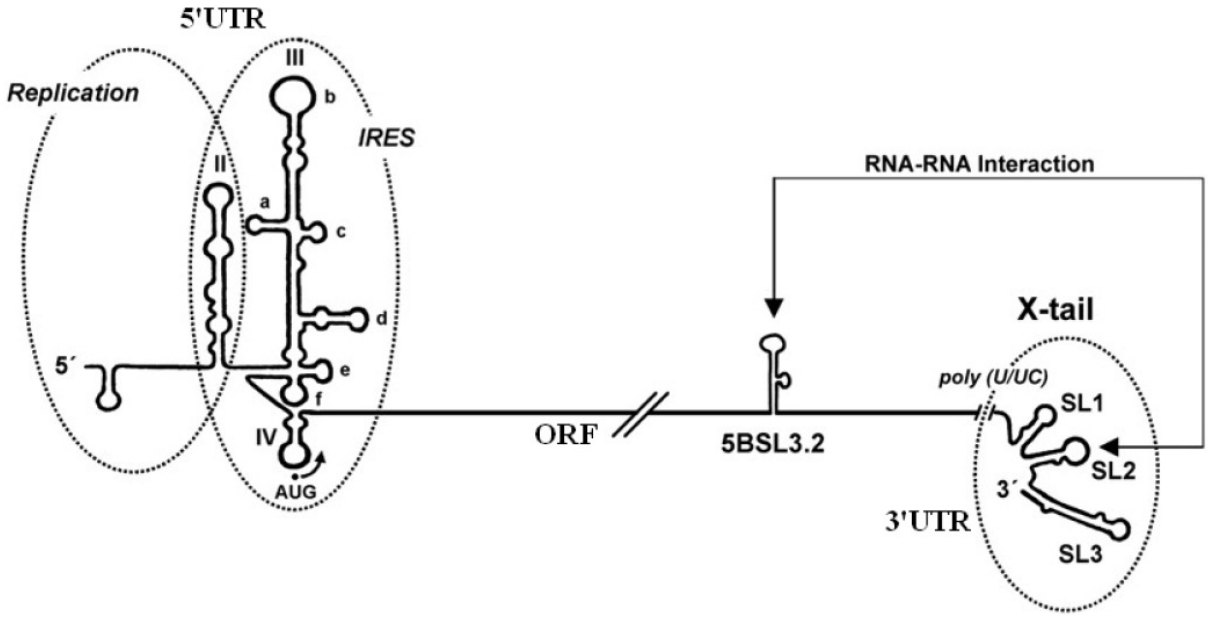


Figure 5: Secondary structure of HCV genome.

Additionally, the single ORF encoded polyprotein is co- and post-translationally processed by host and viral proteases to produce at least 10 viral proteins [61]. The structural proteins (core, E1, E2 and p7) are located within the amino-terminal of the polyprotein, while the non-structural proteins (NS2, NS3, NS4A, NS5A and NS5B) are found within the carboxy-terminal (Figure 6) [62, 61].

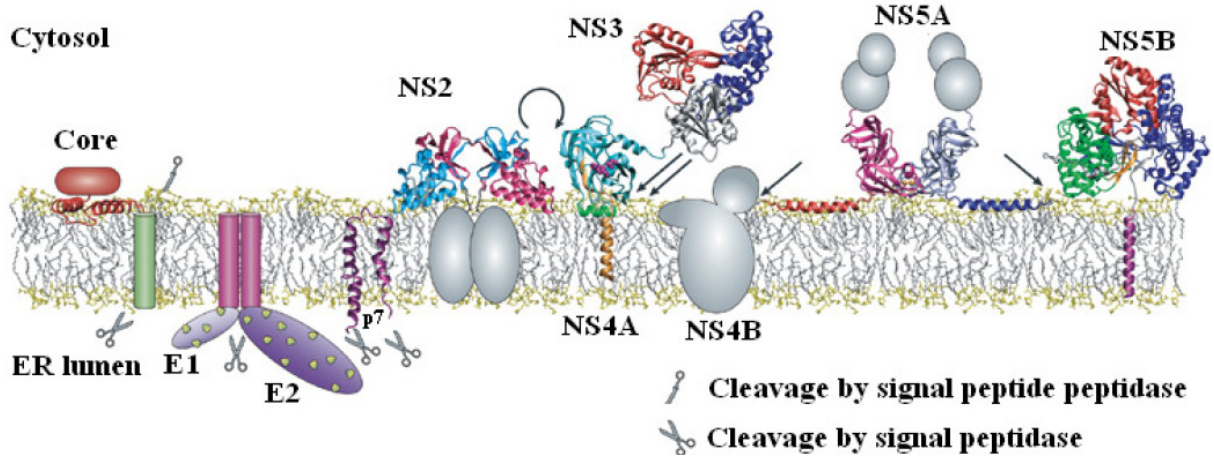


Figure 6: HCV proteins and their association with the ER membrane (From [63]).

The structural proteins are processed by host proteases while the non-structural proteins by viral proteases [64]. Furthermore, the 3'UTR is a tripartite structure consisting of a conventional 3'end, a poly(U) tract and a 3' X-tail [12]. The X-tail is highly conserved even

between the most divergent HCV isolates, suggesting its important role in HCV replication [12, 65], as proven by the requirement of the 3' X-tail for replication of the HCV replicon [66].

6.b: 5'UTR

The 5'UTR forms a secondary structure comprising four major separate structural domains I-IV. Domains II-IV alongwith the first 24-40 nucleotides of ORF sequence comprises internal ribosome entry site (IRES) (Figure 7) [10, 67]. HCV undergoes internal initiation of translation via IRES in a cap-independent manner [60, 68]. This translation is similar to that of *pestiviruses* [69] and antagonistic to the members of the *Flavivirus* genus, where they tend to have smaller 5'UTRs and undergo 5' cap-dependent translation (like eukaryotic mRNAs). Translation efficiency from the IRES has been shown to vary between different genotypes [44]. Thus, by analogy with other RNA viruses using IRES, it was suggested that the 5'UTR contains not only the determinants for translation, but also elements for RNA replication [70-72].

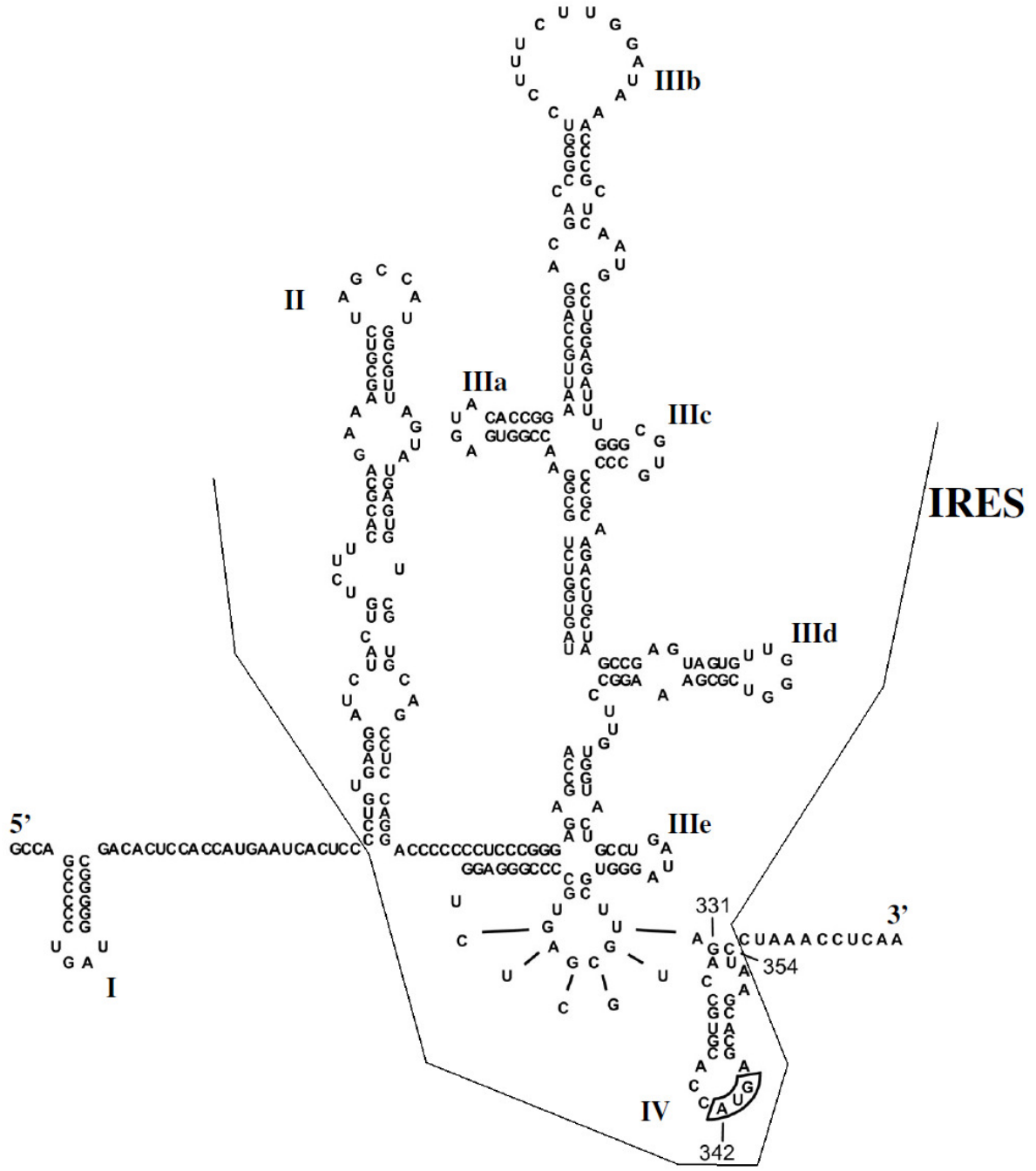


Figure 7: Proposed secondary and tertiary structure of the HCV 5'UTR, including a small portion of the core coding sequence, from a genotype 1b infectious clone (HCV-N). Major structural domains are labeled with Roman numerals and the initiator codon within stem-loop IV is highlighted (based on [73]).

However, the interaction between IRES and ribosome is likely to require several protein co-factors. Interestingly, a number of cellular proteins have been shown to bind the 5'UTR like polypyrimidine tract-binding protein (PTB) [74], the RNA binding protein La [75] and poly(C)-binding proteins (PCBs) 1 and 2 [76]. In addition, a microRNA (miR-122), specifically

expressed in hepatocytes, has been shown to facilitate the replication of HCV RNA by interacting with the 5'UTR region in cell culture [77].

6.c: Viral proteins

6.c.1: Core

The general properties, putative pathogenic roles and interactions of the core protein are discussed in details in later sections**.**

6.c.2: E1 and E2

E1 and E2 are transmembrane proteins anchored to the viral envelope. These proteins are thought to be involved in host-cell entry by binding to cell surface receptors and fusing with host-cell membranes [78]. They both are type I transmembrane (TM) glycoproteins, each comprising a large N-terminal ectodomain of 160 and 334 amino acids respectively, and a short C-terminal transmembrane domain (TMD) [79]. In their functional form, these proteins assemble into non-covalent heterodimers, dependent on specific transmembrane areas (Figure 8) [80].

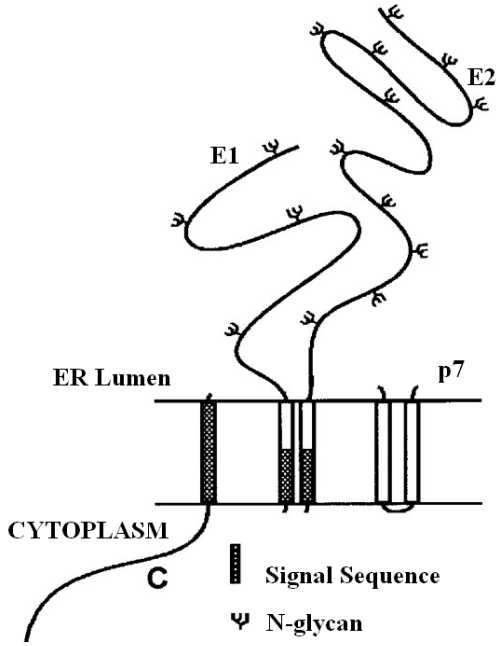


Figure 8: Schematic representation of envelope glycoproteins E1 and E2 of HCV and their glycosylation sites (based on [80]).

Both glycoproteins are modified by N-linked glycosylation (the recognition of a specific type of amino pattern - Asn-X-Ser/Thr-Y) with E1 and E2 having up to 6 and 11 gycosylation

sites, respectively [80]. These gycosylation sites are highly conserved among HCV genotypes. On the other hand, the TMDs, composed of 2 short stretches of hydrophobic residues separated by a short polar segment, are essential for heterodimerisation [81]. These TMDs are also essential for glycoprotein retention in the Endoplasmic Reticulum (ER) (*due to at least one charged residue within the polar segment*), where virion assembly takes place [82, 83, 79].

Furthermore, E2 contains a hypervariable region of about 27 amino acids (384 to 410) at the N-terminus, called HVR-1, which is a target for anti-HCV neutralizing antibodies. This region is highly tolerant to amino acid substitution and is subject to strong positive selection pressure [62, 84]. With this high variability, the virus can easily escape the immune system and persist in its host [85]. Although the amino acid sequence of HVR-1 is highly variable, still its conformation is well conserved. The sequence made up of basic residues is possibly required for interaction with negatively charged molecules such as lipids, proteins, glycosaminoglycans [86] and virus receptor SR-BI [87]. Moreover, it has been shown that HVR-1 targeting antibodies inhibited entry of HCV pseudo particles (HCVpp- which will be described in later section) as well as binding of Virus like particles (VLPs- will be described in later section) to cells [88].

However, it is also possible that other hypervariable regions may play a role in viral entry like HVR-2 (residues 474-482) [89, 84] and HVR-3 (residues 431-466) [90].

On the other hand, E1 has been proposed to be involved in the membrane fusion required for internalization of the virus into the host cell [91], still its role in HCV infection is poorly understood.

6.c.3: p7

The identification of two distinct forms of N-deglycosylated E2 led to the discovery of a small hydrophobic protein between E2 and NS2 termed p7 [92]. It is composed of 63 amino acids (amino acids 747-809) and made up of mostly uncharged polar or hydrophobic residues interspersed by a few conserved charged residues. This composition allows a double-transmembrane spanning conformation for p7, separated by a few residues on the cytoplasmic side and with the N-and C-terminal directed towards the ER lumen [93, 92] (Figure 9). It is post-translationally processed (*due to the presence of an E2-p7-NS2 precursor and stable E2-p7 form*) and sub-optimaly cleaved (*because of a weak signal peptide cleavage site*) from the polyprotein by a host signal peptidase [94, 92].

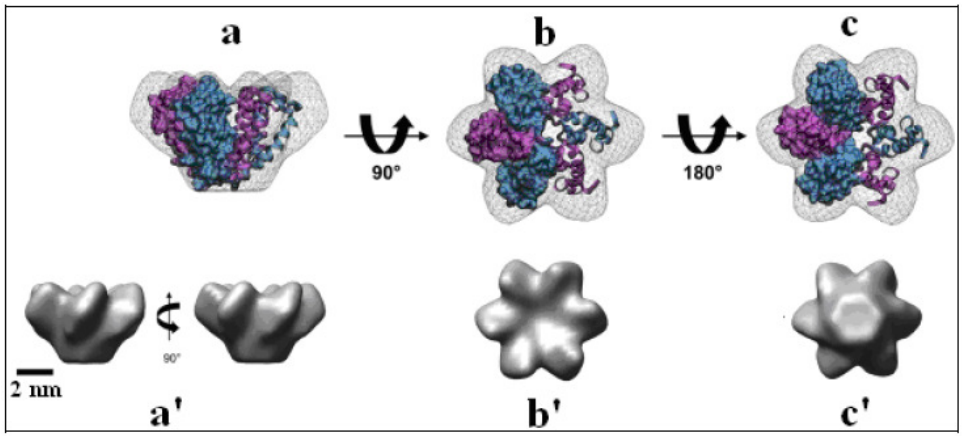


Figure 9: Atomistic (a, b, c) and 3D (a', b', c') models of p7 monomers (from [95]). Three p7 proteins assemble into hexamer, forming an ion channel. The six protuberances or flower petals form the upper channel, oriented towards the ER lumen and the base forms the channel.

It has been reported that this protein forms an oligomer with all the characteristics of an ion channel [93]. It is essential for HCV infection in chimpanzees [96] and also acts at an early stage of the morphogenesis of infectious particles [97]. The combination of p7 to the ER also appears to be important for assembly and secretion of infectious viral particles of HCV [98] as it has been shown that a single mutation in p7 increases the yield of infectious virus from an H77S/JFH-1 chimeric clone [99].

The fact that this protein has been reported to be localized in both ER and mitochondria, suggests that it exists in multiple populations and possibly exhibits more than one role in the viral life cycle. One such role of p7 is as an ion channel to regulate the flow of calcium ions from the ER to the cytoplasm [100, 101]. This ion channel is formed by the assembly of three p7 molecules. As revealed by high resolution electron microscopy (Single-particle electron microscopy) [95] (Figure 10), this protein assemble into a hexamer of 42 kDa, with an architecture of flower petals, facing the ER lumen. The widest part of the channel has a relatively large number of potential sites for interaction with cellular or viral proteins present in the ER membrane [95]. As more is known about the properties of p7, it is becoming clear that this is an important factor in the virus life cycle, presumably at a post-replication stage. This makes p7 an interesting target for future antiviral studies as it may be possible to block the proposed ion channel and thus inhibit the production of infectious particles.

6.c.4: NS2

The NS2 protein (amino acids 810 to 1026) is a transmembrane protein (Figure 6) of 21-23 kDa [102], with possibly 4 TM domains [103]. In conjunction with the N-terminal end of NS3, this protein, in the presence of zinc ions, is responsible for the cleavage at the NS2/3 junction [104, 55, 56]. Residues *His_{952}* and *Cys_{993}* have been found essential for the NS2 protease function but have no effect on downstream cleavage events [104], presumably because downstream cleavage is carried out by the NS3 protease. The crystallographic analysis of the catalytic domain of NS2 has revealed a protease cysteine dimer with two active sites (Figure 10). These catalytic sites are formed by His_{143} and Glu_{13} residues of one monomer and Cys_{184} residue of the other monomer [105].

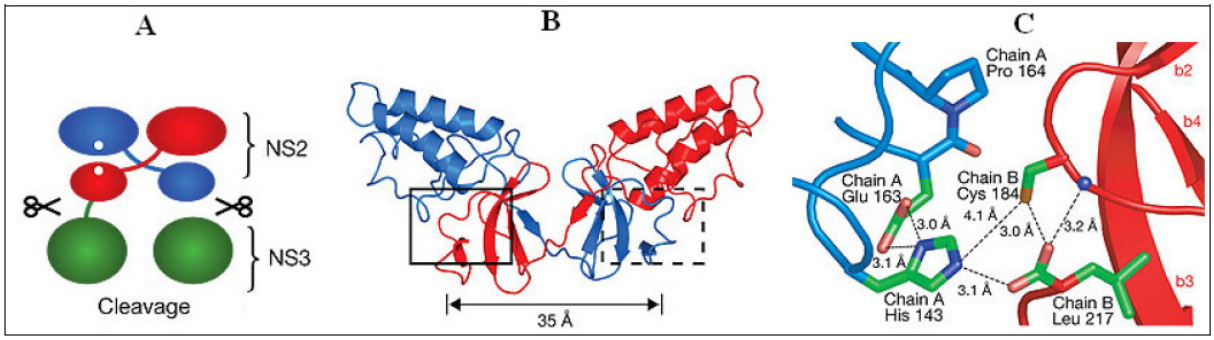


Figure 10: Structure of the catalytic domain after cleavage by the NS2 protease NS2/3 (according to [105]). (A) Model of action of the protease NS2/3 (B) NS2 forms a dimer: monomers are shown in blue and red. (C) Active site of the protease before cleavage, cleavage requires the C-termini of two monomers.

Although, NS2 is not essential for replication, as shown by replication of sub-genomic replicons lacking the NS2 gene [106], it exhibits other functions apart from NS2/3 cleavage. These functions are not directly involved in replication. For instance, it plays an essential role in the phosphorylation of NS5A [107] and is thus involved indirectly in infectivity of HCV [108]. It has been shown that the first transmembrane region of NS2 is important for production of infectious virions. Several residues are involved, but especially a glycine residue (Gly_{10}), at a central position of the first transmembrane region, is essential for the assembly of HCV particles [109]. However, catalytic activity of NS2 is not required for the formation of infectious virus but nevertheless its presence is essential [109, 105]. This was confirmed when the mutation of residues at the dimer interface or in the C-terminal region did not affect

replication and NS2 stability or NS2 protease activity, but decreased the production of infectious viruses [110].

6.c.5: NS3

NS3 (amino acids 1027 to 1657) is a multifunctional protein of 68 kDa. Its N-terminal contains a serine protease domain responsible for downstream cleavages [111-113, 55, 56, 114], while the C-terminal contains an NTPase/helicase domain [115-117]. The two functional domains are not cleaved from each other [111] but can act independently [115]. The serine protease domain is required for 4 cleavage events, acting in *cis* to release itself from the HCV polyprotein and in *trans* to produce the N-termini of NS4B, NS5A and NS5B [111, 118, 119, 114]. In addition to the N-terminal of NS3, a C-terminal domain of NS4A has been described as an NS3 co-factor required for efficient cleavage of the downstream polyprotein, at the NS4B/5A cleavage site [120, 121].

While emphasizing the functional helicase role of the C-terminus of NS3 protein, it has been shown that the associated enzymes are capable to unwind RNA-RNA, RNA-DNA and DNA-DNA duplexes, in the presence of ATP. Although the monomeric form of NS3 is capable to bind, a dimer of NS3 is required by RNA sequences [122]. The helicase activity is required for RNA replication. It has been suggested that this domain has a role in unwinding of viral RNA during replication, as it interacts with the 3' terminal sequence of both positive and negative HCV RNA [123]. At the same time an active form of the helicase domain is required for the replication of an HCV subgenomic replicon [124].

NS3 has been shown to bind many cellular factors including protein kinase A, protein kinase C, tumour suppressor p53 and histones H2B & H4 [125] and has effects on various processes such as cell metabolism, differentiation and tumour promotion [126, 127]. This multifunctionality of NS3, together with its interactions with numerous cellular factors, indicates that this protein plays a pivotal role in the life cycle of HCV. Its role is not limited only to replication of the viral RNA but extends to interactions with host-cell proteins which may result in some of the pathogenic effects, associated with HCV infection. Finally, NS3 is also involved in the assembly of infectious particles[128]. Due to its enzymatic activities, NS3 is a possible antiviral target.

6.c.6: NS4A

The hydrophobic region NS4 of HCV is cleaved by the NS3 protease to produce NS4A and NS4B [104]. NS4A acts as a co-factor of NS3 protease, through its central hydrophobic domain (amino acids 21-32) [129, 130]. NS4A helps in anchoring NS3 to the ER membrane. NS4A forms a complex structure, with NS4B, NS5A and NS5B [131], that inhibit protein synthesis at the translation stage [132, 133, 131]. This protein inhibition function of NS4A suggests that this protein is involved in the switch from translation of the HCV polyprotein to replication of viral RNA once sufficient protein production/processing has occurred. Additionally, it also plays a role in abrogating the host defenses. Moreover, the precursor protein NS4A/4B slows down the rate of ER-Golgi trafficking, in a similar manner to that of Poliovirus 3A protein [134], when they are expressed in isolation [135]. The virus would require an alternative route of exit, if ER-Golgi trafficking is slowed down, like Rotaviruses as they exit the cell via a non-classical secretory route that bypasses the Golgi apparatus [136]. Thus, HCV probably exits via a similar non-Golgi route and affects host defenses, such as trafficking of MHC antigen [135]. Finally an enhancement of NS5A phosphorylation by cellular kinases has also been observed in the presence of NS4A [137-139].

6.c.7: NS4B

NS4B protein (amino acids 1712 to 1972) is a hydrophobic protein of 27 kDa, exclusively present as a fine membranous web in the cytoplasm [140, 141] (Figure 6). These webs forming in a specialized membrane compartment, derived from ER membranes, are supposed to be the headquarters for viral RNA replication [142].

This protein contains at least four TMs with its N-and C-termini oriented towards the cytosol. These N-and C-termini are important for the formation of a replication complex [143, 144]. Furthermore, a fifth TM domain is also observed upon the predicted translocation of the N-terminus towards the luminal side of the ER. This translocation is impaired by NS5A which suggests that NS4B has different functions during the viral life cycle [145]. Moreover, this N-terminus seems responsible for the formation of the "membranous web" [145]. In addition, palmitoylation of two cysteine residues, 257 and 261, at the C-terminus of NS4B seems important for the formation of the replication complex [146] since abolishing palmitoylation at Cys_{261} completely inhibits replication of a sub-genomic replicon [146].

A nucleotide-binding motif has also been identified in NS4B. Genetic disruption of this motif impairs GTP binding/hydrolysis and inhibits RNA replication in a replicon system [147]. The

importance of this nucleotide-binding motif for replication suggests that besides simply providing structural scaffolding for the replication complex, NS4B has other roles in replication. Finally, the oligomerization of NS4B is required for the assembly of a functional viral capsid [143, 144].

6.c.8: NS5A

NS5A (amino acids 1973 to 2420) is a phosphoprotein anchored in the ER membrane through its N-terminus folded into an amphipathic α helix (H1) (Figure 11) [148, 78]. This α helix could serve as a platform for protein-protein interactions which are necessary for the assembly of a functional viral capsid [78]. NS5A exists in two phosphorylated forms, a basally phosphorylated 56 kDa form and a hyperphosphorylated 58 kDa form [137, 139], and composed of three distinct domains (Domain I, II and III) separated by sequences of lower complexity (LCSI and II) (Figure 12) [149].

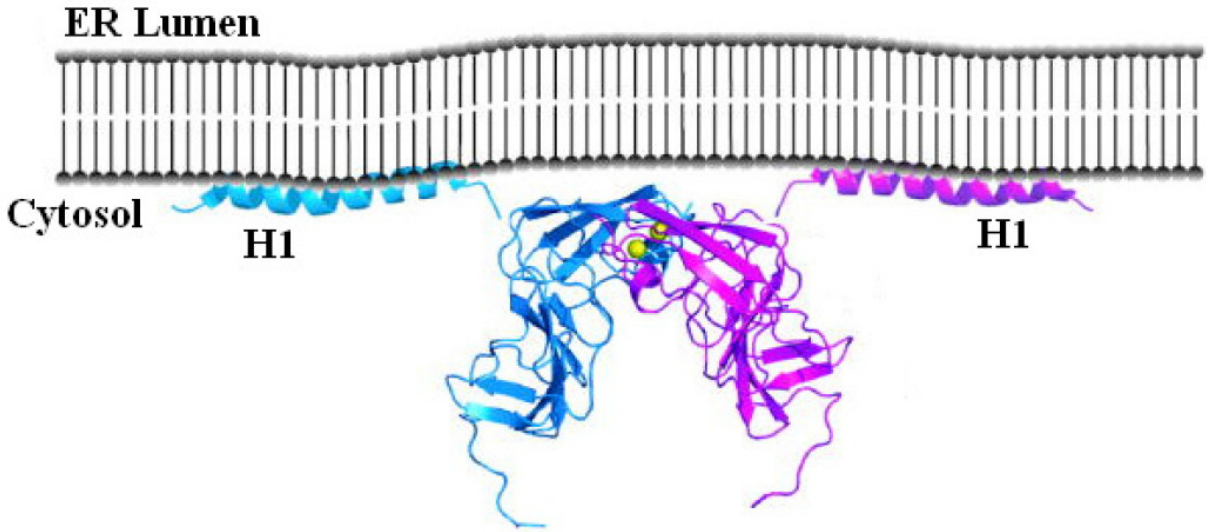


Figure 11: Proposed model of the structure of the NS5A protein bound to the ER membrane (From [150]).

Domain I, responsible for RNA binding [150], includes the basic N-terminal subdomain IA and a predominantly acidic C-terminal subdomain IB [150]. This domain, due to its interaction with core, is also important for localization of NS5A in lipid droplets [151]. The subdomain IA contains four cysteine residues, $Cys_{39, 57, 59 \text{ and } 80}$, and a zinc binding site. This zinc-binding site is critical for proper NS5A folding, and its mutation abolishes viral replication. Domain I forms homodimers through interaction near the N-termini. These homodimers arrange

themselves to form a groove located on the basal membrane. This groove is supposed to be the docking site for RNA. Due to this domain I, NS5A protein strongly interacts with the pyrimidine-rich part of the untranslated strand of viral (+)RNA [152]. Similarly, the subdomain IB contains two cysteine residues, $Cys_{142 \text{ and } 190}$, linked by a disulfide bridge, but this subdomain does not appear essential for replication.

Domain II is involved in the inhibition of protein kinase PKR. These PKR are activated by IFN in response to the presence of dsRNA in the cell [153]. This domain is also important for the replication of HCV RNA [154]. In addition, NS5A blocks the signaling pathway of IFN by inhibiting the phosphorylation of STAT1 [155]. This signaling pathway of IFN is described in later sections.

Domain III is not conserved among genotypes and thus, not necessary for replication [156] but is important for assembly and production of infectious viral particles [157]. On the other hand, deletion of this domain induces a relocation of the NS5A protein that leads to its dissociation from the core protein [157]. However, this domain has been particularly used to insert reporter genes, like GFP, in the construction of subgenomic replicons [156, 63].

Current data indicate that NS5A might function as a molecular switch between replication and assembly. The state of phosphorylation of this protein affects the replication of HCV RNA. However, these phosphorylated adaptive mutations of NS5A for facilitating the replication have an inverse correlation with the virus production [156, 158, 159]. It has been identified that residue Ser_{457} of NS5A (phosphorylated by cellular kinase, casein kinase II - CKII) regulates the production of new infectious particles [159].

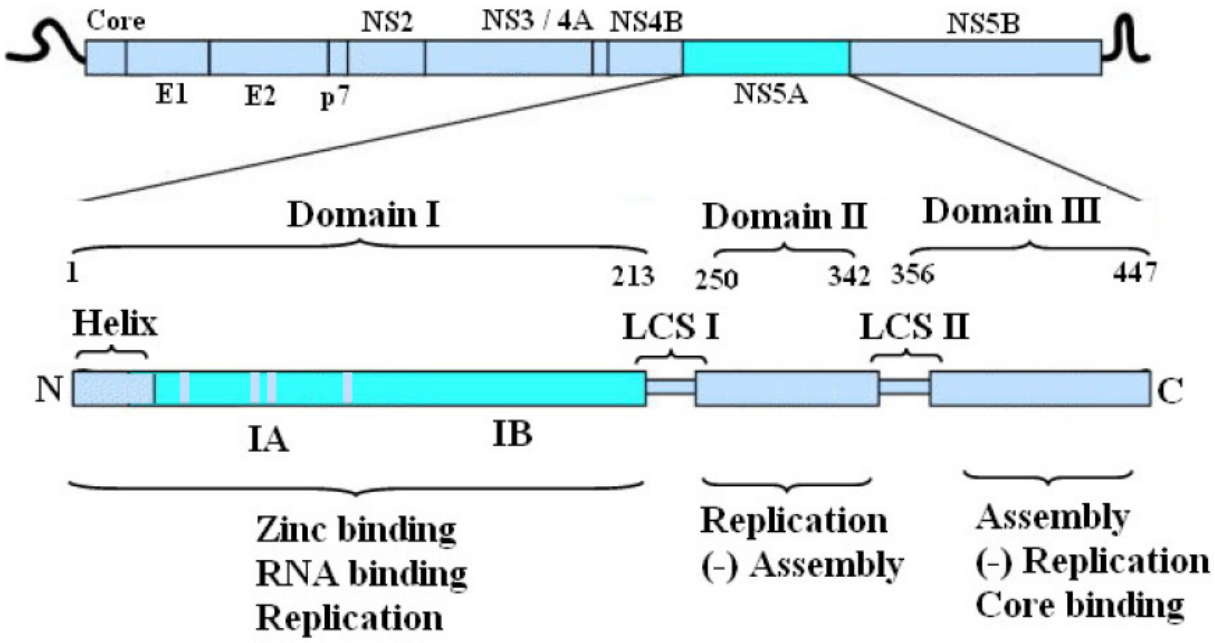


Figure 12: Schematic representation of NS5A protein (based on [159]).

The four cysteine residues corresponding to the coordination site of zinc ion are indicated by gray bars in the domain I of the protein.

Phosphorylation of NS5A regulates its interaction with the human protein VAP-A (vesicle-associated membrane protein-associated protein A), a protein required for replication [158]. On the other hand, its hyperphosphorylation blocks the interaction with VAP-A and therefore limits viral replication. VAP-A is found in the ER membranes and the Golgi apparatus and is involved in mobility of intracellular vesicles. But it is still poorly understood why this protein is diverted by NS5A at certain stages of the viral cycle.

NS5A also interacts with a geranylgeranylated cellular protein FBL-2. The inhibition of FBL-2 by siRNA reduced considerably the replication of viral RNA [160, 161]. Thus, NS5A interacts with proteins involved in many cellular signaling pathways and favours the entry of virus by modulating these signalling proteins [162].

6.c.9: NS5B

Replication of HCV RNA involves synthesis of complementary negative-strand RNA molecule from the genomic template strand. This is followed by synthesis of new genomic positive-strand RNA copies from the negative-strand intermediate RNA template. The protein responsible for this in HCV is NS5B (amino acids 2421-3011). It contains motifs that are

common amongst all RNA-dependent RNA ploymerases and was confirmed as the HCV RNA dependent RNA polymerase (RdRp) [163-165]. Moreover, its crystal structure revealed classical 'fingers', 'palm' and 'thumb' sub-domains [166-168]. The palm represents the catalytic site while thumb and fingers form tunnels. Through one of these tunnels, single-stranded RNA travese directly to the catalytic site while another positively charged tunnel allows the entry of nucleotides in the active site. These nucleotides are required for elongation of viral RNA [169]. It is commonly accepted that the binding of the RNA template and initiation of the synthesis are regulated by a very flexible beta stem-loop which is located in the thumb subdomain of NS5B [170].

Furthermore, NS5B is an integral membrane protein with a cytosolic orientation. NS5B is associated exclusively with with ER membrane via a highly hydrophobic 21 aa transmembrane C-terminus [171, 172] (Figure 6). This association has been shown essential for HCV replication [63]. NS5B can copy full-length HCV RNA, in the absence of a primer, presumably by using the 3' ends of both positive and negative strand HCV RNA as templates [173]. However, the activity of NS5B is modulated by interactions with viral proteins NS5A and NS3 [112]. These modulations in NS5B activity constitute a prime target for the development of new antiviral molecules against HCV [174, 175].

6.d: 3' UTR

The 3'UTR is a tripartite structure consisting of: (1) an upstream variable part with VSL1 and VSL2 stem-loops (28-42 nt), (2) a poly (U/UC) tract, and (3) a highly conserved, 98 nucleotide long 3' X-tail. This 3'X-tail putatively forms three stem-loop structures, known as SL1, SL2 and SL3, that play an important role in the initiation of (-) RNA strand synthesis during replication (Figure 13) [11, 12, 176]. Moreover, this X-tail is highly conserved even between the most divergent HCV isolates [66, 176]. On the other hand, despite the significant heterogeneity in the poly(U)/polypyrimidine tract between isolates, all isolates carry this tract [66]. The presence of 3'X-tail and poly(U)/polypyrimidine tract in all isolates, suggest their important roles in HCV replication.

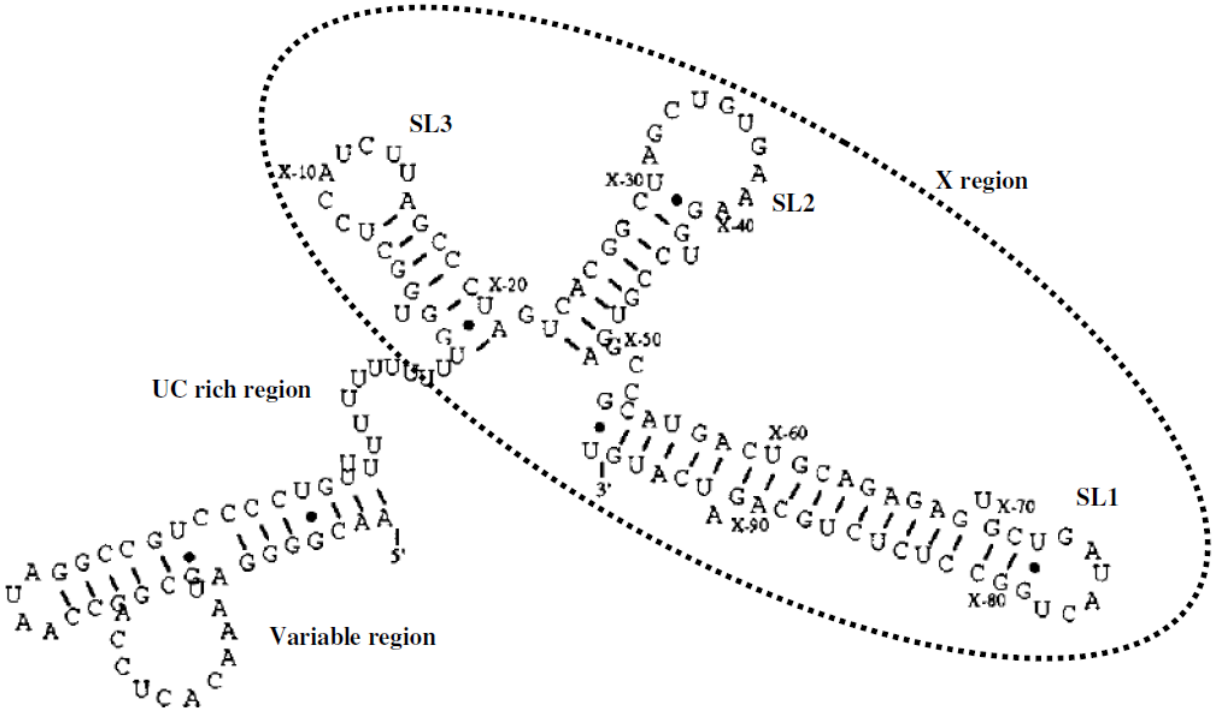


Figure 13: Secondary structure of 3'UTR (based on [177])

It has been suggested that 3'UTR is important for both replication and translation. *In vivo* studies showed that the deletion of the X-tail (*or any of its putative stem-loops*) or any large segment of the 3'UTR (*except for the variable region*) completely destroyed the replicative ability of infectious HCV RNA [178-181]. On the other hand, deletion of the complete variable sequence only reduces replication efficiency [66].

3'UTR influences translation of the HCV polyprotein, as X-tail specifically enhances IRES-dependent translation of HCV viral RNA [182]. This was confirmed by the fact that, it stimulates translation more strongly in hepatoma cell line (Huh-7 and HepG2) as compared to non-hepatoma one (BHk and HeLa) [183].

3'UTR also interacts with a number of host-cell proteins, for example polypyrimidine tract-binding protein (PTB) binds to the poly(U)/polypyrimidine tract and the 5'-end of the X-tail of the 3'UTR [184-186]. Since, PTB can bind to both 5'UTR [74] and 3'UTR and is also capable of dimerisation, PTB is likely required to bring the 5' and 3' UTRs together during replication. Other cellular proteins which bind to 3'UTR are La, heterogenous nuclear ribonucleoprotein C (hnRNPC), glyceraldehye-3-phosphate dehydrogenase (GAPDH) and ribosomal proteins L22, L3 and S3, as well as the mitochondrial homologue of L3 (mL3) [184, 187, 76, 188].

Finally, both NS5B and the helicase domain of NS3 have been shown to specifically interact with the 3'UTR [123, 189]. These two proteins have already been shown responsible for the replication of viral RNA. Their interactions with 3'UTR suggest a formation of a major complex of proteins around the 3'UTR that contributes to the viral RNA replication process.

7: Model Systems to study HCV

Characterization of the HCV viral cycle proved to be difficult for a long time due to the absence of cell model and robust small animal model for the production of infectious viral particles. However, many models to study certain steps of the viral cycle (entry, replication, translation) have been developed over the past decade. In 2005, a cell culture model of HCV (HCVcc) was introduced which opened new perspectives for the studies of HCV viral life cycle. This system was based on the use of the special viral genome JFH-1 (JFH-1 for Japanese Fulminant hepatitis 1), which was capable of producing infectious particles in cell culture [190]. In the few following paragraphs we will describe the various model systems used to study HCV.

7.a: Comparative Studies with HCV-related Viruses

A big load of the HCV current knowledge has been derived from comparative studies with HCV-related viruses like flaviviruses and pestiviruses, as these viruses share similarities at the level of genomic organization and amino acid sequence [46]. The pestivirus BVDV was used as a surrogate model for the evaluation of antiviral agents due to the availability of infectious clones and efficient cell culture systems [191-193]. Another virus used as a surrogate model is GB virus-B, as it is closely related to HCV and as infection results in similar pathological features compared to those of HCV infection [194, 195]. GB virus-B is a hepatotropic virus of unknown natural host that causes acute and sometimes chronic hepatitis in tamarins and marmosets. Furthermore, similarties (~41%) in amino acid sequencing of Domain 2 of GB virus-B core protein with domain 2 of HCV core protein [196], suggested that these two domains might perform similar functions. Also, the putative envelope proteins of GB virus-B and HCV share common structural features. Additionally, significant homology was observed among the NS3 serine protease, the NS3 RNA helicase and the NS5B RNA dependent RNA polymerase regions [195].

7.b: Animal Models

7.b.1: The chimpanzee

For many years, the only animal which was used for HCV infection was the chimpanzee (Pan troglodyte) [33], because of its genomic similarities (more than 98.5%) with humans [197]. This model was used to characterize the physicochemical properties of the then unknown agent [198]. Later, HCV was cloned from plasma collected from a persistently infected chimpanzee [1]. Furthermore, RNA transcripts which were produced from full-length cDNA clones, were found infectious when injected into the liver of chimpanzees [199]. The infection followed a progression similar to that observed in humans, viz., (*i*) the presence of viral RNA in blood could be detected a few days after infection, (*ii*) acute hepatitis characterized by increased serum alanine transferase activity (ALT) after two to twenty weeks of infection and (*iii*) development of an adapted immune response. Similarly like in humans, chronic infection in chimpanzees is associated with histological lesions of chronic hepatitis and increased risk of hepatocellular carcinoma [200]. However, the use of the chimpanzee model is limited because of the differences with human infection (*lower severity of chronic liver disease and reduced immune response as compared to HCV-infected humans*) [201], ethical problems, very high cost of these animals and their maintenance and lastly their availability. The chimpanzee is a protected species now and its use has been prohibited in most countries.

These limitations of the chimpanzee model have stimulated progress towards developing alternative animal models for HCV research.

7.b.2: Human liver cell transplanted uPA/scid mice

To overcome the lack of suitable animal model, transgenic mice with humanized livers have been produced (Figure 14) [202, 203]. These mice have been produced by grafting human hepatocytes and can be infected with HCV [204]. These mice, uPA-SCID, exhibit a severe combined immunodeficiency (SCID) and carries a transgene plasminogen activator (Alb-uPA) that leads to the destruction of their hepatocytes. These destroyed mice hepatocytes are then recolonized by human hepatocytes. These mice can be infected with HCV-positive human serum and support HCV replication within the human portion of their livers at clinically relevant titres. Moreover, HCV can be serially passaged from infected mouse to naïve transgenic mouse [202], due to its long-term persistence ability. Trangenic mice carrying HCV proteins have been used in many studies to investigate the pathogenic properties of these proteins [73, 205, 206].

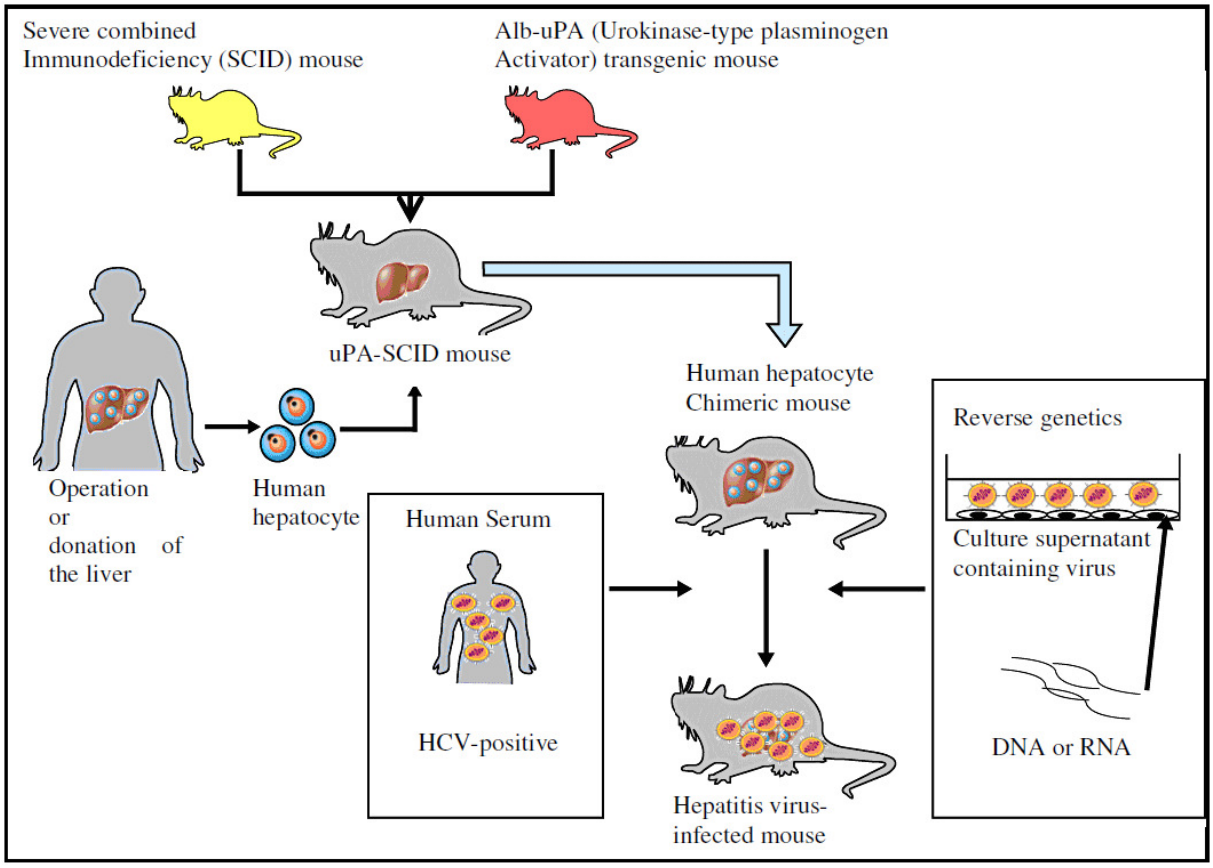


Figure 14: Generation of human hepatocyte chimeric mice and hepatitis virus infection model. A uPA/SCID mouse was created by mating uPA transgenic mouse and SCID mouse. Human hepatocytes obtained by surgical resection or donation were transplanted to newborn mice. The chimeric mice can be infected with hepatitis C virus (HCV) by injecting human serum containing viruses. Alternatively, the mice can also be infected by HCV [207] created in cell culture or by injecting HCV RNA into the mouse liver [208].

Moreover, this mouse model has the advantage of being cheap, easy to maintain and fast reproduction. Even though, the immune response and virus-host interactions are limited by the genetic background of these animals; they have provided the first really useful model for acute and chronic infections of human hepatitis virus.

7.b.3: Other models

Infections with HCV and production of anti-HCV antibodies were also obtained from other primates, such as marmosets [209], tamarind [210] and tree shrews (*Tupaia*) [211]. A study has shown that primary *Tupaia* hepatocytes can be infected *in vitro* by HCV-infected human sera and infectious virus particles can be produced [212]. Efficiency of infection was increased by

whole-body irradiation, suggesting that immunosuppression may facilitate infection [211]. However, the feasibility of using these models to study HCV has yet to be fully investigated.

7.c: In vitro models

7.c.1: Soluble glycoproteins

Biochemical functional analysis of envelope glycoproteins E1 and E2 of HCV were perfomed. However, these studies were limited by the fact that glycoproteins E1 and E2 tend to aggregate and to fold incorrectly when they are expressed *in vitro*. This is due to their high retention in the ER and the presence of heterodimerization sequences in their transmembrane domains [79]. To overcome the problem of aggregation, soluble forms of glycoproteins were produced by deleting transmembrane domains [213] and it was observed that soluble glycoprotein E2 (SE2) binds specifically to hepatocyte cells [214]. SE2 protein was also used to identify cellular factors involved in HCV entry, such as CD81 [215], the scavenger receptor BI (SR-BI) [87] and heparan sulphate [216].

7.c.2: Virus-like particles

Virus-like particles (VLPs) were produced in insect cells. These cells were infected with recombinant baculovirus containing the cDNA of HCV which encodes the structural proteins core, E1 and E2 [217]. Insect cells have activities like acylation and glycosylation of proteins, similar to mammalian cells. Additionally, baculovirus can synthesize many heterologous proteins in large quantities, thus making this system more significant. These proteins self-assemble into VLPs. However, VLPs can not replicate as they do not contain the whole genome. The VLPs were released into the extracellular environment and it was observed that purified-produced particles from lysates of insect cells presented biophysical and morphological characteristics similar to those isolated from serum of HCV-infected patient [218]. This suggested that viral glycoproteins were properly complied in VLPs [219]. Furthermore, these VLPs have antigenic properties similar to those of wild virus, when tested for binding or neutralization, thus making them a potential candidate for vaccine [218, 220]. However, the usefulness of this tool is limited to studying the mechanisms of attachment and virus-host interaction at the receptor level. The absence of marker genomic particles makes it inappropriate for the study of virus entry.

7.c.3: HCV Pseudo-particles (HCVpp)

In an attempt to study the process of HCV cell entry, infectious, genetically tagged HCV pseudo-particles were developed [88]. These particles are chimeric viruses consisting of envelope glycoproteins E1 and E2 of HCV, a virus capsid from murine leukemia virus (MLV) or human immunodeficiency virus (HIV) and a reporter gene encoding GFP or luciferase [221, 222]. These were obtained by co-transfection of abovementioned three vectors in human 293T cells (Figure 15).

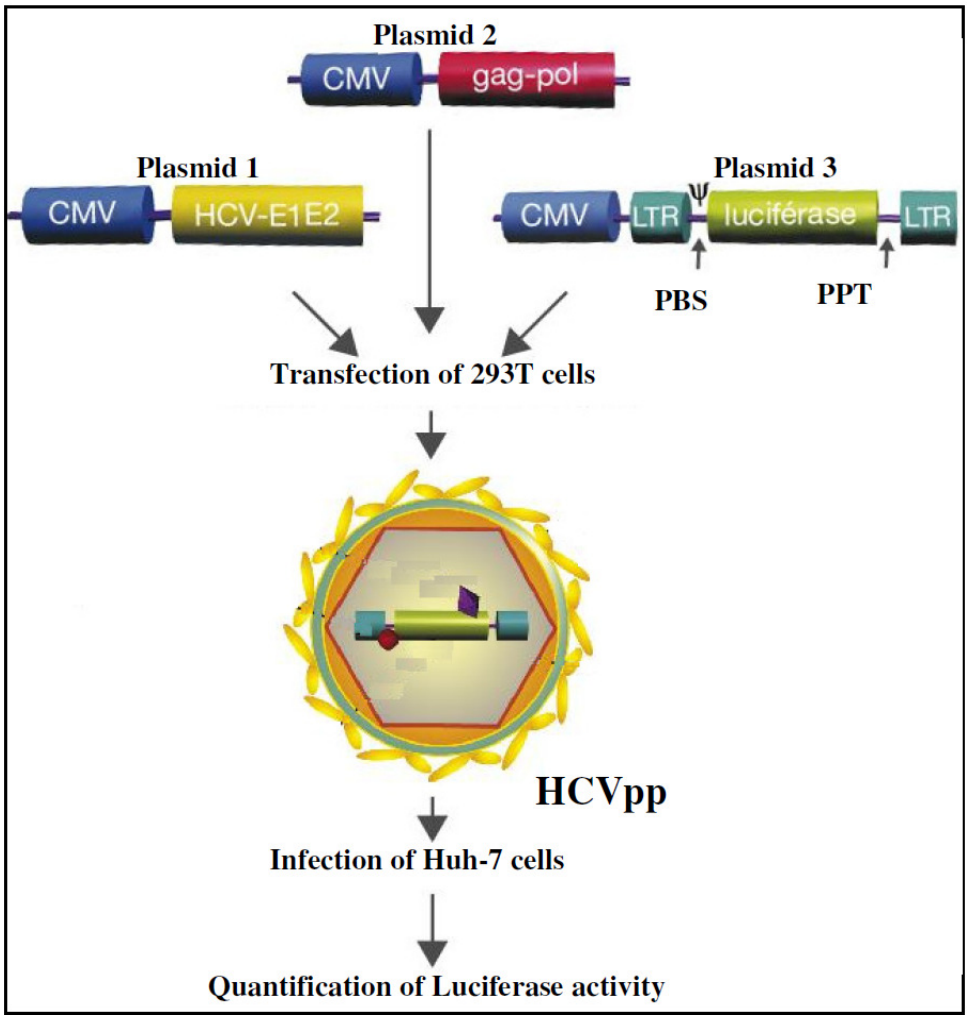


Figure 15: Schematic representation of the production of HCVpp. For the production of HCVpp, the human embryonic kidney cells 293T are transfected with three expression vectors. The first codes glycoproteins E1 and E2 of HCV which are responsible for cell tropism. The second vector expresses the protein encoded by the gag and pol genes of murine leukemia virus (matrix, capsid, nucleocapsid, protease, reverse transcriptase integrase). Matrix proteins, capsid and nucleocapsid allow encapsidation of the RNA with the encapsidation signal (ψ), the particle assembly and budding at the plasma membrane. Finally,

the third vector encodes the RNA that will be packaged into particles. This RNA contains retroviral sequences necessary for reverse transcription and integration of proviral DNA in genomic DNA of infected cells (LTR, long terminal repeat, PBS, primer binding site; PPT, polypurine track) and also contains a reporter gene encoding luciferase. The HCVpp are secreted into the culture supernatant of 293T cells and used to infect cells of human hepatocellular carcinoma, Huh-7. Infected cells will integrate the gene encoding luciferase in their genomic DNA by reverse transcriptase and integrase, and therefore will express luciferase. The infectivity of HCVpp could then be assessed by quantification of luciferase activity in Huh-7 cells. (based on [223]).

These pseudo-particles were secreted into the cell culture medium and shown to infect naïve hepatic cells. It was also observed that they are specifically neutralized by anti-E2 monoclonal antibodies as well as HCV positive sera. Moreover, the production of these particles was relatively efficient, with an average of 105 unit / ml of infectious supernatant [221]. Additionally, Huh7 hepatocyte cell line and human hepatocytes could be infected [221] by these pseudo particles. Infection of target cells can be quantified by measuring reporter gene expression in infected cells. The entry of these particles is neutralized by antibodies against E2 and sera of infected patients [221, 88, 224].

Thus, this model allows the study of virus entry into its target cells and the study of the interaction between HCV glycoproteins and cell surface receptors.

7.c.4: HCV Replicon System

A major breakthrough in HCV research came with the development of the HCV replicon system, which made study of viral replication possible. The replicon is a tool for studying the replication of viral RNA and some virus-host interactions. Replicons have been commonly used for validation of new antiviral drugs. This system is based on the stable autonomous replication of subgenomic and selectable HCV RNAs [106]. Replicons are bicistronic RNAs (*formed due to deletion of the core-NS2 region, and insertion of a gene encoding the selectable marker neomycin phosphotransferase (neo) and the EMCV (EncephaloMyoCarditis Virus) IRES*) containing autonomous replicating 5' HCV IRES and a resistant gene for an antibiotic (neomycin) [106]. These constructs, once transcribed into RNA, are capable of growing colonies containing autonomously replicating HCV genomic RNA with a selection for neomycin sulphate (G418) (Figure 16).

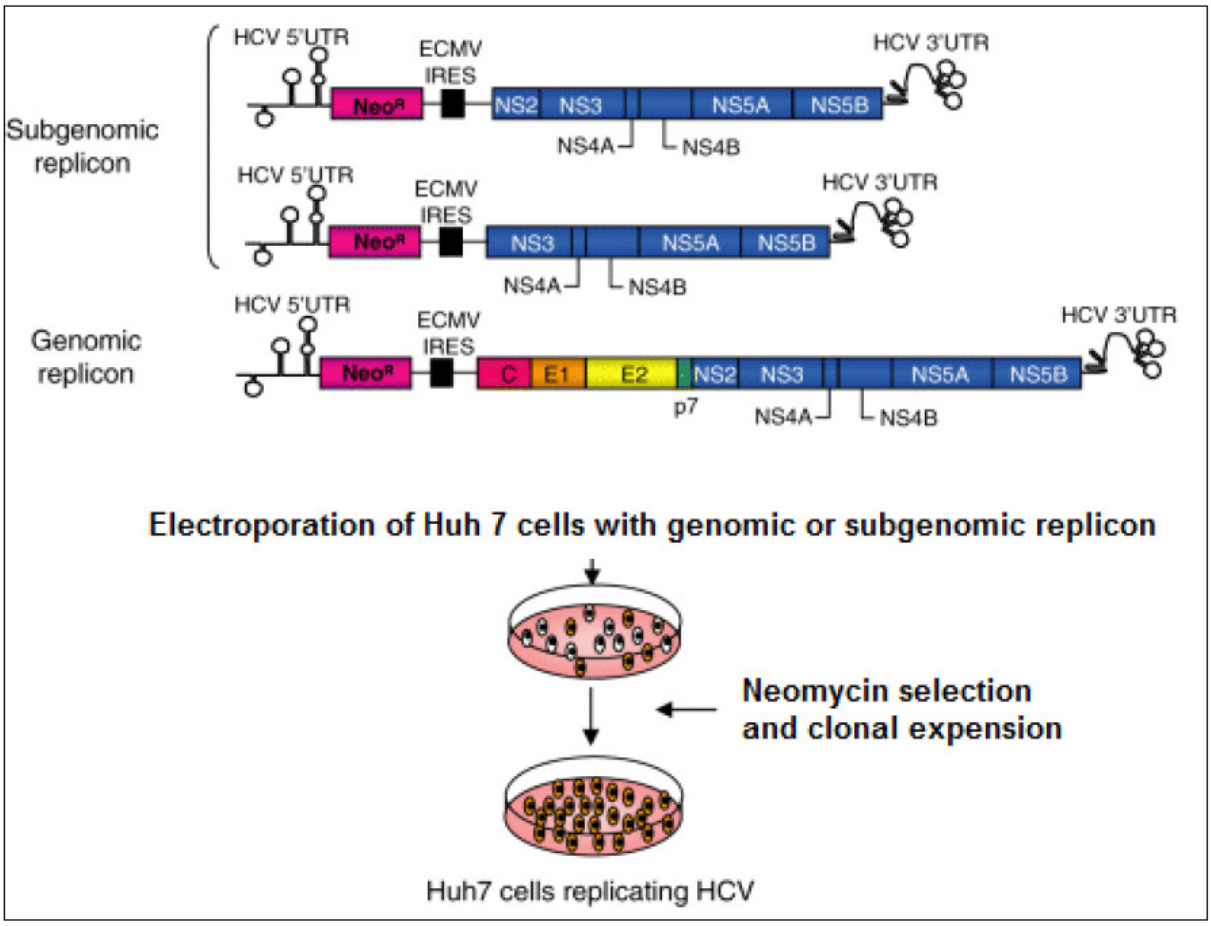


Figure 16: schematic representation of HCV replicon sytem (based on [225])

This approach allowed the selection of strongly virus replicating cell lines with levels of HCV RNA replication being ~10^5 fold higher than in any other *in vitro* system [226]. For example, one of these cell lines called Huh7.5 (*produced by treatment of those cells which can very effectively replicate the replicon for IFN-α*) exhibited a greater ability to replicate virus after re-transfection with viral RNA when compared to Huh7 cell line [227].

The cell culture of these subgenomic replicons led to the appearance of adaptive mutations that increase significantly the rate of replicating transfected cells [228-230]. Among them, several mutations were identified and, in particular, mutations of aa 2884 in NS5B led to ~500 fold increase in replication [230]. Additionally, a cluster of mutations was found in the central region of NS5A [228, 158] and two others identified in NS3 [229].

The replicon systems were further modified in 2002 when selectable full-length HCV genomes were generated in which the HCV structural proteins were efficiently expressed. Unfortunately, no infectious viral particles were produced from this system, suggesting that Huh-7 cells may lack some essential factors required for production of HCV particles [18].

Nevertheless, this theory related to Huh-7 cells laking some essential factors, was dispelled after the production of JFH-1 infectious particles from these cells [190].

7.c.5: JFH-1 Infectious Clone

In 2001, a particular clone of HCV was isolated from a Japanese patient suffering from fulminant hepatitis C [231]. This clone, called JFH-1 (Japanese fulminant hepatitis) was fully infectious in cell culture [190]. Since then, a number of reports have confirmed its infectivity in chimpanzees and humanized mice [232, 190, 233]. Initially, the aim was to construct a genotype 2a replicon construct, as this had not been tried previously. Upon analysis of this replicon it was noted that efficient replication occurred without the need for cell culture-adaptive mutations [234] and produced virions were able to infect naïve Huh-7 cells (Figure 17).

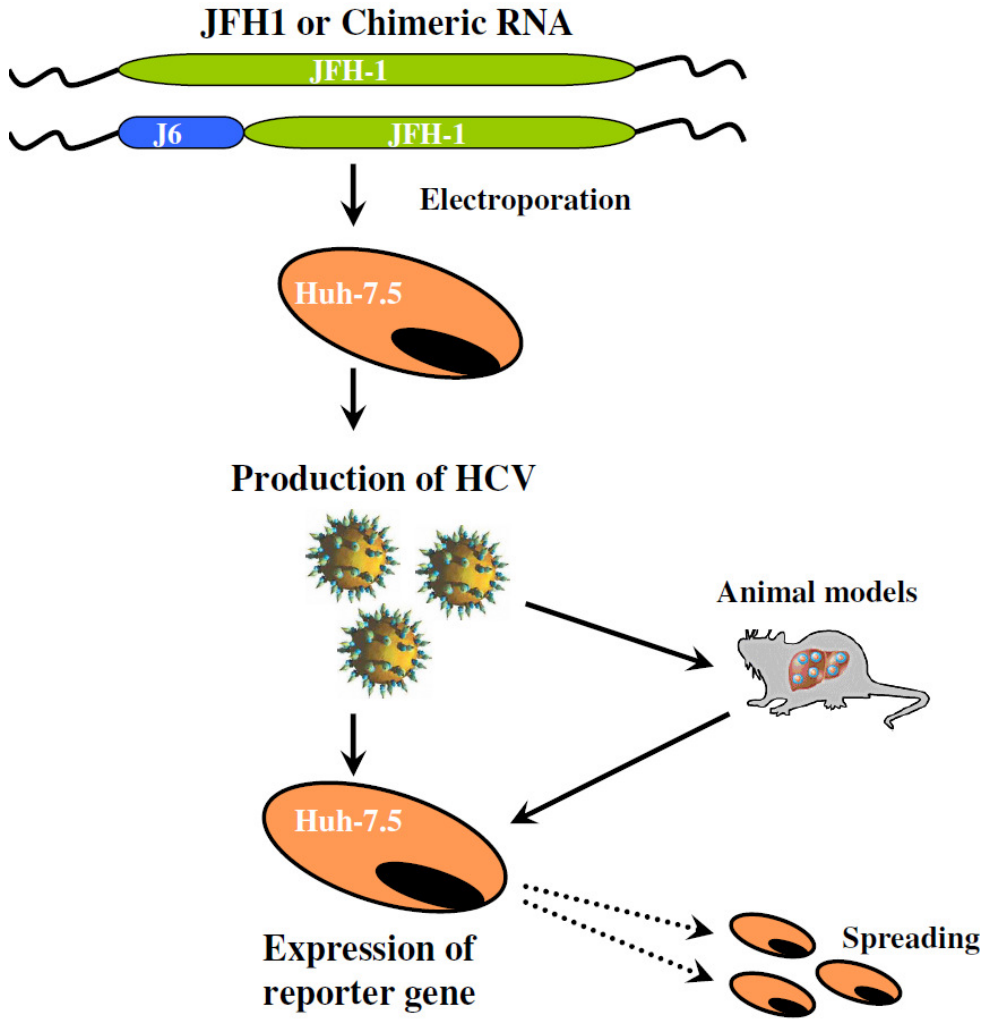


Figure 17: Schematic representation of HCVcc system (Based on [225]).

This system was further improved by the construction of Intra-genotypic [232, 108] or inter-genotypic [108] chimeric viruses that resulted in increased production of HCV infectious particles. In addition, to facilitate the detection of infected cells, viruses expressing a reporter gene (luciferase) were also produced [235].

The development of this model made possible to study HCV and its complete life cycle in cell culture. This system confirmed the validity of previous work on the role of envelope glycoproteins E1 and E2 in viral entry [190], the preferential tropism for liver cells [232, 190, 233] and the role of receptors involved in HCV entry [235, 232, 190, 236, 233]. This system should also enable the study of other steps of the viral cycle like assembly and secretion of infectious viral particles.

8: Viral Life cycle

The study of the viral life cycle of HCV has been hampered until 2005 by the absence of cellular system for the production of infectious viral particles. Nevertheless, with the models described above (HCVcc system and recombinant HCV proteins) partial answers were obtained. Based on these partial answers, a general pattern of HCV viral cycle (Figure 18) has been establised.

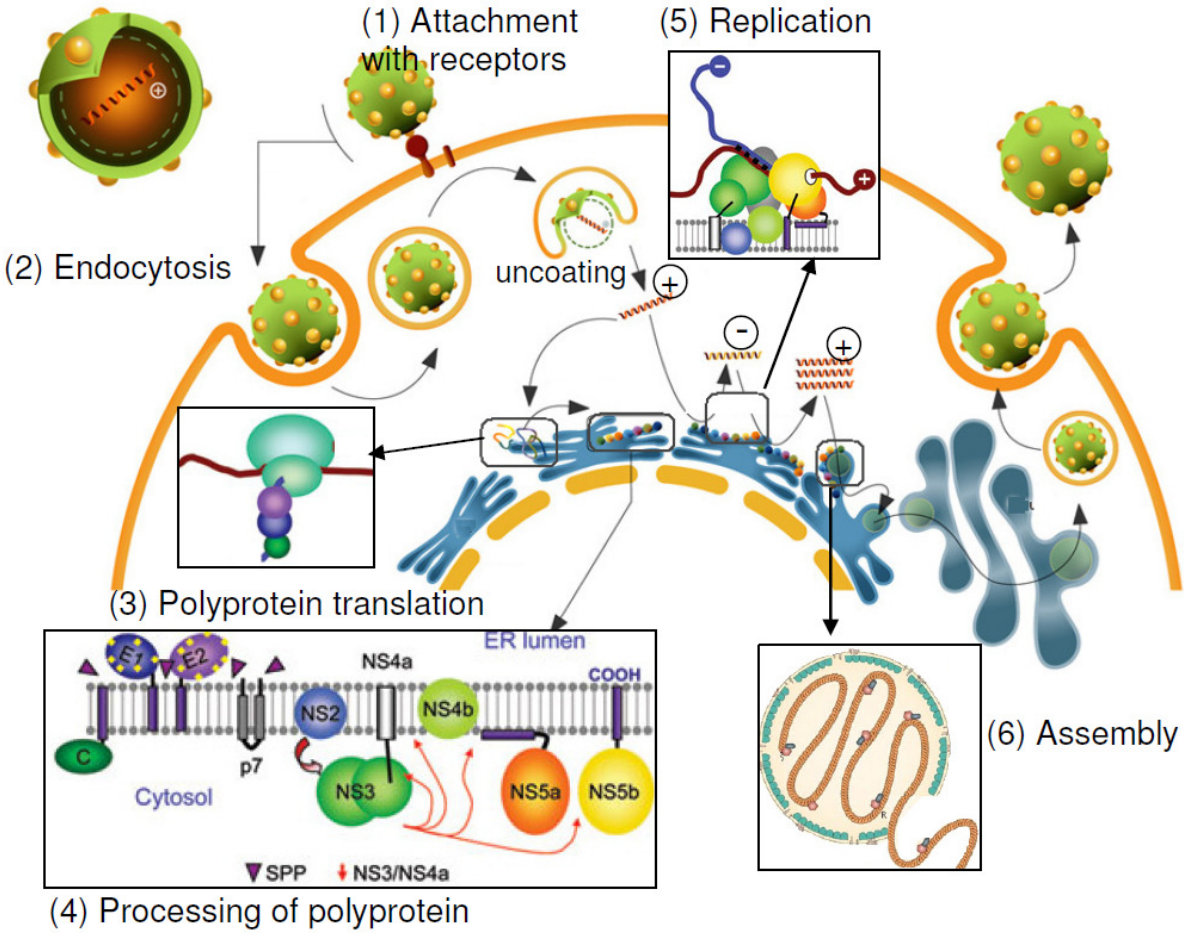


Figure 18: Viral life cycle of HCV (based on [237])

(1) Virions bind to the cell surface via receptors. (2) The particles are then endocytosed by a clathrin-dependent pathway. Next, uncoating allows the release of viral RNA into the cytoplasm. (3) The genome is translated into polyprotein and (4) this polyprotein is processed into different viral proteins that form the RC in the ER. (5) Replication of HCV RNA leads to the synthesis of new molecules of (+)RNA. (6) The virions are then assembled and secreted.

8.a: Attachment and Entry

To initiate its life cycle, a virus must bind to the host cell. Binding occurs via a specific interaction between host cell receptors and an attachment protein on the surface of the virus. This step involves a cascade of events involving the virus via the E1 and E2 glycoproteins, and many other cellular co-factors. As previously stated, the viral particles are present in the external medium in form of LVPs (lipo-viro-particles). Initially the glycosaminoglycans (GAGs) and the LDL receptor (LDL-R) on the surface of the cell will begin the process of anchoring LVPS via lipoproteins. Then, glycoproteins E1 and E2 will recognize the HCV-

specific cellular factors, like the tetraspanin CD81[215], the scavenger receptor BI (SR-BI) [87], junction proteins claudin1, 6 and 9 (CLDN) [238-240] and occludin (OCLN) [241, 242].

In the next sections, we will emphasize on the role of above mentioned HCV-specific cellular cofactors.

8.a.1: Molecules attached on cell surface

8.a.1.1: Glycosaminoglycans (GAGs)

GAGs are linear polysaccharides expressed on the cell surface. These are characterized by strong structural heterogeneity and are binding sites for many viruses. They are involved in the initial interaction of viruses with the cell surface before binding to HCV-specific receptors. Among many types, the sulfated GAGs like heparan sulfate appear highly involved in the binding of many viruses, including HCV. The use of heparin, an analogue of heparin sulfate or heparinase - an enzyme degrading heparin sulfate, inhibits the attachment of HCV and leads to reduced infectivity of the virus [243]. These results showed that the attachment of the virus to heparan sulfate plays an essential role in the early steps of HCV infection. However, the exact role of GAGs in HCV entry remains unclear. A soluble form of E2 (SE2) binds to heparin with high affinity, but studies using HCVpp, failed to confirm these results [244]. These observations suggest that either the binding site for heparin sulfate is not available on HCVpp E1E2 heterodimer or attachment of HCV is mediated by viral particle associated lipoproteins. This discrepancy of SE2 binding with heparin was later solved. It was suggested that lipoprotein lipase (LPL) participates in the interaction between HCV and GAGs [245] by forming a bridge between the virus via associated lipoprotein and heparan sulfates on the surface of cells [245].

8.a.1.2: The LDL receptor

In the serum of infected patients, HCV viral particles in the form of LVP are often associated with LDL or VLDL [53]. Thus, the LDL receptor has also been suggested as a putative HCV receptor [246]. Analysis of HCV particles with heterogenous densities led to the identification of an association between HCV and low density β-lipoprotein [247]. This suggested that β-lipoprotein associated viral particles may enter into host cells via the LDL receptor.

This notion was well supported by anti-LDL receptor antibodies that inhibit HCV infection in a dose-dependent manner [246]. HCV infection was also inhibited by biochemical inhibitors of LDL endocytosis, like heparin sulphate and EGTA [246]. Additionally, LDL and HCV particles

have been shown to co-localise on the surface of Molt-4 cells. However, use of anti-LDL receptor antibodies did not completely block HCV binding to cell surface [248], suggesting that other receptors may be involved in the process.

8.a.1.3: Other molecules

Other molecules thought to bind HCV particles include the asialoglycoprotein receptor (ASGP-R) [249] and the liver (L)- and dendritic cell (DC)-specific intercellular adhesion molecule-3 (ICAM-3)- grabbing nonintegrins (L-SIGN and DC-SIGN) [250-252]. Anti-ASGP-R antibody was shown to inhibit binding of LVPs of Molt-4 and HepG2-cells, while binding to 3T3 cells was only possible in the presence of transfected AAGP-R [249]. L-SIGN and DC-SIGN are type II integral membrane proteins with short N-terminal cytoplasmic domains and a C-terminal calcium dependent lectin domain [250]. L-SIGN is expressed in the liver, although not on hepatocytes, [252] while DC-SIGN is found on dendritic cells. L-SIGN and DC-SIGN are known to bind HIV gp120 with high affinity and further transfer virus particles to adjoining CD4-positive cells [253]. These molecules have also been shown to bind E2, LVPs and infectious HCV from human serum [250-252]. On the other hand, co-culture of cells expressing DC-SIGN/L-SIGN (*pre-incubated with HCVpp*), with Huh-7 cells permits transmission of infectious HCVpp in Huh-7 cells [254, 255]. L-SIGN is expressed on liver sinusoidal endothelial cells which form vessels that separate hepatic blood from hepatocytes. This gives an indication that these cells could be involved in trans-infection by transferring virus from the blood to hepatocytes in the liver [255, 251].

8.a.2: Receptors

8.a.2.1: CD81

The first molecule to be suggested as HCV receptor was CD81 [215]. It belongs to the family of tetraspanins and is a 25 kDa cell-surface molecule, involved in many processes including cell adhesion and signal transduction. It has 4 transmembrane domains, giving rise to 2 extracellular loops, SEL and LEL (small and large extracellular loop), a small intracellular loop and N- and C-termini (Figure 19). The large loop, LEL, contains cysteine-cysteine-glycine patterns which are important for the formation of disulfide bridges.

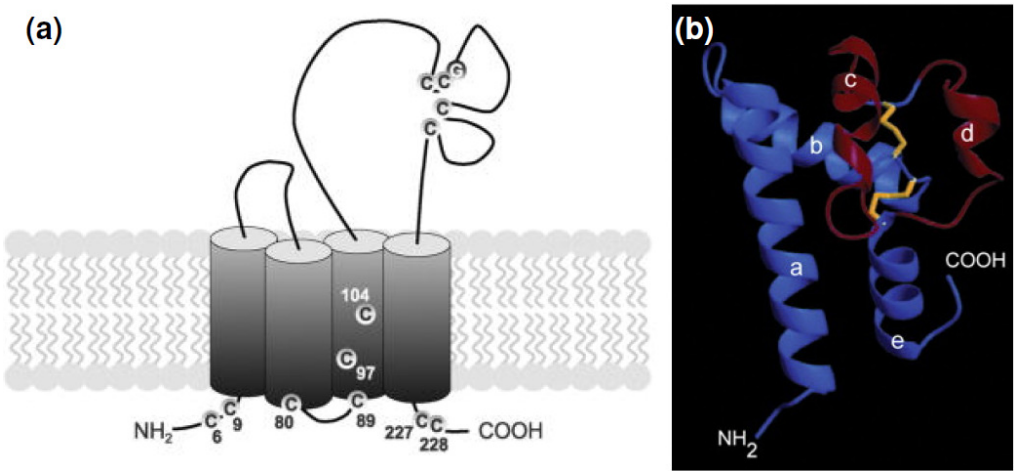


Figure 19: Schematic structure of CD81 (a) CD81 is composed of four transmembrane domains and two extracellular loops designated the small extracellular loop (SEL) and the large extracellular loop (LEL). Conserved cysteines are highlighted in yellow. The conserved CCG motif, which forms disulphide bridges (yellow lines) with additional cysteines, is shown. Conserved cysteines in intracellular domains are potential palmitoylation sites. (b) Three-dimensional representation of the CD81 LEL, which is composed of two subdomains: a conserved three-helix subdomain (shown in blue), in which a variable double-helix subdomain (shown in red) is inserted. (Based on [256])

Initially, in the absence of systems producing infectious HCV particles, the first search of HCV receptors began with the use of soluble SE2 forms. A specific binding to the LEL of CD81 was observed which enabled the discovery of the CD81 receptor [215]. This was confirmed by the use of recombinant GST-fusion polypeptides encompassing the LEL of human CD81. When pre-incubated with HCVpp, these polypeptides specifically precipitated E1-E2 complexes and neutralized HCVpp infection in Huh-7 cells [88].

Additionally, HCVpp infection was also inhibited by monoclonal antibodies against CD81 as well as by knockdown of CD81 expression by siRNA [257]. Finally, involvement of CD81 in viral entry was confirmed in infectious cell culture systems. It was observed that infectivity was blocked by pre-incubation of culture systems with soluble recombinant CD81 LEL molecules [232]. Similarly, anti CD81 monoclonal antibodies were also able to neutralize infectious virus [190].

Later, inhibition of HCV JFH-1 infection at a post-binding stage by anti-CD81, suggested that CD81 is required after an initial binding step [235] in viral entry. However, susceptibility to HCV infection depends on a critical quantity of CD81 molecules on cell surface [258].

Interestingly, comparision of the LEL of human CD81 with that of African Green Monkey (AGM) (having no affinity for E2) showed that residues L_{162}, I_{182}, F_{184} and N_{186} are critical for E2 binding [259].

In summary, the appearance of new model for HCV studies like HCVpp [221] and HCVcc [232, 190, 233] and the use of CD81 neutralizing antibodies, confirmed the involvement of CD81 at a later stage of viral entry. However, the overexpression of CD81 in HCV nonpermissive cells does not make these cells permissive to HCV. Thus, CD81 plays only the role of a co-factor in viral entry and requires the cooperation of several other cofactors [31, 257].

8.a.2.2: SR-BI

The scavenger receptor class B type 1 (SR-B1) [87] was identified as a co-factor that selectively uptakes cholesterol from the hydrophobic cores of lipoproteins. It belongs to the CD36 superfamily which includes cell-surface membrane proteins that bind chemically modified lipoproteins [260]. SR-B1 is a high-density lipoprotein (HDL) receptor containing 2 transmembrane domains with short N- and C-termini located in the cytoplasm and a large extracellular loop [260] (Figure 20). It is highly expressed in the liver and steroidogenic tissues (*ovaries and adrenal glands*) [261, 262]. This protein plays a major role in controlling the plasma concentration of HDL. Due to this function it facilitates the selective entry of esterified cholesterol HDL in the cell and also regulates the cellular cholesterol efflux. SR-BI also plays a role in the metabolism of VLDL and acts as a receptor for these lipoproteins [263].

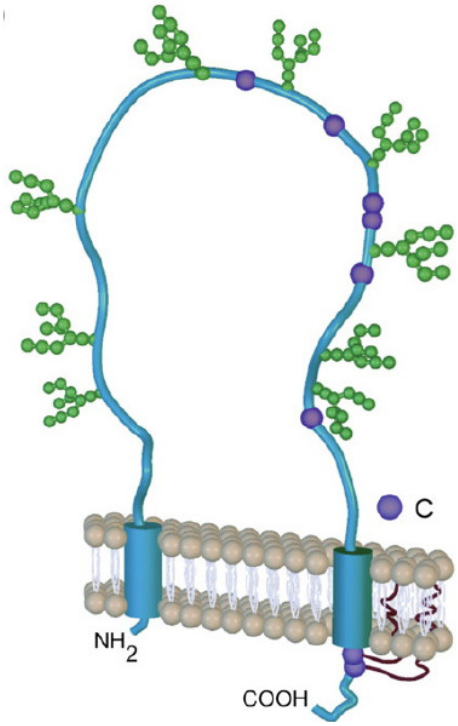


Figure 20: Structure of the protein SR-BI (based on [264]).

The extracellular loop contains nine potential glycosylation sites (green) and six cysteines (Purple). SR-BI is palmitoylated at two cysteines located at the C-terminus.

HepG2 cells which do not express CD81 but efficiently recognize recombinant E2, suggested the presence of an alternative attachment molecule. By reversible cross-linking with E2 and subsequent cell lysis, an 82 kDa, glycosylated molecule was isolated and identified as SR-B1 [87]. Additionally, the hypervariable region-1 of E2 is required for recognition of SR-B1 [265, 266]. The importance of SR-B1 in HCV entry was confirmed by following the use of small interfering RNAs (siRNA) inhibiting the expression of SR-B1. It was observed that they can block infection of cells by HCVpp and HCVcc [266, 223, 236]. On the other hand, the overexpression of the receptor SR-B1can increases infection by HCVpp and HCVcc [267].

It has been observed that HDLs enhance SR-B1 mediated HCV cell entry while oxidized LDL (oxLDL) inhibits infection [268, 269]. In addition, drugs inhibiting the transfer of HDL cholesteryl ester, have also been shown to reduce HCVpp entry [270, 268, 271]. This transfer of HDL cholesteryl ester is mediated by SR-B1. Finally, SR-B1 binds HDLs and lipopolysacharrides (LPS), allowing viral entry via intercellular compartments such as the Golgi complex or endocytotic compartment [272]. HCV may therefore utilize this entry pathway in order to avoid its degradation. Also, fusion of HCV glycoproteins is pH-

dependent, suggesting HCV may use SR-B1 to traffic itself to endosomal compartments for activating the fusion properties of the glycoproteins [221].

8.a.2.3: Claudins (CLDN)

Before discussing about claudins, we will discuss about tight juctions as claudins and occludins are their integral parts.

Tight junctions (TJs)

Tight junctions (TJs) are intercellular contacts that seal the space between the individual cells of an epithelial sheet so that they can collectively separate tissue compartments. The barrier is required to transport ions solute and water through the intercellular spaces and to limit paracellular entry of undesirables like toxins, antigens and microbes [273]. In hepatocytes, these junctions separate bile ducts (Figure 21). They are formed by at least 40 different proteins, including claudins (CLDN), occludin (OCLN), junctional adhesion molecules (JAMs), tricellulin and the scaffold proteins associated with anchoring and polarity. These scaffold proteins form an anchoring platform (*connectors including ZO-1, ZO-2, ZO-3, MAGI-1, PatJ, PALS1 and MUPP1*) for the actin cytoskeleton and proteins involved in polarity [274].

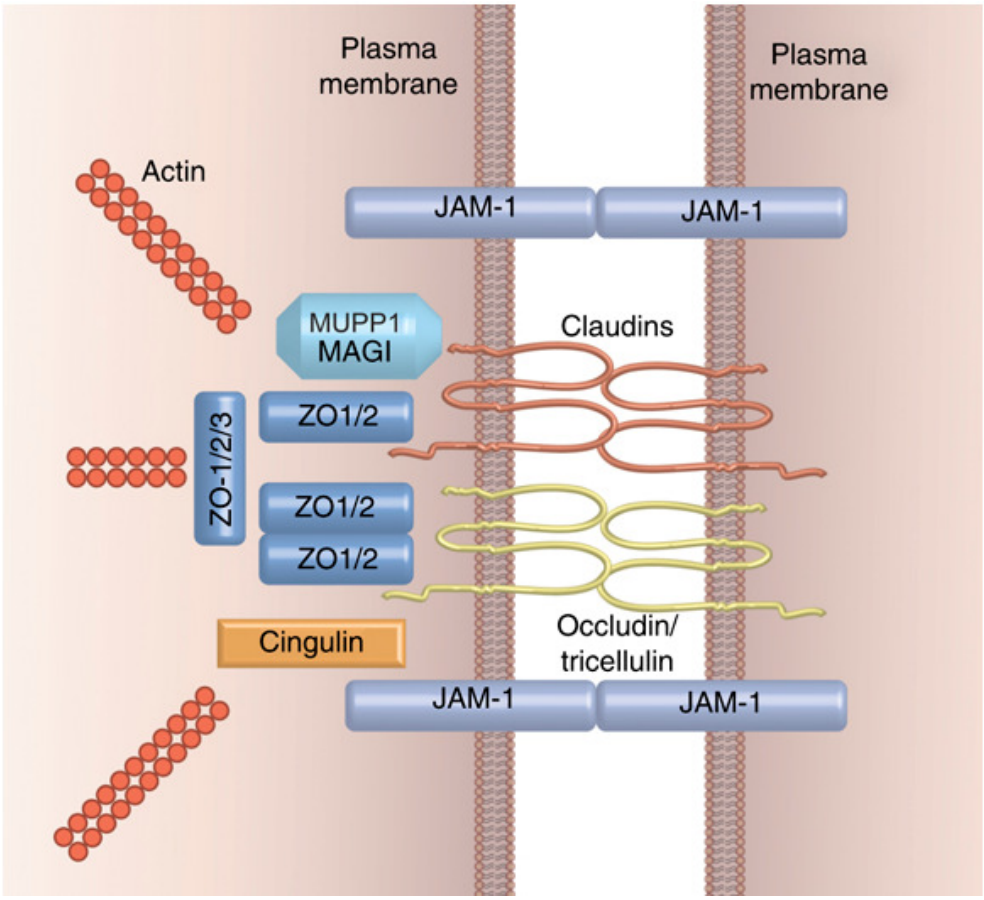


Figure 21: Schematic representation of the basic structural transmembrane components of tight junctions. ZO-1 or ZO-2 is important for clustering of claudins and occludin, resulting in the formation of tight junctional strands. The role of the other scaffolding proteins (ZO-3/MAGI/MUP1) is less clear. The ZOs and cingulin can provide a direct link to the actin cytoskeleton. (based on [273]).

Tight junction platform is important for signal transductions related to the assembly and polarization of epithelia as well as to the proliferation and cell differentiation. These junctions can quickly reshape itself in response to extra-intracellular stimuli, due to the presence of 24% of CLDN1, 70% of OCLN and ZO-1 tight junctions in a mobile fraction [275]. Finally, due to this mobility factor, TJs regulate paracellular permeability and maintain cell polarity in the epithelium and endothelium [276].

CLDN receptor for HCV

Although a number of receptors have been described, the limited HCV tropism could not be explained as some cell lines expressing all above factors, do not support HCV entry. To solve

this limiting step, a cyclic lentivirus based repackage screening of a cDNA library (*derived from the highly HCV-permissive hepatocarcinoma Huh-7.5 cell line, for genes that render the non-permissive CD81+ SR-BI+ 293T cell line infectable with HIV-1 particles pseudotyped with HCVgp (HCVpp)*) was performed. This screening allowed the identification of a new co-factor required for viral entry, the protein CLDN1 [238]. The tight junction-transmembrane protein CLDN1 is one of the 24 known claudins and belongs to the tetraspanins family. CLDN1 forms the backbone of tight junctions through homo- and heterotypic interactions and is highly expressed in liver as well as in other epithelial tissues [277], These proteins contain N- and C-termini extending into the cytoplasm, four transmembrane helices and two extracellular loops (EL1 and EL2) (Figure 22) [278, 276].

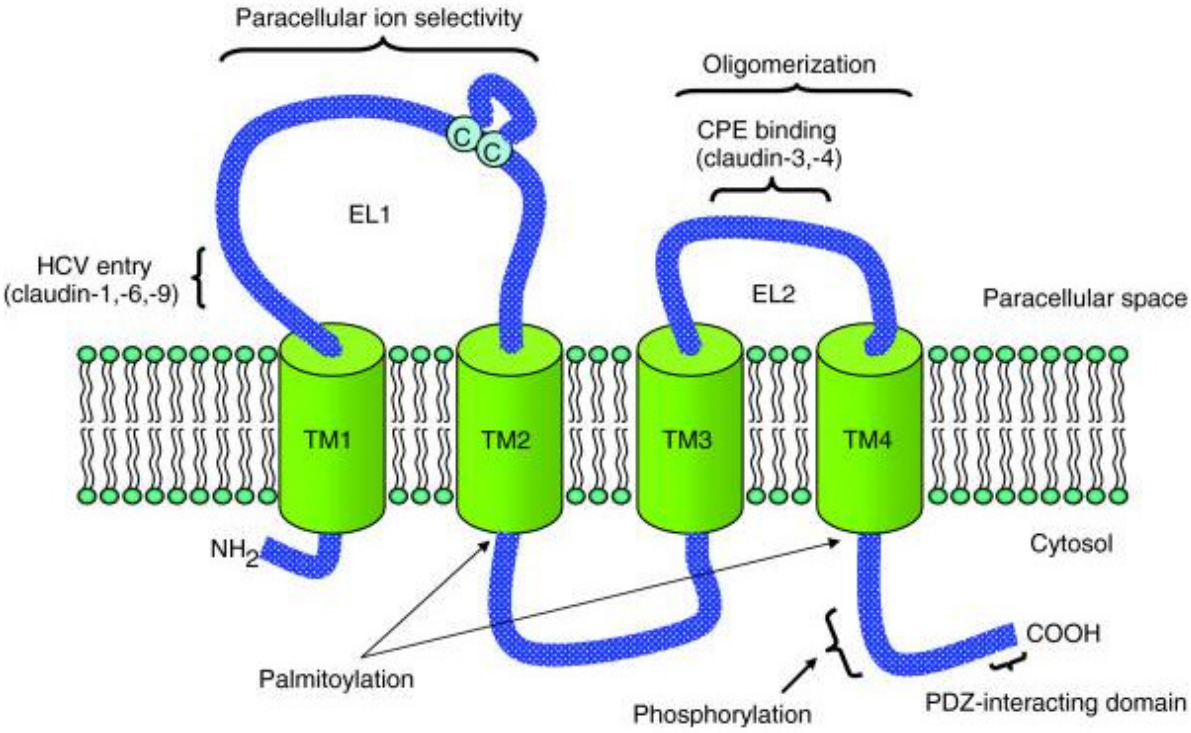


Figure 22: Structure of the junction protein CLDN1 (from [279]).

EL1 influences the selectivity of exchanged ions and its highly conserved clauding motif contains critical residues for entry of HCV (*drawn from six different subtypes isolates*) viz., W_{30}, GLW_{51}-C_{54}-C_{64}. The latter two cysteine residues form a disulfide bridge that stabilizes the protein [280, 281]. Furthermore, it was observed that these critical residues are responsible for *cis-trans* interactions of proteins. Cis-trans intraction is required for cell-cell adhesion [280].

Thus, the contact zones formed by CLDN1 are localized to a specialized membrane area, necessary for HCV entry. Additionally, it was also suggested that EL2 loop can form dimers with CLDN located on the opposite side (Figure 23).

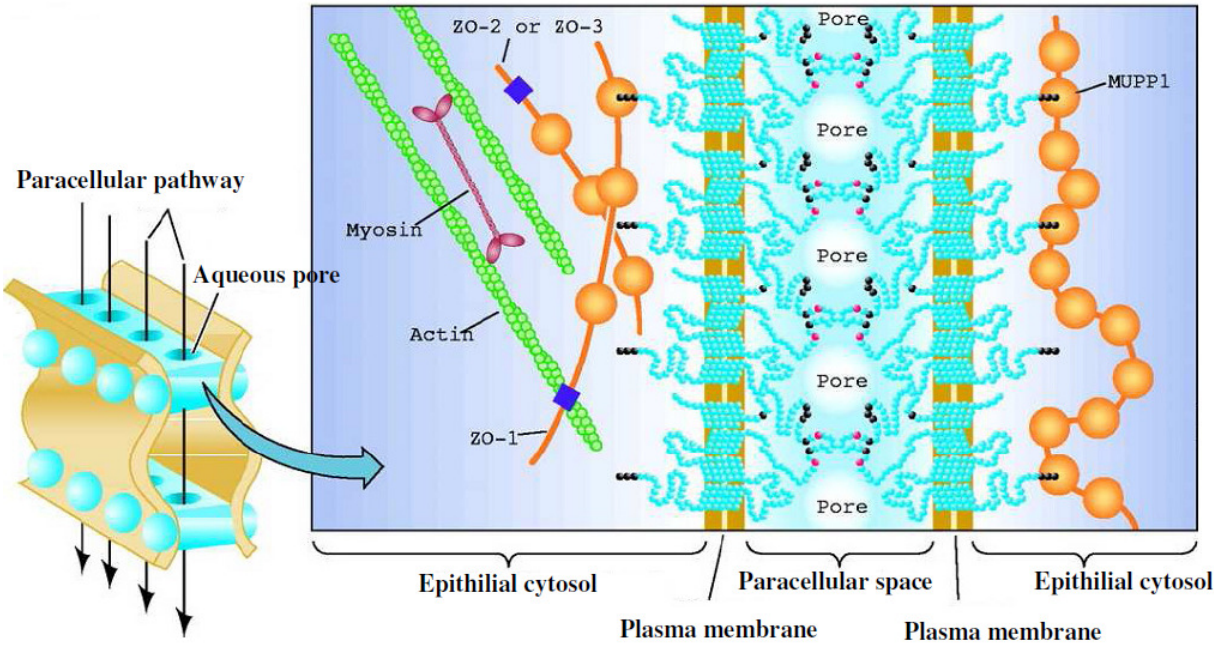


Figure 23: Model for assembly of tight junctions in CLDN (from [276]).

During the process of tight junction formation, two cells associate laterally via CLDN present on the surface of both cells to form a tight junction which allows the passage of selective ions (Figure 23). Interestingly, the carboxy-terminal tail is the target of various post-translational modifications (phosphorylation and palmitoylation). These modifications can significantly alter the localization and function of CLDN.

Overexpression of envelope glycoproteins E1E2 and CLDN1 molecules in HCV non-permissive cells (*293T, HeLa and GHOST*) [282] suggested an interaction of claudins with envelope glycoproteins..

Additionally, overexpression of CLDN1 in liver cells by siRNAs resulted in no change in infectivity by HCVpp and HCVcc. This was in contrast to SR-B1 overexpression. Inversely, a reduction in CLDN1 expression showed a decreased infection by the same viral particles [238, 267].

Further it was showed that antibodies directed against a FLAG TAG epitope were capable of inhibiting HCV infection in a dose-dependent manner [238]. The FLAG TAG epitope is inserted in the CLDN1 EL1. The kinetics of inhibition of HCV entry, with a $t_{1/2}$ inhibition of infection in 73 minutes, suggested that CLDN1 acts late in the process of viral entry. On the

basis of these results, it is proposed that CLDN1 interaction takes place in virus entry, after interaction of the virus with other co-factors like SR-BI and CD81 [238]. Further, substitution of the human CLDN1 by the murine CLDN1 did not affect infection by HCVpp. This leads to the conclusion that CLDN1 is not a key factor in determining the species specificity. In addition, two teams showed that CLDN6 and -9 appear to be involved in HCV entry into the host cell, like H1H and NKNT3 [239, 240].

8.a.2.4: Occludin (OCLN)

Even after the discovery of CLDN, the hepatocarcinoma cell line Huh-7.5 (*these cells are highly permissive for HCV infection*) and a mouse NIH3T3 embryonic fibroblast cell line that overexpressed human CD81, SR-B1, and CLDN1 (termed N3xF26), remained resistant to HCVpp infection. After transduction with the Huh-7.5–derived expression library, N3xF26 cells were exposed to HCVpp encoding a puromycin resistance gene. From this exposure to HCVpp, puromycin-resistant subpopulations were selected and transfected with retroviral packaging plasmids. This transfection resulted in Huh-7.5–derived cDNA elements that further bestowed upon N3xF26 cells and increased their susceptibility to HCVpp infection. Subsequent rounds of transduction and selection in this process resulted in the emergence of subpopulations of cells that displayed increased susceptibility to HCVpp infection. An analysis of Huh-7.5– derived cDNA from clonal populations of these cells identified another gap junction protein, occludin (OCLN) [242]. These experiments also showed the importance of OCLN in the mechanism of HCV entry [241].

OCLN is an approximetly 60 kDa tetraspan membrane protein with two extracellular loops of the same size (EL1 and EL2), a short intracellular turn, and N- and C-termini cytoplasmic domains (Figure 24). Experiments with permissive cells (*hepatocyte cell lines*) or nonpermissive cells (*HepG2 and 293T*), expressing HCV occludin, showed that the OCLN receptor interacts directly with the envelope glycoprotein E2 [283]. This interaction of OCLN is in contrast with the interaction of CLDN1 with E2. On the other hand, like CLDN1, OCLN is also supposed to be a co-factor acting late in the mechanism of HCV entry.

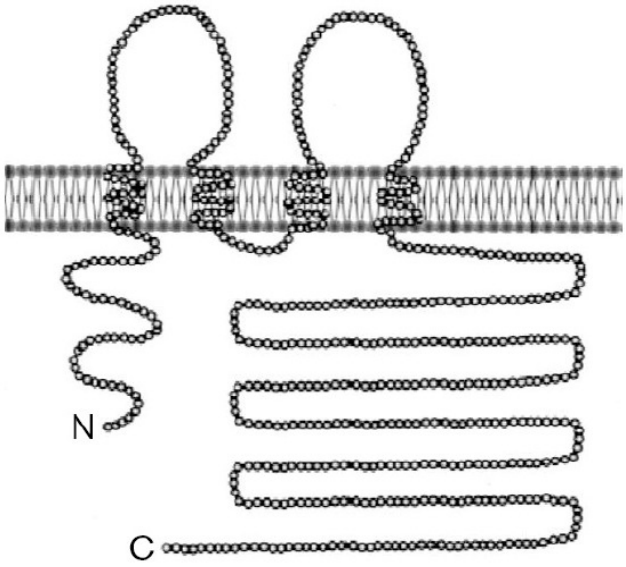


Figure 24: Structure of the junction protein OCLN (based on [284]).

By overexpressing different combinations of the four co-factors SR-B1, CD81, CLDN1 and OCLN in murine and human cell lines of mouse or hamster (*CHO and NIH3T3*), it was observed that the expression of human receptors CD81 and OCLN are important for infection by HCVpp, while receptors SR-B1 and CLDN1 may be of murine origin [242]. However, the rates of HCVpp entry into murine cells are less important than the input rate obtained with human liver cells [242]. This new discovery opened the way to finally use a mouse model for studying HCV.

8.a.2.5: Model of HCV entry

On the basis of already identified co-factors and taking into account the results obtained by different related research groups, a sequential organization of events that governs the entry of the HCV virus can be proposed and summarized in Figure 25.

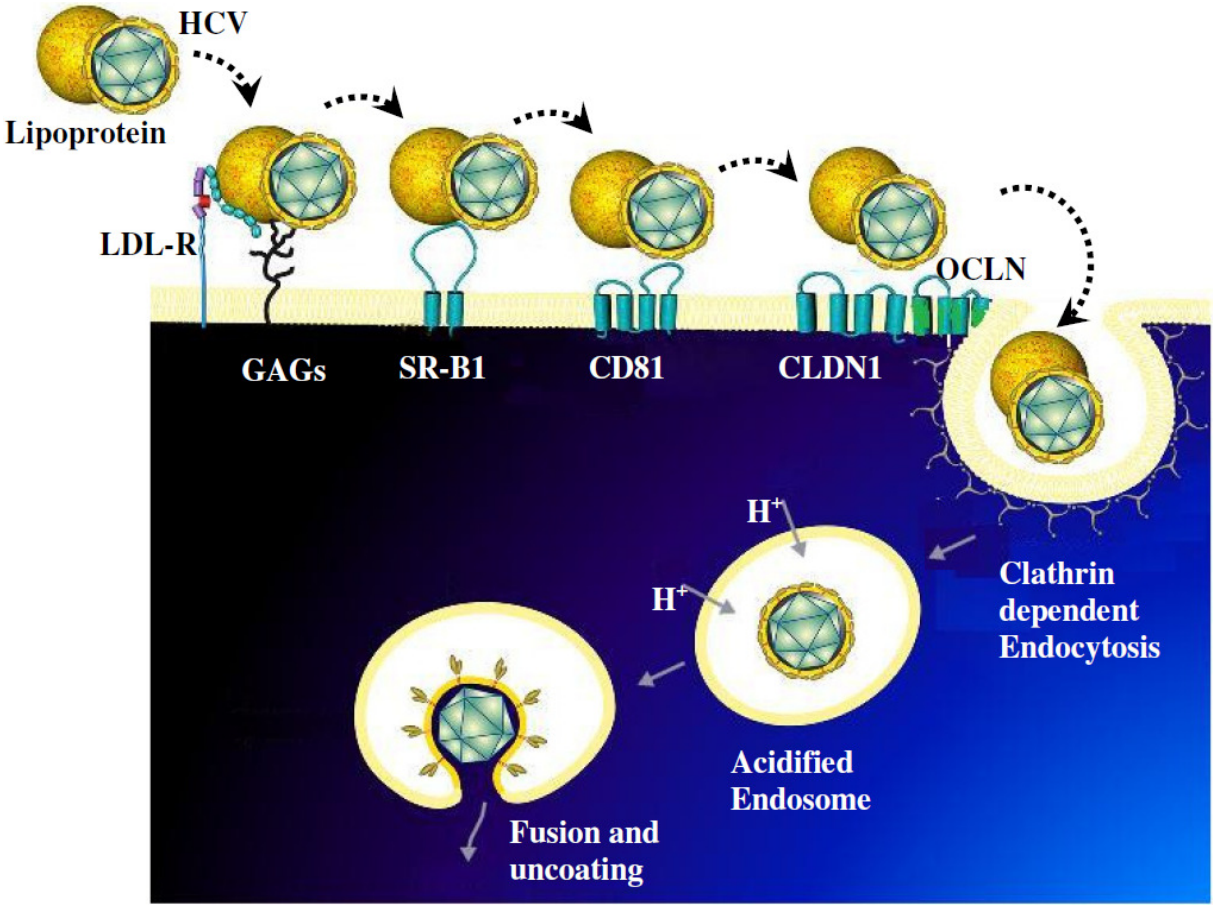


Figure 25: Model of HCV entry (based on [279]). Virus binding to the cell surface and entry involve the low density lipoprotein receptor (LDLR), glycosaminoglycans (GAG), scavenger receptor class B type 1 (SR-B1), the tetraspanin protein CD81 and claudin-1 (CLDN1). CLDN1 functions at a late stage of cell entry, possibly at tight junctions of polarized hepatocytes. Internalization depends on clathrin-mediated endocytosis. Acidification of the endosome induces HCV glycoprotein membrane fusion. Finally, the uncoating process occurs, which results in genome release into the cytosol.

In the bloodstream, the HCV virus particles are coupled with lipoproteins and initially interact with non-specific GAGs and LDL-R on the surface of hepatocytes. These particles are then specifically recognized by SR-B1 at the basolateral pole. SR-B1 brings viral particles at the level of CD81. Further, the interaction of E1E2 envelope glycoproteins with the CD81 receptor induces the virus relocation to cell junctions [285] where it meets the junction protein, CLDN1 and OCLN. This step is followed by endocytosis of the particle via clathrin-mediated pit [286]. Once internalized, the virus is directed to early endosomes where fusion of membranes leads the release of the capsid (Figure 25). This membrane fusion is facilitated by an acidic pH [287].

The presented entry of HCV is based on that of coxsackievirus B. This virus binds to an attachment factor called DAF (decay-accelerating factor) on the apical surface of Caco-2 epithelial cells. This binding induces intracellular signals to relocate the virus at tight junctions [288]. There, the virus interacts specifically with its CAR receptor and enters the cell through a caveolin-dependent mechanism. However, in the case of HCV these specific receptors are not yet fully discovered.

8.b: Internalization and fusion of the virus particle

There are two ways of internalization for enveloped viruses into the host cell. The first is a fusion between the viral envelope and cell membrane, like majority of retroviruses follow. The second one is by the fusion of the viral envelope with the endosomal membrane which is followed by some viruses like influenza virus or vesicular stomatitis virus [289].

In the case of HCV, it was shown that the virus particle is endocytosed via the clathrin-well [286] and only in early endosomes, the membranes fuse to release the viral genome into the cytosol. This process is induced by the conformational changes of the envelope proteins [266, 286, 222, 235, 91, 287, 290].

8.c: Translation and maturation of the viral polyprotein

Upon uncoating, the viral genomic RNA of positive polarity is translated into a polyprotein of ~3011 amino acids by using the cellular translational machinery. This translation takes place at the rough endoplasmic reticulum (rER).

Most eukaryotic and viral messages initiate the translation by a mechanism that involves recognition of a 7-methylguanosine cap at the 5' end of the mRNA. In a few cases, however, translation occurs via a cap-independent mechanism. In this type of mechanism, an Internal Ribosome Entry Site (IRES) in the 5' untranslated region of the mRNA recruits the ribosome. This mechanism is followed in case of HCV. The ~400 nucleotide HCV IRES folds into a magnesium-dependent complex structure, in which loops are thought to interact with the ribosome. These loops are exposed to the RNA surface and enable the translation of viral RNA in a cap-independent manner. IRES consists of three major structural areas (II, III, IV), leaving out the first 24-40 nucleotides of its core region. The 40S subunit of the ribosome binds directly on the area II, through an AUG initiator. This complex then recruits the eukaryotic initiation factor eIF3 along with the ternary complex Met-tRNAi-eIF2-GTP and forms an intermediate non-canonical 48S subunit. This 48S subunit, then associates with 60S

in GTP-dependent manner to form 80S complex (Figure 26). Further, a conformational change is noticed at the level of the 40S subunit [291, 292], due to the binding of eIF3 factor and IRES.

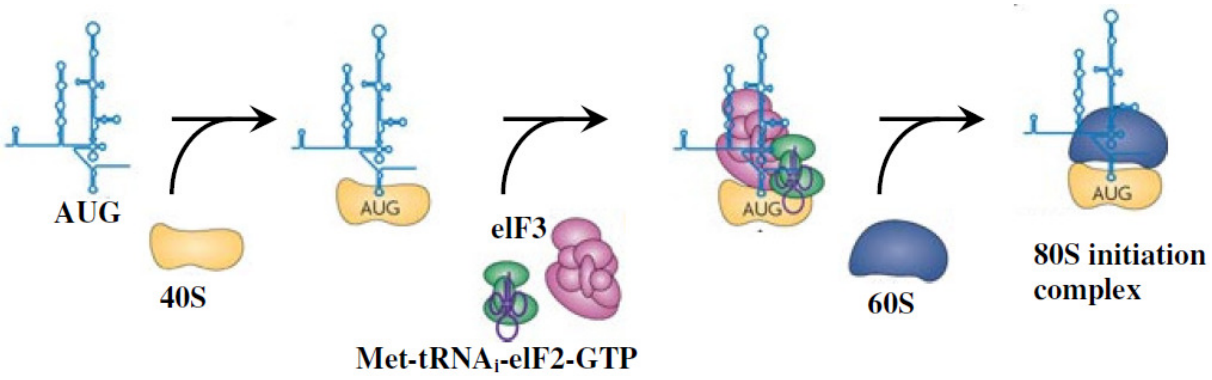


Figure 26: Model of complex formation of translation initiation mediated by HCV IRES. Translation initiation on the HCV IRES begins with the recruitment of the 40S ribosomal subunit directly to the IRES in the absence of any initiation factors. Both eIF2–GTP (bound to Met-tRNAi) and eIF3 are then recruited to the surface of the 40S ribosomal subunit, followed by a GTP hydrolysis step that promotes the joining of the 60S ribosomal subunit to form an 80S initiation complex. (from [293]).

Although, no cellular factor is required for HCV RNA translation [294], several cellular factors, such as polypyrimidine tract binding (PTB) and La proteins, have been shown to bind to the IRES and influence translation [74, 75]. On the other hand, all HCV proteins interact with cellular membranes directly or indirectly and associate with the membraneous web [142, 295-297]. As it has been already described, the translation of the viral genome creates a single polyprotein precursor that is co-and post-translationally cleaved by cellular proteases into structural and nonstructural viral proteins (Figure 27). Structural proteins E1, E2, core and p7 are cleaved by a signal peptidase, while nonstructural proteins are cleaved by two viral proteases, the autocatalytic protease NS2-3 and NS3-4A protease (Figure 27).

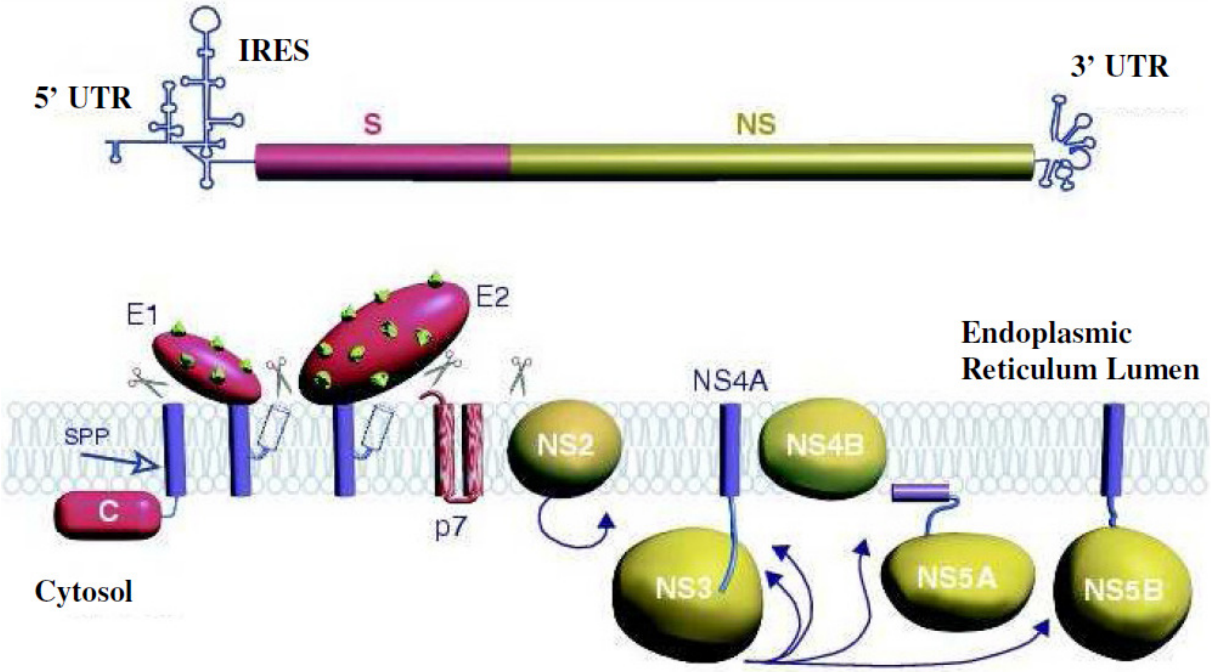


Figure 27: Genome organization, maturation of the HCV polyprotein and localization of proteins in the ER (from [78]).

The arrow locates the site of intramembrane cleavage by the signal peptide peptidase (SPP). Transmembrane domains of proteins E1 and E2 are shown after cleavage, by the signal peptidase, and re-orientation of their hydrophobic C termini (rectangles with dots). Scissors indicate cleavages performed by the signal peptidase (SP) at the Endoplasmic Reticulum and the cyclic arrow shows autocatalytic cleavage of the junction NS2/NS3; arrows in NS3 show NS3/NS4A protease cleavage. The green protrusions represent the sites of glycosylation of the envelope proteins E1 and E2.

The initial cleavages at the core-E1, E1-E2, E2-p7 and p7-NS2 junctions are made by the cellular signal peptidase (SP). Further, the core protein is separated from the signal peptide of E1 by the signal-peptide peptidase (SPP). SPP has a protease activity in cell membranes which leads to the breakage of peptide bonds in the lipid bilayers. Cleavage of the NS2-NS3 junction is a rapid autocatalytic intramolecular reaction directed by the NS2 protease domain and the N-terminal of NS3. Furthermore, the four other junctions are cleaved, probably by a serine protease which is contained within the N-terminus of NS3 [298].

8.d: Replication of viral RNA

The maturation of the viral polyprotein leads to the release of structural and nonstructural proteins. This initiates replication through the replication complex or RC. Like other viruses with positive polarity, replication begins with synthesis of a complementary RNA of negative polarity. Then this newly synthesized strand of negative polarity, serves as a template for synthesizing a new strand of positive polarity. These both steps are catalyzed by a viral enzyme, RNA dependent RNA polymerase (RdRp), NS5B. The recombinant protein NS5B RdRp shows *in vitro* activity but appears to lack specificity and fidelity to its matrix. These characteristics are necessary for faithful viral replication. Thus, cellular factors and/or optimal HCV RNA are required for viral replication and also for the formation of a replication complex (RC). RC binds to the ER membrane, at the "membranous web". This web is a special structure consisting of vesicles and its formation is dependent on the NS4B virus protein (Figure 28).

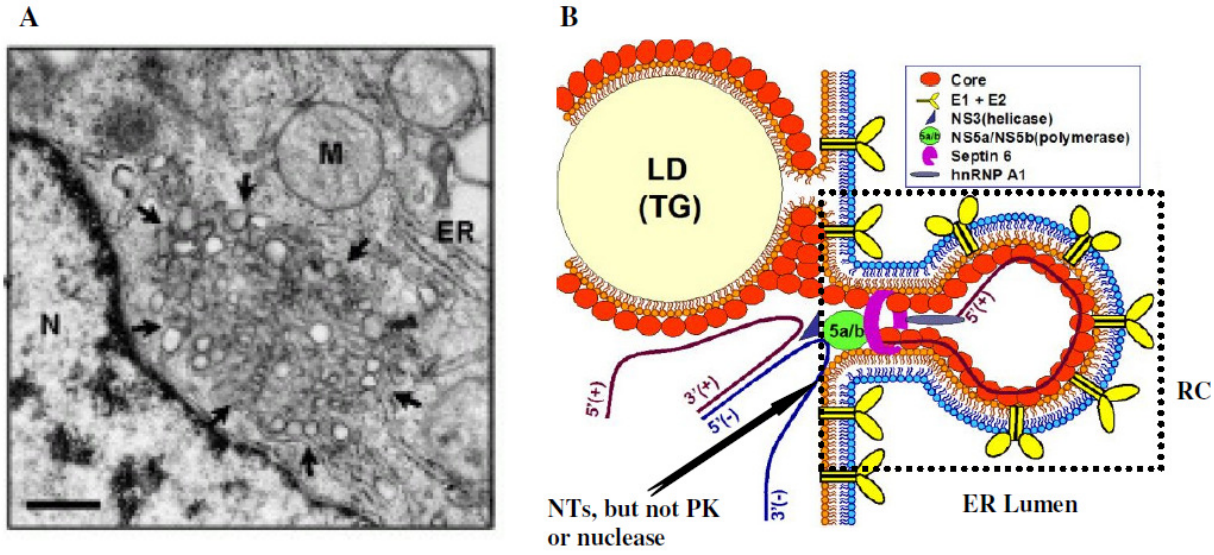


Figure 28: HCV replication based on membraneous network and model of Replication complex (RC) (based on [296, 299])**.**

(A) The "membranous web" of Huh7 cells replicating HCV subgenomic replicon seen by electron microscopy. Bar = 500 nm, N: nucleus, ER: endoplasmic reticulum, M: Mitochondria. (B) Model of replication complex formation. The RC, located in the invagination of the ER membrane, is accessible to nucleotides (NTs) but not to larger molecules such as Proteinase K (PK) or S7 nuclease.

The precise composition of the RC is not known. In cells, only a small portion of nonstructural proteins is associated with the replicon and form RC, the majority of nonstructural proteins which do not interact with replicon form a second complex called nonstructural proteins complex [300, 301]. On the basis of these results, it was suggested that RC is an active compound that consists of a combination of an (-) RNA, two to ten (+) RNAs and several hundred nonstructural proteins. RC is a necessary structural element for the formation of vesicular compartments that serve as sites of HCV replication (Figure 28). This compartmentalization may play a role in the (I) physical support for the RC, (II) increase of the local concentration of viral proteins, to make them more accessible, (III) docking of the viral RNA and facilitating its replication, (IV) reserve lipids which are important for replication and (V) protection of the RNA against the cellular defense [302]. On the other hand, nonstructural protein complex may play a role in the formation of infectious viral particles [151].

There are many other co-factors which regulate the cellular replication of HCV *in vitro* like the miRNAs (miR-122) [77], various host proteins interacting with nonstructural proteins and components of the pathway interfering RNAs [303]. It was demonstrated that the various components of the machinery of RNA interference, like the DICER complex and components of the RISC complex as well as miRNA122, must be functional for effective replication of HCV [303].

Along with these functionalities, other cellular proteins that bind to the proteins NS5A and NS5B are also essential for replication [304]. Thus, membrane-associated proteins, like VAP-A and-B, appear to be involved in the formation of a RC function [305]. VAP-A and B are localized in the ER membranes and Golgi apparatus. Apart for this, many kinases, interacting with the NS5A protein, such PDPK1, RAF1, EIF2AK2 and GRB2, are also important for HCV replication. However, their role in this process is not yet clearly defined [303].

Additionally, studies have revealed that replication of HCV RNA and hepatocyte lipid metabolism, are closely linked. It was observed in cell culture that replication can be stimulated by saturated and monounsaturated fatty acids and inhibited by polyunsaturated fatty acids. These findings again validated the importance of membrane fluidity for membrane network formation and proper functioning of RC.

Further, it has been shown that geranylgeranylation of one or more host proteins is required for HCV replication [306] like FBL-2. The FBL-2 contains a F-box motif that interacts with the NS5A protein. The F-box motif is used in targeting the protein degradation pathway by the

proteasome, thus, FBL-2 could be involved in the degradation of proteins that inhibit replication [160]. Replication of HCV RNA is also dependent on the path of mevalonate. Addition of an inhibitor of this mevalonate pathway (lovostatine, fluvastatin) leads to the blocking of replication [307].

In addition, other host factors like cyclosporine A (CsA) also inhibited RNA replication of HCV *in vitro* [308]. Cyclophilin B targets CsA and interacts with NS5B leading to the boosting of the protein RNA binding activity [309]. Cyclophilin B is a peptidyl-prolyl cis-transisomerase. This enzyme activity has a role that has not yet been defined in the replication of HCV. Based on these findings, non-immunosuppressive CsA analogues are currently under development as antivirals against hepatitis C [310, 311].

Finally, the physiological state of the cell may have a significant effect on efficiency of replication. Confluence, nucleotide distortivity, temperature change and production of ROS (reactive oxygen species) are all factors that can inhibit replication or infection *in vitro* [312-315].

8.e: Factors contributing to HCV assembly

A hallmark of the HCV replication cycle is its tight link with lipid synthesis of host cell. This is best illustrated by the peculiar pathway used for the assembly of infectious HCV particle. Assembly of these particles involves various viral and host factors. The viral proteins involved in HCV assembly are core, E1 and E2, p7, NS2, NS5A, NS4B and NS3 [316, 97, 317-320, 96, 98, 159, 321].

The possible related role of these proteins (except core) has already been described earlier, thus, we will describe the role of host and other factors contributing to HCV assembly.

8.e.1: Cytosolic lipid droplets (cLDs) and their possible role

A prominent accumulation of core protein in close proximity of cLDs in cells, stably or inducibly expressing the core protein [16, 322], was observed. On the basis of this observation, it was concluded that this association is essential for virion assembly. Several independent observations supported this conclusion: (i) inhibition of core cleavage by signal peptide peptidase inhibits mobilization of the protein to cLDs and blocks the particle production [323], (ii) mutations in the core domain D2 as well as any disruption in interaction with cLDs leads to blockage of virus production [324, 151], (iii) mutations in the core domain D2 affecting its binding strength with cLDs alter the efficiency of virion production [325] and (iv) nonstructural

proteins, especially NS5A, are recruited to cLDs and colocalize with the core protein [157, 151]. Mutational ablation of either NS5A-cLD localization or core–NS5A colocalization at cLDs blocks virus production.

Although these data provide compelling evidence that cLDs play a crucial role in HCV assembly, for example by serving as a platform for virion formation, a firm proof for this hypothesis is still missing. In fact, a comparative analysis of core association with cLDs using HCV isolates of high and low assembly competence suggested that cLDs might not be the major assembly sites. Moreover, virus-like particles associated with ER membranes have been detected in close proximity to cLDs in cells transfected with the JFH-1 genome [151]. Thus, it is still not clear, whether assembly initiates on the surface of cLDs or at the ER membrane tightly associated with cLDs.

8.e.2: Host factors contributing to HCV assembly

Host cell factors are often involved in VLDL synthesis, thus supporting the concept of the HCV Lipo-viro-particles (LVPs). Among them, the most notable is apolipoprotein E (apoE) which is a component of infectious HCV particles and essential for assembly [326-329]. Apo E is an exchangeable apolipoprotein with an apparent molecular weight of 34 kDa. It plays an important role in lipid transport by interacting with LDL receptor family proteins [330]. The N-terminal region of Apo E constitutes the receptor binding site (residues 136–150), whereas the C-terminal region contains the lipid binding domain (residues 244–270). However, only apoE proteins associated with lipids can bind LDL receptors with high affinity. Upon binding with lipids, apoE undergoes major conformational changes. Its interaction with NS5A, suggests that this host factor might be required for an early step of assembly [326, 331].

Although, the functional importance of apoB for assembly and infectivity is controversial, it is considered as another host factor which contributes to HCV assembly. It is a nonexchangeable apolipoprotein with an apparent molecular weight of 500 kDa. It is produced in hepatocytes as a full length apoB100 isoform, whereas cells in the intestine release the smaller isoform apoB48. There is a single apoB molecule per lipoprotein particle. It is the only apolipoprotein that remains associated after conversion of VLDL into LDL. A reduction in HCV titres has been reported when the cells were treated either with inhibitors of the microsomal triglyceride transfer protein (MTP) or by RNA interference (RNAi) for knocking down apoB expression [57, 58]. MTP appears as a key enzyme involved in VLDL production. Along with these positive results, there exist contradictory results in respect to the

functions of apoB. Three other groups reported that, in non-cytotoxic concentration range, the RNAi-mediated knockdown of apoE expression, but not that of apo B, reduces the virus production (*whereas MTP inhibition had no effect*) [326-328]. The reasons for these discrepancies are not yet clear but might be due, in part, to the use of different cell culture systems.

Additional proteins implicated in HCV morphogenesis are the heat shock cognate protein 70 (hsc70) (Parent et al., 2009), annexin A2 (ANXA2) [332] and diacylglycerol acyltransferase-1 (DGAT1) [333]. hsc70 colocalizes with HCV proteins on cytosolic lipid droplets (cLDs) and is crucial for both cLD size as well as for the production of infectious virus particles. Antibodies targeting hsc70 reduced infectivity, arguing that a fraction of hsc70 is accessible on the outside of the virion. Surprisingly, the RNAi-mediated knockdown of ANXA2 expression did not affect RNA replication but impaired assembly. NS5A appears to recruit ANXA2, presumably via Domain III. However, the mechanism by which this host factor contributes to assembly remains to be determined. Additionally, DGATI, an enzyme, is an essential enzyme for LD biogenesis. It binds the core protein and localizes it to LDs and thus to assembly sites. Inhibition of DGAT1 activity profoundly inhibits HCV particle production arguing that assembly requires DGAT1-mediated LD formation.

8.f: Assembly of HCV particle

The assembly of virus particles is probably triggered by the multimerisation of core protein and its interaction with genomic RNA. These results in the formation of the nucleocapsid by mechanisms which are not yet elucidated [334]. For simplicity, we divide the assembly process into an early and a late step, corresponding to the formation of the nucleocapsid and its envelopment, respectively.

8.f.1: Nucleocapsid formation

On the basis of fact that viral RNA replication occurs in distinct membranous compartments, located remotely from the assembly site, two general principles of HCV assembly can be envisioned. These principles are differing with regard to where assembly starts, either at the ER membrane or the surface of the cLD (Figure 29).

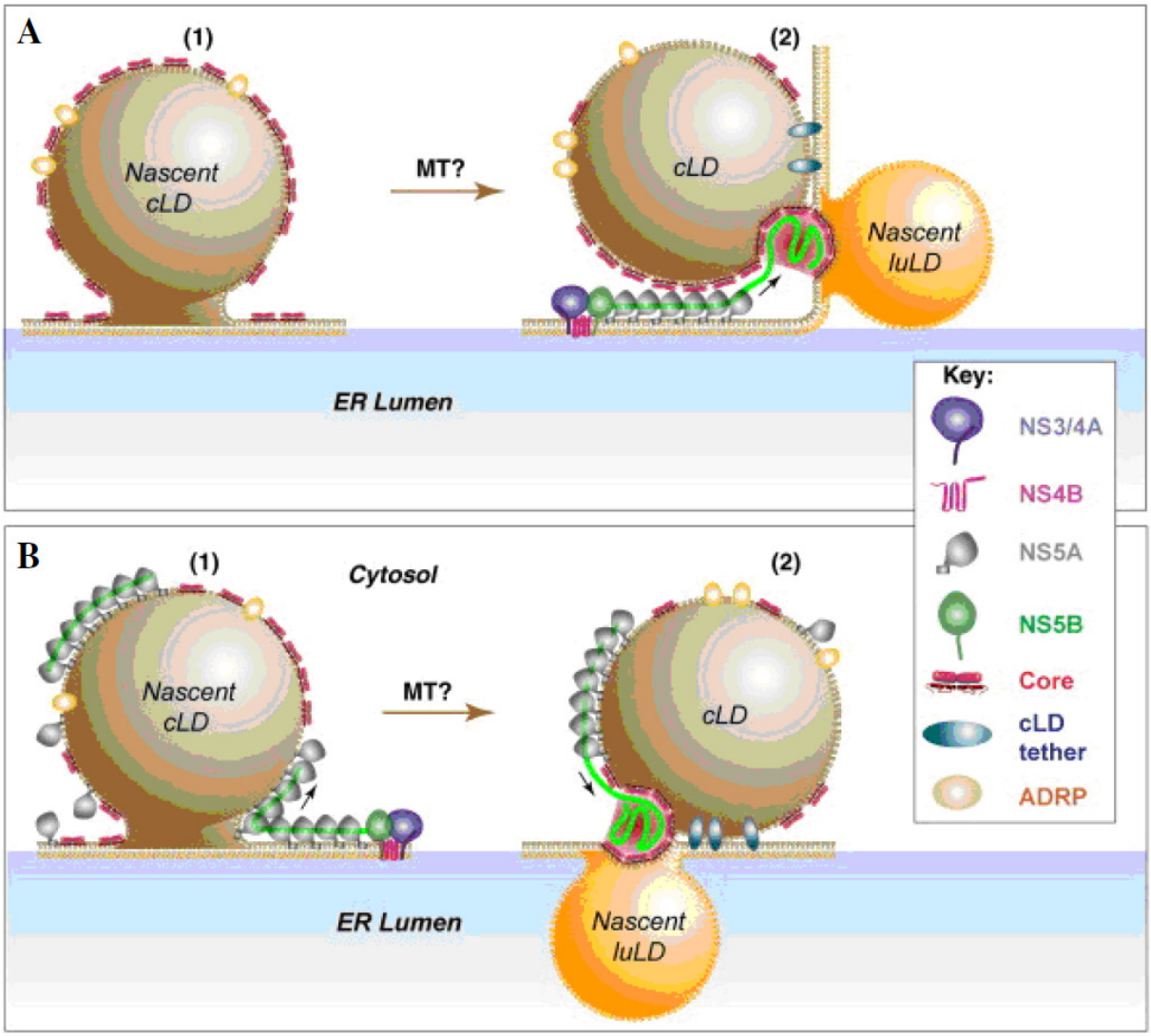


Figure 29: Models of HCV assembly. The biogenesis of cytosolic LDs (cLDs) is thought to take place at distinct domains of the rER with the formation of an intra-ER membrane lens of apolar lipids that finally buds off towards the cytosol. **(A)** Continuity between ER membrane and cLDs should allow their 'loading' with core from the ER membrane (1). cLDs might associate with microtubule motors (MTs) and migrate to a specialized smooth ER (sER)-derived membranous web compartment. There, core-loaded cLDs might attach to the ER membrane via tethering proteins (e.g. Rab 18) at sites where luminal (lu) LDs form and in close proximity of ER-bound replicase (2). Nucleocapsid formation might occur by transfer of the RNA from the ER-resident replicase or (as shown here) from oligomeric complexes of NS5A [70] to the core protein that is released back from the cLD to the ER membrane. However, the interface between cLDs and the ER membrane provides a site favoring the translocation of the hydrophobic core proteins to luLDs that are budding into the ER lumen. **(B)** In this model, cLDs are loaded with core and NS5A (1). ER-bound replication complexes

could deliver HCV RNA to NS5A that can freely move onto cLDs, thus forming an NS5A–RNA array. As in the previous model, these cLDs could associate with MTs and move to specialized ER sites (2). Tethering factors might trap cLDs at the ER membrane where viral RNA is transferred from NS5A to core, thus triggering nucleocapsid formation. This step is linked to the delivery of the nucleocapsid into luLDs. It is also possible that nucleocapsids assemble and bud into cLDs at an early step prior to delivery to luLDs. Both models postulate that the hydrophobic core protein is pulled out of the cLD phospholipid monolayer and inserted into an ER membrane limiting lens of apolar lipids corresponding to nascent luLDs. This process would not require bending of membrane bilayer and thus only limited energy [335].

In the first model, the core protein is initially transferred onto the surface of cLDs. The core protein is then re-recruited to the ER membrane at the assembly sites, or membranous web, where it interacts with the replicase or (RNA-loaded) NS5A (Figure 29A). Hence, cLDs might serve as transport vehicles by transferring the core protein from sites of RNA translation and replication to assembly sites. In fact, cLDs are highly mobile and transported along microtubules [336]. The inhibition of transport by nocodazole or dynein-specific antibodies impaired HCV assembly [337], thus validating this model.

In the second model, the nucleocapsid formation is initiated at the surface of cLDs (Figure 29B). Here, viral RNA is delivered to the core protein by NS5A, which is mobilized onto the surface of cLDs. It is interesting to note that the N-terminal amphipathic helix of NS5A is sufficient for cLD targeting and also well-suited for the monolayer membrane surrounding cLDs [148, 338, 78]. In a variant of this model, the nucleocapsid formation is initiated on the surface of cLDs, but delivery of viral RNA is mediated from the replicase (or NS5A) that remains bound to the ER membrane.

Currently, it is difficult to discriminate between these two possibilities because of the overall low assembly efficiency. This precluded the unambiguous detection of assembly events either in live cells or by using high-resolution (electron) microscopy. As deduced from infectivity titers, assembly efficiency of HCV on a per cell basis is very low and to detect these events, large titers are required.

8.f.2: Envelopment and lipid incorporation

Upon formation of the nucleocapsid, envelopment by cellular membranes containing viral glycoproteins is likely. Although the mechanism of budding for HCV is not well understood, by comparison with the related flaviviruses, HCV is thought to acquire its envelope at the ER [339] where the HCV glycoproteins localize predominantly. Maturation and release of HCV particles, at the late stage of assembly, are tightly linked to the VLDL pathway [327, 57, 58, 328]. Thus, HCV envelopment and maturation could take place in a specialized lipid-rich microdomain at the ER membrane. This microdomain is enriched for LDs and supports synthesis of luminal LDs (luLDs), the precursors of VLDL. Additionally, the core protein accumulates in detergent-resistant lipid fractions containing high amounts of cholesterol and sphingolipids [340]. This suggests that cholesterol and sphingolipids play a crucial role in infectivity of HCV particles [341].

Next, during assembly, the nucleocapsid can be inserted within the core of luLDs (Figure 30). However, the targeting of the envelope glycoproteins to these assembly sites and their incorporation into virions is still unclear. It is assumed that either NS2 alone or together with p7 control the elements of interactions between the replicase, NS3 and the envelope glycoproteins [320, 321]. To strengthen this hypothesis, it has been recently shown that NS2 might 'pull together' envelop glycoproteins and nonstructural proteins (p7, NS3 and NS5A) in close vicinity of LD by forming multiple (direct or indirect) protein–protein interactions between these components [316, 318].

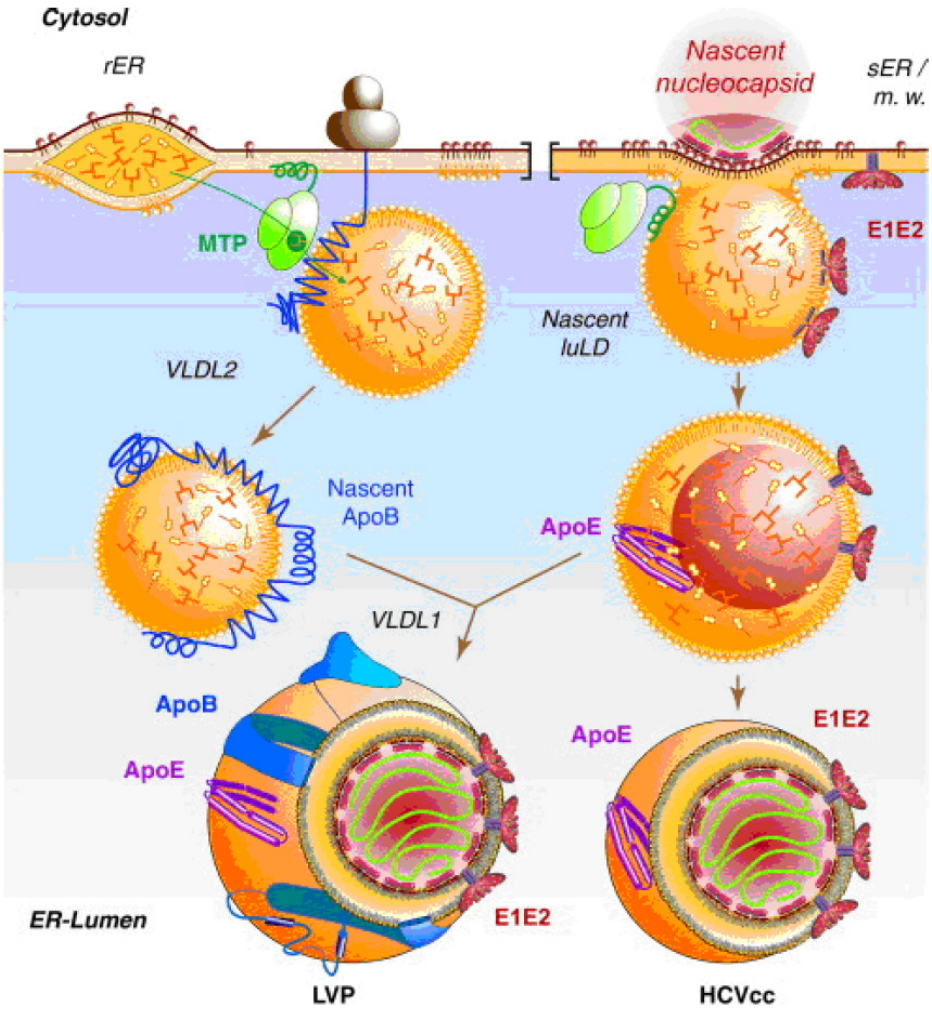


Figure 30: Models of structures of infectious HCV particles and their biogenesis.

During translation at the rER, nascent apoB (blue line) is translocated into the ER lumen and loaded by MTP with phospholipids and triglycerides (left panel). This leads to the formation of a neutral lipid core that is converted into a spherical particle (VLDL2) acquiring exchangeable apoE and apoC. In the smooth ER (sER) or membranous web (m.w.), a second precursor (the luminal LD; luLD) is formed from the ER membrane and by MTP-mediated triglyceride enrichment (right panel). E1 and E2 retained at the ER membrane might slide onto this luLD prior to pinching-off [342]. The nucleocapsid would be inserted into the hydrophobic lipid core of the pinching-off luLD due to the hydrophobic nucleocapsid surface (formed by domain D2 of the core protein). In VLDL competent cells such as primary human hepatocytes, this precursor could fuse with VLDL2 to form the LVP. Alternatively in Huh-7 cells where VLDL1 formation is inefficient [342], HCVcc is secreted predominantly as particles lacking apoB [335].

Furthermore, the way of p7 contribution to assembly is still unknown. However, it was recently proposed that p7 acts as an inhibitor of acidification in intracellular compartments and is probably required to protect virus particles during maturation or exit step [343].

Again, incorporation of apolipoproteins into mature infectious particles is still puzzling. In one model, it has been shown that mature HCV particles (LVPs) result from fusion of the apoB-positive precursor and an apoB-negative precursor which are formed in the rER and in a late compartment of the ER or the ER-derived membranous web, repectively (Figure 30). In this case, HCV particles remain associated with the nonexchangeable apoB, suggesting that the exchangeable apoE and apoC might then be acquired in the lumen of the ER. Mature HCV particles containing apoB, apoE and eventually other apolipoproteins are then transported along the VLDL secretory pathway. Virions thus have the overall structure of triglyceride-rich lipoproteins (TRLs) with a core of neutral lipids surrounded by a monolayer of phospholipids and stabilized by the apolipoproteins [344].

8.g: Release of HCV particles

Following particle formation at intracellular membranes, viral egress to the cell surface is expected. Flaviviruses commonly egress through the host cell constitutive secretory pathway [345]. This pathway is a highly conserved route for proteins to reach the outer membrane of the cell. It involves a complex series of membrane-bound subcellular compartments including the ER, intermediate compartments and Golgi. Most proteins traversing the secretory pathway are modified by N-linked glycosylation, like in case of HCV [80], which takes place in the lumen of the ER. However, during transport through the secretory pathway, it is essential that the integrity and fusogenic form of the glycoproteins remain maintained, so as to allow future host-cell attachment and entry. To prevent premature acid induced inactivation of the glycoproteins, several viruses employ a viroporin to prevent acidification of vesicles. In HCV, this protective function is performed by a viroporin in its p7 gene [100].

Next, members of the flavivirus genus adopt a mechanism of delay cleavage to prevent irreversible conformational changes in the acidic compartment of the secretory pathway [346, 347]. It is likely that HCV employs a similar mechanism in order to retain the fusogenic properties of its glycoproteins. Interestingly, a recent report has described differences in buoyant densities between infectious intracellular virus particles and infectious extracellular particles. This suggests that the biological composition of these two forms of infectious

particles may differ [57]. It is possible that this difference in buoyant density may be due to factors involved in avoiding premature fusion during viral egress.

9: Modes of transmission

The most common factors responsible for HCV transmission worldwide are blood (Transfusion from unscreened donors), intravenous drug abuse, unsafe therapeutic injections and other healthcare related procedures [348]. However, in developed countries, the introduction of blood-screening tests for HCV [349, 350] has effectively eradicated transmission by blood transfusion. Instead, injection drug use has been the predominant mode of transmission in recent times [351]. On the contrary, in the developing world, unsafe therapeutic injections and blood transfusions are still major modes of transmission [348]. HCV transmission via occupational, perinatal or sexual exposure is much less common. Perinatal transmission is estimated to occur in 2.7–8.4% of infants born from HCV infected mothers, with even higher rates in those born to HIV/HCV co-infected mothers [352, 353]. Sexual mode of transmission has also been reported [354], however it is far less efficient than for other sexually transmitted viruses [355].

10: Physiopathology

Hepatitis C is a highly progressive disease. The acute hepatitis phase progresses to chronic hepatitis and in long term can lead to cirrhosis and finally hepatocellular carcinoma (Figure 31).

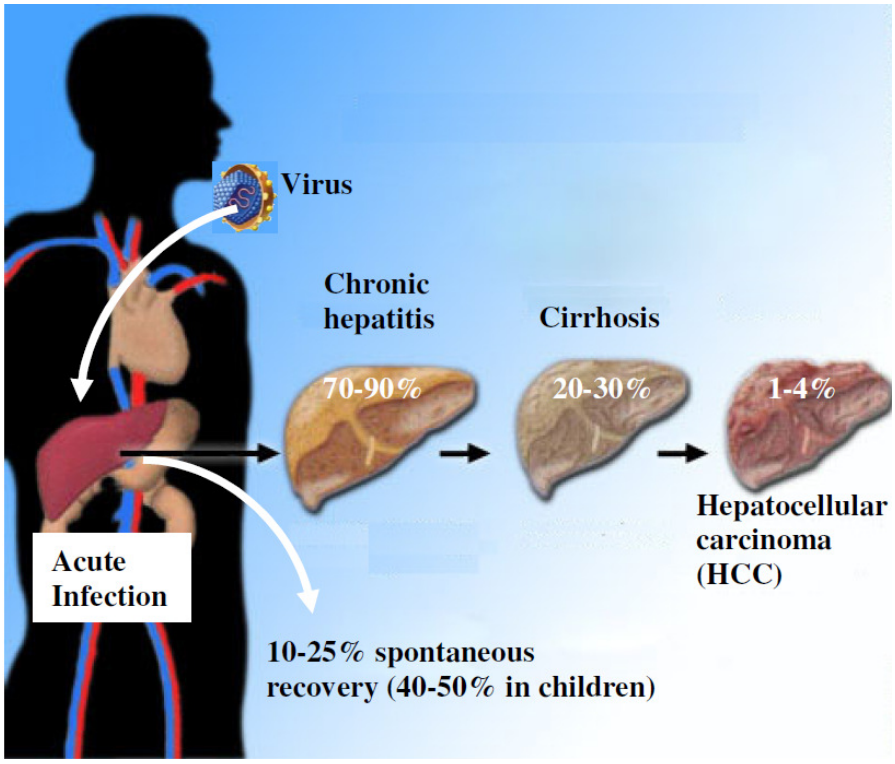


Figure 31: Natural history of HCV infection

10.a: Evolution of the disease and clinical symptoms

10.a.1: Acute Hepatitis C

During hepatitis C, viral RNA can be detected between one to three weeks after infection. Acute hepatitis C is characterized by a positive viremia (viral RNA detection in blood) and negative serology. During this infection a short incubation period of 4 to 12 weeks is followed by an elevation in the level of alanine aminotransferase (ALT). However, the infection is usually self-limiting in less than six months and 15-30 % cases resolve spontaneously due an effective immune response. Although the infection is asymptomatic in most cases, ~20% of patients develop symptoms related to the next stage of hepatitis (nausea, jaundice, vomiting, and anorexia) and 80% of those passed to chronicity [37].

10.a.2: Chronic hepatitis C

Chronic hepatitis C is defined by the viral RNA persistence, in excess of six months, with continuous ALT elevations and the appearance of histological signs [37]. In majority of cases, HCV escapes the immunity system and spreads chronic infection in asymptomatic conditions. However, histological abnormalities due to infection can be observed for most chronic carriers by liver biopsies. It's the biopsy that determines the condition of damaged liver tissue.

Reports indicated that 5% of infected persons remain healthy carriers while 90% develop chronic hepatitis. Among these, 20% evolve to cirrhosis, over an average period of 20 years, with a possible development of hepatocellular carcinoma (HCC) in 4% cases [37, 351]. This transition from acute to chronic phase depends on a number of factors like gender, age, consumption of alcohol, some genes of Major Histocompatibility Complex (MHC) and co-infection with HIV or HBV [37, 351, 348, 356]. Furthermore, metabolic disorders like obesity, diabetes and insulin resistance are associated with more rapid progression of fibrosis in HCV infection and also poor prognosis for antiviral therapy [357].

10.b: Host immune response and viral persistence

Various immune system components, like innate immunity (*including humoral and cellular responses*), are involved in response to pathogens. Thus, the quality of the immune response is essential for the healing or persistence of HCV infection.

10.b.1: Non-specific immune response

Non-specific immune reponse is induced in early stages of HCV infection. During this response, Type I interferons (IFN-α/β) are produced by hepatocytes which are post-infected with HCV and plasmacytoid dendritic cells (pDCs) [358]. These interferons are then recognized by specific cellular receptors (Figure 32) like TLR3 (Toll-like receptor 3) on the surface of cells, and RIG-1 (retinoic acid inducible gene 1) inside the cells (Figure 32). The engagement of these receptors leads to activation of transcription factors, IRF-3 (Interferon regulatory factor 3) and NF-κB (nuclear factor κB) for secretion of IFN-α/β from infected cells (Figure 32) [359]. Activation of NF-κB induces the production of pro-inflammatory cytokines and chemokines. This production amplifies the inflammatory response and also facilitates the recruitment of leukocytes [359]. These leukocytes act in concert with IFN type I in host immune response against HCV.

The host response is triggered when a pathogen-associated molecular pattern (PAMP) presented by the infecting virus is recognized and engaged by specific PAMP receptor factors expressed in the host cell. This interaction with PAMP receptor factors induces the expression of antiviral effector genes [360]. For RNA viruses, viral PAMPs include protein and nucleic acid products of infection or replication like single/double-stranded (ss/ds) and polyuridine signatures. These viral PAMPs are then engaged by specific Toll-like receptors (TLRs) or nucleic acid-binding proteins that serve as PAMP receptors (Figure. 32) [361, 362]. The viral

RNA of HCV contains each of these PAMP signatures and thus, sufficient to trigger the host response when introduced into naïve cells [363, 364].

PAMP receptor's largest effect is the activation of latent cellular transcription factors that mediate the rapid onset of gene expression, thus marking the immediate-early phase of the host response [365]. Furthermore, secreted IFN-β engages the local tissue through paracrine and autocrine processes of binding IFN-α/β receptors. This leads to the activation of Jak-STAT pathway (Figure 32) [359]. In this pathway, the receptor-associated Jak and Tyk1 protein kinases catalyse the phosphorylation of signal transducer and activator of transcription (STAT) proteins on critical serine and tyrosine residues. Jak-STAT pathway in turn activates the transcription of genes induced by IFN called Interferon-stimulated genes (ISGs). These genes are genetic effectors of the host response to viral infection and the human genome encodes hundreds of these ISGs. Progressively, the activation of paracrine pathway by ISGs in neighboring un-infected cells, can limit the spread of the virus [359].

Furthermore, the production of IFN type I is designed to enhance the expression of major histocompatibility complex (MHC) class I on the surface of cells presenting antigen, like DCs, and thus, enhancing the cellular immune response [358]. The cellular response includes the activation of natural killer cells (NK) and cytotoxic CD8+ T lymphocytes (CTL) which lead to liver damage. The death of hepatocytes by NK cells and CTL stimulates myeloid dendritic cells (mDC) which promote the secretion of IFN-γ. Finally, IFN-γ activates the liver macrophages, in order to improve the local inflammation [358].

Interestingly, HCV appears to use several mechanisms to subvert the IFN response and the endogenous CDs. This virus is able to modulate antiviral defenses of infected cells, like the signaling pathways leading to the synthesis of potential antiviral IFN type 1 and other ISGs [359] (Figure 32).

Another way to subvert is through the HCV NS3/4A protease that antagonises the IRF-3 activation as well as the expression of IFN-β by blocking the RIG-1 signaling and TLR-3 [359]. This has multiple consequences related to decreased production of IFN in hepatocytes and blocking the route of IFN-associated PAMP amplification. Among the ISGs, the modification of the MHC-1 expression induces disruption of the presentation pattern of antigens [366]. On the other hand, the loss of NF-κB activation leads to deregulation of the expression of many cytokines and chemokines [367].

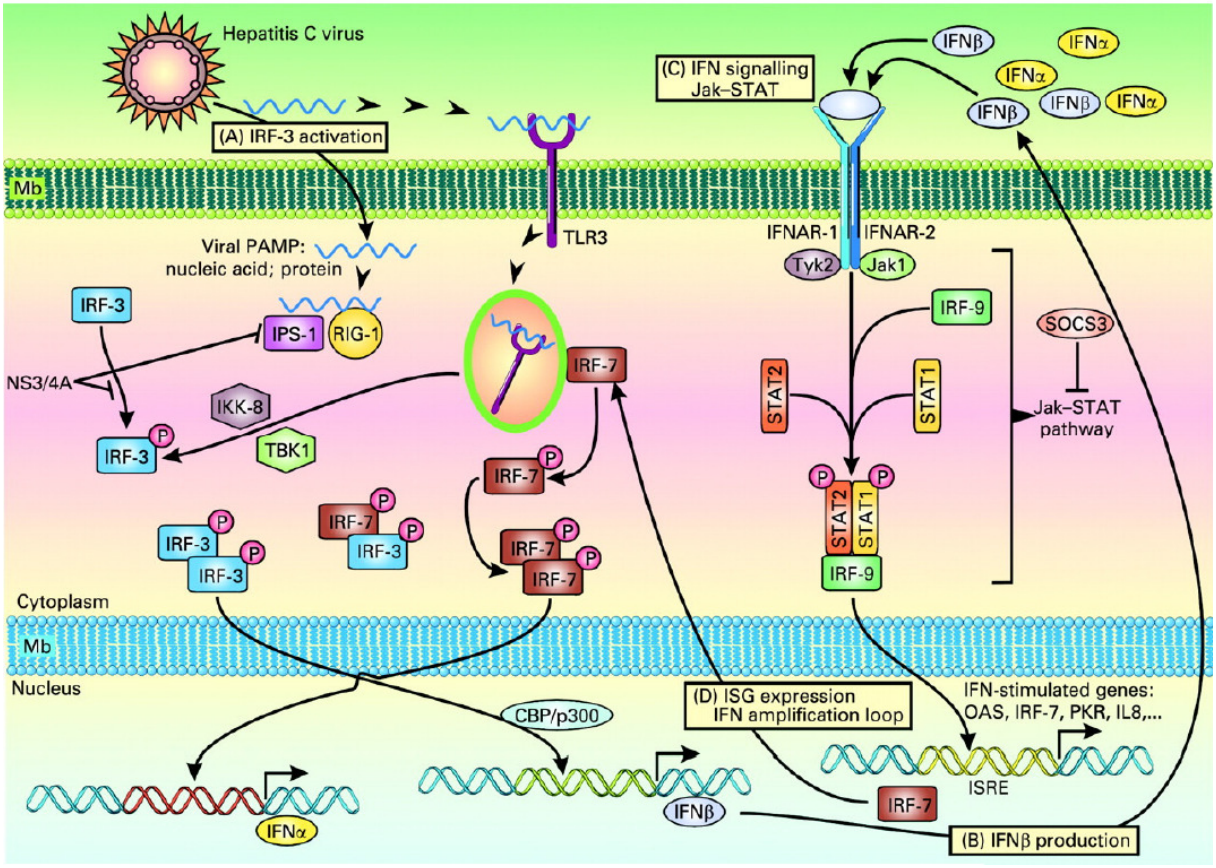


Figure 32: Host response to infection with HCV (from [359]).

(A), Viral PAMP (HCV RNA) binding to RIG-I or TLR3 results in the phosphorylation and activation of IRF-3 by the TBK1 or IKK8- protein kinases [368-370]. The dimer of phospho-IRF-3 translocates to the cell nucleus, interacts with its transcription partners, including CBP/p300 [371, 372] and binds to the cognate-DNA positive regulatory domain (PRD) in the promoter region of IRF-3 target genes, including IFN-β. (B), IRF-3 activation results in IFN-β production and secretion from the infected cell. (C), IFN-β binding to the IFN-α/β receptor signals the activation of the associated Tyk2 and Jak1 protein kinases to direct the phosphorylation and assembly of a STAT1–STAT2 heterodimer and trimeric ISGF3 complex containing IRF-9 [360]. The ISGF3 complex locates to the cell nucleus, where it binds to the ISRE on target genes to direct ISG expression. (D), ISGs are the genetic effectors of the host response. IRF-7 is a transcription factor and an ISG. It is activated after expression through viral PAMP signalling pathways that overlap with the pathways of IRF-3 activation. IRF-7 phosphorylation, dimerization and heterodimerization with IRF-3 allow it to bind its cognate virus-responsive element (VRE) in the promoter region of IFN-α genes, resulting in the production of various IFN-α subtypes that further signal ISG expression [373]. This increases the abundance of RIG-I and viral PAMP signalling components whose continued signalling serves to amplify IFN production and the host response. The therapeutic administration of

IFN-α provides antiviral action against HCV by signalling ISG expression through the IFN-α/β receptor and the Jak–STAT pathway. RIG-I and TLR3 signalling ablation by the HCV NS3/4A protease blocks IRF-3 activation and attenuates the host response to infection.

Furthermore, the expression of core protein in cultured cells is associated with an increase in the level of expression of the protein SOCS-3 (suppressor of cytokine signaling) [374, 375]. SOCS are inhibitors of the Jak-STAT pathway and also negatively regulate the receptor signaling pathway to IFN-α/β. However, their exact role in HCV infection is still unknown. Likewise, the NS5A protein is also an IFN antagonist and appears to regulate the expression of the ISGs [374] and its *in vitro* expression can suppress the action of IFN-α [162].

On the other hand, IFN-λ, type III interferon, also hampers the replication of certain viruses like HBV and HCV [376, 377]. It induces antiviral activity against HCV via the same kind of receptor which activates the signaling pathway in case of IFN-α (Figure 33). The distribution of this receptor for IFN-λ is smaller, as compared to the receptor of IFN-α/β, on the surface of different cell types and inhibits viral replication by activating ISGs [376]. Additionally, *in vitro* experiments reveled that IFN-λ induces a constant increase in ISGs level [376].

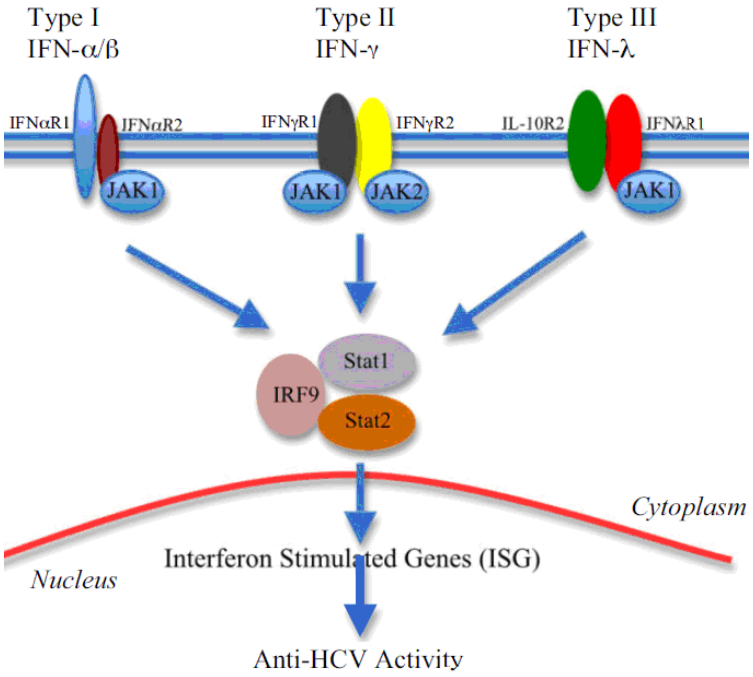


Figure 33: Signaling pathway of IFN type I, II and III (from [378])

The interferon-signaling pathway has three different receptor complexes (Type I, II, III), which mediate distinct differences in JAK (Janus kinase) and Stat (Signal transducer and acitivator of transcription) signaling within the cell cytoplasm. Subsequent interferon

stimulated gene (ISG) expression is differentially expressed, and this in turn determines the relative anti-hepatitis C virus activity of the pathway. The identified IL28B polymorphisms are near the gene that encodes interferon-λ3, which signals through the Type III interferon receptor complex.

Importantly, pegylated interferon- λ (PEG-IFN-λ) is being tested as a monotherapy, and also in combination with ribavirin, in patients non-responsive towards IFN-α treatment. The initial results of the clinical trial showed that PEG-IFN-λ reduces the viral load without inducing side effects, which are commonly observed with IFN-α [379]. However, recently, three major genetic associations demonstrated a link between the lack of response to combination therapy and genetic variations present in the IL28B gene encoding IFN-λ3 [380-382]. This discovery could lead to make better clinical decisions for treatment of chronic infection on the basis of individual perticularity, as all individuals do not repond in a similar way to a particular treatment.

10.b.2: Specific immune response

Specific immune responses include humoral and cellular immune responses.

Humoral immune response

Development of neutralizing antibodies is a hallmark of clearance in many viral infections and induction of these antibodies through immunization with viral subunits or inactivated virus is a classic strategy for the induction of protective immunity [383]. HCV-specific antibodies usually become detectable in the serum within several weeks after primary HCV infection, although the range is highly variable. The first detectable antibodies against HCV antigens in serum usually target NS3 protein (anti-c33 Ab) and core protein (anti-capsid Ab or anti-22c Ab) [384]. Later, these antibodies target NS4 and envelope glycoproteins (E1 and E2) [383, 384]. The hypervariable region-1 (HVR-1) of the glycoprotein E2 of HCV is considered the major target for neutralizing antibodies. *In vivo* studies in chimpanzees revealed that HCV antibody specific for the HVR-1 region, with a capacity of neutralization *in vitro*, have protective effects against HCV infection [37, 385]. On the other hand, anti-HVR-1 antibodies do not confer protective immunity against reinfection [386, 387]. However, patients remain persistently infected despite the presence of high titers of neutralizing antibodies which

indicate their uneffectiveness. This is probably due to the genetic variability of HCV and thus selection of variants that escape constantly to antibodies.

T cell response or cellular immune response

Strong HCV-specific CD4+ and CD8+ T cell responses against HCV multiple epitopes are necessary for spontaneous viral clearance during the acute phase. But at the same time, the virus appears to have multiple strategies to evade these defenses [386, 388].

CD8+ T cells recognize viral antigens presented by MHC class I on antigen-presenting cells and infected target cells. Their antiviral activity and cytotoxicity involves the secretion of antiviral cytokines as interferon-gamma (IFN-γ). In antiviral immune response, CD4+ T cells have effector functions such as the secretion of antiviral cytokines and activation of viral-specific B and CD8+ T cells. Numerous studies has shown that during the acute phase, infection can be limit by strong responses of CD4+ and CD8+ T cell [389, 388, 390, 391]. However, there exists a mismatch between the onset of viremia and T-cell response. But still, decrease in viremia coincides with specific T-cell response and induction of IFN-γ in peripheral blood and liver [391]. It has been shown that without CD4+ T cell responses, HCV-specific CD8+ T cell and neutralizing antibodies could develop, but the subsequent viremia cannot be controlled [392]. Further, studies in humans and chimpanzees have shown that when CD8+ T response is vigorous and targets many epitopes, it can suppress HCV infection during acute phase. Inversely, the infection becomes chronic when the CD8+ T response is low and target few epitopes. Additionally, using the chimpanzee model, in which it is possible to deplete specific cell populations, it has been shown that CD4+ and CD8+ T cells have critical role in primary protective immunity [393, 394]. Although, CD8+ T cells are primary effector cells but without the help of CD4+ T cells, they can not control viral replication and persistant infection is more likely to develop [393-395]. CD8+ T cells have both cytolytic and non-cytolytic effector functions, the latter of which is mediated by production of cytokines such as IFN-γ and INF-α. Among them, the non-cytolytic effector function has been described as a mechanism important for viral clearance [396]. On the other hand, the mechanism by which CD8+ T cells may control HCV replication is still poorly known. Most researchers have focused on the role of IFN-γ as a key cytokine which is both necessary and sufficient to inhibit replication of HCV. Experiments revealed that both IFN-γ and IFN-α/β can inhibit replication [397]. Further, it has been observed that the level of IFN-γ expression correlates positively with the degree of inflammation within liver tissue [398]. But still, there is no

evidence that the level of IFN-γ expression is related to HCV viral load. The positive correlation suggests that other cytokines may play a role in controlling HCV replication. Thus, viral persistence appears to be a multifactorial mechanism and HCV developed several strategies to counter the host immune response.

11: Therapy

In the absence of an effective vaccine against HCV, the current treatment for patients suffering from chronic hepatitis is based on a combination of a stabilized form of IFN-α, IFN-α pegylated (PEG-IFN-α) and a nucleoside analogue, ribavirin. PEG-IFN-α provides a longer half-time, better pharmacokinetics and a better rate of antiviral response compared to interferon-α. However, genotyping of HCV patients is essential because duration and response to treatment is depending on virus genotype. This therapy can treat about 80% of patients infected with genotype 2 or 3, and about 50% of those infected with genotype 1. The duration of treatment is 24 weeks in case of genotype 2 or 3, and 48 weeks for genotype 1 [399].

Inspite of its effectivity, this treatment has many drawbacks viz., a high rate of non-responding patients, time consumption, costly, weight loss, nausea, flu-like symptoms (fever, myalgia, and chills), depression, disorders neuropsychiatric and other hematologic abnormalities. Treatment is also contraindicated in pregnancy or renal failure. It is therefore an urgent need to identify new therapy targets to develop new treatments against HCV. In this regard, the HCV life cycle has a number of potential targets for targeted molecular therapy. Moreover, several new antiviral therapies are currently at varying stages of preclinical and clinical developments [174].

Early studies focused primarily on improving the current treatment processes because it induces many side effects. This leads to the discoveries of Albuferon-alpha and viramidine, which are now in phase III of development. The Albuferon-alpha is an IFN-α2b fused to albumin, which gives longer half-life. On the other hand, viramidine is a precursor of ribavirin with the same antiviral effect, but without hemolytic effect, as induced by treatment with ribavirin.

Another targeting agent could be the NS3 protease inhibitors as they can block viral replication, by preventing the maturation of the polyprotein, and formation of replication complex (RC). In addition, NS3 protease inhibition would increase the effectiveness of host antiviral response by blocking the activation of the host's innate immune response. The

molecule VX-950 (telaprevir) (Vertex / Mitsubishi) is able to block the NS3 protease and effectively inhibit the replication of HCV in cell culture. It has been associated with a good tolerance ratio in laboratory animals. Unfortunately, it is the only molecule that is currently in phase III of development [400]. The results of Phase I of clinical trials show that VX-950 has a very strong antiviral activity [401]. The emergence of resistant mutants was also delayed *in vitro* and *in vivo* [402, 401, 403].

Moreover, clinical phase 2b studies have recently shown encouraging results while combining telaprevir with IFN-α and ribavirin [404-406]. But this combination treatment induces many side effects and the emergence of resistant mutants [404-406].

Another possibility could be the use of NS5B polymerase inhibitors in the course of targeting the virus replication. These are of two types, nucleoside, like ribavirin, and non-nucleoside inhibitors. Nucleoside analogues are converted into nucleotides by cells and then incorporated into the viral RNA during synthesis which in turn blocks the RNA synthesis. The molecule R7128, an analogue of cytidine, shows very strong antiviral activity in clinical trials [407]. On the other hand, non-nucleoside inhibitors are allosteric inhibitors of NS5B. These prevent the conformational changes of the enzyme required for initiation of RNA synthesis. The molecule GS9190 showed a strong antiviral activity in Phase I clinical trial [407] by associating with pegylated interferon therapy. This association provided a concept of dual therapy with ribavrine.

There are also other approaches, specifically designed to target cellular proteins which are necessary for replication and morhogenesis of infectious viral particles. These approaches include the inhibition of cyclophilin, by Debio-025 [311] and MIM811 [310]. Another possibility is the inhibition of glycosylation of envelope proteins by sugar-imines like celgosovir which is an alpha glucosidase I inhibitor. These sugar-imines will be concentrated in the endoplasmic reticulum (ER) and inhibit α-glucosidase [408].

Another possibility could be the use of TLRs-agonist molecules like SM360320. This molecule is an agonist of TLR7 and reduces *in vitro* HCV RNA levels by inducing the expression of IFN type 1 [409]. Molecules containing the CpG nucleotide motifs (*Actilion*) targeting TLR9 receptor is able to stimulate receptors and subsequently, inducing an effective immune response [410]. These molecules are present in the CDs.

Another area of research focuses on vaccine development therapies designed to stimulate cellular immune response against the virus. The development of therapeutic vaccines is necessary, due to the limited effectiveness of current available treatments. HCV antigenic

proteins, expressed in different forms (*proteins alone, DNA, recombinant virus-like particles*) are used to boost the host immune response. For example, the IC41 (Intercell, Vienna, Austria) is a vaccine that contains multiple epitopes of HCV genome and synthetic adjuvant polyarginine. In a phase II study, the IC41 induced immune responses and significant reductions in transient HCV RNA in serum of patients. The test was performed with patients who had not responded to standard therapies [411]. The TG4040 (Transgene, Lyon, France) is another therapeutic vaccine based on a MVA (*modified vaccinia Ankara virus*) viral vector encoding proteins NS3, NS4 and NS5B and stimulating T cell response specific for HCV [412]. Moreover, the virus-like particles, VLPs, with their surface envelope glycoproteins expressing antigenic properties similar to those of wild-type virus [218], can also be used as therapeutic vaccines [413]. Interestingly, recent discovery has shown that monoclonal anti-claudin 1 antibodies prevent HCV infection of primary human hepatocytes [548].

In addition to that, the development of a preventive vaccine to eradicate HCV is also under study. The vaccine must be capable of inducing the production of neutralizing antibodies effective against HCV. It should be able to enhance the response of memory T cells to specifically recognizing the virus. However, due to the high genetic variability of the virus, the construction of such effective vaccine against all genotypes of HCV is a real challenge.

12: Intracellular Distribution of core protein

Core has been shown to exhibit a cytoplasmic and granular localization in liver biopsies [414, 415]. Subcellular fractions indicated that the core is associated with membranes while immmunofluorescence data from tissue cultures has shown the core protein to be cytoplasmic, and to associate with granular structures [16, 15, 416]. These granular structures have been shown to be LDLDs by EM [16] and domain D2 of core, consisting of two amphipathic alpha-helicase separated by a hydrophobic loop, is responsible for targeting core to LDLDs [417, 418]. Although, core has mostly been reported to be bound to cytoplasmic membranes, a portion may localize to the nucleus [419, 420]. This nuclear species of core was reported to be conformationally distinct from the cytoplasmic species [421], but the presence and relevance of this nuclear species has yet to be confirmed *in vivo*.

13: HCV Core Protein (AF009606 – Isolate HCV- H77)

During translation of the HCV polyprotein, the nascent polypeptide is targeted to the host ER membrane for translocation of the E1 ectodomain into the ER lumen, a process mediated by

an internal signal sequence located between the core and E1 sequences. Cleavage of the signal sequence by the host signal peptidase yields the 191 aa long immature form of the core protein, which contains the E1 signal sequence at its C-terminus. This cleavage is predicted to occur on the luminal side of the ER [17, 15]. This signal peptide is processed further by a host signal peptide peptidase, yielding the mature core protein of 179 aa or so of 21 kDa [422]. Most of the core protein is found in the cytoplasm, where it is bound to ER membrane or located at the surface of LDLDs [16, 418, 193]. This maturation of core by SPP is necessary for the formation of infectious viral particles [323]. Apart from LDs, the protein can be retained at the ER membrane and on mitochondria [322, 19] as well as in the nucleus. The mature protein is a dimeric, alpha-helical protein exhibiting features that are consistent with those of a membrane protein [423].

Figure 34: **Amino acid sequence of the HCV core protein**

The N-terminal D1 domain comprises three areas rich in basic residues (BD1 to BD3) and a region rich in tryptophan residues (WD). The C-terminal D2 domain is hydrophobic and binds to membranes.

The first 117 N-terminal amino acids of the core protein constitute a RNA-binding hydrophilic domain, named D1 domain, that contains three highly basic clusters (BD1-BD3) (Figure 34) [24], bearing predicted nuclear localization signals (Figure 34). The N-terminal region of the D1 domain contains immunodominant antigenic sites, and its structural analysis has revealed a helix–loop–helix motif at amino acids 17–37, carrying at least one conformational epitope [424]. Additionally, it also contains amino acids (between residues 36 and 91) involved in the homodimerization of the core [425, 426] required to form the viral nucleocapsid [427, 428]. The region between amino acids 82-102 contains a tryptophan-rich sequence (WD) believed to be involved in homotypic core interactions. The D1 domain shares the characteristics of natively unfolded proteins, which are believed to undergo

"induced folding" on binding to their natural ligands [429]. This conformational plasticity allows the core protein to interact with many different cellular partners and may account for the wide range of functions attributed to this protein.

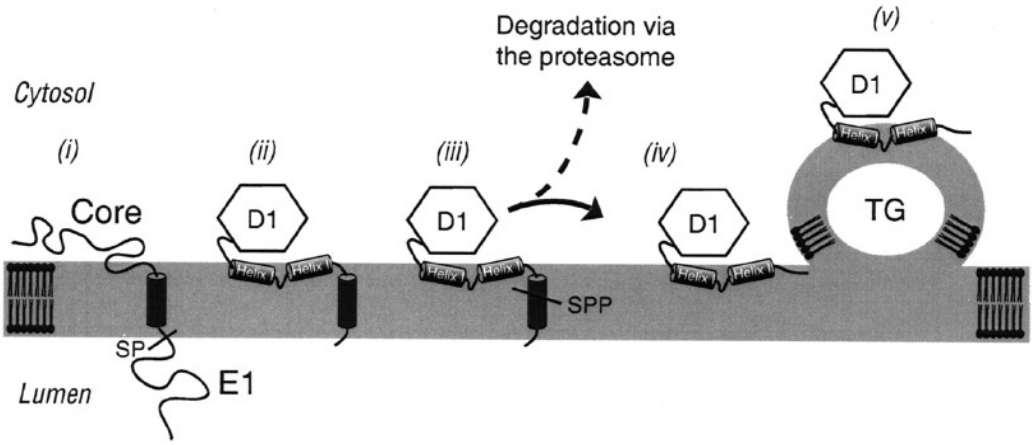


Figure 35: Model depicting the role of the D2 domain in the folding of HCV core and LD targeting. *i*, the HCV polyprotein is translocated to the ER membrane and cleaved by SP to generate the N terminus of E1. *ii*, interaction of D2 with the ER membrane induces folding of Helix I and II, which promotes folding of D1. *iii* and *iv*, SPP cleavage within $SP_{core\text{-}E1}$ releases the mature form of core that remains attached to the ER membrane via the structural motifs in D2. Alternatively, after SPP cleavage, incorrectly folded protein is targeted to the proteasome for degradation. *v*, the mature core protein transfers from the ER membrane to the surface of LDs that are located at the ER membrane [417].

The C-terminus of the mature core protein (segment 118–179 or so) named domain D2 is strongly hydrophobic. This domain is predicted to fold into two amphipathic α-helices (Helix I and II) separated by a hydrophobic loop and is responsible for core association with LDs and ER membranes [16, 417, 418, 322] (Figure 35). The introduction of mutations in this C-terminus part affects the distribution of core, resulting in a decreased production of infectious particles [324].

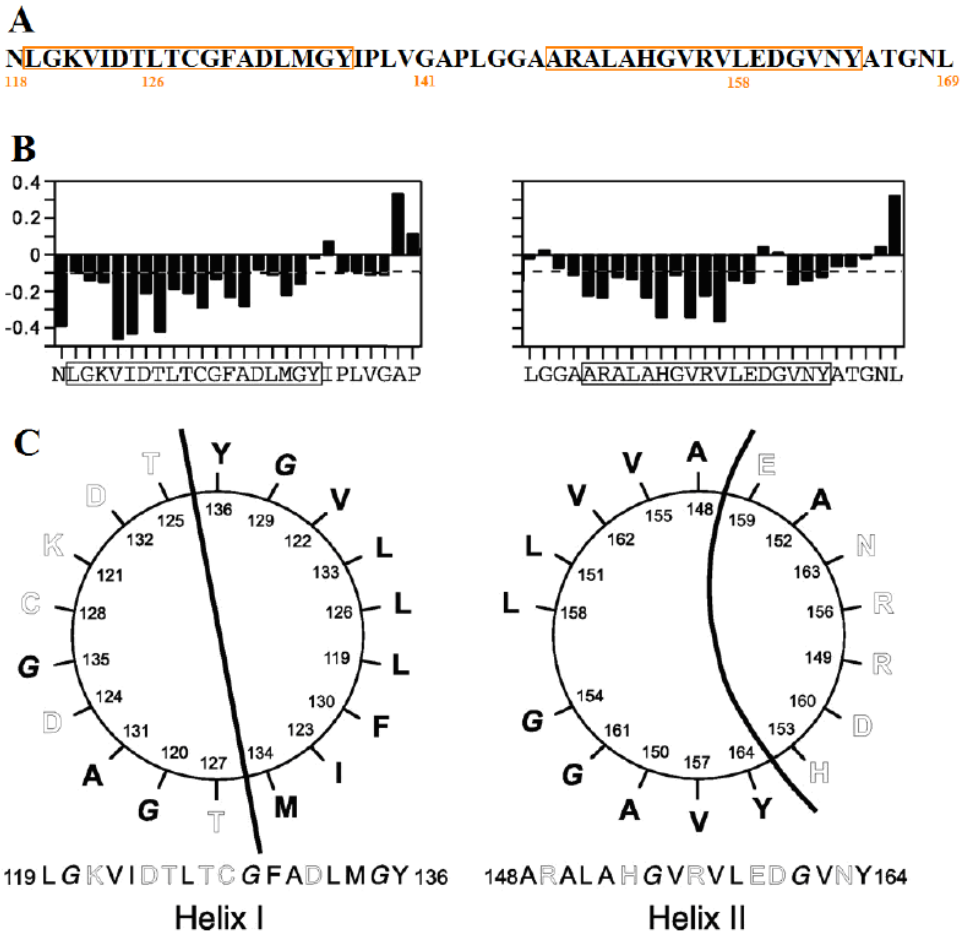


Figure 36: ***Sequence and structure of the D2 domain****(A)* Amino acid sequence of the D2 domain. *(B)* ^{1}Hα chemical shift differences (in ppm) for each residue. The *dotted line* indicates the standard threshold value for an α-helix (ΔHα of -0.1 ppm). *Boxed* amino acids correspond to residues located in α-helical regions. (*C)* helix projections of α-helices 119-136 (*left*) and 148-164 (*right*) showing their amphipathic nature. *Outlined* and *boldface letters* correspond to polar and hydrophobic residues, respectively.

Furthermore, ^{1}Hα chemical shift differences (Figure 36B) and helix projections of α-helices (Figure 36C) of D2 have already been proposed. The first amino acid in D2 has been positioned at residue 118, but its C-terminal limit is less well defined and lays in the region between amino acids 171 and 182, which encompasses the beginning of the signal peptide between core and E1 and the Signal Peptide Protease cleavage site. From studies with a panel of deletion mutants, removing regions from domain D1 of HCV core does not impair LD association [418]. By contrast, deleting segments of D2 completely abolishes the attachment to LDLDs [418]. Thus, the key residues for binding to LDLDs resided within D2. This hypothesis

was confirmed by linking D2 to the C-terminal end of GFP, creating a GFP-D2 chimeric protein that is targeted to LDLDs [417]. Modelling of D2 predicts that the two helices interact in-plane with the membrane surface, with the hydrophobic residues contacting lipids at the membrane interface (Figure 37) [417].

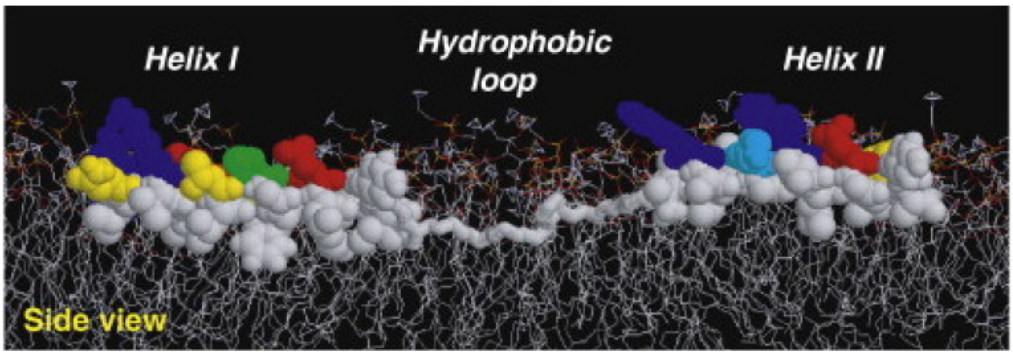


Figure 37: Model of the interaction between the helices of D2 and the membrane surface. Hydrophobic amino acids directed towards the lipid interior are shown in grey; coloured residues represent hydrophilic amino acids (based on [430]).

Folding of the helical regions in D2 may rely on interaction with membrane surfaces since mutations of hydrophobic amino acids in Helix I and II lead to loss of its efficiency to attach with LDs [417]. Both helices require a hydrophobic environment for folding, indicating that lipid interactions contribute to their structural integrity [417]. Alterations of these residues also induce degradation of the entire core protein [417]. It is thought that mutations of hydrophobic residues in the helices may perturb their correct interaction with membranes and as a result impair folding. Additionally, D2 requires its C-terminal limit till amino acid 161 for optimal activity [417]. Any defects in D2 folding may be transmitted to the D1 region of core, creating a misfolded protein that is recognised by the degradative machinery within the cell. Recently, an extensive mutagenesis study across the entire core-coding region has confirmed the importance of the D2 domain for the stability of the core protein [431].

The precise nature of the interactions between the D2 domain and membrane surfaces, such as those on LDs has not been determined yet. However, no single structure motif within domain D2 is responsible for LD association since a combination of both amphipathic alpha-helices and the hydrophobic loop is required [417].

Finally, the core exerts multiple roles in infected cells and virus replication. The core notably interacts with a number of cellular proteins which impact on cell proliferation and differentiation [4, 20] and also on the lipid metabolism with a possible influence on

hepatocellular carcinomas [21]. As a major viral component tightly binding to genomic RNA sequences, the core is also thought to drive genomic RNA packaging and nucleocapsid formation [217, 432, 433]. Moreover, the core has potent nucleic acid chaperoning activities [26, 24].

13.a: Host Cell Proteins Interacting with Core

There are number of host cell proteins which have been identified as interacting partners with HCV core protein. These host cell proteins were identified by yeast-2-hybrid screening of human cDNA libraries using core sequences as bait and the interactions were confirmed by biochemical analyses, including GST-pull down assay and *in vivo* colocalization by immunoflurescence, and other recent techniques [434, 435].

13.a.1: Lymphotoxin-β Receptor (LT-βR)

The core protein binds the cytoplasmic tail of LT-βR [436, 437], which is a member of the tumour necrosis factor receptor family. Although, the exact function of LT-βR is unknown, it is thought to be involved in cytolytic and NF-κB activation in certain cell types. The region of core including residues 1-91 or domain 1 is required for interaction with LT-βR [436, 437]. Additionally, core protein when expressed in HeLa cells, enhances the cytolytic effects of LT-βR ligand (lymphotoxin-α_1-β_2) while showed no such effect in either Huh7 or HepG2 cells [436].

13.a.2: Tumour Necrosis Factor Receptor 1 (TNFR1)

Core protein has also been shown to bind the prototype tumour necrosis factor receptor, TNFR1 [438] as shown by GST-fusion protein pull-down assays. TNFR1 is the primary receptor mediated TNF induction and is involved in cell death signaling and NF-κB activation. The interaction requires residues 1-117 of core protein and the cytoplasmic tail region of TNFR1 (residues 345-407). Finally, in the presence of core, HepG2 and HeLa cells were more sensitive to TNF- or anti-TNFR1 antibody-induced cell death [438].

13.a.3: Heterogeneous Nuclear Ribonucleoprotein K (hnRNP K)

Core protein specifically interacts with hnRNP K as shown by yeast-2-hybrid, GST-fusion protein binding and colocalisation studies [439]. It is a transcriptional regulator with both RNA- and DNA-binding properties and is a component of hnRNP complex. Although it is predominantly found in the nucleus, it also shuttles to the cytoplasm and plays a role in

processing and transport of pre-mRNA. Residues 1-115 of HCV core domain I are required for the interaction with hnRNP K. Core protein was also shown to partially reverse the suppressive effect of hnRNP K on the human thymidine kinase gene, possibly by binding the proline-rich regions of hnRNP K which blocks the interaction of hnRNP K with other cellular factors [439].

13.a.4: Apolipoprotein All (apoAll)

While describing the cytoplasmic localization of core protein, it was observed that core and apoAll colocalize on the surface of LDs in HepG2 cells [16]. The region involved in the interaction was later identified as residues 160-173 of core and its truncation leads to loss of interaction with apoAll and localization to the nucleus [440]. Association between apoAll and HCV core may be involved in the modulation of lipid metabolism by core protein, thus playing a role in HCV pathogenicity.

13.a.5: p53

Results showed that core was able to enhance the gene transactivation activity of exogenous p53 in p53-negative Hep3B cells as well as endogenous p53 in Hep G2 cells [441]. Enhancement of transactivation by p53 in the presence of core resulted in enhanced expression of the downstream p53 effector gene, the cdk inhibitor $p21^{waf1/Cip1/Sdi1}$. Further, this increased expression of $p21^{waf1/Cip1/Sdi1}$ can suppress the cell cycle and, in the presence of core protein Hep3B cell growth was suppressed in a p53-dependent way. This dependent suppression of hepatocellular growth may have important implications in HCV pathogenesis. The binding requires residues 1-151 of core and residues 366-380 of p53 [441].

13.a.6: 14-3-3ε protein

The 14-3-3 protein family associates with components of several signal transduction pathways such as the Raf-1 kinase cascade. An interaction between core protein and the epsilon isoform of 14-3-3 protein (14-3-3ε) has been observed [442] with the involvement of residues 49-97 of core protein and residues 165-234 of 14-3-3 protein. In HepG2 cells expressing core, phosphorylation of serine-53 of core is essential for interaction with the cellular protein, thus, suggesting that the interaction may occur in a phosphor-serine dependent manner. Activation of Raf-1 kinase, a central component of the mitogen-activated protein (MAP) kinase pathway,

requires this interaction. Interestingly, an enhanced MAP kinase activation in HCC [443] is reported, suggesting that the core may play a role in the progression of HCV to HCC [442].

13.a.7: p21$^{Waf1/Cip1/Sdi1}$ (p21)

It was suggested that core protein may be involved in de-regulation of the cell cycle via interaction with the cell cycle regulator p21 [444]. The interaction between core and p21 involved residues 24-52 of core protein and residues 139-164 of p21. As p21 is a regulator of cell cycle, interaction with core protein may inhibit nuclear transport of newly synthesized p21 and its subsequent involvement in cell cycle regulation [444].

13.a.8: Leucine Zipper Protein (LZIP)

LZIP is a transcription factor that has been shown to bind HCV core protein [445]. Exogenously expressed LZIP was shown to localize to the nucleus in HepG2 and HeLa cells and was relocalised from the nucleus to the cytoplasm, once it is co-expressed with HCV core protein. Loss of LZIP function (by over-expression of a transcriptionally incompetent LZIP mutant) resulted in loss of contact inhibition in NIH3T3 cells, resulting in dense foci of cells. This dysfunctioning of cell growth was enhanced in the presence of core protein, suggesting that by sequestration of LZIP, core protein can prevent LZIP function and act as a co-factor in cell transformation [445].

13.a.9: Complement Receptor gC1qR

It was identified as a core-binding protein by screening of human lymphocyte-expression library [446] and it was confirmed that the residues required for interaction include amino acids 26-124 of core and 188-259 of gC1qR. Binding of gC1qR to its natural ligand and complement protein C1q, specifically inhibits T-cell proliferation in a dose-dependent manner [447]. Interestingly, naked core protein has been shown to circulate in the plasma of HCV-infected patients [448] and be secreted from transfected cell lines [440]. Furthermore, it was showed that core protein could specifically inhibit T-cell proliferation in a dose-dependent manner and that inhibition could be blocked by either anti- gC1qR or anti-core antibody [446]. This inhibition of T-cell proliferation by circulating naked core protein may be a mechanism by which the virus evades the immune system and establishes persistence.

13.a.10: p73

Core protein has also been shown to bind another member of the p53 superfamily, known as p73 [449]. This protein transactivate endogenous targets of p53 such as the $p21^{Waf1/Cip1/Sdi1}$ promoter (Jost et al., 1997) and core protein can modulate the transcriptional acitivity of p73 on the p21 promoter[449]. Residues 321-353 of p73 are required for interaction. However, the core protein is able to inhibit p73 dependent cell growth arrest in HepG2 cells, suggesting an involvement of this interaction in the pathogenesis of HCV [449].

13.a.11: Sp110b

Sp110b was identified as a core-interacting protein during a study on the molecular mechanisms behind core modulation of all-*trans*-retinoic acid (ATRA)-induced cell death [450]. Within the nucleus, Sp110b acts as a transcriptional corepressor of the retinoic acid response element (RARE) and prevents the enhancement of downstream proapoptotic gene expression [450]. For interaction with core, residues 21-80 of core and 389-453 of Sp110b are needed. Core protein is responsible for the transfer of Sp110b from the nucleus to the cytoplasmic surface of the ER, thus releasing the suppressive function of Sp110b and activating RARα-mediated transcription and ultimately leading to ATRA-induced cell death [450].

13.a.12: DEAD-box RNA Helicase (DDX3/CAP-Rf/DBX)

Finally, core protein has been shown to bind a DEAD-box RNA helicase termed DBX/DDX3/CAP-Rf [451, 452]. The interaction between HCV core and DDX3 was discovered using yeast-2-hybrid screening and confirmed by both *in vitro* binding studies and immunofluorescence analysis. The core-binding domain of DDX3 was mapped to the C-terminal 409-622 residues [451, 452]. DDX3 has been suggested to be involved in translation. Moreover, interaction of DX3 with core protein may inhibit translation of capped mRNA [451].

13.b: Possible structural role of the Core Protein in nucleocapsid formation

Several studies have indicated an interaction between core and the HCV 5'UTR. Residues 1-75 of core, containing 4 clusters of basic acids, have been shown to have RNA binding capacity [15], while a number of groups have identified HCV RNA sequences capable of binding core [433]. Interestingly, Surface Plasmon resonance (SPR) was used to confirm a stable interaction between core and the HCV 5'UTR and it was observed that core

preferentially bind structural sequences with high G content. On the other hand, residues 1-20 of core have been shown to inhibit translation from the HCV IRES [453], suggesting that core may be involved in the switch from translation of RNA to replication and resultant packaging of viral RNA.

Additionally, if core is multimerizing to form the capsid, it would require homotypic interactions. This has been shown by using various studies, which showed that core contains many homotypic interaction sites [426, 454]. The hydrophilic region encompassing amino acids 1-117 of core, was shown to be sufficient for core-core interaction, however no interaction was detected using full-length core protein. This could be due to the effect of the hydrophobic domain of core that affects the nuclear transport of the fusion proteins [425]. These results also suggest that the homotypic interacting domain may be masked in full-length core and some conformational change is required to allow core-core interaction. Interestingly, this notion was validated soon after a conformational change in core has been identified upon interaction with tRNA [455]. The N-terminus of core is proteinase-sensitive and the C-terminal end is required to stabilize the free core protein. However, core complexed with RNA or assembled into nucleocapsid particles was proteinase-resistant, suggesting conformational changes in core occur upon binding RNA and assembling into nucleocapsids, in order to possibly protect core from cellular degradation [455].

13.c: Possible Pathogenic Roles of the Core Protein

13.c.1: Effects on Apoptosis

Expression of HCV core affects apoptosis, or programmed cell death. Apoptosis is an orderly cellular process resulting in cell death via several morphological phases including cell shrinkage, nuclear condensation, blebbing of the plasma membrane and 'leddering' of genomic DNA by digestion [17]. Apoptosis can be due to a number of stimuli including oxidative stress, heat shock, ionizing radiation, cytokines and viral infection, which normally occurs via cell surface molecules such as Fas or TNF receptors. Apoptosis results in a cascade of intracellular events including death domain clustering and activation of caspases, which are responsible for many of the morphological changes during apoptosis.

Apoptosis is modulated by core protein via 3 receptor pathways, *Fas* receptor-mediated, TNFα receptor-mediated and lymphotoxin β receptor-mediated apoptosis.

There have been conflicting reports on the effect of core protein on *Fas*-mediated apoptosis. On one hand, it was shown that the interaction between anti-*Fas* antibody with the *Fas*

receptor induced marked apoptosis in the presence of core expressed in HepG2 cells [456]. And on other hand, core protein was also shown to protect HepG2 cells from *Fas*-mediated apoptosis induced by the same anti-*Fas* antibody [457]. This contradiction can be explained on basis of the use of clonal cells lines stably expressing core [456], as opposed to transient transfection of the full HCV open reading frame[457].

Like Fas-mediated apoptosis by core, contradictory results have also been reported for TNFα-mediated apoptosis. Transiently transfected core protein was initially shown to inhibit TNFα-induced apoptosis in MCF7 cells [458] and protect HepG2 cells from TNFα-mediated apoptosis [457]. Core was also shown to activate NF-κB and activation was enhanced upon stimulation by either anti-*Fas* or anti-TNFα antibody, thus preventing apoptosis [457]. However, core protein was also shown to sensitize HepG2 and HeLa cells to TNFα-induced apoptosis [438] by recruiting FADD to TNFR1 [459]. Residues 1-115 of core were identified as the binding region for TNFR1, while the death domain within the cytoplasmic tail of TNFR1 was shown to be required for interaction with core. Since TNFR1 has no catalytic activity, the death domain is thought to act as a signal transducer in the apoptotic pathway. Further, upon oligomerisation of TNFR1, TRADD subsequently recruits FADD. In turn, FADD then interacts directly with downstream apoptotic proteases leading to apoptosis.

However, this work was later contradicted since TNFα-induced apoptosis was found to be inhibited by core protein and the core does not interact with TNFR1 or TRADD [460]. This could be possibly due to the fact that core maintains a certain level of c-FLIP, a dual function regulator for caspase-8 activation and CD95-mediated apoptosis, that acts as an apoptosis inhibitor only at high ectopic expression levels. This high expression of c-FLIP inhibits the cleavage and activation of caspase-8 which resulted in inhibition of apoptosis [460]. Again the explanation for this contradiction could be in the use of different cell lines or mode of core expression.

Finally, core protein can enhance LT-βR-mediated apoptosis in a cell-type dependent manner. An enhancement of LT-βR-mediated apoptosis was observed in HeLa cells expressing core, while it was not observed in HepG2 or Huh-7 cells [436].

13.c.2: Effects on Cell transformation

Primary rat embryo fibroblast (REFs) studies have shown that HCV core, can convert cells to a transformed phenotype. Cells showed rapid proliferation, anchor-independent growth and tumor formation in athymic mice [461] as well as in REF rat-1[462]. However, core could

perform this effect only in combination with the oncogene *H-ras*. Additionally, core alone was shown to transform established NIH3T3 cells, when these cells were injected into nude mice. The transformation resulted in anchor-independent growth and tumorigenicity [461, 463]. This transformation was the result of core interacting with signal transducer and activating (through phosphorylation) the transcription 3 (STAT3) activator [464]. Core, in combination with H-*ras*, can promote anchorage-independent growth of *BALB/3T3 A31-1-1* cells by stimulating the activation of the *Erk1/2*-serum response element pathway. This pathway is involved in signal transduction of growth stimuli [465]. However, abnormal cellular proliferation is also seen due to the effects of core protein on the transcription factor LZIP. In the context of core protein, LSIP presents activities consistent with that of a tumor suppressor. LSIP is a nuclear CRE-activation factor whose transcriptional activity is repressed by core. However, subcellular sequestration of LSIP by core results in inactivation of LZIP and subsequent abnormal cellular proliferation [445].

13.c.3: Effects of Transcription and Regulatory Factors

Core protein can modulate expression from cellular and viral promoters [17]. It was observed that HCV core protein could suppress expression and replication of hepatitis B virus (HBV) in Huh-7 cells. This was evident between 6- 9 days post-transfection and also coincided with intracellular relocalization of HCV core from the cytoplasm to the nucleus [466]. Interestingly, HCV core protein may trancriptionally regulate other viruses since it transactivates expression from the HIV-1 long terminal repeat (LTR) and from the SV40 promoter. Core also suppresses expression from the rous sarcoma virus (RSV) LTR [467].

However, contradictory effects of core were reported on p21 which is a promoter of a negative regulator of cell cycle progression. At first, it was reported that core suppresses p21 promoter activity by enhancing cell growth [468]. This activity of core was observed during liver regeneration. On the contrary, an enhancement of activity from the p21 promoter by core protein has also been shown [441].

This discrepancy was later solved by showing that core affected p21 promoter activity in different ways that depend on the used cell lines [469]. It was observed that the core transactivated the p21 promoter activity in HepG2 and Hep3B cells while it suppresses the expression in NIH3T3 cells and primary hepatocytes from transgenic mice [441, 468].

Furthermore, it has been reported that core may counteract the antiviral effect of IFN by suppressing the activity from the promoters of IFN effector protein MxA, PKR and 2'-5' oligoadenylate synthetase (2'-5' OAS) [470].

13.c.4: Effects on Immune Presentation

Since HCV establishes chronic infections in the majority of cases; hence, the virus must be able to avoid clearance by the immune system. In this regard, core protein may play a role by inhibiting the immune response to infection. Cytotoxic T cells are primed against HCV as a result of interaction between antigen-presenting cells (APC), HCV antigen and the T cells themselves. APCs are a site of HCV replication [471] and have also been studied to analyze the role of HCV proteins and stimulation of cytotoxic T cells during immune presentation.

Expression of HCV protein core, E1 and E2, or core protein alone in T dendritic cells shows that the latter possess reduced stimulatory capacity for cytotoxic T cells. This was followed by the production and release of lower levels of T cell stimulatory cytokine IL-12 [472, 473]. Moreover, addition of exogenous IL-12 to dendritic cell/T cell cultures did not restore T-cell proliferation. This indicated that lack of IL-12 is not the main cause of abnormal T cell priming [473]. Moreover, studies in mice using recombinant vaccinia virus showed that those expressing core protein had elevated virus titer 5 days post-infection compared to those expressing HCV non-structural proteins. In addition, mice expressing non-structural proteins overcame infection while mice expressing core protein succumbed to lethal infection [474].

This failure to clear vaccinia virus infection in the presence of HCV core protein, coined the theory that core protein plays a role in inhibition of the cytotoxic T cell response. Moreover, interferon-γ production by immune splenocytes was also profoundly suppressed in the presence of core protein. This lack of interferon-γ production was also accompanied with the reduction in the production of antigen-stimulated IL-2 [474].

Next, the core protein also modulates the immune response by interfering in the complement pathway [446]. This was evidenced by the binding of complement protein C1q to its receptor, gC1qR. This interaction blocked the proliferation of cytotoxic T cells [447, 446]. Additionally, core protein exposure inhibits T cell proliferation in a dose-dependent manner [446], due to inhibition of activation of the ERK/MER MAP kinase signaling pathway [475]. Inhibition of activation of the ERK/MER MAP kinase by C1q/gC1qR interaction inhibits transcription of early genes involved in T cell activation (such as IL-2). This inhibition, finally, leads to suppression of proliferation. Production of IL-2 is also inhibited in the presence of core

protein, suggesting that core may inhibit T cell proliferation by inhibiting the ERK/MER MAP kinase signaling pathway [475].

13.c.5: Effects on Lipid Metabolism

HCV core protein localizes to the surface of LDs and has been suggested to regulate expression of cellular genes involved in lipid metabolism [16]. This has been shown by the stable expression of core protein in both chimpanzees and HepG2 cells [16], as well as in transgenic mice [206].

Transgenic mice presented steatosis at 2 months that coincided with an increased frequency of large (in comparison to small) LDs. These mice were expressing HCV core from birth. Thus, the coinciding of steatosis with frequency of LDs, suggested a role for core in steatosis. This role of core could be dependent on the binding of enzymatic molecules or apolipoproteins involved in lipid metabolism [206]. Interestingly, core was shown to colocalize with apolipoprotein II on LDs within HepG2 cells [16]. However, comparison of lipid levels and fatty acid composition of lipids in core-expressing transgenic mice, non-transgenic mice and non-transgenic obese mice indicated that triglyceride levels in the livers of transgenic mice and non-transgenic obese mice were significantly higher than compared to non-transgenic mice. On the other hand, the concentration of 18 carbon mono-unsaturated fatty acids (oleic and vaccenic acids) was increased only in transgenic mice. This suggested that HCV core may affect a specific pathway in lipid metabolism.

Similar results on lipid composition were also observed in human livers [476]. The core protein impaired hepatic assembly and secretion of triglyceride-rich, very low-density lipoproteins (VLDL) [477]. Additionally, a marked reduction in number of normal sized lipoprotein particles was observed in core expressing transgenic mice. Microsomal triglyceride transfer protein (MTP) and apolipoprotein B (apo B) are major regulators of VLDL assembly. Transgenic mice expressing core show significantly lower MTP activity compared to non-transgenic mice, suggesting that core protein may impair secretion of VLDL by decreasing MTP activity and thus VLDL assembly [477].

13.d: Possible role of Core Protein as a nucleic-acid Chaperone

Several hundred molecules of the structural core protein are thought to coat the genome in the viral particle, as do the nucleocapsid (NC) protein molecules in Retroviruses, another class of enveloped viruses containing a positive-sense RNA genome. Retroviral NC proteins possess

nucleic acid chaperone properties that play critical roles in the structural remodelling of the genome during retrovirus replication [478-484]. This analogy between HCV Core and retroviral NC proteins provided an indication that HCV Core protein could be a putative nucleic acid chaperone.

13.d.1: Nucleic Acid Chaperone

Nucleic acid molecules (*Ribonucleic Acid (RNA)/Deoxy-ribonucleic acid (DNA)*) face difficulties when folding into their native structures. They become easily trapped in inactive conformations, because of structural and functional flexibilities. It was postulated that these unproductive misfoldings of nucleic acids can be solved by specific proteins [25, 485]. These proteins can assist Nucleic acids (NAs) in reaching their functionally active states by binding and stabilizing a specific structure or, in a quite opposite way, by interacting in a non-specific manner [486]. These proteins can either facilitate RNA-RNA, RNA-DNA or DNA-DNA interactions in a reaction termed nucleic acid annealing, or they can resolve non-functional inhibitory structures. The latter is defined as "RNA chaperone activity" and related proteins are defined as chaperone proteins. RNA chaperone activity entails the disruption of interactions between two NAs and the loosening of their structures. The interaction with these proteins is needed for the unfolding of the RNA but not to maintain its structure [486]. These proteins do not require ATP-binding or hydrolysis for their activity. They have been assigned distinct roles in diverse cellular processes such as the regulation of transcription, RNP assembly and stabilization, RNA export, virus replication and histone-like nucleoid structuring.

NA chaperone activity cannot be predicted based on the protein domain structure or the existence of discrete motifs. Nevertheless, it was observed that NA chaperones have the highest frequency of disordered regions. It has been proposed that they act according to an entropy transfer model, allowing correct RNA folding by successive cycles of protein–substrate order–disorder (Figure 38) [487-489].

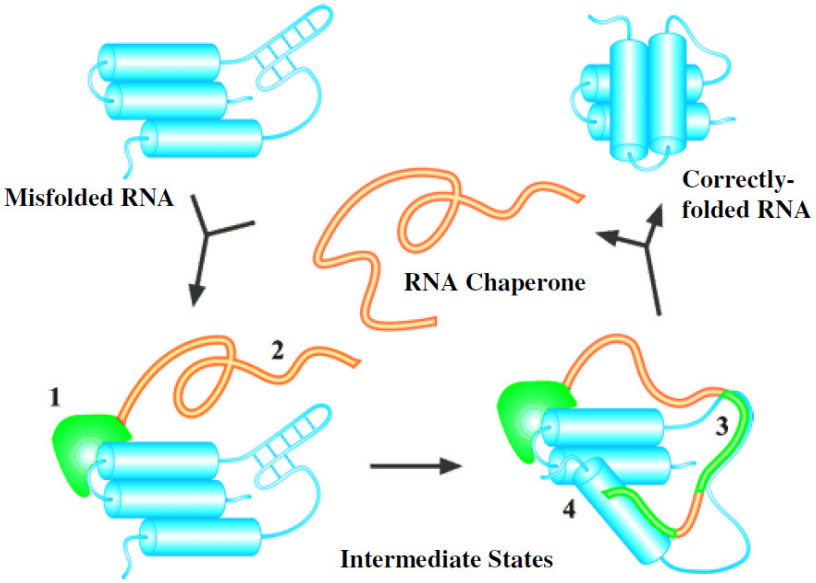


Figure 38: Entropy transfer model of NA chaperone mechanism (based on [489]). The entropy transfer model indicates the possible role of structural disorder in chaperone function. The model incorporates most of the mechanistic details described for the action of disordered proteins/regions in chaperones; making no distinction between RNA and protein chaperones, as their critical mechanistic features are similar. Decisive elements are as follows. (1) A fully disordered chaperone (orange) binds a partially misfolded substrate (RNA or protein, blue) in a relatively nonspecific manner. The recognition segment locally folds (illustrated with green) and anchors the chaperone to the substrate. Naturally, this binding element may already be folded prior to binding (2) A disordered appendage projects away from the substrate and provides a solubilizing effect due to its highly hydrophilic character and entropic exclusion of other molecules. This disordered segment thus prevents aggregation of the substrate. (3) A disordered segment contacts the misfolded part of the substrate. As a result, part of the chaperone becomes ordered (green) whereas the substrate becomes disordered (i.e., locally unfolds due to the reciprocal entropy transfer process). Of course, this secondary binding need not be mechanistically separated from the primary binding event (1), as entropy transfer may already take place upon primary recognition of the substrate. (4) In this state of increased flexibility, the substrate is facilitated in its search through conformational space toward the native conformation by keeping its unfolded segments in close proximity by the folded chaperone. The chaperone may contact and release the substrate several times in rapid succession until it finally releases the properly folded substrate to resume the catalytic cycle.

The number of proteins with NA chaperone activity is steadily growing [490]. There is no consensus on the definition of NA chaperone activity or the minimum assays required to

establish the nucleic acid activity of the protein. In the next few paragraphs, we will discuss the assays that have been used to define a protein as a chaperone protein.

13.d.1.1: Chaperone assays

A number of assays, with different technical difficulty, have been used to analyze the RNA chaperone activity of a protein [486, 490]. Here, we will discuss different assays used in *in vitro* and *in vivo* conditions to analyse the NA chaperone activities of these proteins.

In vitro assays

In vitro assays employ model RNA substrates or ribozyme systems, because RNA chaperones typically show non-specific RNA binding properties. These assays are well suited for the general assessment of a candidate protein and to search for RNA chaperone activity mutants. These assays could be divided into i) simple assays, with relatively low technical difficulty as they are focused on a single aspect of chaperone activity (i.e., annealing, destablization and strand transfer assays), and ii) advanced assays, in which several chaperone activities are analyzed simultaneously (i.e., ribozyme or intron cleavage) [486, 491, 490]. Although there is no consensus, at least two different positive assays should be required to establish the RNA chaperone activity of a protein. Additionally, protein should be released after its acts as nucleic acid chaperone, because this is the main difference between an RNA chaperone and an RNA binding protein that stabilizes a determined RNA folding while remaining bound.

In certain cases, biophysical techniques, such as fluorescence resonance energy transfer (FRET), NMR spectroscopy, fluorescence correlation spectroscopy (FCS), or single-molecule spectroscopy (SMS), and Surface Plasmon resonance (SPR) have been applied to investigate the chaperone mechanism in *in vitro* assays providing precise information on the RNA chaperone mode of action [480, 492-494, 481, 491, 495-497].

Short RNAs are generally employed to monitor the RNA chaperone mediated helix destabilization of double-stranded RNAs. This destabilization can result in strand dissociation, unwinding and strand displacement (Figure 39A) [26, 498-501, 483, 502]. Similarly, the annealing of RNAs of varying length, sequence and structure has been used to examine the RNA chaperone activity of proteins. To identify chaperone proteins, two strategies have been developed: (i) use of highly structured nucleic acid molecules. These NAs require a previous unwinding step by the RNA chaperone prior to the annealing, or (ii) protein elimination after base pairing has taken place (i.e., by treatment with proteinase K).

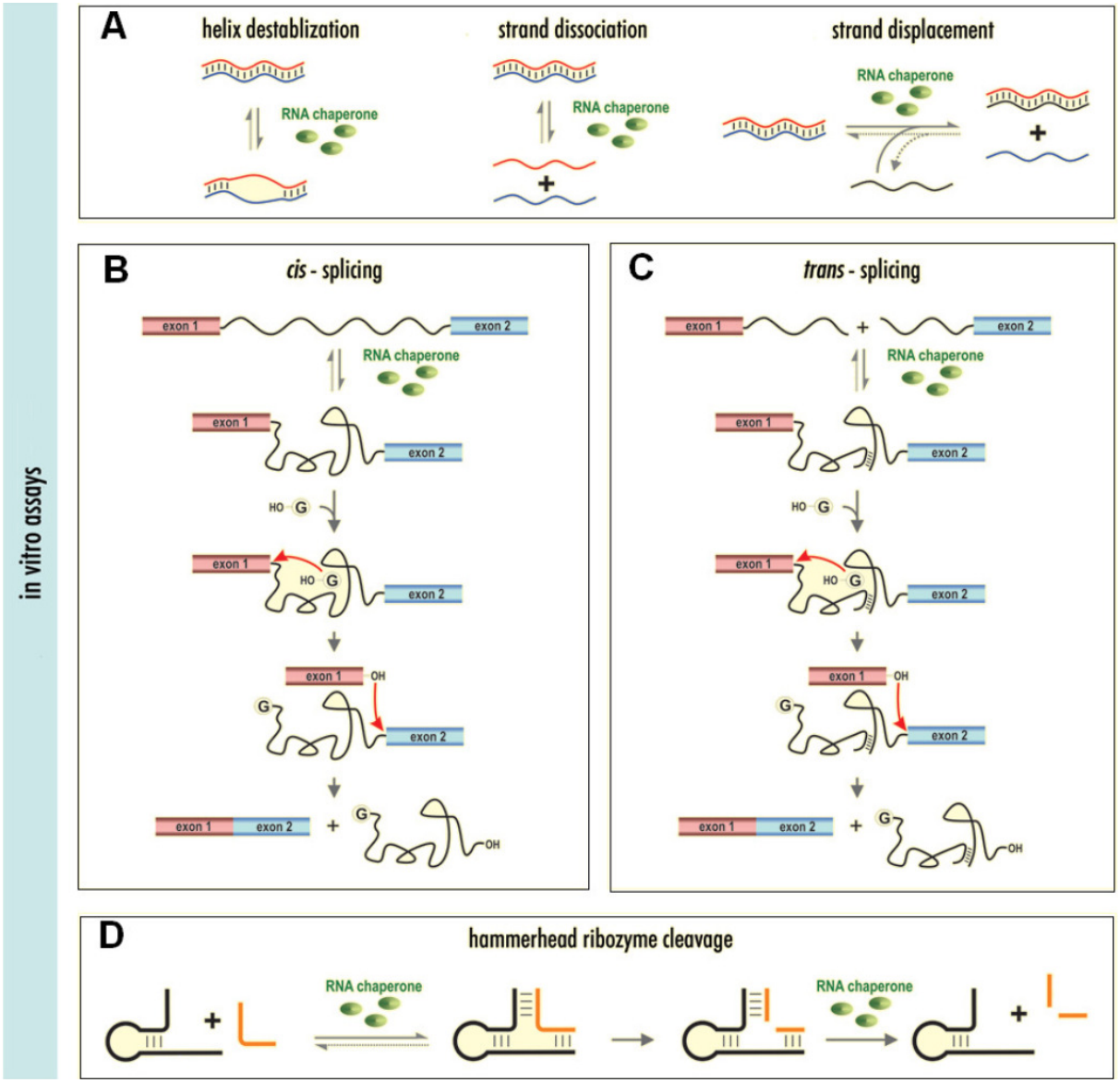


Figure 39: *In vitro* Assays for RNA chaperone activity. (A) RNA chaperone activity (RCA) on double-stranded RNA can result in helix destabilization (duplex unwinding), in full strand dissociation (RNA melting) or—in the presence of a competitor RNA—in strand displacement (strand exchange). (B) Pre-RNAs containing the thymidylate synthase group I intron have to fold correctly to undergo splicing. For the cis-splicing assay, the purified transcript is folded by heat-renaturing, and the reaction is initiated by addition of a guanosine cofactor. Proteins with RCA significantly increase the population of molecules with a catalytically active, splicing competent conformation. (C) To further strengthen the requirement for RNA-RNA interactions and correct folding, the pre-RNA can be transcribed in two pieces, which have to anneal for the formation of an active ribozyme. Trans-splicing is initiated by an exogenous guanosine cofactor that is ligated 5' to the 5'-intron part in the first transesterification step. The second reaction step leads to ligated exon and intron release.

Proteins with RCA facilitate this reaction, especially at low temperatures. (D) The hammerhead ribozyme assay monitors the cleavage of a substrate RNA that has to anneal to the ribozyme. RCA may resolve misfolded hammerhead ribozyme complexes and is beneficial for the release of the cleaved product and for multiple turn-over. Thus, two activities—annealing and strand displacement—can be traced in one assay.

To monitor complex refolding events, group I intron cis- or trans-splicing or hammerhead ribozyme-mediated RNA cleavage can be assessed. Group I intron splicing is a special case of splicing that may occur within the introns of some organisms, in the absence of proteins. It depends on the folding of the RNA in a defined three-dimensional structure, as these types of introns are able of self-splicing. Proteins with RNA chaperone activity increase the fraction of RNA molecules in the native, catalytically active conformation (Figure 39B–D) [503-507, 25, 500, 508-510, 502]. Another assay called template switch has also been used to study the RNA chaperone activity [511-513, 499, 494, 514]. In this assay, the synthesis of different cDNA species is analyzed. This includes generation of a cDNA by a switch from the donor RNA template to an acceptor RNA. This assay is stringent since it requires unwinding of very stable secondary RNA structures, efficient RNase H activity, inhibition of self-priming, and annealing of separate RNAs.

In vivo assays

However, comparing the activity of different proteins in a single assay harbors potential pitfalls, since typically purified proteins are tested with a unique set of RNAs and in the absence of biological partners. Therefore, proteins have also been tested in vivo, mainly by the "folding trap assay," and by the "transcriptional assay". The folding trap assay examines the RNA chaperone dependent group I intron splicing in the cell [515, 516] (Figure 40A) while the transcriptional assay observes RNA chaperone assisted destabilization of a terminator stem [517] (Figure 40B).

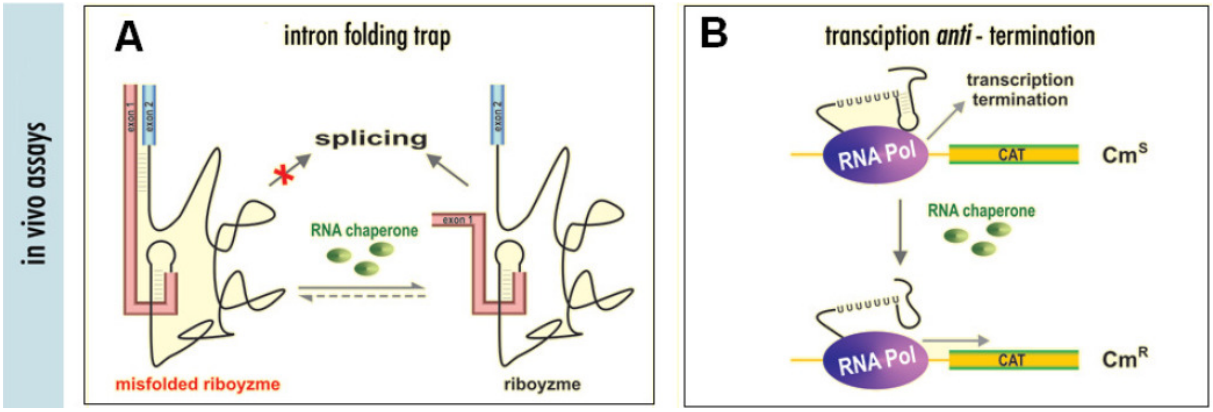


Figure 40: *In vivo* Assays for RNA chaperone activity. (A) For the folding trap assay, a mutant of the thymidylate synthase gene containing a stop codon upstream from the 5' splice site in exon 1 pre-RNA is expressed within E. coli cells. The premature stop codon prevents the ribosome from resolving an aberrant base pairing between 3'-terminal intron and exon 1 sequences, which precludes the folding of a native intron structure. Co-expression of RNA chaperones partly alleviates this splicing deficiency by resolving the misfolded structure. The RNA then folds correctly and the intron splices. (B) Transcription termination can be caused by the formation of a stem-loop in the nascent transcript, followed by a poly(U) stretch. This prevents the RNA polymerase from reaching a downstream reporter gene (chloramphenicol acetyl transferase, CAT), making the cells chloramphenicol sensitive. With RNA chaperones overexpressed in the cell, the terminator stem is 'melted', CAT is transcribed and the cells become chloramphenicol resistant (CmR).

The first system described was based on an in vivo RNA folding trap existing in the bacteriophage T4 thymidylate synthase (td) group I intron [515]. Several RNA chaperones, overexpressed in *Escherichia coli*, are able to resolve the kinetic trap or impede the formation of misfolded structures, therefore promoting intron splicing. This system has been successfully used to test the nucleic acid chaperone activity in vivo [503, 515]. Nevertheless, its physiological relevance remains uncertain [485]. It also has a great limitation when overexpression of the RNA chaperone is toxic for *E. coli*.

Other reported assays, with more physiological relevance are based on: (i) transcription anti-termination *in vivo*, for *E. coli* CspA-family nucleic acid chaperone proteins [518]; (ii) *Neurospora crassa* group I intron splicing by CYT-19 protein, data that were also supported by the splicing phenotype of several mutants [519]; (iii) in vivo hepatitis delta ribozyme

activity; and (iv) in vivo reverse transcriptase template switching assay, using a MLV-based retroviral vector.

Nevertheless, the analysis of RNA chaperone activity *in vivo* is quite difficult mainly because of the pleiotropic effects that RNA chaperones could have in the cell. Additionally, RNA chaperones act in different steps of a single biological process. All these reasons are responsible for the lack of appropriate *in vivo* systems for the analysis of RNA chaperone activity.

13.d.2: HCV core protein as NA chaperone

The RNA chaperone activity of core has been demonstrated using different approaches [26]. The core protein enhances hybridization of complementary ODNs and/or RNAs and allows the formation of the most stable structure by strand exchange. Moreover, the HCV core protein shows a broad range of sequence specificity, and once nucleic acid molecules have been refolded, the core is no longer required to maintain the new conformation. This confirms its bona fide chaperone function. In addition, based on disorder conservation in *Flaviviridae* core proteins, the RNA chaperone activity of core proteins from different members of each genus has been analyzed. It has been found that, although the core proteins from different *Flaviviridae* genera have low sequence similarity, all of them have RNA chaperone activity [500]. However, it was also suggested that remodelled nucleic acids in a nucleoprotein complex with core proteins, remain accessible to other processes and are not hidden in a protected inactive state.

Furthermore, the N-terminal basic amino acid clusters of core protein were found to retain its nucleic acid chaperone activity and induce dimerization of the HCV UTR RNAs (Figure 41) [24].

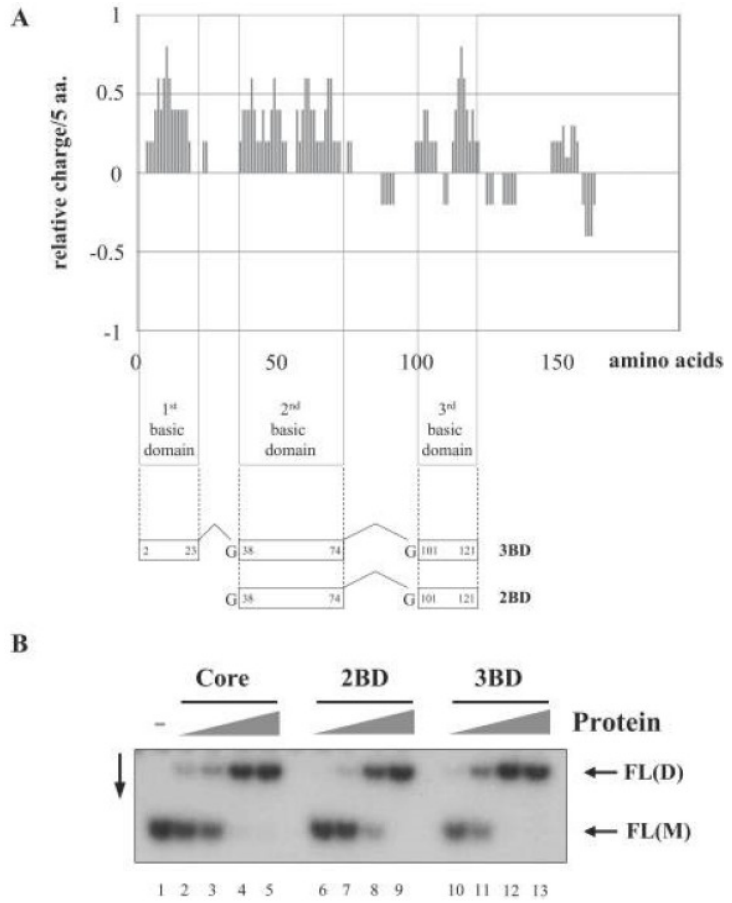


Figure 41: Activity of core peptides in RNA dimerization. (**A**) The bar chart illustrates the charge distribution of amino acids in the core protein, as calculated by the charge function of the EMBOSS package, using default parameters and a sliding window of five amino acids. Synthetic peptides 2BD (BD1+BD2) and 3BD (BD1+BD2+BD3) correspond to two or three basic amino acid clusters of the core protein, respectively. (**B**) Dimerization of FL(+) RNAs induced by core peptides 2BD and 3BD. ^{32}P-labelled RNA molecules were incubated with increasing amounts of the corresponding polypeptides (at polypeptide to nucleotide molar ratios 1:80, 1:40, 1:20 and 1:10). Following incubation, polypeptides were removed and RNA were analyzed by native gel electrophoresis and autoradiography. FL(D) corresponds to dimeric RNA, while FL(M) to the monomeric form [24].

However, it has been already shown that the N-terminal part of the protein is sufficient for RNA-binding [15] and particle formation [427, 432, 428]. Moreover, the uncharged regions connecting the basic clusters of N-terminal domain of core were found to be dispensable for assembly [427]. This indicated that RNA chaperoning and capsid assembly require the same basic regions of core protein.

Binding of core protein to the genomic RNA of HCV triggers important structural rearrangements by inducing the dimerization of the plus-strand 3′-UTR as well as that of the complementary negative-strand RNA region [26, 24]. This dimerization is supposed to be mediated through a 16 nt long palindromic sequence (residues 23-38) within the SL2-loop in the X-tail region, called the dimer linkage sequence (DLS). This RNA motif is universally present in the 3'X-RNA-Tail region of all HCV sequences reported so far, thus suggesting that it plays an important role in the viral life cycle. Moreover, dimerization of HCV RNA could facilitate the release of the (+) RNA from the translation machinery and allow its encapsidation in the viral particle. However, by homology with the crucial role of dimers in RNA recombination in retroviruses [28], dimerization of the X region may promote the hybridization of genomic subtypes. Thus, it would be one of the factors that, concomitantly with the lack of fidelity of the RNA dependent RNA polymerase activity, promote the production of a heterogeneous population of virus responsible for the impaired immune response during infection [29, 27, 30]. Due to dimerisation, HCV RNA 3′ UTR adopts different conformations during the viral life cycle. The interconversion between different structures could regulate transitions between translation and replication as well as between replication and packaging of the genomic RNA [11, 177]. In this context, the RNA chaperone activity of the core protein could regulate these riboswitches by facilitating interconversion of the different RNA structures [24].

In addition to HCV core, a numerous other cellular and viral proteins have been described to bind to the X RNA region of the HCV genomic RNA. This list includes the polypyrimidine-tract binding protein, several ribosomal proteins, and the viral non-structural NS3 and NS5B proteins [123, 173, 186, 188]. Furthermore, the actual conformational state of X RNA and the resulting function probably reflect the complex interplay of these proteins.

Research Aim

The mechanism of chaperone activity of HCV core protein in the dimerization of (+) and (-) HCV 3' UTR is still unknown. The chaperone activity typically comprises three phases: i) the binding of protein to nucleic acids, ii) the possible destabilization of the secondary structure of ODNs and iii) the stimulation of intermolecular hybridization of the complementary nucleic acid sequences.

These three components of the chaperone activity of the core can be studied by spectroscopic techniques at sub-saturating concentrations of peptide, where aggregation is limited and chaperone activity is optimal.

So, the objective of this research is to characterize and quantify the molecular chaperone activity of the HCV core protein by using fluorescence spectroscopy techniques and Surface Plasmon Resonance (SPR).

By using the native D1 domain and peptides derived from this domain, we will first characterize the binding parameters and the conformational changes associated with the binding of these peptides to the native and mutated sequences from HCV 3' UTR sequences.

Next, we will investigate the destabilization of model and HCV ODNs by the D1 domain and its mutants.

Further, we will describe the molecular mechanisms of the core chaperone properties by characterizing the hybridization kinetics by using various mutated HCV and model oligonucleotides.

This study should enable us to better understand the role of the core in the encapsidation step, the synthesis of the complementary strand of the genomic RNA and the recombination mechanisms participating to the genetic variability of the virus. Accordingly, we could thus suggest new strategies to identify by screening or rational designing, inhibitors targeting the interaction of the core protein with its nucleic targets.

Lastly, we will extend our findings by revealing molecular mechanisms and structural determinants of other chaperone proteins like the Human prion protein, Tat and NCp7 from HIV-1 virus and Nucleocapsid protein from Bunyamwera Orthobunyavirus (Appendix part).

Materials and Methods

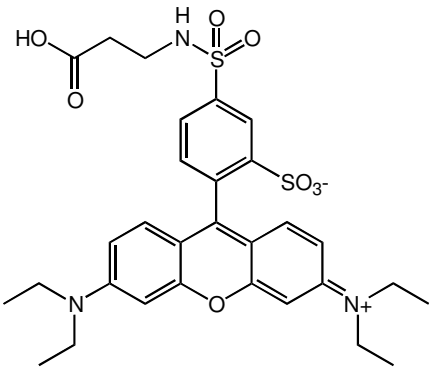
1: Materials

1.1: Peptide Synthesis

Peptides were synthesized by solid phase peptide synthesis on a 433A synthesizer (ABI, Foster City, CA). The synthesis was performed at a 0.1 mmole scale using the standard side-chain protected fluorenylmethoxycarbonyl (Fmoc)-amino acids and HBTU/HOBt coupling protocol. HMP resin (ABI, 0.44 mmol/g reactive group concentrations) was used as solid support. Deprotection steps were performed by piperidine and automatically controlled by UV absorbance. At the end of the synthesis, peptidylresin was isolated and twice washed by NMP.

1.1.1: Labeling

Labeling procedures were performed in flasks. Four equivalents (0.4 mmole) of the label (Lysamine) were dissolved in 1 ml of NMP mixed with four eq. of HBTU/HOBt coupling solution (in DMF) and added to Fmoc-deprotected peptidylresin (0.1 mmole) swelled in 1 mL of NMP. After a few minutes of shaking, five equivalent of DIEA solution was added. Then the reaction mixture was stirred at 40°C for overnight. Resin was filtrated and washed by NMP. In the case of lipophilic peptides, this procedure was repeated, while in case of other peptides, the resin was washed by methanol and DCM, dried and weighted.



MW = 630 Rh(ba)-COOH

Figure 1: Chromophore (Lysamine = Rh-) used to label core peptides

1.1.2: Cleavage and deprotection

Cleavage and deprotection of peptidylresin were performed for 2 h using a 10 ml trifluoroacetic acid (TFA) solution containing water (5%, v/v), phenol (2%, w/v), thioanisole (5%, v/v) and ethanedithiol (2.5%, v/v). The solution was concentrated *in vacuo* and the peptide was precipitated by using ice-cold diethyl ether and then pelleted by centrifugation. The pellet was washed with diethyl ether and dried.

1.1.3: Purification

Before purification, the peptides were solubilized with aqueous TFA (0.05 %, v/v). In the case of lipophilic peptides, the necessary amount of acetonitrile was added. HPLC purification was carried out on a C8 column (Uptisphere 300A, 5µm; 250X10, Interchim, France) in water/acetonitrile mixture containing 0.05% TFA with linear gradients depending on the peptide (typically 20 to 50% of acetonitrile for 30 min) and monitored at 210 nm (detection of all peptides including non-labeled) and 370 nm (detection of labeled peptides only). After purification, the fractions containing pure peptide were combined and lyophilized.

All peptides were characterized by ion spray mass spectrometry and the expected molecular masses were found. Prior to use, the peptides were dissolved in buffer, to be aliquoted and stored at -20°C.

1.2: Amino acid sequences of peptides used in this thesis

1.2.1: Peptides from HCV Core protein

PeptideE

R_1RGPRLGVRATRKTSERSQPRGRRQPIPKARRPEGRGRGSRPSWGPTDPRRRSRNLG KG$_{59}$

Peptide F

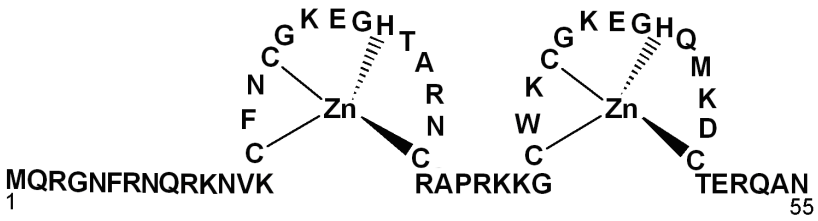
S_1TNPKPQRKTKRNTNRRPQDVKGRRGPRLGVRATRKTSERSQPRGRRQPIPKARRPE GRGRGSRPSWGPTDPRRRSRNLGK$_{83}$

Domain D1

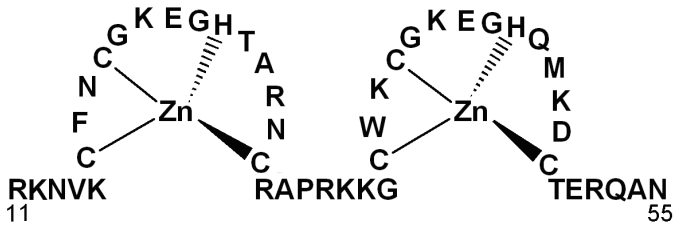
M_2STNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERS QPRGRRQPIPKARRPEGRTWAQPGYPWPLYGNEGMGEGMGWAGWLLSPRGSRPSW GPTDPRRRSR$_{117}$

1.2.2: Peptides from HIV-1 NCp7 and Tat proteins

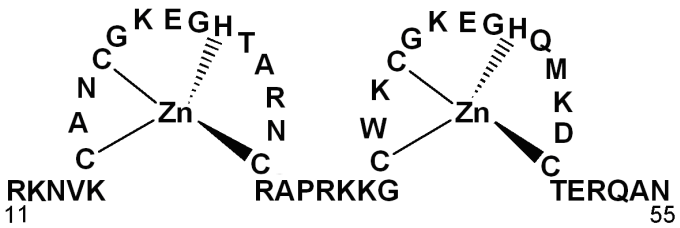
NC(1-55)



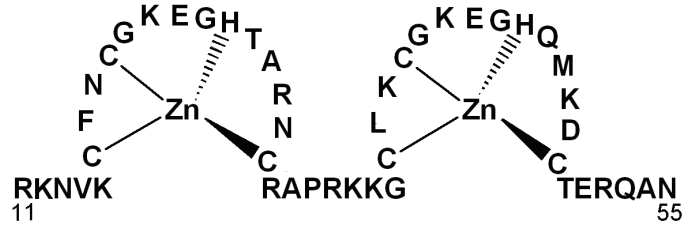
NC(11-55)



A_{16} NC(11-55)



L_{37} NC(11-55)



$SSHS_2$NC(11-55)

K_{11}NVK-SFNSGKEGHTARNS-RAPRKKG-SWKSGKEGHQMKDS-TERQAN$_{55}$

Tat(1-86)

M_1EPVDPRLEPWKHPGSQPKTACTTCYCKKCCFHCQVCFTTKALGISYGRKKRRQRR RPPQGSQTHQVSLSKQPTSQPRGDPTGPKE$_{86}$

Tat(44-61)

G_{44}ISYGRKKRRQRRRPPQG$_{61}$

1.2.3: Peptides from Human Prion Protein

(23-110) HuPrp

R_{23}PKPGGWNTGGSRYPGQGSPGGNRYPPQGGGGWGQPHGGGWGQPHGGGWGQPH GGGWGQPHGGGWGQGGGTHSQWNKPSKPKTNM$_{110}$

1.2.4: Peptides from Bunyamwera Orthobunyavirus Nucleocapsid Protein (BUNV-N)

M_1IELEFHDVA ANTSSTFDPE VAYANFKRVH TTGLSYDHIR IFYIKGREIK TSLAKRSEWE VTLNLGGWK ITVYNTNFP GNRNNPVPDD GLTLHRLSGF LARYLLEKML KVSEPEKLI IKSKIINPLA EKNGITWND GEEVYLSFFP GSEMFLGTFR FYPLAIGIYK VQRKEMEPKY LEKTMRQRYM GLEAATWTVS LTEVQSALT VVSSLGWKKT NVSAAARDFL AKFGINM$_{233}$

(provided by Jean-Luc Darlix, Lyon, France)

1.3: Preparation of the protein solutions

Proteins were lyophilized, dissolved in buffer (25 mM TRIS, 0.2mM $MgCl_2$, 30mM NaCl, pH 7.5, degassed, or 50 mM HEPES, 0.2mM $MgCl_2$, 30mM NaCl, pH 7.4, degassed) and their concentration was determined by absorption at 280 nm. The molar extinction coefficient of the proteins can be estimated by using the following formula:

$$\varepsilon_{280nm} = (n_{Trp} \times 5700) + (n_{Tyr} \times 1300) \quad \textbf{(Equation 1)}$$

Where:

-ε_{280nm} ($M^{-1}.cm^{-1}$) is the molar absorption of the protein at 280 nm

- n_{Trp} and n_{Tyr} correspond to the numbers of Trp and Tyr residues present in the protein sequence

- 5700 $M^{-1}.cm^{-1}$ and 1300 $M^{-1}.cm^{-1}$ are the respective molar extinction coefficients of Trp and Tyr at 280 nm.

Absorption spectra were recorded on a double-beam spectrophotometer Cary 4000 (Varian). The absorbance is characterized by:

$$A = \log(I_0/I) \quad \textbf{(Equation 2)}$$

where I_0 and I are the incident and transmitted intensities, respectively.

A correction for the cuvettes should also be done, as the two cuvettes are never perfectly identical. To this end, the baseline of the instrument is first recorded with both cuvettes filled with the solvents. Then, the Peptide/DNA/Dye is added into the solvent of the sample cuvette and the true absorption spectrum is recorded.

1.4: Preparation of Zn-bound peptides (NCp7 and its mutants)

The complexation of zinc with protein was achieved by adding a concentrated solution of $ZnCl_2$. Approximately 2.5 moles of Zn^{2+} per mole of protein were added to saturate the zinc binding sites of NCp7 and its mutants. The zinc binding induces the folding of protein zinc finger, which enhances photophysical and biological properties of NCp7.

The zinc binding by the protein can be verified by the variation of the tryptophan fluorescence quantum yield induced by the addition of excess EDTA, acting as zinc chelator. The ratio of quantum yields between "holo" and "apo" forms was compared to already published results [520, 521].

1.5: Nucleic Acids synthesis

All the used unlabelled and the labelled ODNs were synthesized by IBA GmbH Nucleic Acids product Supply (Göttingen, Germany). In the case of the doubly labeled ODNs, the 5′ terminus was labelled with carboxytetramethylrhodamine (TMR) or ethyl 2-[3-(ethylamino)-6-ethylimino-2,7-dimethylxanthen-9-yl]benzoate hydrochloride (Rh6G) via an amino-linker with a six carbon spacer arm, while the 3′ terminus was labelled with either 4-(4′-dimethylaminophenylazo) benzoic acid (Dabcyl) or 5(and 6)-carboxyfluorescein (Fl) using a special solid support with the dye already attached. ODNs were purified by the manufacturer by reverse-phase HPLC and polyacrylamide gel electrophoresis.

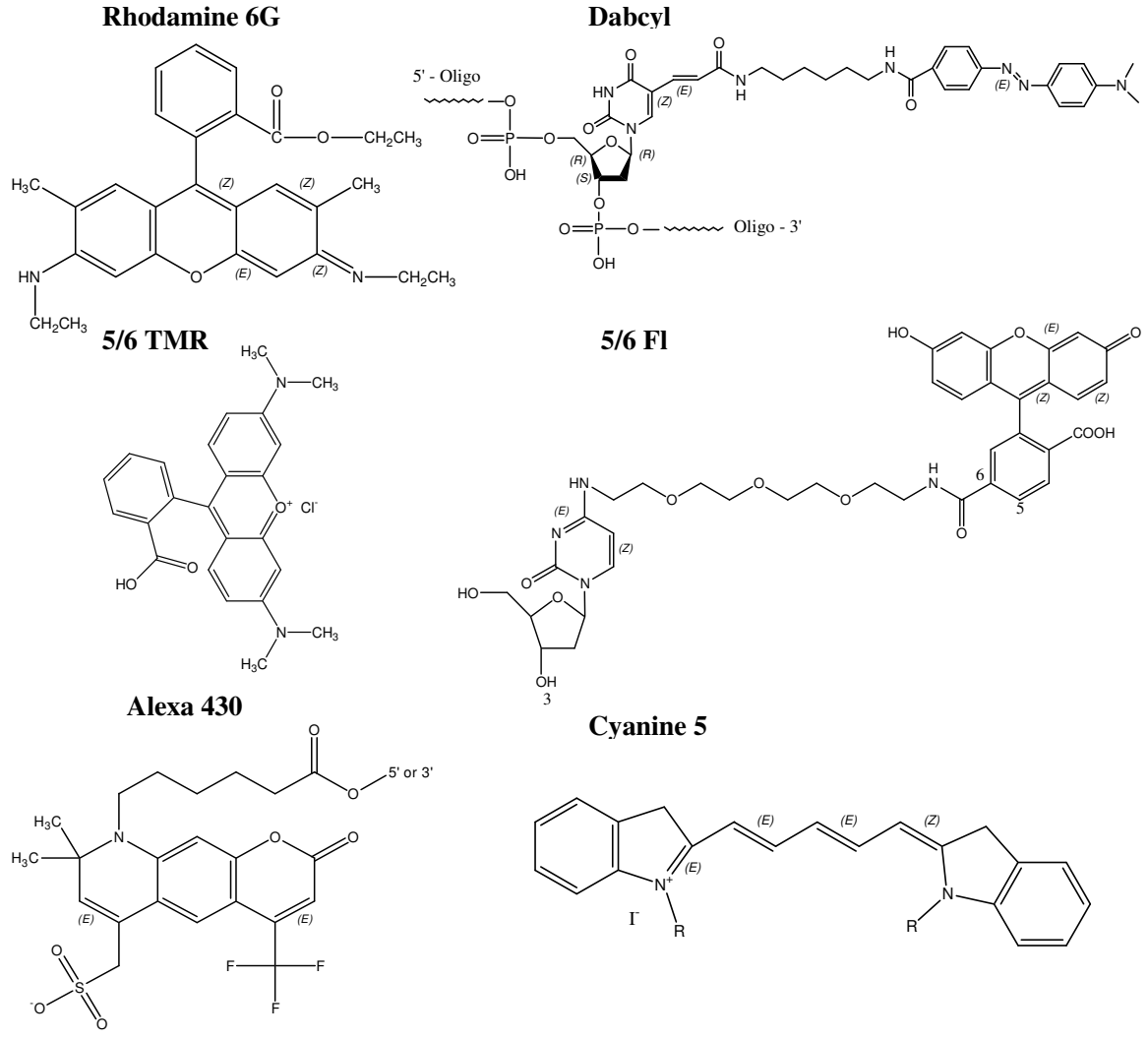


Figure 2: Chromophores and linker used in the present work.

1.5.1 TAR oligonucleotide sequences from HIV-1

To study the chaperone properties of Core protein, Human prion protein, Andes Hanta Virus Nucleocapsid protein and HIV-1 Transactivation response protein (Tat) and NCp7, complementary DNA sequences represented by the canonical model of the HIV-1 transactivation response (TAR) element in a DNA form (dTAR) and the complementary sequence cTAR (Figure 3) were used. Their respective mutants were also used.

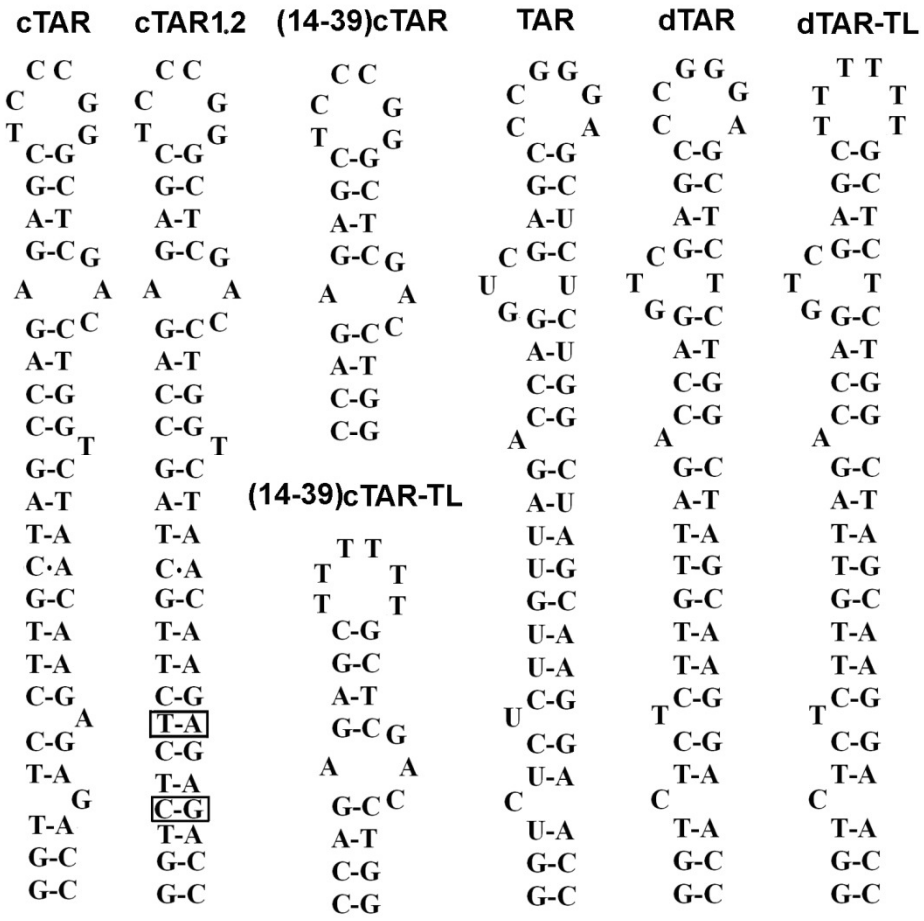


Figure 3: Oligonucleotides secondary structures were predicted from TAR structure [522] and by using the mfold program (http://mfold.dna.albany.edu/?q=mfold).

1.5.2: PBS oligonucleotide sequences from HIV-1

To characterize the chaperone activity of NCp7 and its mutants in the second strand reverse transcription, we used the (-)PBS and (+)PBS sequences and their respective mutants (Figure 4).

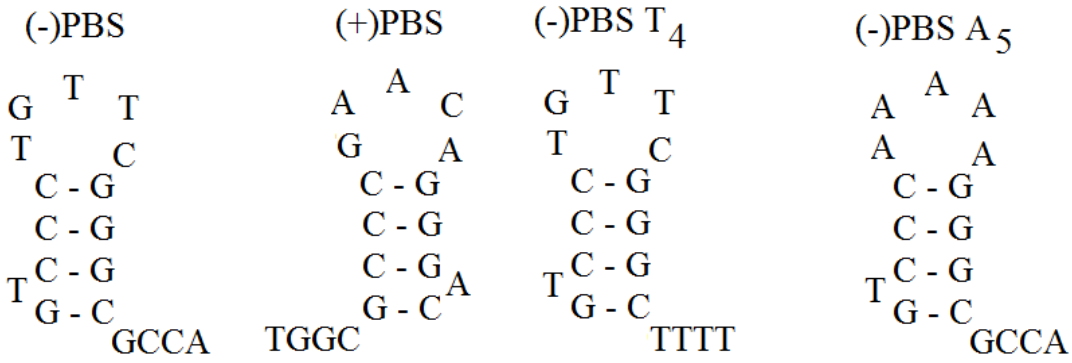


Figure 4: Oligonucleotides secondary structures were predicted from PBS structure [522] and by using the mfold program (http://mfold.dna.albany.edu/?q=mfold).

1.5.3: DLS and SL2 oligonucleotide sequences from HCV

To characterize the chaperone activity of Core peptides on the 3'X tail region of HCV genome, DLS, SL2 and their respective mutant sequences were used (Figure 5).

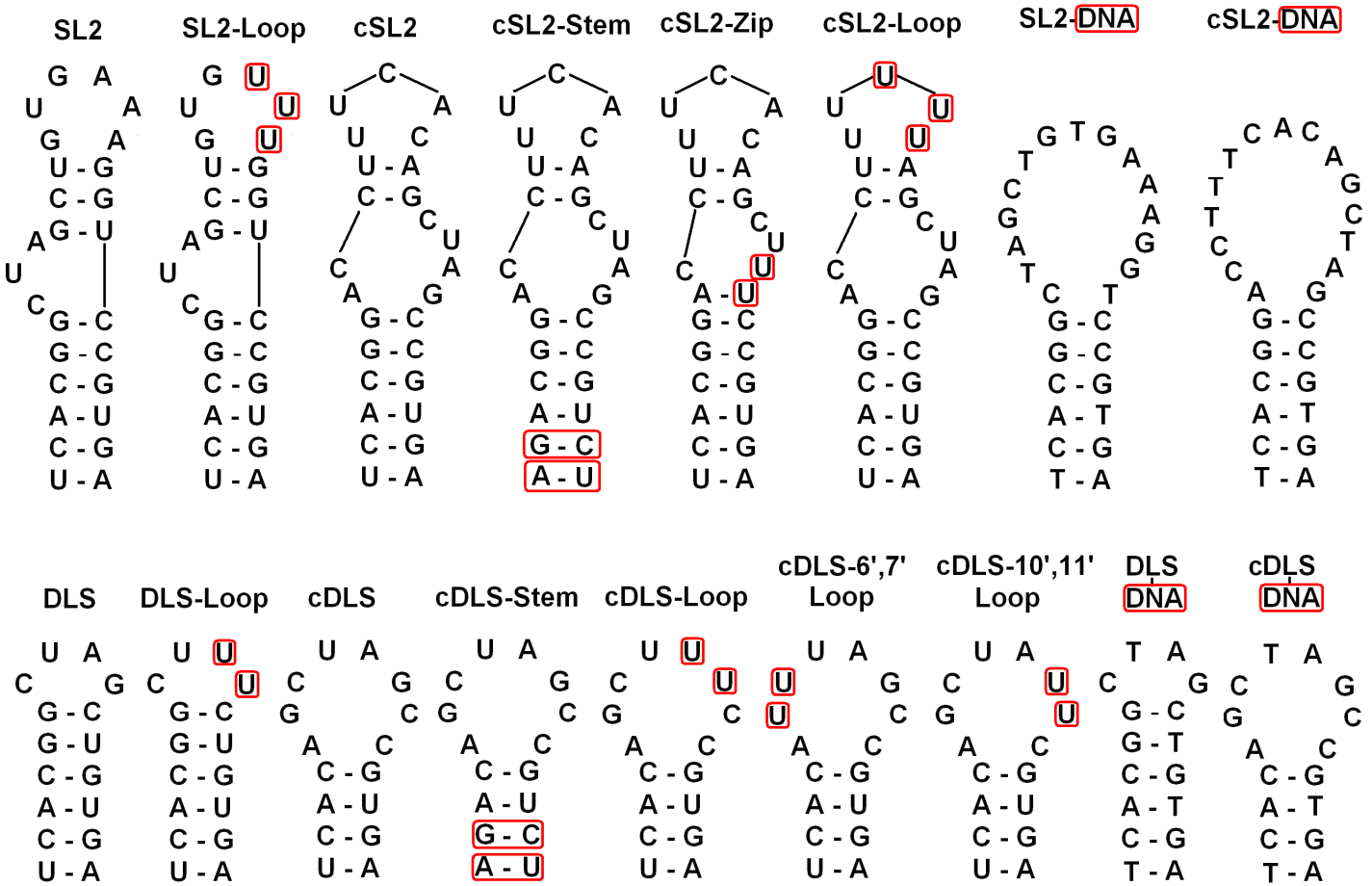


Figure 5: HCV oligonucleotides sequences used in this study. Oligonucleotides secondary structures were predicted by using the mfold program (http://mfold.rna.albany.edu/?q=mfold).

1.5.4. Preparation of nucleic acid solutions

All Supplied lyophilized oligonucleotides were dissolved in Milli-Q and stored at high concentration at -20 ° C.

The concentration of oligonucleotides in solution was determined by their absorbance at 260 nm by using the extinction coefficient provided by the supplier.

2: Steady State Fluorescence Spectroscopy

Steady state fluorescence experiments (spectra and kinetics) were performed by using either FluoroMax-3 spectrofluorimeter (Jobin Yvon Instruments SA Inc.) or Fluorolog-3 (Jobin Yvon Instruments SA Inc.), equipped with a thermostated cell holder.

The excitation source was a POWER xenon lamp. Excitation and emission wavelengths were selected by two identical single grating monochromators equipped with a pair of adjustable slits. The emission signal is corrected for the lamp fluctuations by a calibrated photodiode and for the photomultiplier response versus wavelength by a calibration curve provided by the manufacturer.

2.1: Tryptophan fluorescence

Among the aromatic amino acids, tryptophan (Trp) is the most interesting because of its higher absorbance as compared to tyrosine and phenylalanine. Moreover, the fluorescence of Trp, which is due to its indole moiety, is very sensitive to the physico-chemical properties of its environment (Figure 6).

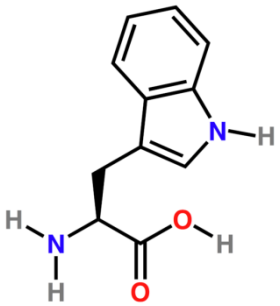


Figure 6: Chemical structure of Tryptophan.

The absorption spectrum of the indole ring in the near UV shows a broad and structured band. This band corresponds to two transitions: $^1A \rightarrow ^1L_a$ and $^1A \rightarrow ^1L_b$, where ^1L_a and ^1L_b are the two excited states of lowest energy, and ^1A indicates the fundamental singlet state [523]. The

characteristic peak at 280 nm is mainly due to the ^1L_b state, while the state ^1L_a is normally involved beyond 300 nm.

The two excited states are characterized by very similar energies, but their transition dipole are almost perpendicular (Figure 7) [523].

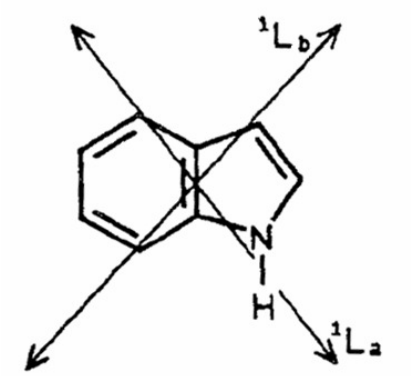


Figure 7: The two dipole electronic transitions of Trp [524].

The emission maximum of Trp, positioned at 350 nm in water (neutral pH), is highly dependent on the environment polarity. In a hydrophobic environment, its emission is blue-shifted. The Trp quantum yield is equal to 0.13 at 23° C in water at neutral pH. Its fluorescence decay is characterized by two lifetimes, a 3 ns major component and a minor component, shorter than 0.6 ns, which emits at shorter wavelengths.

The photophysics of proteins depends on several amino acids residues. The absorption spectrum of a protein between 230 and 300 nm is not only due to Trp residues (ε_{280} = 5700 $M^{-1}.cm^{-1}$), as other residues are also contributing, such as tyrosine residues (ε_{274} = 1,300 $M^{-1}.cm^{-1}$), histidine (ε_{244} = 3.000 $M^{-1}.cm^{-1}$) and phenylalanine (ε_{250} = 200 $M^{-1}.cm^{-1}$). At alkaline pH, tyrosine is ionized as tyrosinate (pKa = 10.9) and presents two absorption bands (ε_{240} = 11,800 $M^{-1}.cm^{-1}$ and ε_{295} = 2,400 $M^{-1}.cm^{-1}$). Additionally, disulfide bridges (ε_{250} = 300 $M^{-1}.cm^{-1}$) and thiol ($\varepsilon_{230\text{-}240}$ = 4500 $M^{-1}.cm^{-1}$) also absorb. However, the fluorescence emission of proteins is mainly governed by Trp fluorescence, which can be selectively excited at 295-305 nm. As previously mentioned in the case of isolated Trp, the fluorescence decay of Trp in proteins is multi-exponential. This heterogeneity is due to the existence of conformational isomers of Trp, called rotamers [524, 525].

In this work, we followed the quenching of Trp fluorescence resulting from the stacking of Trp residues with the nucleotides bases. The static nature of this quenching leads to a decrease in fluorescence intensity without change in lifetimes [524].

2.2: Processing and analysis of titration curves

The measured fluorescence intensity at 350 nm, after exciting at 295nm, was first corrected and then plotted against the concentration of nucleic acids.

The first correction of the fluorescence intensity reflects the so-called screening effect. The corresponding correction factor overcomes the absorption of light from the excitation beam, due to molecules absorbing at either excitation or emission wavelengths [526]. This factor, named *f*, expressed as:

$$f = \frac{\left(A_p + A_s + A_r/2\right)\left[1-10^{-Ap}\right]}{A_p\left[1-10^{-\left(A_p+A_s+A_r/2\right)}\right]} \quad \textbf{(Equation 3)}$$

Where:

- A_p and A_s are the absorbance and excitation wavelengths of the protein, respectively.

A_r is the absorbance at the emission wavelength.

The fluorescence intensity was also corrected for buffer residual emission.

The corrected fluorescence intensity was then plotted versus the concentration of nucleic acids. The titration curve obtained was treated according to the Scatchard model, where the binding of oligonucleotide (ODN) and protein involves n number of independent and equivalent sites. The formation of the ODN-Protein complex can be modelled by the following equilibrium:

$$ODN + n\,\mathrm{Protein} \xleftrightarrow{K} ODN(\mathrm{Protein})_n$$

With,

$$K = \frac{\left[ODN(\mathrm{Protein})_n\right]}{\left[ODN\right]\left[\mathrm{Protein}\right]^n} \quad \textbf{(Equation 4)}$$

Where:

- K is the macroscopic affinity constant for binding of ODN and Protein
- n is the number of binding sites available by ODN for Protein.

This assumption of independent and equivalent sites allows the use of the microscopic affinity constant ($K\mu$) for characterizing the binding of the protein on a given ODN binding site. Therefore, the average number of bound proteins by ODN can be expressed by the equation:

$$\nu = \frac{[\mathrm{Pr}\,otein]_{bound}}{[ODN]_{tot}} = \frac{nK_{\mu}[\mathrm{Pr}\,otein]}{1 + K_{\mu}[\mathrm{Pr}\,otein]} \quad \textbf{(Equation 5)}$$

Where:

- [Protein]$_{bound}$ represents the concentration of proteins bound to ODN
- [ODN]$_{tot}$ represents the total concentration of nucleic acids in the solution
- n represents the number of binding sites of protein to ODN,
- [Protein] represents the concentration of free Protein
- $K\mu$ (M^{-1}) represents the microscopic affinity constant for the binding of Protein for a given binding site.

During the titration, the fluorescence intensity decreases based on the respective concentrations of [Protein]$_{free}$ and [Protein]$_{bound}$, according to the equation:

$$I = I_0 \frac{[\mathrm{Pr}\,otein]}{[\mathrm{Pr}\,otein]_{tot}} + I_f \frac{[\mathrm{Pr}\,otein]_{bound}}{[\mathrm{Pr}\,otein]_{tot}} \quad \textbf{(Equation 6)}$$

Where:

- I represent the corrected fluorescence intensity at 350 nm
- I_0 and I_f represent the fluorescence intensity associated with the free and bound protein respectively
- [Protein]$_{tot}$ represents the total concentration of Protein,
- [Protein] and [Protein]$_{bound}$ represent the concentration of free and bound protein, respectively .

By combining equations (5) and (6), we can deduce the expression of the overall fluorescence intensity depending on the total concentration of added nucleic acids:

$$I = I_0 - \frac{I_0 - I_f}{[\mathrm{Pr}\,otein]_{tot}} \times \frac{\left(1 + K_\mu([\mathrm{Pr}\,otein]_{tot} + n[ODN]_{tot})\right) - \sqrt{\left(1 + K_\mu([\mathrm{Pr}\,otein]_{tot} + n[ODN]_{tot})\right)^2 - 4nK_\mu{}^2[\mathrm{Pr}\,otein]_{tot}[ODN]_{tot}}}{2K_\mu}$$

(Equation 7)

The experimental points were fitted to this model in order to determine the affinity constant $K\mu$. The number of binding sites of the protein on the oligonucleotide, n, can be determined from the intersection of the initial slope of the titration curve with the plateau of fluorescence intensity.

3: Time-resolved fluorescence measurements: fluorescence lifetimes and anisotropy decays

3.1: Technique used

During these experiments, we used single photon counting technique [527, 528]. In the first technique, the sample is excited using a continuous beam of light whose intensity is sinusoidally modulated with time. In our case, we used the second method in which the sample is excited with a pulsed light source (more generally a pulsed laser) and the time during which the molecule remains in the excited state is directly measured.

The single photoelectron technique consist in exciting fluorophores under conditions where no more than a single photon could be detected (less than 5 fluorescence photons detected per 100 excitation pulses). The measured time interval between the initial pulse and the detection of the each single photon is the time during which a molecule remains in its excited state. It can be defined as:

$$\tau = \frac{1}{k_r + k_{nr}} \quad \textbf{(Equation 8)}$$

where k_r is the radiative rate constant, and k_{nr} the non-radiative rate constant. $\tau_0 = k_r^{-1}$ is the radiative lifetime (also called natural lifetime) which is related to the fluorescence lifetime τ via the fluorescence quantum yield φ:

$$\Phi = \frac{\tau}{\tau_0} = \frac{k_r}{k_r + k_{nr}} \quad \textbf{(Equation 9)}$$

Eqns. (8) and (9) do not assume any quenching processes. If some fluorescence quenching induced by a foreign molecule Q, so-called quencher, occurs, an additional term ($k_Q[Q]$) must

be introduced in the denominator, where k_Q is the quenching rate constant and [Q] the quencher concentration.

While the fluorescence decay time depends on the intrinsic characteristics of the fluorophore itself, it also depends in a measurable way upon the local environment. In a general way, the local viscosity, the pH, or the refractive index (Figure 8a), as well as the interactions with other molecules, e.g. by collision or energy transfer (Figure 8b), can all affect the fluorescence lifetime [524].

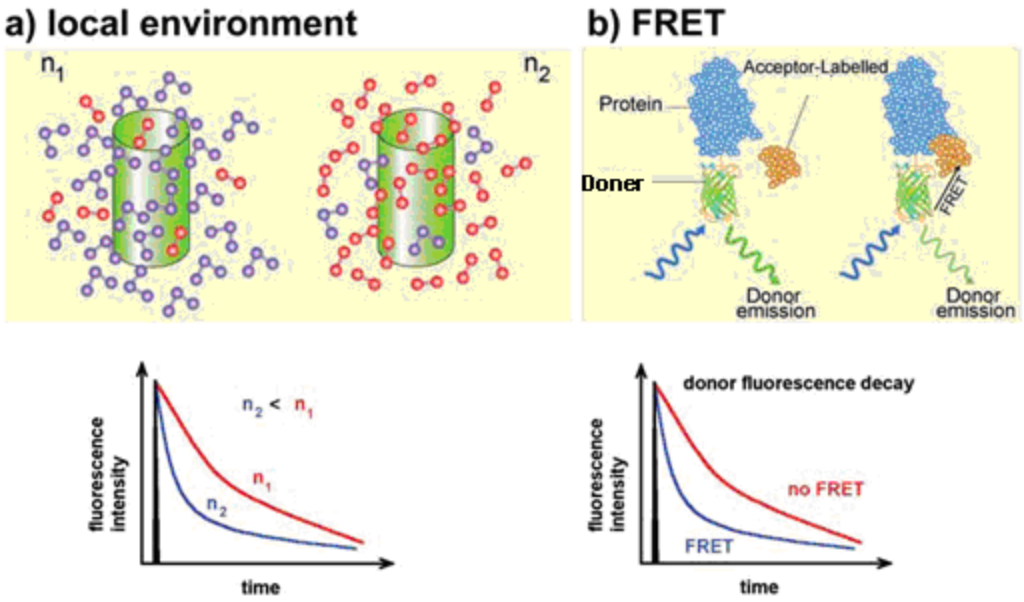


Figure 8: Schematic representation of the photophysical phenomena and its effect on the fluorescence decay. (a)The fluorescence decay is a function of the environment of the fluorophore. For example, the fluorescence lifetime of GFP is a function of its local refractive index. A low refractive index around GFP leads to a longer lifetime, whereas a high refractive index environment causes a shorter lifetime. The fluorescence lifetime of other probes can be sensitive to pH, ions, oxygen etc. (b) FRET occurs when a suitable donor and acceptor are in close proximity, usually below 10 nm.

3.2: Fluorescence anisotropy

On the other hand, fluorescence anisotropy is particularly sensitive to variations in size and shape of macromolecules. It is a method for measuring the binding interaction between two molecules, and can be used to measure the binding constant (or the inverse, the dissociation constant) for the interaction. The basic idea is that a fluorophore excited by polarized light (light whose "waves" vibrate in only one direction) will also emit polarized light. However, if a molecule is moving, it will tend to "scramble" the polarization of the light by radiating at a different direction from the incident light. The "scrambling" effect is greatest with fluorophores freely tumbling in solution and decreases with decreased rates of tumbling.

Protein interactions can be detected when one of the interacting partners is fused to a fluorophore: Upon binding of the partner molecule a larger and more stable complex is formed which will tumble more slowly (thus, increasing the polarization of the emitted light and reducing the "scrambling" effect).

A fluorescent molecule in its ground state is characterized by a transition dipole moment of absorption, $\vec{m}$ A. When a population of fluorophores is illuminated by a linearly polarized light, the molecules which have their absorption transition dipole moments oriented in a parallel direction to that of the electric vector of the incident beam are preferentially excited. This phenomenon is called Photoselection. The absorption probability of each molecule is proportional to *Cos* θ, θ being the angle between the absorption dipole moment of the emitting molecule at the moment of excitation and the incident electric field vector.

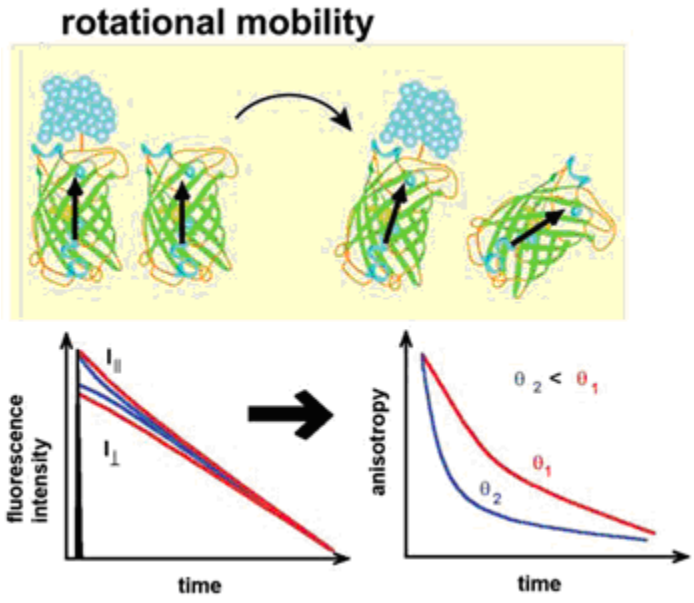


Figure 9: Time-resolved fluorescence anisotropy imaging reveals the rotational mobility of a fluorophore. This is affected by the viscosity of its surroundings, or by binding and conformational changes that affect the rotational mobility. The latter is characterized by the rotational correlation time which can be calculated from the difference between the polarization-resolved fluorescence decays $I_{//}$ and $I_{\perp}$. A fast rotational motion leads to a rapid depolarization.

The emitted fluorescence light is oriented along an electric field vector, parallel to the direction of the transition dipole moment of emission. This transition dipole moment of emission characterizes the fluorescent molecule in its excited state. The distribution of molecules in the excited state is anisotropic. Fundamental anisotropy, r_0, corresponds to the case where emitting molecules are completely immobilized in a rigid medium:

$$r_0 = (3\cos^2\alpha\text{-}1)/5 \quad \textbf{(Equation 10)}$$

where, α is the angle between the moments of absorption $\vec{m}$A and emission $\vec{m}$B.

This r_0 value is the maximum value for the fluorescence anisotropy of the molecule and in some cases it reaches a value equal to 0.4 (when $\vec{m}$A and $\vec{m}$B are strictly parallel; α = 0).

The Brownian rotation of fluorescent particles results in a change in the direction of their emission moments, $\vec{m}$E, by an angle ω(t), which is related to its excited state lifetime. The movement causes a decrease in the anisotropy of distribution fluorescent particle population. This induces a partial or total depolarization of fluorescence. In some cases, fluorescence depolarization can also be induced by the energy transfer from a "donor" molecule to a differently oriented "acceptor" molecule.

In the general case of Brownian motion, r is equal to:

$$r = (3\cos^2(\alpha+\omega(t))\text{-}1)/5 \quad \textbf{(Equation 11)}$$

where α is the angle between the absorption, $\vec{m}$A, and emission $\vec{m}$E moments, and ω(t) is the rotation angle between the direction of Brownian emission vector, $\vec{m}$E, at time zero and its orientation at time t.

For fluorescence anisotropy measurements, the incident light is oriented vertically (along the Z axis) (Figure 10). Fluorescence decays measured at parallel and perpendicular directions versus the excitation can measure the rotational mobility of the fluorophore in its environment (Figure. 9). Thus, the viscosity of the fluorophore environment, binding events or hindered rotation can be examined by this method (Lidke DS et al., 2003, Suhling K, 2004).

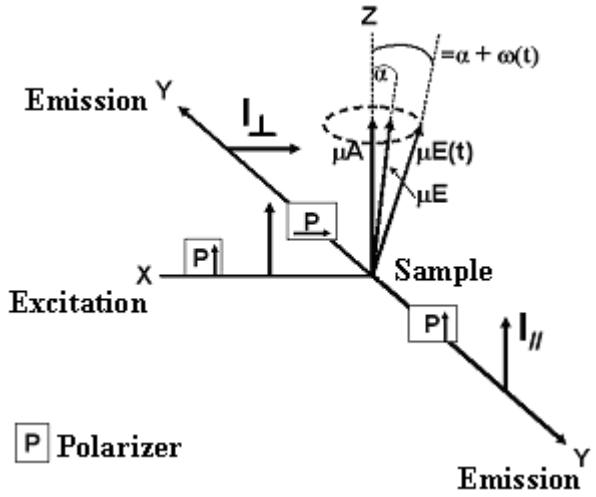


Figure 10: Diagram of the T configuration for the measurement of anisotropy fluorescence.

The time-resolved fluorescence anisotropy r(t) can be defined as:

$$r(t) = \frac{I_{II} - I_{\perp}}{I_{II} + 2I_{\perp}}$$ **(Equation 12)**

where $I_{//}(t)$ and $I_{\perp}(t)$ are the fluorescence intensity decays parallel and perpendicular to the polarization of the exciting light.

This technique works best if a small molecule is fused to a fluorophore and binds to a larger partner (this maximizes the difference in fluorescence anisotropy between bound and unbound states). If the fluorophore is attached to the larger protein in a binding pair, the difference in polarization between bound and unbound states will be smaller (because the unbound protein will already be fairly stable and tumble slowly to begin with) and the measurement will be less precise.

3.2: Experimental

Time-resolved fluorescence measurements were performed with the time-correlated, single-photon counting technique. Excitation pulses were generated by a pulse-picked frequency-tripled Ti-sapphire laser (Tsunami, Spectra Physics) pumped by a Millenia X laser (Spectra Physics) [478]. Excitation wavelength was set at 315 nm, with a repetition rate of 4 MHz. The fluorescence emission was collected through a polarizer set at magic angle 57.7° and a 16 mm band-pass monochromator (Jobin Yvon) at 370 nm. The single-photon events were detected with a micro-channel plate photomultiplier (Hamamatsu) coupled to a pulse pre-amplifier (Philips) and recorded on a multi-channel analyser (Ortec) calibrated at 25.5 ps/channel. The

instrumental response function (IRF) was recorded using a polished aluminium reflector, and its full-width at half-maximum was ~ 40 ps. The mean lifetime $<\tau>$ was calculated from the individual fluorescence lifetimes (τ_i) and their relative amplitudes (α_i) according to $<\tau> = \Sigma\ \alpha_i\ \tau_i$. The population, α_0, of dark species of 2-AP was calculated by: $\alpha_0 = 1 - \tau_{free} / (\tau_{sample}\ R_m)$, where τ_{free} is the lifetime of free 2-Ap, τ_{sample} is the measured mean lifetime of 2-Ap within the ODN and R_m is the ratio of their corresponding quantum yields. The remaining amplitudes, α_{ic} were recalculated from the measured amplitudes according to

$\alpha_{ic} = \alpha_i / (1- \alpha_0)$

Time-resolved anisotropy, resulting from the measurement of the fluorescence decay curves recorded in directions parallel ($I_{//}$) and perpendicular ($I\perp$), alternatively, to the excitation beam polarization, was analysed by the following equations:

$I_{//} = I(t)[1 + 2\ r(t)]/3$

$I\perp = I(t)[1 - r(t)]/3$

$$r(t) = \frac{I_{//} - G \times I_{\perp}}{I_{//} + 2G \times I_{\perp}} = r_0 \sum_i \beta i \times \exp\left(-\,t/\Phi i\right) \quad \textbf{(Equation 13)}$$

where β_i are the amplitudes of the rotational correlation times Φ_i , and G is the geometry factor at the emission wavelength, determined in independent experiments. Theoretical values of the rotational correlation times were calculated from the molecular mass (M) of the complexes, assuming a spherical shape, by:

$$\Phi = \eta\ M\ (\upsilon + h) / RT \quad \textbf{(Equation 14)}$$

where η is the viscosity (assumed to be 1 cP), T is the temperature (293K), υ is the specific volume of the particle (assumed to be 0.83 mL/g [524]), h is the hydration degree (assumed to be 0.2 mL/g for proteins) and R is the molar gas constant.

The cone semi-angle θ_0 of the local motion of 2Ap [529, 530] was calculated as:

$$\theta_0 = \cos^{-1}(0.5\,(1 + 8S)^{0.5} - 1))$$ **(Equation 15)**

where S is the generalized order parameter

$$S = \left(\sum_{i=2}^{n} \beta' i \Big/ \sum_{i=1}^{n} \beta' i \right)^{0.5}$$ **(Equation 16)**

in which β'_i are the amplitudes of the rotational correlation times, recalculated such that $\Sigma\, \beta'_i = r_0$.

Time-resolved intensity and anisotropy data were treated according to the maximum entropy method (Pulse 5 software) [531, 532] or according to a non-linear least-square analysis using a homemade software (kindly provided by G. Krishnamoorthy). In all cases, the χ^2 values were close to 1 and the weighted residuals as well as their autocorrelation were distributed randomly around 0, indicating an optimal fit.

3.3: Measuring the binding of proteins to nucleic acids by direct titration using Steady State Anisotropy

Direct assays were performed by adding increasing concentrations of peptides at a fixed concentration of labelled nucleic acid.

To analyze the titration curves measured by anisotropy, we adapted the model used for Scatchard titration in case of fluorescence intensity measurements (previous equation 7):

$$r = r_i + \frac{r_p - r_i}{n} \times \frac{\left(1 + K_\mu\left(n[ODN]_{tot} + [\mathrm{Pr}otein]_{tot}\right)\right) - \sqrt{\left(1 + K_\mu\left([\mathrm{Pr}otein]_{tot} + n[ODN]_{tot}\right)\right)^2 - 4nK_\mu{}^2[\mathrm{Pr}otein]_{tot}[ODN]_{tot}}}{2K_\mu[ODN]_{tot}}$$

(Equation 17)

Where, $[ODN]_{tot}$ and $[Protein]_{tot}$ are the total concentrations of labelled ODN and unlabelled peptide, respectively. r_p represents the anisotropy when all peptide is bound, while r_i and r correspond to the anisotropy of Labelled DNA in the absence and presence of protein, respectively. The experimental points were fitted to this model to determine the affinity

constant K_μ. The number of binding sites, n, of the protein on the oligonucleotide is determined using indirect assays by measuring the fluorescence of Trp peptides.

4: Hybridization kinetics of nucleic acids

4.1: Doubly labelled oligonucleotides

4.1.1: General Principles

Dual-labelled oligonucleotide sequences were derived on the basis of "Molecular Beacon" type construction. Described by Tyagi and Kramer [533], "Molecular Beacons (MB) are a class of ODN probes used to detect the presence of specific sequences. The MB are stem-loop structured nucleic acids covalently labelled at their 5 'and 3' with a fluorophore and a fluorescence quencher, respectively. This stem-loop structure allows the probes to remain in close proximity, resulting in a strong quenching of fluorescence. Hybridization with a complementary sequence causes the separation of the probes, which induced a significant recovery of fluorescence. Between the two conformations, the fluorescence can increase by a factor of 20- to 60- fold, thus, making the MB very sensitive detection of target sequences [478, 533]. It should be noted that this factor not only depends on the labelling yield of both probes and the stability of the stem but also on the complementarity of the target sequence, since the fluorescence after hybridization depends on the stability of the MB Hybrid by target sequence [533]. The latter property makes the MB interesting, particularly in the mechanistic study of nucleic acid hybridization with target complimentary sequences.

4.1.2: Theoretical Background on the concept of exciton

The absorption spectrum of a doubly labelled stem-loop sequence shows significant differences from the spectrum of an equimolar mixture of singly labelled species. Furthermore, the fluorescence intensity of the doubly labelled sequence is strongly inhibited compared with that of an equimolar mixture of singly labelled species. These spectral changes are defined by the theory of exciton [534].

When the sequence is doubly labelled in its stem-loop conformation, complementary ends promotes the formation of a heterodimer between the two probes, leading to a strong coupling between the dipole transitions of the two probes [535]. The coupling energy (U) between the dipole moments of the two probes depends on the distance between the chromophores and their respective orientations. The energy U is given by:

$$U = \frac{\mu_A \mu_B}{4\pi\varepsilon_0 n^2 R^3}\kappa \quad \textbf{(Equation 18)}$$

Where:

- R is the distance between the transition dipole moments
- ε_0 is the permittivity in vacuum, n is the refractive index of the medium (for water: 1.33)
- μ_A is the intensity of the dipole moment A and κ is an orientation factor.

κ is defined by the equation:

$$\kappa = \cos\theta_{AB} - 3\cos\theta_{AR}\theta_{BR} \quad \textbf{(Equation 19)}$$

With:

- θ_{AB}, the angle between the dipole moments A and B
- θ_{AR} and θ_{BR} are the respective angles between the dipole moment A and the vector distance between the R and B and the dipole moment vector distance R.

In our case, the heterodimer is type H, indicating that the dipole moments is parallel and perpendicular to the axis that joins the two chromophores.

Thus, $\kappa \approx 1$.

The coupling between the chromophores leads to a spectral shift, Δλ, in the absorption spectrum. This discrepancy is related to the energy (U) according to the following equation:

$$U = \frac{hc}{\Delta\lambda} \quad \textbf{(Equation 20)}$$

Where:

- h is the Planck constant
- c is the speed of light
- Δλ is the spectral shift.

The energy U is strongly dependent on the distance between the chromophores. The absorbance of the heterodimer can be used as a tool to monitor changes at very short distance (less than 15 Å).

4.1.3: Theoretical Background of the non-radiative energy transfer (FRET)

For distances beyond 15 Å, the non-radiative energy transfer is replaced by the so-called exciton energy transfer. In this case, the excited molecule (donor) transfers its excess energy to the second molecule (acceptor) either in a radiative or nonradiative mode. Between 15-100 Å, the mechanism of energy transfer involves a weak dipole - dipole coupling without photon emission. This mechanism is named as non-radiative energy transfer or resonance energy transfer (FRET).

The energy transfer between donor and acceptor can be modelled by the following balance:

$$D_1 + A_0 \underset{k_{-T}}{\overset{k_T}{\rightleftharpoons}} D_0 + A_1$$

Where:

- D_0 and D_1 represent the donor at ground and excited state, respectively
- A_0 and A_1 represent the acceptor at ground and excited state, respectively
- k_T represents the rate constant characterizing the energy transfer from D_1 to A_0,
- k_{-T} represents the rate constant characterizing the energy transfer from A_1 to D_0.

The vibrational relaxation of the acceptor in its excited state A_1 is fast enough to render any return to its ground state unlikely (represented by k_{-T}).

In 1949, Forster expressed k_T according to the recovery of emission spectra of donor and acceptor, the distance between the probes, and the relative orientation of transition dipole moments.

He deduced the following equation:

$$k_T = 8.71\times10^{23}\frac{\kappa^2\Phi_D}{R^6\tau_D}n^{-4}J_{AD} = \frac{1}{\tau_D}\left(\frac{R_0}{R}\right)^6 \quad \textbf{(Equation 21)}$$

Where:

-Φ_D and τ_D represent the quantum yield and lifetime of the donor.

-n is the refractive index of the medium

-κ^2 represents the geometric factor characterizing the orientation of the dipole moments of two chromophores. The difficulty of determining this factor for chromophores in solution led us to assume a spatially averaged value of 2/3. In these circumstances, the donor and acceptor are free to take all possible orientations during the lifetime of the excited state of the donor.

-J_{AD} represents the integral overlap between the emission spectrum of the donor and the absorption spectrum of the acceptor. It can be defined by:

$$J_{AD} = \int \varepsilon_A(\lambda) f_D(\lambda) \lambda^4 d\lambda \quad \textbf{(Equation 22)}$$

With:

- ε_A (λ) represents the molar absorption of the acceptor and f $_D$ (λ) represents the fluorescence intensity of the donor.
- R is the distance between the probes,
- R_0 is the so-called Förster critical distance, namely the distance for which transfer efficiency equals 0.5. It can be calculated by the following equation:

$$R_0 = \left(8.79 \times 10^{23} n^{-4} \Phi_D \kappa^2 J_{AD}\right)^{\frac{1}{6}} \quad \textbf{(Equation 23)}$$

Knowing k_T, it is then possible to estimate the efficiency of energy transfer (E) as the fraction of donor molecules in the excited state which deactivates by non-radiative energy transfer, by:

$$E = \frac{k_T}{k_T + \tau_D^{-1}} \quad \textbf{(Equation 24)}$$

Combining equations 23 and 26, it is possible to show that E can be written as:

$$E = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6} \quad \textbf{(Equation 25)}$$

Note: R_0 is 20 and 60 Å for Rhodamine6G/DABCYL and Fluorescein/TetraMethyl Rhodamaine couples respectively [478].

It is possible to determine the efficiency of energy transfer since E is proportional to the ratio of intensities of the donor fluorescence in presence of acceptor (F_{DA}) and in absence of the acceptor (F_D). This can be stated as:

$$E = \frac{1}{1+\left(\frac{R}{R_0}\right)^6} = 1 - \frac{F_{DA}}{F_D} \quad \textbf{(Equation 26)}$$

or from lifetime τ_D and τ_{DA} by the relationship:

$$E = \frac{1}{1+\left(\frac{R}{R_0}\right)^6} = 1 - \frac{\tau_{DA}}{\tau_D} \quad \textbf{(Equation 27)}$$

4.2: Application of FRET in annealing kinetics

To determine the mechanism of hybridization between two complementary ODNs in absence and in presence of protein, we used the MB properties. We mixed a doubly labelled (+)ODN marked at both ends 5 'and 3' with an excess of complementary (-)ODN. The stem-loop (+)ODN brings probes in close proximity, and places (+)ODN in its non-fluorescent conformation, (+)cODN. The pair of chromophores chosen for the kinetic experiments were carboxytetramethylrhodamine (TMR) or ethyl 2-[3-(ethylamino)-6-ethylimino-2,7-dimethylxanthen-9-yl]benzoate hydrochloride (Rh6G) at 5' terminus and either 4-(4′-dimethylaminophenylazo) benzoic acid (Dabcyl) or 5(and 6)-carboxyfluorescein (Fl) at 3' terminus. The hybridization of a complementary sequence, (-)ODN, with (+)cODN, induces the formation of a double stranded ODN sequence, extended away Rh6G from DABCYL or Fl from TMR. This results in a recovery of fluorescence that can be followed in real time.

Theoretically, the hybridization of two complementary stem-loop structured sequences depends partly on their nucleation, and secondly by the merging of their respective strands.

4.3: Analytical resolution of the kinetic pathway described in scheme 1 based on annealing of cTAR and dTAR.

$$\text{cTAR} + \text{dTAR} \underset{k_{diss}}{\overset{k_{ass}}{\rightleftharpoons}} \text{ED*} \xrightarrow{k_2} \text{ED}$$

The reaction kinetics in scheme 1 can be described using the following differential equations:

$$\frac{d[cTAR]}{dt} = -k_{ass}[cTAR][dTAR] + k_{diss}[ED*] \quad \text{(S1)}$$

$$\frac{d[ED*]}{dt} = k_{ass}[cTAR][dTAR] - (k_{diss} + k_2)[ED*] \quad \text{(S2)}$$

$$\frac{d[ED]}{dt} = k_2[ED] \quad \text{(S3)}$$

Solving the set of equations S1 to S3 provides the time dependence of [cTAR], [ED*] and [ED], the concentrations of cTAR, ED* and ED, respectively:

$$[cTAR] = [cTAR]_0 \frac{1}{2\sqrt{\Delta}}\left[(\sqrt{\Delta} + k_{diss} + k_2 - k_{ass}[dTAR])e^{r1t} + (\sqrt{\Delta} + k_{ass}[dTAR] - k_{diss} - k_2)e^{r2t}\right] \quad \text{(S4)}$$

$$[ED*] = [cTAR]_0 \frac{k_{ass}[dTAR]}{\sqrt{\Delta}}(e^{r1t} - e^{r2t}) \quad \text{(S5)}$$

$$[ED] = [cTAR]_0 (1 - \frac{1}{2\sqrt{\Delta}})\left[(\sqrt{\Delta} + k_{diss} + k_2 + k_{ass}[dTAR])e^{r1t} + (\sqrt{\Delta} - k_{ass}[dTAR] - k_{diss} - k_2)e^{r2t}\right] \quad \text{(S6)}$$

where $\Delta = (k_{ass}[\text{dTAR}] + k_{diss} + k_2)^2 - 4k_{ass}[\text{dTAR}]\, k_2$ (S7)

and $r_{1,2} = \frac{-(k_{ass}[dTAR] + k_{-1} + k_2) \pm \sqrt{\Delta}}{2} = -k_{obs1,2}$ (S8)

From equation S8 we obtain:

$k_{obs1} + k_{obs2} = k_{ass}[\text{dTAR}] + k_{-1} + k_2$ (S9)

Substitution into equation 4 of the expressions of [cTAR], [ED*] and [ED], and using $[cTAR]_0 = 1$ leads to:

$$I_{(t)} = I_f - (I_f - I_0)\left[\frac{1}{I_f - I_0}\left(I_f C_{20} - I_0 A_{20} - I_i AB_{20}\right)e^{-k_{obs1}t} + \frac{1}{I_f - I_0}\left(I_f C_{10} - I_0 A_{10} - I_i AB_{10}\right)e^{-k_{obs2}t}\right] \quad \text{(S10)}$$

where

$$A_{10} = \frac{\sqrt{\Delta} + k_{diss} + k_2 - k_{ass}[dTAR]}{2\sqrt{\Delta}} \quad \text{(S11)}$$

$$A_{20} = \frac{\sqrt{\Delta} - k_{diss} - k_2 + k_{ass}[dTAR]}{2\sqrt{\Delta}} \quad \text{(S12)}$$

$$AB_{10} = \frac{k_{ass}[dTAR]}{\sqrt{\Delta}} = -AB_{20} \quad \text{(S13)}$$

$$C_{10} = \frac{\sqrt{\Delta} + k_{diss} + k_2 + k_{ass}[dTAR]}{2\sqrt{\Delta}} \quad \text{(S14)}$$

and

$$C_{20} = \frac{\sqrt{\Delta} - k_{diss} - k_2 - k_{ass}[dTAR]}{2\sqrt{\Delta}} \quad \text{(S15)}$$

The time dependence of the fluorescence intensity (equation S10) is thus described by a double-exponential function, with k_{obs1} and k_{obs2} characterizing the two kinetic phases. Moreover, the amplitude a of the fast kinetic component corresponds to:

$$a = \frac{1}{I_f - I_0}\left(I_f C_{20} - I_0 A_{20} - I_i AB_{20}\right) \quad \text{(S16)}$$

4.4: Kinetics of hybridization of (+)cODN with (-)ODN as a function of temperature

To go further in the mechanistic study of the hybridization of complementary sequences, the temperature dependence of the annealing kinetics was carried out with 10 nM doubly labelled (+)cODN derivatives and a fixed excess of non-labelled (-)ODN derivatives, at different temperatures (10°C, 20°C, 25°C, 30°C, 35°C and 40°C) in the presence and absence of protein. The data was first processed according to the Arrhenius law:

$$k(T) = A\exp\left(-\frac{E_A}{RT}\right) \quad \textbf{(Equation 28)}$$

Where:

- k(T) represents the rate constant determined at the temperature T
- A is the pre-exponential factor of Arrhenius
- EA is the activation energy
- R is the gas constant
- T is the temperature at which the reaction was carried out.

In practice, the natural logarithm of the rate constant determined experimentally is plotted against the inverse of temperature. The graph obtained by adjusting a straight line allows us to deduce the Arrhenius pre-exponential factor (A) and the activation energy (E_A). For a given reaction, E_A quantifies the energy barrier that the reactants must cross to reach the transition state. Once in this configuration, the activated complex can evolve into a more thermodynamically stable state. One of the properties of a catalyst is to facilitate the initiation of this reaction by lowering the E_A. E_A kinetics without and with the presence of protein were therefore compared to determine the catalytic potential of protein in this reaction.
Arrhenius pre-exponential factor is used because it depends on the collision as indicated by the following equation:

$$A = PxZ_{AB} \quad \textbf{(Equation 29)}$$

Where:
- A is the pre-exponential factor of Arrhenius
- P is the steric factor taking into account the relative orientation of molecules during the collision
- Z_{AB} is the collision frequency of a given particle A with its reaction partner B.

According to collision theory, it is possible to calculate the theoretical maximum value of A (A_{th}), assuming A and B as solid spherical particle diameter dA and dB respectively, and assuming that all collisions are productive (P = 1):

$$A_{th} = \frac{\pi}{4}(dA + dB)^2 \sqrt{\frac{8k_B T}{\pi \mu_{AB}}} N_A{}^3 n_A n_B \quad \textbf{(Equation 30)}$$

Where:

- $(\pi/4)\ (dA + dB)^2$ describes the cross section associated with the collision between particles A and B

- dA and dB are the respective diameters of the particles A and B.

- $(8kBT/\pi\mu AB)^{1/2}$ is the mean particle velocity deduced from the kinetic theory of gases, with k_B being the Boltzmann constant, T is temperature, and μ_{AB} being the reduced mass calculated from the relationship: $\mu_{AB} = (mA.mB)/(mA + mB)$ where mA and mB are the respective mass of particles A and B.

- N_A is Avogadro's number

- n_A and n_B are the respective amounts of moles of particles A and B.

On this basis, it is then possible to determine the steric factor associated with the reaction by:

$$P = A/A_{th} \quad \textbf{(Equation 31)}$$

P values provide information on particle geometry and report on the effectiveness of collisions between the reactants. Comparison of P in the absence and presence of protein gives information on the influence of the protein on the collisions between (+)cODN and (-)ODN.

The influence of temperature on annealing kinetics has also been analyzed according to Eyring theory of the transition state,:

$$k(T) = \frac{k_B T}{h} \exp\left(-\frac{\Delta G^{\neq}}{RT}\right) \quad \textbf{(Equation 32)}$$

Where:

- k(T) is the bimolecular rate constant at temperature T

- k_B, h and R are respectively the Boltzmann constant, Planck's constant, and the gas constant

- $\Delta G^{\neq}$ is the Gibbs energy of activation

Given that:

$$\Delta G^{\neq} = \Delta H^{\neq} - T\Delta S^{\neq} \quad \textbf{(Equation 33)}$$

It becomes possible to decompose the Eyring relationship by:

$$\ln\left(\frac{k(T)}{T}\right)=\left(\frac{\Delta S^{\neq}}{R}+\ln\frac{k_B}{h}\right)-\left(\frac{1}{T}\right)\left(\frac{\Delta H^{\neq}}{R}\right) \quad \textbf{(Equation 34)}$$

In practice, the natural logarithm of the rate constant obtained at temperature T divided by temperature is plotted against the inverse of temperature. The graph obtained by adjusting a straight line allows us to deduce the change in enthalpy activation ($\Delta H^{\neq}$) and entropy activation ($\Delta S^{\neq}$).

Theoretically, the $\Delta H^{\neq}$ provides information on the hydrogen bonds to be broken in the double-stranded region of ODN, so that the reactants reach the state of activated complex. In the case of nucleic acid hybridization, the $\Delta H^{\neq}$ reflects the energy barrier to be crossed by the protagonists of the reaction to reach the transition state. $\Delta H^{\neq}$ reflects, at least in part, the secondary or tertiary structure of the hybridization partners including the melting of base pairs.

The $\Delta S^{\neq}$ is a thermodynamic parameter related to the collision of the particles. It also reflects the change in the number of configurations, associated with the passage of the state of the reactive transition state. This is based on the Boltzmann formula that shows:

$$\Delta S^{\neq}=k_B\ln\frac{\Omega^{\neq}}{\Omega^{0}} \quad \textbf{(Equation 35)}$$

Where:

- $\Delta S^{\#}$ is the entropy of activation
- k_B is the Boltzmann constant
- $\Omega^{\neq}$ and Ω^{0} represent the number of possible occupied configurations (or states) by particles at the transition state and initial state, respectively.

$\Delta S^{\#}$ in the absence and presence of the protein can be used to estimate changes in transition states induced by the protein. This can be estimated by applying the the following equation:

$$\Delta\Delta S^{\neq}=\Delta S^{\neq}_{+protein}-\Delta S^{\neq}_{-protein}=k_B\ln\frac{\Omega^{\neq}_{+protein}}{\Omega^{\neq}_{-protein}} \quad \textbf{(Equation 36)}$$

- $\Delta S^{\#}_{-protein}$ and $\Delta S^{\#}_{+protein}$ represent the variations of entropy of activation in the absence and presence of the protein, respectively.
- kB is the Boltzmann constant
- $\Omega^{\neq}_{-Protein}$ and $\Omega^{\neq}_{+Protein}$ represent the number of configurations (or states) occupied by particles at the transition state in the absence and presence of the protein.

The ratio $\Omega^{\neq}_{+Protein}/\Omega^{\neq}_{-Protein}$ represent the relative number of configurations that can occupy the transition states in the presence of the protein. The report reflects how the protein affects the collision efficiency. Thus, a ratio $\Omega^{\neq}_{+Protein}/\Omega^{\neq}_{-Protein} > 1$ means that the protein would allow a greater number of possible configurations of the transition state.

5: Fluorescence correlation spectroscopy

5.1: Principle

The fluorescence correlation spectroscopy (FCS) is based on the analysis of the fluctuations of the fluorescence intensity, measured in a finite focal volume. Analysis of these fluorescence fluctuations provides information on the diffusion time and average number of particles passing through that finite focal volume, as well as all other processes affecting the fluorescence intensity of the particle during its passage through the measurement volume. Eventually, both the concentration and size of the particle (molecule) are determined. The florescence intensity is fluctuating due to Brownian motion of the particles. Since the method consists in observing a small number of molecules in a very tiny spot, it is a very sensitive analytical tool. Both parameters are very important and essential in biochemical research, biophysics and chemistry. Moreover, the method enables us to observe fluorescence-tagged-molecules in the biochemical pathway within intact living cells.

5.2: Processing and Analysis FCS

The focal volume is generated by photon excitation obtained by focusing a femtosecond pulsed laser beam produced by a Titanium Sapphire pulsed laser (Tsunami, Spectra-Physics) (Figure 49 A). Focusing of the beam at the sample is ensured by a 60X Olympus objective lens mounted on an inverted microscope Olympus IX70. The measurements are made in 400µL sample solution in eight wells Lab-Tek II. The fluorescence signal is collected (Figure 49 B) and passed to an avalanche photodiode which is coupled to an ALV-5000E correlator (ALV, Germany). This allows real-time calculation of correlation curves via the treatment of

a mathematical application called normalized autocorrelation function G(τ) whose expression is:

$$G(\tau) = \frac{\langle \delta F(t) \cdot \delta F(t+\tau) \rangle}{\langle F(t) \rangle^2} \quad \textbf{(Equation 37)}$$

Where:

- <F(t)> represents the average fluorescence signal
- $\delta F(t)$ represents the fluctuation of florescence, and τ is the latency.

However, a calibration step for the system is necessary to determine the dimensions of the focal volume. These dimensions are measured by using a FCS solution containing 50 nM of 5-carboxytetramethylrhodamine (TMR). The diffusion coefficient of TMR in water is $2.8 \times 10^{-6}.cm^2.s^{-1}$. The determined excitation volume is order of 0.3 μm^3.

Autocorrelation curves are then adjusted by using a model which takes into account the diffusion of fluorescent particles in a Gaussian excitation volume, and the triplet state (Figure 49 C). Autocorrelation curves were adjusted according to previously proposed models (Azoulay, J. et al. 2003; Hess, ST et al., 2002):

$$G(\tau) = \frac{1}{N}\left(1+\frac{\tau}{\tau_{da}}\right)^{-1}\left(1+\frac{1}{S^2}\frac{\tau}{\tau_{da}}\right)^{\frac{-1}{2}}\left(1+\left(\frac{f_t}{1-f_t}\right)\exp\left(\frac{-\tau}{\tau_t}\right)\right) \quad \textbf{(Equation 38)}$$

Where:

- N is the apparent mean number of fluorescent molecules in the focal volume
- τ_{da} represents the apparent diffusion time
- S represents the ratio between the major and minor axes of the excitation volume
- f_t is the fraction of fluorophores in the triplet state
- τ_t represent associated life time with the triplet state.

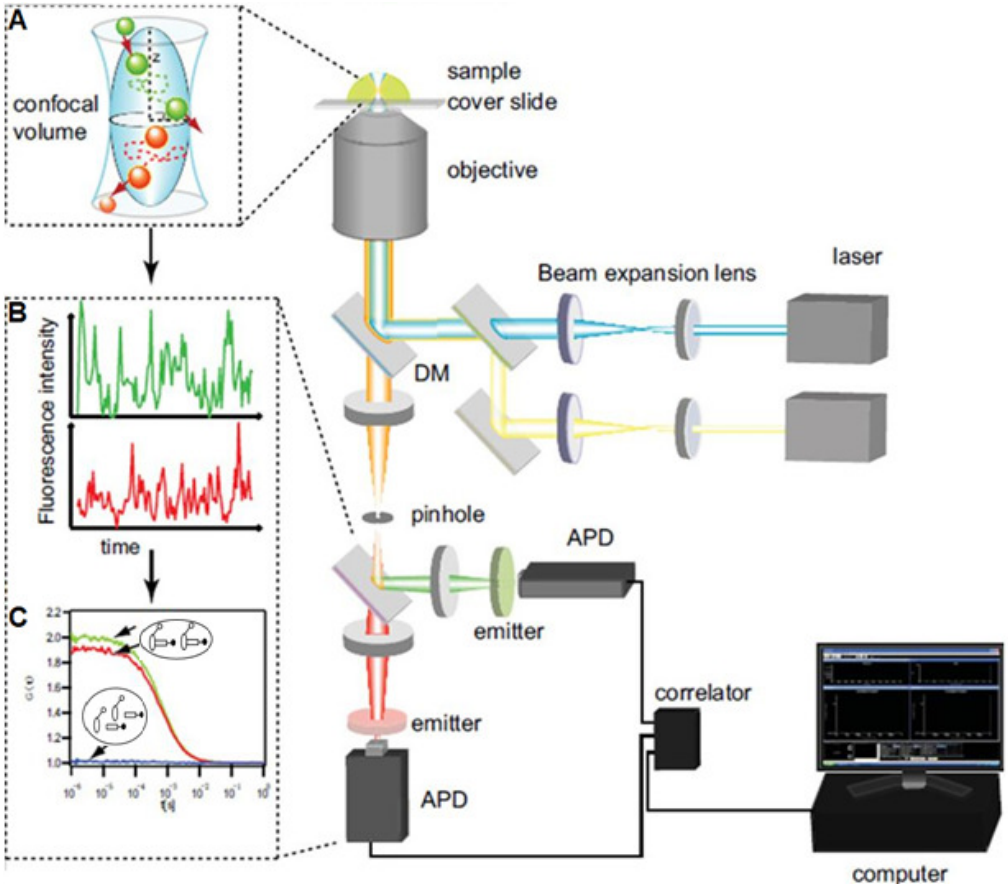


Figure 11: Instrumental setup of FCS. (A) Diagram of the observation focal volume (B) Recording of changes in fluorescence (C) autocorrelation curve.

5.3: Experimental

The experiments were performed in an eight-well lab-Tek II coverglass system, using a 400-µl volume per well. The focal spot was set about 20 µm above the coverslip. Peptides were added to the labelled or non-labelled oligonucleotide sequences at different nucleotide to peptide ratios. To avoid high local concentrations during mixing, both reactants were of the same volume. Assuming that peptides diffuses freely in a Gaussian excitation volume, the normalized autocorrelation function, G(τ), calculated from the fluorescence fluctuations was fitted according to (39):

$$G(\tau) = \frac{1}{N}\left(1+\frac{\tau}{\tau_{da}}\right)^{-1}\left(1+\frac{1}{S^2}\frac{\tau}{\tau_{da}}\right)^{\frac{-1}{2}} \quad \textbf{(Equation 39)}$$

where τ_d is the diffusion time, N is the mean number of molecules within the excitation volume and S is the ratio between the axial and lateral radii of the excitation volume. The point spread function of the set-up was determined from a z-scan on one fluorescent bead (20 nm in diameter). The measured lateral and axial resolutions were respectively 0.3 and 1 µm.

Typical data recording times were 10 min. Twenty autocorrelation curves were recorded for each sample. When spikes of high fluorescence intensity, most likely associated to aggregates, were observed in the fluorescence fluctuation profiles, the corresponding autocorrelation curves were discarded. In these conditions, the average autocorrelation curve calculated from the sum of the remaining autocorrelation curves could be adequately fitted by a single population model with a diffusion time close to that of Oligonucleotide alone.

6: Surface Plasmon Resonance (SPR)

SPR is a physical phenomenon which occurs when a polarized light beam is projected through a prism onto a thin metal film (gold or silver) (Figure 12).

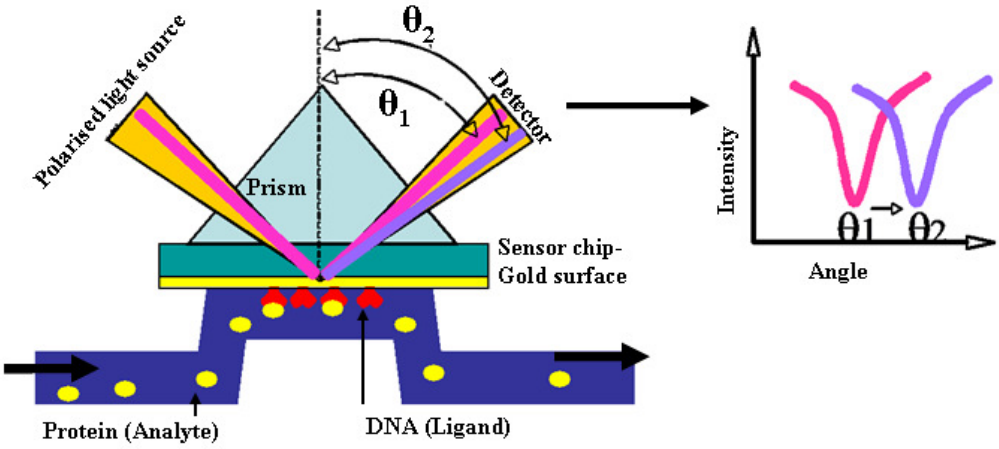


Figure 12: The basic principle of surface Plasmon resonance and analysis.

At a specific angle of the projected light, resonance coupling between light photons and surface plasmons of the gold can occur since their frequencies match. Because the resonance leads to an energy transfer, the reflected light shows a sharp intensity drop at the angle where SPR is taking place. Resonance coupling of the plasmons generates an evanescent wave that extends 100nm above and below the gold surface. For SPR as an analytical tool it is most important that a change in the refractive index within the environment of the evanescent wave causes a change of the angel where the sharp intensity drop can be observed (Compare θ_1 and θ_2). Binding of one biomolecule to another one immobilized on top of the sensor chip's gold surface will lead to a change of refractive index and will be recorded as a change in the reflected light by a detector. This setup enables real-time measurement of biomolecular interactions, with refractive index changes proportional to mass changes.

The standard SPR instrument contains a sensor chip with four flow cells, which enables measurement of four different ligands at a time. One side of the chip is attached to the SPR detection system. On the other side is microfluidic system, which forms four parallel channels on the sensor chip surface and facilitates buffer flow and injections. The standard chip carries on the top of the gold layer, a carboxymethylated dextran surface, which is used for the immobilization of biomolecules. This surface is characterized by low non-specific binding, except for very basic compounds, which demands consideration since DNA binding proteins frequently carry a basic charge. Several other chips are available, including chips with immobilized streptavidin for easy capture of biotinylated DNA fragments and chips with reduced negative charge, which can help to analyze extremely basic proteins.

6.1: Ligand immobilization

Ligand has to be immobilized on a sensor chip surface by either chemical cross-linking or with an affinity interaction like biotin-streptavidin, nickel chelate-His-Tag or antibody-antigen. To ensure the measurement of accurate date the immobilization procedure should not interfere with the ligand-analyte interaction. The most common immobilization procedure for protein is chemical cross linking, which can reduce the activity or affinity of the interaction due to the modification of the protein or by blocking the binding site.

Immobilization of DNA, on the other hand, can be achieved with biotinylated DNA and streptavidin sensor chips, which results in accessible DNA and a uniform attachment of the ligand. Due to the strong interaction of biotin and streptavidin these chips withstand fairly harsh regeneration conditions, eg., high salt and low concentration of sodium dodecyl sulphate (SDS). The size of the immobilized DNA should correspond to the length of the DNA footprints plus a few extra bases (3-6 bp) on each side as a spacer. The biotin should be positioned at the very end of the DNA so it will not interfere with binding. Double stranded DNA is produced by annealing of two oligonucleotides.

6.2: Experimental

During the experiments ligand (DNA) is immobilized on the chip surface (Figure 13). A constant buffer flow over the chip surface precedes injection of a analyte (protein solution). Binding of the protein to DNA can be observed as change in the position of the reflected light minimum, which can be translated into a change of resonance units (RUs); 1 RU corresponds to a change in angle of 0.0001° and 1 $pgmm^{-2}$ of protein. The RUs were recorded and

displayed in real time by a computer. Once the reaction reached equilibrium, the binding and dissociation rates of the protein on DNA were supposed to be equal. Subsequent to the protein injection, buffer was introduced and dissociation of protein was monitored. This led the gradual return of observed minimum in the reflected light to its original position. However, at any point of the dissociation phase a regeneration fluid can be injected into the system, which will allow rapid release of any protein form the chip surface prior to a new round of analyte injection. According, a real time interaction experiment will contain the following parts: an association phase with the protein binding to the immobilized DNA, an equilibrium phase with equal association/dissociation rates for the protein, and a dissociation phase with buffer flow washing away any protein that dissociates for the DNA.

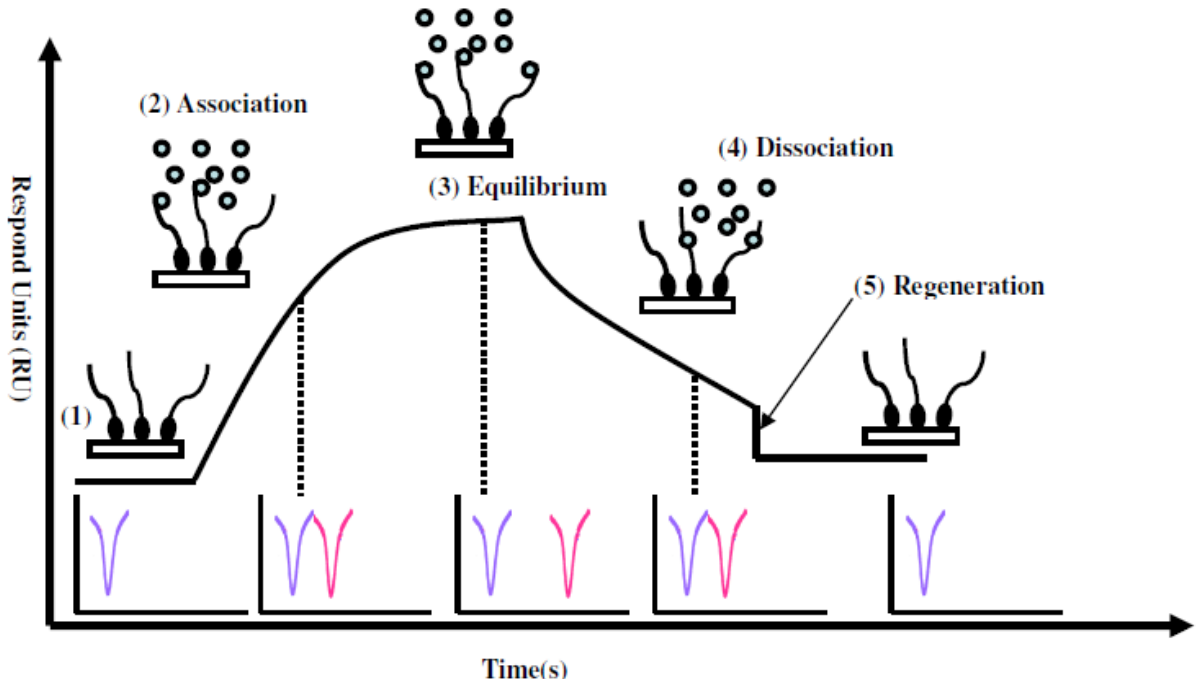


Figure 13: The sensorgram is a graphical representation of the signal produced during a binding experiment. The sensor chip carries immobilized ligand (DNA). (1) Initially buffer is running over the sensor chip, and no change in signal is observed. (2) A DNA binding protein is injected and associated with DNA. This leads to a change in the refractive index, which is monitored in the SPR instrument as change in angle where the drop in intensity of the reflected light is observed. This change in angle is visualized in the sensorgram. (3) Association and dissociation of the protein and the DNA occur at the equal rates during the equilibrium. (4) Protein injection is replaced by buffer flow and dissociation of the protein from the DNA is observed. This leads to a reduction in refractive index and the angle where the drop of intensity is observed. (5) At any point of the dissociation regeneration liquid can be injected, which will remove all the protein from the chip surface.

6.3: Analysis of Data

Before data analysis, the experimental response curves require processing. The unwanted part of a sensorgram, such as a very long baseline before injection and regeneration, can be removed. The baseline must be adjusted to zero since most of the fitting algorithms require that. The response curves must be aligned, usually by setting t = 0s as the start of the injection for each curve, and finally the response in the reference cell should be subtracted. The subtraction can be carried out "on line" during the experiment or afterwards. Often, even though both curves were aligned on the time axis, the subtraction of the reference cell data results in the appearance of spikes in the sensorgrams, especially at the beginning of both association and dissociation phases. These spikes can be esthetically removed or can be left in the sensorgrams, since these regions of the response curves are usually left out during the quantitative analysis. These sensorgram distortions should be removed for the analysis.

6.3.1: Association

During the injection phase of an SPR experiment, termed association, the kinetic processes of DNA/PROTEIN complex formation and its decay can be represented by the scheme:

$$DNA + \mathrm{Pr}otein \underset{k_{off}}{\overset{k_{on}}{\Leftrightarrow}} DNA/PROTEIN$$

The rate of DNA/PROTEIN complex formation is expressed as:

$$\frac{d(DNA/PROTEIN)}{dt} = k_{on}[DNA][PROTEIN] - k_{off}[DNA/PROTEIN]$$

where [DNA] and [PROTEIN] are the concentrations of free or unbound DNA and protein, respectively. Assuming that DNA is immobilized at the chip surface and protein is injected, this equation can be written using SPR terminology as:

$$\frac{d(R)}{dt} = k_{on}((R_{\max} - R) * C - k_{off} * R \quad \textbf{Equation (41)}$$

Where R is the response, R_{max} is the response at DNA saturation by protein, and C is the concentration of injected protein (the amount of protein bound to DNA is negligible.)

Taking the natural logarithm of Eq. 41

$$\ln\frac{d(R)}{dt} = \ln(k_{on} * R_{\max} * C) - (k_{on} * C + k_{off}) * t \quad ,$$

And substituting

$$k_{obs} = k_{on}*\mathrm{C} + k_{off} \quad , \qquad \textbf{Equation (42)}$$

results in the equation:

$$\ln\frac{d(R)}{dt} = -k_{obs} * t + \ln(k_{on} * R_{\max} * C) \quad \textbf{Equation (43)}$$

Since for a given protein concentration C, k_{obs} and $\ln(k_{on}*R_{max}*C)$ are constant, Eq. (42) represents a linear function. The $\ln\frac{d(R)}{dt} = f(t)$ function can be used for evaluation of DNA/PROTEIN complex formation rate constant k_{on}. The plot k_{obs} vs C should be linear (Eq. 42) and the slope of this function represents the k_{on} rate constant.

6.3.2: Dissociation

During the dissociation phase of an experiment, the protein solution is replaced by buffer and the response change represents solely the decay of the DNA/PROTEIN complex.

$$DNA/PROTEIN \xrightarrow{k_{off}} DNA + \mathrm{Pr}otein$$

The rate of this proces is expressed as:

$$\frac{d(R)}{dt} = -k_{off} * R \qquad \textbf{Equation (44)}$$

The solution of this differential equation is:

$$R(t) = R_0 * e^{-k_{off}*t}$$

This can be used for evaluation of the k_{off} rate constant by fitting to a single exponential or it can be converted to equation:

$$\ln\frac{R_0}{R_t} = k_{off}(t_0 - t_0) \quad \textbf{Equation (45)}$$

Where R_0 is the response at the analysis start point t_0 (which is not necessary at the beginning of the dissociation phase)

The slope of this function represents the rate constant k_{off}.

6.3.3: Equilibrium

This phase of the experiment can be analyzed as a special case of the association phase when the rate of DNA/PROTEIN complex formation equals the rate of its decay. By definition, at this point no change of response is observed:

$$\frac{d(R)}{dt} = 0$$

At equilibrium Eq. (1) can be written as:

$0 = k_{on}*(R_{max} - R_{eq})*C - k_{off} - R_{eq}$

Rearranging for R_{eq} and dividing by k_{on} results in:

$$R_{eq} = \frac{C * R_{\max}}{C + \frac{k_{off}}{k_{on}}}$$

Finally, subsituting dissociation constant $K_d = \frac{k_{off}}{k_{on}}$ produces the equation:

$$R_{eq} = \frac{C * R_{\max}}{C + K_d} \qquad \textbf{Equation (46)}$$

Equation 6 can be used for evaluation of the dissociation constant K_d, since K_d is the protein concentration at which $R_{eq} = 0.5*R_{max}$ (Fig). Equation (46) can be rearranged, especially in the case where saturation is not reached, using the Scatchard plot:

$$\frac{R_{eq}}{C} = -\frac{1}{K_d} R_{eq} + \frac{1}{K_d} R_{\max} \quad \textbf{Equation (47)}$$

Where the slope of the linear function $\frac{R_{eq}}{C} = f(R_{eq})$ represents $-\frac{1}{K_d}$.

6.3.4: Stoichiometry Determination

The response measured in SPR measurements is not only proportional to the amount of protein-DNA bound to a chip surface (a prerequisite of kinetic analysis), but also the change in the bulk refractive index per unit change in protein concentration (specific refractive index increment) is closely similar for a wide range of proteins and nucleic acids. In other words, an equal mass of two different protein (or protein and DNA) bound to the chip surface will give the same response value. The SPR technology can therefore be used for estimation of the DNA-protein complex stoichiometry according to the equation:

$$n = R_{\max} \frac{MW_{DNA}}{R_{DNA} * MW_{Protein}} \qquad \textbf{Equation (48)}$$

Where,

n = the number of protein molecules bound to DNA

= the number of protein binding sites on DNA (assuming that a single protein molecule binds a single binding site)

R_{max} = response for Saturating concentration of protein

R_{DNA} = amount of immobilized DNA (RU) and

MW_{DNA}, $MW_{Protein}$ = molecular weight of DNA and protein, respectively.

7: Single Molecule Fluorescence Spectroscopy

7.1: Imaging plateform

7.1.1: Two Photon Excitation Scanning Fluorescence Microscope (TPE-SFM)

This platform has already been described elsewhere (ref). Briefly, the two-photon excitation beam passes through a beam expander that increases the diameter of the laser so that it fills the back aperture of the objective. The resulting excitation volume is about 0.3 fL and almost gaussian (0.3µm radi; ~ 1.1µm on the z axis). Two galvanometer mirrors are added on the excitation pathway in order to allow the scanning of the sample. The fluorescence emission of the sample is collected through a band pass filter (585/40) with an Avalanche PhotoDiode (APD). Images are reconstituted using a lab-written Labview procedure. The excitation wavelength is adjusted by setting the monochromator at 415nm (Pulse ~ 100fs; Repetition rate 80MHz). The laser power is kept at a low level (~20mW). The speed of scanning is 4µs per voxel, sufficient for collecting the needed photons and to limit the bleaching of the sample.

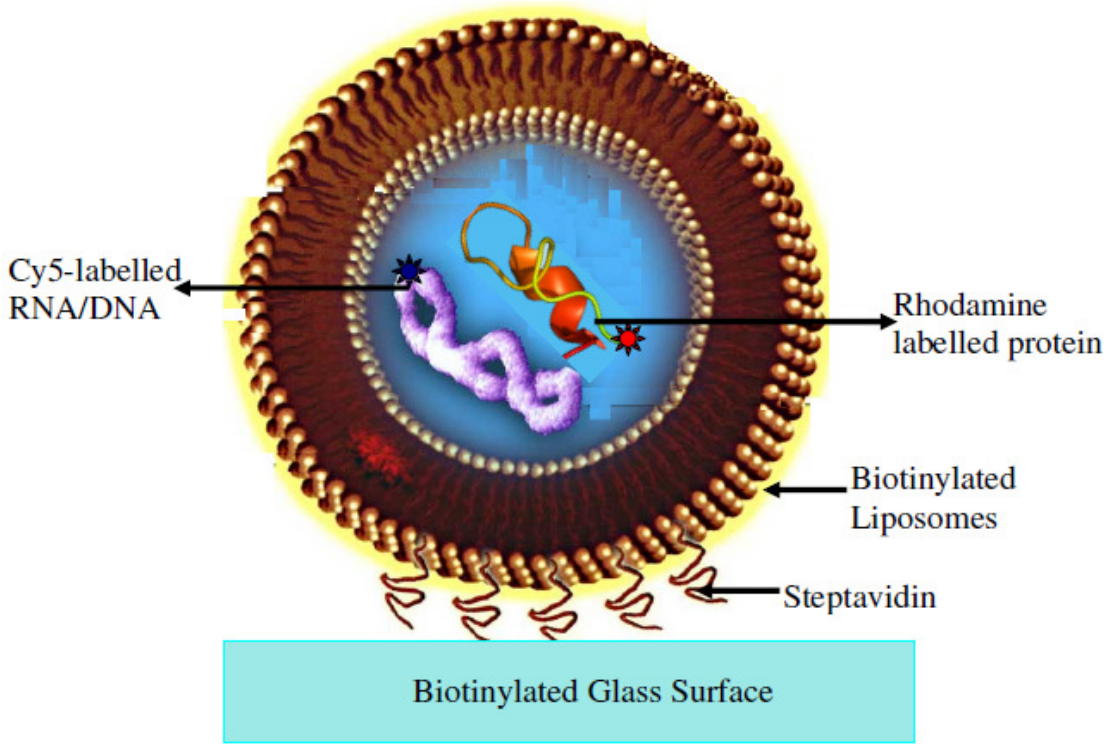
7.1.2: Widefield / TIRF Fluorescence Microscope

This platform has been built in order to perform fluorescence resonance energy transfer (FRET) at the single molecule level. Two CW lasers (532nm (100mW) for the excitation of donors of fluorescence within FRET couples and 635nm (60mW) for the direct excitation of Cy5) (Cobolt) are available. The excitation light is selected by computer-controlled electromechanical shutters. A telescope (f'_1 = 3cm and f'_2 = 15cm) is used to expand the laser beam. The laser is then directed into the microscope with an adjustable mirror that allows switching between the wide-field and the TIRF illumination mode. The fluorescence emission

of the sample is collected through the oil-immersed objective (X100, N.A.= 1.4), spatially restricted by a home-made slit and splitted (dichroic mirror 650nm) in two symmetric detection paths where the fluorescence emissions are band-pass filtered (585±30nm, 680± 20nm). The two signals are finally imaged together on the same EM-CCD camera (Hammatsu).

7.2: Molecule immobilization

Single pair molecule are trapped inside lipid vesicles which are tethered to a glass surface via biotin/streptavidin coupling, in order to overcome the problem of molecule-surface interaction.



7.2.1: Preparation of the anchoring surface

Microscope coverslips were first deeply cleaned in a piranha-like solution of freshly prepared peroxodisulfuric acid. Coverslips were then thorough rinsed with Milli-Q water and dried. All remaining fluorescent particles on the surface were then bleached by treating coverslips for one hour in a UV-Cleaner. The streptavidin sub-layer was built one layer by one layer from the nude glass surface according to the following steps: (i) grafting of aminosilane ((3 Aminopropyl)-triethoxysilane) on the glass support, (ii) biotinylation of the surface amino groups ((+)-Biotin N-hydroxysuccinimide ester = biotin amidocaproic acid 3-sulfo-N-

HydroxySuccinimide ester) and finally (iii) docking of streptavidin. This surface preparation allows the immobilization of intact vesicles, immobile at the micrometer scale, which can be maintained without fusion of the LUV for hours.

7.2.2: Preparation of Large Unilamellar Vesicles (LUV)

18:1 Biotinylated-PE was mixed with Egg Yolk PC in 1:100 molar ratio in a chloroform solution which then dried under vacuum by rotary evaporation in order to obtain a lipid film onto the wall of a round-bottom flask. This lipid film was maintained under vacuum for at least 30 min to remove any residual solvent, and then hydrated with buffer containing the fluorescently-labelled partners at the appropriate concentration to induce swelling of the film lipid. 2 mg/mL of Lipid concentration for multilamellar vesicles (MLV) were obtained by vigorous vortex shaking for two minutes. Regular large unilamellar vesicles (LUV) were obtained by extruding the MLV through polycarbonate membranes with 200-nm or 100-nm calibrated pores (Avanti Mini Extruder). LUVs were then purified by size exclusion chromatography.

7.2.3: Liposome coupling on the coverslips

In order to immobilize the probe containers for single molecule measurements, liposomes were attached to the glass coverslips by using the biotin/avidin chemistry. The final streptavidin coating of the coverslips was obtained by the successive layer by layer treatment of the initially nude glass surface. The liposome solution was finally diluted to the appropriate concentration and incubated for a few minutes on the streptavidin layer. After rinsing with buffer to eliminate the inefficiently bound vesicles, coverslips covered with the encapsulated LUVs were mounted on a glass slide using a 0.5mm silicon spacer (Press-to-Seal, Invitrogen).

7.2.4: Oxygen Scavenger System

A usual oxygen scavenger system was used to deplete the media from O_2, which is responsible for faster bleaching of the dyes due to the oxidation of their activated states (more reactive than the ground state). This system is composed of Glucose Oxidase (GOD) associated to Catalase (CAT), acting in a two step reaction in which Glc is first oxidized, so that one molecule of O_2 is consumed, according to:

$$\beta\text{-D-Glc} + \text{GOD} \rightarrow H_2O_2 + \text{D-glucono-1,5-lactone}$$

then, the generated hydrogen peroxide is transformed by the CAT

$H_2O_2 \rightarrow H_2O + ½\ O_2$

The global yield of one single turn of this reaction is resulting in the disappearance of half the molecules of O_2.

7.3: Imaging of construct tethered liposomes by TPE

Imaging of the construct (Acceptor or Donor labeled ODN/protein encapsulated in liposome) is recorded using a TPE-SFM set-up depicted in 7.1.1. Vesicles encapsulating the TMR appeared as an immobile bright point, just above the glass/water interface. Since the size of the vesicles are well below the resolution of the SFM (= λ/(2.NA); ~ 400 nm), no conclusion can be obtained on their exact shape and size, but the homogeneity of the preparation can be observed. A threshold was applied to the image to discard the background noise. Spots were then selected if they were included within a minimum and maximum area of concomitant pixels. At the exception of some bigger spots (probably fused vesicles), most of the spots presented a homogeneous brightness and area. These spots are stable for at least a few hours at room temperature.

7.3.1: Fluorescence traces of single liposome encapsulating few TMR molecules

The fluorescence emission of the vesicle was then investigated. The photon flux from the sample was recorded. As the bleaching of all the dyes was very rapid, the record of the photon collection was started before the laser was manually triggered to focus on the sample, to collect all the photon emitted. The time needed to bleach the 10 expected TMR molecules was less than 100ms, suggesting that the time before bleaching of a single TMR was about 15ms. Integration of the peak gives around 1500 photons emitted per molecule, in line the collection efficiency expected from the about 10^6 photons emitted before bleaching usually considered for organic dyes.

7.4: Image processing and data analysis

Data are acquired with HCImage software (Hamamatsu) generating .cdx files. For image processing, these files are exported in time-series of .tif images and are then processed with a

home-made macro written under ImageJ. The purpose of the data processing was to generate a 3D matrix ((x; y) images versus time (z)) and to generate a Z-profile on a selected area to reconstruct fluorescence traces of spots. Spots coordinates were selected by a composite image on which a threshold was manually applied. This composite image resulted from a pixel-by-pixel z-standard-deviation image multiplied by a summed-intensities image, both generated from the $(n+1)^{1/2}$ first frames of the sequence. The use of this composite image thus allowed to select brighten spots, for which the fluorescence was the most fluctuating. The choice of this image resulted from different trials (maximum intensity, summed intensities or different composite images) to find an optimal spots selection.

Chapter 1: Fluorescence Techniques to Characterize Ligand Binding to Proteins

Fluorescence Techniques to Characterise Ligand Binding to Proteins

PASCAL DIDIER, KAMAL KANT SHARMA AND YVES MÉLY*

Laboratoire de Biophotonique et Pharmacologie, UMR 7213 CNRS, Université de Strasbourg, Faculté de Pharmacie, 74, Route du Rhin, 67401, ILLKIRCH Cedex, France

To characterise protein/ligand interactions, several critical questions need to be addressed. Among these questions, the most important ones are: (i) Which ligand is likely to bind to a particular protein target?, (ii) Which particular amino acids or structure are required to bind the ligands? and (iii) What are the thermodynamic and kinetic parameters that govern the interaction? These three basic questions could be summarised as the who, where and why questions. Among the biophysical methods used to answer these questions, fluorescence techniques are particularly interesting, due to the specific advantages they offer and that will be reviewed herein. The "who" question is generally answered by medium- and high-throughput techniques, mainly by using fluorescence anisotropy and FRET-derived techniques. Since several excellent reviews already exist on the application of fluorescence techniques for screening large libraries of molecules in order to identify potential

RSC Biomolecular Sciences No. 22
Biophysical Approaches Determining Ligand Binding to Biomolecular Targets: Detection, Measurement and Modelling
Edited by Alberto Podjarny, Annick Dejaegere and Bruno Kieffer

Published by the Royal Society of Chemistry, www.rsc.org

ligands for a given protein,[1–13] we will not cover this aspect in this review. Thus, we will mainly focus on the where and why questions. As we will see, while fluorescence techniques can hardly compete with structural techniques such as NMR and X-ray on crystals for answering the "where" question, they are extremely competitive to address the "why" question in dilute solutions.

5.1 Basic Principles of Fluorescence Spectroscopy and Main Fluorescence Parameters Used for Monitoring Molecular Interactions

5.1.1 Basic Principles of Fluorescence Spectroscopy

Fluorescence describes the emission of light by a substance that has been excited by a light of lower wavelength than the wavelength of the emitted light. To describe absorption and emission of light, one has to consider light as a flux of particles called **photons** that carry an energy $E = hc/\lambda$ (where h is the Planck constant, c/the speed of light and λ the wavelength of light).[14] The relevant spectral domain of wavelengths in fluorescence spectroscopy extends from the UV to the near IR (300–900 nm).

For absorption, the photons interact with the electrons involved in chemical bonds, inducing an electronic transition from an occupied orbital to an unoccupied orbital of higher energy. In the ground state, a molecule is mainly in the lowest electronic and vibration level, denoted E_0, v_0. The electronic transition is characterised by a transition moment that represents the transient dipole resulting from the displacement of charges during the transition. The transition moment is of major importance in experiments using linearly polarised light, since molecules with absorption transition moments parallel to the electric vector of the polarised incident light are preferentially excited. The probability of excitation is proportional to the square of the scalar product of the transition moment and the electric vector.[15]

Depending on the nature of the compounds, various types of molecular orbital can be distinguished. σ orbitals are associated with single bonds. In this case, the electrons are strongly bound and thus high-energy photons (< 180 nm) must be used to induce electronic transitions ($\sigma \rightarrow \sigma^*$). π-orbitals are associated with double and triple bonds and are characterised by much lower energies. Depending on the conjugation of the π-electrons, the electronic transitions ($\pi \rightarrow \pi^*$) extend from the UV to the visible range. Organic groups with conjugated π-electrons giving strong absorption of light are called **chromophores**. Molecules may also possess n-orbitals associated with non-bonding electrons on heteroatoms such as oxygen or nitrogen. The associated transition denoted $n \rightarrow \pi^*$ requires high-energy photons (170 to 250 nm).[15]

The absorption properties of a compound can be measured from the transmission of a light beam through a solution containing the molecules of

interest. At a given wavelength, the absorbance of a compound is given by the Beer–Lambert law:

$$A = \log\left(\frac{I_0}{I}\right) = \varepsilon c l \tag{5.1}$$

where I_0 and I are the incident and transmitted beam, ε the absorption coefficient, l the length of the optical path and c the concentration of the absorbing molecule. The absorbance is an absolute and additive parameter, linearly related to the concentration of the absorbing molecule. Absorption spectroscopy is thus a major technique for determining concentrations of molecules in solution. However, the major disadvantages of this technique are its weak sensitivity (limited to μM range), limited concentration dynamics (mM to μM) and the possible contribution of other species in the spectral range of interest.

Once the molecule is excited to a given vibrational level of the first excited state E_1, v_i, the system relaxes very rapidly (<1 ps) to the lowest vibrational state of the electronically excited state E_1, v_0. From this state, several processes compete to release the energy excess of the molecule, as described in the Jablonski diagram (Figure 5.1A). Fluorescence is generally the only radiative mode and corresponds to the transition from E_1, v_0 to one of the vibrational states of E_0: E_0, v_i. The emitted photons thus have a lower energy than the absorbed ones, resulting in a shift of the wavelengths to the red, called Stokes shift (Figure 5.1B). Numerous non-radiative modes, such as internal conversion, intersystem crossing, quenching by internal and external quenchers and non-radiative energy transfer can compete independently with fluorescence. Since changes of the local environment of fluorescent molecules can affect in a complicated way the different modes, interpretation of the changes in the fluorescence intensity for instance during a binding process will be difficult. Moreover, for most molecules the de-excitation pathways are so efficient that only a few of them are fluorescent and can be used as fluorophores.[16,17]

Two absolute parameters can be determined from the kinetic constants. The first one is the fluorescence lifetime τ:

$$\tau = \frac{1}{k_f + k_{NR}} \tag{5.2}$$

where k_f is the kinetic constant associated to fluorescence emission and k_{NR} the sum of all the non-radiative processes. For most of the fluorescent molecules, τ ranges in nanosecond time scale. Fluorescence lifetimes can be measured by using a pulsed light source (*e.g.* picosecond or femtosecond pulsed laser with pulse duration much shorter than fluorescence lifetimes) and a Time-Correlated Single Photon Counting (TCSPC) device.[16,17] By measuring individually the arrival time of each fluorescence photon on the detector with respect to the excitation pulse, it is possible to reconstruct the fluorescence decay. In the ideal case, the fluorescence decay displays only one fluorescence lifetime, as for

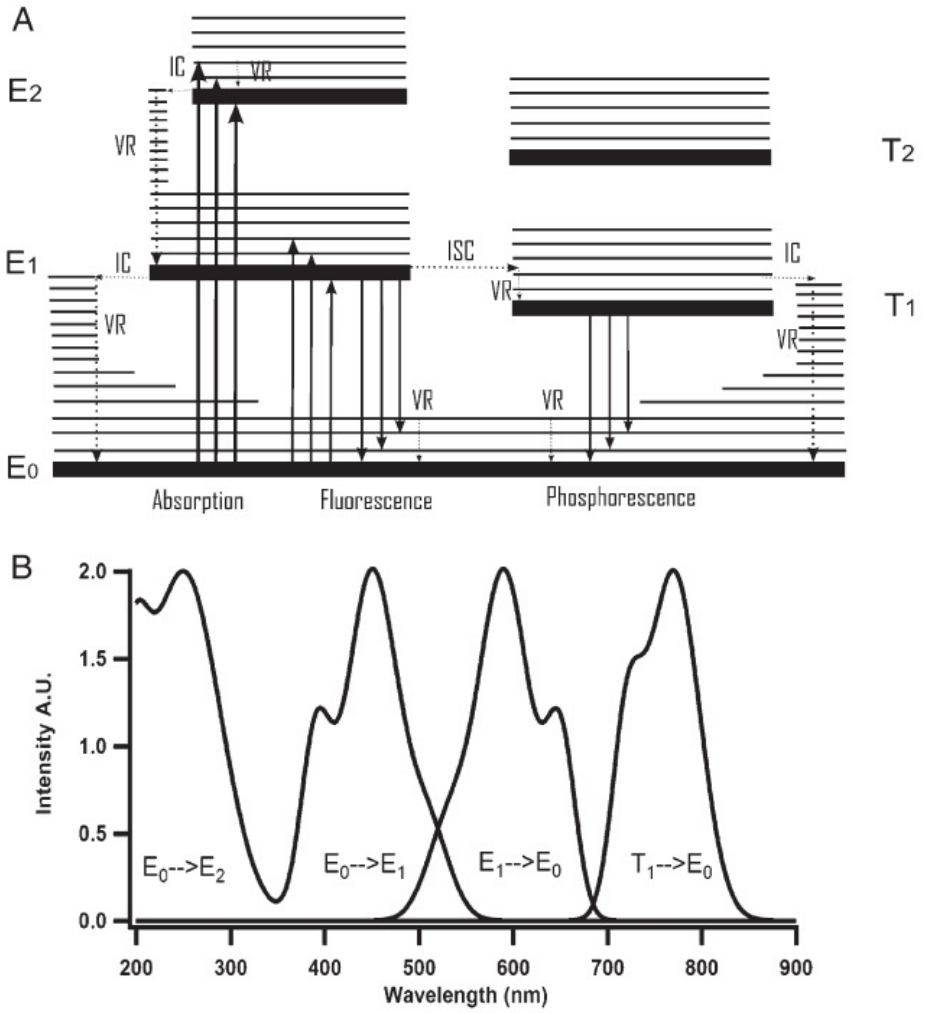


Figure 5.1 Jablonski diagram and spectra resulting from transition between electronic states. (A) Jablonski diagram illustrating the electronic states of a molecule and the transitions between them. E_1 and T_1 designate the singlet and triplet states, respectively. Radiative and non-radiative transitions are indicated by solid and dotted arrows, respectively. IC, internal conversion, VR, vibrational relaxation, ISC, intersystem crossing. The lowest vibrational state of each electronic state is indicated with a thick line. Higher vibrational states are indicated with thinner lines. (B) Absorption and emission spectra corresponding to the different transitions displayed on the Jablonski diagram.

instance with fluorophores that are insensitive to their environment. But, in general, fluorophores exhibit several lifetimes, each one being associated with a particular conformation in respect to the surrounding medium. The second absolute parameter is the fluorescence quantum yield defined as the ratio between the emitted and absorbed photons:

$$\phi = \frac{k_f}{k_f + k_{NR}} = k_f \tau \tag{5.3}$$

Quantum yields ϕ are usually determined with respect to a reference chromophore:

$$\phi = \phi_R \frac{\int_0^\infty I(\lambda_E, \lambda_F) d\lambda}{\int_0^\infty I_R(\lambda_E, \lambda_F) d\lambda} \times \frac{A_R(\lambda_E)}{A(\lambda_E)} \times \frac{n^2}{n_R^2} \tag{5.4}$$

where ϕ_R is the quantum yield of the reference, the two integrals describe the emission spectra of the compound and the reference, respectively, A and A_R are the absorbance of the chromophore and the reference at the excitation wavelength, λ_E, and n and n_R are the refraction indexes of the solvents.

5.1.2 Main Fluorescence Parameters Used for Monitoring Molecular Interactions

The simplest parameter used to monitor molecular interactions is the fluorescence intensity I_F, which at a given excitation wavelength λ_E and emission wavelength λ_F, is given by:

$$I_F(\lambda_E, \lambda_F) = kF(\lambda) I_0(\lambda_E)(1 - 10^{-A(\lambda_E)}) \tag{5.5}$$

where k is a factor depending on the apparatus used, $F(\lambda)$ is the total fluorescence intensity emitted by the sample, $I_0(\lambda_E)$ the excitation intensity and $A(\lambda_E)$ the absorbance at the excitation wavelength. I_F is thus a relative parameter (expressed in arbitrary units), which is linear with respect to the concentration of the fluorescent species, only for absorbance values lower than 0.1. For higher absorbance values, I_F tends to a plateau value and then decreases due to inner filter effects. The fluorescence signal is usually measured with a spectrofluorimeter, which measures the fluorescence emission perpendicularly to the excitation beam to minimise its contribution to the recorded signal. Fluorescence intensities of environment-sensitive fluorophores are frequently used for monitoring molecular interactions.[16–18] The fluorophore environment can change either due to the binding of the interacting partner in its close vicinity or due to remote conformational changes evoked by the binding process. Fluorescence intensities are easy to measure and do not require sophisticated instrumentations. However, due to the multiple pathways involved in the deexcitation of an excited molecule, the fluorescence intensity changes accompanying a given binding process are difficult to predict. These predictions are even more difficult when the structure of the complex and, thus, the position of the fluorophore in respect to the binding partner are unknown. In addition, fluorescence intensities can be affected by inner filter effects, contributions from background fluorescence as well as Raman and Rayleigh scattering and binding-induced events such as aggregation. Finally, being a relative value, fluorescence intensity can hardly be used in the cellular context, to quantitatively investigate molecular interactions.

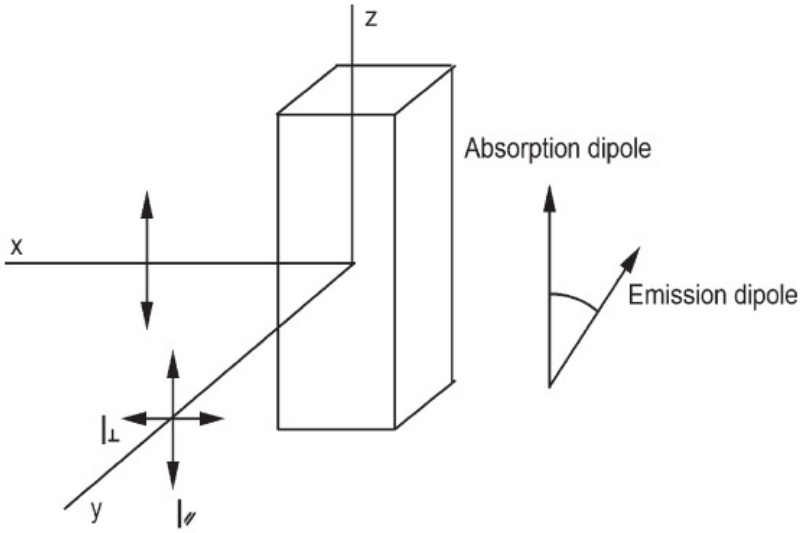


Figure 5.2 Principle of fluorescence anisotropy measurements. The molecules in solution are excited by a vertically polarised beam and the fluorescence emitted at 90° is recorded with a polariser to measure $I_{\perp}$ and $I_{//}$ to calculate the anisotropy parameter *r*. Excitation by polarised light photoselects the molecules with absorption transition moments parallel to the light electric field. The fluorescence anisotropy measures the depolarisation of light, which results notably from the rotational diffusion of the labelled molecules during their excited state lifetime.

Another common fluorescence parameter used for monitoring molecular interactions is fluorescence anisotropy (Figure 5.2), which requires linearly polarised light.[1,10,19–21] Due to the preferred absorption of chromophores along their transition dipole moment, illumination of a population of fluorophores with a linearly polarised incident light creates a photoselection. As a consequence, the distribution of excited molecules is anisotropic, leading to a polarised emission. The extent of polarisation of the emitted light is measured through the fluorescence anisotropy, *r*, given by:

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \tag{5.6}$$

where $I_{\parallel}$ and $I_{\perp}$ describe the fluorescence intensity recorded through a polariser oriented parallel and perpendicular, respectively, to the direction of the polarised excitation. One major cause of depolarisation of the emitted light is the rotational diffusion of the fluorescent molecules during their excited state lifetime. Thus, *r* measures the average angular displacement of the fluorescent molecule between the absorption and subsequent emission of a photon. As a consequence, this parameter depends critically on the size and shape of the rotating molecule, as well as on the viscosity of the medium and the fluorescence lifetime of the fluorophore. Using time-resolved techniques, it is possible to obtain separately the rotational correlation time associated to the local motion of the fluorophore and that associated with the overall tumbling of the

molecule on which the fluorophore is bound. Importantly, the last correlation time is directly proportional to the molecular weight of the tumbling species, thus directly giving access to the size of the complexes formed in an interaction process. This applies only if the correlation time is comparable with the fluorescence lifetime. For instance, fluorophores with a fluorescence lifetime of 5 ns will not allow measuring a correlation time larger than 100 ns, which corresponds to the tumbling of spherical species or complexes with a molecular weight of about 200 kDa.[22]

Resonance energy transfer (RET) is also commonly used to monitor molecular interactions. RET is based on a non-radiative energy transfer between a molecule acting as an energy donor (D) and a chromophore acting as an energy acceptor (A).[16–17,23,24] The transfer results from a long-range dipole–dipole interaction between D and A, and strongly depends on the distance between the two. As a consequence, RET could be used as a spectroscopic ruler to determine intermolecular distances between 1 nm and 10 nm. Several variants based on this principle have been developed. The most familiar one is Fluorescence (or Förster) Resonant Energy Transfer (FRET), in which a D/A couple of chromophores is used (Figure 5.3A). As a result of FRET, the fluorescence of D decreases and the fluorescence of A increases (Figure 5.3B). The rate of energy transfer, k_T, from D to A is given by:

$$k_T = \frac{1}{\tau_d^0}\left(\frac{R_0}{R}\right)^6 \tag{5.7}$$

where τ_d^0 is the fluorescence lifetime of D in the absence of A, and R is the interchromophore distance. The Förster distance R_0 is the D/A distance, for which the energy transfer efficiency is 50%. It is calculated as:

$$R_0 = (8.79 \times 10^{23} n^{-4} Q_D \kappa^2 J_{AD})^{1/6} \tag{5.8}$$

where n designates the refractive index of the medium (a value of 1.333 is usually taken), Q_D is the quantum yield of the donor and J_{AD} is the overlap integral between the emission spectrum of the donor and the absorbance spectrum of the acceptor. The orientation factor $\acute{e}^2$ depends on the relative orientation of the emission transition dipole of D and the absorption transition dipole of A. The values of κ^2 vary between 0 (when the transition dipoles are perpendicular) to 4 (when the transition dipoles are parallel). Frequently, the $\acute{e}^2$ value is assumed to be 2/3, which corresponds to the case where donors and acceptors randomise prior to energy transfer. The R_0 values are typically between 20 and 60 Å. The FRET efficiency E is measured independently, through several approaches. The most common one is through the decrease of the quantum yield or the fluorescence intensity of D that resulted from the addition of A:

$$E = 1 - \frac{\phi_{DA}}{\phi_D} = 1 - \frac{I_{DA}}{I_D} \tag{5.9}$$

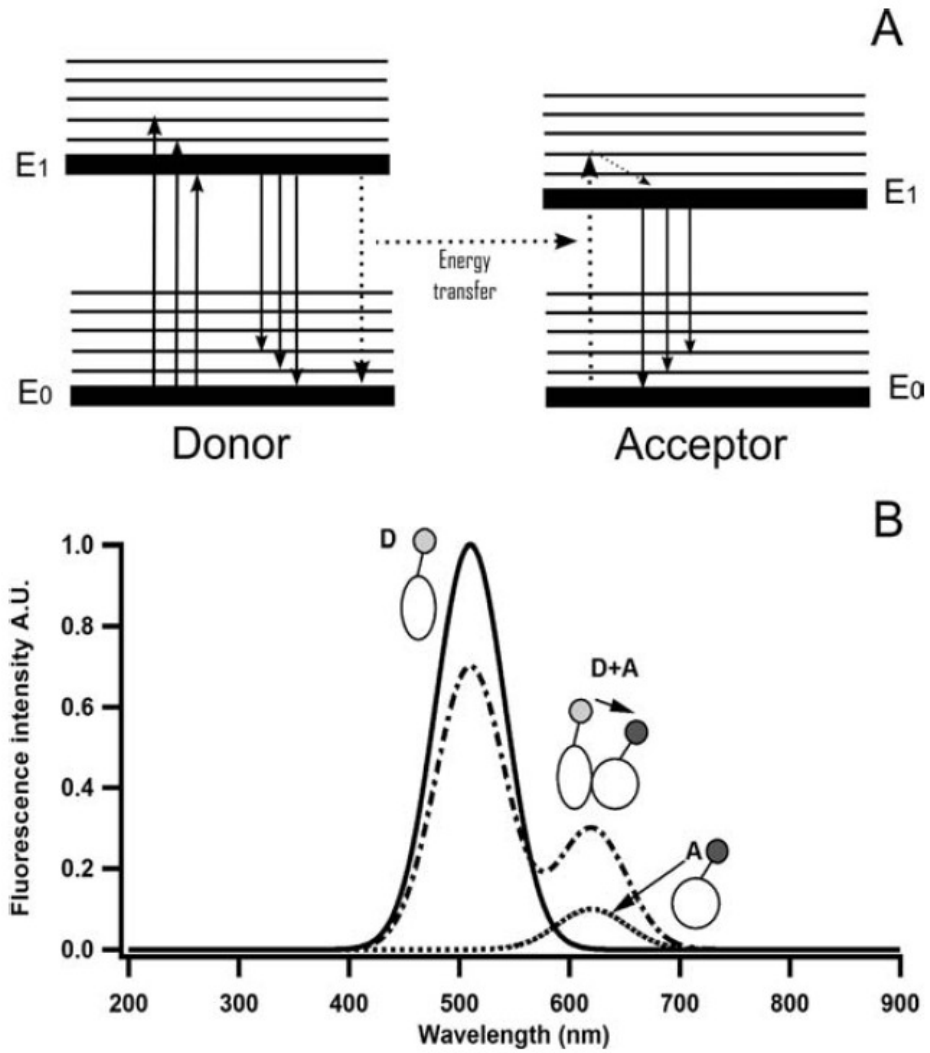


Figure 5.3 Fluorescence Resonance Energy Transfer (FRET). (A) Principle of FRET between a donor/acceptor couple. After excitation, the donor (D) can transfer non-radiatively its energy to an acceptor (A) through dipole–dipole coupling, provided that the donor and acceptor are close in space (< 10 nm) and that the emission spectrum of D and the absorption spectrum of A overlap. (B) Evidence of FRET from emission spectra. FRET between A and D decreases the emission of D and enhances the emission of A. Emission of A, when alone, results from its direct excitation at the excitation wavelength of D.

where ϕ_D and ϕ_{DA} are the quantum yields in the absence and the presence of the acceptor, respectively, while I_D and I_{DA} are the corresponding fluorescence intensities. It should be noted that intensities (usually recorded at the maximum emission wavelength) could be used only if there is no shift in the emission spectrum of D subsequent to the binding of A. Alternatively, E could also be calculated from the enhanced A emission by:

$$E = \frac{A_A(\lambda_D)}{A_D\lambda_D}\left[\frac{I_A(\lambda_A^{em})}{I_A^0(\lambda_A^{em})} - 1\right] \tag{5.10}$$

where A_A and A_D are the absorbance of A and D, respectively, at the excitation wavelength of D. I_A^0 and I_A are the fluorescence intensities of A in the absence and the presence of D, respectively. Finally, E can be determined from time-resolved measurements of D, through:

$$E = 1 - \frac{\tau d}{\tau_d^0} \tag{5.11}$$

where τ_d^0 and τ_d are the fluorescence lifetimes of D in the absence and in the presence of A, respectively. This approach is usually the most powerful one, since in contrast to steady-state measurements, lifetimes are absolute parameters insensitive to inner filter effects and do not require to compare D alone and the D/A pair at rigorously the same concentrations, which is a critical advantage in cell measurements. Finally, the interchromophore distance R is calculated from the R_0 and E values using:

$$R = R_0 \left(\frac{1}{E} - 1 \right)^{1/6}. \tag{5.12}$$

The major disadvantage of FRET experiments relies on their limited signal/noise ratio, especially at dilute concentrations, where the emission of both D and A can be contaminated by the background fluorescence of the other components in solution or the autofluorescence in cells. Moreover, FRET measurements can be biased when D is photobleached. These limitations can be overcome with Bioluminescence Resonant Energy Transfer (BRET), where the fluorescent D is replaced by the luciferase enzyme which emits light as a result of the oxidation of its substrate.[9,25,26] Since no excitation light is needed in this case, no background fluorescence or photobleaching can occur, giving an excellent signal/noise ratio. However, the BRET technique is limited by the low number of photons emitted by the luciferase/substrate system, which increases the measurement time and limits its application in single-cell imaging techniques.

Another alternative to FRET is Lanthanide-based (or Luminescence) Resonance Energy Transfer (LRET), which uses lanthanide complexes as energy donors.[27–30] Due to their poor absorption coefficients, lanthanides need to be complexed with chelates and cryptates that harvest light and transfer it to the lanthanide through intramolecular, non-radiative processes. These lanthanide complexes exhibit unique spectral properties with respect to conventional fluorophores, such as large Stoke shifts and extremely long emission lifetime half-lives (from μs to ms). As a consequence, using time-resolved measurements in the μs to ms range allows full elimination of the background fluorescence, which occurs on the ns time scale. Thus, LRET offers an excellent signal/noise ratio as compared to FRET, and can thus be used even when the energy transfer efficiency is low. Moreover, in contrast to FRET, the unpolarised nature of lanthanides and the ample rotational motions allowed by their long-lived lifetimes makes κ^2 close to 2/3 and reduces the uncertainty in

distance measurements. Moreover, lanthanides can be associated with strongly absorbing chromophores, such as allophycocyanin and phycoerythrin, which provides R_0 values up to 100 Å and thus allows measuring larger inter-chromophore distances than FRET. Finally, LRET being measured mainly through the sensitised emission of the acceptor is insensitive to incomplete probe labelling. Due to its distinctive advantages, LRET is a major technique for detection of binding in HTS, under the acronym HTRF (Homogeneous Time-resolved Fluorescence).[31,32]

Finally, another more and more frequently used technique for investigating molecular interactions is Fluorescence Correlation Spectroscopy (FCS).[33–37] This method is based on the analysis of the temporal fluctuations of the fluorescence intensity recorded in the small open volume, defined by a focused laser beam (Figure 5.4A). These fluctuations are analysed using an auto-correlation function calculated by:

$$G(\tau) = \frac{\langle I(t)I(t+\tau)\rangle}{\langle I\rangle^2} \tag{5.13}$$

where $\langle\rangle$ denotes time averaging, $I(t)$ is the fluorescence intensity at time t and τ is the lag time. To get significant fluctuations, FCS requires that the number of molecules in the observation volume should be small. This can be achieved by using the observation volume defined by a confocal microscope or the excitation volume defined by two-photon excitation. Both volumes are generally lower than a femtolitre, which limits the number of fluorescent molecules in this volume to a few copies, when their bulk concentration is in the nM range. Assuming a 3D Gaussian observation volume, the autocorrelation curve is analysed using:

$$G(\tau) = \frac{1}{N}\frac{1}{1+\frac{4D\tau}{\omega_1^2}}\left(\frac{1}{1+\frac{4D\tau}{\omega_2^2}}\right)^{1/2} \tag{5.14}$$

where N is the average number of fluorescent molecules in the observation volume, D is the translational diffusion coefficient and, ω_1 and ω_2 are the small and long axes of the observation volume. From N, ω_1 and ω_2, it is possible to determine the average concentration of the fluorescent species in the observation volume, which provides a unique means to determine local concentrations of molecular species in cells. Moreover, assuming a spherical shape for the fluorescent species, the Stokes–Einstein equation relates D to the hydrodynamics radius of the molecule by:

$$D = \frac{kT}{6\pi\eta r_H} \tag{5.15}$$

where k is the Boltzmann constant, T is the temperature, η is the viscosity and r_H the hydrodynamics radius of the diffusing molecule. D can thus be used to

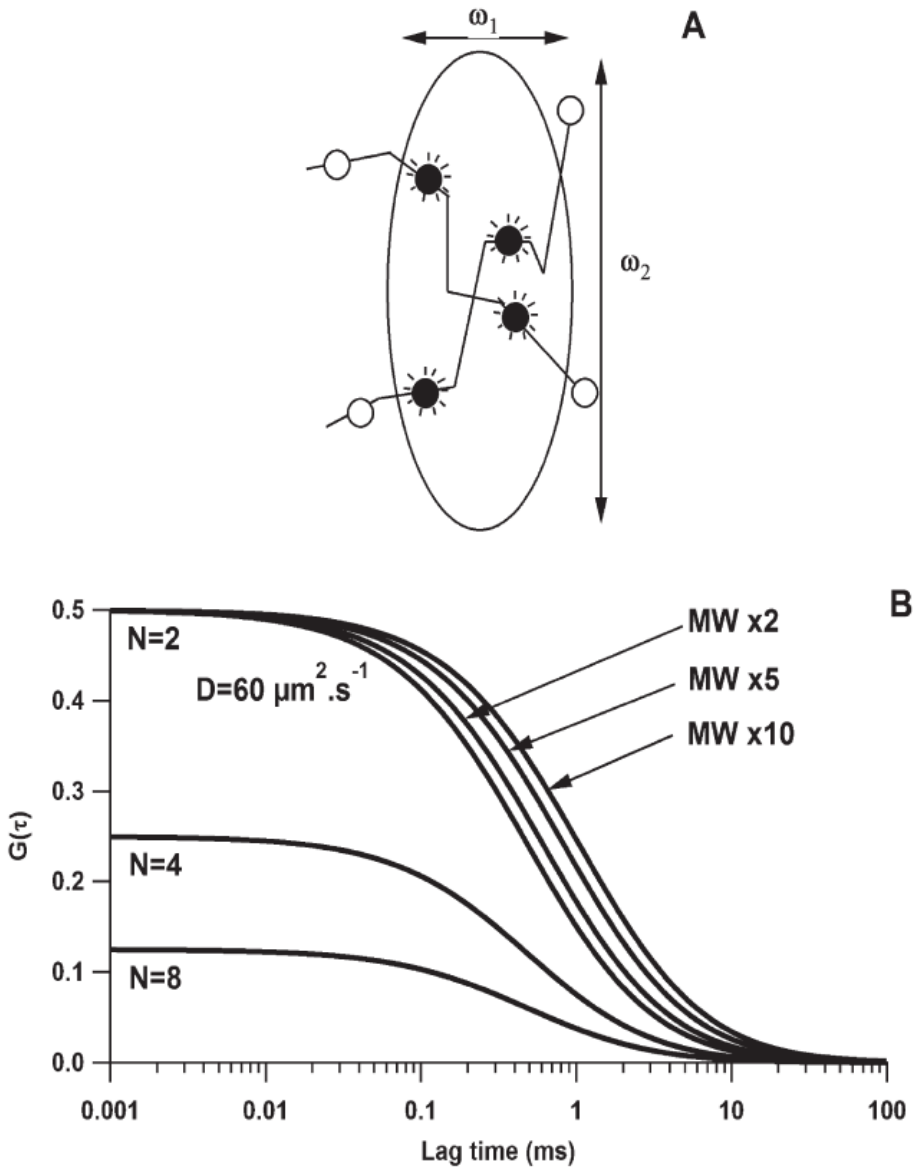


Figure 5.4 Fluorescence Correlation Spectroscopy (FCS). (A) Principle. FCS measurements are based on the analysis of the fluorescence intensity fluctuations in a small observation volume. These fluctuations are mainly related to the diffusion of the fluorescent molecules in and out of this volume. (B) Autocorrelation curves. The autocorrelation curves allow the determination of the average number N of fluorescent molecules in the observation volume, as well as the diffusion constant D of these fluorescent molecules. The value of D being proportional to $MW^{-1/3}$ for spherical molecules, only a limited shift in the autocorrelation curve is observed when the MW of the labelled species increases by a factor of 2. Significant shifts and, thus, clear discrimination in the diffusion constants can only be obtained when the MW increases by a factor of 5, at least.

monitor the increase of size that follows a binding event, but the dependence of D on the increase in molecular weight MW is quite low, since D is proportional to $MW^{-1/3}$ (Figure 5.4B). This means that, for instance, a dimerisation process will only modify D by a factor of 1.25, which is not sufficient for discriminating

dimers from monomers in a mixture. In fact, FCS requires that MW increases at least by a factor of 5 to 8, to provide a clear distinction between free and bound forms.

FCS can also be used to determine the number, n, of fluorescently labelled ligands bound to a biomolecule. To this aim, the molecular brightness, corresponding to the average fluorescence intensity in the observation volume divided by the average number of molecules in this volume, is used. It results that the brightness of a biomolecule/ligand complex should be n times larger than that of the free ligand alone, provided that the quantum yield of the ligand in its free and bound forms does not change. This last point requires an independent measurement in solution. FCS can also be used with a pair of spectrally separated fluorophores, in an approach called Fluorescence Cross-Correlation Spectroscopy (FCCS),[38,39] which cross-correlates the temporal fluctuations of the fluorescence intensities I_1 and I_2 of the two fluorophores measured in two different channels through:

$$G_{cc}(\tau) = \frac{\langle I_1(t) I_2(t+\tau) \rangle}{\langle I_1 \rangle \langle I_2 \rangle} \tag{5.16}$$

With this method, a cross-correlation will be observed only if the two fluorescent species diffuse simultaneously within the observation volume, which requires that the two species are bound (Figure 5.5). FCCS is thus extremely powerful to evidence an interaction and measure the concentration of complexes. FCCS

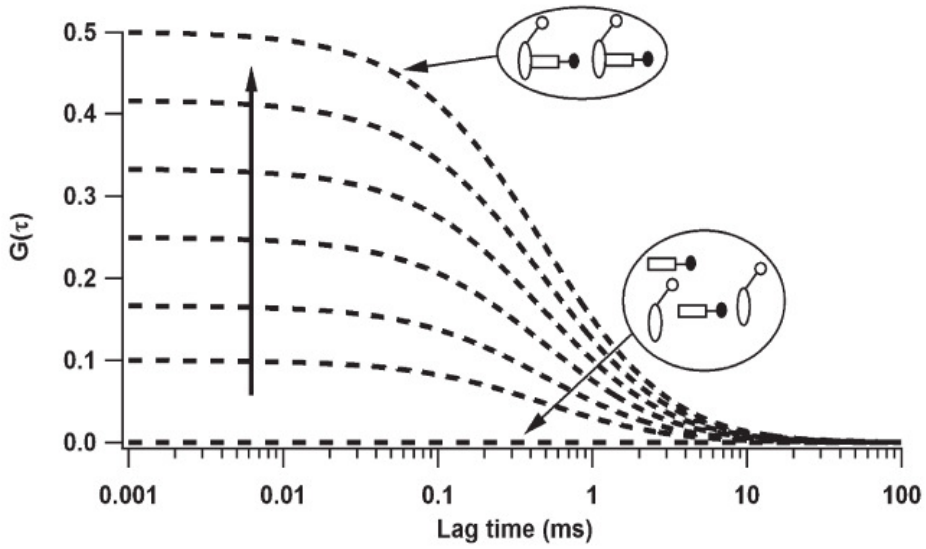


Figure 5.5 Fluorescence Cross-Correlation Spectroscopy. A cross-correlation between the two binding partners labelled with spectrally different labels is observed only if the two partners diffuse simultaneously within the observation volume, which requires that the two partners are bound together. Thus, the amplitude of the cross-correlation function measures the degree of binding (the arrow indicates an increasing degree of binding).

appears as an alternative to FRET since it does not require a close proximity of the chromophores in the complexes.

5.2 Fluorophores

To investigate protein–ligand interactions by using fluorescence-based techniques, at least one of the partners must be fluorescent, either naturally or through the introduction of an external fluorophore. Proteins are generally fluorescent, due to their Trp residues which constitute intrinsic fluorophores exhibiting an emission highly sensitive to environment.[16,17] Binding of a ligand can affect the environment and thus the emission of the Trp residues either by direct interaction with the ligand or through remote ligand-induced conformational changes. Though a large range of proteins/ligands could be monitored through the changes of one of the fluorescence parameters of Trp, this approach has several limitations. The extinction coefficient and the quantum yield of Trp are low, preventing the analysis of highly affine complexes (low nM range and less). Moreover, the absorption and emission of Trp being in the UV range, background fluorescence and inner filter effects from the containers and the reagents in the solution are a major concern. In addition, Trp residues have limited photostability, so that long-term illumination should be avoided. Finally, this approach is limited to proteins of reasonable size since a too large number of Trp averages the effects and makes them difficult to interpret. In contrast to proteins, nucleic acids are almost not fluorescent, with the exception of a few special bases in RNAs. Other natural compounds are fluorescent such as NADH, flavins and pyridoxal phosphate, but their use is limited to specific applications.

As a consequence of the limited abundance and the poor spectroscopic properties of the natural fluorophores, extrinsic fluorophores are frequently used to label proteins or ligands. A large range of compounds is commercially available (Figure 5.6), with generally high quantum yields, high extinction

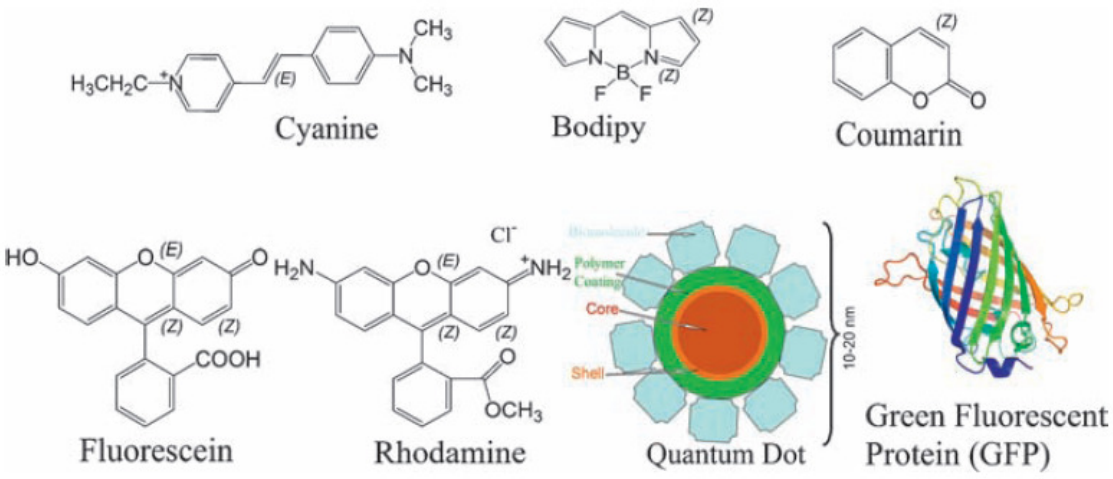


Figure 5.6 Fluorophores commonly used in fluorescence spectroscopy.

coefficients, good photostability and red-shifted excitation wavelength to limit the fluorescence background. Among the most popular derivatives, the fluorescein, rhodamine, cyanine, coumarine, Bodipy, Atto and Alexa families can be cited. Several of these families cover nearly all the UV-visible range. Proteins in solution are generally labelled covalently through a cysteine residue. To specifically label the cysteines, the labelling time should be carefully controlled to limit the reaction of the functionalised probe with other less-reactive groups of the protein such as the amine groups. Moreover, the labelling position should be carefully chosen, so that coupling to the external fluorophore will not affect the interaction with the binding partner. In cells, the most common strategy is based on the use of fluorescent proteins, such as the derivatives of the Green Fluorescent Protein (GFP), which was discovered in the jellyfish Aequorea Victoria.[40,41] By transfecting cells with a plasmid where the gene coding for the fluorescent protein is coupled to the gene coding for a protein of interest, fluorescently tagged chimer proteins will be expressed. Thus, a non-invasive, specific and stoichiometric tagging of proteins of interest can be obtained in live cells, tissues and animals. A large range of mutants of these fluorescent proteins has been designed, covering the full visible range, with good quantum yield and photostability. In addition, fluorescent proteins with specific properties such as photoswitching or photoconversion properties as well as sensitivity to given physico-chemical parameters of the environment have been engineered, giving access to a large palette of tools for monitoring interactions in cells.[42] Nevertheless, these fluorescent proteins present several pitfalls, such as their large size, which can affect or prevent the interaction of the fusion protein with its partner and the transfection protocols which can lead to strong overexpression of the fusion proteins, even in cell compartments where the native protein is not expressed, and thus give false positive.

Another interesting class is constituted by the quantum dots (QDs), that ;are extremely bright and photostable nanocrystals.[43–45] Through their semiconductor properties, QDs exhibit a large absorption spectrum but a narrow Gaussian emission at wavelengths controllable by the size of the material; the emission being shifted to the red when the QD size increases. The full visible range is covered by changing the size of the nanocrystal from 3 to 6 nm. QDs are usually composed of zinc sulfide, lead sulfide, cadmium selenide and indium phosphide. In order to limit their toxicity for biological applications and to make them soluble in water, QDs are coated with a protective polymer, which further increases their size. Due to their extreme brightness and photostability, QDs have major applications in single particle tracking over extended periods of time, but their use in live cells is still limited by toxicity problems. Moreover, while they can be easily used to label extracellular species, the labelling of intracellular biomolecules in live cells is still an issue. Another concern is that functionalised QDs frequently react with more than one molecular species in the labelling process, so that it is difficult to obtain labelled species with a 1:1 stoichiometry. Finally, the size of QDs, being much larger than that of conventional fluorophores, can strongly perturb the interaction processes.

5.3 Advantages and Pitfalls of Fluorescence-based Methods

One of the main reasons for the extreme popularity of fluorescence in its numerous applications is based on its exquisite sensitivity, which allows going down to the single molecule level. This extreme sensitivity is due in part to the recent progress in material science and electronics, which allows single photon counting with high signal-to-noise ratio. Moreover, in contrast to radioactive elements which emit a single radiation, fluorophores can emit several thousands of photons. Also, due to the Stokes shift, the wavelength range of emitted photons is distinct from that of the excitation photons. Thus, using appropriate filters or monochromators, the emitted photons can be recorded without contamination from the excitation photons. Another advantage of fluorescence is that it is conditioned by an external illumination, so that the switching on and off of fluorescence is controlled by the user. In addition, fluorescence techniques allow exploration of a very large dynamic range of concentrations (from pM to mM), since the intensity of the source, the monochromator slits and the sensitivity of the detectors can be adjusted as a function of the emission level of the sample. Furthermore, a large range of fluorophores is available commercially, covering the full range of wavelengths from UV to near IR, which allows selecting the appropriate fluorophore for a given application and performing multiplex analysis with spectrally different fluorophores. It should also be stressed that the large range of both fluorophores and dedicated instruments enable applications from the molecular level up to the live animal level. Another strong point is that fluorescence is a multi-parametric spectroscopy enabling monitoring of a process of interest through multiple channels: intensity at a single emission wavelength, position of the maximum emission wavelength, quantum yield, lifetime, anisotropy, rotational correlation time, diffusional correlation time *etc*. Finally, fluorescence shows also the key advantage to be a solution and homogeneous technique, which can be performed with diluted samples of small volumes in a large range of conditions (temperature, pH, salt concentrations, *etc*.) and does not require immobilising or separating the binding partners in a binding process.

While the sensitivity of fluorescence-based techniques is outstanding, their signal-to-noise ratio shows some limitations, especially at low concentrations of the fluorescent species, due mainly to the background fluorescence originating from buffers in solutions and the autofluorescence from naturally fluorescent components in cells. The fluorescence signal can also be contaminated by diffusion of the excitation light, especially when large particles are present. Diffusion is notably a strong concern in cells due to their numerous diffusing structures. Two types of diffusion can be observed: (i) Rayleigh diffusion that corresponds to elastic scattering of the photons at the excitation wavelength and (ii) Raman diffusion that is associated to inelastic scattering of the photons, yielding a red-shifted peak in respect to the Rayleigh peak, which can be easily identified by changing the excitation wavelength. Due to these different sources of contaminant photons, measurements with sub-nanomolar concentrations of

fluorescent species are difficult. Another major limitation of fluorescent-based techniques is the need to use external fluorophores in most applications. Due to their non-negligible size, fluorophores can affect the folding and activity of the labelled species. This is of course particularly critical for the largest fluorescent probes, such as the fluorescent proteins and QDs. An additional drawback, notably in living organisms, is related to the photodegradation or photobleaching of the fluorophores. Indeed, only a finite number of absorption/emission cycles can be performed with organic fluorophores before they reach a permanent photobleached dark state. Thus, for an ensemble of chromophores under continuous excitation, the number of fluorescent molecules will irreversibly decrease with time. The mechanism of photobleaching is frequently related to transitions from the excited singlet state to the excited triplet state, which is long lived and chemically more reactive. At room temperature, oxygen also plays an important role in photobleaching, by reacting in its singlet excited state, itself generated by reaction with the triplet state of the fluorophore. Each fluorophore has different photobleaching characteristics. For instance, for fluorescent proteins (like GFP), the number of absorption–emission cycles before photobleaching is around 10^5 [46] and can reach 10^8 or 10^9 for the best organic chomophores.[47] Chemical agents, such as oxygen scavengers, can reduce the photobleaching of organic chromophores. Photobleaching can be avoided by using QDs, which do not undergo irreversible transition through a dark state.

5.4 Determination of Binding Parameters from Fluorescence-based Techniques

Binding processes can be monitored through the fluorescence changes of either the protein or the ligand. The most favourable case is when the binding can be monitored through the intrinsic fluorescence of one of the binding partners. While only a limited number of ligands are naturally fluorescent, proteins are generally fluorescent, due to their intrinsic Trp residues which exhibit an emission highly sensitive to environment.[16] However, due to the aforementioned intrinsic spectroscopic limitations of Trp residues, the proteins or ligands are frequently labelled with extrinsic fluorophores. This requires, of course, to check that the latter do not affect the binding process, through for instance competitions with the unlabelled partner.

5.4.1 Binding Stoichiometry

Usually, the first parameter to be determined is the binding stoichiometry. Fluorescence techniques can be used for this purpose by titrating a high concentration of one partner by the other one (stoichiometric conditions). By reporting the changes of the measured fluorescence signal as a function of the ratio of the concentrations of the two binding partners, the binding stoichiometry can be inferred from the intersection of the initial slope of the titration

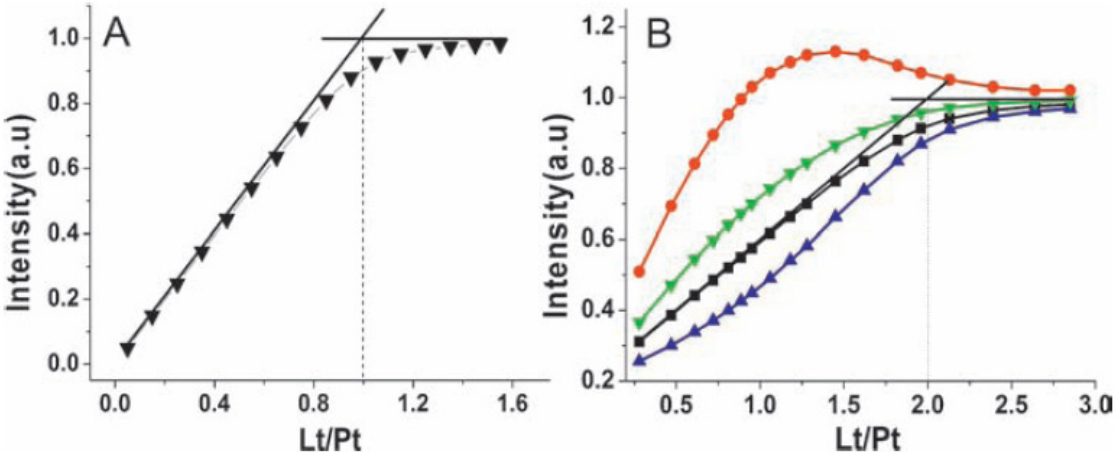


Figure 5.7 Determination of the binding stoichiometry for the interaction between a protein and a ligand, as monitored by fluorescence spectroscopy. (A) The binding stoichiometry is determined from the intercept of the initial slope with the plateau. Lt and Pt designate the total concentrations of ligand and protein, respectively. The intercept in this case suggests that the protein binds the ligand with a 1:1 stoichiometry. (B) Influence of the signal changes associated to individual sites on the determination of the binding stoichiometry. The binding curves were simulated with eqn (5.23) or a protein/ligand interaction with a 2:1 stoichiometry, a protein concentration of 0.5 μM and association binding constants for the individual sites of $3\times10^8\,M^{-1}$ and $5\times10^7\,M^{-1}$. The fluorescence intensities $I_P, I_{P^1L}, I_{P^2L}, I_{PL_2}$ associated to the individual binding states were, respectively, 0.2, 1.5, 0.6 and 1 (red circles), 0.2, 0.8, 0.3 and 1 (green inverted triangles), 0.2, 0.6, 0.6 and 1 (black squares) and 0.2, 0.3, 0.8 and 1 (blue triangles). An appropriate binding stoichiometry can only be obtained when the fluorescence changes associated to each binding site were equal and correspond to half of the total fluorescence change associated with full saturation (black squares).

with the tangent to the plateau. This approach is fully straightforward when the stoichiometry is 1:1 (Figure 5.7A). With higher stoichiometries, this approach is only valid if the changes in the fluorescence signal that accompany the binding to the different sites are identical (Figure 5.7B). Since the signal changes associated to individual sites are difficult to predict or to determine experimentally, the binding stoichiometry needs to be cross-checked by monitoring other fluorescence parameters or using different techniques, such as mass spectrometry. One of the most straightforward fluorescence-based approaches to determine the binding stoichiometry is time-resolved fluorescence anisotropy. This technique gives access to the rotational correlation time associated with the tumbling motion of the complex, which is directly proportional to its molecular mass. However, this requires that (i) the molecular weights of the free and bound labelled partners are significantly different, (ii) the shape of the complex is known or does not deviate much from a sphere and (iii) the molecular weight of the complex does not exceed 100–200 kDa when fluorophores with nanosecond lifetimes are used. FCS could also be used as a confident means to determine the binding stoichiometry, by monitoring the changes in the average number and the brightness of the fluorescent entities that move

through the focal volume. For instance, binding of two labelled ligands to a protein should decrease the average number and increase the average brightness of the diffusing species by a factor of two. The changes in diffusion times could also be used, but their dependence on the cubic root of the molecular mass of the labelled species enables only large changes to be perceived.

5.4.2 Binding Constants

The binding constants are usually recovered by titrating one partner at a concentration close to $1/k$ (where k is the expected equilibrium binding constant in M^{-1}) with increasing concentrations of the second partner. In the simplest case, there is a single binding site for the ligand. To illustrate this, let's assume that a protein/ligand interaction with a 1:1 stoichiometry and an equilibrium association constant k: $P + L \overset{k}{\Leftrightarrow} PL$ is monitored through the intrinsic fluorescence of the protein. The fluorescence intensity I of the protein can be expressed by:

$$I = \frac{[P]}{P_t} I_P + \frac{[PL]}{P_t} I_{PL} \tag{5.17}$$

where I_P and I_{PL} are the intensities associated to the free and bound forms of the proteins and P_t is the total concentration of protein.

Expressing the concentration of bound ligand [PL], as: [PL] = k[P][L] and substituting it in eqn (5.17) gives:

$$I = I_P + (I_{PL} - I_P)\frac{k[L]}{1 + k[L]} \tag{5.18}$$

The concentration of free ligand can be deduced from the mass conservation laws:

$$P_t = [PL] + [P] \Rightarrow [P] = P_t/(1 + k[L]) \tag{5.19a}$$

$$L_t = [PL] + [L] = [L](1 + k[P]) \tag{5.19b}$$

By substituting eqn (5.19a) into eqn (5.19b), we obtain a second-order equation:

$$k[L]^{2.} + (1 + k(P_t - L_t))[L] - L_t = 0 \tag{5.20}$$

which gives the concentration of free ligand:

$$[L] = \frac{-(1 + k(P_t - L_t)) + \sqrt{(1 + k(P_t - L_t))^2 + 4kL_t}}{2k} \tag{5.21}$$

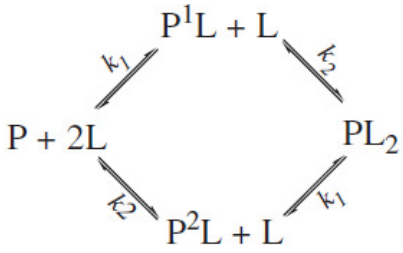
The final fitting equation is obtained by substituting eqn (5.21) into eqn (5.18). This equation is then used to fit the couples of values (L_{ti}, I_i) obtained for each addition of ligand, using a non-linear least-squares method. This fit allows recovering the k value, as well as the intensity of the saturated protein I_{PL}. The intensity of the free peptide I_P is usually measured experimentally in the absence of ligand and, thus, is fixed.

When the protein/ligand interaction is monitored by fluorescence anisotropy, the following equation applies:

$$r = \frac{r_0[P] + r_{PL}R[PL]}{[P] + R[PL]} \tag{5.22}$$

The final fitting equation is obtained by substituting eqn (5.21) into eqn (5.22), which allows recovering the k equilibrium constant, as well as the fluorescence anisotropy of the saturated protein r_{PL}. The anisotropy of the free peptide r_0 and the ratio R of the intensities of the free and bound forms are measured experimentally and thus fixed. Anisotropy titrations work best when no change in the protein quantum yield accompanies the binding of the ligand, since the free and bound forms will contribute equally in this case to the anisotropy.

In the case of multiple binding sites, the binding curves will depend not only on the binding constant associated to each binding site but also on the fluorescence signal associated with each bound state.[48,49] In general, only the fluorescence signal associated to the free and fully bound states could be measured experimentally. As a consequence, the fluorescence signals associated with intermediate binding states are treated as additional unknowns in the fitting procedure. Since the number of independent microscopic binding constants is given by 2^n–1 (n being the number of binding sites), straightforward determination of the binding constants is limited to systems with low binding stoichiometries. To illustrate a system with multiple binding sites, let's for instance consider a protein/ligand interaction, where two ligands bind independently to two non-interacting and distinct protein binding sites, with k_1 and k_2 equilibrium binding constants. This binding can be described by Scheme 5.1:



Scheme 5.1

where P^1L and P^2L are the intermediate complexes where the ligand is bound to the first or second binding sites, respectively, and PL_2 is the fully saturated

protein. By monitoring the binding through the changes in the protein fluorescence intensity, the fitting equation can be written as:

$$\begin{aligned} I &= \frac{[P]}{P_t} I_P + \frac{[P^1L]}{P_t} I_{P^1L} + \frac{[P^2L]}{P_t} I_{P^2L} + \frac{[PL_2]}{P_t} I_{PL_2} \\ &= \frac{(I_p + (k_1 I_{P^1L} + k_2 I_{P^2L})[L] + k_1 k_2 I_{PL_2}[L]^2)}{(1 + (k_1 + k_2)[L] + k_1 k_2 [L]^2)}. \end{aligned} \tag{5.23}$$

Using the conservation law for the total concentrations of protein and ligand, the concentration of free ligand can be obtained from the cubic equation:

$$\begin{aligned} k_1 k_2 [L]^3 + (k_1 + k_2 + 2k_1 k_2 P_t - k_1 k_2 L_t)[L]^2 \\ + (1 + (k_1 + k_2)(P_t - L_t))[L] - L_t = 0. \end{aligned} \tag{5.24}$$

By substituting the concentration of free ligand into eqn (5.23), the couples of (L_{ti}, I_i) values were then fitted with this equation, to provide the k_1 and k_2 values, as well as the values of I_{P^1L}, I_{P^2L} and I_{PL_2}, the fluorescence intensities associated to P^1L, P^2L and PL_2, respectively. As can be seen from this example, the complexity of the fitting equations and the number of unknowns grow rapidly with the number of binding sites. This generally leads to instable solutions with interdependent values and large error bars. Interestingly, to avoid developing the analytical equations needed for more complex binding models, numerical resolution software packages, such as Dynafit or Matlab, may be used. Nevertheless, this does not solve the problem of instability in the solutions when the number of unknowns is too high.

5.5 Applications: Example of the Binding of Zinc and Oligonucleotides to the Nucleocapsid Protein of HIV-1

To illustrate the application of fluorescence techniques to answer the who, where and why questions in protein/ligand interactions, we will use NCp7, the nucleocapsid protein of HIV-1 as an example. NCp7 is a small basic protein, characterised by a central globular domain formed of two highly folded zinc fingers (ZFs) flanked by basic residues (Figure 5.8A and B). In the HIV-1 viral particle, about 1500 molecules of NCp7 coat the two single stranded RNA genomes.[50] This protein plays key roles in both the early and late phases of the virus life cycle,[51–54] notably through its propensity to specifically bind the viral nucleic acids[55–57] and to chaperone their necessary transconformation during the processes of reverse transcription and virus assembly.[52] During reverse transcription, NCp7 chaperones viral DNA synthesis by reverse transcriptase, which necessitates two obligatory strand transfers.[58,59] The proviral DNA is then integrated by the viral integrase into the cellular genome, in a reaction also

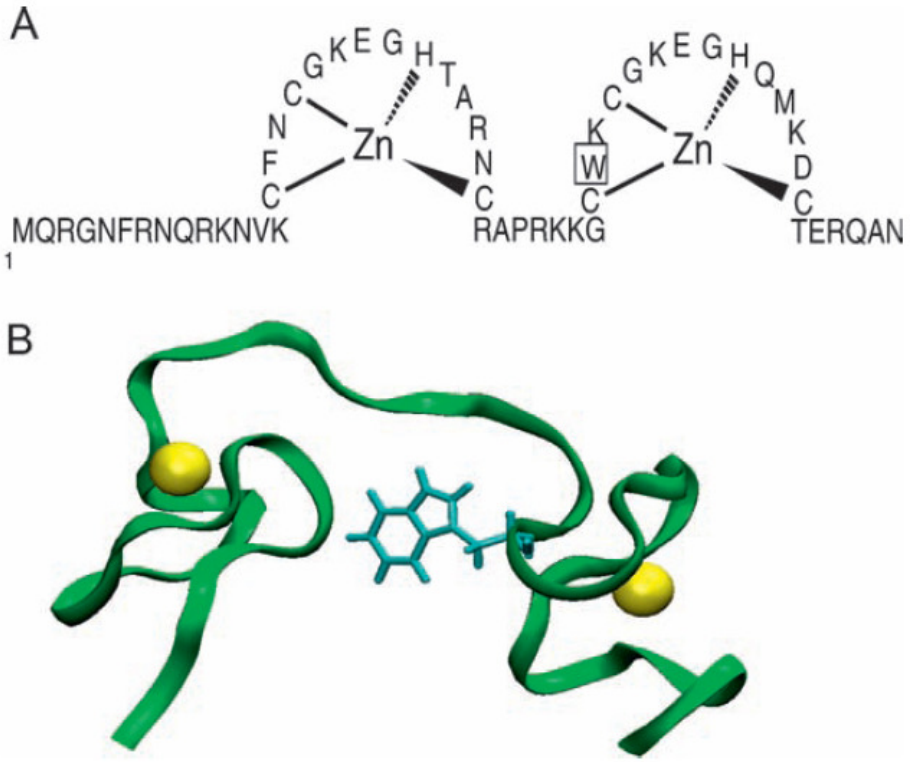


Figure 5.8 Sequence (A) and 3D-structure (B) of the HIV-1 nucleocapsid protein. In (B), the Trp residue is in blue and the zinc ions in yellow.

assisted by NCp7.[60] During the late phase of virus replication, NCp7 pilots genomic RNA selection and dimerisation and thus is considered to be an essential determinant of virus assembly. The importance of NCp7 is illustrated by the complete loss of infectivity caused by point mutations in the ZFs.[61–65] These mutations affect the proper folding of the ZFs[66–68] that play a key role in providing specific hydrophobic and H-bonding interactions with the RNA or DNA bases. Thus, NCp7 represents a major target for the development of new anti-HIV-1 agents that could impair both the early and late steps of HIV-1 replication and complement the "highly active anti-retroviral therapies" (HAART).[69,70]

5.5.1 Binding of Zinc to the Nucleocapsid Protein of HIV-1

NCp7 is characterised by a single Trp residue at position 37, in the distal ZF. The binding of zinc being critical for most NCp7 functions, the intrinsic fluorescence of Trp37 was first used to determine the binding parameters of NCp7 for zinc. As a first step, we monitored the binding of zinc to the Trp^{16}NC(13–30) and NC(34–51) peptides, which correspond to the sequences of the proximal and distal fingers, respectively.[71] In the proximal ZF, the weakly fluorescent Phe16 was substituted by a Trp37 residue. This substitution was not expected to affect the physicochemical properties of the peptide, since Phe or Trp residues were indifferently found at this position in various related

nucleocapsid proteins. The maximum emission wavelength of both Trp residues in the apo-peptides was close to 350 nm, indicating their full exposure to solvent (Figure 5.9A). Moreover, their quantum yields were low (0.06), in line with both peptides being unfolded, with an efficient dynamic quenching of the Trp fluorescence by the proximal Cys residues. Addition of zinc to Trp^{16}(13–30)NCp7 and NC(34–51) only marginally changed their emission maximum, but increased their quantum yield by 2.9- and 3.5-fold, respectively. This dramatic increase in the quantum yield is likely related to the zinc-induced folding of the peptides into a highly constrained structure,[72,73] where the Trp residue can no longer interact with the nearby quenching groups (Figure 5.9A). Using these large fluorescence changes and 1 mM EDTA to buffer very low concentrations of free zinc, binding constants of $5.7\times10^{13}\,M^{-1}$ and $1.7\times10^{13}\,M^{-1}$ were found respectively, for the proximal and distal ZF at pH 7.5 (Figure 5.9B).[71]

To determine the binding parameters of the full-length protein, the binding scheme in Figure 5.9C was used. At any zinc concentration, the measured fluorescence intensity I is given by:

$$I = I_0\frac{[P]}{[P_t]} + I_1\frac{[PZn^1]}{[P_t]} + I_2\frac{[PZn^2]}{[P_t]} + I_3\frac{[PZn_2]}{[P_t]} \tag{5.25}$$

Since I_0 and I_3, corresponding to the fluorescence of the apopeptide and the holopeptide, respectively, were easily measurable, eqn (5.25) can be rearranged into:

$$\Delta I = \frac{\Delta I_i k_i[Zn] + k_1k_3[Zn]^2}{1 + (k_1 + k_2)[Zn] + k_1k_3[Zn]^2} \tag{5.26}$$

where $\Delta I_i = (I_i - I_0)/(I_3 - I_0)$, with $i = 1$ for Trp^{16}-containing peptides and $i = 2$ for Trp^{37}-containing peptides. The concentration of free zinc [Zn] buffered using $[E_t] = 1$ mM EDTA was calculated by : $[Zn] = [Znt]/[K_E([E_t] - [Zn_t])]$, where $[Zn_t]$ is the total concentration of zinc, and K_E is the affinity of EDTA for zinc at pH 7.5. The binding constant values of the proximal and distal ZFs in the native protein were close to those of the isolated ZFs, indicating that the fingers possess most of the information to strongly bind zinc. Moreover, a slight negative cooperativity, likely related to the spatial proximity of the zinc-bound fingers, was observed.[67,74]

Further thermodynamic parameters can be recovered from the temperature dependence of the binding parameters. Plotting the experimental zinc binding constants for the two isolated ZFs *versus* the reciprocal of the temperature in a van 't Hoff plot gives a straight line that allows recovering the enthalpy changes ΔH^0 (Figure 5.9D). Moreover, since the binding constants k_i give access to the free energy changes $\Delta G^0 = -RT\ \ln k_i$, the entropy changes ΔS^0 can also be deduced by: $\Delta S^0 = (\Delta H^0 - \Delta G^0)/T$. The enthalpy changes $\Delta H^0 = 35$ kJ/mol were identical in both peptides while the entropy change ΔS^0 associated to the proximal finger (148 J/K/mol) was larger than those associated to the distal

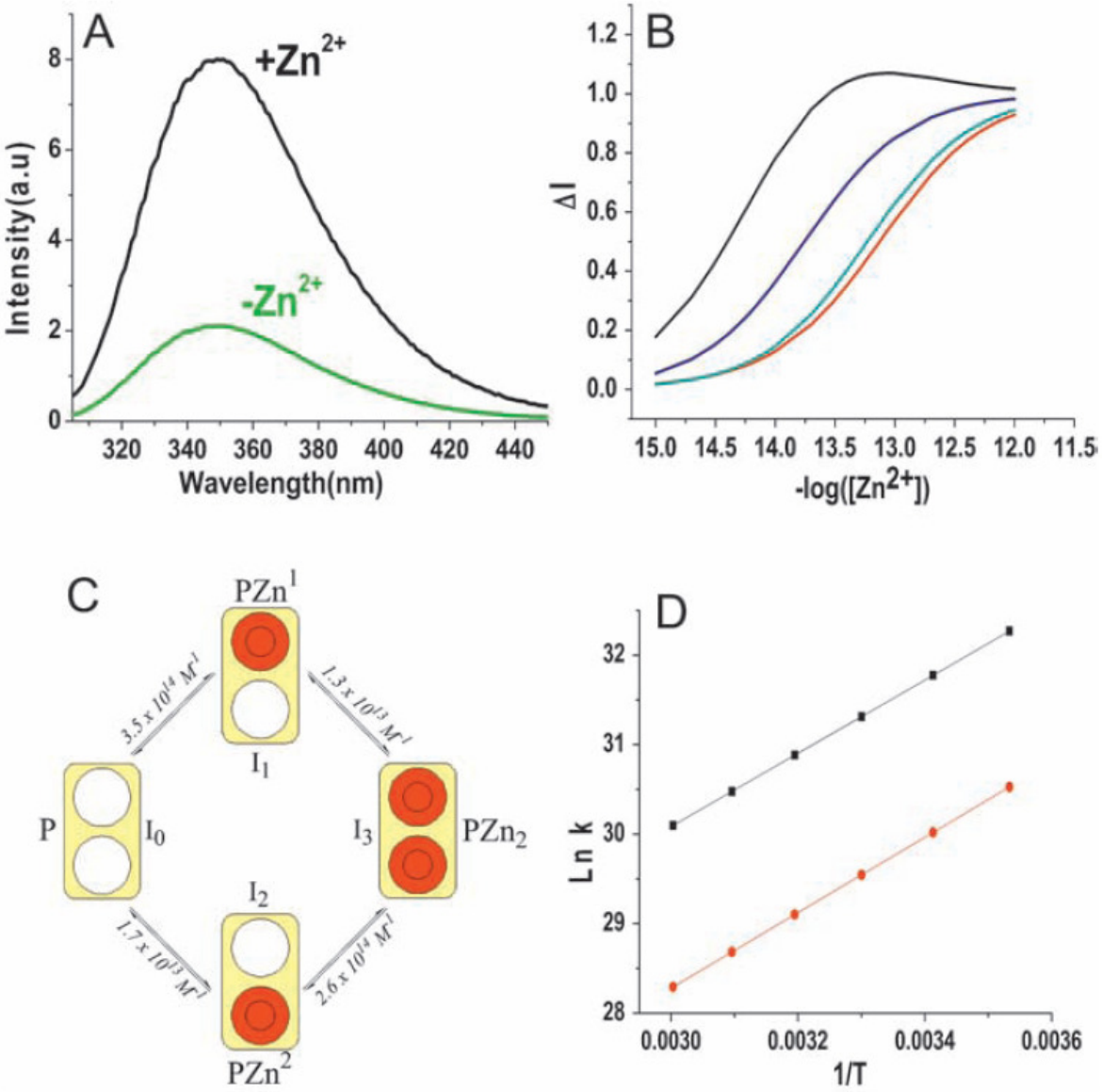


Figure 5.9 Zinc binding to the HIV-1 nucleocapsid protein NCp7. (A) Emission spectra of 1 μM NCp7, in the absence (green) and in the presence (black) of zinc. The spectrum of the apoprotein was recorded in the presence of 1 mM EDTA. Excitation wavelength was 295 nm. (B) Zinc binding isotherms of NCp7 derivatives. Titrations were performed with 10 μM (1–55) NCp7 (red), $Trp^{16}Phe^{37}Phe^{61}$(13–64) NCp7 (black), Trp16(13–30) NCp7 (blue) and (34–51) NCp7 (green) in 50 mM HEPES, 0.1 M KCl, 1 mM EDTA, pH 7.5. The solid lines were obtained by using eqn (5.26) and the binding constants given in C and in the text. (C) Zinc binding scheme and equilibrium binding constants of NCp7. Zinc-free and zinc-saturated finger motifs are represented by open and closed circles, respectively. (D) van 't Hoff plots for the binding of zinc to the isolated zinc finger motifs of NCp7. The temperature dependence of the equilibrium binding constants of Zn^{2+} to the Trp^{13}(13–30) NCp7 (black squares) and (34–51) NCp7 (red disks) peptides were recorded in the temperature range 10 °C to 60 °C. The solid lines correspond to the fits of the experimental data using $k = -\frac{\Delta H^\circ}{R} \times \frac{1}{T} + \frac{\Delta S^\circ}{R}$ with the ΔH^0 and ΔS^0 values given in the text.

finger (130 J/K/mol). The entropy changes contributed more than half to the ΔG^0 values in both peptides, despite the transition of the ZF from a random-coil to a highly constrained structure. The unfavourable structural entropy associated to this transition is thus likely compensated by the very strong entropic contribution due to loss of hydration water from both the zinc ion and the highly hydrophilic peptides during complexation. Though useful and easy to perform, this approach through the temperature dependence of the binding parameters exhibits some limitations, since it assumes negligible changes in the heat capacity. Moreover, in contrast to NCp7 which stably binds zinc ions and remains folded even at rather high temperatures,[75] many proteins are unfolded by the temperature, making the analysis irrelevant.

Fluorescence techniques can also be used to determine kinetic rate constants, using a stopped-flow instrument. The sensitivity of the fluorescence detection is very useful to limit the amount of material needed by the stopped-flow technique. Using Trp37 fluorescence, the dynamics of Zn^{2+} association and dissociation to the NC(35–50) peptide, corresponding to the distal ZF, was examined under anaerobic conditions to avoid oxidation of the Cys residues.[76] Using pseudo-first-order conditions, individual time courses could be fitted to a monoexponential function, providing one observed rate constant k_{obs} (Figure 5.10A). Plotting the k_{obs} values as a function of the Zn^{2+} concentration at a given pH value revealed a saturation behaviour (Figure 5.10B), consistent with a two-step reaction scheme, where a fast pre-equilibrium binding step is followed by a rate-limiting inter-conversion step:

$$L + Zn^{2+} \overset{K_M^*}{\rightleftharpoons} LZn^{2+*} \underset{k_b}{\overset{k_f}{\rightleftharpoons}} LZn^{2+}$$

Accordingly, the Zn^{2+} concentration dependence of k_{obs} was fitted to:

$$k_{obs} = \frac{k_f K_M^* [Zn^{2+}]}{1 + K_M^* [Zn^{2+}]} + k_b \qquad (5.27)$$

in order to recover the values for K_M^*, the association equilibrium constant of the LZn^{2+*} intermediate and the forward (k_f) and backward (k_b) interconversion rate constants. pK_a values of 6.4, 8.0, 8.8 and 9.3 were independently determined by ^{1}HNMR for the four Zn^{2+}-coordinating residues, His44, Cys36, Cys39 and Cys49, respectively.[77] Accordingly, the deprotonation is almost sequential and likely induces a sequential binding to the four coordinating residues. To further investigate this point, the kinetics of Zn^{2+} binding to the peptide was investigated at various pH values. Moreover, the backward interconversion rate constants k_b were determined from dilution-induced dissociation experiments. Additional information was obtained from the pH dependence of the fluorescence of the intermediates. From all these data, the whole set of kinetic and equilibrium constants characterising the complete reaction scheme (Figure 5.11A) could be determined. In this reaction scheme, there are five protonation states of the apopeptide, but the fully protonated

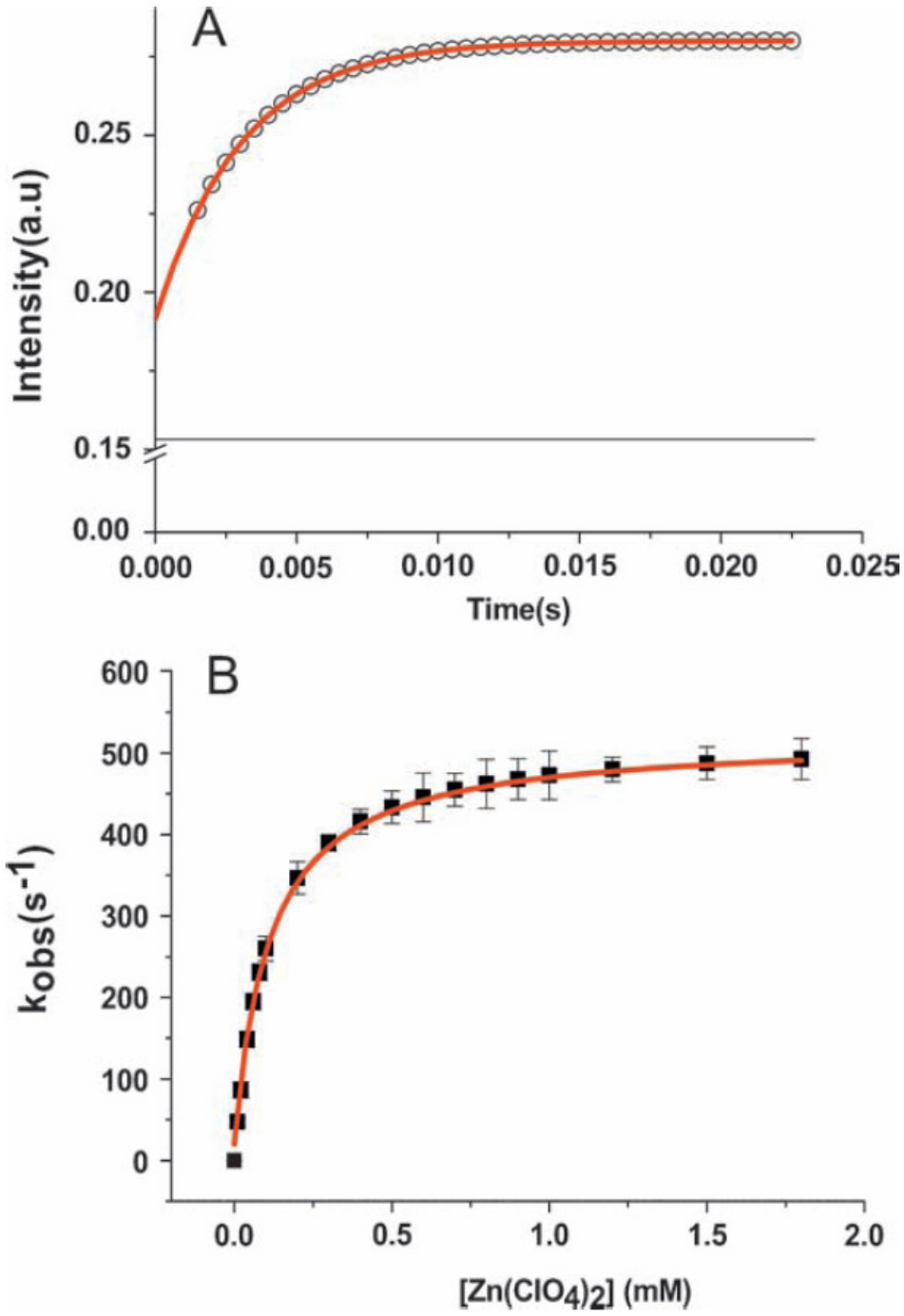


Figure 5.10 Zn^{2+} binding kinetics to the distal zinc finger motif of NCp7 at pH 6.8, as monitored through the intrinsic fluorescence of the peptide. The (35–50) NCp7 peptide was 1.5 μM in 50 mM HEPES, pH 6.8 at a total ionic strength of 0.15 M. (A) Representative kinetic experiment. The peptide was reacted with 0.35 mM $Zn(ClO_4)_2$ at 20 °C. Excitation wavelength was 280 nm. The solid line is the best fit to a monoexponential function with $k_{obs} = 330\ s^{-1}$. The black solid line is the fluorescence of the apopeptide. (B) Dependence of the observed kinetic rate constant, k_{obs} on the total Zn^{2+} concentration. The solid line is the fitted curve using eqn (5.27) with $k_f = 500\ s^{-1}$ and $K_M^* = 9000\ M^{-1}$ (the k_b value was too small to be evaluated).

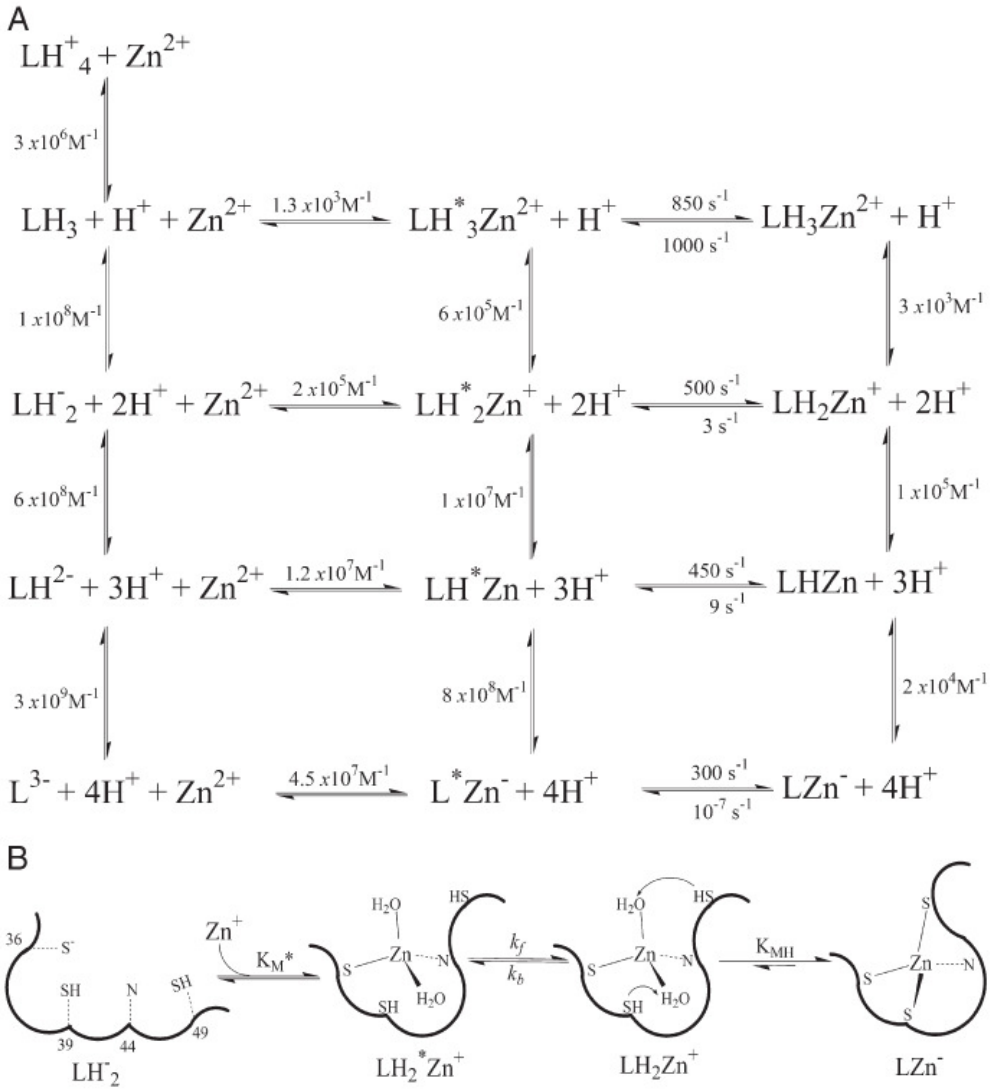


Figure 5.11 Mechanism of Zn^{2+} binding to the distal finger of NCp7. (A) Reaction scheme for Zn^{2+} binding to (35–50) NCp7. The various kinetics and equilibrium constants have been determined from stopped-flow measurements, using fluorescence detection, with the exception of the four protonation constants of the apopeptide, which have been obtained by ^{1}HNMR. (B) Proposed mechanism of Zn^{2+} coordination and induced folding of (35–50) NCp7 in the neutral pH range. Only the species most relevant for Zn^{2+} binding are shown. The interconversion between $LH^*_2Zn^+$ and LH_2Zn^+ is assumed to correspond to a rearrangement of the peptide backbone, enabling the formation of an adequately folded form that rapidly leads to the final LZn^- complex.

LH^+_4 species is assumed not to coordinate Zn^{2+}. Binding of Zn^{2+} to the apopeptide forms is thought to rapidly lead to the $LH^*_3Zn^{2+}$ to L^*Zn^- intermediates, which are then converted into the four states of the final Zn^{2+} complex, differing in coordination number: LH_3Zn^{2+} to LZn^-. From this scheme, the major pathway for the Zn^{2+}-driven folding of NC(35–50) at neutral pH (Figure 5.11B) is thought to start mainly with the bidentate LH^-_2

species where both His44 and Cys36 residues are deprotonated. Binding of both residues to Zn^{2+} gives the $LH_2^*Zn^+$ intermediate, which is likely converted into the appropriately folded LH_2Zn^+ species through a rearrangement of the peptide backbone around Zn^{2+} to optimise the binding geometry. Then, the final LZn^- complex is rapidly generated through deprotonation of Cys39 and Cys49 residues, which are likely close to Zn^{2+} in LH_2Zn^+, so that they can readily substitute the remaining solvate molecules.

5.5.2 Binding of Oligonucleotides to the HIV-1 Nucleocapsid Protein

In addition to the binding of Zn^{2+}, the intrinsic fluorescence of NCp7 can also be used to monitor the binding of oligonucleotides (ODNs). For instance, we investigated the binding parameters of NCp7 to the d(AACGCC) sequence, whose structure in complex with NCp7 has been solved by ^{1}H NMR.[78] Titration of NC(12–53), a peptide corresponding to the central finger domain, by d(AACGCC) in the absence of salt (Figure 5.12A inset) showed a very strong decrease (up to 90%) in the peptide fluorescence as well as the presence of a single protein-binding site on d(AACGCC), as revealed from the intersection of the initial slope with the plateau.[79] To get more information,

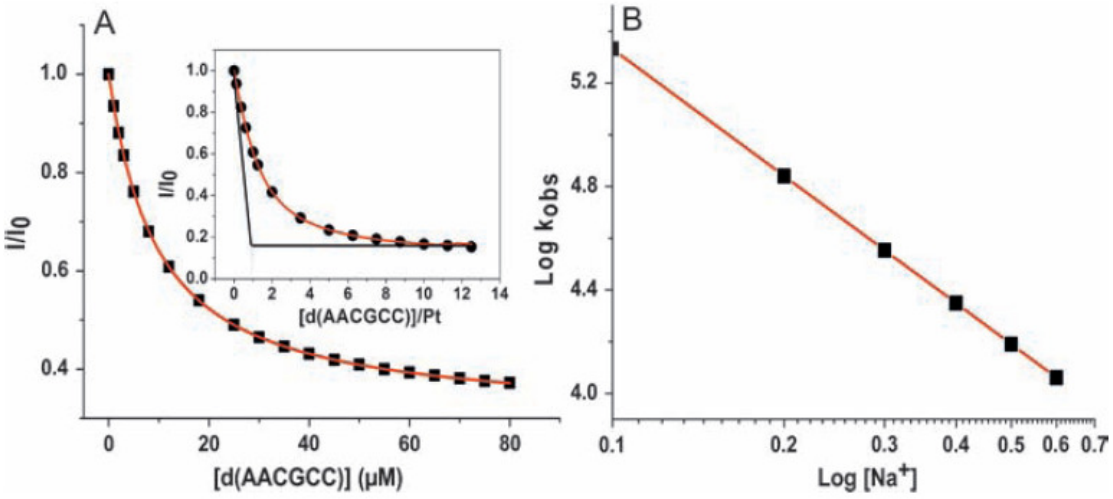


Figure 5.12 Binding of d(AACGGCC) to NC(12–53), as monitored through the intrinsic fluorescence of the peptide. (A) Titration of NC(12–53) with d(AACGCC). Peptide concentration was 0.8 μM in 50 mM HEPES, pH 7.5, in the absence (inset, circles) or in the presence (squares) of 100 mM NaCl. Solid lines correspond to the fit of the experimental points with eqn (5.18) and a binding constant of $1.7\times10^6\,M^{-1}$ and $1.1\times10^5\,M^{-1}$, in the absence and the presence of 100 mM NaCl, respectively. In the inset, the intercept of the initial slope with the plateau gives the binding stoichiometry. (B) Salt-back titration of NC(12–53) with Na^+ in the 100–600 mM concentration range. The solid line is a fit with eqn (5.28) and the m′ and K(1M) values are given in the text.

a time-resolved fluorescence investigation was performed. In the absence of ODN, the peptide decay was characterised by three lifetimes (0.64 ns, 3.4 ns and 7.3 ns). In the presence of a saturating d(AACGCC) concentration, the three lifetimes were not or only poorly affected, but their relative proportions strongly decreased to the benefit of an ultra-short lifetime (60 ps), which suggested a stacking of Trp37 with d(AACGCC). Using the fluorescence changes of NC(12–53) on titration with d(AACGCC), a binding constant of $1.1\times10^5\,M^{-1}$ was obtained in 50 mM HEPES, 0.1 M NaCl, pH 7.5 (Figure 5.12A).

Interestingly, substitution of NC(12–53) by the NC(1–72) precursor (which contains 17 additional C-terminal amino acids with respect to the mature NCp7) only led to a 2.5-fold increase in the binding constant, suggesting that the N- and C-terminal domains flanking the finger domain of NCp7 were not critical for binding to d(AACGCC), in line with the perpendicular orientation of the ODN in respect with the ZF domain.[78] In contrast, removal of zinc by EDTA decreased the binding constant by about two orders of magnitude, indicating that zinc in NCp7 was critical for the binding to d(AACGCC). To identify the key nucleotides in the peptide/d(AACGCC) complex, each nucleotide in interaction with the peptide was systematically substituted. Since G4 was shown to stack with Trp37 in the complex with d(AACGCC),[78] G4 was changed to either A, C or T. In each case, a dramatic decrease in the binding constant was observed, indicating that the stacking interaction contributed to about 20% of the total binding energy in the complex. In contrast, the final fluorescence at the plateau was only poorly affected by G4 substitution, suggesting that Trp37 also stacks with A, C, or T bases but probably with a lower stability. Interestingly, while the binding constant of the peptide to $d(C_6)$ was very low, it was only two-fold reduced for $d(GC_5)$ as compared to d(AACGCC), suggesting that the stacking of Trp37 with a G drives the positioning of NCp7, irrespective of the position of G in the ODN. From the additional substitutions, we finally concluded that NCp7 binds to single-stranded ODNs with the following preference: $XiTGXj > XiGXGXj \approx XiTXGXj > XiGXj \gg XiXj$, where X corresponds to either A or C.

Fluorescence techniques are also well suited to determine the electrostatic and non-electrostatic contributions to the binding energy. These contributions can be obtained from the dependence of the binding constants on the salt concentration[80] by calculating the number m′ of ion pairs between the protein and the ligand and the non-electrostatic binding constant K(1M), respectively. Since no anion uptake or release was found to accompany NCp7 binding to various tRNAs,[81] the dependence of the observed binding constants as a function of the NaCl concentration was expressed as:

$$\log K_{obs} = \log K(1M) - \psi_{Na^+} m' \log[Na^+] \qquad (5.28)$$

where Ψ_{Na+}, the fraction of Na^+ thermodynamically bound per phosphate group, is assumed to be 0.71. Since the plots of log K_{obs} *vs.* log[NaCl]

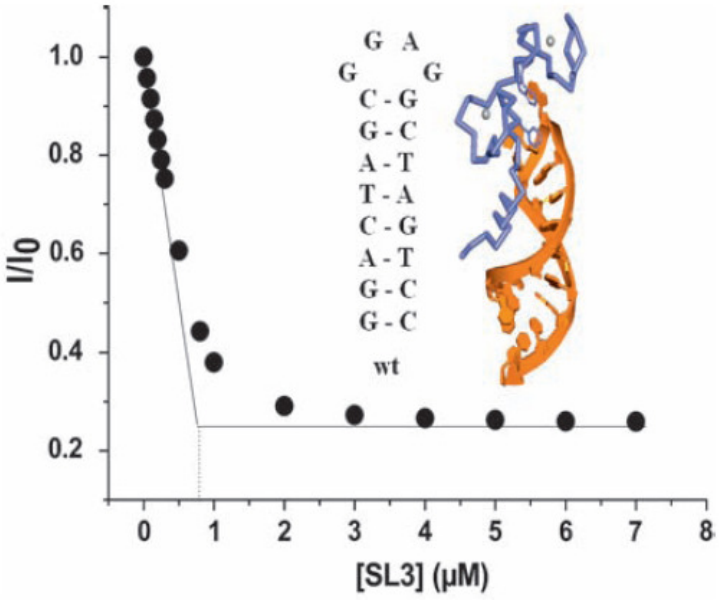


Figure 5.13 Binding curves of the SL3 stem-loop to NCp7. NCp7 concentration was 0.8 μM in 50 mM HEPES, 0.1 M NaCl, pH 7.5. The intercept of the initial slope with the plateau gives the number of protein binding sites on SL3. In the NMR-derived structure of the SL3/NCp7 complex,[82] the protein is in blue and SL3 in brown and the Trp37–G^7 stacking interaction is shown.

were linear in the 100–600 mM salt concentration range (Figure 5.12B), $m' = 2.3(\pm 0.2)$ ion pairs were determined for the NC(12–53)/d(AACGCC) complex, in good agreement with the three electrostatic interactions deduced from the NMR structure.[78] Moreover, in line with the numerous H bonds and hydrophobic interactions of the (12–53)NCp7–d(ACGCC) complex, the non-electrostatic interactions as deduced from the K(1M) value ($5\times10^3\ M^{-1}$) represented about 70% of the binding energy at 0.1 M NaCl.

We also investigated the interaction of NCp7 with the DNA analogue of the SL3 stem-loop of the ψ-RNA recognition element (Figure 5.13). In line with the ^{1}H-NMR structure of the NCp7/SL3 complex showing a single binding site on the SL3 loop,[82] we found that at 0.1 M salt, NCp7 binds to SL3 with a 1:1 stoichiometry and a rather high binding constant ($1.4\times10^7\ M^{-1}$). In contrast to d(AACGCC), both N- and C-terminal domains of NCp7 significantly stabilised its complex with SL3. Using various mutants of SL3, Phe16–G^9 and Trp37–G^7 interactions were found to contribute equally to NCp7–SL3 stability.

The intrinsic fluorescence of Trp37 was also used to characterise the binding of NCp7 to cTAR, a 55nt DNA sequence corresponding to the cDNA copy of the transactivation response element.[83] In a 25 mM Tris (pH 7.5), 30 mM NaCl, 0.2 mM $MgCl_2$ buffer, the binding stoichiometry was found to be about $n = 8$ (Figure 5.14A inset).[83] Assuming that the binding sites are identical and

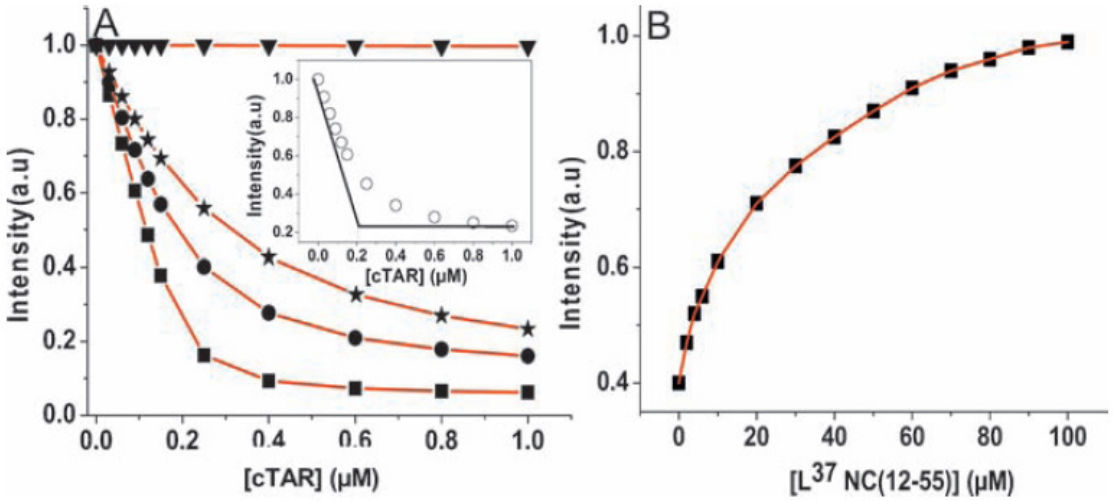


Figure 5.14 Binding curves of NCp7 derivatives to cTAR. (A) Binding curves monitored from the intrinsic fluorescence of the NCp7 derivatives. The peptide concentration was 0.2 μM to 1.5 μM in 25 mM Tris (pH 7.5), 30 mM NaCl, 0.2 mM $MgCl_2$. Excitation and emission wavelengths were 295 nm and 350 nm, respectively. Experimental points with NC(1–55) (squares), NC(12–55) (circles), $W^{16}F^{37}$NC(12–55) (stars) and NC(35–50) (triangles) were fitted with eqn (5.29) and the parameters of Table 5.1. Inset: Determination of the binding stoichiometry for NC(12–55). The intercept of the initial slope with the plateau gives n = 8 NC(12–55) molecules bound to cTAR. (B) Determination of the binding constant of L^{37}NC(12–55) for cTAR, by competition with NC(12–55). The non-fluorescent L^{37}NC(12–55) peptide was added at increasing concentrations to dissociate a cTAR/NC(12–55) complex. The binding constant of L^{37}NC(12–55) given in Table 5.1 was derived from eqn (5.30).

independent, the native NCp7 was found to bind with an affinity $K_{app} = 10^8$ M^{-1} (Figure 5.14A and Table 5.1), using:

$$I = I_0 - \frac{(I_0 - I_t)}{L_t} \times \frac{(1 + (L_t + nN_t)Kapp) - \sqrt{(1 + (L_t + nN_t)K_{app})^2 - 4L_t nN_t K_{app}^2}}{2K_{app}} \quad (5.29)$$

where I_t describes the fluorescence at the plateau when all the peptide is bound, whereas I_0 and I correspond to the fluorescence intensities of the peptide in the absence and in the presence of a given concentration of cTAR. Moreover, L_t and N_t designate the total concentration of peptide and cTAR, respectively. This approach is only semi-quantitative, since NCp7 binding depends on the ODN sequence and strandedness[79,84] and may be cooperative.[81,85] Nevertheless, this approach is adequate for comparing the binding affinities of cTAR to various NCp7 mutants and thus for identifying the structural determinants for binding (Table 5.1). A 5-fold decrease in affinity was observed when NCp7 was substituted by NC(12–55), suggesting that the N-terminal

Table 5.1 Binding parameters of NC mutants to cTAR[a].

Peptides	K_{app} $(M^{-1})^a$	n^a
NC(1–55)	$1\ (\pm 0.2)\times 10^8$	7.8 ± 0.4
NC(12–55)	$1.7\ (\pm 0.2)\times 10^7$	7.6 ± 0.6
C^{44}NC(1–55)	$1.5\ (\pm 0.3)\times 10^8$	8.0 ± 0.6
C^{23}NC(12–55)	$3.9\ (\pm 0.5)\times 10^5$	8^b
NC(35–50)	$1.5\ (\pm 0.2)\times 10^4$	18^b
S_3NC(13–64)	$2.6\ (\pm 0.5)\times 10^6$	6.0 ± 0.6
NC(13–64)	$4.0\ (\pm 0.8)\times 10^7$	6.0 ± 0.8
L^{37}NC(12–55)	$3.4\ (\pm 1.0)\times 10^{5c}$	8
$W^{16}F^{37}$NC(12–55)	$6\ (\pm 2)\times 10^6$	7.4 ± 0.6

Binding experiments were performed in 25 mM Tris (pH 7.5), 30 mM NaCl, 0.2 mM $MgCl_2$.
[a]The number, n, of binding sites was determined with 1.5–2 mM peptide. The K_{app} values were obtained with 0.2 to 1.5 mM peptide by fitting the data in Figure 5.14A to eqn (5.29) and fixing the number of binding sites to its closest integer. The n and K_{app} values are expressed as means and standard errors of the mean for at least two experiments.
[b]Due to the low binding constants of C^{23}NC(12–55), NC(35–50) and L^{37}NC(12–55), the n values were fixed for these mutants.
[c]The binding constant of L37NC(1–55) was determined by fitting the data of Figure 5.14B with eqn (5.30).

(1–11) basic domain stabilises the binding of NCp7 to cTAR. The similarity in the K_{app} values of C^{44}NC(1–55) and NC(1–55) and, in contrast, the two orders of magnitude decrease in K_{app} for C^{23}NC(12–55) with respect to NC(12–55) indicate that the native folding of the proximal ZF, but not the distal ZF, is critical for tight binding to cTAR. Large decreases in K_{app} values are also observed with the isolated ZFs, indicating that both ZFs are required for high-affinity binding to cTAR. In addition to the fingers, the central basic residues ^{32}RKK^{34} of the linker contribute to the binding process, since their substitution by three Ser residues in the NC(13–64) peptide decreases K_{app} by one order of magnitude. Finally, the K_{app} value of the L^{37}NC(12–53) peptide, which does not contain any Trp residue, was determined through competition, by monitoring the dissociation of a cTAR/NC(12–55) complex in the presence of increasing concentrations of L^{37}NC(12–55) (Figure 5.14B). The competition curve was fitted with:

$$I = \left(L_t + \frac{1}{2}\left(\frac{I_t}{I_0} - 1\right) \left(L_t + nN_t + \frac{1}{K_{12-55}} + A_t \frac{K_{L37}}{K_{12-55}} - \sqrt{\left(L_t + nN_t + \frac{1}{K_{12-55}} + A_t \frac{K_{L37}}{K_{12-55}} \right)^2 - 4nL_tN_t} \right) \right) \frac{I_0}{L_t} \tag{5.30}$$

where A_t designates the total concentration of L^{37}NC(12–55), while K_{L37} and K_{12-55} designate the binding constants of the L^{37}NC(12–55) and NC(12–55) peptides, respectively. The fit yielded a K_{app} value as low as that of C^{23}NC(12–55), showing that Trp37 plays a critical role in cTAR binding. This example also illustrates that competition experiments constitute an interesting approach when a

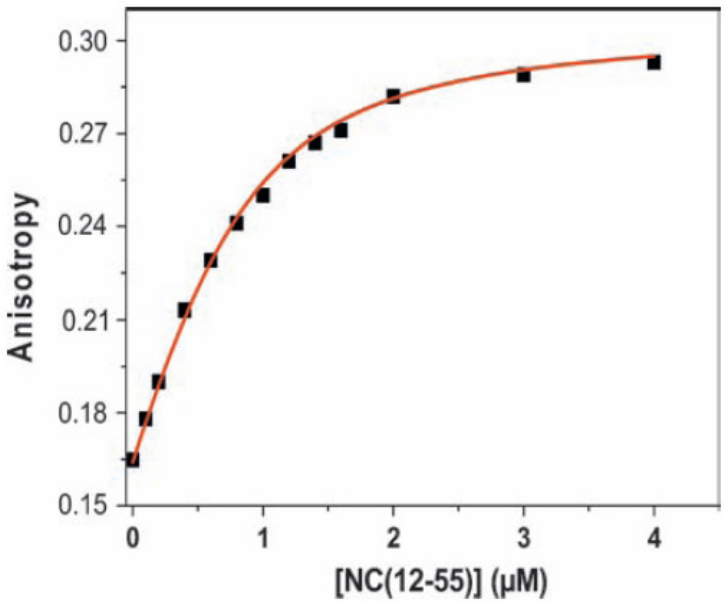


Figure 5.15 Titration of cTAR with NC(12–55), as monitored by fluorescence anisotropy. The concentration of Rh6G-5′-cTAR was 100 nM. The titration was performed by adding increasing concentrations of NC(12–55) (squares) to the labelled cTAR, in 25 mM Tris (pH 7.5), 30 mM NaCl, 0.2 mM $MgCl_2$. Excitation was 480 nm. The emission (>550 nm) was monitored through a high-pass filter. Solid line corresponds to the fit of the experimental points with eqn (5.31) and the parameters given in the text.

non-fluorescent partner has to be characterised. In contrast, only a 3-fold decrease in the K_{app} value accompanies the permutation of the aromatic residues at positions 16 and 37, indicating that in contrast to the non-aromatic Leu residue, a Phe residue can partly substitute for Trp in cTAR binding.

As an alternative to the intrinsic protein fluorescence, protein/ODN interaction can be monitored by fluorescence anisotropy using fluorescently labelled ODNs. In this case, a fixed concentration of labelled ODN is titrated by increasing concentrations of protein. Such a titration was performed by adding NC(11–55) to cTAR labelled at its 5′ end by 6-carboxyrhodamine (Rh6G) (Figure 5.15). The dependence of the anisotropy r of the ODN as a function of the peptide concentration was fitted to:

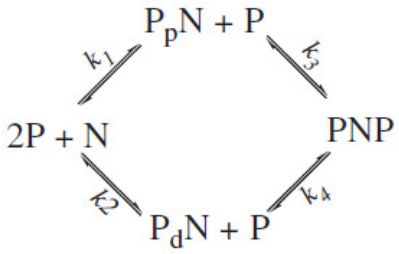
$$r = r_0 + \frac{(r_f - r_0)}{n}\left(\frac{\left(1 + K_{app}(nN_t + L_t)\right) - \sqrt{\left(1 + K_{app}(nN_t + L_t)\right)^2 - 4nK_{app}^2 N_t L_t}}{2K_{app}N_t}\right) \tag{5.31}$$

where L_t and N_t designate the total concentration of peptide and 5′-Rh6G-cTAR, respectively. r_f represents the anisotropy at the plateau when all the peptide is bound, whereas r_0 and r correspond to the anisotropy values of 5′-Rh6G-cTAR in the absence and in the presence of a given concentration of

peptide, respectively. Assuming n = 8 protein binding sites on cTAR, a K_{app} value of $2.2\times10^{7}\,M^{-1}$ was determined, in excellent agreement with the $1.7\times10^{7}\,M^{-1}$ value determined by titrating the peptide with non-labelled cTAR (Table 5.1).

As can be seen from this example, the analysis of NC binding to ODNs with multiple binding sites is hampered by the intrinsic inability of most techniques that monitor the binding in a global manner only, to discriminate between the different binding sites. To circumvent this limitation, ODNs labelled with site-specific reporters such as 2-aminopurine (2Ap), an environmentally sensitive fluorescent analogue of adenine, can be used. This fluorescent probe is highly quenched, due to collisions with the neighbour bases. In contrast, binding of NCp7 restricted the ODN flexibility on the picosecond to nanosecond time scale, impeding the collisions of 2Ap with its immediate neighbours, which in turn results in a strong fluorescence increase.[66] To characterise the potential use of 2Ap to site-selectively characterise the binding of NCp7 to an ODN with multiple binding sites, we used three 2Ap-labeled dodecanucleotides that constitute the simplest presumably "multi-valent" ODNs (Figure 5.16). To limit the contributions of overlapping sites in the binding process, all ODNs contained two TG motifs that constitute strong preferential binding sites for NCp7. The three ODNs differed by the position and proximity of the TG motifs within their sequence and were non-self-complementary and unfolded. Through competition with non-labelled sequences, substitution of A with 2Ap was checked to not significantly affect NCp7 binding. Moreover, NCp7 was found to increase substantially the quantum yield of 2Ap at all labelled positions.

In order to determine the stoichiometry of the NCp7/ODN complexes, time-resolved anisotropy decay experiments were performed. The rotational correlation time associated with the tumbling of the complexes ranged from 7.5 to 10.7 ns. These values are consistent with the theoretical correlation time of 7.5 ns for a 2:1 NCp7/ODN complex, suggesting that all tested ODNs bind two NC molecules, as expected from the two high-affinity TG motifs in their sequences. To model the interaction of NCp7 with the labelled ODNs, a binding scheme with two individual binding sites was used (Scheme 5.2):



Scheme 5.2

where P is the free protein, N is the free ODN, P_pN and NP_d are the 1:1 NCp7/ODN complexes with the protein bound to the proximal (5′-terminal) and

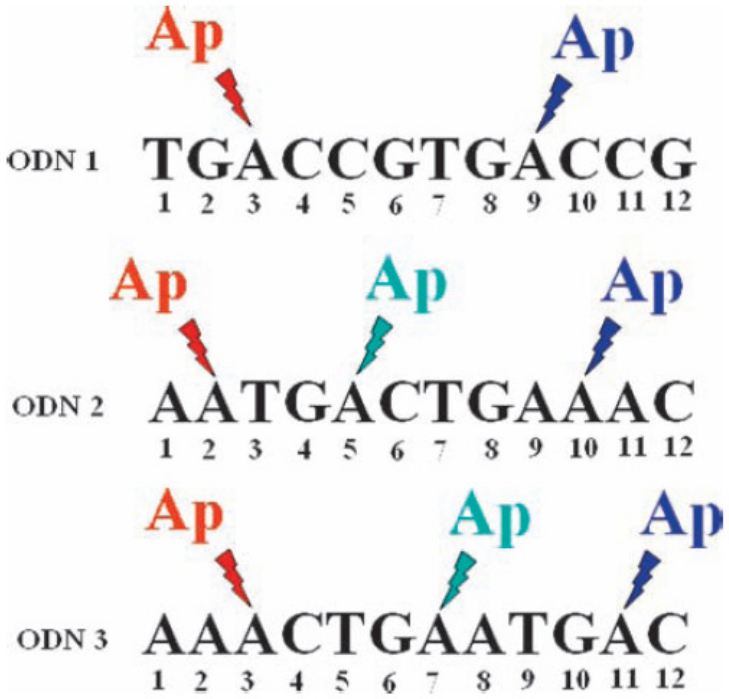


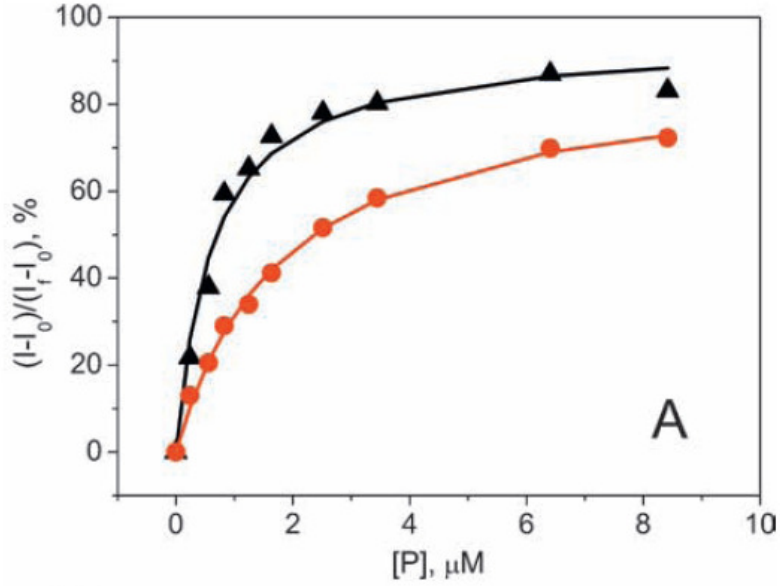
Figure 5.16 Sequences of dodecanucleotides used to site-selectively monitor the binding of NCp7. The ODNs were labelled by 2-aminopurine (2Ap) at the indicated positions.

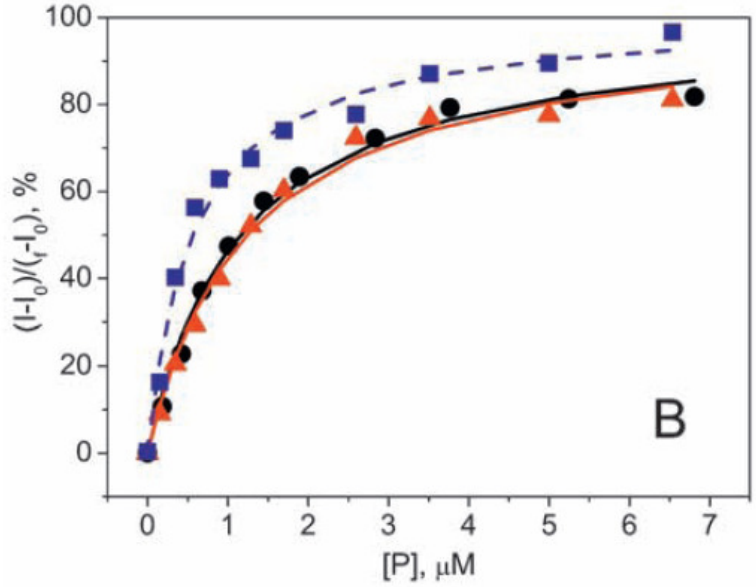
distal (3′-terminal) binding sites, respectively (termed "p" and "d" sites, respectively), PNP is the 2:1 NCp7/ODN complex, and k_1–k_4 are the microscopic equilibrium binding constants. Due to its sensitivity to changes in its immediate vicinity, 2Ap is expected to report on NCp7 binding only to the site in which it is included. As a consequence, for an ODN labelled at binding site p, the concentration of bound protein at this site can be calculated from:

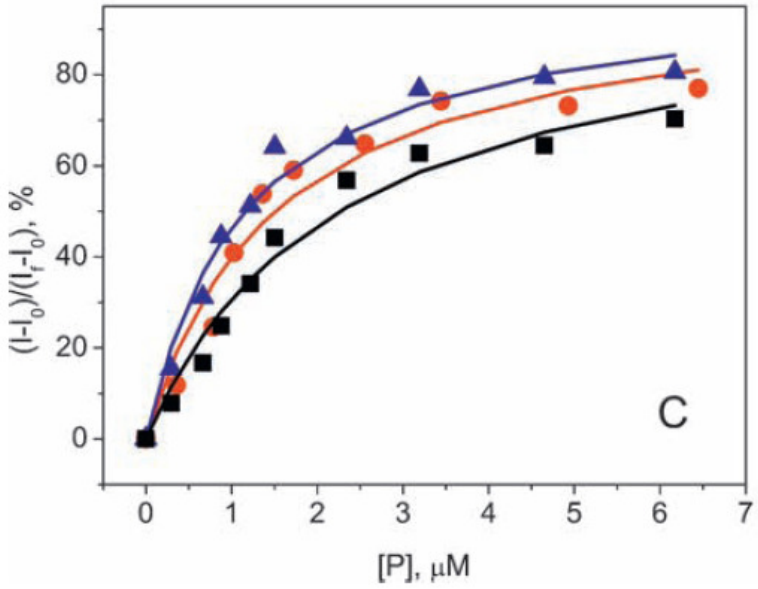
$$\frac{PN_p + PNP}{N_t} = \frac{I_p - I_{p0}}{I_{pf} - I_{p0}} \tag{5.32}$$

where N_t is the total ODN concentration, I_p is the 2Ap fluorescence intensity at a given total protein concentration P_t, I_{p0} and I_{pf} are the 2Ap fluorescence intensities in the absence and presence of saturating NCp7 concentrations, respectively. In contrast, since the concentration P of free NCp7 also depends on the binding of NCp7 to the non-labelled binding site, it cannot be determined from a single titration curve. To get the P concentrations, equal concentrations of an ODN labelled with 2Ap at the p binding site and the same ODN labelled at the d site are titrated in parallel. Under these conditions, P can be calculated by:

$$P = P_t - N_t\left(\frac{I_p - I_{p0}}{I_{pf} - I_{p0}} + \frac{I_d - I_{d0}}{I_{df} - I_{d0}}\right) \tag{5.33}$$


100
80
60
40
20
0
$(I-I_0)/(I_f-I_0)$, %
A
0
2
4
6
8
10
[P], μM


100
80
60
40
20
0
$(I-I_0)/(_f-I_0)$, %
B
0
1
2
3
4
5
6
7
[P], μM


80
60
40
20
0
$(I-I_0)/(I_f-I_0)$, %
C
0
1
2
3
4
5
6
7
[P], μM

Next, the titration data were replotted as $(I - I_0)/(I_f - I_0)$ *versus* P (Figure 5.17 A–C) and the microscopic affinity constants were determined by:

$$\frac{I_p - I_{p0}}{I_{pf} - I_{p0}} = \frac{k_1 P + k_1 k_3 P^2}{1 + (k_1 + k_2)P + k_1 k_3 P^2} \tag{5.34a}$$

$$\frac{I_d - I_{d0}}{I_{df} - I_{d0}} = \frac{k_2 P + k_1 k_3 P^2}{1 + (k_1 + k_2)P + k_1 k_3 P^2} \tag{5.34b}$$

where k_1 and k_2 are the microscopic affinity constants for NCp7 binding to sites p and d, respectively. For all sequences, the substantial differences between the titration curves as a function of the 2Ap position confirmed that 2Ap reports NCp7 binding at a given site. For the ODN1 sequence (Figure 5.17A), good fits were obtained, assuming that the binding sites were independent, and thus that $k_3 = k_2$ and $k_4 = k_1$. The d site was found to bind NCp7 with a 3.5-fold higher affinity than the p site even though the ODN1 sequence is a repeat of two TGACCG motifs, suggesting that NCp7 binding to the two TGACCG motifs is context-dependent. The binding curves with the labelled ODN2 sequences suggest that 2Ap at positions 2 and 5 sense the binding of NCp7 to the same binding site and that this site has a two times lower affinity than the one with 2Ap at position 10 (Figure 5.17B). For ODN3, the ODN3Ap3 and ODN3Ap11 binding curves (Figure 5.17C) suggested that the affinity of the proximal binding site was 50% higher than that of the distal binding site. Surprisingly, the binding curve of ODN3Ap7 was clearly shifted to lower P concentrations, suggesting a higher affinity. In fact, through its central position, 2Ap at position 7 is sensitive to the binding to both sites. As a consequence, the binding was analysed using:

$$\frac{I - I_0}{I_f - I_0} = \frac{I_p - I_{p0}}{I_{pf} - I_{p0}} + \frac{I_d - I_{d0}}{I_{df} - I_{d0}} = \frac{(k_1 + k_2)P + k_1 k_3 P^2}{1 + (k_1 + k_2)P + k_1 k_3 P^2} \tag{5.35}$$

The k_1 and k_2 values obtained through this fit were in excellent agreement with the values obtained from the fits of the ODN3Ap3 and ODN3Ap11 binding curves. As a consequence, ODN labelling by 2Ap appears adequate to determine the microscopic binding constants of individual binding sites. This constitutes a substantial advantage over other techniques such as surface

Figure 5.17 Titration curves of 2Ap-labelled dodecanucleotides with NCp7. The relative 2Ap fluorescence intensity changes $(I - I_0)/(I_f - I_0)$ were plotted as a function of the free protein concentration calculated with eqn (5.33). The ODN concentration was 2 μM in 50 mM HEPES (pH 7.5) and 0.1 M NaCl. Excitation was at 315 nm and emission at 370 nm. Titrations were performed with (A) ODN1 labelled by 2Ap at position 3 (red) or 9 (black), (B) ODN2 labelled by 2Ap at position 2 (black), 5 (red) or 10 (blue) and (C) ODN3 labelled at position 3 (red), 7 (blue) or 11 (black). Solid lines correspond to the fitted curves with eqns (5.34a) and (5.34b), respectively. The blue line in (C) was obtained from a fit with eqn (5.35).

plasmon resonance, fluorescence anisotropy, isothermal titration calorimetry, mass spectroscopy and ELISA, which do not respond site-specifically. This approach may be extended to any nucleic acid binding protein with ODNs containing multiple binding sites provided that those binding sites are non-overlapping.

Another privileged means to monitor protein/ligand interactions is to use FRET. The Trp residues in proteins can be used as fluorescence donors, but their low absorption coefficient and quantum yield do not allow R_0 values greater than 30 Å, requiring that the acceptor should be less than 50 Å apart from the Trp residue. Moreover, the presence of multiple Trp residues complicates the analysis and prevents calculating interchromophore distances. Thus, FRET is generally performed with proteins and ligands site-selectively labelled with strongly absorbing and emitting fluorophores. The D/A couples most frequently used (such as fluorescein/tetramethyl rhodamine, Cy3/Cy5 *etc.*) exhibit R_0 values up to 60 Å, which allow monitoring binding even in rather large complexes and calculate accurately the interchromophore distances. The FRET technique was, for instance, used to monitor the binding of Cy5-labelled AATGCC to a fixed concentration of Rh6G-labelled NC(11–55). The occurrence of FRET in the titration is clearly evidenced by the gradual decrease of the emission of the donor (Rh6G) in parallel to the increase in the emission of the Cy5 acceptor (Figure 5.18A). The transfer efficiency, calculated by $E_{FRET} = 1 - I_{DA}/I_D$ (where I_D and I_{DA} are the intensities of the donor in the

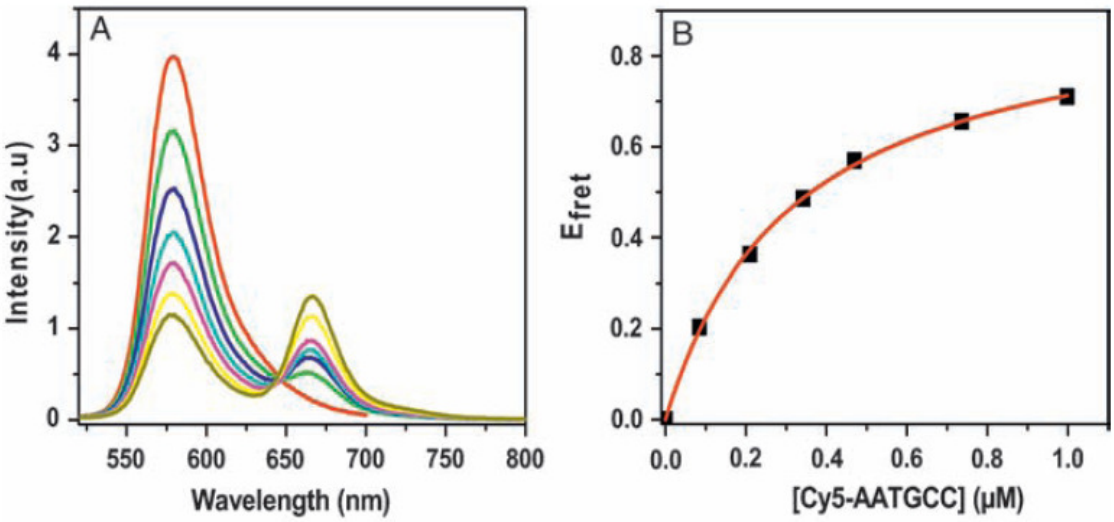


Figure 5.18 Titration of NC(11–55) by AATGCC, as monitored by FRET. The concentration of Rh6G-NC(11–55) was 1 μM. (A) Emission spectra. The red spectrum corresponds to the free peptide. The remaining spectra (from top to bottom) were obtained with concentrations of Cy5-5′-AATGCC from 0.1 to 1 μM. Excitation wavelength was 514 nm. (B) Titration curve. The E_{FRET} values were calculated from the emission of Rh6G at the maximum (580 nm). The data points (squares) were fitted to eqn (5.36), using the K_{app} and E_F values given in the text. The buffer was as in Figure 5.15.

absence and in the presence of acceptor, respectively), was then plotted as a function of the total concentration of the labelled ODN (Figure 5.18B) and fitted with the following equation:

$$E_{FRET} = E_F \left\{ \frac{1 + K_{app}(N_t + L_t) - \left\{ \left[1 + K_{app}(N_t + L_t) \right]^2 - 4K_{app}N_tL_t \right\}^{\frac{1}{2}}}{2K_{app}N_t} \right\} \tag{5.36}$$

giving $K_{app} = 3.1 \times 10^6\,M^{-1}$, very close to that measured from titrations with the intrinsic fluorescence of the peptide.[79] At binding saturation, the value of the final FRET efficiency, E_F was 0.94, giving an interchromophore distance of 31 Å, in good agreement with the expected distance calculated from the NMR structure of the NC(12–53)/ACGCC complex.[78]

FCS can also be used to monitor protein/ligand interactions, but in classical FCS a significant change in the diffusion constant can be evidenced only if the mass ratio between bound and free forms is at least five, due to the cubic root dependence of the diffusion coefficient on the molecular mass and the uncertainty on the excitation volume.[36] Such a large change in the molecular mass is rarely achieved on binding. To avoid this limitation, FCCS can be used. If the two partners labelled by two spectrally distinct fluorophores are bound to each other, they diffuse together through the focal volume, giving a positive cross-correlation which is a direct measure of the binding degree. However, this technique requires that the two partners are labelled and also the two labels do not give FRET, to avoid additional corrections. A more elegant way to circumvent the FCS limitations is to use the recently introduced two-focus FCS (2fFCS)[86] that requires only one labelled partner. In 2fFCS, the fluorescence fluctuations from two overlapping excitation volumes are monitored and the resulting cross-correlation is used to determine absolute diffusion constants. The power of the method relies on the introduction of an external ruler based on the distance between the overlapping beams generated by using a Nomarski prism with two cross-polarised laser beams[86] or two adjustable overlapping beams generated by using a two-photon excitation microscope coupled to a Michelson interferometer.[87] 2fFCS was used to monitor the binding of NC(11–55) to 5′-TMR-AATGCC. Diffusion constants of 216 $\mu m^2\,s^{-1}$ and 172 $\mu m^2\,s^{-1}$ were obtained for free 5′-TMR-AATGCC and 5′-TMR-AATGCC bound to NCp7(11–55), respectively (Figure 5.19A), showing that 2fFCS can discriminate two forms that differ by a factor 3 in their molecular mass. The brightnesses of the free and bound forms were identical indicating that TMR emission was not affected by NC(11–55). Next, the correlation curves obtained by titrating the labelled ODN with NC(11–55) were analysed with a two-population model and fixing the diffusion constants of the free and bound species. An affinity constant $K_{app} = 2.4 \times 10^6\,M^{-1}$ was obtained from the fit of Figure 2.19B, in good agreement with the value

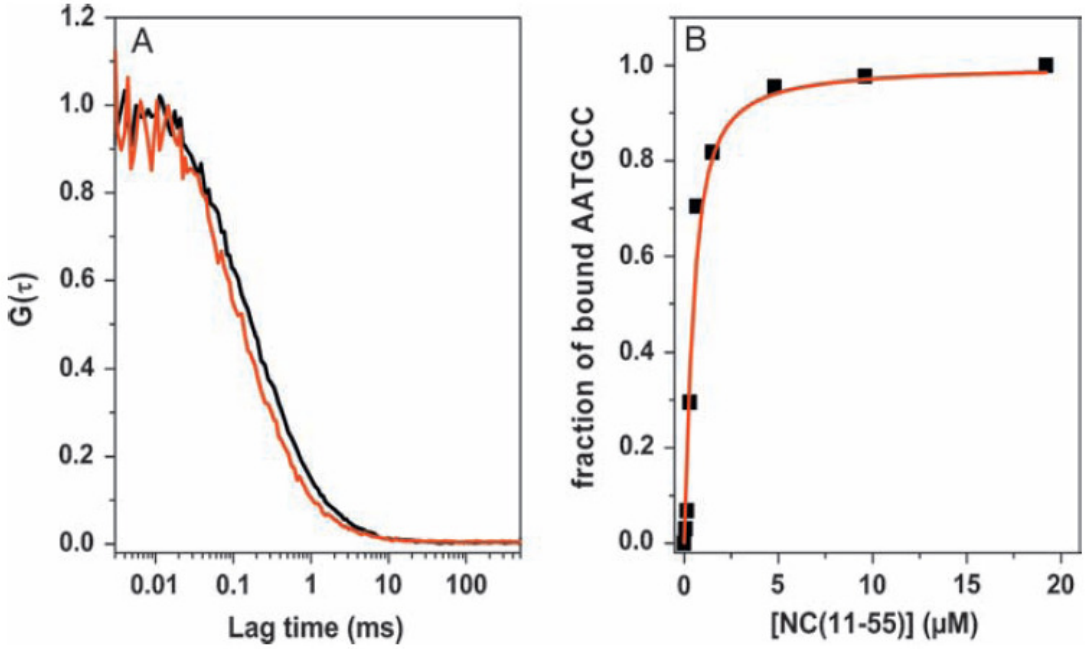


Figure 5.19 Titration of AATGCC by NC(11–55), as monitored by 2fFCS. (A) Correlation curves of 0.3 μM 5′-TMR-AATGCC in the absence (red) and in the presence (black) of 20 μM NCp7(11–55). (B) Fraction of bound AATGCC as a function of the NC(11–55) concentration. The solid line corresponds to the fit of the experimental data (squares) using a Scatchard-derived equation and the K_{app} value given in the text. Typical acquisition time per point was 10 min.

measured from titrations using the intrinsic fluorescence of the peptide[79] or a FRET signal (Figure 5.18B).

5.6 Conclusion

It clearly appears from this overview that the highly sensitive and versatile fluorescence techniques can address through a large number of channels the who, where and why questions in protein/ligand interactions. For answering the "who" question and identifying the potential ligands of a given protein, fluorescence techniques and especially FRET-based techniques and fluorescence anisotropy have been proven to be extremely efficient. Indeed, these techniques being readily amenable to automation and allowing straightforward miniaturisation of the assays, they are largely used in HTS with throughput rates in excess of 100 000 assays per day.[2,88] Moreover, fluorescence-based HTS is not limited to cell-free assays but can also be easily adapted to design cell-based assays. Fluorescence techniques can also ideally answer the "why" question, in order to characterise the thermodynamic parameters governing the interaction. Due to the high sensitivity and large dynamic range of these techniques, affinities in the pM to mM range could be

determined; the lower limit being mainly imposed by the background fluorescence of the sample. In addition to the equilibrium binding parameters (stoichiometry, binding constants, cooperativity) and kinetic parameters (association and dissociation rate constants), information on the entropy and enthalpy changes, as well as on the relative contribution of electrostatic and non-electrostatic contributions, can be obtained. One of the main limitations is the need in most cases of labelling at least one of the partners by an external fluorophore that can perturb the monitored interaction. Thus, great care should be taken to label as selectively as possible a given site, and carefully control the effects of the label on the properties and functions of the labelled partner. Moreover, while the binding parameters could be easily recovered in complexes with 1:1 stoichiometry (provided that the binding process is accompanied by a substantial change of the measured signal), complications associated to the unknown fluorescence of the intermediate binding states appear in systems with multiple binding sites. Thus, determination of the binding constants is limited to systems with low binding stoichiometries. Fluorescence techniques are less suited for answering the ''where'' question, since they provide mainly information on the close environment of the fluorophore. Characterisation of the contribution of individual atoms or residues in the binding process requires site-selective mutations and comparison of the binding parameters for these mutants with those of the unlabeled molecule. This approach provides only indirect information, with a limited spatial resolution. Moreover, even point mutations in proteins can affect not only the binding process, but can also strongly perturb the protein folding and functions, thus giving meaningless results.

Acknowledgments

This work was supported by grants from the French agency against AIDS (ANRS), Sidaction, Centre National de la Recherche Scientifique (CNRS) and Université de Strasbourg. KKS is supported by an ANRS fellowship. We acknowledge Julien Godet for providing data and suggestions.

References

1. J. R. Sportsman, *Methods Enzymol.*, 2003, **361**, 505.
2. S. A. Sundberg, *Curr. Opin. Biotechnol.*, 2000, **11**(1), 47.
3. S. Metcalfe, A. Weeds, A. L. Okorokov, J. Milner, M. Cockman and B. Pope, *Oncogene*, 1999, **18**(14), 2351.
4. C. Eggeling, L. Brand, D. Ullmann and S. Jager, *Drug Discov. Today*, 2003, **8**(14), 632.
5. S. Jager, L. Brand and C. Eggeling, *Curr. Pharm. Biotechnol.*, 2003, **4**(6), 463.

6. L. Zemanova, A. Schenk, M. J. Valler, G. U. Nienhaus and R. Heilker, *Drug Discov. Today*, 2003, **8**(23), 1085.
7. L. A. de Jong, D. R. Uges, J. P. Franke and R. Bischoff, *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.*, 2005, **829**(1–2), 1.
8. R. Heilker, L. Zemanova, M. J. Valler and G. U. Nienhaus, *Curr. Med. Chem.*, 2005, **12**(22), 2551.
9. J. Bacart, C. Corbel, R. Jockers, S. Bach and C. Couturier, *Biotechnol. J.*, 2008, **3**(3), 311.
10. D. S. Smith and S. A. Eremin, *Anal. Bioanal. Chem.*, 2008, **391**(5), 1499.
11. C. Chakraborty, C. H. Hsu, Z. H. Wen and C. S. Lin, *Curr. Pharm. Des.*, 2009, **15**(30), 3552.
12. M. E. Bosch, A. J. Sanchez, F. S. Rojas and C. B. Ojeda, *Comb. Chem. High Throughput Screen.*, 2007, **10**(6), 413.
13. B. Liu, S. Li and J. Hu, *Am. J. Pharmacogenomics*, 2004, **4**(4), 263.
14. J. W. Harris, W. Benenson, H. Stocker and H. Lutz, *Handbook of Physics*, Springer-Verlag, New York, 2002.
15. G. M. Barrow, *The Structure Of Molecules: An Introduction to Molecular Spectroscopy*, ACS Publications, New York, 1964.
16. J. R. Lakowicz, *Principles Of Fluorescence Spectroscopy*, Springer, Baltimore, 1999.
17. B. Valeur, *Molecular Fluorescence – Principles and Applications*, Wiley-VCH Verlag GmbH, Germany, 2002.
18. D. M. Jameson, J. C. Croney and P. D. Moens, *Methods Enzymol.*, 2003, **360**, 1.
19. S. H. Park and R. T. Raines, *Methods Mol. Biol.*, 2004, **261**, 161.
20. S. H. Park and R. T. Raines, *Methods Mol. Biol.*, 2004, **261**, 155.
21. B. J. Anderson, C. Larkin, K. Guja and J. F. Schildbach, *Methods Enzymol.*, 2008, **450**, 253.
22. E. Deprez, P. Tauc, H. Leh, J. F. Mouscadet, C. Auclair and J. C. Brochon, *Biochemistry*, 2000, **39**(31), 9275.
23. A. Hillisch, M. Lorenz and S. Diekmann, *Curr. Opin. Struct. Biol.*, 2001, **11**(2), 201.
24. E. A. Jares-Erijman and T. M. Jovin, *Curr. Opin. Chem. Biol.*, 2006, **10**(5), 409.
25. T. Issad, C. Blanquart and C. Gonzalez-Yanes, *Expert Opin. Ther. Targets*, 2007, **11**(4), 541.
26. A. Prinz, M. Diskar and F. W. Herberg, *Chembiochem*, 2006, **7**(7), 1007.
27. H. Bazin, E. Trinquet and G. Mathis, *J. Biotechnol.*, 2002, **82**(3), 233.
28. H. L. Handl and R. J. Gillies, *Life Sci.*, 2005, **77**(4), 361.
29. H. L. Handl, J. Vagner, H. I. Yamamura, V. J. Hruby and R. J. Gillies, *Anal. Biochem.*, 2005, **343**(2), 299.
30. P. R. Selvin, *Annu. Rev. Biophys. Biomol. Struct.*, 2002, **31**, 275.
31. G. Mathis, *J. Biomol. Screen.*, 1999, **4**(6), 309.
32. G. Zhou, R. Cummings, J. Hermes and D. E. Moller, *Methods*, 2001, **25**(1), 54.
33. K. M. Berland, *Methods Mol. Biol.*, 2004, **261**, 383.

34. D. Grunwald, M. C. Cardoso, H. Leonhardt and V. Buschmann, *Curr. Pharm. Biotechnol.*, 2005, **6**(5), 381.
35. E. Haustein and P. Schwille, *Annu. Rev. Biophys. Biomol. Struct.*, 2007, **36**, 151.
36. S. T. Hess, S. Huang, A. A. Heikal and W. W. Webb, *Biochemistry*, 2002, **41**(3), 697.
37. A. Pramanik, *Curr. Pharm. Biotechnol.*, 2004, **5**(2), 205.
38. K. Bacia, S. A. Kim and P. Schwille, *Nat. Methods*, 2006, **3**(2), 83.
39. L. C. Hwang and T. Wohland, *Cell Biochem. Biophys.*, 2007, **49**(1), 1.
40. R. Y. Tsien, *Annu. Rev. Biochem.*, 1998, **67**, 509.
41. R. Y. Tsien, *Angew. Chem. Int. Ed. Engl.*, 2009, **48**(31), 5612.
42. D. A. Zacharias and R. Y. Tsien, *Methods Biochem. Anal.*, 2006, **47**, 83.
43. R. Bakalova, Z. Zhelev, I. Aoki, K. Masamoto, M. Mileva, T. Obata, M. Higuchi, V. Gadjeva and I. Kanno, *Bioconjug. Chem.*, 2008, **19**(6), 1135.
44. Y. Ghasemi, P. Peymani and S. Afifi, *Acta. Biomed.*, 2009, **80**(2), 156.
45. U. Resch-Genger, M. Grabolle, S. Cavaliere-Jaricot, R. Nitschke and T. Nann, *Nat. Methods*, 2008, **5**(9), 763.
46. W. E. Moerner, E. J. Peterman, S. Brasselet, S. Kummer and R. M. Dickson, *Cytometry*, 1999, **36**(3), 232.
47. C. Julien, A. Debarre, D. Nutarelli, A. Richard and P. Tchenio, *J. Phys. Chem. B*, 2005, **109**(49), 23145.
48. M. R. Eftink and M. C. Shastry, *Methods Enzymol.*, 1997, **278**, 258.
49. M. R. Eftink, *Methods Enzymol.*, 1997, **278**, 221.
50. J. Pager, D. Coulaud and E. Delain, *J. Virol.*, 1994, **68**(1), 223.
51. J. L. Darlix, J. L. Garrido, N. Morellet, Y. Mely and H. de Rocquigny, *Adv. Pharmacol.*, 2007, **55**, 299.
52. J. L. Darlix, M. Lapadat-Tapolsky, H. de Rocquigny and B. P. Roques, *J. Mol. Biol.*, 1995, **254**(4), 523.
53. A. Rein, C. Yang, J. A. Haynes, J. Mirro and R. W. Compans, *J. Virol.*, 1998, **72**(4), 3432.
54. J. A. Thomas and R. J. Gorelick, *Virus Res.*, 2008, **134**(1–2), 39.
55. J. Clever, C. Sassetti and T. G. Parslow, *J. Virol.*, 1995, **69**(4), 2101.
56. J. L. Clever, D. Miranda Jr and T. G. Parslow, *J. Virol.*, 2002, **76**(23), 12381.
57. J. Dannull, A. Surovoy, G. Jung and K. Moelling, *Embo. J.*, 1994, **13**(7), 1525.
58. J. G. Levin, J. Guo, I. Rouzina and K. Musier-Forsyth, *Prog. Nucleic Acid Res. Mol. Biol.*, 2005, **80**, 217.
59. A. Rein, L. E. Henderson and J. G. Levin, *Trends Biochem. Sci.*, 1998, **23**(8), 297.
60. L. Poljak, S. M. Batson, D. Ficheux, B. P. Roques, J. L. Darlix and E. Kas, *J. Mol. Biol.*, 2003, **329**(3), 411.
61. H. Demene, C. Z. Dong, M. Ottmann, M. C. Rouyez, N. Jullian, N. Morellet, Y. Mely, J. L. Darlix, M. C. Fournie-Zaluski and S. Saragosti *et al*, *Biochemistry*1994, **33**(39), 11707.

62. T. Dorfman, J. Luban, S. P. Goff, W. A. Haseltine and H. G. Gottlinger, *J. Virol.*, 1993, **67**(10), 6159.
63. R. J. Gorelick, W. Fu, T. D. Gagliardi, W. J. Bosche, A. Rein, L. E. Henderson and L. O. Arthur, *J. Virol.*, 1999, **73**(10), 8185.
64. R. J. Gorelick, S. M. Nigida Jr, J. W. Bess Jr, L. O. Arthur, L. E. Henderson and A. Rein, *J. Virol.*, 1990, **64**(7), 3207.
65. N. Morellet, N. Jullian, H. De Rocquigny, B. Maigret, J. L. Darlix and B. P. Roques, *Embo. J.*, 1992, **11**(8), 3059.
66. S. V. Avilov, E. Piemont, V. Shvadchak, H. de Rocquigny and Y. Mely, *Nucleic Acids Res.*, 2008, **36**(3), 885.
67. N. Morellet, H. de Rocquigny, Y. Mely, N. Jullian, H. Demene, M. Ottmann, D. Gerard, J. L. Darlix, M. C. Fournie-Zaluski and B. P. Roques, *J. Mol. Biol.*, 1994, **235**(1), 287.
68. R. H. Stote, E. Kellenberger, H. Muller, E. Bombarda, B. P. Roques, B. Kieffer and Y. Mely, *Biochemistry*, 2004, **43**(24), 7687.
69. H. de Rocquigny, V. Shvadchak, S. Avilov, C. Z. Dong, U. Dietrich, J. L. Darlix and Y. Mely, *Mini Rev. Med. Chem.*, 2008, **8**(1), 24.
70. R. H. Goldschmidt and B. J. Dong, *J. Am. Board Fam. Pract.*, 2001, **14**(4), 283.
71. Y. Mely, H. De Rocquigny, N. Morellet, B. P. Roques and D. Gerad, *Biochemistry*, 1996, **35**(16), 5175.
72. T. L. South, P. R. Blake, D. R. Hare and M. F. Summers, *Biochemistry*, 1991, **30**(25), 6342.
73. M. F. Summers, T. L. South, B. Kim and D. R. Hare, *Biochemistry*, 1990, **29**(2), 329.
74. Y. Mely, N. Jullian, N. Morellet, H. De Rocquigny, C. Z. Dong, E. Piemont, B. P. Roques and D. Gerard, *Biochemistry*, 1994, **33**(40), 12085.
75. S. Ramboarina, N. Srividya, R. A. Atkinson, N. Morellet, B. P. Roques, J. F. Lefevre, Y. Mely and B. Kieffer, *J. Mol. Biol.*, 2002, **316**(3), 611.
76. E. Bombarda, E. Grell, B. P. Roques and Y. Mely, *Biophys. J.*, 2007, **93**(1), 208.
77. E. Bombarda, N. Morellet, H. Cherradi, B. Spiess, S. Bouaziz, E. Grell, B. P. Roques and Y. Mely, *J. Mol. Biol.*, 2001, **310**(3), 659.
78. N. Morellet, H. Demene, V. Teilleux, T. Huynh-Dinh, H. de Rocquigny, M. C. Fournie-Zaluski and B. P. Roques, *J. Mol. Biol.*, 1998, **283**(2), 419.
79. C. Vuilleumier, E. Bombarda, N. Morellet, D. Gerard, B. P. Roques and Y. Mely, *Biochemistry*, 1999, **38**(51), 16816.
80. M. T. Record Jr., M. L. Lohman and P. De Haseth, *J. Mol. Biol.*, 1976, **107**(2), 145.
81. Y. Mely, H. de Rocquigny, M. Sorinas-Jimeno, G. Keith, B. P. Roques, R. Marquet and D. Gerard, *J. Biol. Chem.*, 1995, **270**(4), 1650.
82. R. N. De Guzman, Z. R. Wu, C. C. Stalling, L. Pappalardo, P. N. Borer and M. F. Summers, *Science*, 1998, **279**(5349), 384.
83. H. Beltz, C. Clauss, E. Piemont, D. Ficheux, R. J. Gorelick, B. Roques, C. Gabus, J. L. Darlix, H. de Rocquigny and Y. Mely, *J. Mol. Biol.*, 2005, **348**(5), 1113.

84. R. J. Fisher, A. Rein, M. Fivash, M. A. Urbaneja, J. R. Casas-Finet, M. Medaglia and L. E. Henderson, *J. Virol.*, 1998, **72**(3), 1902.
85. R. Khan and D. P. Giedroc, *J. Biol. Chem.*, 1992, **267**(10), 6689.
86. T. Dertinger, V. Pacheco, I. von der Hocht, R. Hartmann, I. Gregor and J. Enderlein, *Chemphyschem*, 2007, **8**(3), 433.
87. P. Didier, J. Godet and Y. Mely, *J. Fluoresc.*, 2009, **19**(3), 561.
88. A. J. Pope and R. Hertzberg, *J. Biomol. Screen.*, 1999, **4**(5), 231.

Results and Discussion

Chapter 1: Characterization of the binding parameters of the D1 domain to HCV oligonucleotides

Nucleic acid binding activity of core protein was analyzed by using synthetic core peptide consisting of a cluster of two basic domains (peptide E: BD1+BD2), as this peptide was easy to synthesize in large quantities and was previously shown to mimic the core chaperone properties [26, 24].

Peptide E exhibits a quantum yield of 0.098. Accessibility of Trp residue in peptide E by iodide ion (using KI) was studied and quenching constants, *kq*, were observed as 1.9 x 10^{-9} $M^{-1}s^{-1}$ and 2.0 x 10^{-9} $M^{-1}s^{-1}$, in the absence and presence of urea, respectively. The absence of any significant difference in the values of quenching constants suggests that the peptide is poorly folded.

Next, to investigate the binding properties of peptide E, we used a 16-nt long palindrome (DLS) sequence labelled with Alexa-430 at its 3' end . This sequence is relevantly taken from the second Stem-loop (SL2) of the HCV 3'X-tail region and has previously shown to be responsible for the dimerisation of HCV genomic RNA [24, 536].

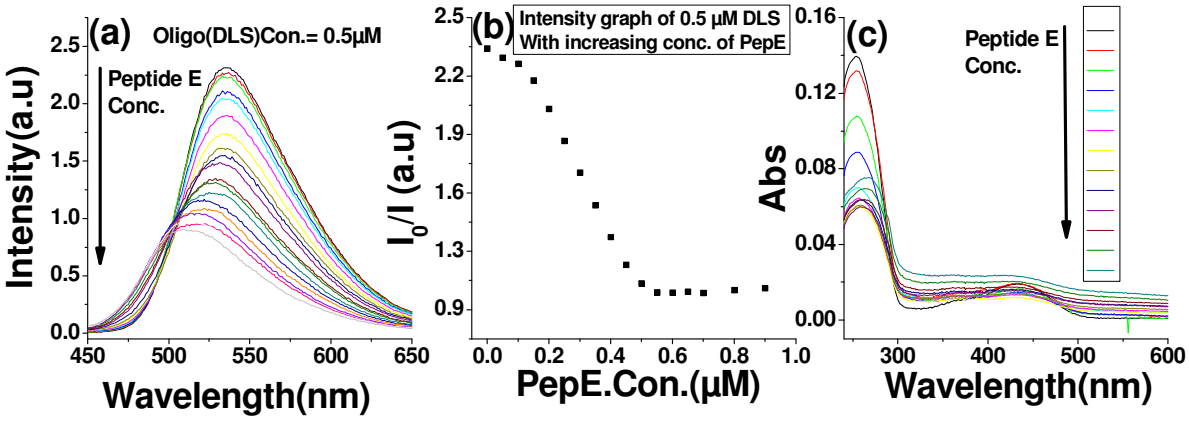


Figure 1: Binding of peptide E to DLS. (a) 0.5 μM DLS labelled with Alexa-430 at its 3'end was titrated with increasing concentrations of peptide E. (b) Intensity curve was plotted against increasing concentration of peptide E *vs* the ratio of fluorescence intensity in absence (I_0) and presence (I) of peptide, at 535 nm and the corresponding (c) absorption spectra.

During the binding course of peptide E to the fluorescently-labelled DLS, a shift in the maximum emission peak (Figure1a) of Alexa-430 and a complicated sigmoidal intensity curve (Figure 1b) were observed. In addition, a strong diffusion with the disappearance of the fluorophore absorbance peak (Figure1c) was also observed. These data suggested formation of aggregates in the presence of peptides E[537]. These aggregates provided complicated

binding curves and thus prevented any quantitative characterization of the binding parameters by fluorescence techniques.

To overcome this hindrance, we examined, in collaboration with the group of D. Altschuh (ESBS, France), the binding of peptide E to native and mutated DLS and SL2 sequences (ODNs) (Figure 3a) by Surface Plasmon Resonance (SPR). This technique can overcome the aggregation problems through the immobilization of one of the binding partners on a chip. SPR experiments were performed on BIAcore 3000 and BIAcore S51 instruments (BIAcore Inc., Piscataway NJ) in HEPES buffer (HEPES 10 mM, pH 7.5, NaCl 150 mM, EDTA 3.4 mM, Tween 20 0.01%).

In our first attempt, we immobilized the peptide E on the surface by amine coupling and ODNs was flushed into SPR, as an analyte. SPR responses were observed only at high peptide density ($\geq$ 300 RU) and at low salt concentration ([NaCl] $\leq$ 50 mM), but those responses showed complex kinetic profiles and progressive loss of binding activity with time during repetitive cycles. These results suggested an irreversible aggregation of immobilized peptides in the presence of oligonucleotides.

To overcome this unexpected behavior, in next step, ODNs were immobilized (by biotine-streptavidine capture) on chip and flushed with peptide E, as an analyte. Solution of peptide E was injected during 1 min over the surface at 50 µl/ min for two or three times to have homogenized interaction. This process in turn provided reproducible sensorgrams under conditions that minimize peptide aggregation, (very low oligonucleotide density (Rmax $\leq$ 30 RU)) and non specific binding (150 mM NaCl) (Figure 2a).

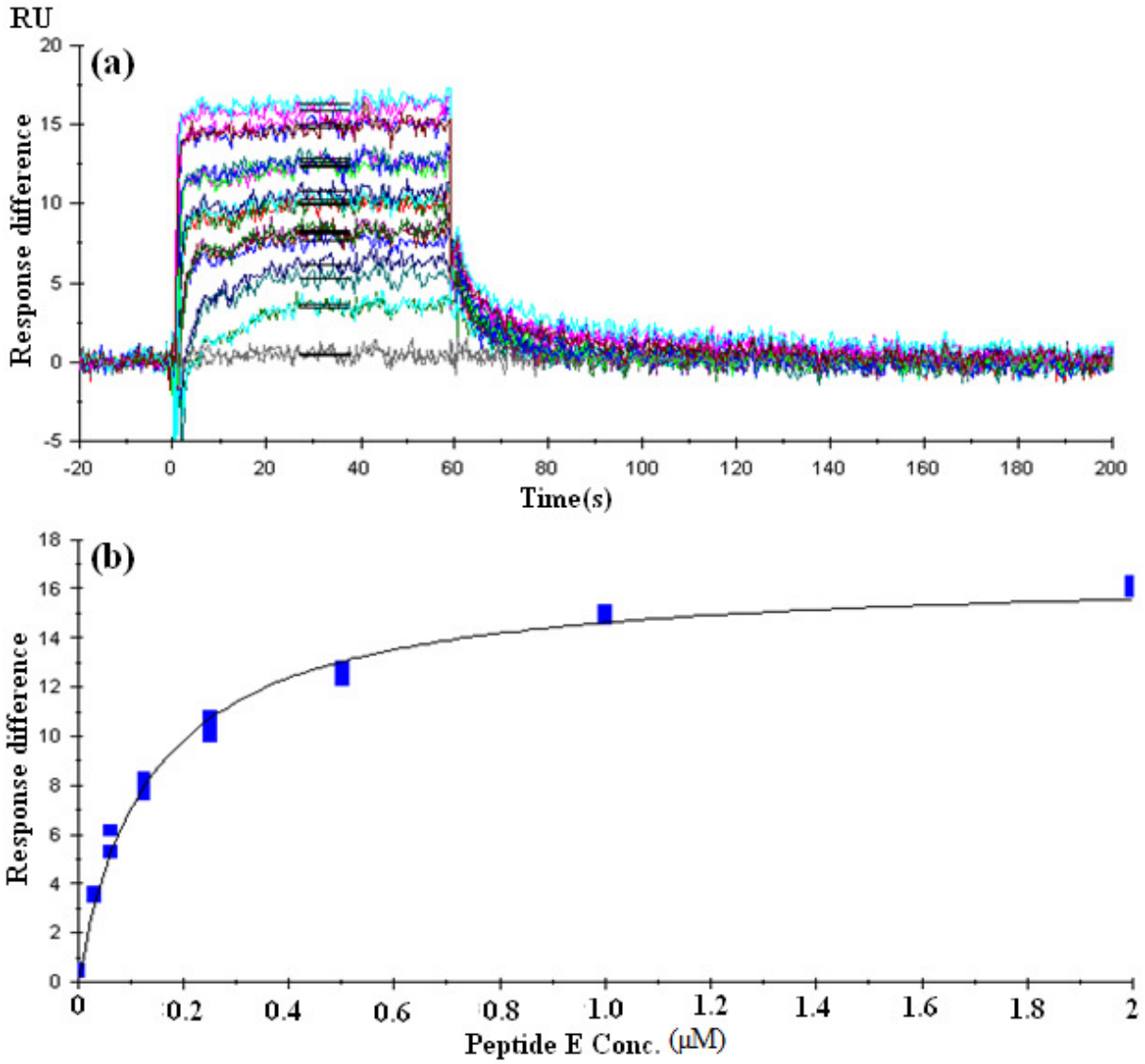


Figure 2: Interaction of peptide E with immobilized biotinylated DLS by SPR. (a) Binding of peptide E to 17 RU of DLS in HEPES buffer (Hepes 10 mM, pH 7.5, NaCl 150 mM, EDTA 3.4 mM, Tween20 0.01%). Solution of peptide E was injected during 1 min over the surface at 50 µl/ min in two to three attempts. (b) The binding response for peptide E at concentrations of 31, 62, 125, 250, 500, 1000 and 2000 nM. To obtain affinity and stoiechiometry of binding, the responses at equilibrium, from figure 2a, were plotted against concentration of peptide E. The black line correspond the best fit to a simple binding site isotherm, described in the materials and methods section.

These sensorgrams in turn provided the value of K_d and its related stoiechiometry (Figure 3b) in aggregation free conditions.

(a) The tested oligonucleotide sequences:

1. **DLS-DNA:** TCACGGCTAGCTGTGA
2. **DLS-RNA:** UCACGGCUAGCUGUGA
3. **Pal2-DNA:** TGACGCCTAGCTGTGA
4. **Pal2-RNA:** UGACGCCUAGCUGUGA
5. **SL2-RNA:** UCACGGCUAGCUGUGAAAGGUCCGUGA
6. **cSL2-RNA:** UCACGGACAUUUCACAGCUAGCCGUGA
7. **cDLS-RNA:** UCACAGCUAGCCGUGA

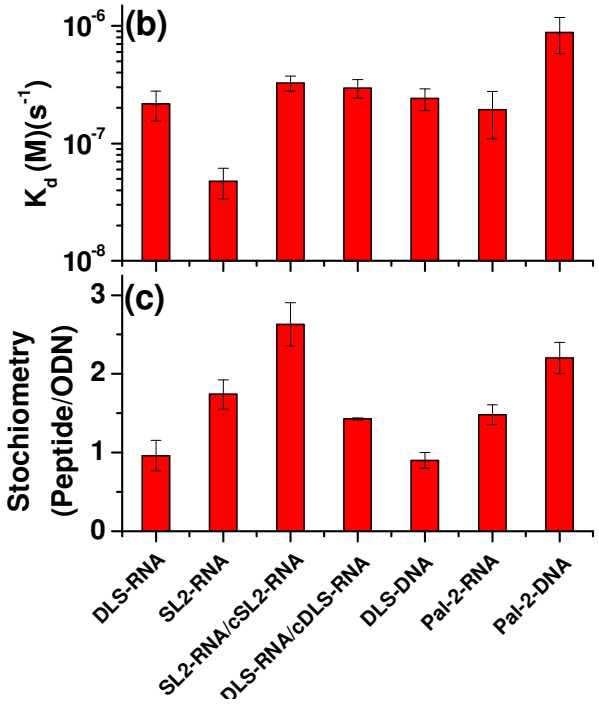


Figure 3: Binding of peptide E with DLS and SL2 and their mutants.

(a) Oligonucleotide sequences used in this study.

(b) Equilibrium dissociation constant (K_d) and (c) number of binding sites for peptide E on single stranded (DLS-RNA, SL2-RNA, DLS-DNA, Pal-2-RNA and Pal-2-DNA) and double stranded (SL2-RNA/cSL2-RNA and DLS-RNA/cDLS-RNA) oligonucleotides.

To test the specificity of the HCV ODNs for peptide E, DLS was flushed with other non-related basic peptides such as scFvQ34S (pI = 10) or lysozyme (pI = 9.2), at similar concentrations to that of peptide E (Data not shown). Interestingly, no SPR response was observed thus, proving its specificity towards peptide E. Moreover, the number of sites occupied by peptide E was dependent on ODNs sequences (Figure3a). A single binding site was revealed for ODNs which adopt a stable hairpin structure, like DLS-RNA or DLS-DNA (Figure 3a). These sequences bind peptide E with similar affinities (K_d) of 0.24 µM. On the contrary, binding of peptide E to unstructured ODNs (like Pal2-RNA, Pal2-DNA) (Figure 3a) showed a higher binding stoiechiometry as compared to structured ODNs, but with a 4-5 fold decrease in affinity.

Additionally, when the same experiments were performed with another stable hairpin structure, SL2-RNA, it was observed that peptide E can occupy two binding sites, with an order of magnitude higher affinity as compared to DLS (Figure 3b). However, taken together SPR results indicated that the interaction between peptide E and SL2-RNA is potentially quite

complex. The binding of peptide E at a level greater than equimolar implies that under these conditions, an additional peptide E can bind to an SL2-RNA:PeptideE complex. In addition, the fact that the SPR profile is function of the SL2-RNA density on the chip suggests that when Peptide E is bound to oligonucleotide, it can still interact with a second oligonucleotide molecule, provided that the latter is in close enough proximity.

We further analyzed the binding affinity of peptide E to double stranded sequences. For this we annealed DLS-RNA and SL2-RNA with their complimentary cDLS-RNA and cSL2-RNA (Figure 3a), respectively, to obtain DLS-RNA/cDLS-RNA and SL2-RNA/cSL2-RNA duplexes. Interestingly, binding of peptide E to SL2-RNA/cSL2-RNA showed a higher binding stoiechiometry (Figure 3c) as compared to structured SL2 but with an order of magnitude decrease in affinity (Figure 3b). On the other hand, an increase of stoiechiometry with no change in affinity was observed for interaction of peptide E with DLS-RNA/cDLS duplex. The increase in stoiechiometry was along the expected lines, but the decrease in affinity needed some explanations. This could probably be explained by the unfolding of structured nucleotides (DLS and SL2), upon hybridization. This hybridization increases the probability of non-specific binding, which leads to an increase in stoiechiometry. On the other hand, hydrogen bonds in RNA or DNA backbone become involved in duplex formation, thus probably leading to low affinity for protein.

Further, data from SPR prompted us to again fit the unsuccessful fluorescence data by taking into account the parameters obtained from SPR experiments. Thus, we analyzed the binding of the peptide E to DLS-RNA, which was studied by steady-state fluorescence intensity as well as steady-state anisotropy in 50mM HEPES pH 7.5, 30mM NaCl and 0.2mM $MgCl_2$. During this measurement, the concentration of the Alexa430-labelled DLS was kept at constant concentration and the peptide E was added in increasing concentrations. The binding curves were fitted by equation 1:

$$r = r_i + \frac{r_p - r_i}{n} \times \frac{\left(1 + K_\mu\left(n[ODN]_{tot} + [\mathrm{Pr}otein]_{tot}\right)\right) - \sqrt{\left(1 + K_\mu\left([\mathrm{Pr}otein]_{tot} + n[ODN]_{tot}\right)\right)^2 - 4nK_\mu{}^2[\mathrm{Pr}otein]_{tot}[ODN]_{tot}}}{2K_\mu[ODN]_{tot}}$$

Equation (1)

Where, $[ODN]_{tot}$ and $[Protein]_{tot}$ are the total concentrations of labelled ODN and unlabelled peptide, respectively. r_p represents the anisotropy when all peptide is bound, while r_i and r

correspond to the anisotropy of Labelled DNA in the absence and presence of protein, respectively. The experimental points were fitted to this model to determine the affinity constant K_μ. The number of binding sites, n, of the protein on the oligonucleotide was fixed on the basis of SPR data (Figure 3c).

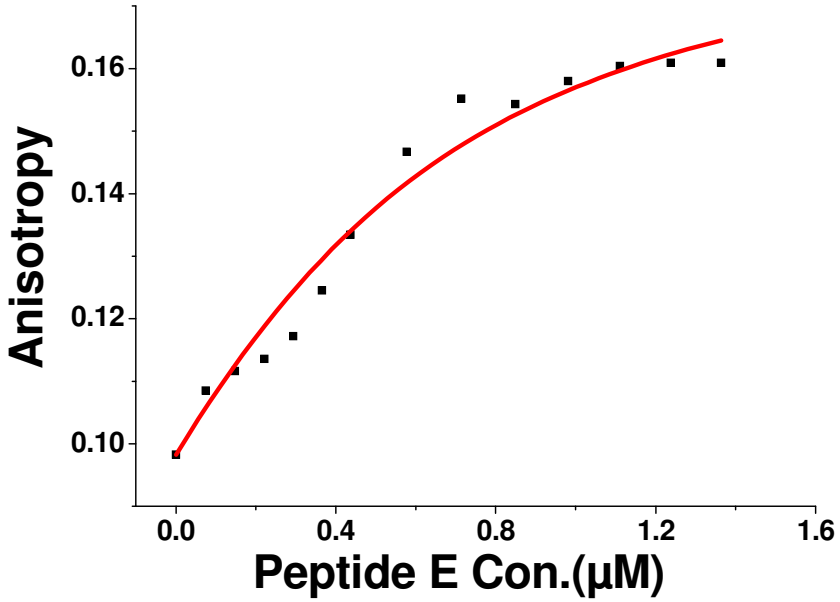


Figure 4: Anisotropy data of binding of peptide E to DLS. 0.5 µM DLS-Alexa 430 was titrated by increasing concentrations of peptide E. The solid red line corresponds to best fit according to equation (1). The number of binding sites was fixed to n = 1, based on SPR data (Figure 3c).

The value obtained for affinity constant was 0.27µM, similar to the value obtained by SPR experiments. This suggests that, although peptide E leads to an aggregation of ODNs, it is still possible to get some relevant data. But the significance and authenticity of the acquired data needed to be verified with other techniques, as the results are biased due to the presence of aggregates (Figure 1c).

We extended our efforts to find more binding parameters of DLS-RNA and SL2-RNA with peptide E and analyzed their binding by using single molecule Fluorescence resonance energy transfer (smFRET) technique. For this, total internal reflection fluorescence microscopy, with a 532 nm laser excitation, was used to measure smFRET of single interacting pairs of Cy5-labelled ODNs (Cy5-DLS-RNA or Cy5-SL2-RNA) and Rhodamine-labelled Peptide E (Rh-pepE). Single interacting pairs were trapped in liposomes as explained in material and

methods section.

smFRET analysis

Rh-pepE at single molecule level

Initially, the fluorescence of Rh-pepE molecules at single molecule level was analyzed by encapsulating them in EYPC liposomes. Liposomes were formed and immobilized on the observation surface as described in material and methods section. Fluorescent spots corresponding to liposomes encapsulating Rh-pepE were observed, after illuminating the surface at 532nm with a laser.

Fluorescence traces corresponding to these spots have the shape of one (or more) step (s) (Figure 5). It was observed that the fluorescence emitted by the labeled protein was constant over a certain period of time until the photobleaching of probe. After photobleaching, no more fluorescence signal was recorded. Some observations show two or even three events of photobleaching, suggesting that the liposome can encapsulate two or three labeled molecules (Data not shown). But the majority of the liposomes encapsulated only one molecule of Rh-pepE (Figure 5).

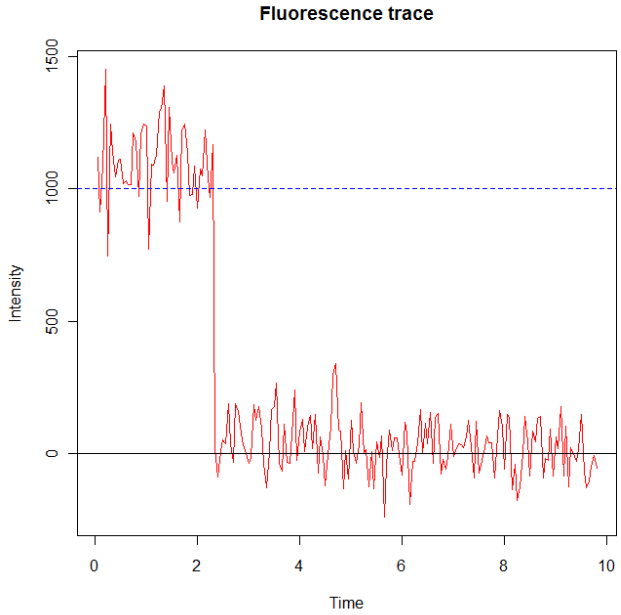


Figure 5: Fluorescence traces of encapsulated Rh-pepE in liposomes with 100 nm diameter. The presented trace corresponds to the fluorescence associated with a liposome encapsulating one molecule of Rh-peptide E.

The analysis of combined data of these fluorescence traces (Figure 6) from different experiments in turn provided the fluorescence parametres of the Rh-pepE population (or

populations). The analysis of data was performed as described in material and methods section.

It was observed that the fluorescent population of peptide E was rapidly depopulated (within 7-8 seconds under these conditions of acquisition) to a state which corresponds to the state after photobleaching (Figure 6).

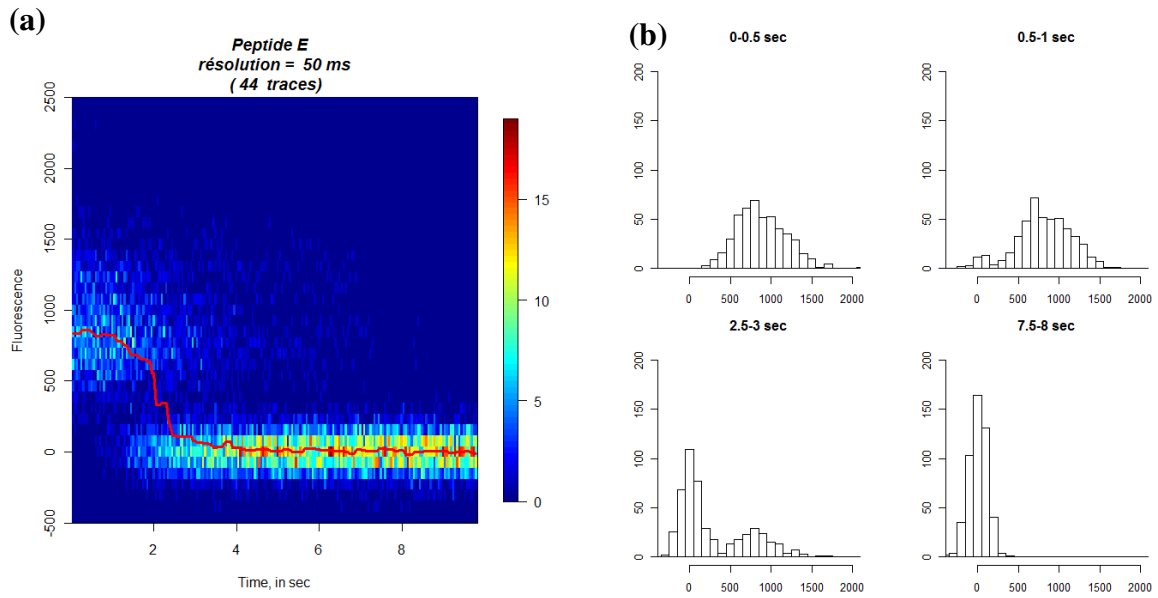


Figure 6: (a) Combined representation of the 44 fluorescent vesicles encapsulating Rh-pepE *vs* time. Each pixel represents a number of molecules associated to the provided color code (the number of molecules increases with the heat of the color). (b) Data representing the shift of the fluorescent population of Rh-pepE to a non-fluorescent population in around 7-8 seconds.

However, the combined distribution of fluorescence traces confirmed that the majority of the vesicles contains single fluorescent molecules of Rh-pepE (Figure 6a) that photobleached only after 2 seconds (Figure 6b).

In the end, all these experiments showed that:

- The encapsulation of individual flurescent molecules within immobilized liposomes is possible.
- Peptide E does not interact with the liposome surface.
- Rhodamine labelling of the peptide E, is photostable and sufficiently fluoresces to be used as fluorescence donor in a FRET pair in single molecule conditions.
- The fluorescence signal of the Rh-peptide E behaves in a biphasic manner by initially emitting stable fluorescence intensity for a few seconds before photobleaching.

In a next attempt, FRET as a consequence of binding of peptide E to Cy5-labelled HCV ODNs was monitored by the same procedure.

To this, we first analyzed FRET between Rh-pepE and Cy5-DLS-RNA. The encapsulated complex of rhodamine labeled peptide E and cyanine-5 labeled DLS-RNA, Cy5-DLS/pepE, showed fluorescent spots of both donor and acceptor channels. Interestingly, fluorescence traces of the donor and acceptor were found clearly anti-correlated. The time evolution of the apparent FRET transitions showed two levels of fluorescence, one where the fluorescence was strong (close to 1) and another where the fluorescence was virtually zero. These two fluorescence levels or states were associated to Rh-pepE in free and bound form, respectively. Next, unlike liposomes enclosing only the Rh-pepE, the presence of Cy5-DLS-RNA resulted in a strong fluctuating signal of Rhodamine, before the photobleaching step. This signal of the rhodamine varied between two fluorescent states viz. a non-fluorescent state in presence of Cy5-DLS-RNA and a fluorescent state with Rh-pepE remained alone. These traces showed multiple transitions between non-fluorescent and fluorescent states (Figure 7a).

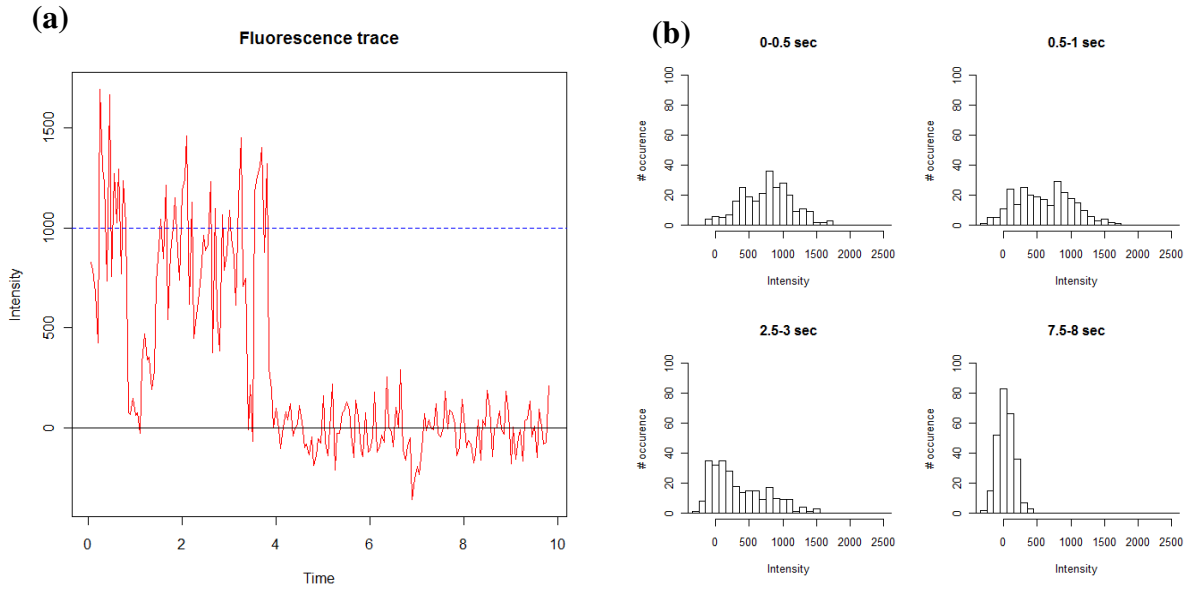


Figure 7: Fluorescence traces of encapsulated Cy5-DLS-RNA/Rh-pepE complex in liposomes of 100 nm diameter. (a) The fluorescence trace corresponds to fluorescence fluctuations associated with a liposome encapsulating one pair of Cy5-DLS RNA and Rh-pepE. (b) Fluorescene distribution states established by analyzing a large number of combined fluorescence traces.

The distribution showed that the polydispersity in the fluorescent population as compared to Rh-pepE alone (Figure 7b), is probably attributed to the effect of interaction with Cy5-DLS-

RNA.

Next, from these combined obtained traces, dwell-time histograms of residence time were plotted. These histograms represented the occurrence of a particular transition (corresponding to either dissociation or association event) as a function of the time spend in the initial state before the transition occurred.

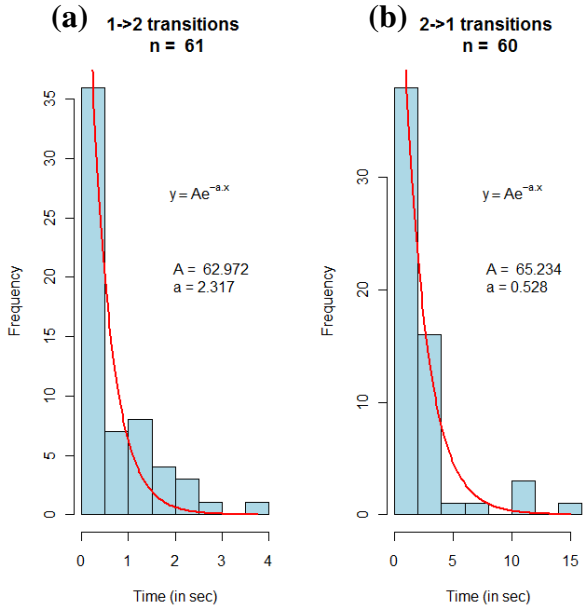


Figure 8: Dwell-time Histogram of Cy5-DLS-RNA/Rh-pepE complex.
Histograms represent the mean residence time (in sec^{-1}) of the system during the transition from (a) bound to free state ($\tau_{bound} \rightarrow \tau_{free}$), and from (b) free to bound state ($\tau_{free} \rightarrow \tau_{bound}$). The red line corresponds to the single exponential decay fit as described as $y = Ae^{-ax}$ and n represents number of treated fluorescence traces.

The decay calculated from the $\tau_{bound} \rightarrow \tau_{free}$ residency time distribution (Figure 8a) directly corresponded to the monomolecular dissociation rate and was about ~2.3 s^{-1}. To retrieve the association rate constant, the observed exponential rate decay was divided by the effective concentration of the complex (3 µM) inside the 100 nm vesicle. This provided a k_{on} value of 1.76×10^{7} $M^{-1}s^{-1}$. The ratio between k_{on} and k_{off} (k_{off}/k_{on}), in turn showed an apparent binding constant of $k_d = 1.3 \times 10^{-7}$ M. Interestingly, the value obtained by single molecule for binding of peptide E to DLS, was close to that observed with SPR and binding studies.

Similar analysis of the Cy5-SL2-RNA/pepE complexes by the single molecule platform, provided fluorescence traces which were associated to liposomes encapsulating 2 peptides E

and one SL2-RNA. To fit the acquired data, a three state model was used, which was able to define the presence of multiple steps in the fluorescence trace (see figure 9 for an illustration).

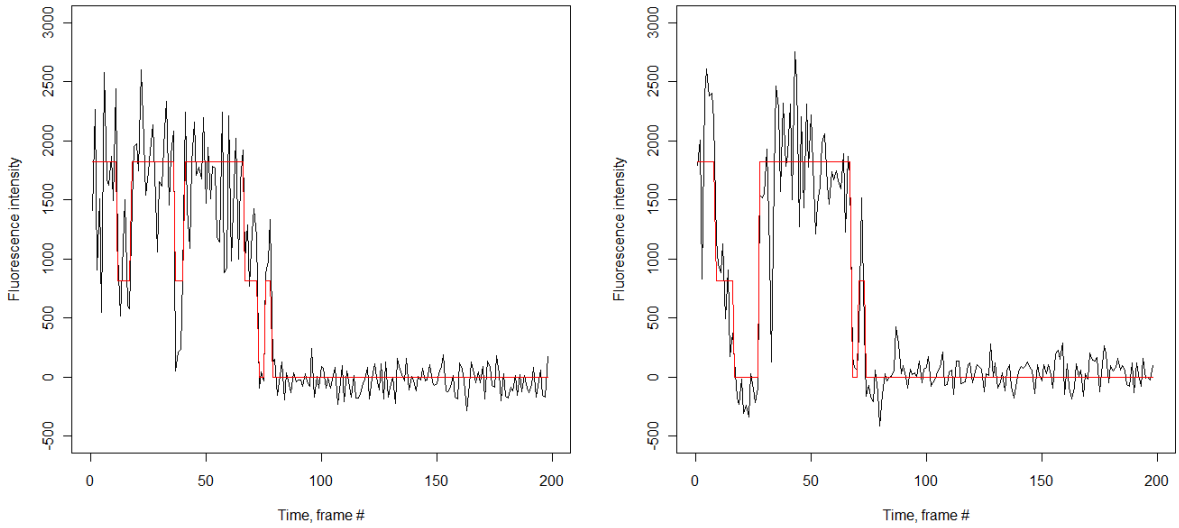


Figure 9: An illustration of multistep fluorescence traces.

In this three state model, we assumed that the lowest fluorescent level corresponds to a state where the two peptide E were bound (state 1) to a single SL2-RNA molecule. The intermediate fluorescent (state 2) state was associated to the binding of a single peptide E, while the second remain free and the highest fluorescent level (state 3) corresponds to the two peptides E in their free form. We modeled these states transitions with a hidden Markov model and retrieved the residence time from the dwell-time histograms (Figure 10). We observed that most of the transitions occur within states 2 and 3. The probability of direct transitions between the states 1 and 3 was very low (Figure 10 as 1->3 and 3->1), in line with the low probability of observing simultaneous dissociation or association of 2 peptides E. Thus, we estimated the binding constants of peptide E to SL2 from the transitions between states 2 and 3 (Figure 10 as 2->3 and 3->2), for which a k_{on}= 3.1 × 10^{7} $M^{-1}s^{-1}$ and k_{off} =1.8 s^{-1} were observed. The ratio of dissociation and association rate constants provided K_{d1} = 5.8 × 10^{-8} M, which was in good agreement with SPR experiments. Using the same calculations for the transitions between state 1 and 2 (Figure 10 as 1->2 and 2->1), which represent the binding of second peptide E to Cy5-SL2-RNA/Rh-pepE complex, we observed K_{d2} = 4.2 × 10^{-6} M for the second binding site.

The two orders of magnitude difference between k_{d1} and k_{d2}, confirms the possibility of two different binding sites for SL2-RNA, which is again in line with the observed SPR data. The

difference of binding affinity of these sites can probably be explained by the sequential attachment of peptide E to SL2-RNA. It is possible that the two binding sites are really different in their binding affinity or that attachment of one peptide molecule may affect the affinity of other binding site.

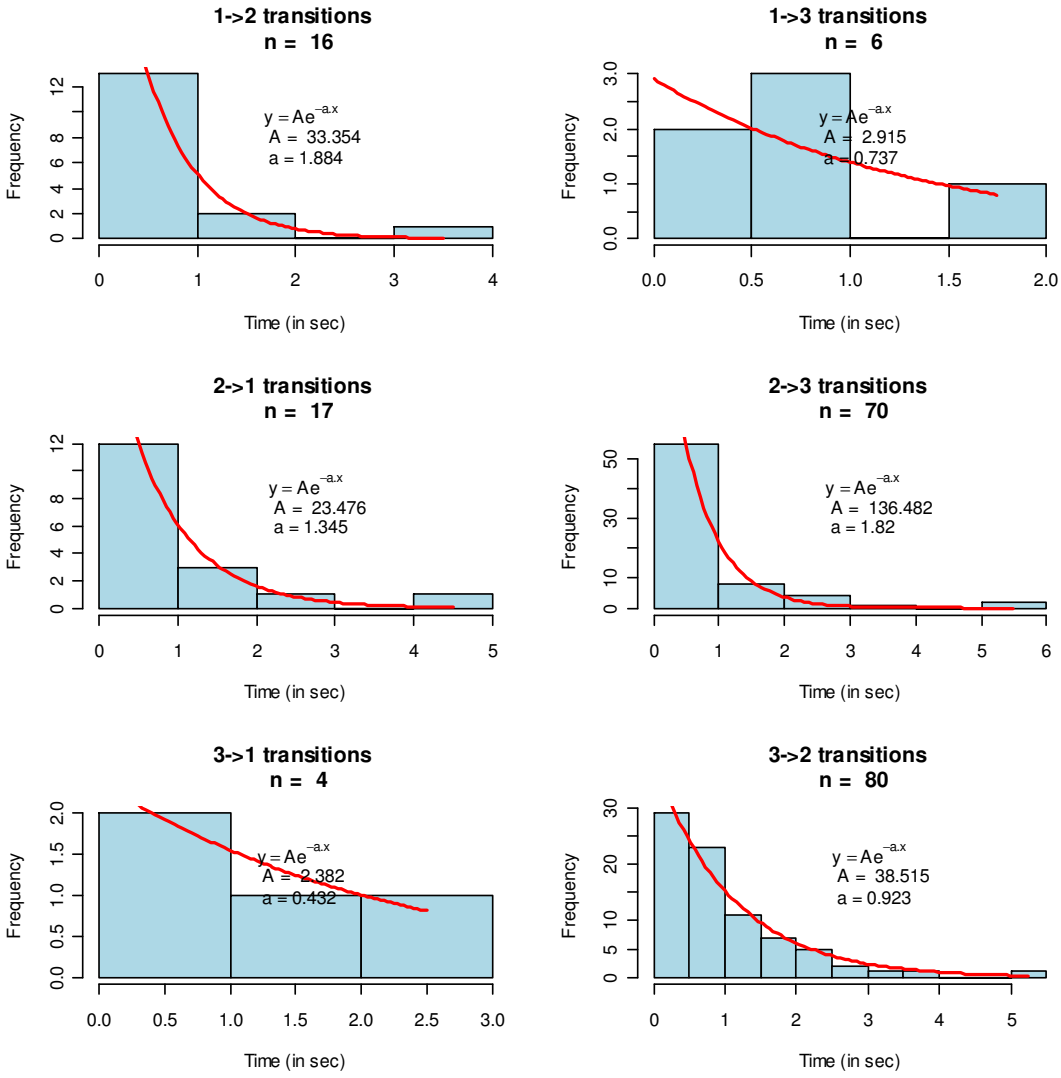


Figure 10: Dwell-time Histograms of fluorescence fluctuations associated to Cy5-SL2-RNA/Rh-pepE complexes in peptide to ODN ratio of 2:1.

Histograms represent the mean residence time (in sec^{-1}) of the system during different transitions. The red line corresponds to the single exponential decay fit described as $y = Ae^{-ax}$ and n represents number of treated fluorescence traces.

Binding parameters of the basic domains of the core protein with HCV oligonucleotides had never been determined before by a direct method due to its high aggregation properties. Thus,

these results are of high importance. The data suggested that peptide E can bind specifically to the structured oligonucleotides that are in abundance in 3' X-Tail region of HCV genome.

Chapter 2: Kinetic investigation of model DNA annealing in the presence of core protein

We characterized the mechanism by which the core protein can activate the annealing of complementary DNA sequences by using fluorescence techniques. For this, a model system based on the annealing of the DNA copy of the Transactivation response element (dTAR) of HIV-1 and its complementary sequence cTAR was used. These sequences were selected since their annealing in the presence of several chaperone proteins has been extensively investigated. To analyze the mechanism of chaperone activity of core protein we used synthetic core peptides consisting of a cluster of the three basic domains (peptide F: BD1+BD2+BD3) or only of two of them (peptide E: BD1+BD2) and Domain D1.

Peptide E, F and Domain D1 with labelled cTAR and non-labelled dTAR mutants were used to analyze the chaperone properties of the D1 domain of core protein. Since positively charged peptides are known to aggregate nucleic acids in a concentration-dependent manner, we determined aggregation-free conditions by using Fluorescence Correlation Spectroscopy and performed experiments in these conditions. Core peptides strongly accelerated the annealing reaction, since the reaction was completed in less than 30 min, instead of more than 1 day in the absence of peptides. Variation in dTAR concentration indicated a saturation behaviour of both fast, k_{obs1}, and slow, k_{obs2}, kinetic rate constants, which is in consistent with a two-step reaction mechanism where a fast pre-equilibrium intermediate complex precedes the formation of the final stable extended duplex (ED). Moreover, different mutants of cTAR and dTAR showed that the core peptides nucleate cTAR/dTAR annealing through their respective ends.

On the basis of our kinetic data, a reaction mechanism with two parallel kinetic pathways involving two different intermediate complexes was proposed. In the proposed model, the fast and slow pathways differed by the number of base pairs, which should be melted to nucleate the intermediates. In addition, we confirmed that the chaperone activity of the core protein is mainly mediated through its three basic clusters. Furthermore, comparison of HCV peptide E with NC (11-55) peptide, which mimics the activities of NCp7 without aggregation, showed that one peptide E molecule exhibits the same chaperone activity as four NC(11-55) peptide molecules and that the two peptides can chaperone in concert the annealing of the two complementary ODNs.

Further, efficient chaperone properties of the core peptides on the model TAR oligonucleotides used in this study further suggest that the core protein is a nucleic chaperone with broad sequence specificity. These properties may be critical for genomic RNA dimerization in HCV replication and packaging, as well as for facilitating recombination between various HCV genotypes and subtypes to increase viral variability.

Nucleic Acids Research Advance Access published February 18, 2010

Nucleic Acids Research, 2010, 1–11
doi:10.1093/nar/gkq094

Kinetic analysis of the nucleic acid chaperone activity of the Hepatitis C virus core protein

Kamal kant Sharma[1], Pascal Didier[1], Jean Luc Darlix[2], Hugues de Rocquigny[1], Hayet Bensikkadour[1], Jean-Pierre Lavergne[3], François Pénin[3], Jean-Marc Lessinger[1] and Yves Mély[1,*]

[1]Laboratorie de Biophotonique et Pharmacologie, UMR 7213 CNRS, Faculté de Pharmacie, Université de Strasbourg, 74, Route du Rhin, 67401, Illkirch, Cedex, [2]LaboRétro, Unité de Virologie Humaine INSERM, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364, Lyon and [3]Institut de Biologie et Chimie des Proteines, UMR 5086 CNRS, Université de Lyon, IFR 128 BioSciences Gerland Lyon Sud, 7 Passage du Vercors, 69367 Lyon, France

Received January 8, 2010; Revised February 1, 2010; Accepted February 3, 2010

ABSTRACT

The multifunctional HCV core protein consists of a hydrophilic RNA interacting D1 domain and a hydrophobic D2 domain interacting with membranes and lipid droplets. The core D1 domain was found to possess nucleic acid annealing and strand transfer properties. To further understand these chaperone properties, we investigated how the D1 domain and two peptides encompassing the D1 basic clusters chaperoned the annealing of complementary canonical nucleic acids that correspond to the DNA sequences of the HIV-1 transactivation response element TAR and its complementary cTAR. The core peptides were found to augment cTAR-dTAR annealing kinetics by at least three orders of magnitude. The annealing rate was not affected by modifications of the dTAR loop but was strongly reduced by stabilization of the cTAR stem ends, suggesting that the core-directed annealing reaction is initiated through the terminal bases of cTAR and dTAR. Two kinetic pathways were identified with a fast pre-equilibrium intermediate that then slowly converts into the final extended duplex. The fast and slow pathways differed by the number of base pairs, which should be melted to nucleate the intermediates. The three peptides operate similarly, confirming that the core chaperone properties are mostly supported by its basic clusters.

INTRODUCTION

Hepatitis C virus (HCV) (1) is an important human pathogen transmitted mainly through blood contacts and affecting ~170 million people worldwide (2). Chronic infections are frequent and a major cause of liver cirrhosis and cancer (3). HCV is an enveloped virus of the Flaviviridae family, which has a positive-sense, single-stranded RNA genome with an open-reading frame encoding a polyprotein of 3010 amino acids or so (4). The 5′- and 3′-ends of the viral RNA correspond to non translated, highly structured and conserved regions (5′ UTR and 3′ UTR) of 340 and 230 nucleotides, respectively. The 5′ UTR consists of five stem–loops (5) and includes an internal ribosome entry site (IRES) that directs translation of the HCV genome in a cap-independent manner. The 3′ UTR is composed of three distinct parts: (i) a variable part with the VSL1 and VSL2 stem–loops, (ii) a poly (U/UC) tract and (iii) a sequence X of 98 nucleotides (6,7). Deletion of the U-rich and X sequence results in a loss of virus replication and infectivity (8).

The polyprotein precursor is co- and post-translationally processed by cellular and viral proteases at the endoplasmic reticulum (ER) membrane to yield the mature structural and nonstructural proteins. The structural proteins include the core and the envelope glycoproteins E1 and E2 (9). The N-terminal end of the polyprotein, which is a major HCV antigen (10,11), is cleaved by a cellular signal peptidase, generating the core of 191 residues which binds the ER membrane (12). This protein is then processed by a signal peptide peptidase (13,14), giving a mature protein of 179 residues or so, which is targeted to the lipid droplets (LDs) (15–17). Apart from LDs, the protein can be retained at the ER membrane and localized on mitochondria (15,18). The mature protein is a dimeric, alpha-helical protein exhibiting features that are consistent with those of a membrane protein (19) . It consists of a basic N-terminal RNA binding domain D1 (aa 2–117)

*To whom correspondence should be addressed. Tel: +33 (0)3 68 85 42 63; Fax: +33 (0)3 68 85 43 13 Email: Yves.mely@pharma.u-strasbg.fr

with three highly basic amino acid clusters (BD1-BD3) (20), and a hydrophobic C-terminal domain D2 (aa 118–179 or so) involved in targeting the protein to LDs (21). Core attachment to LDs is critical for assembly and release of infectious HCV particles (HCVcc) (22).

The core exerts multiple roles in infected cells and virus replication. The core notably interacts with a number of cellular proteins which impacts on cell proliferation and differentiation (10,23) and also on the lipid metabolism with a possible influence on hepatocellular carcinomas (24). As a major viral component tightly binding to genomic RNA sequences, the core is also thought to drive genomic RNA packaging and nucleocapsid formation (25–27). Moreover, the core has potent nucleic acid chaperoning activities (20,28), similar to retroviral nucleocapsid (NC) proteins (29–35). Through this activity, the core can resolve RNA misfolding and promote annealing of complementary sequences and strand exchanges (28). This activity appears to be mediated by the three basic clusters (BD1–BD3) of the D1 domain (20,28).

To further characterize the core RNA chaperone properties, we investigated how the D1 domain and peptides encompassing the basic clusters direct the annealing of complementary DNA sequences represented by the canonical model of the HIV-1 transactivation response (TAR)(36) element in a DNA form (dTAR) and the complementary sequence cTAR. The extremely slow annealing of these two sequences is drastically accelerated by addition of a retroviral NC protein (30,32,34) or the HCV core (28). The nucleocapsid protein NCp7 of HIV-1 activates the transient opening of the cTAR terminal base-pairs (29,37) causing a partial melting of the stem (33,38). These partly melted TAR DNAs anneal then through their termini to form an intermediate complex which is further converted into the final extended duplex (ED) (30,33,35). To determine whether the HCV core promotes cTAR/dTAR annealing through a similar molecular mechanism, we monitored the real-time annealing kinetics of fluorescently labelled cTAR with dTAR and mutants thereof in the presence of D1 Domain, and peptides E and F corresponding to the two (BD2 and BD3) and all three (BD1, BD2 and BD3) basic clusters, respectively. Results indicate that the three core peptides destabilize cTAR and dTAR secondary structures to a small extent, but efficiently promote dTAR/cTAR annealing through the dTAR and cTAR ends.

MATERIALS AND METHODS

Core peptides and oligonucleotides

Peptides E and F were synthesized by solid-phase peptide synthesis on a 433 A synthesizer (ABI, Foster City, CA, USA) as described (20). Purification by HPLC was carried out on a C8 column (Uptisphere 300 A, 5 μm; 250×10, Interchim, France) in 0.05% TFA with a linear gradient of 10–70% of acetonitrile for 30 min. The peptide purity and molecular weight (6754 for peptide E and 9444 for Peptide F) were checked by LC/MS. Absorption coefficient of 5700 $M^{-1} cm^{-1}$ at 280 nm was used to determine their concentration.

The D1 domain consisting of amino acids 1–117 of the core protein fused to a C-terminal 6xHis tag was prepared as previously described (19).

The unlabelled and the labeled ODNs were synthesized by IBA GmbH Nucleic Acids product Supply (Gottingen, Germany). In the case of the doubly labeled ODNs, the 5′ terminus was labeled with carboxytetramethylrhodamine (TMR) or ethyl 2-[3-(ethylamino)-6-ethylimino-2, 7-dimethylxanthen-9-yl]benzoate hydrochloride (Rh6G) via an amino-linker with a six carbon spacer arm, while the 3′ terminus was labeled with either 4-(4′-dimethylaminophenylazo) benzoic acid (Dabcyl) or 5(and 6)-carboxyfluorescein (Fl) using a special solid support with the dye already attached. ODNs were purified by the manufacturer by reverse-phase HPLC and polyacrylamide gel electrophoresis.

Experiments were performed in 25 mM Tris–HCl (pH 7.5), 30 mM NaCl and 0.2 mM $MgCl_2$ at 20°C.

Fluorescence measurements

Emission spectra and kinetic traces were recorded with Fluorolog and FluroMax spectrofluorimeters (Jobin Yvon Instruments, S.A. Inc.) equipped with a temperature-controlled cell compartment. All fluorescence intensities were corrected for buffer emission and Lamp fluctuations.

Kinetic measurements were performed in pseudo first-order conditions by using concentrations of unlabelled TAR and dTAR at least 10-fold higher than the concentration of the labeled complementary sequence. Excitation and emission wavelengths were 520 and 550 nm, respectively, to monitor the Rh6G fluorescence. The corresponding wavelengths were 480 and 520 nm, respectively to monitor the Fl fluorescence. All reported concentrations correspond to those after mixing. To avoid high local concentrations during mixing, both reactant mixtures were prepared of the same volume. Peptides were added to each reactant separately, and then the reaction was initiated by mixing the peptide-coated ODNs together. The kinetics was fast enough to monitor the fluorescence intensities continuously without photobleaching. The apparent rate constants k_{obs} and the amplitudes were determined from the kinetic traces by including a dead-time correction t_0 to take into account the delay between the mixing of reactants and the start of the measurements. All fitting procedures were carried out with Origin™ 7.5 software based on nonlinear, least-square methods and the Levenberg-Marquardt algorithm.

The temperature dependence of the annealing kinetics was carried out with 10 nM doubly labeled cTAR derivatives and 300-nM nonlabeled dTAR derivatives at different temperatures (10°C, 20°C, 25°C, 30°C, 35°C and 40°C) in the presence of either peptide E, peptide F and Domain D1, added at a peptide to oligonucleotide ratio of 1.4:1.

Fluorescence correlation spectroscopy setup and data analysis

Fluorescence correlation spectroscopy (FCS) measurements were performed on a two-photon platform

including an Olympus IX70 inverted microscope, as described previously (37,39). Two-photon excitation at 850 nm is provided by a mode-locked Tsunami Ti:sapphire laser pumped by a Millenia V solid-state laser (Spectra Physics, USA). The experiments were performed in an eight-well lab-Tek II coverglass system, using a 400-µl volume per well. The focal spot was set about 20 µm above the coverslip. Peptide E was added to the labeled cTAR sequences at different nucleotide to peptide ratios. To avoid high local concentrations during mixing, both reactants were of the same volume. Assuming that peptide E diffuses freely in a Gaussian excitation volume, the normalized autocorrelation function, $G(\tau)$, calculated from the fluorescence fluctuations was fitted according to (40):

$$G(\tau) = \frac{1}{N}\left(1+\frac{\tau}{\tau_d}\right)^{-1}\left(1+\frac{1}{S^2}\frac{\tau}{\tau_d}\right)^{-1/2} \quad (1)$$

where τ_d is the diffusion time, N is the mean number of molecules within the excitation volume and S is the ratio between the axial and lateral radii of the excitation volume. The point spread function of the set-up was determined from a z-scan on one fluorescent bead (20 nm in diameter). The measured lateral and axial resolutions were respectively 0.3 and 1 µm. Typical data recording times were 10 min. Twenty autocorrelation curves were recorded for each sample. When spikes of high fluorescence intensity, most likely associated to aggregates, were observed in the fluorescence fluctuation profiles, the corresponding autocorrelation curves were discarded. In these conditions, the average autocorrelation curve calculated from the sum of the remaining autocorrelation curves could be adequately fitted by a single population model with a diffusion time close to that of cTAR alone.

RESULTS

Investigating the interaction of cTAR DNA with the core peptides by FCS

To characterize in depth the nucleic acid chaperone properties of the core protein, we first had to find experimental conditions where the core peptides did not cause DNA aggregation in a manner similar to other nucleic acid chaperones (32,34,41,42). Indeed, molecular aggregation can cause strong bias when using fluorescence-based techniques (30,32). Since nucleic acid aggregation by positively charged peptides is concentration dependent (42,43), we investigated this dependence by means of FCS. In FCS, the fluorescence intensity arising from a very small volume (about 0.2 fL) is correlated to obtain information about the processes that give rise to fluorescence fluctuations. These fluctuations are mainly governed by the diffusion of fluorescent species throughout this volume, and thus parameters such as the average number of fluorescent species within this volume and their diffusion constant can be determined. As a consequence, aggregation of the TMR-labeled 5′-cTAR molecules by the core peptides would be expected to decrease the number of fluorescent species.

Experiments were performed at two different ODN concentrations (100 and 300 nM). In the absence of peptide, the number of fluorescent TAR molecules in the excitation volume was fully consistent with the theoretical number of molecules expected from their concentration. By adding increasing concentrations of either core peptide, we found no change in the number of fluorescent species up to a peptide/ODN molar ratio of 1.4:1, indicating that no aggregation occurred under these conditions (Figure 1). In contrast, aggregation took place at higher ratios (2.8:1 and 5.6:1) as evidenced by the sharp drop in the number of fluorescent species. Changes in the diffusion constant were also observed at these higher ratios but were less prominent (data not shown), due to the dependence of the diffusion constant on the cubic root of the molecular mass of the diffusing species. As a consequence, we selected a peptide/ODN molar ratio of 1.4:1 to characterize the chaperone properties of the core peptides.

Destabilization of cTAR secondary structure by the core peptides

Since destabilization of nucleic acid secondary structures is a feature of the chaperone protein activity (29,32,37,41,44), we characterized the nucleic acid destabilizing activity of the core peptides, using cTAR labeled at its 5′ and 3′ ends by Rh6G and Dabcyl, respectively. In the absence of peptide, the proximity of cTAR ends induces a strong fluorescence quenching of Rh6G by the Dabcyl group (37), while the melting of the lower half of the cTAR stem increases the distance between the two dyes and thus, restores Rh6G fluorescence. Thus, the destabilizing activity of the core peptides can be evaluated from the ratio of the fluorescence intensity with or without the peptide. Addition of core peptides E and F at a peptide/cTAR molar ratio of 1.4:1 led to a 1.5-fold increased fluorescence intensity for Rh6G-5′-cTAR-3′-Dabcyl, while no change was observed with the D1 domain (Figure 2). Since full melting of cTAR would lead to a 20- to 25-fold increase of the Rh6G fluorescence, <10% of the Rh6G fluorescence was restored upon binding of the core peptides. No further increase was observed with either peptide at higher ratios (data not shown), possibly due to the fluorescence quenching related to formation of aggregates. Thus, in contrast to retroviral NC proteins (30–32,37), the core peptides can only slightly destabilize the secondary structure of cTAR, in the present experimental conditions.

Kinetics of cTAR/dTAR annealing in the presence of peptide E

To characterize the mechanism by which the core protein can activate the annealing of the complementary cTAR and dTAR sequences, we first used the peptide E, which was easy to synthesize in large quantities and was previously shown to mimic the core chaperone properties (20,28). The real-time annealing kinetics of cTAR to dTAR was monitored by mixing Rh6G-5′-cTAR-3′-Dabcyl with an excess of non-labeled dTAR, in the presence of peptide E added at a peptide/ODN ratio of 1.4:1. Formation of the 55-bp ED strongly increases the

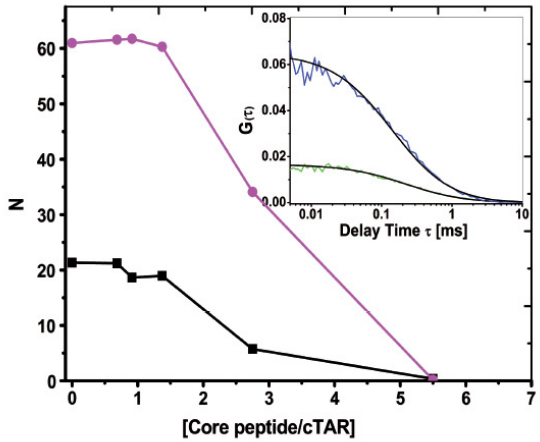


Figure 1. Evidence by fluorescence correlation spectroscopy of core peptide-induced aggregation of TMR-5′-cTAR. The average number of fluorescent species, *N*, for 100 nM (black squares) or 300 nM (magenta circles) of TMR-5′-cTAR within the excitation volume was obtained by fitting the autocorrelation curves to Equation (1). The curves are given with peptide E. Data with peptide F and D1 domain were very similar and were thus not represented. (Inset) Autocorrelation curves of 100 nM (blue line) and 300 nM TMR-5′-cTAR (green line) and their fits (black line) with Equation (1) in the presence of 140 nM and 420 nM of peptide E, respectively.

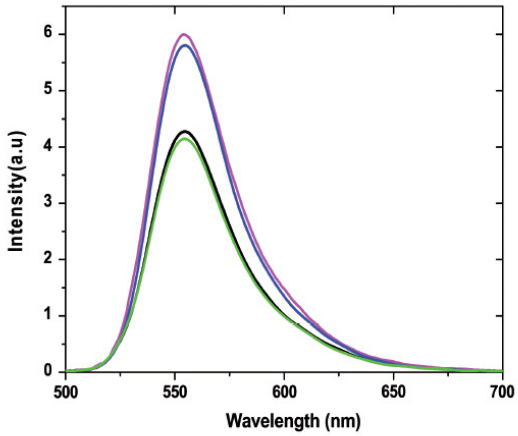


Figure 2. Destabilization of cTAR secondary structure by core peptides. Emission spectra of Rh6G-cTAR-Dabcyl (100 nM) were recorded in the absence (black) and presence of peptide E (magenta), peptide F (blue) and D1 Domain (green) at a molar ratio of 1.4 peptide per cTAR in 25 mM Tris, 30 mM NaCl and 0.2 mM $MgCl_2$, pH 7.5. Excitation wavelength was 520 nm.

interchromophore distance, leading to a full recovery of Rh6G emission (29). The same fluorescence plateau was obtained as in the absence of peptide, indicating that ED formation went to completion (Figure 3A). Peptide E strongly accelerated the annealing reaction, since it was complete in <30 min, while it needed more than 24 h in the absence of peptide (30). As shown by the nearly random distribution of residuals around the zero value (Figure 3B), the annealing kinetic traces could be adequately fitted using a biexponential function:

$$\mathrm{I}(t) = \mathrm{I_f} - (\mathrm{I_f} - \mathrm{I_0})\left(a\, e^{(-k_{\mathrm{obs1}}(t-t0))} - (1-a)\, e^{(-k_{\mathrm{obs2}}(t-t0))}\right) \quad (2)$$

where, t_0 being the dead time, $k_{\mathrm{obs1,2}}$ are the observed kinetic rate constants, a is the amplitude of the fast component, and I_0 and I_f are the fluorescence intensities of the stem–loop and the ED, respectively.

The dTAR concentration dependence indicates a saturation behavior for both k_{obs1} and k_{obs2} values (Figure 4), consistent with a two-step reaction, where a rate-limiting interconversion step is coupled to a much faster, preceding binding step, considered as a pre-equilibrium:

$$\mathrm{cTAR} + \mathrm{dTAR} \overset{K_M}{\rightleftharpoons} \mathrm{IC} \underset{k_b}{\overset{k_f}{\rightleftharpoons}} \mathrm{ED} \quad (3)$$

where IC corresponds to the intermediate complex. The dTAR concentration dependence of both k_{obs} values was fitted using (45):

$$k_{\mathrm{obs}} = \frac{k_f K_M[\mathrm{dTAR}]}{1 + K_M[\mathrm{dTAR}]} + k_b \quad (4)$$

Values of 4 $(\pm 2) \times 10^6\,\mathrm{M}^{-1}$ and 1.4 $(\pm 0.7) \times 10^6\,\mathrm{M}^{-1}$ were obtained for K_{M1} and K_{M2}, the equilibrium binding constants of the intermediate complexes associated with the fast and slow kinetic component, respectively. Moreover, values of 0.045 $(\pm 0.008)\,\mathrm{s}^{-1}$ and 0.016 $(\pm 0.002)\,\mathrm{s}^{-1}$ were obtained for the sum of the forward (k_{f1}, k_{f2}) and backward (k_{b1}, k_{b2}) interconversion rate constants associated to the two kinetic components. The values of k_{b1} and k_{b2}, given by the intercept, were very low ($<0.001\,\mathrm{s}^{-1}$) and could not be determined from the experimental data, indicating that peptide E is unable to dissociate the ED, and that the $0.045\,\mathrm{s}^{-1}$ and $0.016\,\mathrm{s}^{-1}$ values mainly correspond to k_{f1} and k_{f2}, respectively.

Noticeably, substitution of the Rh6G and Dabcyl dyes at the 5′- and 3′-ends of cTAR by TMR and Fl, respectively, had a marginal influence on the kinetic parameters (Table 1), indicating that the peptide E-promoted cTAR/dTAR annealing kinetics does not depend on the nature of the dyes.

In contrast to the k_{obs} values, the amplitudes of the two components (Table 1) were independent of dTAR concentration (data not shown), suggesting that peptide E-promoted annealing reaction involves two parallel pathways. To confirm this, we repeated the annealing reaction at a peptide/ODN molar ratio of 1:1. At this lower ratio, the amplitudes of the two components were also independent of dTAR concentration, but with a decreased value for the fast component as compared to the 1.4:1 ratio, in line with the hypothesis of two parallel pathways. Moreover, while the K_{M1} and K_{M2} values were three to four times lower than at the higher ratio, the interconversion rate constants were similar (Table 1), suggesting that peptide E mainly affects the pre-equilibrium step.

On the basis of our kinetic data, a reaction mechanism with two parallel kinetic pathways involving two different

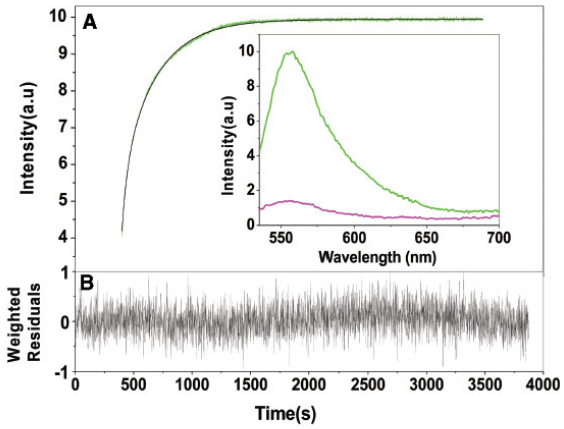


Figure 3. Kinetics of cTAR/dTAR annealing in the presence of peptide E. (**A**) Kinetic trace of 10 nM doubly labeled cTAR reacted with 300 nM dTAR in the presence of peptide E added at a peptide/ODN molar ratio of 1.4:1. The continuous line describes the best fit of the data according to equation (2) with $k_{obs1} = 1.8 \times 10^{-2}\,s^{-1}$, $k_{obs2} = 2.8 \times 10^{-3}\,s^{-1}$ and $a = 0.7$. Inset: emission spectra of the doubly labeled cTAR in the presence of peptide E added at a peptide/ODN ratio of 1.4:1 before (magenta) and after completion of the annealing reaction with dTAR (green). (**B**) Weighted residuals for the fit of the experimental data to Equation (2).

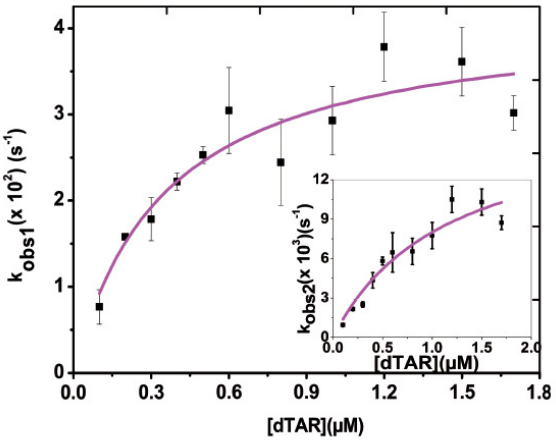
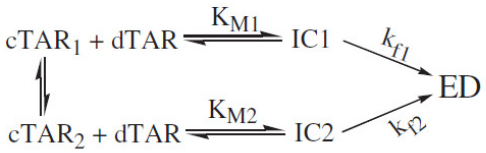


Figure 4. Kinetic parameters of cTAR/dTAR annealing in the presence of peptide E. The fast (k_{obs1}) and slow (k_{obs2}) (inset) components were determined in pseudo-first-order conditions from real-time kinetics, as described in Figure 3. The solid line corresponds to the fit of the data with Equation (4) and the K_M and k_f values given in Table 1.

cTAR/peptide E complexes ($cTAR_1$ and $cTAR_2$) can be proposed:

Scheme 1



where IC1 and IC2 are the intermediate complexes of the fast and slow pathways, respectively, and k_{f1} and k_{f2} are the corresponding interconversion rate constants. Further insights into nature of the two pathways were obtained from the temperature dependence of the k_{obs} values, using the Arrhenius equation:

$$k_i = A_i \exp^{(E_{a,i}/RT)} \quad (5)$$

where the rate constant k_i is given by k_{obsi}/[dTAR], A_i is the pre-exponential Arrhenius factor, $E_{a,i}$ is the activation energy, R is the universal gas constant and T is the temperature (in Kelvin).

Both reaction rates as well as the amplitude of the fast pathway increased with temperature (Figure 5). Positive enthalpy values for the transition state of 9.5(±0.6) kcal/mol and 24(±1) kcal/mol for the fast and slow pathways, respectively, were obtained. These values indicate that cTAR/dTAR annealing promoted by the core peptide involves pre-melting of ~2 and 5 bp, for the fast and slow pathways, respectively (46,47). Moreover, since the fast pathway represents 100% at 40°C (Figure 5), it is necessarily associated with ED formation and thus differs from the unproductive fast pathway observed for the cTAR/dTAR annealing reaction promoted by the HIV-1 NC (11-55) mutant (30).

Effects of oligonucleotide sequence and stability on the annealing kinetics promoted by the core peptide E

To further characterize how the core peptide influenced the annealing kinetics, we investigated the impact of the oligonucleotide sequence and stability on the reaction.

First, dTAR was substituted by the dTAR T-L mutant where the 6 nt of the loop were changed to T residues (Figure 6), thus preventing base-pairing with the cTAR loop. These nucleotide substitutions did not significantly change the kinetic parameters of the two pathways (Table 1) and ED formation (Figure 7), indicating that loop–loop interactions do not play a significant role in peptide E-promoted cTAR/dTAR annealing reaction.

To investigate the role of the cTAR stem in the annealing reaction, we used the cTAR1,2 derivative where bases complementary to the bulged bases at positions 49 and 52 have been introduced in order to stabilize the lower half of the stem (Figure 6). This mutant was previously shown to be about four times more stable than cTAR (44). Peptide E did not cause a fluorescence change of the doubly labeled cTAR1,2 (data not shown), indicating that peptide E was unable to destabilize the mutated stem. Moreover, the annealing of this mutant to dTAR in the presence of peptide E was extremely slow (Figure 7) with one kinetic pathway and a 10-fold decrease in the values of the pre-equilibrium constant, K_M, and the interconversion rate constant, k_f (Table 1). Thus, in a manner similar to NCp7 (30,35,48), peptide E nucleates cTAR/dTAR annealing probably through the cTAR and dTAR ends. This is confirmed by the increased value of the transition state enthalpy (30(±2) kcal/mol), indicating that the rate-limiting step of the peptide-promoted cTAR1,2/dTAR annealing involves melting of the stabilized cTAR1,2 stem (46). Moreover, the decreased

Table 1. Kinetic parameters of cTAR/dTAR and cTAR/TAR annealing in the presence of core peptides[a]

Oligo-nucleotide	Fluorophores	Complementary sequence	Peptide	Ratio of peptide/oligo	*a*	K_{M1} (M^{-1}) $\times 10^{-6}$	K_{M2} (M^{-1}) $\times 10^{-6}$	k_{f1} (s^{-1}) $\times 10^{2}$	k_{f2} (s^{-1}) $\times 10^{2}$	k_{ass1} ($M^{-1}s^{-1}$) $\times 10^{-4}$	k_{ass2} ($M^{-1}s^{-1}$) $\times 10^{-4}$
cTAR	3′Rh6G – 5′Dabcyl	dTAR	E	1.4	0.70 (±0.01)	4 (±2)	1.4 (±0.7)	4.5 (±0.9)	1.6 (±0.2)	18 (±2)	2.2 (±0.1)
cTAR	3′Fl – 5′TMR	dTAR	E	1.4	0.79 (±0.02)	1.5 (±0.9)	1.1 (±0.7)	6.4 (±2)	1.1 (±0.2)	10 (±2)	1.2 (±0.1)
cTAR	3′Rh6G – 5′Dabcyl	dTAR	E	1	0.60 (±0.01)	1.1 (±0.4)	0.55 (±0.1)	5.3 (±0.6)	1.7 (±0.2)	6 (±2)	1 (±0.02)
cTAR	3′Rh6G – 5′Dabcyl	dTAR-TL	E	1.4	0.60 (±0.01)	3.1 (±0.8)	1.3 (±0.6)	5 (±1)	1 (±0.1)	16 (±1)	1.3 (±0.06)
cTAR	3′Rh6G – 5′Dabcyl	TAR	E	1.4	0.35 (±0.02)	1.6 (±0.5)	1.2 (±0.8)	2 (±0.1)	0.13 (±0.01)	3 (±0.05)	0.16 (±0.008)
(14-39) cTAR	3′Rh6G – 5′Dabcyl	dTAR	E	1.4	–	0.33 (±0.02)		[b]	0.07 (±0.03)	[b]	0.030 (±0.003)
cTAR 1,2	3′Rh6G – 5′Dabcyl	dTAR	E	1.4	[b]	[b]	0.4 (±0.2)	[b]	0.02 (±0.01)	[b]	0.008 (±0.002)
cTAR	3′Rh6G – 5′Dabcyl	dTAR	F	1.4	0.85 (±0.02)	3 (±1)	2.7 (±0.7)	17 (±3)	4.2 (±0.6)	51 (±3)	11.3 (±0.4)
cTAR	3′Rh6G – 5′Dabcyl	dTAR	D1	1.4	0.80 (±0.02)	2 (±1)	1.5 (±0.6)	4.1 (±0.4)	0.6 (±0.05)	8 (±0.8)	1 (±0.03)

[a]The association equilibrium constants (K_{M1} for fast and K_{M2} for slow components) and interconversion rate constants (k_{f1} for fast and k_{f2} for slow component) are calculated from Equation (4), as described in Figure 4; k_{ass1} and k_{ass2} are given by: $k_{assi} = K_{Mi} \times k_{fi}$, while *a* corresponds to the amplitude of the fast component.
[b] A faster kinetic pathway with an amplitude of <5% is observed, but not reported.

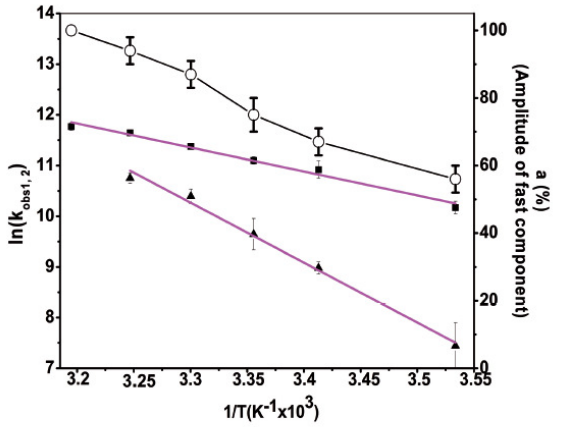


Figure 5. Temperature dependence of cTAR/dTAR annealing kinetics in the presence of peptide E at ratio 1.4:1. The reaction was performed with 10 nM doubly labeled cTAR and 300-nM non-labelled dTAR. The natural logarithm of the rate constant values for the fast (black squares) and slow (black triangles) components as well as the amplitude of the fast component (open circles) are indicated at six different temperatures. The solid magenta lines are the best fit to Equation (5) with $E_{a1} = 9.5(\pm 0.6)$ kcal/mol and $E_{a2} = 24(\pm 1)$ kcal/mol, for the fast and slow components, respectively. It was checked that no aggregation was associated with the changes in temperature.

stability of the cTAR1,2/dTAR intermediate as compared with the cTAR/dTAR one, most probably results from the base pair mismatches in the intermediate due to the additional bases in cTAR1,2 stem, further supporting the notion that the annealing reaction is nucleated through the stem ends.

Substitution of dTAR by TAR RNA also caused a significant decrease of the annealing reaction promoted by peptide E (Figure 7), due to a decrease in the amplitude of the fast component together with a 10-fold decrease of the interconversion rate, k_{f2} of the slow pathway (Table 1). This reduced annealing rate is likely caused by the higher stability of the TAR stem (29), suggesting that as with HIV-1 NCp7 (30,33,35), the cTAR/TAR annealing reaction promoted by peptide E is probably starting from the stem. This conclusion is further supported by the similar values of the transition state enthalpies (data not shown) and the equilibrium binding constants of the reaction intermediates (Table 1) in the cTAR/dTAR, cTAR/dTAR T-L and cTAR/TAR annealing reactions, indicating that the ICs in all these systems are stabilized by a similar number of base pairs.

To further confirm the critical role of the cTAR stem end, we investigated the annealing kinetics with the (14-39)cTAR derivative, also called mini-cTAR (49), in which the lower half of cTAR stem has been deleted (Figure 6). This cTAR mutant was previously shown to be about twice more stable than the entire cTAR (50). The annealing of (14–39)cTAR/dTAR in the presence of core peptide E was much slower than that of cTAR/dTAR (Figure 7), and exhibited only one kinetic pathway, confirming that the peptide E-promoted cTAR/dTAR annealing is nucleated through the cTAR ends.

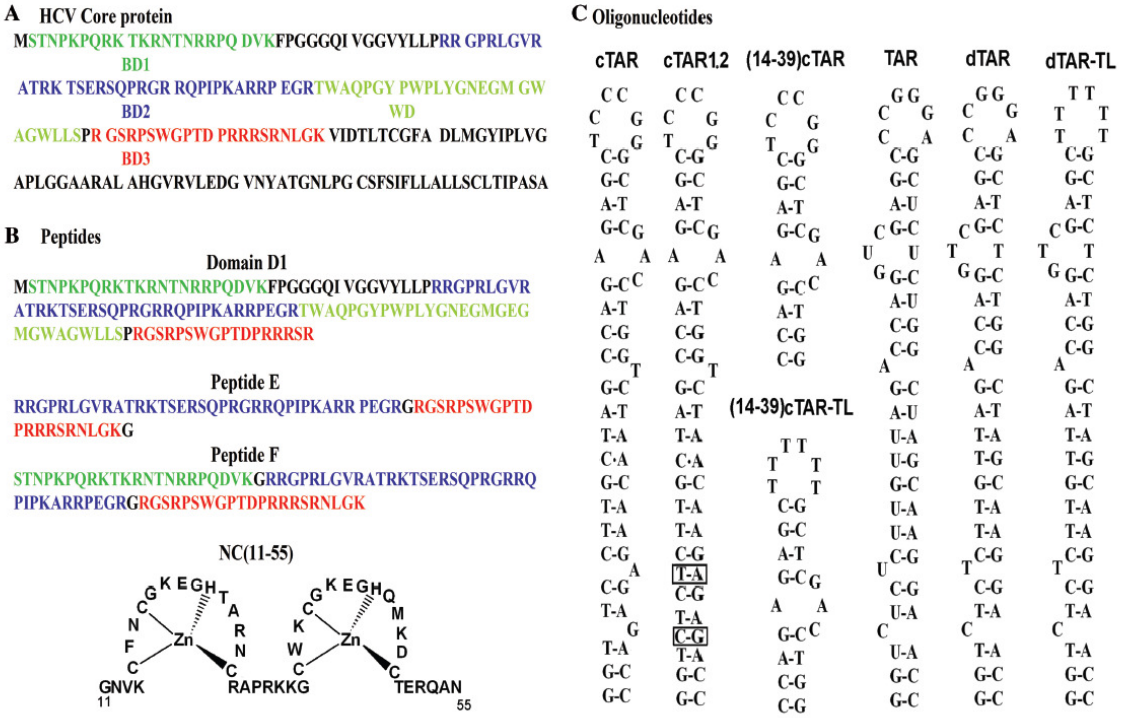


Figure 6. Sequences of the HCV core protein (**A**) and the peptides (**B**) and oligonucleotides (**C**) used in this study. The N-terminal D1 domain of the core protein (B) comprises three regions rich in basic residues (BD1 to BD3) and a region rich in tryptophan residues (WD). Peptides E and F correspond to the association of two (BD2 and BD3) and all three (BD1, BD2 and BD3) basic clusters, respectively. The NC(11–55) peptide corresponds to the zinc finger domain of the HIV-1 nucleocapsid protein. The cTAR and TAR RNA sequences (c) are from the HIV-1 MAL strain. The secondary structures of the oligonucleotides were predicted from the structure of TAR (36) and the mfold program (http//www.bioinfo.rpi.edu/applications/mfold/old/dna/form1.cgi).

Substitution of (14-39)cTAR by (14-39)cTAR-TL, in which the 6 nt of the (14-39)cTAR loop were substituted with T residues (Figure 6), further decreased the annealing reaction rate (Figure 7). This suggests that as for NCp7 (51), the (14-39)cTAR/dTAR annealing reaction promoted by peptide E may partially be nucleated through kissing loop intermediates. This conclusion was further supported by the value of the transition state enthalpy 15(± 5) kcal/mol, that perfectly matches the value found with the NCp7-promoted annealing of mini-cTAR (51).

Taken together, our data indicate that both kinetic pathways of the peptide E-promoted cTAR/dTAR annealing reaction are nucleated through the stems.

Promotion of cTAR/dTAR annealing by peptide F and D1 domain

Next, we investigated cTAR/dTAR annealing in the presence of peptide F and the D1 domain to determine whether they promote the annealing reaction through the same mechanism as peptide E (Figure 6B). The annealing kinetics in their presence was also biphasic with a hyperbolic dependence of $k_{obs1,2}$ values and a constant amplitude value of about 0.8, strongly suggesting that the three core peptides promote cTAR/dTAR annealing by the same mechanism. While the three peptides provide

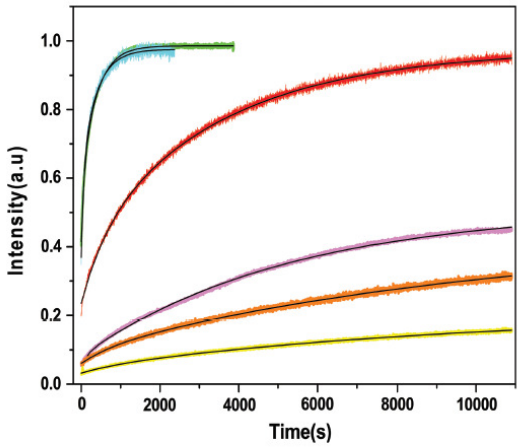


Figure 7. Kinetics of peptide E-promoted annealing of cTAR derivatives with TAR or dTAR derivatives. Kinetic traces of 10 nM doubly labeled cTAR derivatives with 300 nM nonlabeled TAR or dTAR derivatives [cTAR/dTAR (cyan), cTAR/dTAR-TL (green), cTAR/TAR RNA (red), (14-39)cTAR/dTAR (magenta), (14-39)cTAR-TL/dTAR (orange) and cTAR1,2/dTAR (yellow)]. Black lines correspond to fits of the kinetic curves with Equation (2) and the parameters in Table 1. Peptide E was added at a peptide/oligonucleotide ratio of 1.4:1 in all the cases and experimental conditions were as in Figure 3.

similar values for the pre-equilibrium binding constants, they differ in the interconversion rate constants. Indeed, the interconversion rates of the intermediates to ED were about 3- to 7-fold faster with peptide F than that with peptide E and the D1 domain (Table 1).

Interestingly, the temperature dependence of the k_{obs} values with peptide F and domain D1 yielded enthalpy values that were similar to those with peptide E (data not shown), confirming that the three peptides follow the same mechanism by nucleating ICs with a similar number of base pairs.

Comparison of the three peptides suggests that the third basic domain (BD3) favors the interconversion step while the additional sequences in D1, notably the tryptophan-rich domain, WD, have a counter effect.

Comparison of HCV core peptide E and HIV-1 NCp7 in promoting cTAR/dTAR annealing

Since both peptide E and NCp7 promote cTAR/dTAR annealing, this prompted us to compare their efficiencies in this activity. Because wild-type NCp7 strongly aggregates oligonucleotides (30,42,52,53), we used the truncated NC (11-55) peptide (Figure 6B), which mimics the activities of NCp7 without aggregation (30,38). Comparison of peptide E with NC(11-55) at several molar ratios revealed that 1 and 1.4 molecules of peptide E provided the same annealing activity as that of 4 and 6 NC (11-55) molecules (Figure 8A), respectively. This suggests that one peptide E molecule exhibits nearly the same chaperone activity as 4 molecules of NC(11-55). Furthermore, the kinetic traces in the presence of both NC(11-55) added at a ratio of 4 peptides per oligonucleotide and peptide E added at ratios of 1 and 1.4 peptides per oligonucleotide were similar to the kinetic traces obtained with ratios of 8 and 10 NC(11-55) per oligonucleotide, respectively (Figure 8B). This confirmed that one peptide E exhibits the same chaperone activity as four NC(11-55) and favored the notion that the two peptides can chaperone in concert the annealing of the two complementary ODNs.

DISCUSSION

The aim of the present study was to characterize the mechanism by which the basic domain of the HCV core chaperones the annealing of two canonical complementary stem–loop sequences, namely the HIV-1 cTAR and dTAR DNA, which have been extensively used to study the chaperone properties of retroviral NC proteins (30,32–35,54–56). This study was performed with the D1 core domain and two peptides that encompass the basic clusters critical for the chaperone properties of the core (20,28). To maintain the nonaggregating conditions required for the present fluorescent experiments, we used a limited peptide to oligonucleotide molar ratio of 1.4:1.

In contrast to retroviral NCs, such as HIV-1 NCp7 and MuLV NCp10 (29,32), the core peptides exhibit only a slight destabilizing activity, suggesting that this component plays a limited role in the core-promoted annealing reaction. Nevertheless, these peptides efficiently activate the cTAR/dTAR annealing reaction as shown by the up to three orders of magnitude difference in the values of the overall kinetic constants ($k_{ass} = K_M \times k_f$) in the presence of the core peptides (Table 1) as compared to the bimolecular rate constant ($85\,M^{-1}\,s^{-1}$) in their absence (30). The promotion of cTAR/dTAR annealing by the core peptides was found to start from the cTAR end and to proceed through two parallel kinetic pathways that include a fast pre-equilibrium intermediate and a rate-limiting conversion into the final ED. The fast and slow pathways likely differ by the number of base pairs (≈2 and 5, respectively) that should be melted in the original cTAR secondary structure to nucleate the ICs. Since thermally driven fraying of cTAR termini occurs spontaneously at room temperature (29,37,44), the fast pathway is probably associated with a cTAR species where the 3 bp of the terminal stem are melted, while the slow pathway may be associated with the fully structured oligonucleotide (Figure 9). Therefore, formation of IC1 in the fast pathway is thought to involve the melting of the 2 bp from the penultimate double-stranded segment while the 5 bp of both terminal and penultimate double

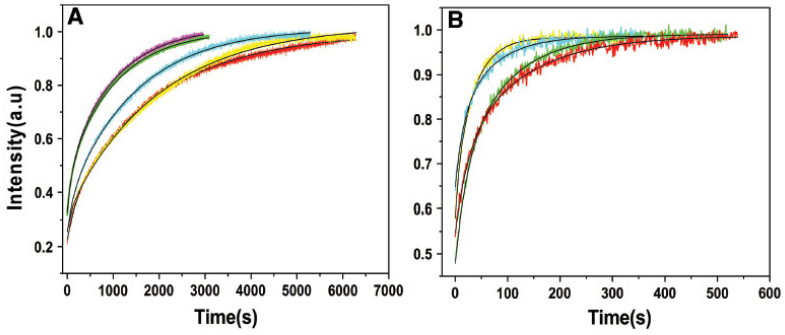


Figure 8. Comparison of the chaperoning properties of peptide E and HIV-1 NC(11-55). Kinetic traces were recorded with 10 nM doubly labeled cTAR and 100 nM nonlabeled dTAR. (**A**) Kinetic traces in the presence of NC (11-55) added at a peptide/oligonucleotide molar ratio of 4 (red), 5 (cyan) and 6 (magenta) or in the presence of peptide E added at a molar ratio of 1:1 (yellow) and 1.4:1 (green). (**B**) Kinetic traces with NC (11-55) added at a peptide/oligonucleotide molar ratio of 4:1 in the presence of peptide E, added at a peptide/oligonucleotide molar ratio of 1:1 (red) and 1.4:1 (cyan) compared with only NC (11–55) added at a molar ratio of 8:1 (green) and 10:1 (yellow) respectively. Black lines correspond to fits of the kinetic curves with Equation (2).

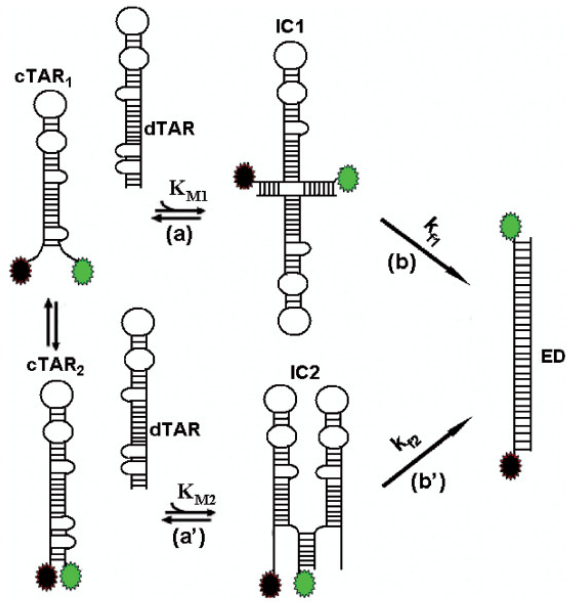


Figure 9. Proposed mechanism for core-promoted cTAR/dTAR annealing. The annealing reaction at sub-saturating protein concentrations is thought to involve two pathways that rely on thermal fraying of cTAR. This fraying leads to a fast equilibrium (microsecond range) between the fully closed cTAR species ($cTAR_2$) and a partially melted cTAR species ($cTAR_1$) (37). The upper pathway (steps a and b) and the lower pathway (steps a′ and b′) are associated with the fast and the slow kinetic components, respectively. In both pathways, the cTAR species nucleate an intermediate complex (IC) that subsequently, converts in a rate-limiting step to the final extended duplex (ED). In both pathways, the ICs are nucleated through the stem termini.

stranded DNA segments should melt to form IC2 in the slow pathway. As previously shown with HIV-1 NCp7 (35) and Tat (submitted), IC1 may result from the annealing of both 3′ and 5′ terminal strands of cTAR with the complementary terminal strands of dTAR and be stabilized by 12 intermolecular base pairs. Since the IC2 complex in the slow pathway is less stable than IC1, it likely involves a smaller number of intermolecular base pairing. A possible explanation is that IC2 results from an invasion mechanism, in which only one cTAR strand anneals with the complementary dTAR strand (54,56), allowing the formation of 5–7 intermolecular base pairs. Both IC1 and IC2 are further converted into the ED, which most probably relies on the conformational rearrangement and melting of the stable upper part of both TAR species. In this respect, the 3- to 4-fold faster interconversion rate of IC1 as compared to IC2 probably results from its larger number of intermolecular base pairs and/or more favorable conformation for the subsequent conversion into the ED.

Interestingly, the degree of coating of cTAR and dTAR by the peptides was found to strongly affect the equilibrium binding constant of the ICs but not their interconversion rate. Moreover, the amplitude of the fast component increases with the coating degree, suggesting that a single pathway can probably take place at a saturating concentration of core peptide, as in the case of retroviral proteins (35,49,51). In the proposed model reaction (Figure 9), the amplitude of the fast component is a complex function of the fast equilibrium between $cTAR_1$ and $cTAR_2$ species, their stability, their initial melting by the peptides and the fluorescence intensity of the intermediates. Accordingly, the amplitude of the fast component decreased, when either of the two partners, cTAR or dTAR, was substituted by a more stable one (~35% for cTAR/TAR and <5% cTAR1,2/dTAR) as compared to 70% for cTAR/dTAR. On the other hand, a decrease in stability of partners results in an increase of amplitude (100% at 40°C for cTAR/dTAR). Finally, the core chaperone mechanism was found to be similar for peptide E, peptide F and D1 domain, confirming that the basic clusters are the main determinants in this mechanism (20,28).

Comparison with the truncated HIV-1 NC(11-55) peptide revealed that one core peptide molecule exhibits the same chaperone activity as four NC(11-55) molecules in cTAR/dTAR annealing (Figure 8). A similar conclusion applies for cTAR/TAR RNA annealing with the full-length NCp7, since the K_M values of the ICs with peptide E (about $10^6\,M^{-1}$) are similar to the corresponding K_M value obtained with NCp7 added at a ratio of six proteins per oligonucleotide (35). This suggests that a 1.4 equivalent of peptide E could be as active as 6 NCp7 equivalents in the present experimental conditions. This is likely a consequence of the stronger 'nucleic acid aggregating' properties of the core peptides, in line with the efficient oligonucleotide aggregation observed by FCS (Figure 1). This propensity of the core peptides to neutralize the negatively charged nucleic acids and to promote their aggregation is probably related to the highly flexible nature of the core peptides (19,57) as compared to the folded NCp7 structure with two zinc fingers (58,59). In addition, NCp7 and the core peptides likely differ on how they promote the interconversion reaction. For cTAR/TAR, the interconversion rate constant, in the absence of any peptide is about $1.5 \times 10^{-4}\,s^{-1}$, while it is about 3-fold higher in the presence of NCp7, at any protein/oligonucleotide ratio (35). For the core peptides, the corresponding values for the fast and slow pathway are respectively, two and one order of magnitude higher than in the absence of protein. Thus, in contrast to NCp7, the core peptides strongly favor the interconversion reaction, which is thought to be rate-limited by a conformational change involving cooperative melting of a large fragment of one or both hairpin stems. Moreover, while NCp7 also promotes cTAR/TAR annealing through kissing loop intermediates when added at sub-saturating concentrations to ODNs (35), the core peptides mainly chaperone the annealing through the stems.

Taken together, our data suggest that though at sub-saturating conditions, the core peptides only marginally destabilize the secondary structure of oligonucleotides, they are more efficient than the truncated HIV-1 NC(11-55) peptide in nucleating IC formation through their nucleic acid aggregation properties and promoting the subsequent IC conversion into ED. Moreover, in

contrast to NC(11-55) and native NCp7 (35) which requires high peptide/ODN ratios for optimal activity, the core peptides already show strong nucleic acid annealing activity at low peptide/ODN ratios, suggesting that the core protein can exert its strong chaperone activity at low fractional saturation of nucleic acids. Importantly, since the core protein and NCp7 are structurally different proteins from two different virus families, the conserved nucleic acid chaperone properties suggest that they are required in a large panel of viruses, in line with the recently demonstrated conservation of RNA chaperoning in *Flaviviridae* core proteins (60). The potent chaperone properties of the core peptides in the TAR oligonucleotides model used in this study further suggest that the core protein is a nucleic chaperone with broad sequence specificity. These properties may be critical for genomic RNA dimerization in HCV replication and RNA packaging, as well as for facilitating recombination between various HCV genotypes and subtypes to increase viral variability (61–63).

FUNDING

The French National Agency for Research on AIDS and Viral Hepatitis (CSS n°4 to K.K.S.). Funding for open access charge: The funds from the French Agency against AIDS could partly support the publication charges.

Conflict of interest statement. None declared.

REFERENCES

1. Choo,Q.L., Kuo,G., Weiner,A.J., Overby,L.R., Bradley,D.W. and Houghton,M. (1989) Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science*, **244**, 359–362.
2. Recommendations from the National Institutes of Health consensus development conference statement: management of hepatitis C: 2002. *Hepatology*, **36**, 1039.
3. Lauer,G.M. and Walker,B.D. (2001) Hepatitis C virus infection. *N. Eng. J. Med.*, **345**, 41–52.
4. Moradpour,D., Penin,F. and Rice,C.M. (2007) Replication of hepatitis C virus. *Nat. Rev.*, **5**, 453–463.
5. Honda,M., Beard,M.R., Ping,L.H. and Lemon,S.M. (1999) A phylogenetically conserved stem-loop structure at the 5′ border of the internal ribosome entry site of hepatitis C virus is required for cap-independent viral translation. *J. Virol.*, **73**, 1165–1174.
6. Blight,K.J. and Rice,C.M. (1997) Secondary structure determination of the conserved 98-base sequence at the 3′ terminus of hepatitis C virus genome RNA. *Journal of virology*, **71**, 7345–7352.
7. Kolykhalov,A.A., Feinstone,S.M. and Rice,C.M. (1996) Identification of a highly conserved sequence element at the 3′ terminus of hepatitis C virus genome RNA. *Journal of virology*, **70**, 3363–3371.
8. Friebe,P. and Bartenschlager,R. (2002) Genetic analysis of sequences in the 3' nontranslated region of hepatitis C virus that are important for RNA replication. *Journal of virology*, **76**, 5326–5338.
9. Penin,F., Dubuisson,J., Rey,F.A., Moradpour,D. and Pawlotsky,J.M. (2004) Structural biology of hepatitis C virus. *Hepatology*, **39**, 5–19.
10. Ray,R.B. and Ray,R. (2001) Hepatitis C virus core protein: intriguing properties and functional relevance. *FEMS Microbiol. Lett.*, **202**, 149–156.
11. McLauchlan,J. (2000) Properties of the hepatitis C virus core protein: a structural protein that modulates cellular processes. *J. Viral Hepat.*, **7**, 2–14.
12. Hijikata,M., Kato,N., Mori,S., Ootsuyama,Y., Nakagawa,M., Sugimura,T., Ohkoshi,S., Kojima,H., Meguro,T., Taki,M. *et al.* (1990) Frequent detection of hepatitis C virus US strain in Japanese hemophiliacs. *Jpn J Cancer Res*, **81**, 1195–1197.
13. Lemberg,M.K. and Martoglio,B. (2002) Requirements for signal peptide peptidase-catalyzed intramembrane proteolysis. *Mol. Cell*, **10**, 735–744.
14. McLauchlan,J., Lemberg,M.K., Hope,G. and Martoglio,B. (2002) Intramembrane proteolysis promotes trafficking of hepatitis C virus core protein to lipid droplets. *EMBO J.*, **21**, 3980–3988.
15. Moradpour,D., Englert,C., Wakita,T. and Wands,J.R. (1996) Characterization of cell lines allowing tightly regulated expression of hepatitis C virus core protein. *Virology*, **222**, 51–63.
16. Barba,G., Harper,F., Harada,T., Kohara,M., Goulinet,S., Matsuura,Y., Eder,G., Schaff,Z., Chapman,M.J., Miyamura,T. *et al.* (1997) Hepatitis C virus core protein shows a cytoplasmic localization and associates to cellular lipid storage droplets. *Proc. Natl Acad. Sci. USA*, **94**, 1200–1205.
17. Hope,R.G. and McLauchlan,J. (2000) Sequence motifs required for lipid droplet association and protein stability are unique to the hepatitis C virus core protein. *J. Gen. Virol.*, **81**, 1913–1925.
18. Schwer,B., Ren,S., Pietschmann,T., Kartenbeck,J., Kaehlcke,K., Bartenschlager,R., Yen,T.S. and Ott,M. (2004) Targeting of hepatitis C virus core protein to mitochondria through a novel C-terminal localization motif. *J. Virol.*, **78**, 7958–7968.
19. Boulant,S., Vanbelle,C., Ebel,C., Penin,F. and Lavergne,J.P. (2005) Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features. *J. Virol.*, **79**, 11353–11365.
20. Ivanyi-Nagy,R., Kanevsky,I., Gabus,C., Lavergne,J.P., Ficheux,D., Penin,F., Fosse,P. and Darlix,J.L. (2006) Analysis of hepatitis C virus RNA dimerization and core-RNA interactions. *Nucleic Acids Res.*, **34**, 2618–2633.
21. Boulant,S., Montserret,R., Hope,R.G., Ratinier,M., Targett-Adams,P., Lavergne,J.P., Penin,F. and McLauchlan,J. (2006) Structural determinants that target the hepatitis C virus core protein to lipid droplets. *J. Biol. Chem.*, **281**, 22236–22247.
22. Shavinskaya,A., Boulant,S., Penin,F., McLauchlan,J. and Bartenschlager,R. (2007) The lipid droplet binding domain of hepatitis C virus core protein is a major determinant for efficient virus assembly. *J. Biol. Chem.*, **282**, 37158–37169.
23. Giannini,C. and Brechot,C. (2003) Hepatitis C virus biology. *Cell Death Differ.*, **10 (Suppl. 1)**, S27–S38.
24. Lerat,H., Honda,M., Beard,M.R., Loesch,K., Sun,J., Yang,Y., Okuda,M., Gosert,R., Xiao,S.Y., Weinman,S.A. *et al.* (2002) Steatosis and liver cancer in transgenic mice expressing the structural and nonstructural proteins of hepatitis C virus. *Gastroenterology*, **122**, 352–365.
25. Kunkel,M., Lorinczi,M., Rijnbrand,R., Lemon,S.M. and Watowich,S.J. (2001) Self-assembly of nucleocapsid-like particles from recombinant hepatitis C virus core protein. *J. Virol.*, **75**, 2119–2129.
26. Baumert,T.F., Ito,S., Wong,D.T. and Liang,T.J. (1998) Hepatitis C virus structural proteins assemble into viruslike particles in insect cells. *J. Virol.*, **72**, 3827–3836.
27. Shimoike,T., Mimori,S., Tani,H., Matsuura,Y. and Miyamura,T. (1999) Interaction of hepatitis C virus core protein with viral sense RNA and suppression of its translation. *J. Virol.*, **73**, 9718–9725.
28. Cristofari,G., Ivanyi-Nagy,R., Gabus,C., Boulant,S., Lavergne,J.P., Penin,F. and Darlix,J.L. (2004) The hepatitis C virus Core protein is a potent nucleic acid chaperone that directs dimerization of the viral (+) strand RNA in vitro. *Nucleic Acids Res.*, **32**, 2623–2631.
29. Bernacchi,S., Stoylov,S., Piemont,E., Ficheux,D., Roques,B.P., Darlix,J.L. and Mely,Y. (2002) HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. *J. Mol. Biol.*, **317**, 385–399.
30. Godet,J., de Rocquigny,H., Raja,C., Glasser,N., Ficheux,D., Darlix,J.L. and Mely,Y. (2006) During the early phase of HIV-1

DNA synthesis, nucleocapsid protein directs hybridization of the TAR complementary sequences via the ends of their double-stranded stem. *J. Mol. Biol.*, **356**, 1180–1192.
31. Ramalanjaona,N., de Rocquigny,H., Millet,A., Ficheux,D., Darlix,J.L. and Mely,Y. (2007) Investigating the mechanism of the nucleocapsid protein chaperoning of the second strand transfer during HIV-1 DNA synthesis. *J. Mol. Biol.*, **374**, 1041–1053.
32. Egele,C., Piemont,E., Didier,P., Ficheux,D., Roques,B., Darlix,J.L., de Rocquigny,H. and Mely,Y. (2007) The single-finger nucleocapsid protein of moloney murine leukemia virus binds and destabilizes the TAR sequences of HIV-1 but does not promote efficiently their annealing. *Biochemistry*, **46**, 14650–14662.
33. Liu,H.W., Cosa,G., Landes,C.F., Zeng,Y., Kovaleski,B.J., Mullen,D.G., Barany,G., Musier-Forsyth,K. and Barbara,P.F. (2005) Single-molecule FRET studies of important intermediates in the nucleocapsid-protein-chaperoned minus-strand transfer step in HIV-1 reverse transcription. *Biophys. J.*, **89**, 3470–3479.
34. Stewart-Maynard,K.M., Cruceanu,M., Wang,F., Vo,M.N., Gorelick,R.J., Williams,M.C., Rouzina,I. and Musier-Forsyth,K. (2008) Retroviral nucleocapsid proteins display nonequivalent levels of nucleic acid chaperone activity. *J. Virol.*, **82**, 10129–10142.
35. Vo,M.N., Barany,G., Rouzina,I. and Musier-Forsyth,K. (2009) HIV-1 nucleocapsid protein switches the pathway of transactivation response element RNA/DNA annealing from loop-loop "kissing" to "zipper". *J. Mol. Biol.*, **386**, 789–801.
36. Baudin,F., Marquet,R., Isel,C., Darlix,J.L., Ehresmann,B. and Ehresmann,C. (1993) Functional sites in the 5' region of human immunodeficiency virus type 1 RNA form defined structural domains. *J. Mol. Biol.*, **229**, 382–397.
37. Azoulay,J., Clamme,J.P., Darlix,J.L., Roques,B.P. and Mely,Y. (2003) Destabilization of the HIV-1 complementary sequence of TAR by the nucleocapsid protein through activation of conformational fluctuations. *J. Mol. Biol.*, **326**, 691–700.
38. Beltz,H., Clauss,C., Piemont,E., Ficheux,D., Gorelick,R.J., Roques,B., Gabus,C., Darlix,J.L., de Rocquigny,H. and Mely,Y. (2005) Structural determinants of HIV-1 nucleocapsid protein for cTAR DNA binding and destabilization, and correlation with inhibition of self-primed DNA synthesis. *J. Mol. Biol.*, **348**, 1113–1126.
39. Clamme,J.P., Azoulay,J. and Mely,Y. (2003) Monitoring of the formation and dissociation of polyethylenimine/DNA complexes by two photon fluorescence correlation spectroscopy. *Biophys. J.*, **84**, 1960–1968.
40. Thompson,N.L. (1991) *Fluorescence Correlation Spectroscopy, in Topics in Fluorescence Spectroscopy*. Plenum Publishers, NY.
41. Cristofari,G. and Darlix,J.L. (2002) The ubiquitous nature of RNA chaperone proteins. *Prog. Nucleic Acid Res. Mol. Biol.*, **72**, 223–268.
42. Stoylov,S.P., Vuilleumier,C., Stoylova,E., De Rocquigny,H., Roques,B.P., Gerard,D. and Mely,Y. (1997) Ordered aggregation of ribonucleic acids by the human immunodeficiency virus type 1 nucleocapsid protein. *Biopolymers*, **41**, 301–312.
43. Bloomfield,V.A., He,S., Li,A.Z. and Arscott,P.B. (1991) Light scattering studies on DNA condensation. *Biochem. Soc. Transact.*, **19**, 496.
44. Beltz,H., Azoulay,J., Bernacchi,S., Clamme,J.P., Ficheux,D., Roques,B., Darlix,J.L. and Mely,Y. (2003) Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the nucleocapsid protein NCp7. *J. Mol. Biol.*, **328**, 95–108.
45. Bombarda,E., Grell,E., Roques,B.P. and Mely,Y. (2007) Molecular mechanism of the Zn2+-induced folding of the distal CCHC finger motif of the HIV-1 nucleocapsid protein. *Biophys. J.*, **93**, 208–217.
46. Rouzina,I. and Bloomfield,V.A. (1999) Heat capacity effects on the melting of DNA. 1. General aspects. *Biophys. J.*, **77**, 3242–3251.
47. Cantor,C. and Schimmel,P. (1980) *Biophysical Chemistry Part 2: Techniques for the Study of Biological Structure and Function*. Academic Press, NY.
48. Cosa,G., Zeng,Y., Liu,H.W., Landes,C.F., Makarov,D.E., Musier-Forsyth,K. and Barbara,P.F. (2006) Evidence for non-two-state kinetics in the nucleocapsid protein chaperoned opening of DNA hairpins. *J. Phys. Chem.*, **110**, 2419–2426.
49. Vo,M.N., Barany,G., Rouzina,I. and Musier-Forsyth,K. (2009) Effect of Mg(2+) and Na(+) on the nucleic acid chaperone activity of HIV-1 nucleocapsid protein: implications for reverse transcription. *J. Mol. Biol.*, **386**, 773–788.
50. Beltz,H., Piemont,E., Schaub,E., Ficheux,D., Roques,B., Darlix,J.L. and Mely,Y. (2004) Role of the structure of the top half of HIV-1 cTAR DNA on the nucleic acid destabilizing activity of the nucleocapsid protein NCp7. *J. Mol. Biol.*, **338**, 711–723.
51. Vo,M.N., Barany,G., Rouzina,I. and Musier-Forsyth,K. (2006) Mechanistic studies of mini-TAR RNA/DNA annealing in the absence and presence of HIV-1 nucleocapsid protein. *J. Mol. Biol.*, **363**, 244–261.
52. Le Cam,E., Coulaud,D., Delain,E., Petitjean,P., Roques,B.P., Gerard,D., Stoylova,E., Vuilleumier,C., Stoylov,S.P. and Mely,Y. (1998) Properties and growth mechanism of the ordered aggregation of a model RNA by the HIV-1 nucleocapsid protein: an electron microscopy investigation. *Biopolymers*, **45**, 217–229.
53. Grohmann,D., Godet,J., Mely,Y., Darlix,J.L. and Restle,T. (2008) HIV-1 nucleocapsid traps reverse transcriptase on nucleic acid substrates. *Biochemistry*, **47**, 12230–12240.
54. Chen,Y., Balakrishnan,M., Roques,B.P. and Bambara,R.A. (2003) Steps of the acceptor invasion mechanism for HIV-1 minus strand strong stop transfer. *J.Biol. Chem.*, **278**, 38368–38375.
55. Kim,J.K., Palaniappan,C., Wu,W., Fay,P.J. and Bambara,R.A. (1997) Evidence for a unique mechanism of strand transfer from the transactivation response region of HIV-1. *J. Biol. Chem.*, **272**, 16769–16777.
56. Roda,R.H., Balakrishnan,M., Kim,J.K., Roques,B.P., Fay,P.J. and Bambara,R.A. (2002) Strand transfer occurs in retroviruses by a pause-initiated two-step mechanism. *J. Biol. Chem.*, **277**, 46900–46911.
57. Ivanyi-Nagy,R., Davidovic,L., Khandjian,E.W. and Darlix,J.L. (2005) Disordered RNA chaperone proteins: from functions to disease. *Cell. Mol. Life Sci.*, **62**, 1409–1417.
58. Morellet,N., de Rocquigny,H., Mely,Y., Jullian,N., Demene,H., Ottmann,M., Gerard,D., Darlix,J.L., Fournie-Zaluski,M.C. and Roques,B.P. (1994) Conformational behaviour of the active and inactive forms of the nucleocapsid NCp7 of HIV-1 studied by 1H NMR. *J. Mol. Biol.*, **235**, 287–301.
59. Summers,M.F., Henderson,L.E., Chance,M.R., Bess,J.W. Jr, South,T.L., Blake,P.R., Sagi,I., Perez-Alvarado,G., Sowder,R.C. III, Hare,D.R. *et al.* (1992) Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1. *Protein Sci.*, **1**, 563–574.
60. Ivanyi-Nagy,R., Lavergne,J.P., Gabus,C., Ficheux,D. and Darlix,J.L. (2008) RNA chaperoning and intrinsic disorder in the core proteins of Flaviviridae. *Nucleic Acids Res.*, **36**, 712–725.
61. Colina,R., Casane,D., Vasquez,S., Garcia-Aguirre,L., Chunga,A., Romero,H., Khan,B. and Cristina,J. (2004) Evidence of intratypic recombination in natural populations of hepatitis C virus. *J. Gen. Virol.*, **85**, 31–37.
62. Kalinina,O., Norder,H. and Magnius,L.O. (2004) Full-length open reading frame of a recombinant hepatitis C virus strain from St Petersburg: proposed mechanism for its formation. *J. Gen. Virol.*, **85**, 1853–1857.
63. Kalinina,O., Norder,H., Mukomolov,S. and Magnius,L.O. (2002) A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. *J. Virol.*, **76**, 4034–4043.

Chapter 3: Kinetic investigation of HCV 3' X-tail region annealing in the presence of core peptides (Peptide E, F and Domain D1)

Next, we characterized the mechanism and structural determinants of the core-promoted annealing of HCV nucleotide sequences (relevantly taken from first 55 nt region of 3'X-tail) in the absence and presence of the HCV core peptides. To this end, we investigated the annealing kinetics of the fluorescently labelled 16-nt palindromic sequence (DLS) and 27-nt Stem Loop II (SL2) with their complementary non-labelled cDLS and cSL2 sequences, respectively. In the absence of protein, annealing of the abovementioned oligonucleotides was very slow. Variations in complementary DLS or SL2 concentrations indicated saturation behaviour for both fast, k_{obs1}, and slow, k_{obs2}, kinetic rate constants. This behaviour is consistent with a two-step reaction mechanism in which a fast pre-equilibrium intermediate precedes the formation of the final stable ED through a monomolecular reaction. Additionally, different mutants of DLS/cDLS and SL2/cSL2 suggested that the annealing process for both systems is nucleated by the kissing of the complementary loops.

The three peptides were found to poorly destabilise the hairpin-loop RNA structures, but to strongly promote their annealing through kissing interactions between the loops with the same two-step reaction mechanism as in the absence of protein. Thus, core peptides strongly promote annealing of HCV oligonucleotide sequences, but without changing the pathways and reaction mechanism.

On the basis of our kinetic data, a core promoted-reaction mechanism with a single kinetic pathway involving intermediate complexes of ~12 bp for SL2/cSL2 and ~8 bp for DLS/cDLS was proposed.

Since HCV RNA likely adopts different conformations during the viral life cycle, the demonstrated ability of the core protein to facilitate the interconversion between different RNA structures could explain how this protein can regulate the transitions between translation and replication as well between replication and packaging.

Chaperone activity of the Hepatitis C Virus core protein on the 3' X-tail region of the virus genome

Kamal kant Sharma[1], Hugues de Rocquigny[1], Jean Luc Darlix[2], Jean-Pierre Lavergne[3], François Pénin[3], Jean-Marc Lessinger[1] and Yves Mély[1]

[1]*Laboratorie de Biophotonique et Pharmacologie, UMR 7213 CNRS, Faculté de Pharmacie, Université de Strasbourg, 74, Route du Rhin, 67401, Illkirch, Cedex, France*

[2] *LaboRétro, Unité deVirologie Humaine INSERM, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364, Lyon, France*

[3] *Institut de biologie et Chimie des Proteines, UMR 5086 CNRS, Université de Lyon, IFR 128 BioSciences Gerland Lyon Sud, 7 Passage du Vercors, 69367 Lyon, France*

Abstract

RNA chaperone proteins are essential partners of RNA in living organisms and viruses as they facilitate nucleic acid rearrangements in their most thermodynamically stable conformational states. In Hepatitis C Virus (HCV), the core protein that forms the nucleocapsid with the genomic RNA has been shown to exhibit nucleic acid chaperone properties. The core protein (179 amino acids) consists of a hydrophilic N-terminal domain (Domain D1) and a hydrophobic C-terminal domain (Domain D2). This protein plays key roles in the viral life cycle, being involved in cell proliferation, differentiation, RNA packaging and nucleocapsid formation. To better understand the molecular chaperone properties of the core protein, we characterized the annealing mechanism of the HCV 3' X-tail region in the absence and presence of core peptides. To this end, two oligonucleotide sequences (DLS and SL2), from the first 55nt of X-tail, were reacted to their respective complementary sequences (cDLS and cSL2, respectively). The core peptides were found to enhance annealing kinetics of both systems, DLS/cDLS and SL2/cSL2, by at least one order of magnitude. Interestingly, both systems, in absence as well as in presence of core peptides, showed significant changes in annealing kinetics with modification in their loop region, suggesting that their annealing is initiated through the interactions between complementary loops. Finally, a single kinetic pathway was identified with a fast pre-equilibrium kissing-loop intermediate that then slowly converts into the final extended duplex. Our data on the relevant nucleic acid sequences of HCV gives further credit on the chaperone activities of the core protein. Since, HCV RNA likely adopts different conformations during the viral life cycle, the demonstrated ability of the core protein to facilitate the interconversion between different RNA structures may explain how this protein can regulate the transitions between translation and replication as well between replication and packaging.

INTRODUCTION

Hepatitis C virus (HCV) is a major cause of chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma (1). Recombinant alpha interferon, either alone or in combination with ribavirin, is widely used to treat HCV infection but with an overall efficacy of less than 50% (2-7). The urgent development of more effective drugs has long been hampered by the lack of cell culture systems that are fully permissive for viral replication and by the absence of small-animal models of the disease. However, recently, the development of sub-genomic HCV replicons, capable of replication in human hepatoma (Huh 7) cells, transgenic mice expressing HCV proteins and the successful repopulation of SCID/Alb-uPA mice with human hepatocytes has facilitated drug discovery efforts and paved new pathways to characterize the molecular processes involved in replication of the viral RNA (8-12).

HCV possesses a single stranded RNA molecule of ~9.6 kb length with positive-sense polarity (13). The genomic RNA plays complex, temporally and spatially regulated roles throughout the virus life cycle, by serving both as a template for minus-strand RNA synthesis and as an mRNA directing the translation of viral proteins, and lastly, by specifically packaging itself into newly made progeny virions. The coding part of the RNA is flanked by two non-translated, but highly conserved and structured regions, named 5' UTR and 3' UTR, of 341 and 230 nucleotides, respectively. These non-translated RNA sequences contain *cis*-acting elements that are crucial for RNA replication and translation (Ito and Lai et al . 1997; song et al. 2006). Between these flanked terminal sequences a single ~3011-codon-long open reading frame (ORF) encodes at least 10 proteins, including (in 5'-to-3' direction) four structural proteins (core, E1, E2 and p7) (14).

The 5' UTR folds into a complex, four-domain structure that contains an internal ribosome entry site (IRES), which directs the cap-independent translation of the ORF, as well as upstream signals required for viral RNA synthesis (15,16). The 3' UTR also contains three structurally distinct domains: (1) a highly conserved, 98 nucleotide long 3' X-tail, that putatively forms three stem-loop structures (designated as SL1, SL2 and SL3, in the 3'-to-5' direction), (2) an upstream variable part with VSL1 and VSL2 stem-loops, and (3) a poly (U/UC) tract (17,18). Both X-tail region and a minimal poly(U/UC) tract of about 40 nt are essential for RNA replication (19-23), while the variable region plays some role in RNA replication and *in vivo* infectivity (19-23). The 3'X- tail is also thought to facilitate the translation of viral proteins, directed by the viral IRES (24,25). Two alternate conformations are predicted for the 3'X-tail region (26) that includes a sequence of 55 nucleotides, which is fully conserved among different HCV strains (27). Interestingly, this sequence contains a 16-

nt palindrome (DLS), suggested to be responsible for the homodimerisation of the HCV RNA (26,27).

The HCV core protein is a structural and basic protein that forms the viral nucleocapsid by binding to genomic RNA (28-30). The mature core protein has two domains, (I) a 117 amino acids long basic N-terminal RNA binding domain D1 with three highly basic clusters [BD1-BD3], and (II) a hydrophobic C-terminal domain D2 involved in targeting the protein to lipid droplets (26). The three basic amino acid clusters (BD1-BD3) of the D1 domain are responsible for the nucleic acid chaperone activity (26,31-33) of the HCV core protein. Thus, by analogy to other RNA chaperone proteins (34-39), the core protein can facilitate the correct folding of RNA molecules by preventing their misfolding or by resolving misfolded RNA structures (40,41). Through its chaperone properties, the core as well as peptides consisting of the three basic domains (peptide F) or only of two of them (peptide E), were shown to promote the annealing of complementary sequences (33), strand exchanges as well as the dimerization of the 3'-UTR sequence of the HCV genome (26,32).

Here, we further characterized the RNA chaperone properties of the core protein by investigating the mechanism and structural determinants for the annealing of the 16-nt palindromic sequence (DLS) and the 27-nt Stem Loop II (SL2) sequences, taken from the first 55nt region of the 3'X-tail, with their complementary sequences in the absence and presence of the core domain D1 or peptides E and F. By monitoring the real-time annealing kinetics of native and mutated fluorescently labelled DLS and SL2 sequences with their respective complementary sequences, the three peptides were found to strongly promote the complementary annealing reactions through kissing interactions between the loops. Moreover, to site-selectively monitor the effect of the peptides on the oligonucleotides, we introduced 2-aminopurine (2-Ap), an environment-sensitive fluorescent nucleoside analogue of adenine (42) at several positions of the SLS and SL2 sequences. Using this approach, peptide E was found to preferably bind to the upper region of both oligonucleotides and constrain both the overall flexibility of the loops and the local mobility of the bases.

(a) HCV Core protein

MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGG
VYLLPRRGPRLGVRATRKTSERSQPRGRRQPIPK
ARRPEGRTWAQPGYPWPLYGNEGMGWAGWLLS
PRGSRPSWGPTDPRRRSRNLGKVIDTLTCGFADL
MGYIPLVGAPLGGAARALAHGVRVLEDGVNYAT
GNLPGCSFSIFLLALLSCLTIPASA

(b) Peptides

Peptide E

RRGPRLGVRATRKTSERSQPRGRRQPIPKARRPE
GRGRGSRPSWGPTDPRRRSRNLGKG

Peptide F

STNPKPQRKTKRNTNRRPQDVKGRRGPRLGVRA
TRKTSERSQPRGRRQPIPKARRPEGRGRGSRPSW
GPTDPRRRSRNLGK

Domain D1

MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGG
VYLLPRRGPRLGVRATRKTSERSQPRGRRQPIPK
ARRPEGRTWAQPGYPWPLYGNEGMGEGMGWA
GWLLSPRGSRPSWGPTDPRRRSR

(c) Oligonucleotides

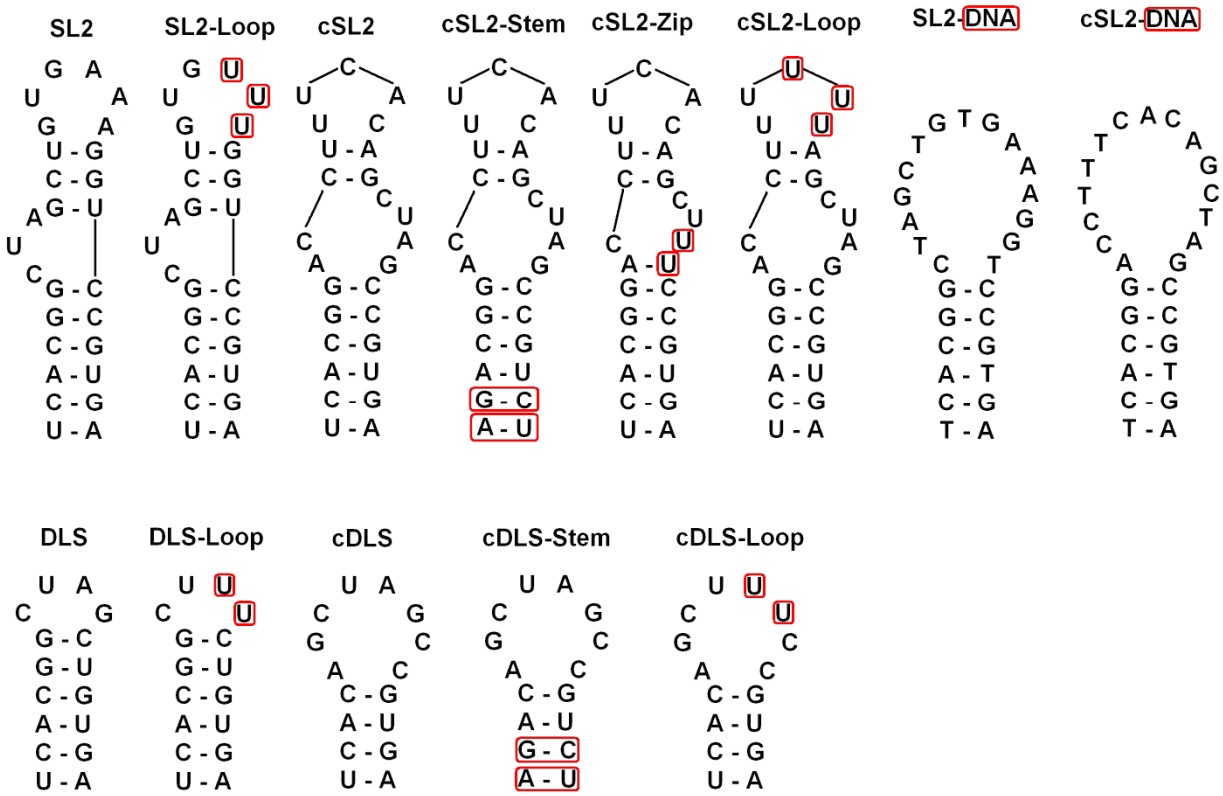


Figure 1: Structure of the HCV core protein (a), the peptides (b) and oligonucleotides (c) used in this study. The ~117 amino acid long and poorly folded N-terminal D1 domain of the core protein (a) comprises three regions rich in basic residues [BD1 (green), BD2 (blue) and BD3 (red)] and a tryptophan rich domain (yellow) while the C-terminal, ~52- amino acid long, D2 domain is hydrophobic and binds to membranes of the endoplasmic reticulum. The ODN sequences are from the 3' X-tail of the HCV genome and their secondary structures were predicted by using the mfold program (http://mfold.rna.albany.edu/?q=mfold).

Materials and Methods

Oligonucleotides and core peptides

The unlabelled and labelled ODNs were synthesized by IBA GmbH Nucleic Acids product Supply (Gottingen, Germany). In the case of the doubly labelled ODNs, the 5'terminus was labelled with carboxytetramethylrhodamine (TMR) or ethyl 2-[3-(ethylamino)-6-ethylimino-2,7-dimethylxanthen-9-yl]benzoate hydrochloride (Rh6G) via an amino-linker with a six carbon spacer arm, while the 3'terminus was labelled with either 4-(4'-dimethylaminophenylazo) benzoic acid (Dabcyl) or 5(and 6)-carboxyfluorescein (Fl) using a special solid support with the dye already attached. ODNs were purified by the manufacturer by reverse-phase HPLC and polyacrylamide gel electrophoresis.

Peptides E and F were synthesized by solid phase peptide synthesis on a 433A synthesizer (ABI, Foster City, CA) as described (26). Purification by HPLC was carried out on a C8 column (Uptisphere 300A, 5µm; 250X10, Interchim, France) in 0.05% TFA with a linear gradient of 10 to 70% of acetonitrile for 30 min. The peptide purity and molecular weight (6754 for peptide E and 9444 for Peptide F) were checked by LC/MS. An absorption coefficient of 5700 $M^{-1}cm^{-1}$ at 280 nm was used to determine their concentration.

The D1 domain consisting of amino acids 1 to 117 of the core protein fused to a C-terminal 6xHis tag was prepared as previously described (43).

Experiments were performed in 50 mM Hepes-HCl (pH 7.5), 30 mM NaCl, and 0.2 mM $MgCl_2$ at 20°C.

Steady-state fluorescence spectroscopy measurements

Emission spectra and kinetic traces were recorded with Fluorolog and FluroMax spectrofluorimeters (Jobin Yvon Instruments, S.A. Inc.) equipped with a temperature-controlled cell compartment. All fluorescence intensities were corrected for buffer emission and lamp fluctuations. Quantum yields of 2Ap-labeled oligonucleotides were calculated using free 2-Ap riboside as a reference (0.68, Ward1969) and an excitation wavelength of 315nm.

For the annealing assays, the oligonucleotides were first heat labelled at 80°C for 2 min and then cooled on ice for 2 min. Real-time kinetic measurements were performed in pseudo first-order conditions by using concentrations of unlabelled cDLS and cSL2 at least tenfold higher than the concentration of the labelled DLS and SL2 sequences. Excitation and emission wavelengths were 520 nm and 550 nm, respectively, to monitor the Rh6G fluorescence. The corresponding wavelengths were 480 nm and 520 nm, respectively to monitor the Fl fluorescence. All reported concentrations correspond to those after mixing. To avoid high

local concentrations during mixing, both reactant mixtures were prepared of the same volume. Peptides were added to each reactant separately in peptide to oligonucleotide ratio of 0.5:1 and 1:1 for DLS and SL2 systems, respectively, and then the reaction was initiated by mixing the peptide-coated ODNs together. The kinetics was fast enough to monitor the fluorescence intensities continuously without photo-bleaching. The apparent rate constants k_{obs} and the amplitudes were determined from the kinetic traces by including a dead-time correction t_0 to take into account the delay between the mixing of reactants and the start of the measurements. All fitting procedures were carried out with Origin™ 7.5 software based on non-linear, least-square methods and the Levenberg-Marquardt algorithm.

The temperature dependence of the annealing kinetics was carried out with 20 nM doubly labelled DLS derivatives/10nM doubly labelled SL2 derivatives and 500-4000 nM non-labelled cDLS derivatives/200nM non-labelled cSL2 derivatives at different temperatures (15°C, 20°C, 25°C, 30°C, 35°C and 40°C) in the presence of either peptide E, peptide F and Domain D1. Same experiments were carried in absence of peptide, but only with SL2/cSL2 system.

Time-resolved fluorescence measurements

Time-resolved fluorescence measurements were performed with the time-correlated, single-photon counting technique. Excitation pulses were generated by a pulse-picked frequency-tripled Ti-sapphire laser (Tsunami, Spectra Physics) pumped by a Millenia X laser (Spectra Physics) (44,45). Excitation wavelength was set at 315 nm, with a repetition rate of 4 MHz. The fluorescence emission was collected through a polarizer set at magic angle and a 16 mm band-pass monochromator (Jobin Yvon) at 370 nm. The single-photon events were detected with a micro-channel plate photomultiplier (Hamamatsu) coupled to a pulse pre-amplifier HFAC (Becker-Hickl) and recorded on a SPC-130 board (Becker-Hickl). The instrumental response function (IRF) recorded using a polished aluminium reflector had a full-width at half-maximum of 40 ps. The mean lifetime $<\tau>$ was calculated from the individual fluorescence lifetimes (τ_i) and their relative amplitudes (α_i) according to $<\tau> = \Sigma\ \alpha_i\ \tau_i$. The population, α_0, of dark species of 2-AP was calculated by: $\alpha_0 = 1 - \tau_{free} / (\tau_{sample} \times R_m)$, where τ_{free} is the lifetime of free 2-Ap, τ_{sample} is the measured mean lifetime of 2-Ap within the ODN and R_m is the ratio of their corresponding quantum yields. The remaining amplitudes, α_{ic} were recalculated from the measured amplitudes according to $\alpha_{ic} = \alpha_i / (1- \alpha_0)$.

Time-resolved intensity data were treated according to the maximum entropy method (Pulse 5 software) (46,47) or according to a non-linear least-square analysis using a homemade

software (kindly provided by G. Krishnamoorthy). In all cases, the χ^2 values were close to 1 and the weighted residuals as well as their autocorrelation were distributed randomly around 0, indicating an optimal fit.

Results

Kinetics of DLS/cDLS and SL2/cSL2 annealing in the absence of core

The real-time annealing kinetics of DLS and SL2 with their complementary cDLS and cSL2 sequences were investigated by mixing excess of non-labelled cDLS and cSL2 to TMR-5'-DLS-3'-Fl and TMR-5'-SL2-3'-Fl, respectively. Formation of the 16-bp and 27-bp ED strongly increases the interchromophore distance, leading to the recovery of Fl emission (Bernacchi, S et al. 2002). To determine the fluorescence intensity of the fully stretched ED, we heated the complementary sequences in pseudo-first order conditions at 80°C for 2 min, followed by slow cooling.

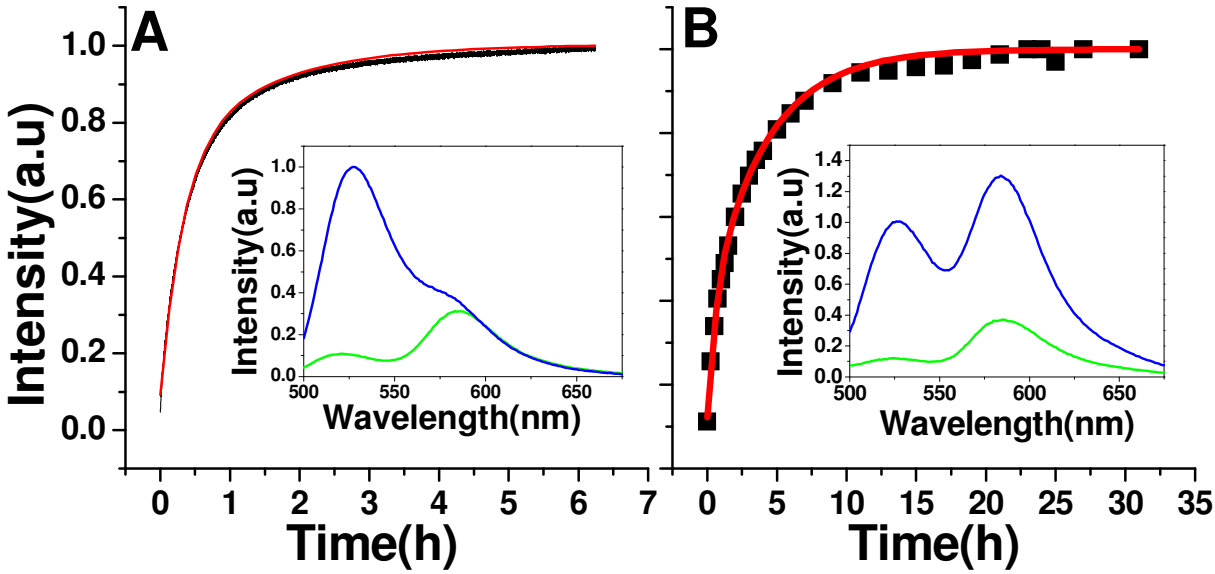


Figure 2: Kinetics of SL2/cSL2 and DLS/cDLS annealing. A) Kinetic trace of 10 nM doubly labelled SL2 with 500 nM of unlabelled cSL2. Inset A: Emission spectra of the doubly labelled SL2 before (green) and after (blue) completion of the annealing reaction with cSL2. B) Kinetic trace of 20 nM doubly labelled DLS with 1 µM of unlabelled cDLS. Inset B: Emission spectra of the doubly labelled DLS before (green) and after (blue) completion of the annealing reaction with cDLS. Experiments were performed in 50 mM Hepes (pH 7.5), 30 mM NaCl, 0.2 mM $MgCl_2$ at 20°C. Excitation and emission wavelengths were 480 nm and 520 nm, respectively, to monitor the fluorescein emission. The continuous red lines correspond to a fit of the experiment al curves with eqn. 1.

A difference of 12-times and 20-times was observed for DLS/cDLS and SL2/cSL2, respectively, between initial closed and final stretched ED fluorescence intensities (Figure 2). The annealing kinetics traces could be adequately fitted using a biexponential function:

$$I(t) = I_f - (I_f - I_0)\left(a\, e^{(-k_{obs1}(t-t0))} - (1-a)\, e^{(-k_{obs2}(t-t0))}\right) \qquad (1)$$

where $I(t)$ is the actual fluorescence intensity at 520 nm, k_{obs1} and k_{obs2} are the apparent pseudo-first order rate constants, a is the relative amplitude of the fast component, and t_0 is the dead time. I_0 and I_f stands for the fluorescence intensities of TMR-5'-DLS-3'-Fl or TMR-5'-SL2-3'-Fl in their free state and in the final ED, respectively.

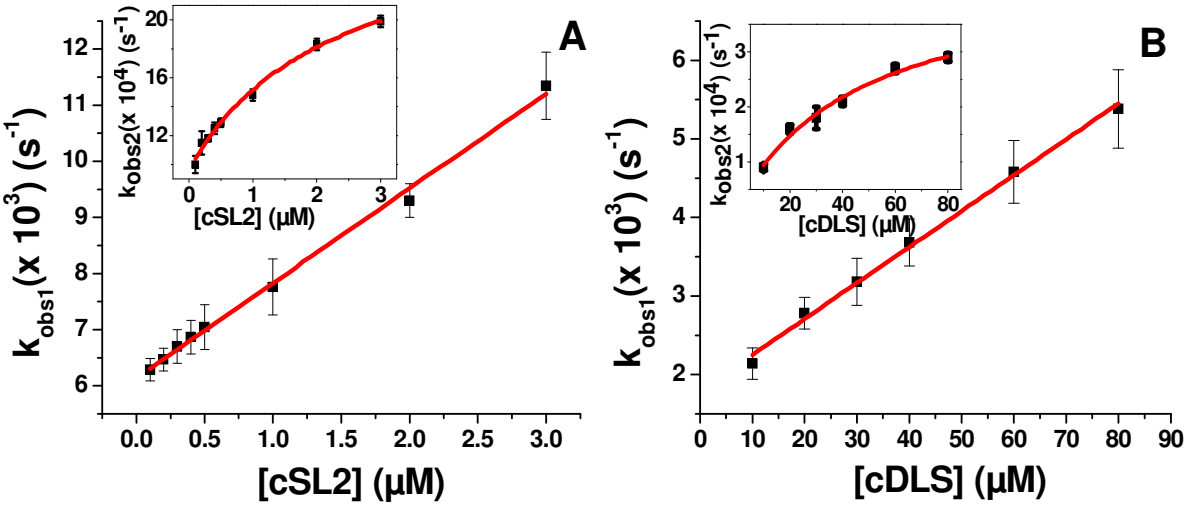
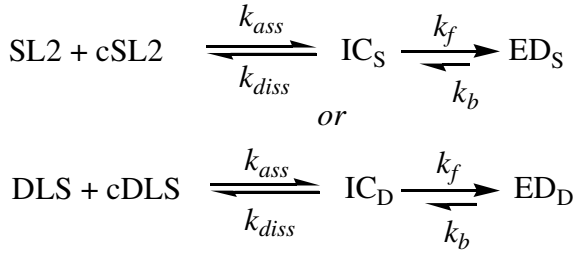


Figure 3: Kinetic parameters of SL2/cSL2 (A) and DLS/cDLS (B) annealing. The fast (k_{obs1}) [A & B] and slow (k_{obs2}) [Inset A & B] components were determined in pseudo-first-order conditions from real-time kinetics, as described in figure 2. The solid lines correspond to the fit of the data with Equation 2 and Equation 3 (for Inset data). Corresponding values are given in table 1 and table 2.

Interestingly, the fast component, k_{obs1}, varies linearly with cDLS or cSL2 concentrations (Figure 3A and B), while k_{obs2} shows a hyperbolic dependence on the complementary sequence concentration (Figure 3 A and B, insets). This behaviour is consistent with a two-step reaction, as described in scheme 1:



Scheme 1

where a fast pre-equilibrium intermediate, IC, precedes the formation of the final stable ED through a monomolecular reaction (39). Formation of IC is governed by the second order association rate constant, k_{ass}, and the first order dissociation rate constant, k_{diss}, while the interconversion of IC into ED is governed by the forward and backward interconversion rate constants, k_f and k_b. The hyperbolic dependence of k_{obs2} on the concentration of non-labeled complementary oligonucleotide may be associated to IC accumulation as a consequence of its slow inter-conversion into ED which is likely the rate-limiting step of the annealing pathway. If the pre-equilibrium is rapidly reached (k_{ass}.[cDLS-or-cSL2] + k_{diss} >> k_f), scheme 1 predicts a linear variation of k_{obs1}, which can thus be described by :

$$k_{obs1} = k_{ass}[cDLS - or - cSL2] + k_{diss} \qquad (2)$$

On the other hand, the hyperbolic variation of k_{obs2} on the complementary oligonucleotide concentration can be described by:

$$k_{obs2} = \frac{k_f K_M [cDLS - or - cSL2]}{1 + K_M [cDLS - or - cSL2]} + k_b \qquad (3)$$

where K_M is the equilibrium association constant governing the IC formation ($K_M = k_{ass}/k_{diss}$).

The k_{ass} values for both SL2/cSL2 (1.7×10^3 $M^{-1}s^{-1}$) and DLS/cDLS ($4.6 \pm 0.04 \times 10^2$ $M^{-1}s^{-1}$) annealing reactions were several orders of magnitude smaller than the 10^5-10^7 $M^{-1}s^{-1}$ rate constants reported for the annealing of unstructured sequences (48). This suggests that for both reactions, there is a low probability of the oligonucleotides to be melted at room temperature and to nucleate the duplex upon collision (49). Moreover, the values of k_b, given by intercepts, were found to be close to zero indicating that the dissociation of ED is negligible. This is in line with the large differences in the stability of the ED and the initial stem-loop. In contrast to the k_{obs} values, the amplitudes of the two kinetic components (Table 1 and 2) for both systems were independent of the complementary sequence concentrations (data not shown), suggesting that the reactions involves only one pathway, as described in scheme 1.

To further check the validity of the postulated two-step annealing mechanism, we used the Dynafit numerical resolution software (50), which allows fitting simultaneously the

experimental progress curves obtained at different complementary sequence concentrations. The best estimates of the elementary rate constants k_{ass}, k_{diss} and k_f (Table A, supplementary data) were in excellent agreement with those found by the empirical approach, (Table 1 and 2), further strengthening the proposed reaction scheme.

To get further insight in the annealing reaction mechanism, we further evaluated the temperature dependence of the k_{obs} values of SL2/cSL2 annealing kinetics by using the Arrhenius equation:

$$k_i = A_i \exp\left(E_{a,i} / RT\right) \qquad (4)$$

where the rate constant k_i is given by k_{obsi}/[cSL2], A_i is the pre-exponential Arrhenius factor, $E_{a,i}$ is the activation energy, R is the universal gas constant, and T is the temperature (in Kelvin).

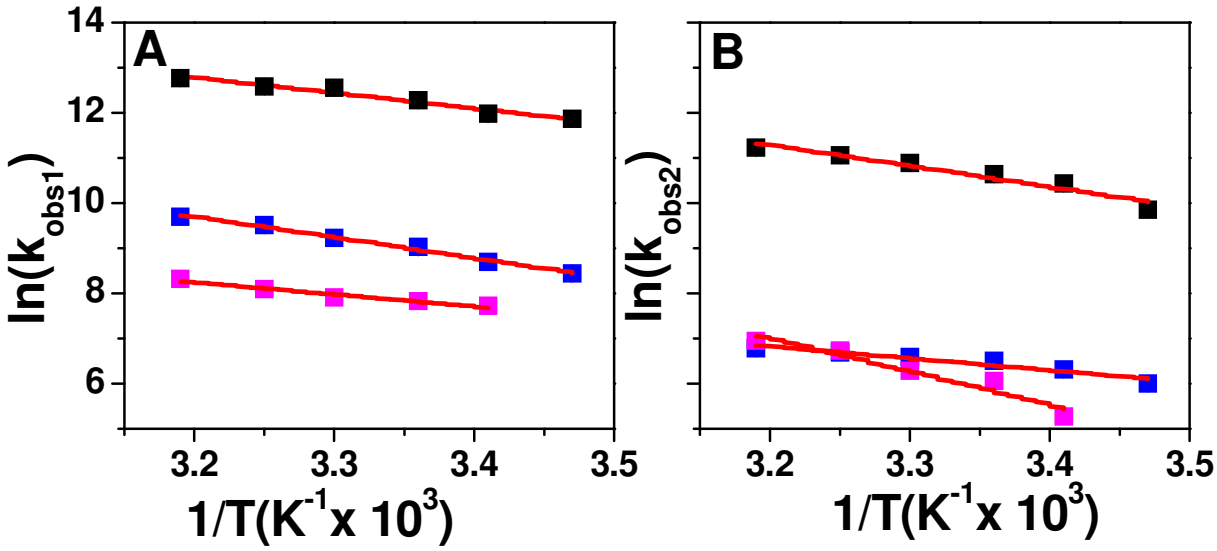


Figure 4: Temperature dependence of SL2/cSL2 in absence and presence of peptide E, and DLS/cDLS annealing kinetics in presence of peptide E. The reactions were performed in absence of peptide E, by mixing 10 nM doubly labelled SL2 with 6µM (pink squares) unlabelled cSL2. On the other hand, same reaction was performed in presence of peptide E at a peptide to oligonucleotide ratio 1:1 by mixing 10nM TMR-3'-SL2-5'-Fl with 200 nM (black squares) cSL2. The DLS/cDLS reaction was performed only in the presence of peptide E at ratio 0.5:1, with 20 nM doubly labelled DLS and 4 µM non-labelled cDLS (blue squares). The natural logarithms of the rate constant values for the fast (A) and slow (B) components are indicated at different temperatures. The solid red lines are the best fit to equation (4) and the fitting values are given in table 3.

Table 1. Kinetic parameters of SL2/cSL2 annealing and their mutants in the absence and presence of Core peptides

Oligo-nucleotide	Fluoro phores	Complementary sequence	Peptide	Ratio of peptide/Oligo	k_{ass} (M^{-1} s^{-1}) $\times 10^{-3}$	k_{diss} (s^{-1}) $\times 10^{4}$	K_M (M^{-1}) $\times 10^{-5}$	k_f (s^{-1}) $\times 10^{3}$
SL2	3'Fl – 5'TMR	**cSL2**	-	-	1.7 (±0.04)	62 (±6)	4.4 (±1)	1.8 (±0.2)
SL2	3'Fl – 5'TMR	**cSL2-Stem**	-	-	1.4 (±0.08)	43 (±5)	4.3 (±1)	2.2 (±0.2)
SL2	3'Fl – 5'TMR	**cSL2-Zip**	-	-	1 (±0.06)	18.3 (±2)	2.6 (±1)	1.1 (±0.2)
SL2	3'Fl – 5'TMR	**cSL2-Loop**	-	-	0.09 (±0.002)	15.1 (±0.4)	0.8 (±0.6)	1.5 (±0.2)
SL2-Loop	3'Fl – 5'TMR	**cSL2**	-	-	0.1 (±0.006)	10.1 (±0.8)	0.9 (±0.2)	0.82 (±0.06)
SL2	3'Fl – 5'TMR	**cSL2**	E	1	56 (±3)	89 (±17)	39 (±12)	19 (±0.1)
SL2	3'Fl – 5'TMR	**cSL2-Stem**	E	1	31 (±2)	86 (±6)	22 (±3)	17 (±0.4)
SL2	3'Fl – 5'TMR	**cSL2-Zip**	E	1	23 (±0.8)	67 (±6)	33 (±10)	6 (±0.4)
SL2	3'Fl – 5'TMR	**cSL2-Loop**	E	1	13 (±0.8)	60 (±6)	28 (±10)	1.6 (±0.1)
SL2-Loop	3'Fl – 5'TMR	**cSL2**	E	1	27 (±3)	280 (±26)	8 (±1)	3 (±1)
SL2-DNA	3'Rh6G – 5'Dabcyl	**cSL2-DNA**	-	-	1.4 (±0.05)	9 (±2)	4 (±1)	2.8 (±0.2)
SL2-DNA	3'Rh6G – 5'Dabcyl	**cSL2-DNA**	E	1	98 (±7)	173 (±5)	43 (±7)	10 (±0.4)
SL2	3'Fl – 5'TMR	**cSL2**	D1	1	31 (±2)	280 (±6)	13 (±4.7)	16.3 (±3)
SL2	3'Fl – 5'TMR	**cSL2**	F	1	308 (±11)	86 (±17)	382 (±24)	21 (±0.5)

Kinetic rate constants were calculated from the dependence of the k_{obs} values on the concentration of the unlabelled oligonucleotide and were calculated from equation (2) and (3).

Table 2. Kinetic parameters of DLS/cDLS annealing and their mutants in the absence and presence of Core peptides

Oligo-nucleotide	Fluoro phores	Complementary sequence	Peptide	Ratio of peptide/ Oligo	k_{ass} (M^{-1} s^{-1}) $\times 10^{-3}$	k_{diss} (s^{-1}) $\times 10^{4}$	K_M (M^{-1}) $\times 10^{-5}$	k_f (s^{-1}) $\times 10^{3}$
DLS	3'Fl – 5'TMR	**cDLS**	-	-	0.46 (±0.01)	180 (±8)	0.21 (±0.07)	0.44 (±0.05)
DLS	3'Fl – 5'TMR	**cDLS**	E	0.5	3 (±0.2)	89 (±4)	2.2 (±0.6)	5.4 (±0.7)
DLS	3'Fl – 5'TMR	**cDLS-Stem**	E	0.5	2 (±0.1)	92 (±4)	2.1 (±0.6)	6.3 (±0.3)
DLS	3'Fl – 5'TMR	**cDLS-Loop**	E	0.5	56 (±4)	95 (±8)	46 (±20)	50 (±21)
DLS-Loop	3'Fl – 5'TMR	**cDLS**	E	0.5	1.4 (±0.1)	168 (±32)	0.85 (±0.1)	3 (±1)

Kinetic rate constants were calculated from the dependence of the k_{obs} values on the concentration of the unlabelled oligonucleotide and were calculated from equation (2) and (3).

Table 3. Arrhenius parameters of SL2/cSL2 and DLS/cDLS annealing and their mutants in presence and absence of peptide E

Oligo-nucleotide	Fluoro phores	Complementary sequence	Ratio of Peptide E	ΔH (Fast) (kcalMol^{-1})	ΔH (Slow) (kcalMol^{-1})
SL2	3'Fl – 5'TMR	**cSL2**	0	5.4 (±0.7)	15 (±2)
SL2	3'Fl – 5'TMR	**cSL2-Stem**	0	4.8 (±0.9)	15.9 (± 2)
SL2	3'Fl – 5'TMR	**cSL2-Loop**	0	26.2 (±3)	8.5 (± 0.8)
SL2-Loop	3'Fl – 5'TMR	**cSL2**	0	11.5 (±1)	10.7 (± 0.5)
SL2	3'Fl – 5'TMR	**cSL2**	1	8.9 (±0.6)	9.3 (±1.1)
SL2	3'Fl – 5'TMR	**cSL2-Stem**	1	7.8 (±0.9)	15.8 (± 1.1)
SL2	3'Fl – 5'TMR	**cSL2-Zip**	1	8.5 (±0.5)	13.3 (± 1.1)
SL2	3'Fl – 5'TMR	**cSL2-Loop**	1	12.2 (±0.6)	12.7 (± 1)
SL2-Loop	3'Fl – 5'TMR	**cSL2**	1	12.8 (±0.6)	13.1 (± 1.7)
DLS	3'Fl – 5'TMR	**cDLS**	0.5	9.2 (±0.2)	5.3 (±0.7)
DLS	3'Fl – 5'TMR	**cDLS-Stem**	0.5	9.6 (±0.2)	9.8 (±0.4)
DLS	3'Fl – 5'TMR	**cDLS-Loop**	0.5	6.3 (±0.6)	5.4 (±0.6)
DLS-Loop	3'Fl – 5'TMR	**cDLS**	0.5	16.6 (±1.6)	20 (±1.5)

Temperature dependence parameters were calculated as described in figure 4.

Both reaction rates were found to increase with temperature (Figure 4 A and B, pink squares), providing positive enthalpy values for the transition state of 5.4(±0.7) kcal/mol and 15(±2) kcal/mol for the fast and slow components, respectively, for SL2/cSL2 annealing. These enthalpy values indicate that the SL2/cSL2 annealing reaction involves premelting of ~1 and ~3 bp, for fast and slow kinetic components, respectively (51,52).

Effect of oligonucleotide sequence on the annealing kinetics

The molecular determinants of the annealing reaction were characterized by investigating the annealing kinetics with different SL2 and cSL2 mutants (Figure 1). To investigate the role of stem, cSL2 was substituted by the cSL2-Stem mutant in which the last and penultimate base pairs at the bottom of the stem were inverted, in order to prevent base-pairing with SL2 through the stem termini. These mutations only marginally changed the kinetics parameters (Table 1, Fig 5A) and the values of the transition state enthalpy (Table 3), suggesting that the SL2/cSL2 annealing is not nucleated through the stem termini. Next, to determine whether the annealing was nucleated through kissing loop interactions, we used SL2-Loop and cSL2-Loop mutants, where $A_{16,\ 17\ and\ 18}$ and C_{13}, A_{14} and C_{15}, respectively, were substituted by U nucleotides (Fig 1) in order to decrease the complementarities between the central loops. By monitoring SL2-Loop/cSL2 and SL2/cSL2-Loop annealing reactions, a one order of magnitude decrease in the values of k_{ass}, and k_M as well as a 4- to 6-fold decrease in the k_{diss} value, coupled with a strongly increased value of the transition state enthalpy (Table 3) for the fast kinetic component were observed. These dramatic changes in the annealing parameters with these mutant sequences strongly suggested that the SL2/cSL2 annealing reaction was nucleated through the loops. Interestingly, the very high value of the transition state enthalpy (26.2 kcal. mol^{-1}) for the fast component of the SL2/cSL2-loop annealing reaction is consistent with the melting of 5 to 6 bp, suggesting that in this case, annealing could be nucleated through the 6bp-long stem bottom. In addition, a significant decrease in the fluorescence plateau was observed for both SL2-Loop/cSL2 and SL2/cSL2-Loop EDs as compared to the SL2/cSL2 ED (Fig. 5A). This fluorescence plateau was found independent on the concentration of the complementary sequence, and thus could not be related to incomplete reaction between partners. A more likely explanation is that the partial complementarity between the loops induces some flexibility in the middle of the extended duplex, allowing its kinking and thus, a closer distance between the probes at the opposite ends of the duplex. Finally, to investigate the contribution of the internal loop in the annealing reaction, we used another mutant, cSL2-Zip, in which the A_{20} and G_{21} residues were

substituted by U nucleotides (Fig 1) in order to prevent base-pairing with the residues of the SL2 internal loop. The values of the kinetic parameters with this mutant show about 2-fold decrease with respect to the native cSL2 sequence (Table 1, Fig 5A), suggesting a moderate contribution of the internal loop in the annealing reaction.

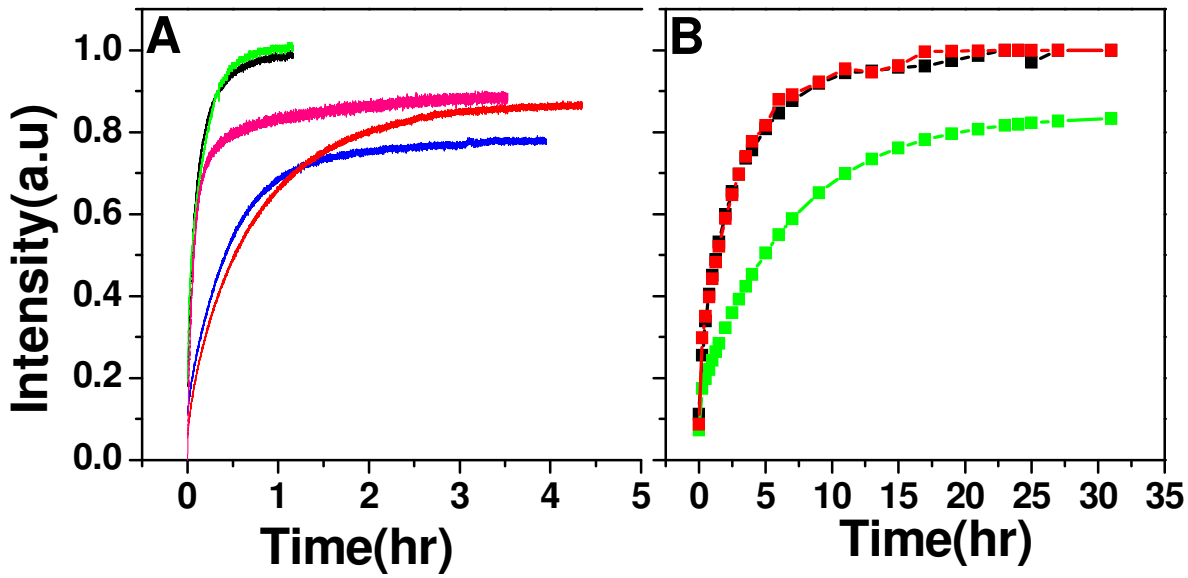


Figure 5: Annealing kinetics of SL2/cSL2 and DLS/cDLS and their respective mutants. (A) Kinetic traces of 10 nM doubly labelled SL2 derivatives with 6 µM non-labelled cSL2 derivatives [SL2/cSL2 (black), SL2/cSL2-Stem (green), SL2/cSL2-Zip (pink), SL2/cSL2-Loop (blue) and SL2-Loop/cSL2 (red)]. (B) Kinetic traces of 20 nM doubly labelled DLS derivatives with 10 µM non-labelled cDLS derivatives [DLS/cDLS (Black), DLS/cDLS-Stem (red) and DLS-Loop/cDLS (green)].

The very slow rates of the DLS/cDLS annealing reaction even at very high concentrations of complementary sequences prevented us to determine the full set of kinetic parameters. Nevertheless, experiments at single spot concentration of DLS with the cDLS-Stem mutant (where last and penultimate base pairs at the bottom of the stem were inverted, Fig 1) indicated almost no changes in the reaction rates as compared to the reaction with the native sequences (Fig. 5B). In sharp contrast, a strong decrease in the reaction rate as well as a decrease in the fluorescence plateau were observed when cDLS was reacted with the doubly labelled DLS-Loop mutant, in which the A_9 and G_{10} residues were substituted by U nucleotides (Figure 1), indicating that the annealing reaction for DLS/cDLS, like SL2/cSL2, is also nucleated through kissing loop interactions.

Further, we checked if substitution of ribose backbone to desoxyribose in form of DNA sequences made any impact on the kinetic parameters. For this, we repeated the annealing experiments with DNA version of SL2 and cSL2 sequences. Observed results (Table 1) indicated that the nature of the nucleotide backbone marginally influences the kinetic parameters and thus, the annealing mechanism.

Taken together, our data indicate that the SL2/cSL2 and the DLS/cDLS annealing reactions are nucleated through loop-loop interactions. Consequently, kissing loop complexes are likely the intermediate complexes in the reaction scheme.

Kinetics of DLS/cDLS and SL2/cSL2 annealing in the presence of peptide E

To characterize the mechanism by which the core protein can activate the annealing of both systems (DLS/cDLS and SL2/cSL2), we first use peptide E, which mimics the core chaperone properties and can be synthesized in large quantities (26,32,33). Addition of peptide E to the labelled DLS or SL2 sequences did not lead to any change in their fluorescence spectrum (Fig. 6A and B, insets), indicating that the peptide was unable to destabilize their secondary structure, as already reported with the model cTAR sequence (33). Annealing reactions were then performed by adding peptide E at a peptide to nucleotide ratio of 0.5:1 and 1:1 to DLS/cDLS and SL2/cSL2 systems, respectively, in order to ensure aggregation-free conditions. A rapid increase in the fluorescence was observed when the doubly-labelled DLS and SL2 sequences were mixed with their complementary cDLS and cSL2 sequences in the presence of peptide E. The final fluorescence reached the same plateau value as in the absence of core peptide, suggesting that the reaction went to completion giving the same ED.

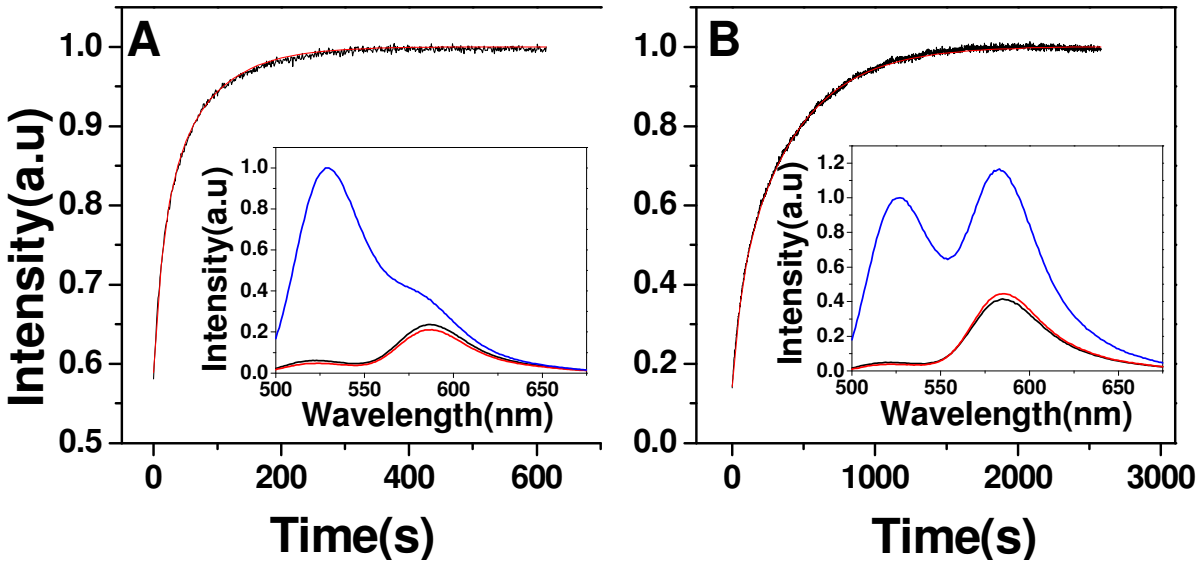


Figure 6: Kinetics of SL2/cSL2 and DLS/cDLS annealing in presence of peptide E. (A) Kinetic trace of 10 nM doubly labelled SL2 with 500 nM of unlabelled cSL2. Inset A: Emission spectra of the doubly labelled SL2 before (black) and after (blue) completion of annealing reaction with cSL2. The spectrum in red corresponding to the emission spectrum of the doubly labelled SL2 in the presence of peptide E, shows marginal destabilisation of SL2 by peptide E. (B) Kinetic trace of 20 nM doubly labelled DLS with 3 µM of unlabelled cDLS. Inset B: Emission spectra of the doubly labelled DLS before (black) and after (blue) completion of annealing reaction with cDLS. The spectrum in red corresponding to the emission spectrum of the doubly labelled DLS in the presence of peptide E, again shows marginal destabilisation of DLS by peptide E. Experiments were performed in 50 mM Hepes (pH 7.5), 30 mM NaCl, 0.2 mM $MgCl_2$ at 20°C with peptide E added at peptide to oligonucleotide ratio of 1:1 and 0.5:1 for SL2/cSL2 and DLS/cDLS annealing, respectively. Excitation and emission wavelengths were 480 nm and 520 nm, respectively, to monitor the fluorescein emission. The red lines correspond to the fits of the kinetic traces with eqn. 1.

Interestingly, peptide E induced a dramatic increase in the reaction kinetics, as the reactions for both systems were completed much faster as compared to their respective reactions in its absence (compare Figure 2 and Figure 6). As in the absence of peptide, the kinetics traces could be adequately fitted with equation (1). Furthermore, like in the absence of core, the fast component, k_{obs1}, varied linearly with [cDLS or cSL2], while k_{obs2} showed a hyperbolic dependence, suggesting a similar two-step reaction pathway (scheme 1). For both systems, peptide E was found to increase the values of k_{ass}, k_f and K_M by about one magnitude, while the k_{diss} value remained almost unchanged (Tables 1 and 2). The values of k_b for both

annealing reactions were found to be very low (<0.0001 s^{-1}), indicating that peptide E is unable to dissociate the ED. Finally, it was observed that the amplitudes of both components were independent of the concentration of the complementary sequence (data not shown). Thus, on the basis of our kinetic data, the same reaction mechanism as described in scheme 1 can also be proposed for peptide E-promoted-SL2/cSL2 and -DLS/cDLS annealing. The validity of this reaction mechanism was checked with the Dynafit numerical resolution software in the same manner as in the absence of core peptide (Table A, supplementary material).

Finally, we further evaluated the kinetic reaction mechanism by analyzing the temperature dependence of the k_{obs} values as a function of the temperature. Analysis of both reaction rates with temperature (Figure 4) provided positive enthalpy values (Table 3) for peptide E-promoted-SL2/cSL2 and -DLS/cDLS annealing. These enthalpy values indicate that peptide E-promoted-SL2/cSL2 annealing involves premelting of ~2 to 3 pb, for both kinetic components (51,52), while peptide E-promoted-DLS/cDLS annealing involves premelting of ~2 to 3 and ~1 to 2 bp, for fast and slow kinetic components, respectively (51,52).

Effect of oligonucleotide sequence on the core peptide E-promoted annealing kinetics

To characterize the molecular determinants in the peptide E-promoted-SL2/cSL2 annealing reaction, we first used the cSL2-Stem mutant. As in the absence of peptide, this mutation was found to have no effect on the kinetics of the peptide E-promoted annealing reaction (Table 1 and Figure 7A), indicating that the annealing is not promoted through the stem termini. Interestingly, an increased value of the transition state enthalpy (Table 3) for the slow kinetic component was observed with this mutant. This likely results from the incremental energy requirement for the melting of the SL2 stem due to its non complementarity with the terminal bases of the cSL2-stem mutant (53), suggesting that the IC interconversion step relies on the melting of the lower half of SL2. In contrast to the stem mutants, the cSL2-Loop and SL2-Loop mutants induced a 2- to 5-fold decrease in the k_{ass} and K_M values, as well as 6- to-10 fold decrease in the k_f value (Table 1), suggesting that the peptide E-promoted annealing reaction is nucleated through a kissing loop intermediate. Next, to determine the role of the internal loop in the peptide E-promoted annealing reaction, we substituted cSL2 by the cSL2-Zip mutant. This substitution did not change significantly the k_{diss} and K_M values but induced a significant 3-fold decrease in the k_f value as well as an increase in the reaction enthalpy of the slow component (Table 3), suggesting that the internal loop is involved in the conversion of the kinetic intermediate into the ED. Alternatively, this decrease in the k_f value (Table 1)

could be related to the increased stability of the cSL2-Zip mutant, as a consequence of the additional A_7:U_{21} base-pair in its stem (Fig. 1). Further, annealing experiments with the DNA versions of SL2 and cSL2, indicated that the nature of nucleotide backbone does not affect the peptide E-promoted annealing mechanism (Table 1).

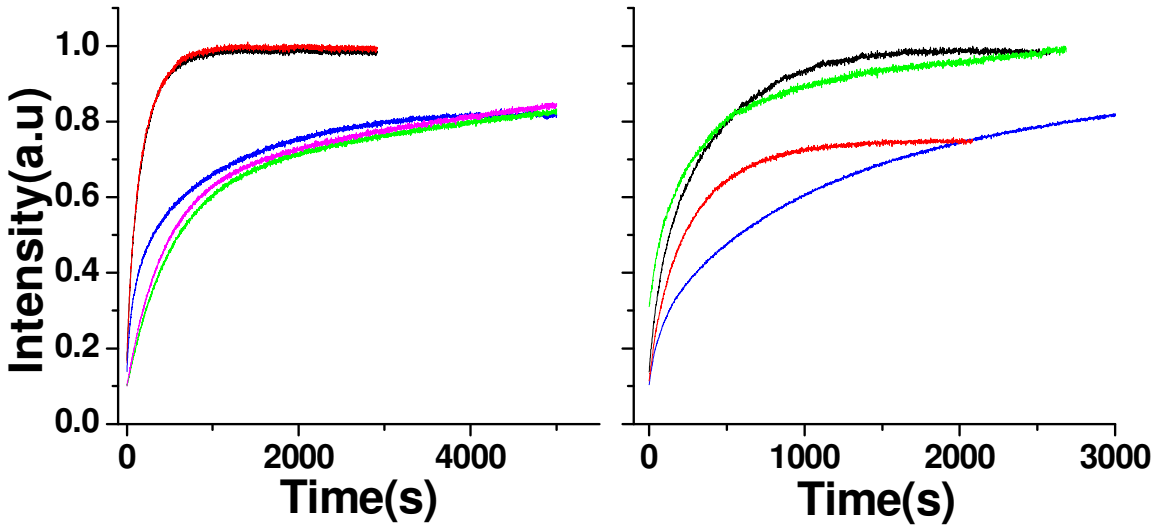


Figure 7: Kinetic of annealing of HCV SL2 and DLS derivatives with their respective complementary sequences in presence of peptide E. (A) Kinetic traces of 10 nM doubly labelled SL2 derivatives with 200 nM non-labelled cSL2 derivatives [SL2/cSL2 (black), SL2/cSL2-Stem (red), SL2/cSL2-Loop (magenta), SL2/cSL2-Zip (green) and SL2-Loop/cSL2 (blue)]. (B) Kinetic traces of 20 nM doubly labelled DLS derivatives with 3 µM non-labelled cDLS derivatives [DLS/cDLS (Black), DLS/cDLS-Stem (green) and DLS-Loop/cDLS (blue)] and 500 nM non-labelled cDLS derivative [DLS/cDLS-Loop (red)].

We then investigated the molecular determinants for the peptide E-promoted-DLS/cDLS annealing reaction. Substitution of cDLS by the cDLS-Stem mutant did not significantly change the kinetics parameters, (Table 2 and Figure 7B), indicating that the peptide E-promoted DLS/cDLS annealing reaction is not promoted through the stem termini. Moreover, similar to peptide E-promoted-SL2/cSL2 annealing, an increased value of the transition state enthalpy was observed for the slow kinetic component (Table 3). As for SL2, this may be linked to the incremental energy needed to melt the DLS stem termini in the absence of complementary bases in the cDLS-stem mutant, indicating that the interconversion of IC into ED is related to the melting of DLS stem.

Surprisingly, substitution of cDLS by the cDLS loop mutant in which A_9 and G_{10}, were substituted by U nucleotide (Figure 1) was found to increase the k_{ass}, K_M and k_f values by about one order of magnitude as compared to the values observed with DLS/cDLS (Table 2). These strong changes in the annealing kinetics suggest that the loop plays a key role in the DLS/cDLS annealing mechanism. In this respect, comparison of the cDLS-Loop data with the native cDLS data suggests the presence of structural constraints in the loop of native cDLS that likely restrict its annealing with the DLS loop, as well as the conversion of the IC into the ED. Due to the large size of the cDLS loop (8 nt), possible base pairing between the ^{6}CG^{7} and the ^{10}CG^{11} nucleotides could constrain the loop of the native cDLS sequence and slow down the annealing with the DLS sequence. In contrast, substitution of G_{10} by U could prevent this base pairing and thus favour an efficient annealing of the resulting flexible mutated loop with DLS. The decreased stability of the mutated loop sequence as compared to the native loop in cDLS was well supported by the decreased value of the transition state enthalpy (Table 3) for the fast kinetic component, which indicates a lower energy barrier for nucleating the IC in the case of the cDLS-loop mutant.

To confirm the role of the loops in the annealing reactions, we mutated the loop of DLS, by substituting the A_9 and G_{10} residues with U nucleotides.

In contrast to the data with the cDLS mutant, we observed a 2 to 3-fold decrease in the values of k_{ass}, K_M and k_f (Table 2) together with a strong increase in the values of the transition state enthalpy (Table 3) for both the fast and slow kinetic components, as compared to the peptide E-promoted-DLS/cDLS annealing reaction. These data unambiguously confirmed that the peptide E-promoted annealing reaction of DLS with cDLS is nucleated through the loops.

Taken together, our data indicate that the peptide E promoted-DLS/cDLS and the -SL2/cSL2 annealing reactions are nucleated through loop-loop interaction, providing a kissing intermediate. This IC is then further converted into the final ED, in a reaction which is rate-limited by the melting of the lower half of the oligonucleotides.

SL2/cSL2 annealing promoted by peptide F and D1 Domain

Next, we investigated the effect of the other two core peptides (peptide F and D1 domain) on the annealing of SL2 with cSL2, to determine whether they promote the annealing reaction through the same mechanism as peptide E (Fig 1B). The annealing kinetics in their presence was also biphasic with a linear and hyperbolic dependence of k_{obs1} and k_{obs2}, respectively, on the concentration of cSL2 (Table 1). These data suggest that the three core peptides promote SL2/cSL2 annealing by the same mechanism. We observed marked differences in the k_{ass} and

K_M values with peptide F and D1 domain as compared to peptide E. While a 2- to 3 -fold decrease of the values both parameters was observed with the D1 domain (Table 1), an increase of about 1 order of magnitude was observed for peptide F (Table 1). In contrast, the k_f values for the three peptides were similar. Thus, the three peptides promote and stabilize the IC at different levels, but once this complex is formed, it is converted with similar kinetics by the three peptides into the ED. The increased k_{ass} and K_M values observed with peptide F as compared to peptide E could likely be attributed to its additional basic domain that additionally screens the negative charges of the ODNs and thus, favour their annealing. In contrast, the opposite effect observed with the D1 domain is likely related to its additional neutral Trp-rich domains, which reduce the density of positive charges on the peptide, and thus the ability to screen efficiently the charges of the ODN.

Effect of core peptide on the dynamics of the 2Ap-labeled HCV nucleotides

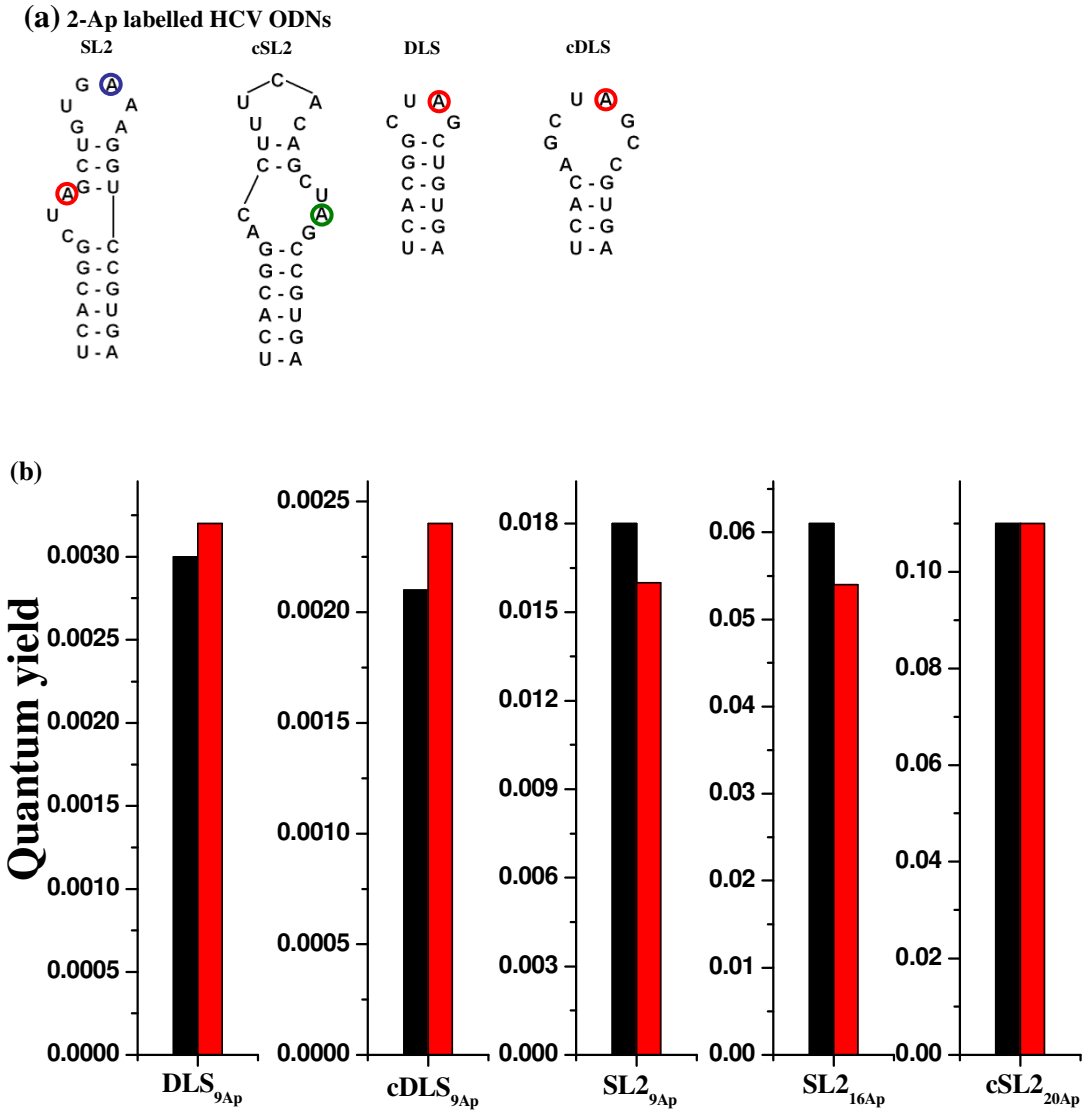


Figure 8: Site-selective monitoring of the interaction of peptide E with SL2 and DLS and their respective complementary sequences labelled with 2-Ap. (a) Labelled positions in HCV oligonucleotides. The sequences were labelled by 2-Ap at position 9 (red circles), 16 (blue circles) and 20 (green circle). Quantum yields of (b) DLS_{9Ap}, $cDLS_{9Ap}$, $SL2_{9Ap}$, $SL2_{16Ap}$ and $cSL2_{20Ap}$ in the absence (black) and presence (red) of peptide E. Quantum yields were calculated assuming a quantum yield of 0.68 for free 2Ap (42). Excitation wavelength was 315 nm and all experiments were performed in 50 mM Hepes (pH 7.5), 30 mM NaCl, 0.2 mM $MgCl_2$ at 20°C.

Further, site-selective interactions of peptide E with DLS and SL2, were characterised by using ODNs labelled by 2Ap, a fluorescence analogue of adenine, highly sensitive to the

environment. To this end, we substituted A residues by 2-Ap in DLS and cDLS at position 9 (DLS_{9Ap} and $cDLS_{9Ap}$) as well as in SL2 at position 9 ($SL2_{9Ap}$) and 16 ($SL2_{16Ap}$) and cSL2 at position 20 ($cSL2_{20Ap}$) (Fig 8a). In the absence of peptide E, the 2Ap-labelled oligonucleotides exhibit a rather low quantum yield in comparison to that of free 2-Ap, as a consequence of the dynamic quenching of 2Ap fluorescence by the neighbouring bases (44,54), and notably the flanking G residues which are the strongest quenchers among the bases (55).

Binding of peptide E to the various 2-Ap labelled oligonucleotides led only to marginal changes in the quantum yield (Fig 8b) and emission maximum (data not shown), suggesting that peptide E does not restrict the dynamic interactions of the 2-Ap residues with their neighbouring bases in the central loop of SL2, DLS and cDLS and the internal loop of SL2 and cSL2. The absence of any effect for the central loop region of these ODNs can hardly be connected with an absence of binding of peptide E to this region, since the annealing reaction for these oligonucleotides was shown to be promoted by peptide E through the loops.

Discussion

RNA chaperone proteins are vital partners of RNA in viruses. They are thought to chaperone, through an entropy driven process, the correct folding and structural rearrangements of RNA molecules by resolving misfolded kinetically trapped RNA species in an ATP-independent manner (41). At various stages of the cellular RNA metabolism, that includes transcription, transport, translation and storage, RNA molecules are associated with a particular set of chaperone molecules that protect them and help in their folding process (40,56). To accomplish its functions, the genomic RNA relies at least in part on short specific cis-acting RNA elements (CREs) (12,56,57). The genomic RNA of HCV contains a short palindromic CRE, named DLS, mediating *in vitro* dimerization of the genomic RNA 3'UTR upon core binding (26,32). The aim of the present work was to characterize the mechanism by which the HCV core chaperones the annealing of DLS and the stem-loop sequence SL2 that includes DLS with their complementary partners (cDLS and cSL2, respectively, and to identify the structural determinants for this mechanism. To this end, we monitored in real-time conditions, the annealing of these RNA sequences in the absence and presence of different core peptides (D1 domain, peptide E and F).

In the absence of Core peptides, the annealing of both SL2 with cSL2 and DLS with cDLS was rather slow, taking several hours to reach completion (Table 1 and Table 2, Fig 1). This slow annealing kinetics can be related with the requirement of the complementary sequences

to be in a reactive conformation and proper orientation in order to nucleate the intermediate complex. Using DLS and SL2 mutants, we found that in both cases, the annealing reaction was initiated through the loops, providing a kissing loop intermediate that further converts into the final ED. Formation of the SL2/cSL2 intermediate complex involves an enthalpy energy of about 5 kcal/mol that likely corresponds to the melting of the $_{11}C\text{-}G_{21}$ base pair surrounded by two poorly stable noncanonical G-U base pairs at the base of the SL2 central loop. Alternatively, this energy may correspond to the melting in cSL2 of the corresponding double stranded segment composed of $_{9}C\text{-}G_{17}$ and $_{10}U\text{-}A_{16}$ base pairs which is certainly partly destabilized by the surrounding central and internal loops. The resulting IC could be stabilized by up to 15 intermolecular base-pairs. However, the actual number of base pairs in the IC is likely less, as a consequence of steric constraints imposed by the conformation of the loop and the stem. The IC is then converted to the ED, in a rate limiting step associated with an enthalpy energy of about 15 kcal/mol. This energy likely corresponds to the melting of the remaining part of the stem or at least the three upper G-C bases pairs, which are probably the key residues for maintaining the stem integrity.

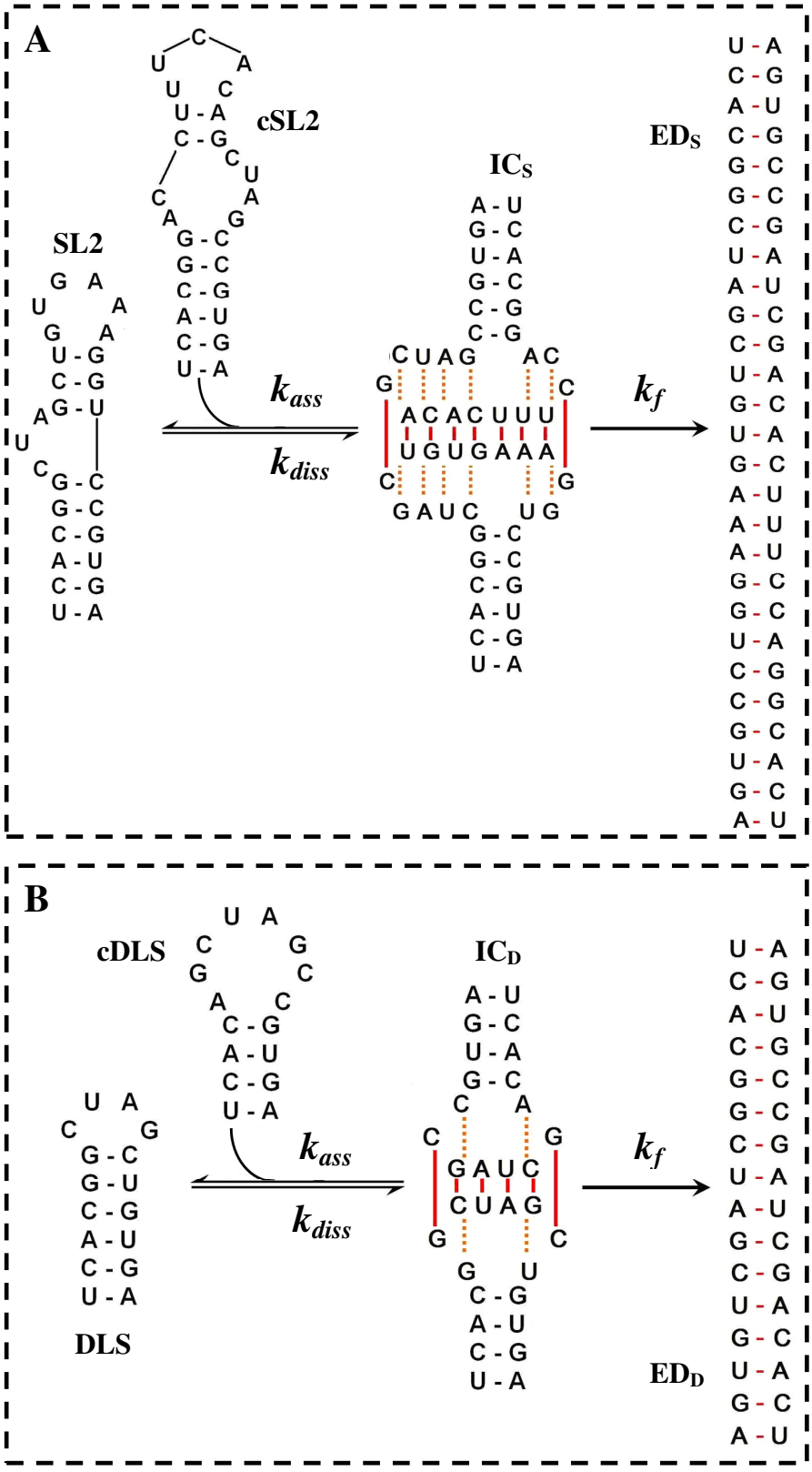


Figure 9: Proposed mechanism for the core-promoted (A) SL2/cSL2 and (B) DLS/cDLS annealing reactions. In both systems, the mechanistic pathway is nucleated through loop-loop

interaction. This leads to the formation of intermediate complex (IC_S and IC_D) that subsequently, converts in a rate-limiting step to the final extended duplex (ED_S and ED_D).

The core peptides were found to efficiently activate the annealing of SL2/cSL2 and DLS/cDLS as shown by the one order of magnitude increase in the values of k_{ass}, K_M and k_f in the presence of the core peptide as compared to their absence (Table 1 and 2). This activation is in line with the previously reported promotion, by the same core peptides, of the annealing of HIV-1 TAR with cTAR, taken as model sequences (33). However, in sharp contrast to the TAR/cTAR system, where the core protein shifted the annealing mechanism from a kissing loop pathway to a zipper pathway through the stem termini, no change in the mechanistic pathway was observed when the core was added to the SL2/cSL2 or DLS/cDLS systems. In fact, the core protein appears to facilitate the already existing pathway in both systems by favouring the nucleation of the IC, mainly through an increase in the association rate constant. This increase is likely related to the core-induced attraction between the complementary oligonucleotides complexed to core, which favours the diffusional search for the complementary sequences (ref 54). A second key issue in the core-promoted annealing is related to the enhancement of the inconversion rate. The protein likely favours the conformational rearrangement and the melting of the remaining stem, which are required for converting the IC into the ED. The role of the protein on the melting of the stem is clearly assessed by the significant decrease in the enthalpy energy associated to the interconversion step.

In the case of DLS/cDLS annealing in presence of core peptide, formation of the IC is likely associated with the melting of the two upper base pairs ($_6$G-C_{11} and $_5$G-U_{12}) of the DLS stem.. The DLS/cDLS intermediate which could involve up to 8 intermolecular base pairs (Fig 9) is then converted into the final ED in a step, likely rate-limited by the melting of the upper $_4$G-C_{13} base pair. The melting of this base pair is then probably followed by the final zipping of the remaining base pairs to complete the structure of the ED.

The mechanism of core in the activation of both annealing reactions is also in line with the 'entropy exchange model' in which a highly flexible protein, like core, undergoes disorder-to-order transition upon binding to RNA that in turn leads to the melting of the RNA structure through an entropy exchange process (58). Interestingly, using 2-Ap labelled oligonucleotides to monitor site-selectively the effect of the core protein, we found that the core did not modify the dynamics of the bases in the central and internal loop of SL2 or cSL2, as well as in the

loop of DLS and cDLS. This observation is in contrast with the effect of the nucleocapsid protein NCp7, an HIV-1 chaperone protein, which freezes the local mobility of bases through interaction with its hydrophobic aminoacids (44,59). This suggests that by analogy to a fingerless and unfolded mutant of NCp7 (44), peptide E likely interacts with the ODNs mainly through electrostatic interactions between its basic amino acids and the negatively charged phosphate groups, leaving the bases unaffected.

In conclusion, our data with the HCV RNA sequences further demonstrated that the core protein exhibits potent *in vitro* nucleic acid chaperone activities, facilitating RNA-RNA and RNA-DNA interactions and structural rearrangements (26,32,33). Since the HCV genomic RNA likely adopts different conformations during the viral life cycle, we can speculate that through its chaperone properties, the core protein can promote the interconversion between these different structures, and thus regulate the transitions between translation and replication as well as between replication and packaging of the genomic RNA (17,60). Moreover, as in the case of NCp7, the core may also contribute to the genetic variability of HCV by enhancing the frequency of copy-choice recombination (57,61,62).

References

1. Giannini, C. and Brechot, C. (2003) Hepatitis C virus biology. *Cell Death Differ*, **10 Suppl 1**, S27-38.
2. Hadziyannis, S.J. and Papatheodoridis, G.V. (2004) Emerging treatments in chronic hepatitis B. *Expert opinion on emerging drugs*, **9**, 207-221.
3. Hadziyannis, S.J., Sette, H., Jr., Morgan, T.R., Balan, V., Diago, M., Marcellin, P., Ramadori, G., Bodenheimer, H., Jr., Bernstein, D., Rizzetto, M. *et al.* (2004) Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. *Annals of internal medicine*, **140**, 346-355.
4. Lindsay, K.L., Trepo, C., Heintges, T., Shiffman, M.L., Gordon, S.C., Hoefs, J.C., Schiff, E.R., Goodman, Z.D., Laughlin, M., Yao, R. *et al.* (2001) A randomized, double-blind trial comparing pegylated interferon alfa-2b to interferon alfa-2b as initial treatment for chronic hepatitis C. *Hepatology (Baltimore, Md*, **34**, 395-403.
5. McHutchison, J.G., Everson, G.T., Gordon, S.C., Jacobson, I.M., Sulkowski, M., Kauffman, R., McNair, L., Alam, J. and Muir, A.J. (2009) Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection. *The New England journal of medicine*, **360**, 1827-1838.
6. McHutchison, J.G., Lawitz, E.J., Shiffman, M.L., Muir, A.J., Galler, G.W., McCone, J., Nyberg, L.M., Lee, W.M., Ghalib, R.H., Schiff, E.R. *et al.* (2009) Peginterferon alfa-2b or alfa-2a with ribavirin for treatment of hepatitis C infection. *The New England journal of medicine*, **361**, 580-593.
7. Muir, A.J., Bornstein, J.D. and Killenberg, P.G. (2004) Peginterferon alfa-2b and ribavirin for the treatment of chronic hepatitis C in blacks and non-Hispanic whites. *The New England journal of medicine*, **350**, 2265-2271.

8. Appel, N., Herian, U. and Bartenschlager, R. (2005) Efficient rescue of hepatitis C virus RNA replication by trans-complementation with nonstructural protein 5A. *Journal of virology*, **79**, 896-909.
9. Appel, N., Pietschmann, T. and Bartenschlager, R. (2005b) Mutational analysis of hepatitis C virus nonstructural protein 5A: potential role of differential phosphorylation in RNA replication and identification of a genetically flexible domain. *Journal of virology*, **79**, 3187-3194.
10. Hiraga, N., Imamura, M., Tsuge, M., Noguchi, C., Takahashi, S., Iwao, E., Fujimoto, Y., Abe, H., Maekawa, T., Ochi, H. *et al.* (2007) Infection of human hepatocyte chimeric mouse with genetically engineered hepatitis C virus and its susceptibility to interferon. *FEBS Lett*, **581**, 1983-1987.
11. Lohmann, V., Korner, F., Koch, J., Herian, U., Theilmann, L. and Bartenschlager, R. (1999) Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science*, **285**, 110-113.
12. Moradpour, D., Penin, F. and Rice, C.M. (2007) Replication of hepatitis C virus. *Nat Rev Microbiol*, **5**, 453-463.
13. Blight, K.J., Kolykhalov, A.A. and Rice, C.M. (2000) Efficient initiation of HCV RNA replication in cell culture. *Science*, **290**, 1972-1974.
14. Penin, F., Dubuisson, J., Rey, F.A., Moradpour, D. and Pawlotsky, J.M. (2004) Structural biology of hepatitis C virus. *Hepatology (Baltimore, Md*, **39**, 5-19.
15. Tsukiyama-Kohara, K., Iizuka, N., Kohara, M. and Nomoto, A. (1992) Internal ribosome entry site within hepatitis C virus RNA. *Journal of virology*, **66**, 1476-1483.
16. Wang, C., Sarnow, P. and Siddiqui, A. (1993) Translation of human hepatitis C virus RNA in cultured cells is mediated by an internal ribosome-binding mechanism. *Journal of virology*, **67**, 3338-3344.
17. Blight, K.J. and Rice, C.M. (1997) Secondary structure determination of the conserved 98-base sequence at the 3' terminus of hepatitis C virus genome RNA. *Journal of virology*, **71**, 7345-7352.
18. Kolykhalov, A.A., Feinstone, S.M. and Rice, C.M. (1996) Identification of a highly conserved sequence element at the 3' terminus of hepatitis C virus genome RNA. *Journal of virology*, **70**, 3363-3371.
19. Friebe, P. and Bartenschlager, R. (2002) Genetic analysis of sequences in the 3' nontranslated region of hepatitis C virus that are important for RNA replication. *Journal of virology*, **76**, 5326-5338.
20. Kolykhalov, A.A., Mihalik, K., Feinstone, S.M. and Rice, C.M. (2000) Hepatitis C virus-encoded enzymatic activities and conserved RNA elements in the 3' nontranslated region are essential for virus replication in vivo. *Journal of virology*, **74**, 2046-2051.
21. Yanagi, M., St Claire, M., Emerson, S.U., Purcell, R.H. and Bukh, J. (1999) In vivo analysis of the 3' untranslated region of the hepatitis C virus after in vitro mutagenesis of an infectious cDNA clone. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 2291-2295.
22. Yi, M. and Lemon, S.M. (2003) 3' nontranslated RNA signals required for replication of hepatitis C virus RNA. *Journal of virology*, **77**, 3557-3568.
23. Yi, M. and Lemon, S.M. (2003) Structure-function analysis of the 3' stem-loop of hepatitis C virus genomic RNA and its role in viral RNA replication. *RNA*, **9**, 331-345.
24. Song, M., Balakrishnan, M., Chen, Y., Roques, B.P. and Bambara, R.A. (2006) Stimulation of HIV-1 minus strand strong stop DNA transfer by genomic sequences 3' of the primer binding site. *The Journal of biological chemistry*, **281**, 24227-24235.

25. Song, Y., Friebe, P., Tzima, E., Junemann, C., Bartenschlager, R. and Niepmann, M. (2006b) The hepatitis C virus RNA 3'-untranslated region strongly enhances translation directed by the internal ribosome entry site. *Journal of virology*, **80**, 11579-11588.
26. Ivanyi-Nagy, R., Kanevsky, I., Gabus, C., Lavergne, J.P., Ficheux, D., Penin, F., Fosse, P. and Darlix, J.L. (2006) Analysis of hepatitis C virus RNA dimerization and core-RNA interactions. *Nucleic acids research*, **34**, 2618-2633.
27. Shetty, S., Kim, S., Shimakami, T., Lemon, S.M. and Mihailescu, M.R. (2010) Hepatitis C virus genomic RNA dimerization is mediated via a kissing complex intermediate. *Rna*, **16**, 913-925.
28. Baumert, T.F., Ito, S., Wong, D.T. and Liang, T.J. (1998) Hepatitis C virus structural proteins assemble into viruslike particles in insect cells. *Journal of virology*, **72**, 3827-3836.
29. Kunkel, M., Lorinczi, M., Rijnbrand, R., Lemon, S.M. and Watowich, S.J. (2001) Self-assembly of nucleocapsid-like particles from recombinant hepatitis C virus core protein. *Journal of virology*, **75**, 2119-2129.
30. Shimoike, T., Mimori, S., Tani, H., Matsuura, Y. and Miyamura, T. (1999) Interaction of hepatitis C virus core protein with viral sense RNA and suppression of its translation. *Journal of virology*, **73**, 9718-9725.
31. Cimarelli, A. and Luban, J. (1999) Translation elongation factor 1-alpha interacts specifically with the human immunodeficiency virus type 1 Gag polyprotein. *Journal of virology*, **73**, 5388-5401.
32. Cristofari, G., Ivanyi-Nagy, R., Gabus, C., Boulant, S., Lavergne, J.P., Penin, F. and Darlix, J.L. (2004) The hepatitis C virus Core protein is a potent nucleic acid chaperone that directs dimerization of the viral (+) strand RNA in vitro. *Nucleic acids research*, **32**, 2623-2631.
33. Sharma, K., Didier, P., Darlix, J.L., de Rocquigny, H., Bensikaddour, H., Lavergne, J.P., Penin, F., Lessinger, J.M. and Mely, Y. (2010) Kinetic analysis of the nucleic acid chaperone activity of the hepatitis C virus core protein. *Nucleic acids research*, **38**, 3632-3642.
34. Bernacchi, S., Stoylov, S., Piemont, E., Ficheux, D., Roques, B.P., Darlix, J.L. and Mely, Y. (2002) HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. *Journal of molecular biology*, **317**, 385-399.
35. Egele, C., Piemont, E., Didier, P., Ficheux, D., Roques, B., Darlix, J.L., de Rocquigny, H. and Mely, Y. (2007) The single-finger nucleocapsid protein of moloney murine leukemia virus binds and destabilizes the TAR sequences of HIV-1 but does not promote efficiently their annealing. *Biochemistry*, **46**, 14650-14662.
36. Egele, C., Schaub, E., Piemont, E., de Rocquigny, H. and Mely, Y. (2005) Investigation by fluorescence correlation spectroscopy of the chaperoning interactions of HIV-1 nucleocapsid protein with the viral DNA initiation sequences. *C R Biol*, **328**, 1041-1051.
37. Godet, J., de Rocquigny, H., Raja, C., Glasser, N., Ficheux, D., Darlix, J.L. and Mely, Y. (2006) During the early phase of HIV-1 DNA synthesis, nucleocapsid protein directs hybridization of the TAR complementary sequences via the ends of their double-stranded stem. *Journal of molecular biology*, **356**, 1180-1192.
38. Ramalanjaona, N., de Rocquigny, H., Millet, A., Ficheux, D., Darlix, J.L. and Mely, Y. (2007) Investigating the mechanism of the nucleocapsid protein chaperoning of the second strand transfer during HIV-1 DNA synthesis. *Journal of molecular biology*, **374**, 1041-1053.

39. Vo, M.N., Barany, G., Rouzina, I. and Musier-Forsyth, K. (2009) HIV-1 nucleocapsid protein switches the pathway of transactivation response element RNA/DNA annealing from loop-loop "kissing" to "zipper". *Journal of molecular biology*, **386**, 789-801.
40. Cristofari, G. and Darlix, J.L. (2002) The ubiquitous nature of RNA chaperone proteins. *Prog Nucleic Acid Res Mol Biol*, **72**, 223-268.
41. Herschlag, D. (1995) RNA chaperones and the RNA folding problem. *The Journal of biological chemistry*, **270**, 20871-20874.
42. Ward, D.C., Reich, E. and Stryer, L. (1969) Fluorescence studies of nucleotides and polynucleotides. I. Formycin, 2-aminopurine riboside, 2,6-diaminopurine riboside, and their derivatives. *The Journal of biological chemistry*, **244**, 1228-1237.
43. Boulant, S., Vanbelle, C., Ebel, C., Penin, F. and Lavergne, J.P. (2005) Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features. *Journal of virology*, **79**, 11353-11365.
44. Avilov, S.V., Piemont, E., Shvadchak, V., de Rocquigny, H. and Mely, Y. (2008) Probing dynamics of HIV-1 nucleocapsid protein/target hexanucleotide complexes by 2-aminopurine. *Nucleic acids research*, **36**, 885-896.
45. Ben Gaied, N., Glasser, N., Ramalanjaona, N., Beltz, H., Wolff, P., Marquet, R., Burger, A. and Mely, Y. (2005) 8-vinyl-deoxyadenosine, an alternative fluorescent nucleoside analog to 2'-deoxyribosyl-2-aminopurine with improved properties. *Nucleic acids research*, **33**, 1031-1039.
46. Brochon, J.C. (1994) Maximum entropy method of data analysis in time-resolved spectroscopy. *Methods Enzymol*, **240**, 262-311.
47. Livesey, A.K. and Brochon, J.C. (1987) Analyzing the distribution of decay constants in pulse-fluorimetry using the maximum entropy method. *Biophys J*, **52**, 693-706.
48. Bloomfield, V.A., He, S., Li, A.Z. and Arscott, P.B. (1991) Light scattering studies on DNA condensation. *Biochem Soc Trans*, **19**, 496.
49. Hargittai, M.R., Gorelick, R.J., Rouzina, I. and Musier-Forsyth, K. (2004) Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site. *Journal of molecular biology*, **337**, 951-968.
50. Kuzmic, P. (1996) Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Analytical biochemistry*, **237**, 260-273.
51. Cantor, C. and Schimmel, P. (1980) *Biophysical chemistry Part 2: Techniques for the study of biological structure and function*. Academic press; NY.
52. Rouzina, I. and Bloomfield, V.A. (1999) Heat capacity effects on the melting of DNA. 1. General aspects. *Biophys J*, **77**, 3242-3251.
53. Kiviniemi, A. and Virta, P. (2010) Characterization of RNA invasion by (19)F NMR spectroscopy. *Journal of the American Chemical Society*, **132**, 8560-8562.
54. Jean, J.M. and Hall, K.B. (2001) 2-Aminopurine fluorescence quenching and lifetimes: role of base stacking. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 37-41.
55. Guest, C.R., Hochstrasser, R.A., Sowers, L.C. and Millar, D.P. (1991) Dynamics of mismatched base pairs in DNA. *Biochemistry*, **30**, 3271-3279.
56. Ivanyi-Nagy, R., Davidovic, L., Khandjian, E.W. and Darlix, J.L. (2005) Disordered RNA chaperone proteins: from functions to disease. *Cell Mol Life Sci*, **62**, 1409-1417.
57. Rein, A., Henderson, L.E. and Levin, J.G. (1998) Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem Sci*, **23**, 297-301.
58. Tompa, P. and Csermely, P. (2004) The role of structural disorder in the function of RNA and protein chaperones. *FASEB J*, **18**, 1169-1175.

59. Godet, J. and Mely, Y. (2010) Biophysical studies of the nucleic acid chaperone properties of the HIV-1 nucleocapsid protein. *RNA Biol*, **7**, 48-60.
60. Ito, T. and Lai, M.M. (1997) Determination of the secondary structure of and cellular protein binding to the 3'-untranslated region of the hepatitis C virus RNA genome. *Journal of virology*, **71**, 8698-8706.
61. Darlix, J.L., Garrido, J.L., Morellet, N., Mely, Y. and de Rocquigny, H. (2007) Properties, functions, and drug targeting of the multifunctional nucleocapsid protein of the human immunodeficiency virus. *Advances in pharmacology (San Diego, Calif*, **55**, 299-346.
62. Darlix, J.L., Lapadat-Tapolsky, M., de Rocquigny, H. and Roques, B.P. (1995) First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. *Journal of molecular biology*, **254**, 523-537.

Supplementary material

Table A: comparison of experimental values with values obtained with Dynafit software for two-step reaction mechanism

		Experimental values	Values from dynafit
DLS/cDLS annealing	k_{ass} ($M^{-1}s^{-1}$)	456	582
	k_{diss} (s^{-1})	0.018	0.015
	k_f (s^{-1})	0.00044	0.00036
DLS/cDLS annealing + 0.5x Peptide E	k_{ass} ($M^{-1}s^{-1}$)	2297	1777
	k_{diss} (s^{-1})	0.0089	0.0051
	k_f (s^{-1})	0.0054	0.0049
SL2/cSL2 annealing	k_{ass} ($M^{-1}s^{-1}$)	1687	1896
	k_{diss} (s^{-1})	0.0614	0.086
	k_f (s^{-1})	0.0018	0.0021
SL2/cSL2 annealing + 1x Peptide E	k_{ass} ($M^{-1}s^{-1}$)	56088	58373
	k_{diss} (s^{-1})	0.0089	0.010031
	k_f (s^{-1})	0.0187	0.0165

The two-step reaction mechanism described in scheme 1, was validated by the similarities in the values of the kinetic constants obtained by both methods for DLS/cDLS and SL2/cSL2 annealing. Standard deviations for the values obtained from dynafit simulations were lower than 2% in all cases.

General Conclusions

The core protein (179 amino acids) of Hepatitis C virus (HCV) is a multifunctional chaperone protein that consists of a hydrophilic N-terminal domain (Domain D1) and a hydrophobic C-terminal domain (Domain D2). The core is involved in many viral processes like cell proliferation, differentiation, RNA packaging and nucleocapsid formation.

This work was mainly concentrated on the understanding of the molecular chaperone properties of this protein. During these studies, we used three representative peptides of core viz., D1 core domain (Domain D1) and peptides E and F, corresponding to the two (BD2+BD3) and all the three (BD1+ BD2+BD3) basic sub-domains of the N-terminal domain, respectively.

To achieve the first step in understanding the chaperone properties of the core, its nucleic acid binding activity was analyzed by using the synthetic core peptide E, as this peptide was previously shown to mimic the core chaperone properties.

However, binding studies of this peptide with HCV ODNs, by conventional fluorescence techniques, were not possible due to its high-aggregation properties. In binding experiments, the high concentrations of protein need to reach saturation and get binding constants cause aggregation. Thus, the binding parameters of the HCV 3'-X-tail region ODNs with peptide E were analyzed by using Surface Plasmon Resonance technology (SPR), by immobilizing biotinylated-ODNs on a streptavidin coated gold surface and by flushing peptide E as an analyte. For this, two ODN sequences (DLS and SL2), from the first 55nt of the HCV 3'X-tail region, and their mutants were used as ligand. A single binding site was revealed for ODNs which adopt a stable hairpin structure, like DLS with an affinity (K_d) of 0.24 µM. On the contrary, binding of peptide E to unstructured ODNs (DLS mutants) showed a higher binding stoiechiometry as compared to structured ODNs, but with a 4-5 fold decrease in affinity. However, with SL2, another stable hairpin structure, it was observed that peptide E can occupy two binding sites, with an order of magnitude higher affinity as compared to DLS. These binding studies were well supported by single-molecule FRET experiments (smFRET). An apparent dissociation constant $k_d = 1.3 \times 10^{-7}$ M was observed for the DLS/peptide E interaction, similar to the results obtained with SPR experiments. Moreover, the two orders of magnitude difference between the two dissociation constants k_{d1} and k_{d2} values observed with SL2, confirmed the possibility of two different binding sites. These interactions of the core protein with the relevant nucleic acid sequences of HCV give credit on the chaperone activities of the core protein and suggest that peptide E can bind specifically to the structured ODNs that are in abundance in the 3' X-Tail region of HCV genome (Chapter 1).

Next, we characterized the chaperone activity of the core protein on complementary cTAR and dTAR DNA sequences from HIV-1, as a model.

To characterize in depth the nucleic acid chaperone properties of this protein, we first analyzed the experimental conditions where the core peptides (Domain D1, peptide E and peptide F) did not cause DNA/RNA aggregation by means of FCS and observed that these peptides do not cause aggregation till peptide/ODN molar ratio of 1.4:1. It was necessary to find out these non-aggregation conditions as molecular aggregation can cause strong bias when using fluorescence-based techniques in *in vitro* conditions.

Since destabilization of nucleic acid secondary structures is a feature of the chaperone protein activity, we characterized the nucleic acid destabilizing activity of core protein, and observed that core peptides (Domain D1, peptide E and peptide F) have only marginal effect on the secondary structure of cTAR in non-aggregating conditions.

After analyzing these properties, the mechanism by which core protein can promote the annealing of cTAR and dTAR DNA sequences was characterized. It was observed that core protein activates the annealing of these complimentary sequences up to three orders of magnitude, by a reaction mechanism of two parallel kinetic pathways that include two distinct intermediate complexes (ICs). The formation of ICs was found to be nucleated through the stem-ends of cTAR and dTAR. Then, the conversion of these ICs to the final extended duplex was enhanced by core peptides, in contrast to other chaperone proteins like NCp7 from HIV-1. This study confirmed the notion that the chaperone activity of core protein is mediated through its three basic sub-domains of hydrophilic Domain D1 (Chapter 2).

On the basis of these results, we put our efforts in finding the annealing parameters of HCV 3' X-tail region in presence of these core peptides. For, this, DLS and SL2 were reacted to their respective complementary sequences (cDLS and cSL2, respectively). The core peptides were found to enhance the annealing kinetics of both systems, DLS/cDLS and SL2/cSL2, by at least one order of magnitude, by initiating their annealing through the interactions between the complementary loops of ODNs. Next, a single kinetic pathway was identified with a fast pre-equilibrium kissing-loop intermediate that then slowly converts into the final extended duplex.

Our data on the relevant nucleic acid sequences of HCV RNA, DLS and SL2, demonstrated that the core protein exhibits potent *in vitro* nucleic acid chaperone activities, facilitating RNA-RNA and RNA-DNA interactions and structural rearrangements and thus regulate the transitions between translation and replication as well as between replication and packaging of the genomic RNA (Chapter 3).

Future Prospectives

The present work can be extended in the following directions:

- The conformational changes associated to the binding of core peptides to the DLS and SL2 sequences and other sequences of 3'X-tail regions (SL3 and whole X-RNA) can be investigated, by labelling peptides at their N- and C-termini with a donor and acceptor couple (For Example Rhodamine and Fluorescein). Using smFRET experiments, it will be possible to monitor whether the peptide can fold on HCV ODNs. In parallel, NMR experiments could be performed in collaboration with the team of O. Mauffret and P. Fossé to determine the structure of SL2 and that of peptide E and/or D1 domain bound to the high affinity binding site. This should give us important data about the folding of the protein on SL2 (thus substantiating the entropy exchange model) and the conformational changes of SL2 on binding of the core peptide.

- Next, the chaperone activity of core can be checked with the whole protein (Domain 1 + Domain 2), when bound to membranes or lipid droplets by its hydrophobic domain D2. This work can be initiated by first analysing the activity of the hydrophobic C-terminal trans-membrane domain (D2 Domain), as this ~60 amino acid long sequence is easy to synthesize by solid support peptide synthesizing techniques. An extended research program based on the use of the 3-hydroxychromone dyes developed in the laboratory, on Atomic Force Microscopy, FCS and wide-field Total Internal Reflection Fluorescence will be used to characterize the insertion depth, orientation and mobility of the D2 domain in the membrane, as well as its effect on the membrane properties. In a second step, the recombinant full-length core will be produced and its chaperone properties when bound to lipid membranes will be explored by the tools developed in our study. This will enable us to conclude whether the chaperone properties are kept in a more integrated system. The last step would be of course to explore the core chaperone properties in the cellular context. Unfortunately, this appears much more challenging since it requires to get labelled oligonucleotides inside the cells and to be able to discriminate the fluorescence of the chaperoned sequences from the cell background.

Further, the study of inhibitors of the core/HCV 3'X-tail region interaction has not yet been reported. The results obtained in this thesis suggest that inhibition of the ODN annealing properties of core can be a target of clinical interest. For this reason, the development of inhibitors of core-mediated ODN annealing may be valuable. In this respect, the presented experimental procedures can be used to identify compounds with high inhibitory activity *in*

vitro. To this end, fluorescence based assay compatible with high-throughput screening technology can be developed. These assays could be of great help to test a large series of products. Secondly, the selected compounds should be tested on cell-based assays and animal models to characterize their antiviral activity and identify the inhibited step in the viral life cycle.

Appendix

Chapter 1: Kinetic investigation of model DNA annealing in the presence of Human Prion Protein (23-110)HuPrp

Prion diseases are neurodegenerative illnesses associated with the conversion of the cellular prion protein (PrP^C) into the aggregated misfolded isoform, named PrP^{Sc}. Recent studies on the physiological role of PrP^C revealed that this protein has probably multiple functions, notably in cell-cell adhesion and signal transduction. In addition, there are indications that PrP^C assists nucleic acid folding and interactions in a manner similar to cellular and viral nucleic acid chaperones, and could restrict retrovirus replication. In an attempt to better understand the relationships between the cellular prion protein and nucleic acids and its role in nucleic acid metabolism, we have investigated the nucleic acid chaperoning activities of the cellular prion protein *in vitro*. For this, we used a truncated N-terminus part of Human prion protein, named (23-110)HuPrp. It is an 86 amino acid long peptide, containing 7 tryptophan and 2 tyrosine residues. Using fluorescence energy transfer (FRET), we analysed the destablization and nucleic acid annealing activity of (23-110)HuPrp peptide.

Nucleic acid destabilizing properties of HuPrp protein was analysed by using a doubly labelled cTAR by Fluorescein at 5' end and TetramethyRhodamine at 3' end. It was observed that (23-110)Huprp does not exhibit any cTAR destabilization activity (Figure 1a), but induces a fluorescence decrease when added over peptide:ODN ratio of 1:1 (ratio of peptide/cTAR > 1). This decrease was associated to an aggregation of cTAR, as confirmed by the increase in light scattering in the corresponding absorption spectra (Figure 1a and 1b).

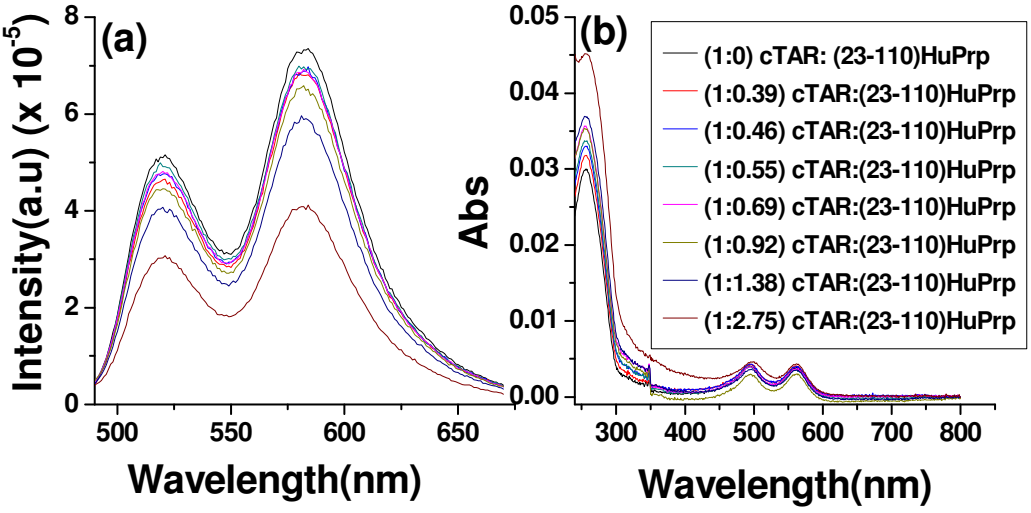


Figure 1: Destabilisation and aggregation of doubly labelled cTAR secondary structure by (23-110) HuPrp

(a) Emission and (b) absorbance spectra of Fl-cTAR-TMR (50 nM) were recorded in the absence (black) and presence of (23-110)HuPrp at different ratios in 25 mM Tris, 30 mM NaCl, and 0.2 mM $MgCl_2$, pH 7.5. Excitation wavelength was 480 nm. Upon continuously adding (23-110)HuPrp, emission spectra (a) shows a decrease in Fluorescein intensity. On the other hand, an increase in light scattering was observed by absorbance profile (b).

On the contrary, (23-110) HuPrp was found to efficiently activate the annealing of dTAR, the DNA analogue of the transactivation response element to its complementary sequence, cTAR DNA from HIV-1.

Annealing dependence on various dTAR concentrations indicated saturation behaviour of both k_{obs1} and k_{obs2} values (Figure 2).

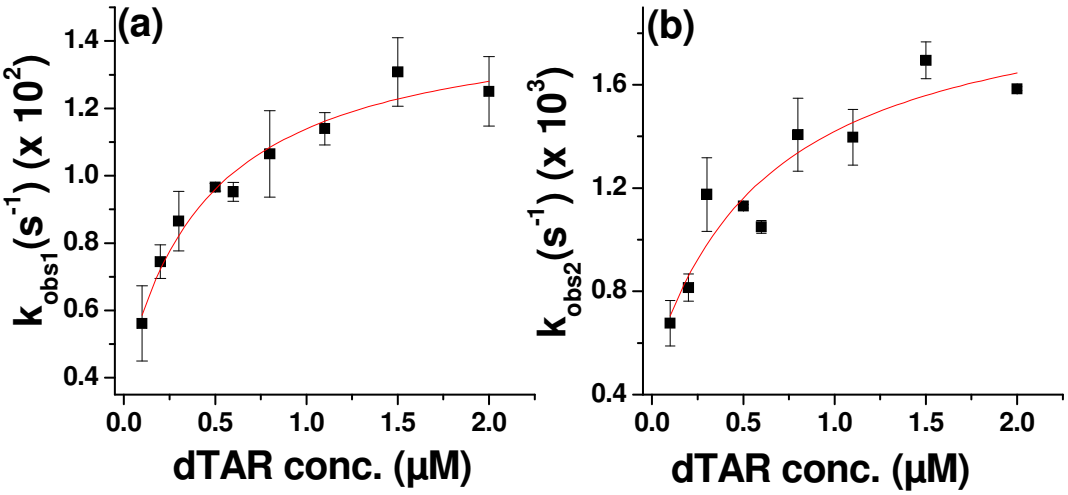


Figure 2: Kinetic parameters of (23-110)HuPrp promoted-cTAR/dTAR annealing. The fast (k_{obs1}) (a) and slow (k_{obs2}) (b) components were determined in pseudo-first order conditions from real-time kinetics. The solid lines correspond to the fit of the data with: $k_{obs} = \frac{k_f K_M [dTAR]}{1 + K_M [dTAR]} + k_b$, where k_{fi} is the forward rate constant for the interconversion of IC into ED, K_{Mi} is the equilibrium binding constant for the intermediate complex, and kbi is the backward rate constant for the conversion of IC into ED, and corresponds to the Y-axis intercept. For both k_{obs1} and k_{obs2}, the kbi values were found to be close to 0.

.

This saturation behaviour is consistent with a two-step reaction scheme as described for HCV core promoted cTAR/dTAR annealing under sub-saturation conditions[1]. Based on the comparison and acquired kinetic data, a reaction mechanism with two parallel pathways can be proposed (scheme 1). In mechanism, both the fast and slow pathways are based on a distinct cTAR/huPrP(23-110) complex ($cTAR_{1,2}$) and involve a rate-limiting interconversion step ($k_{f1,2}$) coupled to a much faster, preceding binding step, governed by an equilibrium constant ($K_{M1,2}$).

Scheme 1

where IC1 and IC2 represent intermediate complexes formed during fast and slow pathway, respectively.

Values of 2.2 (±0.8) × 10^6 M^{-1} and 1.5 (±1)× 10^6 M^{-1} were obtained for K_{M1} and $K_{M2,}$ the equilibrium binding constants of the intermediate complexes associated with the fast and slow kinetic component, respectively. Values of 0.011 (± 0.001) s^{-1} and 0.0015 (± 0.0003) s^{-1} were obtained for the sum of the forward (k_{f1}, k_{f2}) and backward (k_{b1}, k_{b2}) interconversion rate constants associated to the two kinetic components.

The binding constant of the intermediate in the two pathways is about one to two orders of magnitude lower than that with NC(1-55), added at a ratio of 10 peptides per oligonucleotide (K=10^8 M^{-1})[2]. This difference may in part be explained by the incomplete coating of the oligonucleotides by HuPrP, since the binding constant of the intermediate was shown to be strongly dependent on the level of protein coating in the case of NCp7 (with an about 3 orders of magnitude difference between low and full coating). Interestingly, the k_{f1} value was close to the 0.03 s^{-1} value of the corresponding parameter obtained with NC(1-55). The comparison of this parameter for the two proteins is more straightforward, since this parameter was found to be only poorly dependent on the level of NCp7 coating. Thus, HuPrP appears to promote the conversion of the intermediate to the final extended duplex, with the same efficiency than NCp7.

Furthermore, a 2-3fold decrease was observed for K_{M1} and k_{f1} values while a ten-fold decrease was observed for k_{f2} value as compared to HCV core-promoted cTAR/dTAR annealing. In addition to that, the constant value of the amplitudes for the two kinetic components is also in line with the mechanism of the core-promoted annealing. The similar equilibrium constants for the two pathways suggested that the intermediates of both pathways exhibit a similar number of base pairs between cTAR and dTAR. Nevertheless, the intermediates can not be rigorously identical, since otherwise, we should only see one pathway with an interconversion step, governed by the sum of the observed interconversion rates. In fact, the two pathways differ by a ten-fold difference in the kinetics of these interconversion rates. One possible difference could be either the number or the position of the bound proteins on the oligos.

Next, the effect of oligonucleotide sequence and their stability on the (23-110)HuPrp-promoted cTAR/dTAR annealing kinetics was analysed. For this, dTAR was substituted by

the dTAR T-L mutant where the 6 nucleotides of the loop were changed to T residues, thus preventing its base-pairing with the cTAR loop. These nucleotide substitutions did not significantly change the time course of ED formation (Figure 3), indicating that loop-loop interactions do not play a significant role in the (23-110)HuPrp-promoted cTAR/dTAR annealing reaction.

To investigate the role of the cTAR stem in the annealing reaction, we used the cTAR1.2 derivative where bases complementary to the bulged bases at positions 49 and 52 have been introduced in order to stabilize the lower half of the stem. The annealing of this mutant to dTAR in the presence of (23-110)HuPrp E was extremely slow (Figure 3).

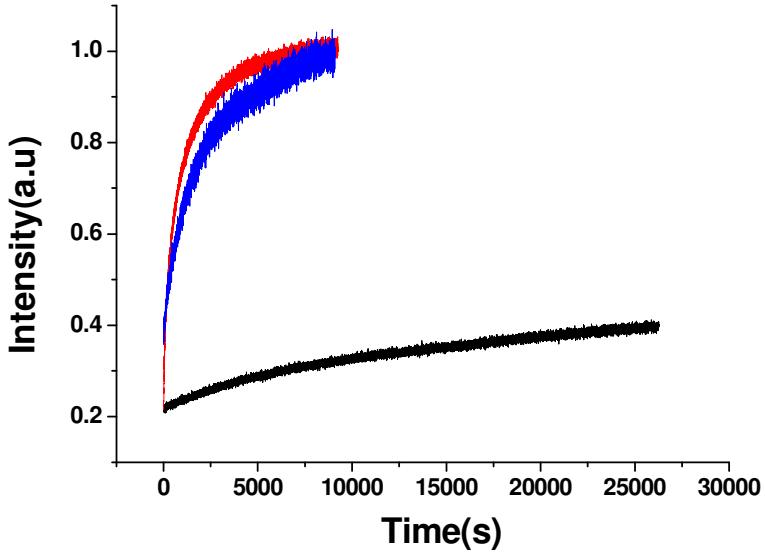


Figure 3: Kinetics of (23-110)HuPrp-promoted annealing of cTAR derivatives with dTAR derivatives. Kinetic traces of 10 nM doubly labelled TMR-cTAR-Fl derivatives with 300 nM non-labelled dTAR derivatives [cTAR/dTAR (red), cTAR/dTAR-TL (blue) and cTAR1.2/dTAR (black)]. (23-110)HuPrp was added at a peptide/oligonucleotide ratio of 1:1 Excitation and emission wavelengths were 480 nm and 520 nm respectively.

This data indicated that both kinetic pathways of the (23-110)HuPrp-promoted cTAR/dTAR annealing reaction are nucleated through the stems of cTAR and dTAR.

The nature of the two pathways was further investigated by analyzing the temperature dependence of the kobs values (Figure D, supplementary materials) using the Arrhenius equation:

$$k_i = A_i \exp\left(E_{a,i} / RT\right) \qquad \text{Equation 1}$$

where the rate constant ki is given by kobsi/[dTAR], Ai is the pre-exponential Arrhenius factor, Ea,i is the activation energy, R is the universal gas constant, and T is the temperature (in Kelvin).

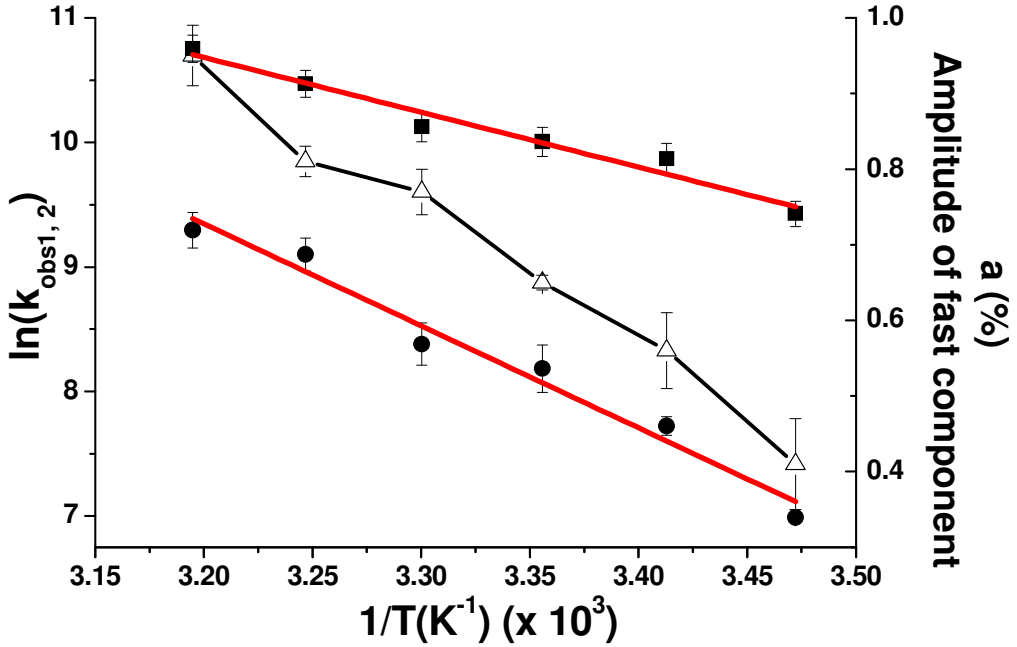


Figure 4: Temperature dependence of cTAR/dTAR annealing kinetics in the presence of (23-110) HuPrp at ratio 1:1. The reaction was performed with 10 nM doubly labelled cTAR and 500-nM non-labelled dTAR. The natural logarithm of the rate constant values for the fast (black squares) and slow (black circles) components as well as the amplitude of the fast component (open tringles) are indicated at six different temperatures. The solid magenta lines are the best fit to Equation (2) with E_{a1} = 8.8(±0.7) kcal/mol and E_{a2} = 17(±1) kcal/mol, for the fast and slow components, respectively.

The temperature dependence of (23-110)HuPrp-promoted cTAR/dTAR annealing provided positive enthalpy values for the transition state of 8.8(±0.7) kcal/mol and 17(±1) kcal/mol for the fast and slow pathways, respectively (Figure 4). These values indicated that cTAR/dTAR annealing promoted by the (23-110) HuPrp peptide involves premelting of approximately 2 and 4 bp, for the fast and slow pathways, respectively (Rouzina I et al., 1999, Cantor C 1980). Further, on the basis of collective data, a schematic reaction mechanism has been proposed and discussed in manuscript.

Next, in search for HuPrP-chaperoning inhibitors, we used a 12-nt long thioaptamer GACACAAGCCGA (GA-12), with five modifications 5' to the dA residues, since it was found to bind with high affinity to Syrian hamster, bovine and human PrP's (Prusiner JMB). We also selected a 11-nt 2' O-methyl oligoribonucleotide GGUUUUUGUGU (mODN) that was shown to be a potent inhibitor of the nucleic acid chaperoning activity of the HIV-1

nucleocapsid protein. We examined the effect of GA-12 and mODN-11, on (23-110)-HuPrp-promoted cTAR/dTAR annealing. Both GA-12 and mODN-11 decreased the time course of (23-110)-HuPrp-promoted cTAR/dTAR annealing, but with a higher efficiency for GA-12, confirming its potential role as a HuPrp inhibitor.

ANALYSIS OF NUCLEIC ACID CHAPERONING BY THE PRION PROTEIN AND ITS INHIBITION BY OLIGONUCLEOTIDES

Cécile Guichard (1)*, Roland Ivanyi-Nagy (1)*,§, Kamal Kant Sharma (2), Caroline Gabus (1), Daniel Marc (3), Yves Mély (2) and Jean-Luc Darlix (1)**

1. Unité de Virologie Humaine INSERM, ENS, IFR 128,
46 allée d'Italie, 69364 Lyon, France.
2. UMR 7213 CNRS, Laboratoire de Biophotonique et Pharmacologie,
Faculté de Pharmacie, Université de Strasbourg, 74 route du Rhin, 67401 Illkirch, France.
3. INRA, Nouzilly, TOURS, France.
* equal contributions; ** corresponding author.
§ Present address: Molecular Parasitology Group, The Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, OX3 9DS, United Kingdom

Abstract.
Prion diseases are unique neurodegenerative illnesses associated with the conversion of the cellular prion protein (PrP^C) into the aggregated misfolded scrapie isoform, named PrP^{Sc}. Recent studies on the physiological role of PrP^C revealed that this protein has probably multiple functions, notably in cell-cell adhesion and signal transduction, and in assisting nucleic acid folding. In fact, *in vitro* findings indicated that the human PrP (huPrP) possesses nucleic acid binding and annealing activities, similarly to nucleic acid chaperone proteins that play essential roles in cellular DNA and RNA metabolism. Here we show by fluorescence resonance energy transfer (FRET) that a peptide, representing the N-terminal domain of huPrP, facilitates nucleic acid annealing. We also show that PrP of human or ovine origin facilitates DNA strand exchange, ribozyme-directed cleavage of an RNA template and RNA *trans*-splicing in a manner similar to the nucleocapsid protein of HIV-1. In an attempt to characterize inhibitors of PrP-chaperoning *in vitro* we discovered that the thioaptamer 5'-GACACAAGCCGA-3' was extensively inhibiting the PrP chaperoning activities. At the same time a recently characterized methylated oligoribonucleotide inhibiting the chaperoning activity of the HIV-1 nucleocapsid protein was poorly impairing the PrP chaperoning activities.

Key words: prion protein, nucleic acid chaperoning, oligonucleotides, inhibition

Introduction.

Transmissible spongiform encephalopathies (TSEs) such as Creutzfeldt-Jakob disease (CJD), kuru and fatal familial insomnia (FFI) in humans, scrapie in sheep and bovine spongiform encephalopathy (BSE) in cattle are a group of fatal neurodegenerative diseases (1-3). A major molecular characteristic of TSEs is the accumulation of a misfolded, aggregated, partially protease-resistant prion protein, named PrPres, in the central nervous system (CNS) (1-3). Accumulation of PrPres appears to take place by recruitment and templated transconformation of the normal cellular prion protein (PrP^C) by PrPres (4-7), and is thought to induce functional damages to the central nervous system. In support of this, generation of spongiform encephalopathy requires the presence of both PrPres and PrP^C since mice devoid of PrP^C are resistant to challenge with the infectious prion agent (8).

The cellular prion protein (PrP^C) is highly conserved in mammals and abundantly expressed in cells of the nervous and lymphoreticular systems but its physiological role has remained for a long time a matter of speculation (2,9). In fact, PrP null mice were found to develop and reproduce normally (10), or manifested only subtle phenotypic effects [reviewed in (11)], suggesting that PrP^C has no essential function(s) that could not be compensated by proteins with overlapping activities. Nevertheless, several functions have been proposed for PrP^C, including superoxide dismutase activity (12,13), participation in copper metabolism (14), signal transduction (15) and neuroprotection [(16,17), and references therein]. Recent studies revealed that PrP^C is also involved in cell-cell adhesion and in signal transduction mediated by Src-related kinases in the zebrafish animal model (18). In addition, there are clear indications that PrP^C assists nucleic acid folding and interactions in a manner similar to cellular and viral nucleic acid chaperones *in vitro* (19-21), and could restrict retrovirus replication (22,23).

In fact, there are many different nucleic acid-binding proteins that recognize DNA and RNA with a broad sequence specificity in any given cell. Among these ubiquitous nucleic acid binding proteins (NABP) there exists a class named nucleic acid chaperones, which provide assistance to the folding of DNA and RNA by preventing and resolving misfolding, and by chaperoning DNA/RNA interactions (24,25). Thus nucleic acid chaperones are considered to be essential co-factors for many basic biological processes including nucleic acid

maintenance, RNA splicing, transport and translation (24,25) and PrP^C would be one of these proteins (19-21).

In an attempt to better understand the relationship between the cellular prion protein and nucleic acids and its possible role in nucleic acid metabolism, we have investigated the nucleic acid chaperoning activities of the recombinant human and ovine PrP *in vitro*. Using fluorescence resonance energy transfer (FRET), we report that a prion peptide representing the N-terminal domain of huPrP is very active in facilitating nucleic acid annealing. Furthermore, by using established nucleic acid chaperoning assays, notably DNA strand transfer (26), ribozyme cleavage of an RNA substrate (27,28) and RNA *trans*-splicing (29,30), we show that the human and ovine prion proteins are potent nucleic acid chaperones. Recently we have been able to characterize a potent oligonucleotide (ODN) inhibitor of the nucleic acid chaperoning activity of the HIV-1 nucleocapsid protein (NC) *in vitro* and its inhibitory effect on virus replication in primary human cells (31). This prompted us to search for ODNs capable of inhibiting the nucleic acid chaperoning activity of PrP. We discovered that a previously identified 5'-GACACAAGCCGA-3' thioaptamer binding to Syrian hamster (SHa) and human PrP (32) was a potent inhibitor of PrP-chaperoning activity *in vitro*, while the non-modified ODN was not, and a methylated ODN specific for HIV-1 NC was slightly inhibitory.

Materials and Methods

Recombinant proteins. Recombinant human PrP (huPrP from residues 23-231), the N-terminal region [huPrP(23-145)] and the C-terminal region [huPrP(122-231)] were expressed in *E. coli* and purified to homogeneity (19). The N-terminal region 23-110 of huPrP was synthesized by fmoc chemistry and purified to homogeneity by HPLC (19). The ovine PrP (ovPrP, residues 25–234) was produced in *E. coli* and purified to homogeneity (33). HIV-1 nucleocapsid protein NCp7 and NC(12-53) lacking the N- and C-terminal regions were obtained as pure proteins as previously described (19,34). Proteins were dissolved at 1 mg/ml in buffer containing 30 mM Hepes pH 6.5, 30 mM NaCl and 0.1 mM $ZnCl_2$. HnRNP A1 and YB-1/p50 were provided by Christiane Branlant (France) and Lev Ovchinnikov (Russia), respectively.

Plasmid DNAs and RNAs. Plasmids pS14, pS20 and pR3, for the ribozyme asssays, were provided by E. Bertrand (Montpellier) (28) and plasmids H1 and H2 for the *trans*-splicing

assays by Renée Schroeder (Vienna, Austria). All plasmid DNAs were amplified in *E. coli* Rec A- cells and purified by affinity chromatography (Qiagen, USA). H1 DNA (549 nt of exon 1 and 131 nt of the 5' part of the intron) was linearized with SalI and H2 DNA (147 nt of the 3' half of the intron and 23 nt of exon 2) was linearized with BamHI, then transcribed. Templates pS14, pS20 and pR3 were digested by Pst I, treated by Klenow polymerase to remove the 3' strand overhang, and then the substrate RNA and the ribozyme were generated by *in vitro* transcription with modifications: for substrate RNA UTP was at 10 □M and 50 □Ci ^{32}P-UTP (Amersham, UK) was added. For the ribozyme, UTP was at 100 □M with 10 □Ci of ^{32}P-UTP. For the *trans*-splicing assays, RNAs H1 and H2 were prepared by *in vitro* transcription with T7 RNA polymerase according to the manufacturer's instructions (Promega) and labelled by incorporation of ^{35}S-UMP during transcription.

Following RNA synthesis, the DNA template was removed by treatment with RNase-free DNase I (Promega) for 20 min at 37 °C, followed by phenol extraction and ethanol precipitation. All RNAs were purified by 8% PAGE in 7 M urea, 0.5x TBE. RNAs were recovered by elution in 0.3 M sodium acetate, 0.1% SDS, for 4 h at 37 °C and ethanol precipitated. RNAs were dissolved in sterile H_2O and their integrity was verified by PAGE-urea.

TAR DNA. ODNs used for DNA annealing corresponded to the HIV-1 TAR sequences, in the sense and anti-sense orientations, respectively. ODNs were purchased from Eurogentec (Belgium). TAR ODNs are 56 nt in length.

TAR(+) (sense):

5'-GGTCTCTCTTGTTAGACCAGGTCGAGCCCGGGAGCTCTCTGGC

TAGCAAGGAACCC-3';

TAR(–) (anti-sense):

5'-GGGTTCCTTGCTAGCCAGAGAGCTCCCGGGCTCGACCTGGT

CTAACAAGAGAGACC-3';

ODNs used for DNA strand transfer assays corresponded to the HIV-1 repeated R sequences, in the sense and anti-sense orientations, already described in (35,36).

TAR(–) and R(+) ODN were ^{32}P-labelled with 50 µCi of ^{32}P-ATP using T4 polynucleotide kinase. Labelled ODNs were purified by 10% PAGE, 7 M urea in 50 mM Tris-borate, 1 mM EDTA, pH 8.3 (0.5x TBE) and recovered, ethanol precipitated and dissolved in sterile H_20 before use.

Oligonucleotides.

The 12-nt long thioaptamer (GA-12; 5'-GACACAAGCCGA-3'), with five modifications 5' to the dA residues, mODN-11 (5'-GGUUUUUGUGU-3' with a 2'O-methyl modification at each residue) and ODN-11 (5'-GGTTTTTGTGT-3', without modifications) were obtained in a highly pure form from Eurogentec (Belgium). Ct-12 corresponds to GA-12 without the modifications.

Destabilization assay. To monitor the nucleic acid destabilizing properties of huPrP, cTAR DNA was labelled either by carboxytetramethylrhodamine (TMR) or ethyl 2-[3-(ethylamino)-6-ethylimino-2,7-dimethylxanthen-9-yl]benzoate hydrochloride (Rh6G) at its 5' end, while the 3' terminus was labelled with either 4-(4'-dimethylaminophenylazo) benzoic acid (Dabcyl) or 5(and 6)-carboxyfluorescein (Fl). The dyes form a nonfluorescent heterodimer when the cTAR stem is closed, while melting of the stem restores the fluorophore fluorescence. Thus, the destabilizing ability of HuPrP peptide (23-110) was evaluated from the ratio of the fluorescence intensity in the presence versus the absence of the peptide. Experiments were performed by adding huPrP(23-110) at increasing peptide to oligonucleotide ratio to 50 nM doubly labelled cTAR, in 25 mM Tris, 30 mM NaCl, and 0.2 mM $MgCl_2$, pH 7.5.

Kinetics of cTAR-dTAR annealing with the huPrP peptide (23-110). The real-time annealing kinetics of cTAR with dTAR was investigated by mixing Rh6G-5'-cTAR-3'-Dabcyl with an excess of nonlabeled dTAR in conditions already published (35,37). Formation of the 55 bp cTAR/dTAR extended duplex (ED) strongly increases the interchromophore distance, leading to a full recovery of Rh6G emission. The dependence of cTAR-dTAR annealing kinetics on temperature was performed at 6 different temperatures, viz. 15, 20, 25, 30, 35, and 40 °C, as previously described (37).

The TAR(–)/TAR(+) annealing assay. Tar(+) and ^{32}P-Tar(–) ODNs were incubated (0.03 pmol each) in 10 µl of buffer A (35,37) in the presence of increasing protein concentrations. Reactions were performed at 37 °C for 5 min except for the positive control which was incubated at 65 °C. To stop the reaction and denature the protein, we added 5 µl of a solution containing 20% glycerol, 20 mM EDTA pH 8.0, 2% SDS, 0.25% bromophenol blue and 0.4 mg/ml calf liver tRNA. Samples were resolved by 8% native PAGE in 50 mM Tris-Borate pH

8.3, 1 mM EDTA at 4 °C. Subsequently, gels were autoradiographed and the amounts of labelled single-stranded and double-stranded DNA were assessed by PhosphorImaging.

Effect of inhibitors on the kinetics of cTAR-dTAR annealing with the N-terminal huPrP. Inhibition of huPrP(23-110)-promoted cTAR-dTAR annealing was performed by adding a five-fold molar excess of GA-12, mODN-11 or Ct-12 relative to cTAR/dTAR. The cTAR (or dTAR)-huPrP(23-110)-ODN mixtures were incubated for 5 min to ensure that the binding equilibrium was reached. Then, cTAR-dTAR hybridisation was initiated by manual mixing of cTAR-huPrP(23-110)-ODN mixture with dTAR-huPrP(23-110)-ODN mixture. To avoid the effects of local concentration fluctuations on the annealing kinetics, equal volumes of the mixtures were used. Experiments were performed in pseudo-first order conditions by mixing 10 nM of doubly labelled cTAR with 300 nM of non-labelled dTAR in presence of huPrP(23-110) at a peptide/oligonucleotide molar ratio of 1:1. To this mixture, inhibitors were added in increasing concentrations (from 0.1 to 10-fold) to follow their concentration dependent inhibitory effect.
All experiments were performed in 25 mM Tris-HCl (pH 7.5), 30 mM NaCl, 0.2 mM $MgCl_2$ at 20°C.

DNA strand transfer assay. ^{32}P-labelled R(+) wt, non-labelled R(-) 3'-modified and R(-) wt were heat denaturated (2 min at 90 °C) and chilled on ice. All components were kept on ice. 0.03 pmole each of R(+) wt and R(-) 3'-modified were mixed with reaction buffer to a final concentration of 20 mM Tris-HCl, pH 7.0, 30 mM NaCl, 0.1 mM $MgCl_2$, 10 µM $ZnCl_2$ and 5 mM DTT in 5 µl final volume, incubated for 30 min at 65 °C and chilled on ice. Subsequently, 0.03 pmole R(-) wt was added together with the chaperone protein at a final protein to nucleotide molar ratio as indicated in figure legends. Reactions were left to proceed for one to five min at 37 °C. The mixtures were then chilled on ice and reactions stopped with 2.5 µl of 20 % glycerol, 20 mM EDTA pH 8.0, 0.2% SDS, 0.25% bromophenol blue and 0.4 mg/ml calf liver tRNA. Samples were resolved by 6% native PAGE in 50 mM TBE at 4 °C. The level of strand exchange as a function of time was quantified by PhosphorImaging

Hammerhead Ribozyme cleavage assay. Ribozyme and substrate RNAs were independently heated for 1 min at 90 °C in H_2O. The reaction buffer was added to yield final concentrations of 5 mM $MgCl_2$, 100 mM NaCl, 20 mM Tris-HCl, pH 7.5. After slowly cooling down to 37 °C, RNAs were further incubated for 5 min at 20 °C. 0.1 pmole of ribozyme and 0.02 to 2

pmoles of RNA substrate were then combined in a final volume of 10 □l, each protein was added at final concentrations as indicated in the figure legend and incubations were for 1 to 30 min at 37 °C. Reactions were terminated by adding 20 □l of stop solution (0.3% SDS, 15 mM EDTA), and RNAs were extracted with 30 □l of phenol and 15 □l of chloroform. The aqueous phase was precipitated with ethanol and the pellet resuspended in 45% formamide, 0.5x TBE, and 0.1% dye. RNAs were then analyzed on 8% PAGE in TBE. All quantitative measurements were done by PhosphorImaging.

***Trans*-splicing assay.** $2x10^{-8}$ M each of ^{35}S-labelled RNAs were incubated for 1 min at 95 °C in H_2O and cooled down to 37 °C. Then, the splicing buffer (40 mM Tris-Cl at pH 7.4, 3 mM $MgCl_2$, 0.4 mM spermidine, 4 mM DTT), 8 U RNasin (Promega) and ^{32}P-GTP were added with the nucleic acid chaperone, at protein to nucleotide molar ratios as indicated in the figure legend, at 37 °C for 30 minutes in 10 µl. Reactions were stopped by adding a final concentration of 40 mM EDTA and 300 µg/ml of tRNA. Proteins were removed by phenol-chloroform extraction, and nucleic acids precipitated by ethanol. Samples were resuspended in 10 µl formamide buffer (97 % formamide, 1 mM EDTA, 0.02 % bromophenol blue, 0.02 % xylene cyanol), denatured for 2 min at 95 °C and resolved by denaturing 5% PAGE-7 M Urea in 0.5x TBE. Subsequently, the gel was fixed, dried and autoradiographed. 5' ^{32}P-labelled FX174 DNA HinfI markers (Promega) were used for size determination (not shown). Bands corresponding to the product of the first step of splicing (G-I1) were measured by PhosphorImager.

RESULTS

In order to examine the nucleic acid chaperoning (24,25,38) properties of the prion protein (PrP), we have used advanced assays [reviewed in (39)] that have been developed for the characterization of key chaperone proteins implicated in the metabolism of cellular RNAs such as hnRNP A1 and FMRP proteins (28,40-42), and of retroviral RNA, like the NC protein of retroviruses [(26,28,42), reviewed in (43,44)]. Although the details of their mechanism of action is incompletely understood, nucleic acid chaperones are thought to facilitate RNA (and DNA) conformational rearrangements by a variety of ATP-independent mechanisms, including destabilization of preformed helices, charge neutralization, and matchmaker activity (24,25,45,46).

Firstly, we have analyzed the helix destabilizing activity of the N-terminal peptide (23-110) of huPrP. Using the doubly labelled cTAR DNA stem-loop (SL), there was no evidence of helix destabilizing activity of huPrP(23-110) in the present conditions, as observed by fluorescence resonance energy transfer (FRET) (Figure 1A, supplementary materials). This differs from HIV-1 NC protein, which shows significant helix destabilizing activity (47-49) (data not shown), but is similar to the behavior of the core protein of hepatitis C virus (HCV), another well-characterized RNA chaperone with limited cTAR melting activity (37). Next, we monitored the impact of huPrP(23-110) on the annealing kinetics of cTAR/dTAR by FRET.

Kinetics of cTAR-dTAR annealing facilitated by the N-terminal peptide of huPrP

The real-time annealing kinetics of cTAR to dTAR was investigated in pseudo-first order conditions by mixing Rh6G-5'-cTAR-3'-Dabcyl with an excess of nonlabelled dTAR (35,37). Formation of the 55 bp cTAR/dTAR extended duplex (ED) strongly increases the interchromophore distance, leading to a full recovery of Rh6G emission. In the absence of huPrP(23-110), the annealing of cTAR with dTAR was very slow, involving two distinct kinetic components with second-order rate constants. In the presence of huPrP(23-110) at a peptide/TAR ratio of 1/1, the same fluorescence plateau was observed as in the absence of peptide, indicating that ED formation went to completion. The huPrP(23-110) strongly accelerated the annealing reaction, since it was completed in about 30 min, instead of more than 1 day without the peptide. An adequate fit of the annealing kinetic traces was obtained using a bi-exponential function:

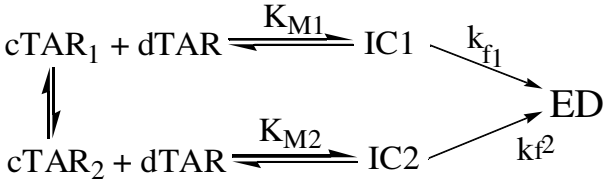
$$I(t) = I_f - (I_f - I_0)\left[a\, e^{(-k_{obs1}(t-t0))} - (1-a)\, e^{(-k_{obs2}(t-t0))}\right] \quad \textbf{Equation 1}$$

where t0 is the dead time, $k_{obs1,2}$ are the observed kinetic rate constants, a is the amplitude of the fast component, and I_0 and I_f are the fluorescence intensities of the stem-loop and the ED, respectively.

We already observed a fast and full fluorescence increase at the lowest tested oligonucleotide concentrations (10 nM cTAR and 100 nM dTAR), suggesting that the peptide probably binds to the TAR sequences with a high affinity (> 10^7 M^{-1}). Indeed, if the binding constant would be lower than 10^6 M^{-1}, and if we assume that, as for other nucleic acid chaperones, PrP does not exhibit any catalytic activity, most of the oligonucleotides would be free and would anneal at the low speed observed for cTAR and dTAR in the absence of protein.

Figure 1

Annealing as a function of the dTAR concentration indicated a saturation behaviour of both k_{obs1} and k_{obs2} values as well as a constant value for their amplitudes (Figure 1B, Supplementary materials) . Moreover, the values of the intercept for both k_{obs1} and k_{obs2} curves were very low, indicating that huPrP(23-110) is unable to dissociate the ED. Based on these kinetic data, a reaction mechanism with two parallel pathways can be proposed (scheme 1), as for the HCV core-promoted cTAR/dTAR annealing (37). Both the fast and slow pathways are based on a distinct cTAR/huPrP(23-110) complex ($cTAR_{1,2}$) and involve a rate-limiting interconversion step ($k_{f1,2}$) coupled to a much faster, preceding binding step, governed by an equilibrium constant ($K_{M1,2}$).



Scheme 1

IC1 and IC2 represent intermediate complexes formed during the fast and slow pathway, respectively. The values of the $K_{M1,2}$ and $k_{f1,2}$ parameters (Table 1) determined from the fits of Fig. 1A (supp informations) were found to be close to the corresponding values reported for the HCV core protein in the same system, strengthening the similarity between the mechanisms of huPrP(23-110)- and HCV core-promoted cTAR/dTAR annealing reaction.

Table 1: Comparison of the kinetic parameters of the huPrP(23-110)- and HCV core-promoted cTAR/dTAR annealing [a]

	K_{M1} (M^{-1}) (´10^{-6})	K_{M2} (M^{-1}) (´10^{-6})	k_{f1} (s^{-1}) (´10^{2})	k_{f2} (s^{-1}) (´10^{2})
HuPrP	2.2 (±0.8)	1.5 (±1)	1.1 (±0.1)	0.15 (±0.03)
HCV Core [b]	4 (±2)	1.4 (±0.7)	4.5 (±0.9)	1.6 (±0.2)

[a] The equilibrium and kinetic constants for HuPrP-promoted cTAR/dTAR annealing reaction are calculated, as described in Figure B of supplementary materials.
[b.] Data from (37) in the same buffer, with a core/oligonucleotide molar ratio of 1.4.

To further characterize the two pathways, the effect of the sequence and stability of the oligonucleotides on the huPrP(23-110)-promoted cTAR/dTAR annealing kinetics was analysed. In this respect, dTAR was first substituted by the dTAR T-L mutant where the six

nucleotides of the loop were changed to T residues, thus preventing its base-pairing with the cTAR loop. These nucleotide substitutions did not significantly change the time course of ED formation, indicating that loop-loop interactions do not play a significant role (Figure 1B, supplementary materials). Further, to investigate the role of the cTAR stem in the annealing reaction, we used the cTAR1.2 derivative where bases complementary to the bulged bases at positions 49 and 52 have been introduced in order to stabilize the lower half of the stem. The annealing of this mutant to dTAR in the presence of huPrP(23-110) was extremely slow (Figure 1B, supplementary information), indicating that both kinetic pathways of the huPrP(23-110)-promoted cTAR/dTAR annealing reaction are nucleated through the stems of cTAR and dTAR.

The nature of the two pathways was further investigated by analyzing the temperature dependence of the k_{obs} values (Figure 1C supplementary materials) using the Arrhenius equation:

$$\kappa_i = \mathcal{A}\exp^{E_{\alpha,i}/PT} \quad \textbf{Equation 2}$$

where the rate constant k_i is given by k_{obsi}/[dTAR], A_i is the pre-exponential Arrhenius factor, $E_{a,i}$ is the activation energy, R is the universal gas constant, and T is the temperature (in Kelvin).

The temperature dependence of huPrP(23-110)-promoted cTAR/dTAR annealing provided positive enthalpy values for the transition state of 8.3(±0.7) kcal/mol and 16.5(±1) kcal/mol for the fast and slow pathways, respectively. These values indicated that cTAR/dTAR annealing promoted by the huPrP(23-110) peptide involves premelting of approximately 2 and 4 bp, for the fast and slow pathways, respectively (50,51). Moreover, the amplitude of the fast component was found to continuously increase with temperature, as observed for the HCV core (37).

Activation of DNA strand exchange by recombinant PrPs.

The strand exchange assay (26) is designed to evaluate the ability of the chaperone protein to facilitate formation of the most stable nucleic acid conformation. It utilizes three DNA molecules, namely the wild type plus strand, (+) wt (^{32}P-labelled), the wild type minus strand, (-) wt, and the mutated minus strand, (-) mut. DNA molecules (+) wt and (-) wt are

complementary, whereas (-) mut is only partially complementary to (+) wt (see scheme in fig. 2A). DNA molecules (+) wt and (-) mut are first heat annealed, giving rise to a double stranded molecule containing mismatches. Then, DNA (-) wt and either PrP, or NCp7 as the positive control, were added to the preformed double stranded DNA [(+) wt:(-) mut] molecules and incubated for 5 min at 37 °C (see materials and methods). Last, nucleic acids were deproteinized and analyzed by PAGE under native conditions. The strand exchange activity of the chaperone protein facilitates formation of the perfect double stranded DNA [(+) wt:(-) wt] molecules at the expense of [(+) wt:(-) mut] molecules with mismatches (fig. 2A). If the recombinant PrPs have general nucleic acid chaperoning activities, they should facilitate formation of the most stable duplex and thus activate the substitution of DNA (-) wt for (-) mut in the double stranded nucleic acid molecule (Fig. 2A).

The ^{32}P-labelled DNA (+) wt is shown in lane 1 while control double-stranded DNA [(+) wt : (-) wt] and [(+) wt : (-) mut] are shown in lanes 3 and 4, respectively (Fig. 2B). Addition of DNA (-) wt to double stranded [(+) wt : (-) mut] and incubation for 30 min at 0 °C (lane 5) or 37 °C (lane 6) did not result in any strand exchange. Addition of HIV-1 NCp7 together with DNA (-) wt resulted in a nearly complete strand exchange under the conditions where NCp7 to nt molar ratios were 1 per 12 nt and 1 per 6 nt (lanes 7 and 8, respectively). Mutant NC(12-53) lacking the N- and C-terminal amino-acids has only limited strand exchange activity (lanes 9 and 10). HuPrP was found to strongly activate the strand exchange (lanes 11 and 12). A polypeptide representing the N-terminal region of huPrP, huPrP(23-145), also exhibited a strong strand exchange activity (lanes 13 and 14), whereas the polypeptide huPrP(122-231), representing the C-terminal region of huPrP, had little activity (lanes 15 and 16). The recombinant ovine PrP (ovPrP) behaved in a manner similar to huPrP, showing a strong strand exchange activity (lanes 17 and 18) at protein to nt ratios of 1:12 and 1:6.

For kinetic analysis, DNA strand exchange assays were performed at 30 °C from one to several min with NC, huPrP and ovPrP proteins at a protein to nt molar ratio of 1 per 12 nt (see materials and methods). For the wild type NCp7, the exchange rate at 30 °C was 0.031 mole of DNA exchanged per mole of protein per min. For the mutant NC(12-53), this value was reduced about 9 times, but still was about three times higher than that observed in the absence of NC protein. The exchange rates for the recombinant prion proteins were very close to that for NCp7 (0.025, 0.03 and 0.03 mole of DNA per mole of protein per min at 30 °C for

huPrP, huPrP(23-145) and ovPrP, respectively. For huPrP(122-231), the value was about 10 times lower than that for huPrP(23-145).

Strand exchange assays have also been carried out with cellular hnRNP A1 and p50, two well characterized RNA chaperone proteins (40-42). Both proteins proved to facilitate the strand exchange reaction, although at a rate about 3 times lower than that observed under the present experimental conditions for huPrP (data not shown).

Figure 2 A, B

Facilitation of ribozyme-directed cleavage of an RNA substrate by PrPs.

Next, we used the hammerhead ribozyme-directed cleavage of an RNA substrate to examine both the RNA annealing and unwinding activities of nucleic acid chaperones (27,28). Nucleic acid chaperones are thought to enhance the rate of ribozyme cleavage by activating the annealing of the substrate RNA to the ribozyme (Fig. 3A step 1) and the unwinding and release of the cleaved RNA products (Fig. 3A step 3), thus allowing recycling of the ribozyme. The ribozyme cleavage assay aims at examining whether PrP accelerates ribozyme cleavage of an RNA substrate in a manner similar to the viral NCp7 chaperone (see Fig. 3A) (27,28). We selected the R3 hammerhead ribozyme and two RNA substrates, namely S14, with a 14 nt substrate-ribozyme duplex length (7 nt either side of the cleavage site) and S20, with 10 nt either side of the cleavage site (28). The above RNA substrate model system was chosen due its likely biological relevance as evidenced by the similarity of data obtained both *in vitro* and in cultured cells (28). ^{32}P-labelled RNA S14, the ribozyme and PrP were mixed and incubated for 30 min. RNAs were deproteinized, recovered and analyzed by PAGE under denaturing conditions. In the absence of a chaperone, ribozyme-directed cleavage of the RNA substrate occurred only slowly at 37 °C (fig. 3B, lanes 1 at 4 °C and 2 at 37 °C; ^{32}P-RNA substrate is S14 and cleaved product □S14). In agreement with previous reports (28), HIV-1 NCp7 facilitated ribozyme cleavage of RNA S14 to reach completion at a concentration as low as $2x10^{-8}$ M (lanes 3-5). On the other hand, NC(12-53) was relatively inactive even at a concentration above $1x10^{-7}$ M (lanes 6-8). Interestingly, PrP of ovine or human origin showed strong enhancement of ribozyme cleavage at $2\text{-}4x10^{-8}$ M (lanes 9-11 and 12-14, respectively). The N-terminal region of huPrP was very active (lanes 15-17 and 18-20) while the C-terminal region was very poorly active (lanes 21-23).

We also examined the effect of PrP using the RNA substrate S20, capable of forming an extended duplex of 20 nt with the hammerhead ribozyme (Fig. 3A) which precludes activation of ribozyme cleavage by HIV-1 NCp7 or hnRNP A1 (28). In the absence of NCp7 or PrP, only minimal ribozyme-directed cleavage of RNA S20 was observed at 37 °C (Fig. 3C, lane 2; ^{32}P-RNA substrate is S20 and product □S20) as seen with RNA S14 (Fig. 3B, lane 2). The enhancement of ribozyme cleavage of RNA S20 by NCp7 remained very modest even at NCp7 concentration of $4x10^{-8}$ M (Fig. 2C; lanes 3-4), in agreement with reported data (28). Mutant NC(12-53) was also very poorly active (lanes 5-6). The ovPrP and huPrP also exhibited very little, if any, activity using RNA S20 (lanes 7-8, 9-10, 11-12 and 13-14, for ovPrP, huPrP, huPrP(23-145) and huPrP(122-231), respectively).

Kinetic assays of ribozyme cleavage of RNA S14 were performed with NCp7, huPrP and ovPrP at a concentration of $4x10^{-8}$ M and cleavage monitored for up to 30 min. Initial rates were 10-12 □moles of RNA S14 cleaved at 37 °C per min per mole of NCp7 and ovPrP. For huPrP, the observed rate was 3-fold slower, while it was approximately 100-fold slower for NC(12-53) at concentrations above $1x10^{-7}$ M, although 2-3 times above rates obtained in the absence of NC or PrP. Similar results were obtained at huPrP or NC concentration of $1x10^{-7}$ M (data not shown).

Figure 3A-3C

HuPrP and ovPrP facilitate RNA *trans*-splicing.

To further examine the RNA chaperoning activity of the mammalian PrPs, we used the previously established *trans*-splicing assay (29,30) (Fig. 4A), where the pre-mRNA of the thymidylate synthase (td) gene containing a group I intron was split into two halves. The first RNA transcript, H1, corresponds to the 5' exon sequence of 549 nt and 131 nt of the intron while the second RNA transcript, H2, represents the 3' part of the intron (147 nt) and part of exon 2 (23 nt) (Figure 4A). The two RNAs were ^{35}S-UMP labelled during transcription and were incubated together. Reaction was started by adding ^{32}P-GTP, so that the resulting spliced RNA was doubly labelled, internally and at the 5' end with ^{32}P-GTP (Figure 4A). In the absence of protein, the reaction was carried out at 55 °C to allow a productive interaction between H1 and H2 RNAs. *Trans*-splicing was indeed found to take place at 55 °C but only to a limited extend [data not shown and (36)] but not at 37 °C according to the accumulation of ^{32}P-GTP-I1 final product (lane 1 in Figure 4B and C). As previously reported (36) HIV-1

NCp7 strongly stimulated *trans*-splicing (lanes 2–5), with an optimal enhancement at a concentration of $8x10^{-7}$ M (panel B, lane 4). Mutant NC(12–53) was found to be poorly active even at $2x10^{-6}$ M (panel B, lane 7).

The recombinant huPrP was found to be a strong activator of *trans*-splicing (Figure 4B, lanes 8-11), with an optimal activation at a concentration of $4x10^{-7}$ M (lanes 10 and 11). Interestingly the N-terminal huPrP peptide (23-144) also activated *trans*-splicing (Figure 4B, lanes 9–12) at concentrations of $2–4x10^{-7}$ M (lanes 13 and 14) while the C-terminal domain (123-231) was clearly inactive (lanes 16 and 17) under the present conditions. The ovine PrP was already active at a concentration of $2x10^{-7}$ M (lanes 7-11 in panel C) and the N-terminal huPrP peptide (23-110) was as active as the full length huPrP (compare lanes 2-6 and 12-16 in panel C).

Figure 4

Inhibition of huPrP chaperoning activity by a thioaptamer.

In a search for PrP-chaperoning inhibitors, we used a 12-nt long thioaptamer (GA-12; GACACAAGCCGA), with five modifications 5' to the dA residues, since it was found to bind with high affinity to Syrian hamster, bovine and human PrPs *in vitro* (32). We also selected a 11-nt 2' O-methyl oligoribonucleotide (mODN-11; GGUUUUUGUGU) because it was shown to be a potent inhibitor of the nucleic acid chaperoning activity of the HIV-1 nucleocapsid protein (31).

We examined the ability of these two selected oligonucleotides, namely GA-12 and mODN-11, and of a control ODN with the non-modified sequence GACACAAGCCGA (Ct-12) to inhibit the annealing activity of huPrP (Fig. 5 A&B). As shown in lanes 7-9, only GA-12 at 100-200 nM was able to completely inhibit huPrP annealing activity, while mODN-11 (lanes 4-6) and ODN Ct-12 (lanes 10-12) were poorly active. The impact of such ODNs on the hammerhead ribozyme directed cleavage of an RNA substrate was monitored under the same conditions. Again the thioaptamer GA-12 at 50-100 nM was found to extensively inhibit the RNA chaperoning of huPrP (lanes 7-9) while mODN-11 at 100-200 nM had moderate effect on huPrP (lane 6) and ODN Ct-12 had no effect (lanes 10-12). Last, we monitored the impact of these three ODNs on the *trans*-splicing reaction facilitated by huPrP, and again the thioaptamer GA-12 was efficiently (lanes 7-9) inhibiting *trans*-splicing activation by huPrP, while Ct-12 and mODN-11 were inactive (lanes 4-6 and 10-11). Results obtained with ovPrP

were very similar, namely an inhibition of ovPrP chaperoning activity by GA-12 at 100 nM but not by mODN-11 and Ct-12 (data not shown).

We also examined the effect of GA-12 and mODN-11 on huPrP(23-110)-promoted cTAR/dTAR annealing using FRET. Both GA-12 and mODN-11 decreased the kinetics of huPrP(23-110)-promoted cTAR/dTAR annealing (Figure 6a), but the decrease was more pronounced with GA-12 (Figure 6a). Finally, we analyzed the concentration dependence of the inhibition of GA-12 on the huPrP-promoted annealing of cTAR/dTAR (Figure 6b). The inhibitory effect of GA-12 increased with its concentration and appeared already when the GA-12 concentration was one order of magnitude lower as compared to cTAR and dTAR concentrations.

Figure 5 A,B,C

Figure 6 a and b

Taken together these results show that the GACACAAGCCGA thioaptamer is a potent inhibitor of the huPrP nucleic acid chaperoning activity *in vitro* and this appears to rely on the backbone modifications.

Discussion

Collectively the data presented herein clearly show that the human and ovine prion proteins can assist the folding of RNA and DNA molecules in physiological conditions and at concentrations of 10^{-7} M or below (Fig. 1-5). Thus, these two mammalian prion proteins have the hallmarks of *bona fide* nucleic acid chaperones (24,25). In addition, the chaperoning properties of the human PrP reside within the N-terminal region of the protein (Fig. 1-4). Interestingly this N-terminal region possesses the characteristic features of a chaperone (52,53) since it contains aromatic residues and clusters of basic amino acids in an intrinsically unstructured environment (54-56). Despite the absence (or relaxation) of structural constraints on mutations in the N-terminal domain of PrP, the sequence and distribution of basic amino acids is highly conserved between various mammalian prion proteins (not shown).

As a consequence of its chaperone activity, the N-terminal region of huPrP was found to activate by two orders of magnitude the kinetics of the cTAR/dTAR annealing reaction, probably through a reaction mechanism similar to that described for the HCV core protein (37). In this mechanism, the fast and slow pathways were found to be nucleated through the

cTAR ends, but differ by the number of base pairs that should be premelted in the original cTAR secondary structure to nucleate the IC formation. In its free form, as a consequence of the thermal fraying that occurs spontaneously at room temperature, the closed cTAR species was shown to be in equilibrium with partially melted cTAR species where either the terminal or both the terminal and penultimate double stranded segments of the stem are melted (47-49). In this respect, the 8.3 kcal/mol enthalpy energy associated to the fast pathway is consistent with the melting of the terminal double-stranded 3 bp segment, which is poorly stable due to its position at the end of the stem and the presence of the neighbouring G52 bulge. On the other hand, the 16.5 kcal/mol enthalpy energy associated with the slow pathway is consistent with the energy needed to melt both the terminal and penultimate double-stranded segments of the cTAR stem (Figure 7). As a consequence of this thermal fraying, the PrP-bound $cTAR_1$ species in the fast pathway is thought to form IC1,stabilized by the seven intermolecular base pairs resulting from the annealing of both the 3′ and 5′ terminal strands of cTAR with the complementary terminal strands of dTAR. Since the K_M values of IC1 and IC2 are similar, we suggest that the PrP-bound $cTAR_2$ species in the slow pathway anneals only through one of its frayed strands with the complementary dTAR sequence to form IC2, also stabilized by 7 base pairs.. Further conversion of both ICs into the ED most probably relies on the conformational rearrangement and melting of the stable upper part of both TAR species. As for the core protein of the hepatitis C virus (HCV) (37,59,60), the interconversion rate of IC1 into ED was found to be much larger than for IC2 (Table 1), probably as a consequence of the more favorable zippering mechanism adopted by IC1 as compared to the invasion mechanism adopted by IC2 (Figure 7).

With respect to NCp7, the natural partner of HIV-1 cTAR and dTAR sequences, data revealed that the equilibrium constant K_M for IC formation was two orders of magnitude lower for the N-terminal region of huPrP than for NCp7 added at a ratio of 10 peptides per oligonucleotide ($K=10^8\ M^{-1}$) (61). This difference may in part be explained by the incomplete coating of the oligonucleotides by huPrP(23-110), since the K_M value of the IC was shown to be strongly dependent on the level of protein coating in the case of NCp7 (with about 3 orders of magnitude difference between low and full coating). Interestingly, the interconversion rate value of the fast pathway for the N-terminal region of huPrP was close to the value ($0.03\ s^{-1}$) of the corresponding parameter obtained with NCp7. The comparison of this parameter for the two proteins is more straightforward, since this parameter was found to be only poorly dependent on the level of NCp7 coating (61). Thus, huPrP(23-110) appears to promote the conversion of IC1 to the final ED with the same efficiency as NCp7. Moreover, as for NCp7

(35,61,62), huPrP(23-110) was found to nucleate the cTAR/dTAR annealing reaction through the stems of cTAR and dTAR but without destabilization of the termini (Figure 1, inset). As for the HCV core protein, this is a likely consequence of the stronger 'nucleic acid aggregating' properties of the huPrP peptide that compensate for the absence of its destabilizing activity. This propensity of the huPrP N-terminal peptide to efficiently neutralize the negatively charged oligonucleotides and promote their aggregation is probably related to the more flexible nature of this peptide, as compared to NCp7 with the structured zinc fingers (63,64).

It is hypothesized that such a disorder status of the N-terminal domain of PrP allows for the recognition of a rather large number of cellular and viral RNA sequences, as well as protein partners, as amply shown for the cellular protein FMRP and the retroviral NC proteins [reviewed in (52)]. Recognition of many partners by nucleic acid chaperones favors the notion that these ubiquitous proteins can achieve several functions as exemplified by retroviral NC proteins in virus structure, genome replication and virus assembly [reviewed in (65,66)] and the cellular p53 and p50 proteins in DNA maintenance and mRNA translation, respectively (40,67).

What is the function of PrP and what could be its role as an RNA chaperone? For a long time the physiological function of PrP^C remained a matter of speculation and debate. Recent data demonstrated that PrP is involved in cell-cell adhesion whereby it controls cell movements in the early gastrula of zebrafish embryos [reviewed in (68)]. In addition, PrP appears to trigger cellular signaling via Src-related kinases especially in neurons (18). These cellular functions of PrP could well be linked to the fact that PrP is, at least in part, a membrane associated protein through its GPI anchor or its transmembrane domain [reviewed in (69)]. Regarding the RNA chaperoning activity, PrP could act at the level of mRNA translation as shown for some fungal prion proteins (70). In agreement with this notion, we have recently found that a fraction of huPrP is associated with translating ribosomes in human cells, impacting on mRNA translation notably in cells in the G2M phase (to be published elsewhere). Thus PrP would be another chaperone protein associated with the cellular translation machinery, thereby regulating the level of mRNA translation in a manner similar to other essential RNA chaperones such as hnRNP A1 in every cell type and FMRP in neurons (to be published elsewhere).

We also report that a small thioaptamer (GA-12) is capable of inhibiting the PrP chaperoning activity *in vitro* but not that of HIV-1 NC (data not shown) At the same time a small methylated oligoribonucleotide (mODN-11), previously found to be a potent inhibitor of

HIV-1 NC *in vitro* and of HIV-1 replication in cells (31), had little inhibitory potential against PrP. It is possible that the small thioaptamer functions by tightly binding to the N-terminal domain of huPrP, in turn causing the formation of a complex where the N-terminal disordered domain adopts a stable ordered conformation. Such thioaptamer derivatives were found to delay the onset of prion disease in the mouse model (71), most probably by binding to PrP^{C}, in turn preventing its recruitment into aggregates of PrP^{Sc}. These results and those recently published on HIV-1 NC pave the way for the future development of highly modified oligonucleotides to treat diseases caused by RNA viruses as well as by a non-conventional agent.

Acknowledgements

We thank W. Surewicz (USA) for huPrP, PrP(23-145) and PrP(123-231), C. Branlant (France) and L. Ovchinnikov (Russia) for hnRNP A1 and p50, respectively, H. Moine (France) for pFlashSV40Luc, E. Bertrand (France) and J. Rossi (USA) for pR3, pS14 and pS20. We also thank D. Ficheux (France) for NCp7 and NC(12-53). Work supported by ANRS, INSERM, GIS Prions (France) and TRIoH (FP6 EC).

Figure legends.

Figure 1: Promotion of cTAR/dTAR annealing by huPrP(23-110).

Kinetic trace of 10 nM Rh6G-5'-cTAR-3'-Dabcyl with 100 nM of dTAR in 25 mM Tris-HCl (pH 7.5), 30 mM NaCl, 0.2 mM $MgCl_2$ at 20 °C. huPrP(23-110) was added at a 1:1 molar ratio to cTAR and dTAR. Excitation and emission wavelengths were 520 nm and 550 nm, respectively. The continuous line corresponds to the best fit of the data according to eqn 1, using $k_{obs1} = 5.6 \times 10^{-3}$ s^{-1}, $k_{obs2} = 6.8 \times 10^{-4}$ s^{-1} and a = 0.7. Inset: Emission spectra of 10 nM doubly labelled cTAR (black line), with 10 nM huPrP(23-110) (red line) and with 100 nM dTAR + 100 nM huPrP(23-110) after completion of the annealing reaction (Blue).

Figure 2. Facilitation of DNA strand exchange by PrPs.

A/ Assay schematic. DNA sequences representing the R region of HIV-1 are 96 nt in length. 5' ^{32}P-labelled R(+) wt was hybridized to R(-) mut to generate a double stranded DNA with mismatches at the 3' end (step 1). Fully complementary R(-) wt was added in the presence or absence of HIV-1 NCp7 or PrP (step 2). Strand exchange is visualized by native PAGE.

B/ Assays with 3×10^{-9} M each of the DNA oligonucleotides were carried out at 37 °C in 10 □l (see methods). Analysis of the reaction products was by 6% PAGE under native conditions. *R = ^{32}P-labelled DNA.

Lanes 1 and 2: *R(+) wt alone and with R(-) mut at 0 °C.

Lanes 3 and 4: *R(+) hybridized to R(-) wt and *R(+) hybridized to R(-) mut.

Lanes 5 and 6: *R(+) : R(-) mut incubated with R(-) wt at 4 °C and 37 °C for 30 min.

Lanes 7 to 18: 5 min incubations with *R(+) : R(-) mut and R(-) wt and with NCp7 (lanes 7-8), NC(12-53) (lanes 9-10), recombinant huPrP (lanes 11-12), huPrP(23-145) (lanes 13-14), huPrP(122-231) (lanes 15-16) or ovine PrP (lanes 17-18) at

6×10^{-8} M (protein to nt molar ratio of 1/12) (odd lanes) or 12×10^{-8} M (protein to nt molar ratio of 1/6) (even lanes).

Proteins are indicated at the top of the figure; double stranded products [*R(+) :R(-) wt] and [*R(+) : R(-) mut] are indicated on the right. Arrow is direction of electrophoresis. Note that the C-terminal region of huPrP is inactive (lanes 15-16).

Figure 3. PrPs facilitate ribozyme cleavage of an RNA.

A/ Assay schematic. A hammerhead ribozyme and a ^{32}P-labelled RNA substrate were generated by *in vitro* transcription and gel purified. Cleavage of the ^{32}P-RNA by the ribozyme appears to first necessitate hybridization of the ribozyme to the substrate (step 1). After RNA cleavage (step 2), the products must be released to allow recycling of the ribozyme (step 3). At the end of the reaction RNAs were deproteinized and analyzed by PAGE under denaturing conditions to visualize the ^{32}P-RNA products. In the absence of a nucleic acid chaperone, hybridization of the RNA to the ribozyme and release of the products appear to be slow. Addition of a nucleic acid chaperone will accelerate hybridization of the substrate to the ribozyme and dissociation of the products and thus ribozyme turnover.

Base pairing between the RNA substrate and ribozyme R3 are underlined on the substrate sequence. Ribozyme mediated cleavage occurs on the 3' side of A (space).

for RNA S14, ...GAUUAAGUAGUA AGAGUGUCUGCA 3'

for RNA S20, ...GAUUAAGUAGUA AGAGUGUCUGCA 3'

B/ Ribozyme-directed cleavage of RNA substrate S14.

$1x10^{-8}$ M of ribozyme R3 and $5x10^{-9}$ M of RNA S14 were incubated as described in the methods. ^{32}P-RNA substrate (S14) and product (□S14) were analysed by denaturing 8% PAGE.

Lanes 1-2: R3 and S14 at 4 °C or 37 °C.

Lanes 3-5: HIV-1 NCp7 at NC concentrations of $2x10^{-8}$, $4x10^{-8}$ and $8x10^{-8}$ M (protein/nt molar ratios of 1/20, 1/10 and 1/5).

Lanes 6-8: NC(12-53) at concentrations of $1.6x10^{-7}$, $3.2x10^{-7}$ and $6.4x10^{-7}$ M (molar ratios of 1/2.5, 1/1.2 and 1/0.6).

Lanes 9-11: ovPrP at $2x10^{-8}$, $4x10^{-8}$ and $8x10^{-8}$ M, respectively.

Lanes 12-14; huPrP at concentrations of $2x10^{-8}$, $4x10^{-8}$ and $8x10^{-8}$ M, respectively.

Lanes 15-17: huPrP(23-110) at concentrations of $2x10^{-8}$, $4x10^{-8}$ and $8x10^{-8}$ M, respectively.

Lanes 18-20: huPrP(23-145) at concentrations of $2x10^{-8}$, $4x10^{-8}$ and $8x10^{-8}$ M, respectively.

Lanes 21-23: huPrP(122-231) at concentrations of $2x10^{-8}$, $4x10^{-8}$ and $8x10^{-8}$ M, respectively.

R3, S14 and the 5' sequences of S14 (□S14) are identified on the right. Markers are on the left. Arrow shows direction of electrophoresis. Note that the RNA products rapidly

accumulate in the presence of a chaperone whereas they do not in the absence of a chaperone.

C/ Ribozyme cleavage of RNA S20. Conditions were as in fig. 3B except that the substrate was S20 and protein concentrations of $4x10^{-8}$ (odd lanes) and $8x10^{-8}$ M (even lanes) were used. Proteins are identified at the top of the figure.

R3, S20 and the 5' sequences of S20 (□S20) are identified on the right. Markers are on the left. Arrow shows direction of electrophoresis. Note that ribozyme-directed cleavage of S14 or S20 did not occur at 4 °C (lane 1).

Figure 4. Facilitation of RNA *trans*-splicing by PrPs.

A. Schematic representation of the *trans*-splicing assay: The two RNA constructs H1 of 680 nt in length (encoding exon 1 of 549 nt and the 5' part of the intron of 131 nt) and H2 of 170 nt in length (encoding the 3' part of the intron of 147 nt and exon 2 of 23 nt) have to fold into a splicing competent structure (step 1). The splicing reaction was started upon addition of the ^{32}P-labelled GTP. The recombinant PrP was added at step 2, where indicated. The final RNA products are represented in step 4. The splicing rates are based on the levels of GTP-I1 (see B).

B and C. Assays where the H1 and H2 RNAs were incubated with or without a nucleic acid chaperone, namely NCp7 or PrP. At the end of the reaction RNAs were phenol treated to remove the chaperone and analysed by PAGE in denaturing conditions (see methods). The RNA substrates (H1, H2), the ligated exons (E1-E2) and the product (guanosine-5'-intron G-I1) are indicated.

Panel B. Lane 1: *trans*-splicing reaction at 37 °C with RNA alone at a concentration of 4x10-8 M. Lanes 2-5: HIV-1 NCp7 at concentrations of 2.5, 5, 10 and $20x10^{-7}$ M, corresponding to protein to nt ratios of 1:128, 1:64, 1:32 and 1:16. Lanes 6-7: NC(12-53) at concentrations of 1 and $2x10^{-6}$ M, corresponding to peptide to nt ratios of 1:32 and 1:16. Lanes 8-11: recombinant huPrP(23-231) at concentrations of 2, 4, 8 and 16 x 10^{-7} M, corresponding to PrP to nt molar ratios of 1:160, 1:80, 1:40 and 1:20. Lanes 12-15: N-terminal huPrP(23-144) at the same protein to nt ratios as for the huPrP(23-231). Lanes 16-17: C-terminal huPrP (123-231) at protein to nt molar ratios of 1:40 and 1:20.

Panel C. Lane 1: splicing reaction at 37 °C with RNA alone at a concentration of 4x10-8 M. Lanes 2-6: huPrP at concentrations of 1, 2, 4, 8 and $16x10^{-7}$ M. Lanes 7-11: recombinant

ovPrP at concentrations of 2, 4, 8, 16 and $32x10^{-7}$ M. Lanes 12-16: huPrP(23-110) at concentrations of 1, 2, 4, 8 and $16x10^{-7}$ M.

Note that HIV-1 NCp7, huPrP and ovPrP strongly activated *trans*-splicing at protein to nt molar ratios of 1:20 to 1:40 . Peptides NC(12-53) and C-terminal huPrP(122-231) were very poorly active (lanes 6, 7, 16, 17; in panel B)

Figure 5. Thioaptamer inhibition of huPrP chaperoning activities *in vitro*

A- Annealing of TAR(+) / TAR(-).

The annealing conditions were exactly as described in the methods section with the TAR(+) and TAR(-) DNA. PAGE analyses are described in methods.

Lane 1: Control at 65 °C for 30 min; lane 2: at 37 °C for 30 min.

Lanes 3-12: with huPrP for 5 min at 37 °C.

Lane 3: control showing the complete hybridization of TAR(+) and TAR(-) DNA.

Lanes 4-6: with mODN-11 added at the start of the incubation period at 50, 100 and 200 nM. No inhibition of huPrP annealing activity was observed.

Lanes 7-9: with ODN GA-12 at 50, 100 and 200 nM. Note that ODN GA-12 causes a complete inhibition of huPrP chaperoning at 200 nM.

Lanes 10-12: with control ODN Ct-12 at 50, 100 and 200 nM.

B- Ribozyme mediated cleavage of an RNA substrate

The conditions for the ribozyme cleavage were exactly as described in the methods section with hammerhead ribozyme and the RNA substrate. PAGE analyses under denaturing conditions are described in methods.

Lane 1: Control at 0 °C for 30 min; lane 2: at 37 °C for 30 min without PrP.

Lanes 3: control with huPrP for 15 min at 37 °C showing a complete cleavage.

Lanes 4-6: with mODN-11 added at the start of the incubation period at 50, 100 and 200 nM. Some inhibition of huPrP activity was observed at 100 nM.

Lanes 7-9: with ODN GA-12 at 50, 100 and 200 nM. Note that ODN GA-12 causes a nearly complete inhibition of huPrP at 50 nM and 100 nM for ovPrP (data not shown).

Lanes 10-12: with control ODN Ct-12 at 50, 100 and 200 nM.

C- The RNA *trans*-splicing reaction.

The conditions for the *trans*-splicing reactions were exactly as described in the methods section with the H1 and H2 RNAs. PAGE analyses under denaturing conditions are

described in methods.

Lane 1: Control at 0 °C for 30 min; lane 2: at 37 °C for 30 min without PrP.

Lanes 3: control with huPrP at $8x10^{-7}$ M for 30 min at 37 °C showing a complete cleavage.

Lanes 4-6: with mODN-11 added at the start of the incubation period at 50, 100 and 200 nM. No inhibition of PrP activity was observed.

Lanes 7-9: with ODN GA-12 at 50, 100 and 200 nM. Note that ODN GA-12 causes a nearly complete inhibition of huPrP at 100 nM.

Lanes 10-12: with control ODN Ct-12 at 50, 100 and 200 nM.

Figure 6: Inhibition of huPrP(23-110)-promoted cTAR/dTAR annealing.

(a) Comparative effects of GA-12 and mODN-11. 10 nM of doubly labelled cTAR was mixed with 300 nM of non-labelled dTAR at a peptide/oligonucleotide molar ratio of 1:1 in the absence (blue) and presence of GA-12 (black) and mODN-11 (red) sequences, added at 5 times molar excess, as compared to cTAR and dTAR. Excitation and emission wavelengths were 520 nm and 550 nm, respectively.

(b) Concentration dependence of the effect of GA-12 on the annealing kinetics. Kinetic traces recorded with 10 nM doubly labelled TMR-cTAR-Fl and 300 nM non-labelled dTAR in the absence (black) and the presence of GA-12, added at 0.25x (green), 1x (cyan), 5x (magenta) and 10x (yellow) molar ratio as compared to cTAR and dTAR concentrations. huPrP(23-110) was added at a peptide/oligonucleotide ratio of 1:1 Excitation and emission wavelengths were 480 nm and 520 nm, respectively. All experiments were performed in 25 mM Tris-HCl (pH 7.5), 30 mM NaCl, 0.2 mM $MgCl_2$ at 20°C.

Figure 7: Proposed mechanism for PrP-promoted cTAR/dTAR annealing.

The two pathways of the annealing reaction are thought to rely on the thermal fraying of cTAR that leads to a fast equilibrium (µs range) between the premelted cTAR species ($cTAR_1$ and $cTAR_2$) (47). The upper and lower pathways are associated with the fast and slow kinetic components, respectively. In both pathways, the cTAR species nucleate an intermediate complex (IC) through the stem termini. Then, the ICs are converted in a rate-limiting step into the final extended duplex (ED).

References

1. Weissmann, C. (2004) The state of the prion. *Nat Rev Microbiol*, **2**, 861-871.
2. Aguzzi, A., Baumann, F. and Bremer, J. (2008) The prion's elusive reason for being. *Annu Rev Neurosci*, **31**, 439-477.
3. Prusiner, S.B., Scott, M.R., DeArmond, S.J. and Cohen, F.E. (1998) Prion protein biology. *Cell*, **93**, 337-348.
4. Basler, K., Oesch, B., Scott, M., Westaway, D., Walchli, M., Groth, D.F., McKinley, M.P., Prusiner, S.B. and Weissmann, C. (1986) Scrapie and cellular PrP isoforms are encoded by the same chromosomal gene. *Cell*, **46**, 417-428.
5. Oesch, B., Westaway, D., Walchli, M., McKinley, M.P., Kent, S.B., Aebersold, R., Barry, R.A., Tempst, P., Teplow, D.B., Hood, L.E. *et al.* (1985) A cellular gene encodes scrapie PrP 27-30 protein. *Cell*, **40**, 735-746.
6. Pan, K.M., Baldwin, M., Nguyen, J., Gasset, M., Serban, A., Groth, D., Mehlhorn, I., Huang, Z., Fletterick, R.J., Cohen, F.E. *et al.* (1993) Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. *Proc Natl Acad Sci U S A*, **90**, 10962-10966.
7. Horwich, A.L. and Weissman, J.S. (1997) Deadly conformations--protein misfolding in prion disease. *Cell*, **89**, 499-510.
8. Bueler, H., Aguzzi, A., Sailer, A., Greiner, R.A., Autenried, P., Aguet, M. and Weissmann, C. (1993) Mice devoid of PrP are resistant to scrapie. *Cell*, **73**, 1339-1347.
9. Westergard, L., Christensen, H.M. and Harris, D.A. (2007) The cellular prion protein (PrP(C)): its physiological function and role in disease. *Biochim Biophys Acta*, **1772**, 629-644.
10. Bueler, H., Fischer, M., Lang, Y., Bluethmann, H., Lipp, H.P., DeArmond, S.J., Prusiner, S.B., Aguet, M. and Weissmann, C. (1992) Normal development and behaviour of mice lacking the neuronal cell-surface PrP protein. *Nature*, **356**, 577-582.
11. Steele, A.D., Lindquist, S. and Aguzzi, A. (2007) The prion protein knockout mouse: a phenotype under challenge. *Prion*, **1**, 83-93.
12. Brown, D.R., Wong, B.S., Hafiz, F., Clive, C., Haswell, S.J. and Jones, I.M. (1999) Normal prion protein has an activity like that of superoxide dismutase. *Biochem J*, **344 Pt 1**, 1-5.
13. Milhavet, O., McMahon, H.E., Rachidi, W., Nishida, N., Katamine, S., Mange, A., Arlotto, M., Casanova, D., Riondel, J., Favier, A. *et al.* (2000) Prion infection impairs the cellular response to oxidative stress. *Proc Natl Acad Sci U S A*, **97**, 13937-13942.
14. Brown, D.R., Qin, K., Herms, J.W., Madlung, A., Manson, J., Strome, R., Fraser, P.E., Kruck, T., von Bohlen, A., Schulz-Schaeffer, W. *et al.* (1997) The cellular prion protein binds copper in vivo. *Nature*, **390**, 684-687.
15. Mouillet-Richard, S., Ermonval, M., Chebassier, C., Laplanche, J.L., Lehmann, S., Launay, J.M. and Kellermann, O. (2000) Signal transduction through prion protein. *Science*, **289**, 1925-1928.
16. Chiarini, L.B., Freitas, A.R., Zanata, S.M., Brentani, R.R., Martins, V.R. and Linden, R. (2002) Cellular prion protein transduces neuroprotective signals. *EMBO J*, **21**, 3317-3326.
17. Bounhar, Y., Zhang, Y., Goodyer, C.G. and LeBlanc, A. (2001) Prion protein protects human neurons against Bax-mediated apoptosis. *J Biol Chem*, **276**, 39145-39149.
18. Malaga-Trillo, E., Solis, G.P., Schrock, Y., Geiss, C., Luncz, L., Thomanetz, V. and Stuermer, C.A. (2009) Regulation of embryonic cell adhesion by the prion protein. *PLoS Biol*, **7**, e55.
19. Gabus, C., Derrington, E., Leblanc, P., Chnaiderman, J., Dormont, D., Swietnicki, W.,

Morillas, M., Surewicz, W.K., Marc, D., Nandi, P. *et al.* (2001) The prion protein has RNA binding and chaperoning properties characteristic of nucleocapsid protein NCP7 of HIV-1. *J Biol Chem*, **276**, 19301-19309.

20. Gabus, C., Auxilien, S., Pechoux, C., Dormont, D., Swietnicki, W., Morillas, M., Surewicz, W., Nandi, P. and Darlix, J.L. (2001) The prion protein has DNA strand transfer properties similar to retroviral nucleocapsid protein. *J Mol Biol*, **307**, 1011-1021.
21. Moscardini, M., Pistello, M., Bendinelli, M., Ficheux, D., Miller, J.T., Gabus, C., Le Grice, S.F., Surewicz, W.K. and Darlix, J.L. (2002) Functional interactions of nucleocapsid protein of feline immunodeficiency virus and cellular prion protein with the viral RNA. *J Mol Biol*, **318**, 149-159.
22. Leblanc, P., Baas, D. and Darlix, J.L. (2004) Analysis of the interactions between HIV-1 and the cellular prion protein in a human cell line. *J Mol Biol*, **337**, 1035-1051.
23. Lotscher, M., Recher, M., Lang, K.S., Navarini, A., Hunziker, L., Santimaria, R., Glatzel, M., Schwarz, P., Boni, J. and Zinkernagel, R.M. (2007) Induced prion protein controls immune-activated retroviruses in the mouse spleen. *PLoS One*, **2**, e1158.
24. Schroeder, R., Barta, A. and Semrad, K. (2004) Strategies for RNA folding and assembly. *Nat Rev Mol Cell Biol*, **5**, 908-919.
25. Cristofari, G. and Darlix, J.L. (2002) The ubiquitous nature of RNA chaperone proteins. *Prog Nucleic Acid Res Mol Biol*, **72**, 223-268.
26. Tsuchihashi, Z. and Brown, P.O. (1994) DNA strand exchange and selective DNA annealing promoted by the human immunodeficiency virus type 1 nucleocapsid protein. *J Virol*, **68**, 5863-5870.
27. Tsuchihashi, Z., Khosla, M. and Herschlag, D. (1993) Protein enhancement of hammerhead ribozyme catalysis. *Science*, **262**, 99-102.
28. Bertrand, E.L. and Rossi, J.J. (1994) Facilitation of hammerhead ribozyme catalysis by the nucleocapsid protein of HIV-1 and the heterogeneous nuclear ribonucleoprotein A1. *EMBO J*, **13**, 2904-2912.
29. Coetzee, T., Herschlag, D. and Belfort, M. (1994) Escherichia coli proteins, including ribosomal protein S12, facilitate in vitro splicing of phage T4 introns by acting as RNA chaperones. *Genes Dev*, **8**, 1575-1588.
30. Galloway Salvo, J.L., Coetzee, T. and Belfort, M. (1990) Deletion-tolerance and trans-splicing of the bacteriophage T4 td intron. Analysis of the P6-L6a region. *J Mol Biol*, **211**, 537-549.
31. Grigorov, B., Bocquin, A., Gabus, C., Avilov, S., Mély, Y., Agopian, A., Divita, G., Gottikh, M., Witvrouw, M. and Darlix, J.L. (2011) Identification of a methylated oligoribonucleotide as a potent inhibitor of HIV-1 reverse transcription. *Nucleic Acids Res*, March 29, 2011..
32. King, D.J., Safar, J.G., Legname, G. and Prusiner, S.B. (2007) Thioaptamer interactions with prion proteins: sequence-specific and non-specific binding sites. *J Mol Biol*, **369**, 1001-1014.
33. Rezaei, H., Marc, D., Choiset, Y., Takahashi, M., Hui Bon Hoa, G., Haertle, T., Grosclaude, J. and Debey, P. (2000) High yield purification and physico-chemical properties of full-length recombinant allelic variants of sheep prion protein linked to scrapie susceptibility. *Eur J Biochem*, **267**, 2833-2839.
34. Gabus, C., Mazroui, R., Tremblay, S., Khandjian, E.W. and Darlix, J.L. (2004) The fragile X mental retardation protein has nucleic acid chaperone properties. *Nucleic Acids Res*, **32**, 2129-2137.
35. Godet, J., de Rocquigny, H., Raja, C., Glasser, N., Ficheux, D., Darlix, J.L. and Mely, Y. (2006) During the early phase of HIV-1 DNA synthesis, nucleocapsid protein

directs hybridization of the TAR complementary sequences via the ends of their double-stranded stem. *J Mol Biol*, **356**, 1180-1192.
36. Kuciak, M., Gabus, C., Ivanyi-Nagy, R., Semrad, K., Storchak, R., Chaloin, O., Muller, S., Mely, Y. and Darlix, J.L. (2008) The HIV-1 transcriptional activator Tat has potent nucleic acid chaperoning activities in vitro. *Nucleic Acids Res*, **36**, 3389-3400.
37. Sharma, K., Didier, P., Darlix, J.L., de Rocquigny, H., Bensikaddour, H., Lavergne, J.P., Penin, F., Lessinger, J.M. and Mely, Y. (2010) Kinetic analysis of the nucleic acid chaperone activity of the hepatitis C virus core protein. *Nucleic Acids Res*, **38**, 3632-3642.
38. Herschlag, D. (1995) RNA chaperones and the RNA folding problem. *J Biol Chem*, **270**, 20871-20874.
39. Rajkowitsch, L., Semrad, K., Mayer, O. and Schroeder, R. (2005) Assays for the RNA chaperone activity of proteins. *Biochem Soc Trans*, **33**, 450-456.
40. Skabkin, M.A., Evdokimova, V., Thomas, A.A. and Ovchinnikov, L.P. (2001) The major messenger ribonucleoprotein particle protein p50 (YB-1) promotes nucleic acid strand annealing. *J Biol Chem*, **276**, 44841-44847.
41. Portman, D.S. and Dreyfuss, G. (1994) RNA annealing activities in HeLa nuclei. *EMBO J*, **13**, 213-221.
42. Herschlag, D., Khosla, M., Tsuchihashi, Z. and Karpel, R.L. (1994) An RNA chaperone activity of non-specific RNA binding proteins in hammerhead ribozyme catalysis. *EMBO J*, **13**, 2913-2924.
43. Darlix, J.L., Lapadat-Tapolsky, M., de Rocquigny, H. and Roques, B.P. (1995) First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. *J Mol Biol*, **254**, 523-537.
44. Rein, A., Henderson, L.E. and Levin, J.G. (1998) Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem Sci*, **23**, 297-301.
45. Ameres, S.L., Shcherbakov, D., Nikonova, E., Piendl, W., Schroeder, R. and Semrad, K. (2007) RNA chaperone activity of L1 ribosomal proteins: phylogenetic conservation and splicing inhibition. *Nucleic Acids Res*, **35**, 3752-3763.
46. Rajkowitsch, L., Chen, D., Stampfl, S., Semrad, K., Waldsich, C., Mayer, O., Jantsch, M.F., Konrat, R., Blasi, U. and Schroeder, R. (2007) RNA chaperones, RNA annealers and RNA helicases. *RNA Biol*, **4**, 118-130.
47. Azoulay, J., Clamme, J.P., Darlix, J.L., Roques, B.P. and Mely, Y. (2003) Destabilization of the HIV-1 complementary sequence of TAR by the nucleocapsid protein through activation of conformational fluctuations. *J Mol Biol*, **326**, 691-700.
48. Beltz, H., Azoulay, J., Bernacchi, S., Clamme, J.P., Ficheux, D., Roques, B., Darlix, J.L. and Mely, Y. (2003) Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the nucleocapsid protein NCp7. *J Mol Biol*, **328**, 95-108.
49. Bernacchi, S., Stoylov, S., Piemont, E., Ficheux, D., Roques, B.P., Darlix, J.L. and Mely, Y. (2002) HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. *J Mol Biol*, **317**, 385-399.
50. Cantor, C. and Schimmel, P. (1980) *Biophysical chemistry Part 2: Techniques for the study of biological structure and function*. Academic Press, NY.
51. Rouzina, I. and Bloomfield, V.A. (1999) Heat capacity effects on the melting of DNA. 1. General aspects. *Biophys J*, **77**, 3242-3251.
52. Ivanyi-Nagy, R., Davidovic, L., Khandjian, E.W. and Darlix, J.L. (2005) Disordered

RNA chaperone proteins: from functions to disease. *Cell Mol Life Sci*, **62**, 1409-1417.
53. Tompa, P. and Csermely, P. (2004) The role of structural disorder in the function of RNA and protein chaperones. *FASEB J*, **18**, 1169-1175.
54. Lysek, D.A., Schorn, C., Nivon, L.G., Esteve-Moya, V., Christen, B., Calzolai, L., von Schroetter, C., Fiorito, F., Herrmann, T., Guntert, P. *et al.* (2005) Prion protein NMR structures of cats, dogs, pigs, and sheep. *Proc Natl Acad Sci U S A*, **102**, 640-645.
55. Lopez Garcia, F., Zahn, R., Riek, R. and Wuthrich, K. (2000) NMR structure of the bovine prion protein. *Proc Natl Acad Sci U S A*, **97**, 8334-8339.
56. Zahn, R., Liu, A., Luhrs, T., Riek, R., von Schroetter, C., Lopez Garcia, F., Billeter, M., Calzolai, L., Wider, G. and Wuthrich, K. (2000) NMR solution structure of the human prion protein. *Proc Natl Acad Sci U S A*, **97**, 145-150.
57. Chen, Y., Balakrishnan, M., Roques, B.P. and Bambara, R.A. (2003) Steps of the acceptor invasion mechanism for HIV-1 minus strand strong stop transfer. *J Biol Chem*, **278**, 38368-38375.
58. Kim, J.K., Palaniappan, C., Wu, W., Fay, P.J. and Bambara, R.A. (1997) Evidence for a unique mechanism of strand transfer from the transactivation response region of HIV-1. *J Biol Chem*, **272**, 16769-16777.
59. Cristofari, G., Ivanyi-Nagy, R., Gabus, C., Boulant, S., Lavergne, J.P., Penin, F. and Darlix, J.L. (2004) The hepatitis C virus Core protein is a potent nucleic acid chaperone that directs dimerization of the viral (+) strand RNA in vitro. *Nucleic Acids Res*, **32**, 2623-2631.
60. Ivanyi-Nagy, R., Kanevsky, I., Gabus, C., Lavergne, J.P., Ficheux, D., Penin, F., Fosse, P. and Darlix, J.L. (2006) Analysis of hepatitis C virus RNA dimerization and core-RNA interactions. *Nucleic Acids Res*, **34**, 2618-2633.
61. Vo, M.N., Barany, G., Rouzina, I. and Musier-Forsyth, K. (2009) HIV-1 nucleocapsid protein switches the pathway of transactivation response element RNA/DNA annealing from loop-loop "kissing" to "zipper". *J Mol Biol*, **386**, 789-801.
62. Godet, J. and Mely, Y. (2010) Biophysical studies of the nucleic acid chaperone properties of the HIV-1 nucleocapsid protein. *RNA Biol*, **7**, 48-60.
63. Summers, M.F., Henderson, L.E., Chance, M.R., Bess, J.W., Jr., South, T.L., Blake, P.R., Sagi, I., Perez-Alvarado, G., Sowder, R.C., 3rd, Hare, D.R. *et al.* (1992) Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1. *Protein Sci*, **1**, 563-574.
64. Morellet, N., de Rocquigny, H., Mely, Y., Jullian, N., Demene, H., Ottmann, M., Gerard, D., Darlix, J.L., Fournie-Zaluski, M.C. and Roques, B.P. (1994) Conformational behaviour of the active and inactive forms of the nucleocapsid NCp7 of HIV-1 studied by 1H NMR. *J Mol Biol*, **235**, 287-301.
65. Muriaux, D. and Darlix, J.L. (2010) Properties and functions of the nucleocapsid protein in virus assembly. *RNA Biol*, **7**.
66. Darlix, J.L., Garrido, J.L., Morellet, N., Mely, Y. and de Rocquigny, H. (2007) Properties, functions, and drug targeting of the multifunctional nucleocapsid protein of the human immunodeficiency virus. *Adv Pharmacol*, **55**, 299-346.
67. Oberosler, P., Hloch, P., Ramsperger, U. and Stahl, H. (1993) p53-catalyzed annealing of complementary single-stranded nucleic acids. *EMBO J*, **12**, 2389-2396.
68. Malaga-Trillo, E. and Sempou, E. (2009) PrPs: Proteins with a purpose: Lessons from the zebrafish. *Prion*, **3**, 129-133.
69. Chakrabarti, O., Ashok, A. and Hegde, R.S. (2009) Prion protein biosynthesis and its emerging role in neurodegeneration. *Trends Biochem Sci*, **34**, 287-295.
70. Wickner, R.B., Edskes, H.K., Shewmaker, F.P., Kryndushkin, D., Nemecek, J.,

McGlinchey, R. and Bateman, D. (2010) The Relationship of Prions and Translation. *WIREs RNA*, **1**, 81-89.

71. Kocisko, D.A., Vaillant, A., Lee, K.S., Arnold, K.M., Bertholet, N., Race, R.E., Olsen, E.A., Juteau, J.M. and Caughey, B. (2006) Potent antiscrapie activities of degenerate phosphorothioate oligonucleotides. *Antimicrob Agents Chemother*, **50**, 1034-1044.

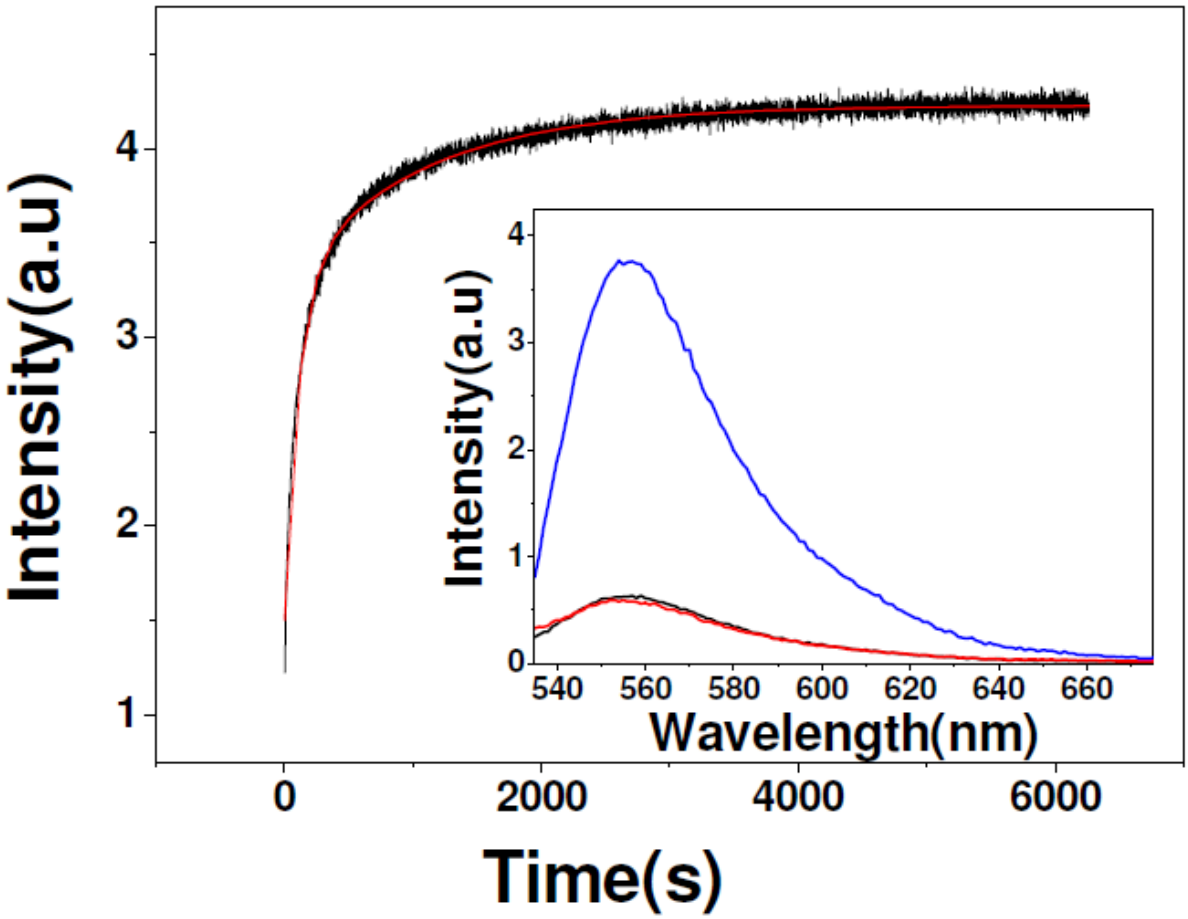
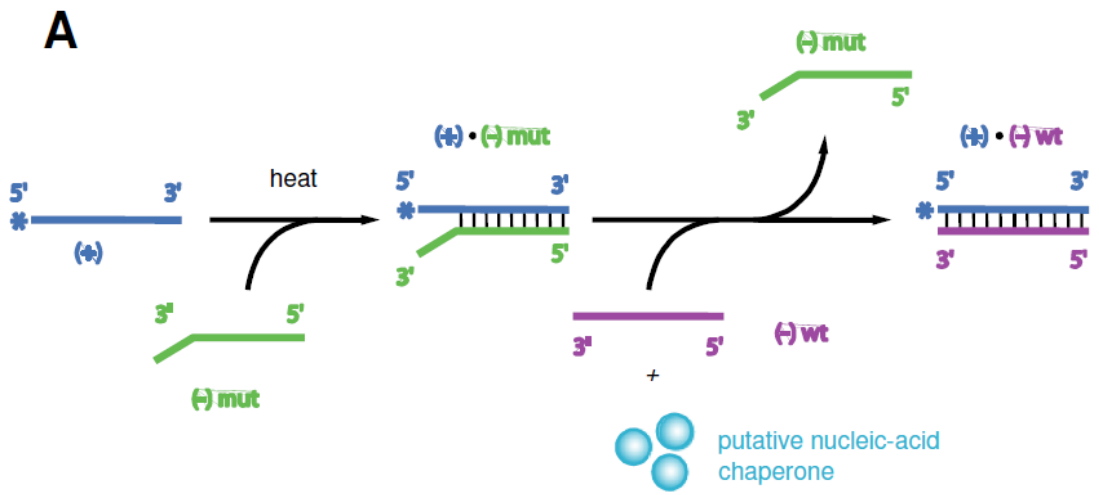

Intensity(a.u)
Time(s)
0
2000
4000
6000
1
2
3
4
Intensity(a.u)
Wavelength(nm)
0
1
2
3
4
540
560
580
600
620
640
660

Figure 1



Scheme of the strand exchange activity

Figure 2A

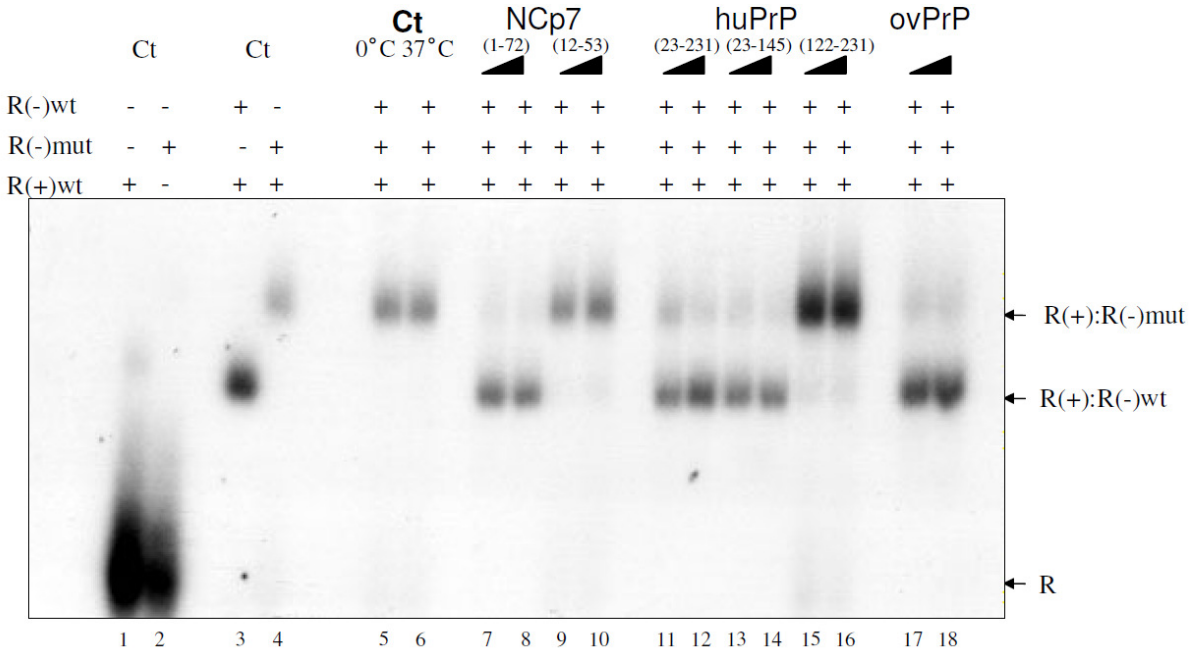
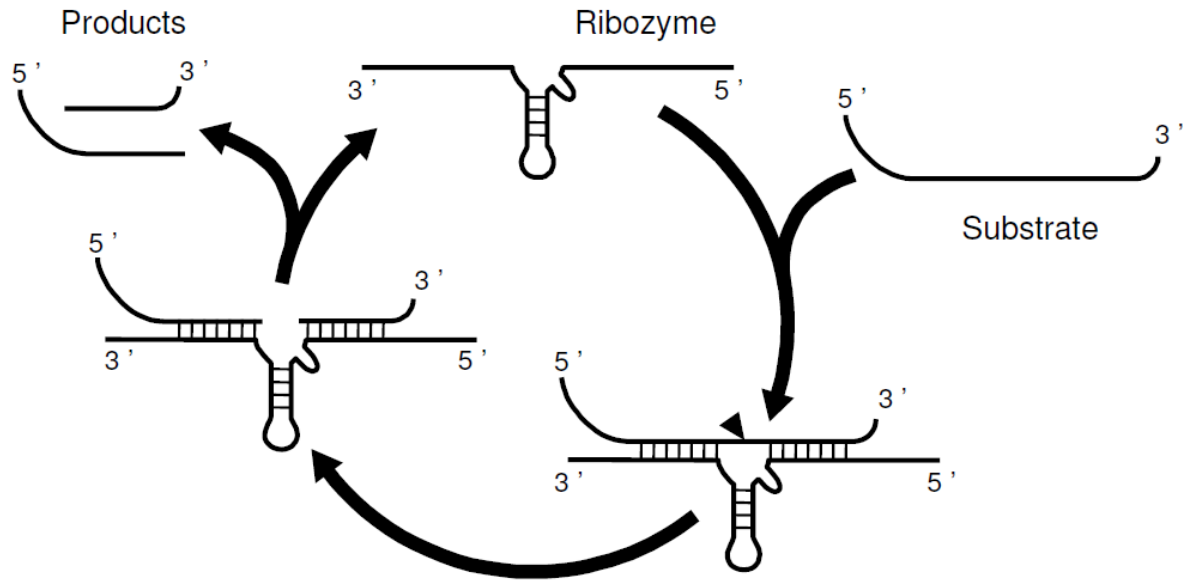


Figure 2B



A. Scheme of ribozyme cleavage

Figure 3A

B. Enhancement of Ribozyme cleavage by HIV-1 NCp7 and huPrP

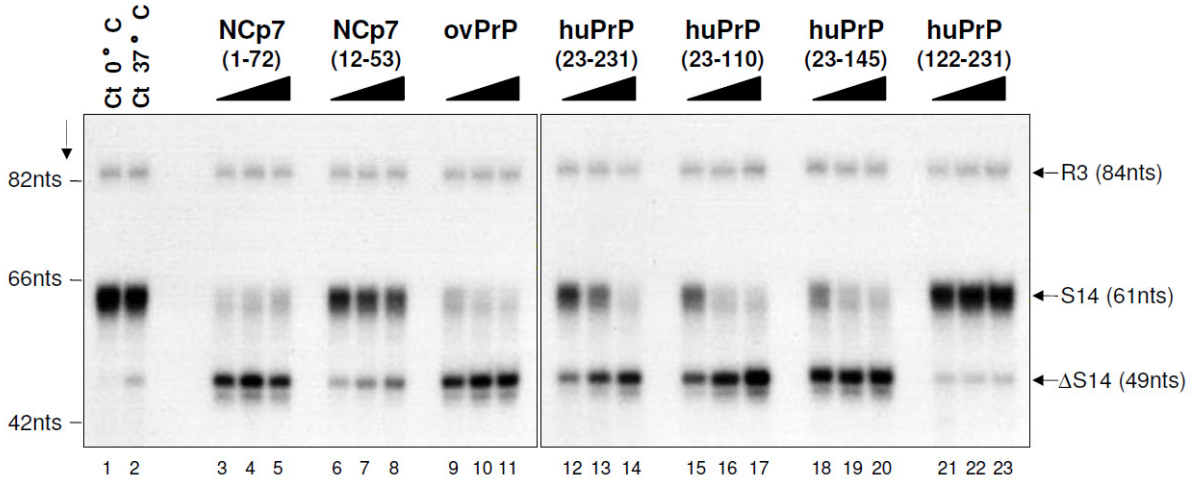


Figure 3B

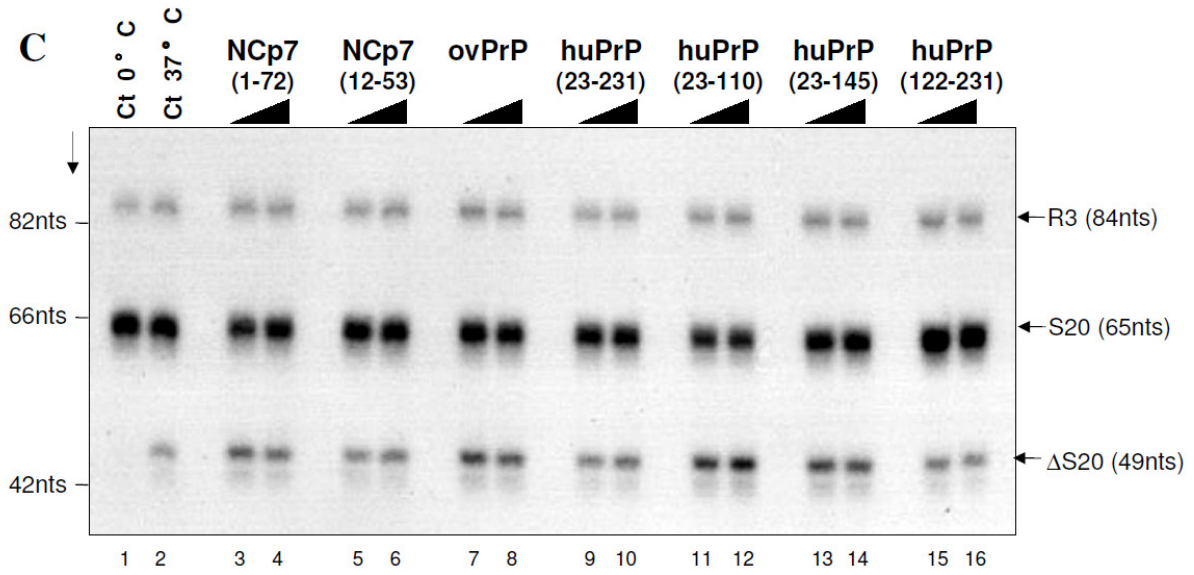


Figure 3C

A. Scheme of the RNA *trans-splicing* reaction.

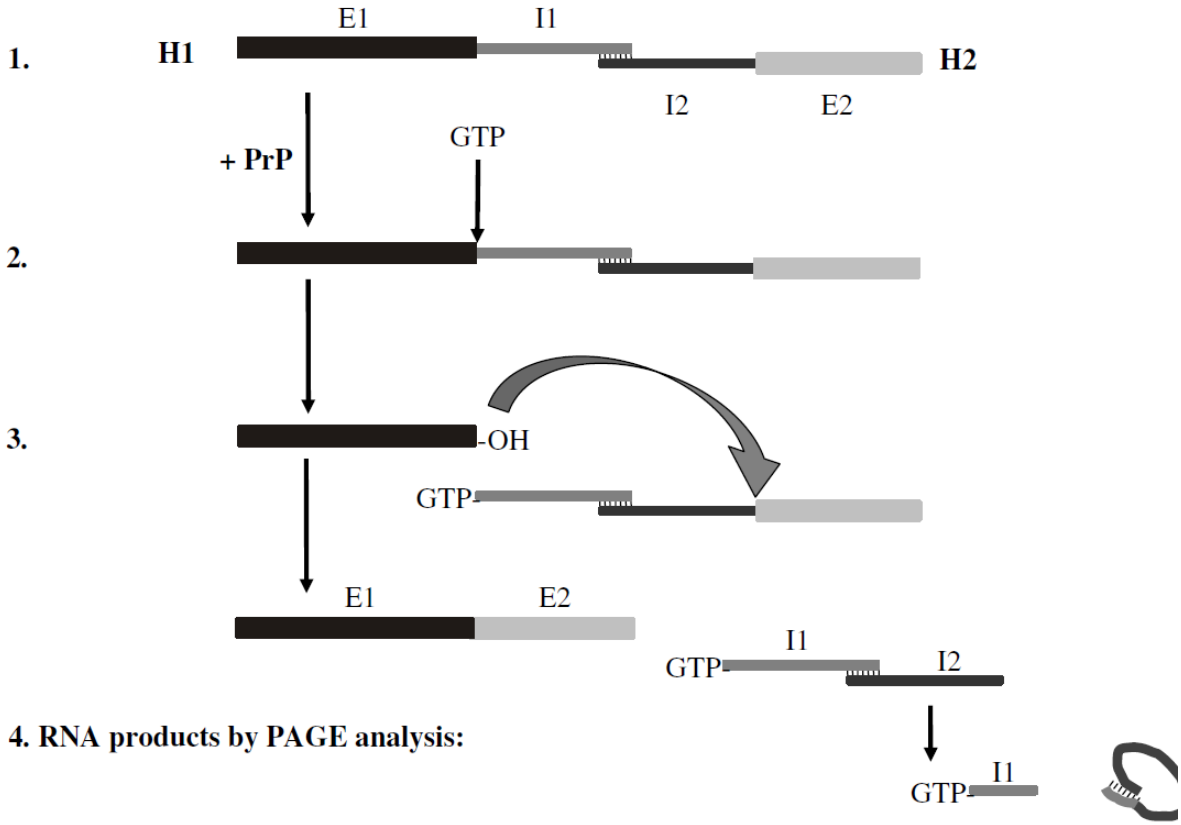


Figure 4A

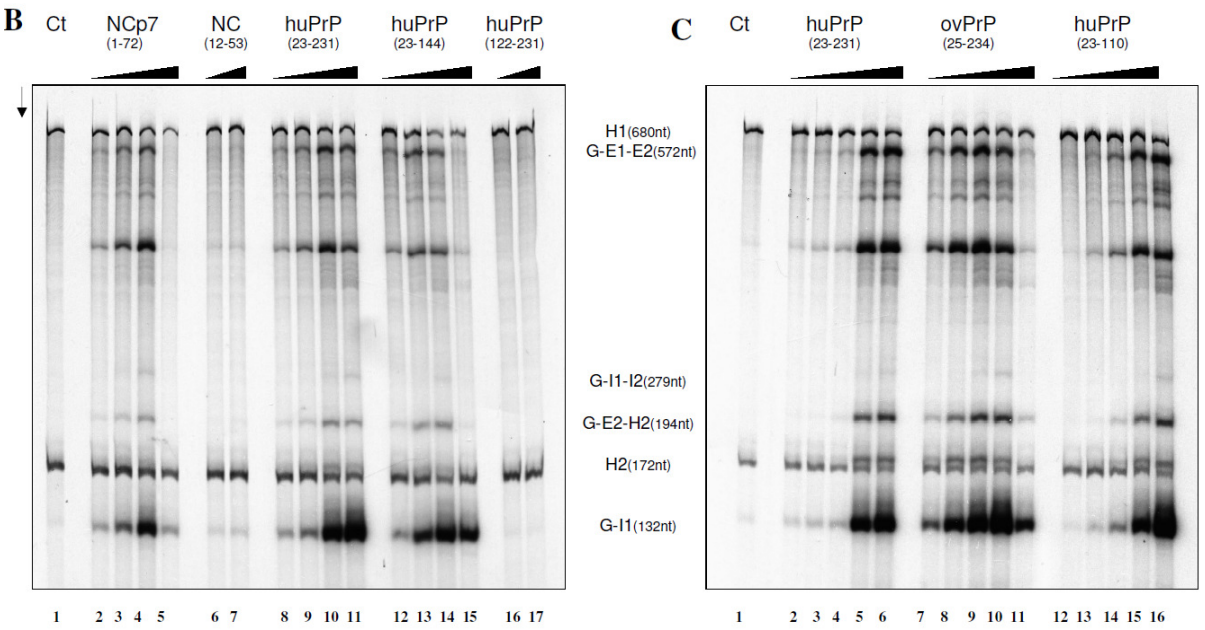


Figure 4B and C

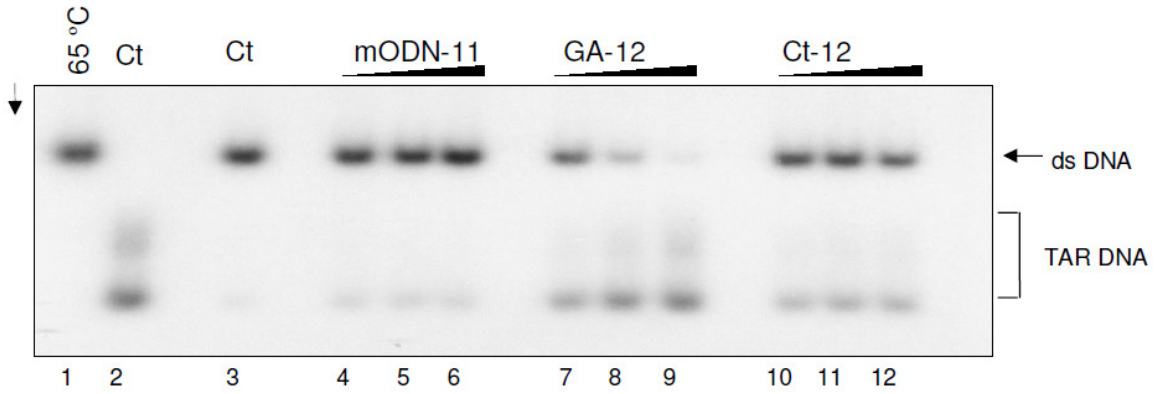


Figure 5A

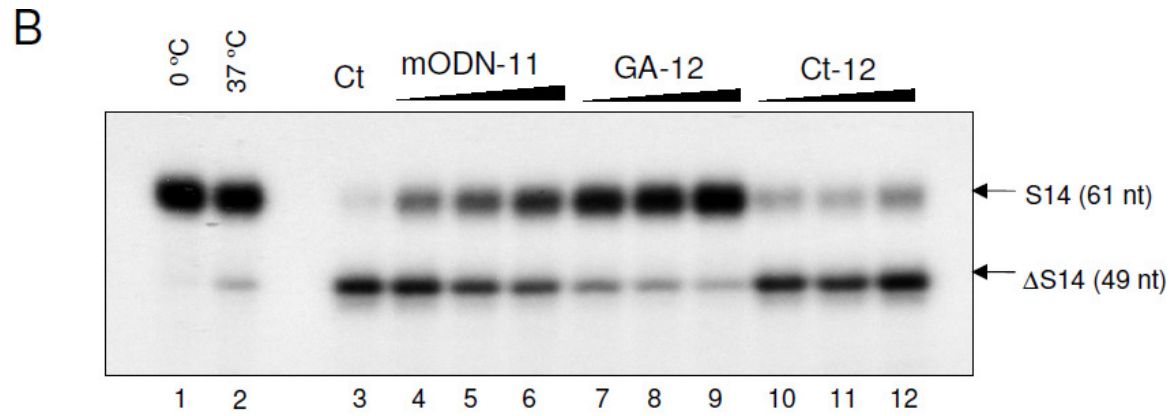


Figure 5B

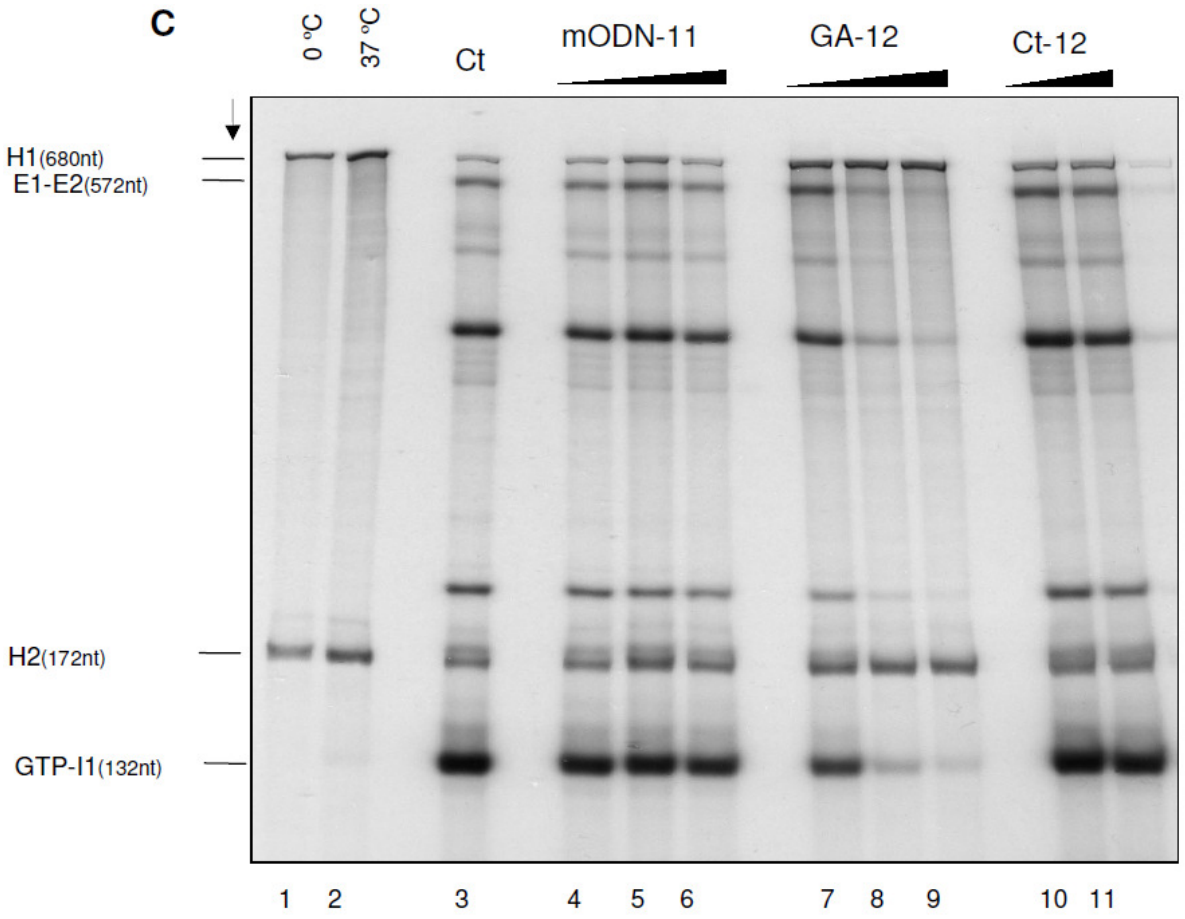


Figure 5C

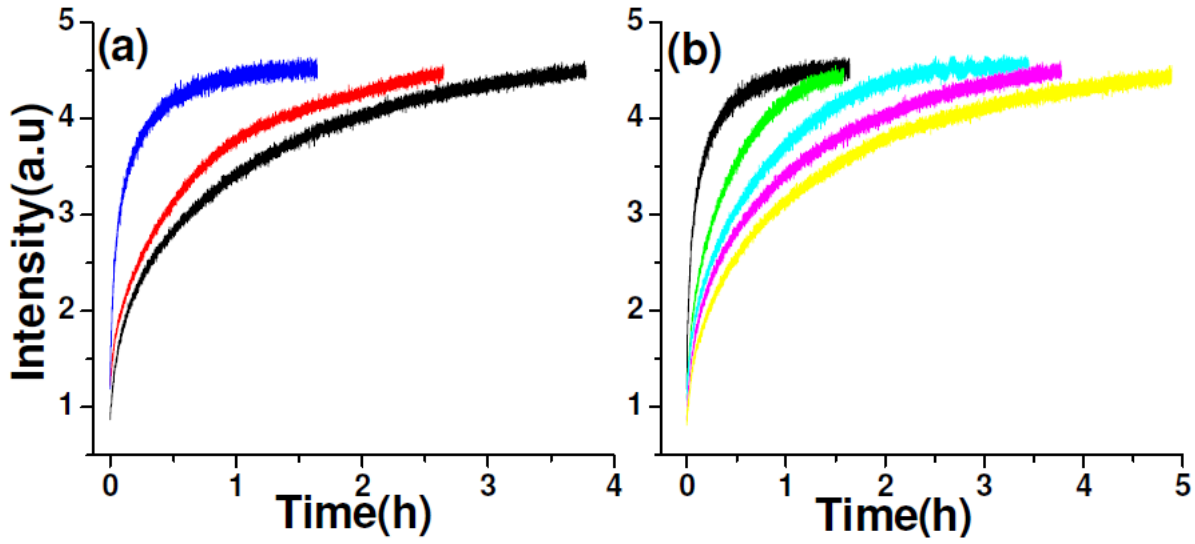


Figure 6

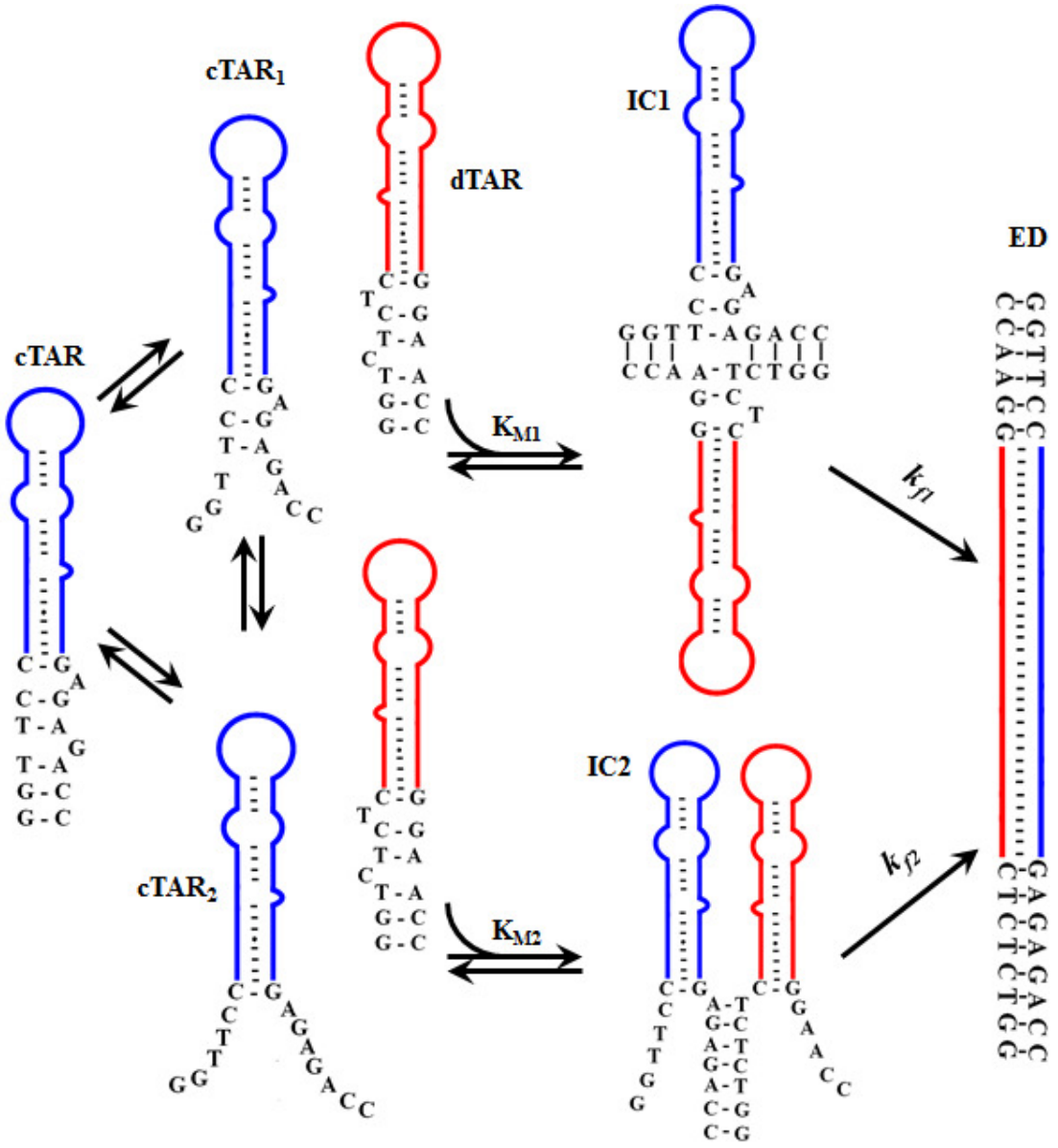


Figure 7

Supplementary Material:

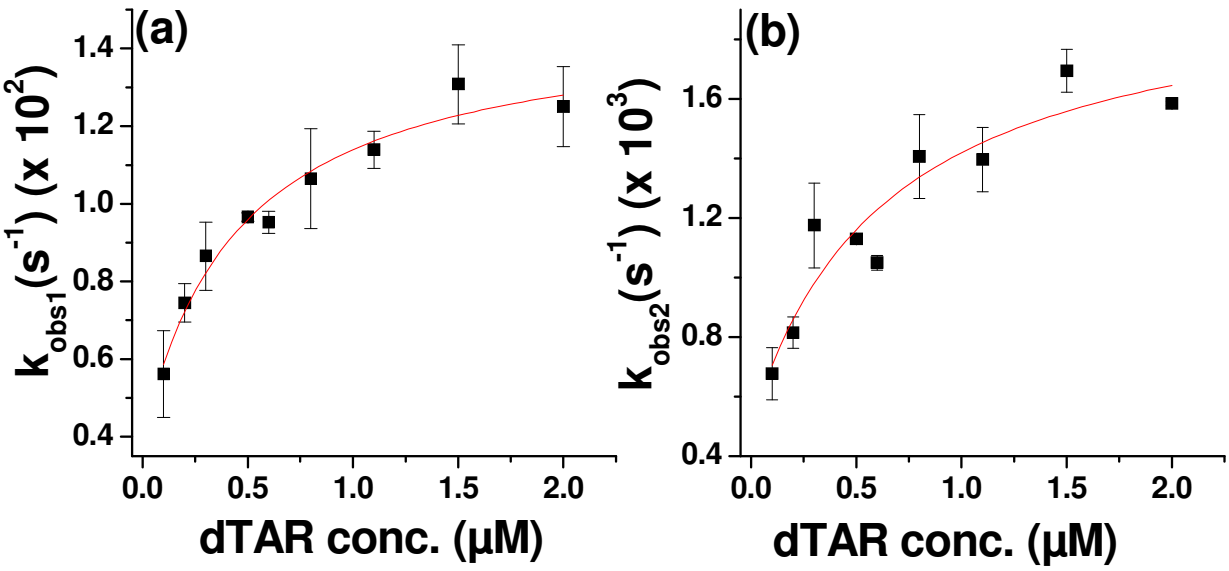


Figure 1A: Kinetic parameters of huPrp(23-110)-promoted cTAR/dTAR annealing. The fast (k_{obs1}) (a) and slow (k_{obs2}) (b) components were determined in pseudo-first order conditions from real-time kinetics. The solid lines correspond to the fit of the data with: $k_{obsi} = \frac{k_{fi} K_{Mi}[dTAR]}{1 + K_{Mi}[dTAR]} + k_{bi}$, where k_{fi} is the forward rate constant for the interconversion of IC into ED, K_{Mi} is the equilibrium binding constant for the intermediate complex, and k_{bi} is the backward rate constant for the conversion of IC into ED, and corresponds to the Y-axis intercept. For both k_{obs1} and k_{obs2}, the k_{bi} values were found to be close to 0.

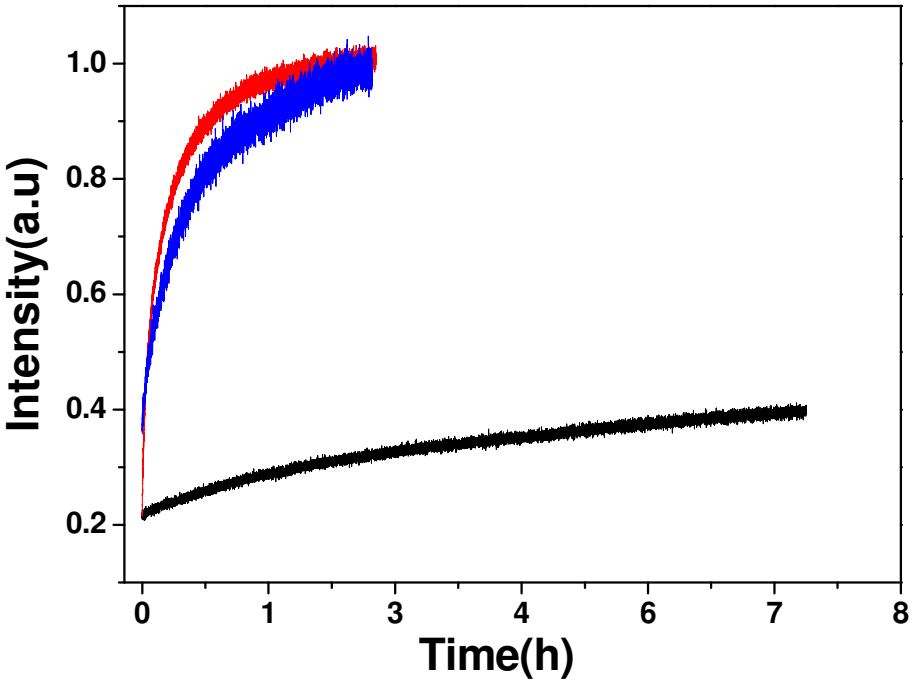


Figure 1B: Kinetics of huPrP(23-110) -promoted annealing of cTAR derivatives with dTAR derivatives. Kinetic traces of 10 nM doubly labelled TMR-cTAR-Fl derivatives with 100 nM non-labelled dTAR derivatives [cTAR/dTAR (red), cTAR/dTAR-TL (blue) and cTAR1.2/dTAR (black)]. huPrP(23-110) was added at a peptide/oligonucleotide ratio of 1:1. Excitation and emission wavelengths were 480 nm and 520 nm, respectively.

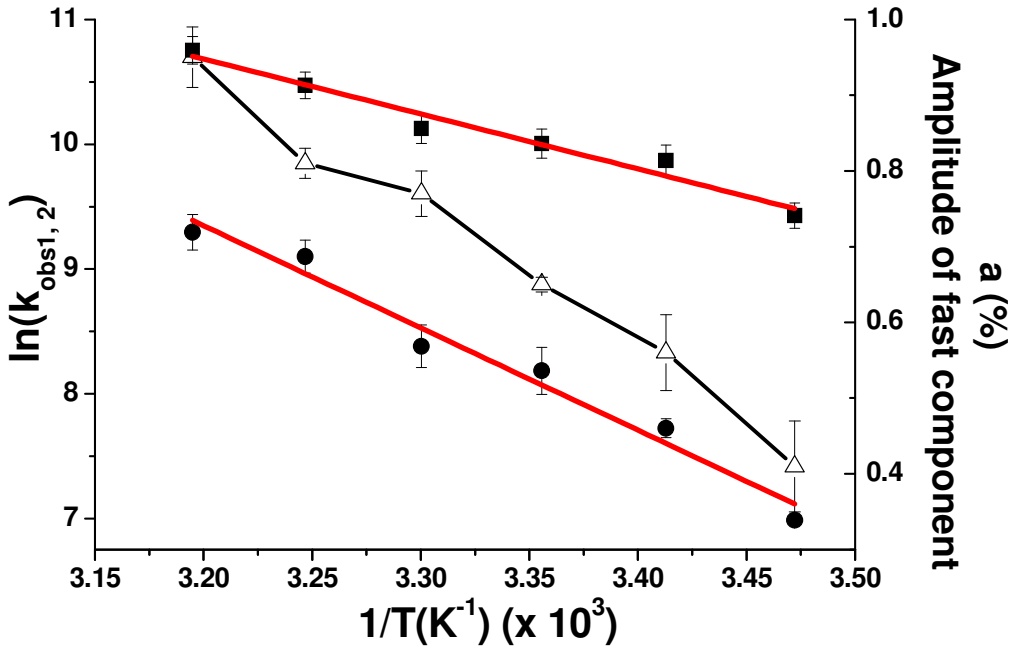


Figure 1C: Temperature dependence of cTAR/dTAR annealing kinetics in the presence of huPrP(23-110). The reaction was triggered by adding 500 nM non-labelled dTAR to 10 nM doubly labelled cTAR, in the presence of huPrP(23-110) added at a protein/oligonucleotide ratio of 1:1. The natural logarithm of the rate constant values for the fast (closed squares) and slow (closed circles) components as well as the amplitude of the fast component (open triangles) are plotted as a function of the inverse of the temperature. The solid red lines are the best fit to Equation (2) with E_{a1} = 8.8(±0.7) kcal/mol and E_{a2} = 17(±1) kcal/mol, for the fast and slow components, respectively.

.

Chapter 2: Kinetics of (-)PBS/(+)PBS annealing in the presence of NC mutants peptides

Conversion of the human immunodeficiency virus type I (HIV-1) genomic RNA into the proviral DNA by reverse transcriptase involves two obligatory strand transfers that are chaperoned by the nucleocapsid protein (NCp7). The second strand transfer relies on the annealing of the (-) and (+) copies of the primer binding site, (-)PBS and (+)PBS, which fold into complementary stem-loops. NCp7 increases the (-)PBS/(+)PBS annealing kinetics and modifies the mechanistic pathway of the annealing reaction by strongly activating the loop-loop kissing pathway, that poorly contributes to the annealing reaction in the absence of NCp7. To understand this mechanism the restriction dynamics of (-) PBS loop were followed by using 2-Amino-purin strategy and annealing reactions of doubly-labeled (+)PBS with non-labelled (-)PBS in the presence of NCp7 and its different mutants were performed. Interestingly, a strong correlation was found between the ability of NCp7 mutants to restrict the dynamics of the bases within the loop and their efficiency in (-)PBS/ (+)PBS annealing, clearly indicating that NCp7 binding on the (-)PBS loop directs the formation of the competent (-)PBS loop conformations for annealing. Data represented in this publication showed the importance of hydrophobic platform as well as structure of two NCp7 zinc fingers in controlling the timing of reverse transcription and the faithful synthesis of viral DNA.

Nucleic Acids Research

Synthesis of HIV-1 DNA requires the specific intervention of the nucleocapsid zinc fingers at the level of the obligatory plus strand transfer

Journal:	*Nucleic Acids Research*
Manuscript ID:	Draft
Manuscript Type:	1 Standard Manuscript
Key Words:	HIV, nucleocapsid, fluorescence, chaperone, strand transfer

Synthesis of HIV-1 DNA requires the specific intervention of the nucleocapsid zinc fingers at the level of the obligatory plus strand transfer

Julien Godet[1,2,#], Nick Ramalanjaona[1,#], Kamal K Sharma[1], Ludovic Richert[1], Hugues de Rocquigny[1], Jean-Luc Darlix[3], Guy Duportail[1], Yves Mély[1]*

[1] Laboratoire de Biophotonique et Pharmacologie, UMR 7213 CNRS, Université de Strasbourg, Faculté de Pharmacie, 74 route du Rhin, 67401 Illkirch, France

[2] Medical Information and Biostatistics Department, Strasbourg University Hospital, 67000 Strasbourg, France

[3] LaboRétro Unité de Virologie Humaine INSERM 758, IFR 128 Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 LYON, France

# JG and NR contributed equally to this work

* Corresponding author

Tel : +33 (0)3 68 85 42 63 Fax : +33 (0)3 68 85 43 12

e-mail: yves.mely@unistra.fr

Synthesis of the HIV-1 viral DNA by reverse transcriptase involves two obligatory strand transfers. The second strand transfer reaction corresponds to the annealing of the (-) and (+) DNA copies of the primer binding site (PBS) sequence and is strongly promoted by nucleocapsid protein (NCp7), which modifies the annealing mechanism by activating a loop-loop kissing pathway that is negligible without NCp7. To characterize in depth the dynamics of the loop in the NCp7/PBS complexes, we investigated the time-resolved fluorescence parameters of a (-)PBS derivative containing the fluorescent nucleoside analogue 2-aminopurine at position 6, 8 or 10. The NCp7-directed switch of (+)/(-)PBS annealing towards the loop pathway was associated to a drastic restriction of the loop dynamics, indicating that NCp7 was able to "freeze" PBS conformations competent for annealing via the loops. Moreover, the modifications of the PBS loop structure and dynamics that govern the annealing mechanism were found strictly dependent on the zinc finger hydrophobic platform. Our data suggest that the two NCp7 zinc fingers are required to ensure the specificity and fidelity of the second strand transfer, further underlining the pivotal role played by NCp7 to control the timing of reverse transcription and the faithful synthesis of viral DNA.

Introduction

Human immunodeficiency virus type 1 (HIV-1) viral DNA synthesis is a complex multi-step process catalyzed by the viral reverse transcriptase (RT). Synthesis of the complete viral DNA requires two obligatory DNA transfer reactions. During the first strand transfer, the minus-strand strong-stop DNA ((-)ssDNA) is translocated to the 3' end of the viral RNA genome, in a reaction mediated by base-pairing of the repeat sequences at the 3' ends of the RNA and cDNA reactants, to allow reverse transcription to resume and to continue up to the 5' end of the primer binding site, (-) PBS. The second strand transfer relies on (-)PBS annealing to the (+)PBS sequence located in the plus strand strong stop DNA (ssDNA(+)), the copy of the minus strand synthesized by RT(1,2). The annealing of the two complementary PBS DNA stem-loops (Figure 1) enables RT to resume and complete viral DNA synthesis. These two obligatory strand transfers are chaperoned by the HIV-1 nucleocapsid protein (NCp7), a potent nucleic acid chaperone that plays major roles in the viral replication cycle (3-6). The mature NCp7 is a 55 amino-acids protein encoded by the HIV-1 Gag polyprotein (Figure 1). It contains two highly conserved CCHC zinc fingers (ZFs) that coordinate zinc ions with high affinity (7). The two folded ZFs are connected by a highly conserved basic sequence and flanked by N- and C-terminal basic domains. NCp7 catalyzes the annealing reactions of the complementary (+)/(-)PBS sequences (8,9), due to its ability to chaperone the rearrangement of nucleic acids into their most thermodynamically stable conformations (6,10). In the absence of NCp7, (+)PBS can spontaneously anneal to (-)PBS *in vitro* (8). This annealing reaction proceeds mainly through the single-strand overhangs of the PBS sequences while nucleation through loop-loop interaction appears negligible (Figure 2). NCp7 was found to increase the (-)/(+)PBS annealing kinetics by about 60-fold by strongly accelerating the loop pathway, which thus becomes the major pathway. This switch in the annealing mechanism probably results from significant changes in the loop structure and/or dynamics upon NCp7 binding.

Structural changes of the loop were evidenced by solving the structure of NCp7 complexed to ΔP(-)PBS, a (-)PBS derivative without the 3' protruding sequence (11). NCp7 was shown to bind preferentially to the 5' end of the loop (5-CTG-7) of ΔP(-)PBS. This binding is mediated by the hydrophobic plateau, involving the Val13, Phe16, Thr24, Ala25, Trp37, Gln45 and Met46 residues that form on the top of the folded ZFs (12-14). The Phe16 and Trp37 residues insert between the T6 and G7 bases, allowing a tight stacking of the Trp37 with G7 (15), as well as the flipping of the T6 and G7 bases towards the exterior of the loop. Through this binding, NCp7 also stretches the loop, which increases the accessibility of the 8-TTC-10 nucleobases and slightly destabilizes the upper base pair of the stem. As a consequence of these structural changes in PBS, NCp7 is thought to favour the kissing interactions between (-)PBS and (+)PBS loops and help disrupting the stem to convert the kissing complex into the final extended duplex.

To further characterize the molecular mechanism and protein determinants responsible for the (-)/(+)PBS annealing chaperoned by NCp7, we studied the annealing mechanism using PBS mutants and monitored the thermodynamic and kinetic parameters with various NCp7 mutants. We also investigated the dynamic changes of both (-)PBS and (+)PBS loops in response to NCp7 binding using PBS sequences labelled with 2-aminopurine (2-Ap), an environment-sensitive fluorescent analogue of adenine (16). We found that the intact ZFs were critical to modify the structure and restrict the dynamics of both PBS loops and to switch the annealing mechanism towards the loop-loop kissing pathway. In contrast, the basic domains of NCp7 were able to modulate the annealing rate constants but unable to modify the annealing mechanism. Thus, our data indicated that the hydrophobic plateau at the top of the folded ZFs is needed to direct the formation of competent PBS loop conformations for annealing, ensuring the specificity of the second strand transfer reaction.

Materials and Methods

Materials

The NCp7, NC(11-55), $(SSHS)_2$NC(11-55), L_{37}NC(11-55) and A_{16}NC(11-55) peptides were synthesized on a Applied Biosystems A433 peptide synthesizer, as described (17,18). Unmodified or labelled ODNs were synthesized and HPLC- or PAGE-purified by IBA Gmbh Nucleic Acids Product Supply (Germany). A 2'-deoxyribosyl-2-aminopurine (2-Ap) was selectively introduced at different positions (6, 7 or 10) within the ΔP(-)PBS loop and at position 11 of ΔP(+)PBS, substituting the corresponding natural base. Doubly-labelled (+)PBS sequences were modified at their 5' terminus with 6-carboxyrhodamine (Rh6G) and their 3' terminus with 4-(4'-dimethylaminophenylazo)benzoic acid (DABCYL), via an amino linker with a six carbon spacer arm. All experiments were performed in 25 mM TRIS-HCl, pH 7.5, 30 mM NaCl and 0.2 mM $MgCl_2$ at 20°C, unless specified otherwise.

Steady-state fluorescence spectroscopy

Fluorescence emission spectra were recorded with a Fluorolog or a Fluoromax-3 spectrofluorimeter (Jobin Yvon) equipped with a thermostated cell compartment. All fluorescence intensities were corrected for screening effects, buffer emission and lamp fluctuations. The quantum yield was calculated using free 2-Ap riboside as a reference (quantum yield = 0.68 (16)). 2-Ap was excited at 315nm.

Kinetics of (-)/(+)PBS annealing was monitored in real-time using fluorescent doubly-labelled (+)PBS sequences and non-labelled (-)PBS. Excitation and emission wavelengths were at 520 nm and 550 nm, respectively, to monitor the fluorescence restoration of Rh6G resulting from the

formation of the (-)/(+)PBS duplex. Concentrations of 5'Rh6G-(+)PBS-3'DABCYL and (-)PBS were 10 nM and 100 nM to 1.1 µM, respectively, to ensure pseudo-first order conditions. Both reactants (in identical volumes) coated by NC peptides were mixed together to trigger the reaction. The apparent rate constants k_{obs} were determined from the kinetic data, as previously described (8). All fitting procedures were carried out with the Microcal Origin 6.1 software based on non-linear least-squares methods, applying the Levenberg-Marquardt algorithm.

Time-resolved fluorescence measurements

Time-resolved fluorescence measurements were performed with the time-correlated, single-photon counting technique. Excitation pulses were generated by a pulse-picked frequency-tripled Ti-sapphire laser (Tsunami, Spectra Physics) pumped by a Millenia X laser (Spectra Physics) (19). Excitation wavelength was set at 315 nm, with a repetition rate of 4 MHz. The fluorescence emission was collected through a polarizer set at magic angle and a 16 mm band-pass monochromator (Jobin Yvon) at 370 nm. The single-photon events were detected with a micro-channel plate photomultiplier (Hamamatsu) either coupled to a pulse pre-amplifier (Philips) and recorded on a multi-channel analyser (Ortec) calibrated at 25.5 ps/channel or coupled to a pulse pre-amplifier HFAC (Becker-Hickl) and recorded on a SPC-130 board (Becker-Hickl). The instrumental response function (IRF) was recorded using a polished aluminium reflector, and its full-width at half-maximum was ~ 40 ps. The mean lifetime $\langle\tau\rangle$ was calculated from the individual fluorescence lifetimes (τ_i) and their relative amplitudes (α_i) according to $\langle\tau\rangle=\Sigma\alpha_i\tau_i$. The population, α_0, of dark species of 2-Ap was calculated by: $\alpha_0 = 1 - \tau_{free} / (\tau_{ODN} R_m)$, where τ_{free} is the lifetime of the free 2-Ap, τ_{ODN} is the measured mean lifetime of 2-Ap within the ODN and R_m is the ratio of their corresponding quantum yields. The remaining amplitudes, α_{ic} were recalculated from the measured amplitudes according to $\alpha_{ic} = \alpha_i / (1- \alpha_0)$.

Time-resolved anisotropy, resulting from the measurement of the fluorescence decay curves recorded in directions parallel ($I_{//}$) and perpendicular ($I_\perp$) alternatively, to the excitation beam polarization, was analysed by the following equations:

$$I_{//}(t) = I(t)\left[1 + 2r(t)\right]/3 \qquad \text{Equation (1)}$$

$$I_{\perp}(t) = I(t)\left[1 - 2r(t)\right]/3 \qquad \text{Equation (2)}$$

$$r(t) = \frac{I_{//}(t) - G \times I_{\perp}(t)}{I_{//}(t) + 2G \times I_{\perp}(t)} = r_0 \sum_i \beta_i \times \exp\left(-t/\Phi_i\right) \qquad \text{Equation (3)}$$

where β_i are the amplitudes of the rotational correlation times Φ_i, and G is the geometry factor at the emission wavelength, determined in independent experiments. Theoretical values of the rotational correlation times and the values of the cone semi-angle θ_0 for the local motion of 2Ap

were calculated as described in the supplementary material. Time-resolved intensity and anisotropy data were treated according to the maximum entropy method (Pulse 5 software) (20,21) or according to a non-linear least-square analysis using a homemade software (kindly provided by G. Krishnamoorthy). In all cases, the χ^2 values were close to 1 and the weighted residuals as well as their autocorrelation were distributed randomly around 0, indicating an optimal fit.

Results

Characterization of the ΔP(-)PBS and ΔP(+)PBS sequences labelled with 2-Ap

2-Ap was introduced at three positions of the ΔP(-)PBS loop, namely at the 5' end (2-Ap6), in the middle of the loop (2-Ap8), within the major binding site for NCp7, or at the 3' end of the loop (2-Ap10) within the low affinity binding site for NCp7 (Figure 1). Since 2-Ap is able to report locally on the dynamics of the ODN sequence in which it is inserted, we aimed to probe the whole loop with these three substitutions. The quantum yield of the 2-Ap residues inserted in ΔP(-)PBS was largely reduced as compared to the free probe (16) (Table 1), indicating a strong fluorescence quenching by the neighbor bases (22). The quenching was the most pronounced for 2-Ap at position 10, suggesting a strong stacking with G11. This inferred stacking is in full line with the stacking of C10 with G11 observed by NMR (11) indicating that substitution of C10 by 2-Ap likely preserves the local structure of the loop. Taking into account that each 2-Ap was flanked by a guanine, the most efficient quencher of 2-Ap among the natural bases (23), the quantum yield of 2-Ap at positions 6 and 8 appeared comparatively higher than the quantum yields observed in single-stranded ODNs (24-26). This suggested that the PBS loop is at least partially ordered, so that efficient collisions with neighbour bases, and notably with the G7 base, are restricted. This is again in line with the NMR data (11) since the T8 base was found to be perpendicular to the G7 residue, while the T6 base could stack with C5 but not with G7. Thus, our data strongly suggest that 2-Ap substitutions of the natural bases only marginally perturb the native folding of the loop.

The time-resolved intensity decays of the 2Ap-labeled ODNs were complex, showing that conformational fluctuations of the loop occurred during the excited state of the 2-Ap residues. These decays were best fitted with four discrete lifetime components, ranging from 0.1 ns to nearly 9 ns (Table 1), indicating that 2-Ap experienced at least four conformational states. Moreover, since the mean lifetime was found to be only 4-6.5 times shorter than the 10.2 ns of the free 2-Ap, far from the ratios (from 25- to 155-fold) of their corresponding quantum yields, additional conformations associated to ultra-short-lived lifetimes below the time resolution of our set-up have to be considered (25). These dark-species were previously shown to result from ultra-fast dynamic quenching (27).

Dark-species and weakly emitting species were by far the most populated conformations since they

represented about 90% of the 2-Ap conformations, as it could be seen from the sum of the α_0 and α_1 amplitudes (Table 1). This efficient dynamic quenching of 2-Ap by its neighbour residues can be related to conformational fluctuations of the loop in the picosecond–nanosecond (ps–ns) range that allowed quenching of 2Ap through a charge transfer mechanism (27-30). The stacked conformations rose to 98% for Ap10, further confirming a strong stacking with G11. The values of the longer-lived lifetime τ_4, associated to extra-helical or unstacked conformations, were close to the lifetime of free 2-Ap, but low-populated, representing from 1 to 4% of the whole conformations (Table 1). The high value of the long-lived τ_4 lifetime confirmed the limited flexibility of the (-)PBS loop, since in short ss-DNAs the corresponding lifetime did not exceed 5 ns (25,31), due to efficient collisions caused by the high flexibility of these sequences during the probe lifetime. Moreover, the low amplitude α_4 observed for 2-Ap6, 2-Ap8 and 2-Ap10, suggested that stable conformations with bases exposed to the solvent were marginal, in full agreement with the orientation of the bases towards the interior of the PBS loop (11,32).

The local dynamics of the 2-Ap residues was further explored by time-resolved fluorescence anisotropy which provides information on their rotational dynamics. Fluorescence anisotropy decays were adequately fitted with a three-exponential model (Table 2 and Figure 3). The two shorter components presumably correspond to the local rotation of the dye and the segmental mobility of the loop, respectively. The slowest correlation time of 2.1 ns to 2.6 ns was attributed to the global tumbling of the DNA sequence. Its deviation from the theoretical correlation time (1.75 ns) calculated for a sphere with the same molecular mass than the ODN (33) was related to the non-spherical shape of the PBS stem-loop. The relative amplitude of the fastest component allowed estimating the angular range of the local motion of 2-Ap modelled as the diffusion in a semi-angle cone (θ_0). These angles were found similar for the 2-Ap residues at the three positions and significantly lower to those observed in small single-stranded ODNs (24-26), confirming that the degree of freedom of the ΔP(-)PBS loop bases was restricted as compared to ssDNAs. The restricted rotation of the 2-Ap bases is in full line with the internal orientation of the corresponding bases in the loop, which provides a relatively crowded environment (11,32).

To determine whether the dynamics of the (+)PBS loop was comparable to that of (-)PBS, we site-specifically modified ΔP(+)PBS at position 11 with 2-Ap (Figure 1). The quantum yield of 2-Ap11 in ΔP(+)PBS was somewhat higher than in the ΔP(-)PBS derivatives, as a result of a sharp decrease of the α_0 value in favour of the amplitudes associated to the τ_1 and τ_2 lifetimes, likely due to the absence of guanine residues flanking the 2-Ap11 residue (24,25). As for the ΔP(-)PBS derivatives, the number of lifetimes and distribution of amplitudes for 2-Ap11 in ΔP(+)PBS were in line with fast conformational fluctuations in the ps-ns range. Moreover, the τ_4 and α_4 values, as well as the

time-resolved fluorescence anisotropy parameters, were very similar to those in ΔP(-)PBS, leading also to the conclusion of a partial order within the ΔP(+)PBS loop. Taken together, our data suggest that the dynamics of the ΔP(+)PBS loop is similar to that of ΔP(-)PBS.

Effect of NCp7 on the dynamics of the Δ P(-)PBS and ΔP(+)PBS loop

To characterize the effect of NCp7 on the dynamics of the ΔP(-)PBS and ΔP(+)PBS loops through the 2-Ap fluorescence changes, we first checked whether the 2-Ap substitutions impacted on the binding of NCp7 to these PBS derivatives. By monitoring the binding of the 2-Ap-labeled ODNs to NCp7 through the quenching of the intrinsic Trp37 residue (34,35), we found that the 2-Ap substitutions did not induce significant changes in the binding parameters (see supplementary data), in line with the aforementioned marginal changes induced by 2-Ap substitutions in the PBS structure.

Upon NCp7 addition to the 2Ap-labelled ODNs at a ratio of 3:1 to ensure saturation of the two protein binding sites on the loop (11), the maximum fluorescence emission wavelength of 2-Ap at about 365 nm was not modified, but its fluorescence quantum yield was increased by a factor of 2 to 5 according to its position in the ODN (Table 1). As a consequence, the binding of NCp7 on ΔP(-)PBS or ΔP(+)PBS loop significantly reduced the level of quenching of 2-Ap fluorescence by its neighbor bases but did not change the polarity of its environment. Furthermore, the time-resolved intensity decays revealed that the increase in the 2-Ap quantum yield was mainly due to a decrease in the populations of the dark species to the benefit of the most emitting species, as well as to an increase in the τ_4 value. These changes in the amplitude and lifetime values suggested that NCp7 severely restricted the stacking and collisions of 2Ap with its neighbour residues. This restriction in stacking and collisions is in line with the ability of NCp7 to stretch the entire ΔP(-)PBS loop and direct the T6 and G7 bases toward the exterior of the loop (11), which markedly increases the distance between the bases. In this respect, the large changes in the amplitudes and τ_4 values observed for 2-Ap at position 8 of ΔP(-)PBS, are likely a consequence of the stacking of the Trp37 residue with G7, which prevents its collisions and stacking with 2-Ap8. Furthermore, the large changes in the amplitudes and τ_4 values of 2-Ap10 at the 3'end of the ΔP(-)PBS loop indicated that NCp7 binding to its second site in the loop induces similar restrictions on the stacking and collisions of 2-Ap.

Time-resolved anisotropy decays further revealed that binding of NCp7 to the ΔP(-)PBS loop induced a strong decrease in the amplitude associated with the local motion of the 2-Ap bases, irrespective of their position in the loop. The restriction of the local motion of the bases at positions 6 and 8, is fully consistent with the numerous contacts observed by NMR between NCp7 and these

bases when NCp7 binds to the 5' end of the loop (11). Interestingly, we observed that binding of NCp7 to the 3' end of the loop induced a similar freezing of the local motion of 2-Ap at position 10, suggesting that NCp7 induced similar restrictions on the local mobility of the bases in its two binding sites. The binding of at least two NCp7 molecules on the 2-Ap-labeled ODNs was clearly confirmed by the 9-10 ns value of the Φ_3 correlation time, which is in line with the theoretical values of 7 ns and 10 ns expected, respectively, for a 2:1 complex and a 3:1 complex with a spherical shape. Interestingly, in all ΔP(-)PBS derivatives, the amplitude associated with the Φ_3 correlation time strongly increased upon NCp7 binding, showing an overall decrease in the loop flexibility, so that the loop and the stem tumble as a whole. Taken together, our data indicate that binding of NCp7 molecules to its two binding sites on ΔP(-)PBS loop strongly restricts the picosecond to nanosecond dynamics of the loop, by constraining both the overall flexibility of the loop and the local mobility of the bases. A major consequence of the restricted dynamics of the loop is that it results in base unstacking. Similar conclusions can be drawn from the time-resolved anisotropy decays of ΔP(+)PBS, suggesting that NCp7 constrains the overall flexibility and the local mobility of the loops of two PBS DNA partners.

Effect of NCp7 mutants on the dynamics of the Δ P(-)PBS and ΔP(+)PBS loop

To identify the protein determinants responsible for the NCp7-induced changes in the structure and dynamics of the Δ P(-)PBS and ΔP(+)PBS loops, we used a series of NCp7 mutants (Figure 1). The contribution of the N-terminal domain was investigated with NC(11-55), a peptide composed of the ZF domain but lacking the basic N-terminal domain. Binding of NC(11-55) to ΔP(-)PBS labelled by 2-Ap at positions 6, 8 or 10 or to ΔP(+)PBS labelled at position 11 induced changes in both steady-state and time-resolved fluorescence parameters similar to those of NCp7 (Tables 1 and 2), indicating that the NCp7-induced modifications in the structure and dynamics of the ΔP(-)PBS or ΔP(+)PBS loop are largely mediated by the ZF domain.

Next, we investigated the role of the Zn^{2+}-induced folding of the NCp7 fingers, by using the $(SSHS)_2$NC(11-55) mutant where all cysteins are substituted for serines, in order to prevent the binding of zinc and thus the folding of the fingers (36,37). Binding of this mutated peptide to the 2Ap-labeled ODNs was evidenced by the large increase of the Φ_3 correlation time (Table 2), which suggests that at least two peptides bind to the ODN. However, in sharp contrast to NCp7 and NC(11-55), $(SSHS)_2$NC(11-55) induced negligible changes in the steady-state and time-resolved parameters of the 2Ap-labeled ODNs, indicating that this mutant only slightly restricts the flexibility of the loops and the local mobility of the bases. Thus, the folding of the ZFs appears critical for the NCp7-induced changes in the structure and dynamics of the ΔP(-)PBS and

ΔP(+)PBS loop.

In a next step, the contribution of the hydrophobic plateau at the top of the two ZFs was investigated by mutating the two conserved aromatic residues of this plateau that play a critical role in ODN binding (38-40). To that end, Phe_{16} and Trp_{37} were substituted by Ala and Leu in the A_{16}NC(11-55) and L_{37}NC(11-55) mutants, respectively. Since these two amino-acids are not involved in zinc chelation, these mutations were expected to not alter the folding of the ZFs (13). Due to their limited affinity for ODNs (25), A_{16}NC(11-55) and L_{37}NC(11-55) peptides were added in large excess (1 peptide per 2 nt and 1 peptide to 1 nt, respectively), to ensure full coating of ΔP(-)PBS, as evidenced by the large Φ_3 values comparable to those obtained with NC(11-55). Both A_{16}NC(11-55) and L_{37}NC(11-55) mutants induced only a limited increase in the quantum yield of the different 2-Ap-substituted ΔP(-)PBS sequences, suggesting that these mutations dramatically alter the ability of NCp7 to reduce base stacking and restrict base collisions. This was confirmed by the limited lifetime redistribution towards the less quenched conformations. In addition, the two mutants induced only a moderate restriction of the local mobility of 2-Ap as compared to NC(11-55), demonstrating the critical role of the hydrophobic platform in this process. Moreover, it should be noted that these two mutants show asymmetric effects in respect to the unstacking of 2-Ap in positions 6 and 8. Indeed, A_{16}NC(11-55) was more efficient than W_{37}NC(11-55) to reduce the amplitude α_0 associated to dark species (0.70 vs 0.79) and the local mobility of 2-Ap in position 8 (β_1=0.18 vs 0.26). In contrast, at position 6, L_{37}NC(11-55) appeared more efficient (α_0 = 0.62 vs 0.78 and β_1 = 0.21 vs 0.28). This asymmetric effect was consistent with NMR data showing that in the NC(12-53)/ΔP(-)PBS complex, F_{16} interacts with T6 while W_{37} interacts with G7(11).

These observations indicate that the ability of NCp7 to unstack and reorient the bases of the loop and to restrict the local and overall dynamics of ΔP(-)PBS or ΔP(+)PBS are mainly mediated by the hydrophobic plateau at the top of the ZFs.

NCp7-mediated restriction of the PBS loop dynamics in the (-)/(+)PBS annealing reaction

To further evaluate the relevance of the NCp7-induced restriction of the dynamics of PBS loop, and notably its contribution to the promotion of (-)/(+)PBS annealing, we investigated the annealing reactions of doubly-labelled (+)PBS with (-)PBS in the presence of NCp7 and its mutants. The initial fluorescence of the folded 5'Rh6G-(+)PBS-3'DABCYL was very low due to its stem-loop structure, which brings the Rh6G fluorophore in close vicinity to the DABCYL group, acting as a quencher. Addition of NCp7 or NC(11-55) to the doubly-labelled (+)PBS at a ratio of one NCp7 per 5 nt induced a small fluorescence increase, in line with the weak destabilizing activity of NCp7 on the PBS stem (41). In contrast, the $(SSHS)_2$NC(11-55) or L_{37}NC(11-55) mutants did not induce any increase in the Rh6G fluorescence, confirming the critical role of the Trp37 residue and the ZFs in

the NCp7 destabilizing activity (38). Addition of (-)PBS to 5'Rh6G-(+)PBS-3'DABCYL allowed the formation of the extended (+)/(-)PBS duplex, inducing an important increase of the dye-to-dye distance and thus, Rh6G fluorescence restoration. In the absence of peptide, the (+)/(-)PBS hybridization occurred spontaneously with a rather slow rate (Table 3). In the presence of NCp7, an about two-order of magnitude increase in the rate of the (+)/(-)PBS annealing reaction was observed. The NC(11-55) and L_{37}NC(11-55) mutants also increased the annealing reaction rate, albeit to a lesser extent than NCp7 (5- to 6-fold increase). Interestingly, a much stronger increase (30-fold) in the annealing rate was obtained with the $(SSHS)_2$NC(11-55) mutant (see Figure S1 of the supplementary data). Since this mutant was shown to primarily promote annealing reactions through its electrostatic nucleic acid aggregating component (42,43) this component appears to play a major role in the (+)/(-)PBS reaction rate.

To further dissect the effect of the NCp7 mutants on the (+)/(-)PBS annealing reaction, we monitored the annealing reaction in the temperature range 5-50°C and plotted the reaction rates through an Arrhenius plot (Figure 4). According to the Arrhenius model, the transition state thermodynamic parameters can be derived from the reaction rates using:

$k = A\ exp\ (-E_a/RT)$ Equation (4)

where A is a preexponential factor, E_a is the activation energy, R the gas constant and T the temperature. In line with the Arrhenius model, the logarithm of the bimolecular rate constants was linearly dependent on the inverse of the temperature. Interestingly, the fits of the experimental data in the absence or the presence of the different NC mutants resulted in parallel lines (Figure 4), indicating that the activation energies (E_a) needed for the reaction to be productive (and their related ΔH^* values) were poorly affected by NCp7 and its mutants (Table 4). In sharp contrast, the annealing rates (k) were strongly dependent on the presence and nature of the peptide, giving significant variation of the Gibbs free energy ΔG^*, according to:

$$\Delta G^* = -RT \ln\left(\frac{k.h}{k_B.T}\right) \quad \text{Equation (5)}$$

where h and k_B are the Planck and Boltzmann constants, respectively. As the ΔH^* value was invariant for all NCp7 derivatives, changes in ΔG^* resulted from changes in the activation entropy of the reaction (ΔS^*) (Table 3). The highest ΔS^* values were observed with the native NCp7 and the $(SSHS)_2$NC(11-55) derivative, while a much lower value was observed for the NC(11-55) mutant. The high ΔS^* values observed with NCp7 and $(SSHS)_2$NC(11-55) are likely related to the high rates at which collisions occur between the ODNs coated by these peptides. The basic $N_{terminal}$ domain of the native NCp7 and the unfolded $(SSHS)_2$NC(11-55) peptide probably screen efficiently the negatively charged ODNs and create short range attractions between the two ODNs, even at sub-aggregating concentrations (5). This "aggregation" or crowding effect is thought to facilitate the

diffusional search for the complementary sequences (44) and increase the rate of the reaction.

Next, we performed a (-)PBS mutational analysis to determine the reaction pathways in the presence of NCp7 and its mutants. The annealing experiments were first performed with the A_5(-)PBS mutant (Figure 5A) which prevents the annealing pathway initiated through the loops (Figure 5A). With both NCp7 and NC(11-55), no accumulation of extended duplexes was observed, due to the ability of both peptides to dissociate the duplexes of (+)PBS with A_5(-)PBS that spontaneously form in the absence of peptide (8). In sharp contrast, both L_{37}NC(11-55) and $(SSHS)_2$NC(11-55) strongly promoted the annealing of (+)PBS to A_5(-)PBS, with annealing rates close to those observed with the native (+)/(-)PBS sequences, suggesting that the (-)PBS loop does not play a critical role in the reaction pathway promoted by these mutants (Table 3).

We also investigated the annealing reaction of (+)PBS with a T_4(-)PBS mutant, in which the 15-GCCA-18 sequence was substituted by a T_4 sequence to prevent the nucleation through the PBS single strand overhangs (Figure 5B). In the presence of NCp7 or NC(11-55), the rate constants with the T4(-)PBS mutant were similar to those with the native (-)PBS, suggesting that the (-)PBS overhang is not critical for the main reaction pathway promoted by NCp7 and NC(11-55). In contrast, a 15-fold and 4-fold decrease in the annealing rates was observed when the corresponding annealing reactions were monitored in the presence of $(SSHS)_2$NC(11-55) or L_{37}NC(11-55), respectively (Table 3). This clearly indicated that the (-)PBS overhang plays a central role in the $(SSHS)_2$NC(11-55)-promoted (+)/(-)PBS annealing reaction, and to a lesser extent in the L_{37}NC(11-55)-promoted (+)/(-)PBS annealing reaction.

Thus, our data strongly suggest that $(SSHS)_2$NC(11-55) or L_{37}NC(11-55) promote the (+)/(-)PBS annealing reaction, mainly through the single-strand overhangs at the bottom of the PBS stems, while, NCp7 and NC(11-55) promote this reaction mainly through the loops. Thus, the activity of NCp7 mutants on the loop dynamics correlates well with their ability to promote the (+)/(-)PBS annealing reaction through the loops. Therefore, the NCp7-induced restriction of the local and overall dynamics of the PBS loop likely constitutes a molecular prerequisite for promoting the (+)PBS/(-)PBS annealing reaction through the loop-loop kissing pathway.

Discussion

In this study, we investigated the mechanism of NCp7-directed (+)/(-)PBS annealing that occurs during the plus DNA strand transfer reaction. To this end, we used ΔP(-)PBS and ΔP(+)PBS stem-loops substituted with 2-Ap to characterize the structure and dynamic changes in the loop induced by NCp7 or NCp7 mutants, and to correlate these changes with the thermodynamic and kinetic parameters of the annealing reaction.

The substitution of natural bases by 2-Ap in the PBS loops was shown to minimally affect the folding of the loop and its binding parameters with NCp7. In the absence of protein, the 2-Ap residues in both (-)PBS and (+)PBS loops experienced multiple conformations and were efficiently quenched through collisions with their neighbour residues, in line with an orientation of the bases toward the interior of the loop and a partial order of the loop (11,32). NCp7 restricted the picosecond to nanosecond dynamics of the PBS loops, by constraining both the overall flexibility of the loops and the local mobility of the bases. Similar effects of NCp7 were previously observed with small flexible single-stranded ODNs (24,25), suggesting that they correspond to a general feature of the interaction of NCp7 with its binding sites. The observed changes in the loop dynamics of both (+)PBS and (-)PBS species and the resulting base unstacking clearly suggested that NCp7 stabilizes conformations in which the loop is stretched. Using various NCp7 mutants, the dynamic changes in the PBS loops were found to result from the specific interaction with the hydrophobic plateau at the top of the folded ZFs that also supports the destabilizing component of the chaperone activity of NCp7. Most interestingly, the NCp7-induced restriction of the dynamics in the PBS loops strongly correlated with the ability of NCp7 to switch the (-)/(+) PBS annealing reaction from a pathway through the single strand overhangs (Figure 6, upper pathway) to a loop-loop kissing pathway (Figure 6, lower pathway). Mutating the critical F_{16} and W_{37} aromatic amino acids in the two ZFs resulted in a very limited restriction of the loop mobility. Since these NC mutants only slightly promote the annealing pathway switch, one can speculate that the dynamics of the whole PBS loop should be restricted to induce the switch. As the 5-nt PBS loop can accommodate two NCp7 protein molecules (11), an appropriate restriction of the loop dynamics likely requires the binding of two native proteins.

From the temperature dependence of the annealing rate, it appears that NCp7 does not lower the activation energy *Ea* of the annealing reaction. In fact, the corresponding positive enthalpy (ΔH^*~77 kJ/mol) (Table 4) is consistent with the melting of about 3-4 base-pairs (42,45), indicating that the melting of the PBS stem probably constitutes the rate-limiting step of the annealing reaction (8). Since the *Ea* value was not affected by NCp7 or its mutants, the transition state of the reaction probably remains the same in all pathways and thus constitutes a saddle point in the energy landscape. Furthermore, the absence of NCp7-induced changes in the ΔH^* value was fully

consistent with the previously reported inability of NCp7 to melt the stable (-)PBS stem (41,46). Comparison of the annealing rates of NCp7 and NC(11-55) confirmed that the N-terminal domain of NCp7 strongly increases the annealing reaction rates and thus, plays a central role in determining the annealing reaction rates. Since this domain largely governs the aggregating properties of the protein (43,47,48), the resulting crowding effect is probably responsible for the high ΔS^* value in the presence of NCp7. The ΔS^* value and the annealing rate were also high in the presence of $(SSHS)_2$NC(11-55), which can likely through its unfolded structure better fit to the ODN structure than the folded NC(11-55) peptide and thus, more efficiently screen the repulsive forces between the phosphate charges in the ODN. A similar improved performance of a SSHS(1-55) mutant as compared to the native NCp7 was previously reported for the annealing of $tRNA^{Lys3}$ to the PBS sequence (42). The annealing rate promotion appears thus poorly related to the destabilizing activity of NCp7.

In sharp contrast, the ability of the mutants to induce a mechanistic switch strongly depends on the integrity of the hydrophobic platform on the ZF domain. Mutations preventing the formation of this platform and thus, abolishing the nucleic acid destabilizing activity of NCp7 (38,43,49) were unable to induce this mechanistic switch, though they can still accelerate the annealing reaction. In this context, the chaperoning activity of NCp7 can be appropriately envisioned only if the annealing pathway is considered. The lack of consideration of the annealing pathway probably explains previous *in vitro* observations where the ZFs of NCp7 were reported to be dispensable in promoting annealing reactions (50-52). In sharp contrast to these *in vitro* experiments, even single mutations within the ZFs resulted in the production of non infectious particles *in vivo*, clearly showing that the ZFs play a critical role (53-56). Mutations in the NCp7 ZFs resulted notably in reverse transcription defects, decreases in plus strand transfer efficiency and integration (56-58), as well as in profound modifications of the spatio-temporal control of the reverse transcription process (59-63). Indeed, these ZF mutations notably allowed reverse transcription in the cytoplasm of virus producer cells, thus turning virions into DNA-containing viral particles that contain defective viral DNA sequences characterized by insertions, deletions and improper ends (57-61,64). In line with our observations where ZF mutations still allow efficient (-)/(+)PBS annealing, the loss of the NCp7 destabilizing activity induced by these mutations does not block *in vivo* the synthesis of viral DNA (vDNA) which can go to completion with a reasonable efficiency (62,63,65). Nevertheless, the resulting disturbances in reverse transcription timing and the probable defects in the vDNA sequence render the virus non-infectious (56,57,64) (reviewed in (61)). The question on how the NCp7 nucleic acid chaperoning activity exerts a control on the timing of vDNA synthesis can find explanations in the fact that non-properly folded ZFs do not modify the dynamics and structure of nucleic acids. As a consequence, mutated NC proteins can only stimulate non-specifically the rearrangements of

nucleic acids through already available existing pathways in the absence of peptides (i.e. mostly through the PBS overhangs here). Such a non-specific promotion also appears in various physiological microenvironments, such as the seminal fluid, that promote the natural endogenous reverse transcription (NERT) (66,67). In contrast, the binding of wild-type NCp7 on both (-)PBS and (+)PBS loops reveals hidden nucleation sites which fuels nucleic acid rearrangements towards specific routes. Both the switch in the annealing mechanism and the inhibition of the annealing of imperfect complementary sequences (see mutant A_5(-)PBS in Table 3) constitute two clear examples of how NCp7 directs the nucleic acid rearrangements towards such specific routes. It also highlights the mandatory role of the NCp7 ZFs that freeze competent (+)/(-)PBS conformations for specific annealing through their loops during the second strand transfer.

NCp7 with properly folded ZFs was also reported to improve the fidelity of reverse transcription by strongly inhibiting non-PPT priming, ensuring the selection of the PPT sequence as the sole primer for initiation of plus-strand DNA synthesis (68,69). In this respect, the ability of NCp7 to promote a specific annealing pathway during the second strand transfer constitutes a supplementary assertion in the understanding of how NCp7 ensures spatio-temporal regulation of the reverse transcription as well as faithful and efficient formation of stable vDNA product. At the molecular level, both the plus priming control and the annealing mechanism selection require the hydrophobic platform at the top of the folded ZFs. Since this hydrophobic plateau is also the protein determinant associated with the destabilizing activity of NCp7, it could be of interest to determine whether NCp7 can also 'freeze' specific reactive states within ODNs known for being strongly destabilized by NCp7 such as cTAR DNA (19,70-74) for which a NCp7-induced mechanistic switch in the annealing with TAR was also evidenced (44,75). This work is under progress.

Fundings:

This work was supported by Agence Nationale de Recherche sur le Sida. Funding to pay the Open Access publication charges for this article was provided by Agence Nationale de Recherches sur le SIDA.

Acknowledgments:

We acknowledge Etienne Piémont for his help in performing time-resolved experiments.

Conflict of interest statement. None declared.

References

1. Basu, V.P., Song, M., Gao, L., Rigby, S.T., Hanson, M.N. and Bambara, R.A. (2008) Strand transfer events during HIV-1 reverse transcription. *Virus Res*, 134, 19-38.
2. Darlix, J.L., Lapadat-Tapolsky, M., de Rocquigny, H. and Roques, B.P. (1995) First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. *J Mol Biol*, 254, 523-537.
3. Cristofari, G. and Darlix, J.L. (2002) The ubiquitous nature of RNA chaperone proteins. *Prog Nucleic Acid Re Mol Biol*, 72, 223-268.
4. Darlix, J.L., Cristofari, G., Rau, M., Péchoux, C., Berthoux, L. and Roques, B. (2000) Nucleocapsid protein o human immunodeficiency virus as a model protein with chaperoning functions and as a target for antiviral drugs. *Adv Pharmacol*, 48, 345-372.
5. Levin, J.G., Guo, J., Rouzina, I. and Musier-Forsyth, K. (2005) Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Prog Nucleic Acid Res Mol Biol*, 80, 217-286.
6. Rein, A., Henderson, L.E. and Levin, J.G. (1998) Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem Sci*, 23, 297-301.
7. Mély, Y., Rocquigny, H.D., Morellet, N., Roques, B.P. and Gérad, D. (1996) Zinc binding to the HIV-1 nucleocapsid protein: a thermodynamic investigation by fluorescence spectroscopy. *Biochemistry*, 35, 5175-5182.
8. Ramalanjaona, N., de Rocquigny, H., Millet, A., Ficheux, D., Darlix, J.-L. and Mély, Y. (2007) Investigating the mechanism of the nucleocapsid protein chaperoning of the second strand transfer during HIV-1 DNA synthesis. *J Mol Biol*, 374, 1041-1053.
9. Wu, T., Guo, J., Bess, J., Henderson, L.E. and Levin, J.G. (1999) Molecular requirements for human immunodeficiency virus type 1 plus-strand transfer: analysis in reconstituted and endogenous reverse transcription systems. *Journal of virology*, 73, 4794-4805.
10. Herschlag, D. (1995) RNA chaperones and the RNA folding problem. *J Biol Chem*, 270, 20871-20874.
11. Bourbigot, S., Ramalanjaona, N., Boudier, C., Salgado, G.F.J., Roques, B.P., Mély, Y., Bouaziz, S. and Morellet, N. (2008) How the HIV-1 nucleocapsid protein binds and destabilises the (-)primer binding site during reverse transcription. *J Mol Biol*, 383, 1112-1128.
12. Lee, B.M., Guzman, R.N.D., Turner, B.G., Tjandra, N. and Summers, M.F. (1998) Dynamical behavior of the HIV-1 nucleocapsid protein. *J Mol Biol*, 279, 633-649.
13. Mély, Y., Jullian, N., Morellet, N., Rocquigny, H.D., Dong, C.Z., Piémont, E., Roques, B.P. and Gérard, D. (1994) Spatial proximity of the HIV-1 nucleocapsid protein zinc fingers investigated by time-resolved fluorescence and fluorescence resonance energy transfer. *Biochemistry*, 33, 12085-12091.
14. Morellet, N., de Rocquigny, H., Mély, Y., Jullian, N., Déméné, H., Ottmann, M., Gérard, D., Darlix, J.L., Fournie-Zaluski, M.C. and Roques, B.P. (1994) Conformational behaviour of the active and inactive forms of the nucleocapsid NCp7 of HIV-1 studied by 1H NMR. *J Mol Biol*, 235, 287-301.
15. Mori, M., Dietrich, U., Manetti, F. and Botta, M. (2010) Molecular Dynamics and DFT Study on HIV-1 Nucleocapsid Protein-7 in Complex with Viral Genome. *J Chem Inf Model*. 50, 638-650.
16. Ward, D.C., Reich, E. and Stryer, L. (1969) Fluorescence studies of nucleotides and polynucleotides. I. Formycin, 2-aminopurine riboside, 2,6-diaminopurine riboside, and their derivatives. *J Biol Chem*, 244, 1228 1237.
17. de Rocquigny, H., Ficheux, D., Gabus, C., Fournie-Zaluski, M.C., Darlix, J.L. and Roques, B.P. (1991) First large scale chemical synthesis of the 72 amino acid HIV-1 nucleocapsid protein NCp7 in an active form. *Biochem Biophys Res Commun*, 180, 1010-1018.
18. Shvadchak, V.V., Klymchenko, A.S., de Rocquigny, H. and Mély, Y. (2009) Sensing peptide-oligonucleotide interactions by a two-color fluorescence label: application to the HIV-1 nucleocapsid protein. *Nucleic Acids Res*, 37, e25.
19. Bernacchi, S., Stoylov, S., Piémont, E., Ficheux, D., Roques, B.P., Darlix, J.L. and Mély, Y. (2002) HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. *J Mol Biol*, 317, 385-399.
20. Brochon, J.C. (1994) Maximum entropy method of data analysis in time-resolved spectroscopy. *Methods Enzymol*, 240, 262-311.
21. Livesey, A.K. and Brochon, J.C. (1987) Analyzing the Distribution of Decay Constants in Pulse-Fluorimetry Using the Maximum Entropy Method. *Biophys J*, 52, 693-706.
22. Jean, J.M. and Hall, K.B. (2001) 2-Aminopurine fluorescence quenching and lifetimes: role of base stacking. *Proc Natl Acad Sci U S A*, 98, 37-41.
23. Guest, C.R., Hochstrasser, R.A., Sowers, L.C. and Millar, D.P. (1991) Dynamics of mismatched base pairs in DNA. *Biochemistry*, 30, 3271-3279.
24. Avilov, S., Godet, J., Piemont, E. and Mély, Y. (2009) Site-specific characterization of HIV-1 nucleocapsid protein binding to oligonucleotides with two binding sites. *Biochemistry*, 48, 2422-2430.
25. Avilov, S.V., Piemont, E., Shvadchak, V., de Rocquigny, H. and Mély, Y. (2008) Probing dynamics of HIV-1

nucleocapsid protein/target hexanucleotide complexes by 2-aminopurine. *Nucleic Acids Res*, 36, 885-896.
26. Jean, J.M. and Hall, K.B. (2002) 2-Aminopurine electronic structure and fluorescence properties in DNA. *Biochemistry*, 41, 13152-13161.
27. Larsen, O.F., van Stokkum, I.H., Gobets, B., van Grondelle, R. and van Amerongen, H. (2001) Probing the structure and dynamics of a DNA hairpin by ultrafast quenching and fluorescence depolarization. *Biophys J*, 81, 1115-1126.
28. Wan, C., Fiebig, T., Schiemann, O., Barton, J.K. and Zewail, A.H. (2000) Femtosecond direct observation of charge transfer between bases in DNA. *Proc Natl Acad Sci U S A*, 97, 14052-14055.
29. Fiebig, T., Wan, C. and Zewail, A.H. (2002) Femtosecond charge transfer dynamics of a modified DNA base: 2-aminopurine in complexes with nucleotides. *Chemphyschem*, 3, 781-788.
30. O'Neill, M.A. and Barton, J.K. (2002) 2-Aminopurine: a probe of structural dynamics and charge transfer in DNA and DNA:RNA hybrids. *J Am Chem Soc*, 124, 13053-13066.
31. Jean, J.M. and Hall, K.B. (2004) Stacking-unstacking dynamics of oligodeoxynucleotide trimers. *Biochemistry*, 43, 10277-10284.
32. Johnson, P.E., Turner, R.B., Wu, Z.R., Hairston, L., Guo, J., Levin, J.G. and Summers, M.F. (2000) A mechanism for plus-strand transfer enhancement by the HIV-1 nucleocapsid protein during reverse transcription. *Biochemistry*, 39, 9084-9091.
33. Lakowicz, J.R. (1999) *Principles of Fluorescence Spectroscopy*. 2nd ed. Kluwer Academic/Plenum Publishers.
34. Mely, Y., de Rocquigny, H., Piémont, E., Déméné, H., Jullian, N., Fournié-Zaluski, M.C., Roques, B. and Gérard, D. (1993) Influence of the N- and C-terminal chains on the zinc-binding and conformational properties of the central zinc-finger structure of Moloney murine leukaemia virus nucleocapsid protein: a steady-state and time-resolved fluorescence study. *Biochim Biophys Acta*, 1161, 6-18.
35. Vuilleumier, C., Bombarda, E., Morellet, N., Gérard, D., Roques, B.P. and Mély, Y. (1999) Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: a fluorescence study. *Biochemistry*, 38, 16816-16825.
36. Green, L.M. and Berg, J.M. (1990) Retroviral nucleocapsid protein-metal ion interactions: folding and sequence variants. *Proc Natl Acad Sci U S A*, 87, 6403-6407.
37. Guo, J., Wu, T., Anderson, J., Kane, B.F., Johnson, D.G., Gorelick, R.J., Henderson, L.E. and Levin, J.G. (2000) Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer. *Journal of virology*, 74, 8980-8988.
38. Beltz, H., Clauss, C., Piémont, E., Ficheux, D., Gorelick, R.J., Roques, B., Gabus, C., Darlix, J.-L., de Rocquigny, H. and Mély, Y. (2005) Structural determinants of HIV-1 nucleocapsid protein for cTAR DNA binding and destabilization, and correlation with inhibition of self-primed DNA synthesis. *J Mol Biol*, 348, 1113-1126.
39. Fisher, R.J., Fivash, M.J., Stephen, A.G., Hagan, N.A., Shenoy, S.R., Medaglia, M.V., Smith, L.R., Worthy, K.M., Simpson, J.T., Shoemaker, R. *et al.* (2006) Complex interactions of HIV-1 nucleocapsid protein with oligonucleotides. *Nucleic Acids Res*, 34, 472-484.
40. Gorelick, R.J., Gagliardi, T.D., Bosche, W.J., Wiltrout, T.A., Coren, L.V., Chabot, D.J., Lifson, J.D., Henderson, L.E. and Arthur, L.O. (1999) Strict conservation of the retroviral nucleocapsid protein zinc finger is strongly influenced by its role in viral infection processes: characterization of HIV-1 particles containing mutant nucleocapsid zinc-coordinating sequences. *Virology*, 256, 92-104.
41. Egelé, C., Schaub, E., Ramalanjaona, N., Piémont, E., Ficheux, D., Roques, B., Darlix, J.-L. and Mély, Y. (2004) HIV-1 nucleocapsid protein binds to the viral DNA initiation sequences and chaperones their kissing interactions. *J Mol Biol*, 342, 453-466.
42. Hargittai, M.R.S., Gorelick, R.J., Rouzina, I. and Musier-Forsyth, K. (2004) Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site. *J Mol Biol*, 337, 951-968.
43. Williams, M.C., Rouzina, I., Wenner, J.R., Gorelick, R.J., Musier-Forsyth, K. and Bloomfield, V.A. (2001) Mechanism for nucleic acid chaperone activity of HIV-1 nucleocapsid protein revealed by single molecule stretching. *Proc Natl Acad Sci U S A*, 98, 6121-6126.
44. Vo, M.-N., Barany, G., Rouzina, I. and Musier-Forsyth, K. (2009) HIV-1 nucleocapsid protein switches the pathway of transactivation response element RNA/DNA annealing from loop-loop "kissing" to "zipper". *J Mol Biol*, 386, 789-801.
45. Rouzina, I. and Bloomfield, V.A. (1999) Heat capacity effects on the melting of DNA. 1. General aspects. *Biophys J*, 77, 3242-3251.
46. Egelé, C., Schaub, E., Piémont, E., de Rocquigny, H. and Mély, Y. (2005) Investigation by fluorescence correlation spectroscopy of the chaperoning interactions of HIV-1 nucleocapsid protein with the viral DNA initiation sequences. *C R Biol*, 328, 1041-1051.
47. Le Cam, E., Coulaud, D., Delain, E., Petitjean, P., Roques, B.P., Gerard, D., Stoylova, E., Vuilleumier, C., Stoylov, S.P. and Mely, Y. (1998) Properties and growth mechanism of the ordered aggregation of a model RNA by the HIV-1 nucleocapsid protein: an electron microscopy investigation. *Biopolymers*, 45, 217-229.
48. Stoylov, S.P., Vuilleumier, C., Stoylova, E., De Rocquigny, H., Roques, B.P., Gerard, D. and Mely, Y. (1997) Ordered aggregation of ribonucleic acids by the human immunodeficiency virus type 1 nucleocapsid protein.

Biopolymers, 41, 301-312.
49. Williams, M.C., Gorelick, R.J. and Musier-Forsyth, K. (2002) Specific zinc-finger architecture required for HIV-1 nucleocapsid protein's nucleic acid chaperone function. *Proc Natl Acad Sci U S A*, 99, 8614-8619.
50. De Rocquigny, H., Gabus, C., Vincent, A., Fournie-Zaluski, M.C., Roques, B. and Darlix, J.L. (1992) Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers. *Proc Natl Acad Sci U S A*, 89, 6472-6476.
51. Lapadat-Tapolsky, M., Gabus, C., Rau, M. and Darlix, J.L. (1997) Possible roles of HIV-1 nucleocapsid protein in the specificity of proviral DNA synthesis and in its variability. *J Mol Biol*, 268, 250-260.
52. Prats, A.C., Housset, V., de Billy, G., Cornille, F., Prats, H., Roques, B. and Darlix, J.L. (1991) Viral RNA annealing activities of the nucleocapsid protein of Moloney murine leukemia virus are zinc independent. *Nucleic Acids Res*, 19, 3533-3541.
53. Dorfman, T., Luban, J., Goff, S.P., Haseltine, W.A. and Gottlinger, H.G. (1993) Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein. *Journal of virology*, 67, 6159-6169.
54. Gorelick, R.J., Benveniste, R.E., Gagliardi, T.D., Wiltrout, T.A., Busch, L.K., Bosche, W.J., Coren, L.V., Lifson, J.D., Bradley, P.J., Henderson, L.E. *et al.* (1999) Nucleocapsid protein zinc-finger mutants of simian immunodeficiency virus strain mne produce virions that are replication defective in vitro and in vivo. *Virology* 253, 259-270.
55. Gorelick, R.J., Chabot, D.J., Rein, A., Henderson, L.E. and Arthur, L.O. (1993) The two zinc fingers in the human immunodeficiency virus type 1 nucleocapsid protein are not functionally equivalent. *Journal of virology*, 67, 4027-4036.
56. Tanchou, V., Decimo, D., Péchoux, C., Lener, D., Rogemond, V., Berthoux, L., Ottmann, M. and Darlix, J.L. (1998) Role of the N-terminal zinc finger of human immunodeficiency virus type 1 nucleocapsid protein in virus structure and replication. *Journal of virology*, 72, 4442-4447.
57. Buckman, J.S., Bosche, W.J. and Gorelick, R.J. (2003) Human immunodeficiency virus type 1 nucleocapsid zn(2+) fingers are required for efficient reverse transcription, initial integration processes, and protection of newly synthesized viral DNA. *Journal of virology*, 77, 1469-1480.
58. Thomas, J.A. and Gorelick, R.J. (2008) Nucleocapsid protein function in early infection processes. *Virus Res*, 134, 39-63.
59. Didierlaurent, L., Houzet, L., Morichaud, Z., Darlix, J.-L. and Mougel, M. (2008) The conserved N-terminal basic residues and zinc-finger motifs of HIV-1 nucleocapsid restrict the viral cDNA synthesis during virus formation and maturation. *Nucleic Acids Res*, 36, 4745-4753.
60. Houzet, L., Morichaud, Z., Didierlaurent, L., Muriaux, D., Darlix, J.-L. and Mougel, M. (2008) Nucleocapsid mutations turn HIV-1 into a DNA-containing virus. *Nucleic Acids Res*, 36, 2311-2319.
61. Mougel, M., Houzet, L. and Darlix, J.-L. (2009) When is it time for reverse transcription to start and go? *Retrovirology*, 6, 24.
62. Thomas, J.A., Gagliardi, T.D., Alvord, W.G., Lubomirski, M., Bosche, W.J. and Gorelick, R.J. (2006) Human immunodeficiency virus type 1 nucleocapsid zinc-finger mutations cause defects in reverse transcription and integration. *Virology*, 353, 41-51.
63. Thomas, J.A., Shulenin, S., Coren, L.V., Bosche, W.J., Gagliardi, T.D., Gorelick, R.J. and Oroszlan, S. (2006) Characterization of human immunodeficiency virus type 1 (HIV-1) containing mutations in the nucleocapsid protein at a putative HIV-1 protease cleavage site. *Virology*, 354, 261-270.
64. Thomas, J.A., Bosche, W.J., Shatzer, T.L., Johnson, D.G. and Gorelick, R.J. (2008) Mutations in human immunodeficiency virus type 1 nucleocapsid protein zinc fingers cause premature reverse transcription. *Journal of virology*, 82, 9318-9328.
65. Coren, L.V., Thomas, J.A., Chertova, E., Sowder, R.C., Gagliardi, T.D., Gorelick, R.J. and Ott, D.E. (2007) Mutational analysis of the C-terminal gag cleavage sites in human immunodeficiency virus type 1. *Journal of virology*, 81, 10047-10054.
66. Zhang, H., Dornadula, G. and Pomerantz, R.J. (1996) Endogenous reverse transcription of human immunodeficiency virus type 1 in physiological microenviroments: an important stage for viral infection of nondividing cells. *Journal of virology*, 70, 2809-2824.
67. Zhang, H., Dornadula, G. and Pomerantz, R.J. (1998) Natural endogenous reverse transcription of HIV-1. *Journal of reproductive immunology*, 41, 255-260.
68. Jacob, D.T. and DeStefano, J.J. (2008) A new role for HIV nucleocapsid protein in modulating the specificity of plus strand priming. *Virology*, 378, 385-396.
69. Post, K., Kankia, B., Gopalakrishnan, S., Yang, V., Cramer, E., Saladores, P., Gorelick, R.J., Guo, J., Musier-Forsyth, K. and Levin, J.G. (2009) Fidelity of plus-strand priming requires the nucleic acid chaperone activity of HIV-1 nucleocapsid protein. *Nucleic Acids Res.*, 37, 1755-1766.
70. Azoulay, J., Clamme, J.P., Darlix, J.L., Roques, B.P. and Mély, Y. (2003) Destabilization of the HIV-1 complementary sequence of TAR by the nucleocapsid protein through activation of conformational fluctuations. *J Mol Biol*, 326, 691-700.
71. Beltz, H., Azoulay, J., Bernacchi, S., Clamme, J.-P., Ficheux, D., Roques, B., Darlix, J.-L. and Mély, Y. (2003) Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the

nucleocapsid protein NCp7. *J Mol Biol*, 328, 95-108.
72. Beltz, H., Piémont, E., Schaub, E., Ficheux, D., Roques, B., Darlix, J.-L. and Mély, Y. (2004) Role of the structure of the top half of HIV-1 cTAR DNA on the nucleic acid destabilizing activity of the nucleocapsid protein NCp7. *J Mol Biol*, 338, 711-723.
73. Cosa, G., Harbron, E.J., Zeng, Y., Liu, H.-W., O'Connor, D.B., Eta-Hosokawa, C., Musier-Forsyth, K. and Barbara, P.F. (2004) Secondary structure and secondary structure dynamics of DNA hairpins complexed with HIV-1 NC protein. *Biophys J*, 87, 2759-2767.
74. Cosa, G., Zeng, Y., Liu, H.-W., Landes, C.F., Makarov, D.E., Musier-Forsyth, K. and Barbara, P.F. (2006) Evidence for non-two-state kinetics in the nucleocapsid protein chaperoned opening of DNA hairpins. *J Phys Chem B*, 110, 2419-2426.
75. Godet, J., de Rocquigny, H., Raja, C., Glasser, N., Ficheux, D., Darlix, J.-L. and Mély, Y. (2006) During the early phase of HIV-1 DNA synthesis, nucleocapsid protein directs hybridization of the TAR complementary sequences via the ends of their double-stranded stem. *J Mol Biol*, 356, 1180-1192.

Table 1. Steady-state and time-resolved fluorescence parameters of 2-Ap-substituted ΔP(-)PBS and ΔP(+)PBS

	Quantum yield	α_0	τ_1	α_1	τ_2	α_2	τ_3	α_3	τ_4	α_4	$\langle\tau\rangle$
Free 2Ap riboside	0.680[a]	-	-	-	-	-	-	-	10.2	1.00	10.2
ΔP (-)PBS 2-Ap6	0.024	0.80	0.10	0.13	0.7	0.02	3.6	0.02	8.0	0.03	1.8
ΔP (-)PBS 2-Ap6 + NCp7	0.044	0.72	0.12	0.10	0.6	0.07	2.8	0.06	8.3	0.05	2.4
ΔP (-)PBS 2-Ap6 + NC(11-55)	0.040	0.75	0.12	0.10	0.6	0.06	3.4	0.04	7.5	0.05	2.3
ΔP (-)PBS 2-Ap6 + $SSHS_2$NC(11-55)	0.024	0.80	0.09	0.08	0.6	0.04	2.9	0.06	7.5	0.02	2.0
ΔP (-)PBS 2-Ap6 + A_{16}NC(11-55)	0.026	0.78	0.11	0.11	0.7	0.04	3.1	0.04	8.3	0.03	1.8
ΔP (-)PBS 2-Ap6 + L_{37}NC(11-55)	0.035	0.62	0.10	0.20	0.7	0.07	3.3	0.06	8.7	0.05	1.9
ΔP (-)PBS 2-Ap8	0.028	0.82	0.09	0.08	0.7	0.03	3.1	0.03	8.3	0.04	2.1
ΔP (-)PBS 2-Ap8 + NCp7	0.087	0.60	0.09	0.13	0.6	0.08	3.8	0.08	8.9	0.11	3.3
ΔP (-)PBS 2-Ap8 + NC(11-55)	0.081	0.60	0.14	0.13	0.6	0.10	3.8	0.07	8.5	0.10	3.0
ΔP (-)PBS 2-Ap8 + $SSHS_2$NC(11-55)	0.032	0.80	0.10	0.09	0.7	0.04	3.2	0.04	8.2	0.03	2.1
ΔP (-)PBS 2-Ap8 + A_{16}NC(11-55)	0.033	0.70	0.07	0.17	0.7	0.05	3.2	0.04	8.2	0.04	2.0
ΔP (-)PBS 2-Ap8 + L_{37}NC(11-55)	0.035	0.79	0.10	0.09	0.9	0.04	3.6	0.04	8.9	0.04	2.5
ΔP (-)PBS 2-Ap10	0.006	0.94	0.13	0.03	0.8	0.01	3.5	0.01	8.2	0.01	2.1
ΔP (-)PBS 2-Ap10 + NCp7	0.028	0.83	0.13	0.08	0.7	0.03	4.3	0.04	8.8	0.02	2.7
ΔP (-)PBS 2-Ap10 + NC(11-55)	0.021	0.86	0.15	0.07	0.7	0.03	3.9	0.02	8.6	0.02	2.3
ΔP (-)PBS 2-Ap10 + $SSHS_2$NC(11-55)	0.007	0.92	0.10	0.03	0.8	0.02	3.9	0.02	8.3	0.01	2.0
ΔP (-)PBS 2-Ap10 + A_{16}NC(11-55)	0.008	0.95	0.07	0.02	0.6	0.01	2.9	0.01	7.5	0.01	2.1
ΔP (-)PBS 2-Ap10 + L_{37}NC(11-55)	0.009	0.95	0.09	0.02	0.8	0.01	3.1	0.01	8.2	0.01	2.6
ΔP (+)PBS 2-Ap11	0.036	0.50	0.09	0.34	0.8	0.09	3.6	0.03	8.2	0.04	1.1
ΔP (+)PBS 2-Ap11 + NCp7	0.064	0.51	0.10	0.21	0.9	0.11	3.4	0.06	8.9	0.11	2.6
ΔP (+)PBS 2-Ap11 + NC(11-55)	0.061	0.51	0.10	0.22	0.7	0.12	3.6	0.07	8.8	0.08	2.1
ΔP (+)PBS 2-Ap11 + $SSHS_2$NC(11-55)	0.036	0.52	0.10	0.29	0.9	0.11	3.5	0.05	8.4	0.03	1.1

[a] data from Ward et al. (16)

τ_i (ns) are the fluorescence lifetimes, α_i their amplitudes. The amplitude values are corrected for the dark species, as described in the Materials and Methods section. $\langle\tau\rangle$ is the mean fluorescence lifetime.

Table 2. Fluorescence anisotropy decay parameters of 2-Ap-substituted ΔP(-)PBS and ΔP(+)PBS

	Φ_1	β_1	Φ_2	β_2	Φ_3	β_3	S	θ_0
Free 2Ap riboside	0.08	1.00	-	-	-	-	0.00	90
ΔP (-)PBS 2-Ap6	0.17	0.38	0.7	0.34	2.2	0.28	0.79	32
ΔP (-)PBS 2-Ap6 + NCp7	0.20	0.15	1.5	0.22	9.7	0.63	0.88	19
ΔP (-)PBS 2-Ap6 + NC(11-55)	0.17	0.18	1.2	0.28	7.4	0.54	0.91	21
ΔP (-)PBS 2-Ap6 + SSHS$_2$NC(11-55)	0.18	0.32	1.2	0.21	9.5	0.47	0.82	29
ΔP (-)PBS 2-Ap6 + A^{16}NC(11-55)	0.08	0.28	1.2	0.31	9.4	0.41	0.85	27
ΔP (-)PBS 2-Ap6 + L^{37}NC(11-55)	0.14	0.21	1.2	0.33	9.0	0.46	0.89	23
ΔP (-)PBS 2-Ap8	0.22	0.36	0.7	0.11	2.1	0.53	0.79	31
ΔP (-)PBS 2-Ap8 + NCp7	0.13	0.13	1.3	0.25	10.2	0.62	0.91	17
ΔP (-)PBS 2-Ap8 + NC(11-55)	0.13	0.13	0.9	0.32	7.7	0.55	0.93	17
ΔP (-)PBS 2-Ap8 + SSHS$_2$NC(11-55)	0.18	0.30	1.1	0.27	9.3	0.43	0.84	28
ΔP (-)PBS 2-Ap8 + A^{16}NC(11-55)	0.10	0.18	0.6	0.24	7.8	0.58	0.87	21
ΔP (-)PBS 2-Ap8 + L^{37}NC(11-55)	0.20	0.26	1.2	0.24	9.1	0.50	0.83	25
ΔP (-)PBS 2-Ap10	0.18	0.46	0.9	0.27	2.6	0.27	0.73	36
ΔP (-)PBS 2-Ap10 + NCp7	0.22	0.16	0.8	0.24	9.2	0.60	0.91	19
ΔP (-)PBS 2-Ap10 + NC(11-55)	0.20	0.21	0.9	0.20	7.5	0.59	0.89	23
ΔP (-)PBS 2-Ap10 + SSHS$_2$NC(11-55)	0.16	0.38	0.9	0.20	9.1	0.42	0.79	32
ΔP (-)PBS 2-Ap10 + A^{16}NC(11-55)	0.10	0.36	0.8	0.24	7.3	0.40	0.80	31
ΔP (-)PBS 2-Ap10 + L^{37}NC(11-55)	0.15	0.32	1.2	0.33	9.1	0.35	0.82	29
ΔP (+)PBS 2-Ap11	0.12	0.32	0.6	0.21	2.2	0.47	0.82	29
ΔP (+)PBS 2-Ap11 + NCp7	0.13	0.22	0.8	0.19	8.2	0.59	0.88	23
ΔP (+)PBS 2-Ap11 + NC(11-55)	0.10	0.25	0.9	0.20	7.8	0.55	0.89	25
ΔP (+)PBS 2-Ap11 + SSHS$_2$NC(11-55)	0.15	0.30	0.9	0.18	8.3	0.52	0.89	28

Φ_i (ns) are the rotational correlation times, β_i their amplitudes, S is the generalized order parameter and θ_0 the cone semi-angle (in °) for 2-Ap local motion (calculated as described in the supplementary data)

Table 3. Kinetic parameters of the annealing of (+)PBS to (-)PBS mutants in the absence and in the presence of NC derivatives

Labeled sequence	Complementary sequence	k (M^{-1} s^{-1})				
		-	NCp7	NC(11-55)	L_{37}NC(11-55)	$(SSHS)_2$NC(11-55)
(+)PBS	(-)PBS	3,800 [a] ± 100	180,000 [a] ± 10,000	18,400 [a] ±1,200	22,600 ±2,000	99,000 ±6,500
	A5(-)PBS	3,200 [a] ± 100	- [b]	- [b]	18,000 ±3,000	64,000 ±4,000
	T_4(-) PBS	~20 [a]	210,000 ± 20,000	11,100 [a] ±200	6,000 ±1,000	7,000 ±500

[a] Data from Ramalanjaona & al. (8)

[b] No rate constant could be defined since NCp7 promotes the dissociation of the A5(-)/(+)PBS duplex

Table 4. Thermodynamic parameters of the NC-promoted (+)/(-)PBS annealing

	k^a (M^{-1} s^{-1})	Ea^b (kJ/mol)	ΔG^{*c} (kJ/mol)	ΔH^{*d} (kJ/mol)	ΔS^{*e} (*eu*)
No peptide	$4.3 \pm 1.0 \times 10^3$	80 ± 8	57	78 ± 5	17 ± 3
NCp7	$3.2 \pm 1.0 \times 10^5$	78 ± 10	45	76 ± 8	26 ± 3
NC(11-55)	$3.1 \pm 0.6 \times 10^4$	78 ± 12	52	76 ± 9	19 ± 2
A_{16}NC(11-55)	$9.4 \pm 0.7 \times 10^3$	80 ± 8	55	78 ± 6	17 ± 2
L_{37} NC(11-55)	$3.2 \pm 0.8 \times 10^4$	77 ± 11	48	74 ± 8	18 ± 2
$SSHS_2$ NC(11-55)	$1.8 \pm 0.6 \times 10^5$	81 ± 9	48	79 ± 7	25 ± 3

[a] Determined at 25°C
[b] Determined from the Arrhenius plot (Figure 4)
[c] Calculated from the annealing rate constant at 25°C according to equation (7)
[d] ΔH^* is given by $\Delta H^* = Ea - RT$ with T=298.15 K.
[e] Calculated according to $\Delta G^* = \Delta H^* - T\Delta S^*$. *eu* corresponds to (cal mol^{-1} K^{-1})

Figures legends:

Figure 1:
Sequences of the oligonucleotides and peptides used in this study.

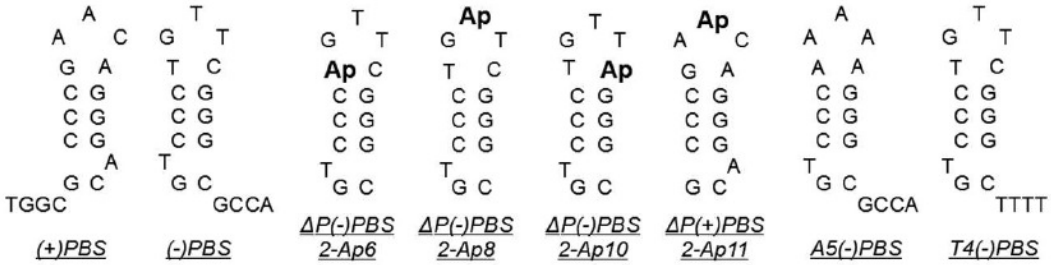
Figure 2:
Mechanism of (−)/(+)PBS annealing, as proposed by Ramalanjaona & al. (8) The (-)(+)PBS annealing reaction can proceed through two major pathways. In these pathways, the final extended duplex can be nucleated either through the single-stranded overhangs (upper pathway) or through loop-loop interactions (lower pathway). In the absence of NCp7, the upper pathway is by far the most contributing one. NCp7 strongly activates the loop-loop kissing pathway, leading to a switch in the reaction mechanism.

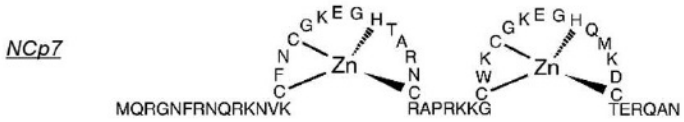
Figure 3:
Experimental anisotropy decay curve of ΔP(-)PBS 2-AP6 in the absence (black) and in the presence of NCp7 (green). The continuous lines (red and blue) correspond to the fit of the data with the parameters in table 2.

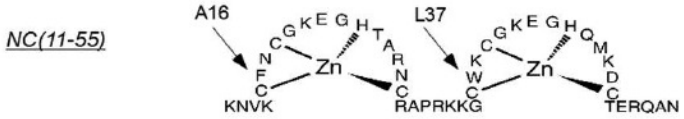
Figure 4:
Arrhenius plot of the annealing of (-)PBS to (+)PBS in the presence of NCp7 mutants. The experiments were performed either in the absence (open squares) or in the presence of NCp7 (open squares) or its mutants, NC(11-55) (closed circles), $SSHS_2$NC(11-55) (closed triangles), A_{16}NC(11-55) (open circles), and W_{37}NC(11-55) (open triangles).

Figure 5:
(-)PBS mutational analysis of the annealing reaction. (A) The A_5(-)PBS derivative (mutated residues are in blue) was designed to prevent loop-loop interactions. No A_5(-)PBS/(+)PBS duplex accumulated in the presence of NCp7. In contrast, NC ZF mutants strongly promoted the A_5(-)PBS/(+)PBS annealing reaction. (B) The T_4(-)PBS derivative (mutated residues are in red) was designed to prevent nucleation of the two complementary PBS sequences through their ss overhangs. Annealing of T_4(-)PBS with (+)PBS occurred at low speed in the absence of NCp7 or in the presence of NCp7 mutants with altered chaperone properties. In contrast, in the presence of NCp7, the T_4(-)PBS/(+)PBS annealing reaction occurred at a rate similar to that observed with the native DNA sequences.

Figure 6:
Proposed mechanism for (−)/(+)PBS annealing. In the absence of NCp7 (upper pathway), the bases of (-)PBS and (+)PBS loops are oriented towards the interior of the loop and not available for loop-loop interaction. As a consequence, the annealing is nucleated through the bases of the flexible single-stranded overhangs and possibly, the exposed bulged nt at the bottom of the stem (red bases). In contrast, through preferential binding to the PBS loops, NCp7 stretches the loop and exposes the loop bases (blue bases, lower pathway). As a result, NCp7 “freezes” PBS conformations competent for annealing via the loops, leading to a strong activation in the loop-loop kissing pathway, and thus, to a switch in the reaction mechanism.





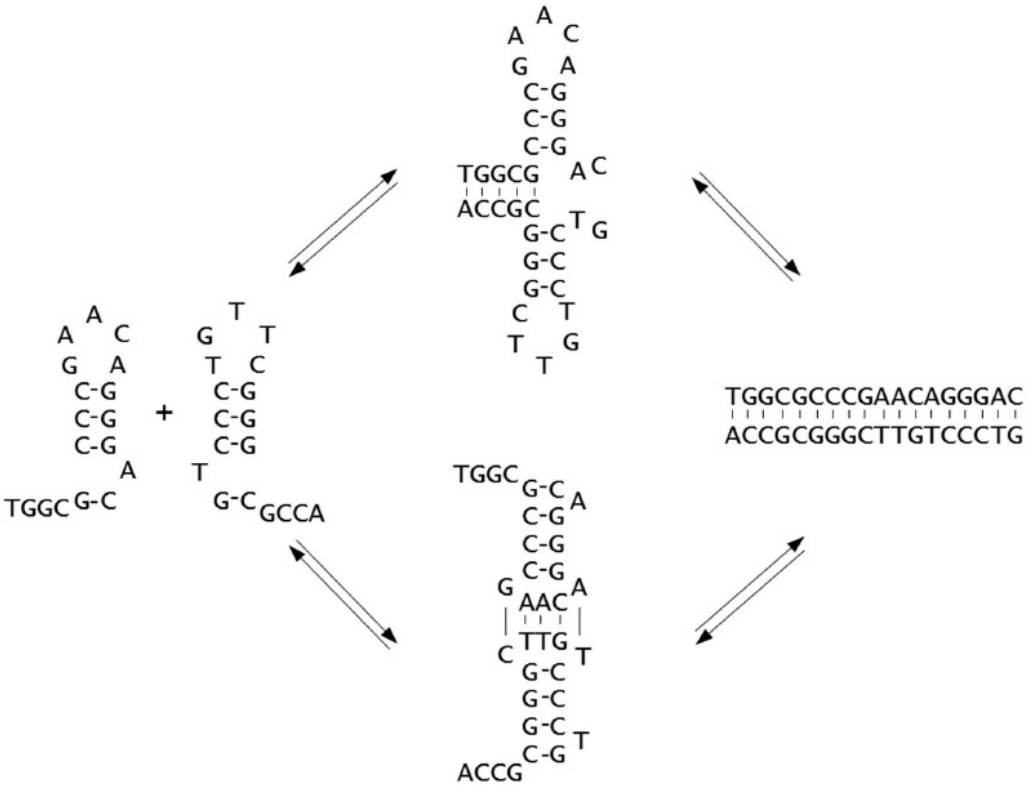


SSHS_NC(11-55)

KNVK-SFNSGKEGHTARNS-RAPRKKG-SWKSGKEGHQMKDS-TERQAN

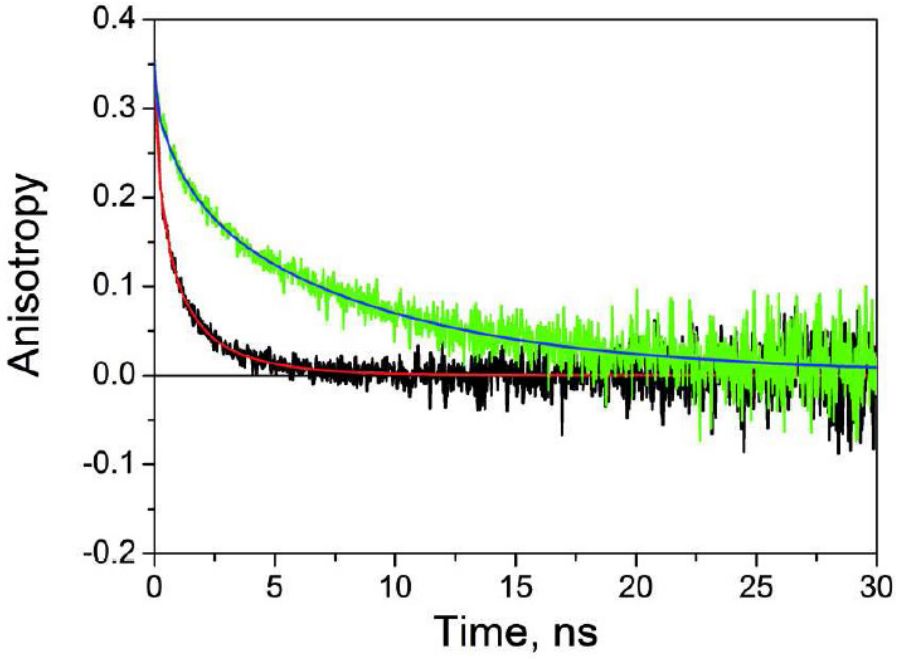
111x83mm (300 x 300 DPI)

Figure 1



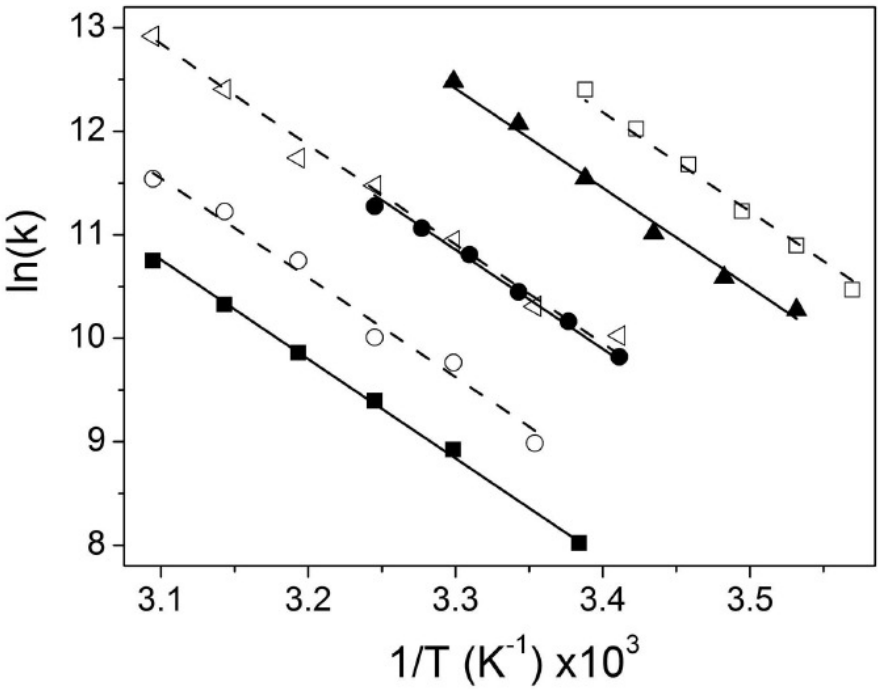
83x63mm (600 x 600 DPI)

Figure 2



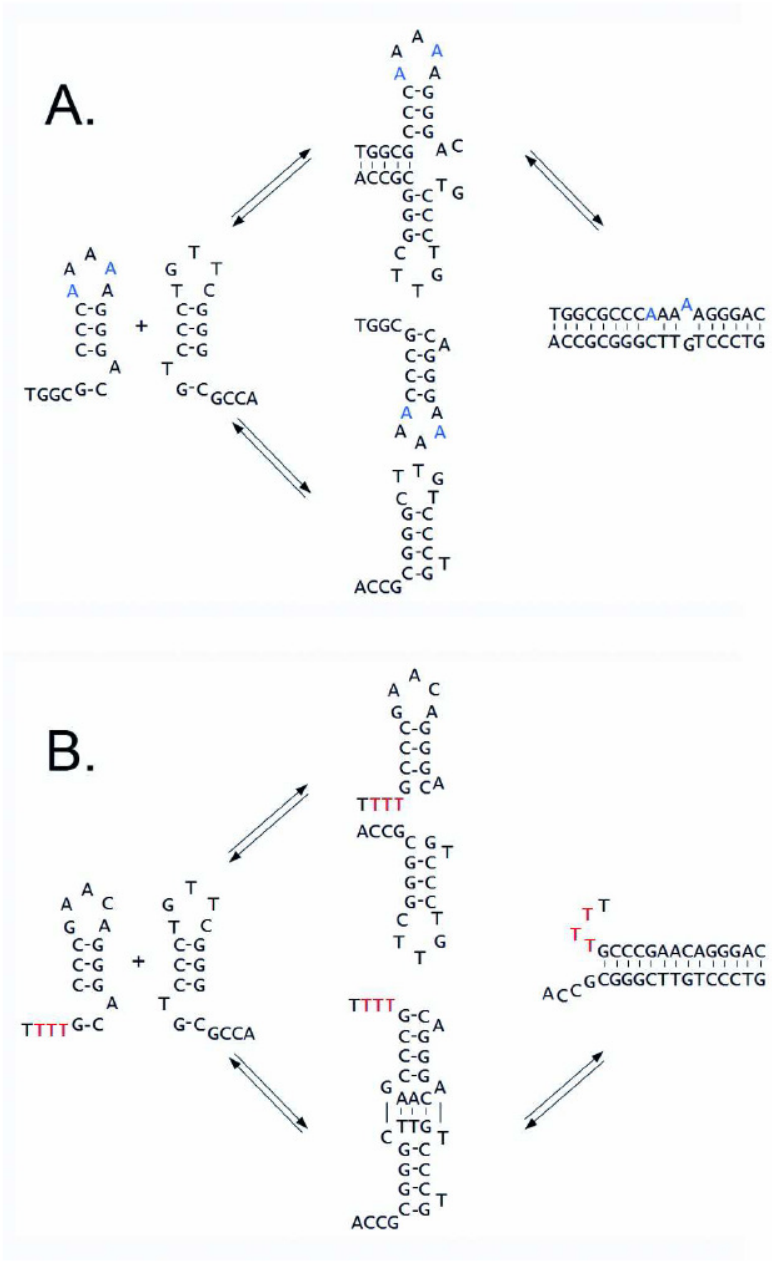
83x59mm (300 x 300 DPI)

Figure 3



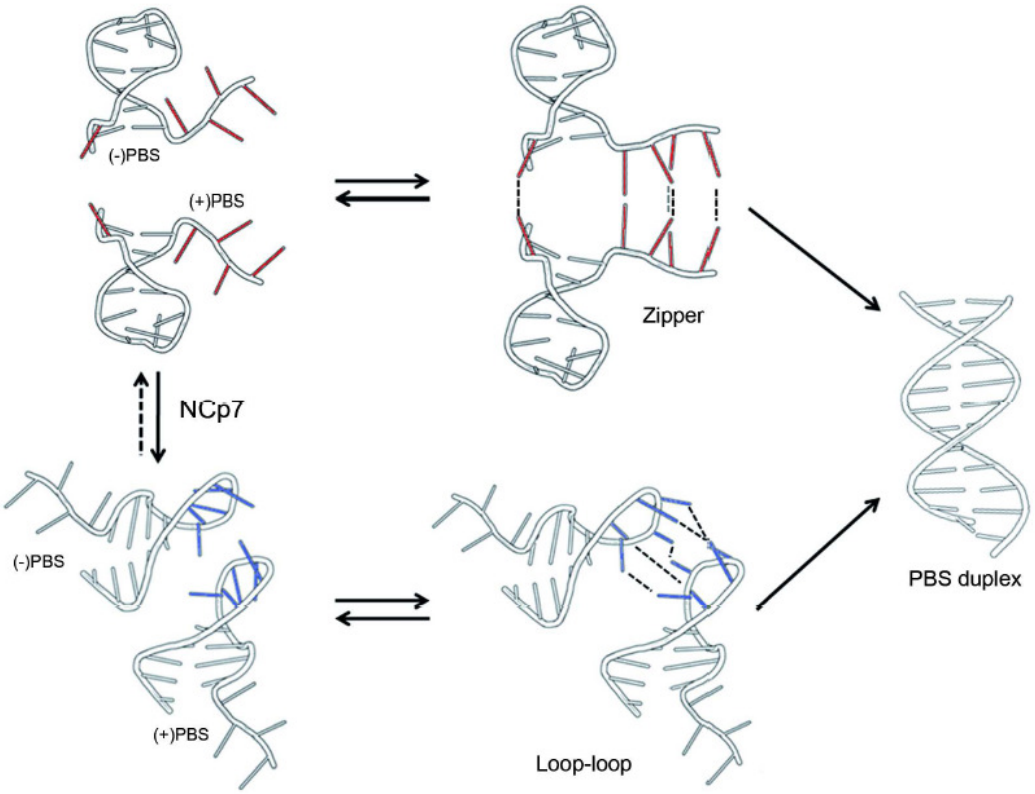
83x59mm (600 x 600 DPI)

Figure 4



83x133mm (300 x 300 DPI)

Figure 5



177x133mm (300 x 300 DPI)

Figure 6

Chapter 3: Kinetic investigation of cTAR/dTAR annealing promoted by HIV-1 Tat

Tat is a small, basic and flexible retroviral protein composed of 101 (clinical virus isolates) or 86 (laboratory adapted virus) residues. Its main function is to promote the transcription of the proviral DNA by the host RNA polymerase to synthesize large quantities of full length viral RNA. The biological activity of Tat relies, at least in part, on its specific binding to the transactivation response element TAR of the nascent transcript, which is a 56 nucleotides (nt) long stem-loop structure located at the 5' end of all HIV-1 RNAs. This in turn promotes the recruitment of a set of transcription factors and the efficient transcription of the provirus DNA. Moreover, Tat is also thought to promote reverse transcription (RTion), possibly through a direct interaction with reverse transcriptase and through its nucleic acid chaperoning activity. In line with the last hypothesis, it has already been shown that Tat through its basic 44-61 sequence drastically accelerates DNA-DNA annealing and DNA strand exchange, ribozyme-catalyzed cleavage of RNA and RNA transplicing. This nucleic acid annealing activity of Tat might explain, at least in part, its role in RTion.

To further investigate this possibility, we carried out a fluorescence study on the mechanism by which the full length Tat protein (Tat(1-86)) and the basic peptide (44-61) direct the annealing of complementary viral DNA sequences, named dTAR and cTAR, essential for the early steps of RTion. It has been observed that both peptides strongly promoted cTAR/dTAR annealing through non-specific attraction between the peptide-bound oligonucleotides. Using cTAR and dTAR mutants, this Tat promoted-annealing was found to be nucleated through the thermally frayed 3'/5' termini, resulting in an intermediate with 12 intermolecular base pairs, which then converts into the final extended duplex. Moreover, Tat(1-86) was as efficient as the nucleocapsid protein NCp7 in promoting cTAR/dTAR annealing, and could act cooperatively with NCp7 during the annealing reaction. Together with the strong similarities in the structure-activity relationships of the nucleic acid chaperone properties of Tat and its stimulatory effects in RTion, the cooperative effects of NCp7 and Tat strongly support a major role of the chaperone properties of Tat in RTion stimulation.

doi:10.1016/j.jmb.2010.05.033 *J. Mol. Biol.* (2010) **400**, 487–501

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

The Mechanism of HIV-1 Tat-Directed Nucleic Acid Annealing Supports its Role in Reverse Transcription

C. Boudier[1]*†, R. Storchak[1]†, K. K. Sharma[1], P. Didier[1], A. Follenius-Wund[1], S. Muller[2], J.-L. Darlix[3] and Y. Mély[1]*

[1]*Laboratoire de Biophotonique et Pharmacologie, UMR-CNRS 7213, Faculté de Pharmacie, Université de Strasbourg, 74 route du Rhin, 67401 Illkirch-Cedex, France*

[2]*CNRS-UPR 9021, Institut de Biologie Moléculaire et Cellulaire, 15 rue René Descartes, 67084 Strasbourg Cedex, France*

[3]*LaboRetro, Unité de Virologie Humaine INSERM #758, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon, France*

Received 23 March 2010; accepted 10 May 2010
Available online
21 May 2010

The main function of the HIV-1 trans-activator of transcription (Tat protein) is to promote the transcription of the proviral DNA by the host RNA polymerase which leads to the synthesis of large quantities of the full length viral RNA. Tat is also thought to be involved in the reverse transcription (RTion) reaction by a still unknown mechanism. The recently reported nucleic acid annealing activity of Tat might explain, at least in part, its role in RTion. To further investigate this possibility, we carried out a fluorescence study on the mechanism by which the full length Tat protein (Tat(1-86)) and the basic peptide (44-61) direct the annealing of complementary viral DNA sequences representing the HIV-1 transactivation response element TAR, named dTAR and cTAR, essential for the early steps of RTion. Though both Tat(1-86) and the Tat(44-61) peptide were unable to melt the lower half of the cTAR stem, they strongly promoted cTAR/dTAR annealing through non-specific attraction between the peptide-bound oligonucleotides. Using cTAR and dTAR mutants, this Tat promoted-annealing was found to be nucleated through the thermally frayed 3′/5′ termini, resulting in an intermediate with 12 intermolecular base pairs, which then converts into the final extended duplex. Moreover, we found that Tat(1-86) was as efficient as the nucleocapsid protein NCp7, a major nucleic acid chaperone of HIV-1, in promoting cTAR/dTAR annealing, and could act cooperatively with NCp7 during the annealing reaction. Taken together, our data are consistent with a role of Tat in the stimulation of the obligatory strand transfers during viral DNA synthesis by reverse transcriptase.

Keywords: Tat protein; NCp7 protein; TAR; fluorescence; nucleic acid chaperone

Edited by M. F. Summers

Introduction

Tat is a small, basic flexible retroviral protein composed of 101 (clinical virus isolates) or 86 (laboratory adapted virus) residues. Its main function is to promote the transcription of the proviral DNA by the host RNA polymerase to synthesize large quantities of full length viral RNA. This biological activity of Tat relies, at least in part, on its specific binding to the transactivation response element TAR of the nascent transcript, which is a 56 nucleotides (nt) long stem-loop structure located at the 5′ end of all HIV-1 RNAs[1,2] (and references therein). This in turn promotes the recruitment of a set of transcription factors and the efficient transcription of the proviral DNA.[3–5] The flexible nature[6,7] of Tat and its interaction with TAR are believed to provide multiple ways to interact with viral and cellular factors, as for example the TATA-binding protein (TBP), cyclin T1, cyclin-dependent kinase 9 (CDK9) and pTEF-beta,[8–11] which are needed for efficient provirus transcription. Moreover, Tat is thought to regulate Rev-dependent mRNA transport[12] mRNA capping,[13] splicing[14] and translation,[15] and to interfere with the cellular RNA interference (RNAi)

*Corresponding authors. E-mail addresses: christian.boudier@unistra.fr; yves.mely@unistra.fr.

† C.B. and R.S. contributed equally to this work.

Abbreviations used: TAR, trans-activation response element; Tat, trans-activator of transcription; RTion, reverse transcription; NC, nucleocapsid protein; ED, extended duplex; Fl, 5 (and-6)carboxyfluorescein; TMR, 5 (and-6) carboxytetramethylrhodamine; Rh6G, 6-carboxyrhodamine; Dabcyl, 4(-4′-dimethylaminophenylazo)benzoic acid.

0022-2836/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.

machinery through interactions with DICER and RNA.[16,17]

Another reported activity of Tat is its stimulatory effect on reverse transcription (RTion). Indeed, HIV-1 mutants deleted of the *tat* gene display a 3 to 5-fold defect in RTion compared to wild type HIV-1, in infected T cells and in endogeneous RTion reactions.[18,19] RTion in HIV-1 can be fully restored if the producer cells are transfected by a *tat* expression plasmid. In line with these data, recombinant Tat stimulates reverse transcriptase-directed *in vitro* DNA synthesis from 2 to 3-fold possibly through direct interaction with the enzyme[20]. The stimulatory activity of Tat appears to be mainly supported by its 60 first amino acids,[18] with a critical role for the Cys-rich (22-38) and the basic (49-58) domains as well as the Y47 residue of the core domain.[21] Ex vivo studies on Tat further reveal that RTion and transactivation are not correlated, suggesting that Tat directly stimulates RTion.[19] This stimulatory effect could possibly occur within HIV-1 virions or in the early steps of infection, since Tat was recently detected by mass spectrometry in highly purified preparations of HIV-1 virions produced by macrophages.[22] Alternatively, due to its ability to permeate the plasma membrane,[23] extracellular Tat could enter in newly infected cells in order to stimulate RTion after disassembly of the viral core[24] or to stimulate the natural endogeneous RTion that occurs during assembly and leads to infectious viral particles containing viral cDNA.[25–27] In addition, Tat was reported to inhibit reverse transcriptase activity at high concentrations,[20,24,28] suggesting a fine tuning of RTion by Tat, depending on its concentration. Tat may also stimulate the RTion reaction through its ability to chaperone the annealing of the primer tRNA onto the viral RNA[24] as well as the first cDNA strand transfer.[29] This hypothesis was further substantiated by a recent report[30] showing that Tat exhibits potent nucleic acid chaperone activities similar to those of the HIV-1 nucleocapsid protein NCp7[31–33] and of the Flaviviridae core protein.[34] This activity of Tat is mainly mediated by the basic 44-61 peptide sequence[2,35] which is the smallest known peptide with potent DNA/RNA annealing activities.

To further examine the molecular basis of the role of Tat in RTion, our aim was to characterize in depth the mechanism by which HIV-1 Tat(1-86) and Tat (44-61) (Fig. 1a) promote the hybridization of complementary TAR sequences, in the form of DNA (dTAR and cTAR), by monitoring the annealing kinetics with a cTAR derivative doubly labelled at its 3′ and 5′ ends by two different fluorophores.[31] The annealing reaction was also investigated using cTAR and dTAR mutants to identify the oligonucleotide determinants in the Tat-directed annealing pathway. Furthermore, since the stimulatory effect of Tat in RTion is functionally similar to that of NCp7,[36] we investigated whether the two proteins cooperate in stimulating the cTAR/dTAR annealing. Taken together, our results indicate that the hybridization reaction can be promoted by either Tat(1-86) or Tat(44-61) *via* a two-step mechanism involving the rapid formation of an intermediate complex through the annealing of cTAR and dTAR stem termini and its slower conversion into the final extended duplex. Moreover, Tat(1-86) and Tat(44-61) were found to chaperone in concert with NCp7 the annealing of two complementary oligonucleotides, suggesting that the nucleic acid chaperoning of Tat contributes to the RTion reaction.

Results

Investigating TAR oligonucleotide/Tat complex formation by dynamic light scattering

Preliminary experiments showed that addition of saturating quantities of Tat(1-86) or Tat(44-61) to a cTAR and/or dTAR solution led to substantial aggregation.[30] Since aggregates preclude further fluorescence analyses in solution, we first determined the experimental conditions where aggregation was negligible. To this end, we used dynamic light scattering to estimate the size of the complexes formed upon reaction of a constant concentration of dTAR (1 μM) with increasing concentrations of Tat ranging from 0 to 8 μM. The apparent particle size was found to be homogeneous, with a value of about 8 nm up to a molar ratio of 1:1. This value is close to that of dTAR alone, indicating that the complexes formed in these conditions did not aggregate. In sharp contrast, at molar ratios of Tat(1-86) to dTAR ≥ 1.5, polydispersed aggregates with dramatically increased sizes (>100 nm) were observed. Similarly, addition of Tat(44-61) to 1 μM dTAR provided complexes of 7–8 nm, up to a molar ratio of 3 peptides per oligonucleotide, while aggregates were observed for molar ratios ≥ 4. According to these results, we used molar ratios of Tat(1-86) and Tat(44-61) to oligonucleotides of 1:1 and 3:1, respectively, in all subsequent experiments. Since one cTAR molecule can bind up to 7 Tat(1-86) and 10 Tat(44-61) peptides with high affinity,[30] the fractional saturation of the oligonucleotides with these peptides is of 0.15 and 0.3, respectively, at these ratios.

Destabilization of cTAR secondary structure by Tat(1-86) and Tat(44-61)

Under non-aggregating conditions, one component of the activity of canonical chaperone proteins, such as HIV-1 NCp7, consists in their ability to destabilize the secondary structure of their nucleic acid targets. For instance, NCp7 binding to cTAR or dTAR induces a substantial melting of the lower half of their stem at temperatures well below their melting temperature (~47 °C).[37–40] To examine whether a similar conformational change of the cTAR stem loop can result from the binding of Tat, we used a cTAR doubly-labelled by 5 (and-6) carboxyfluorescein (Fl) and 5 (and-6) carboxytetramethylrhodamine (TMR) at its 3′ and 5′ ends, respectively. In its free state cTAR essentially exists

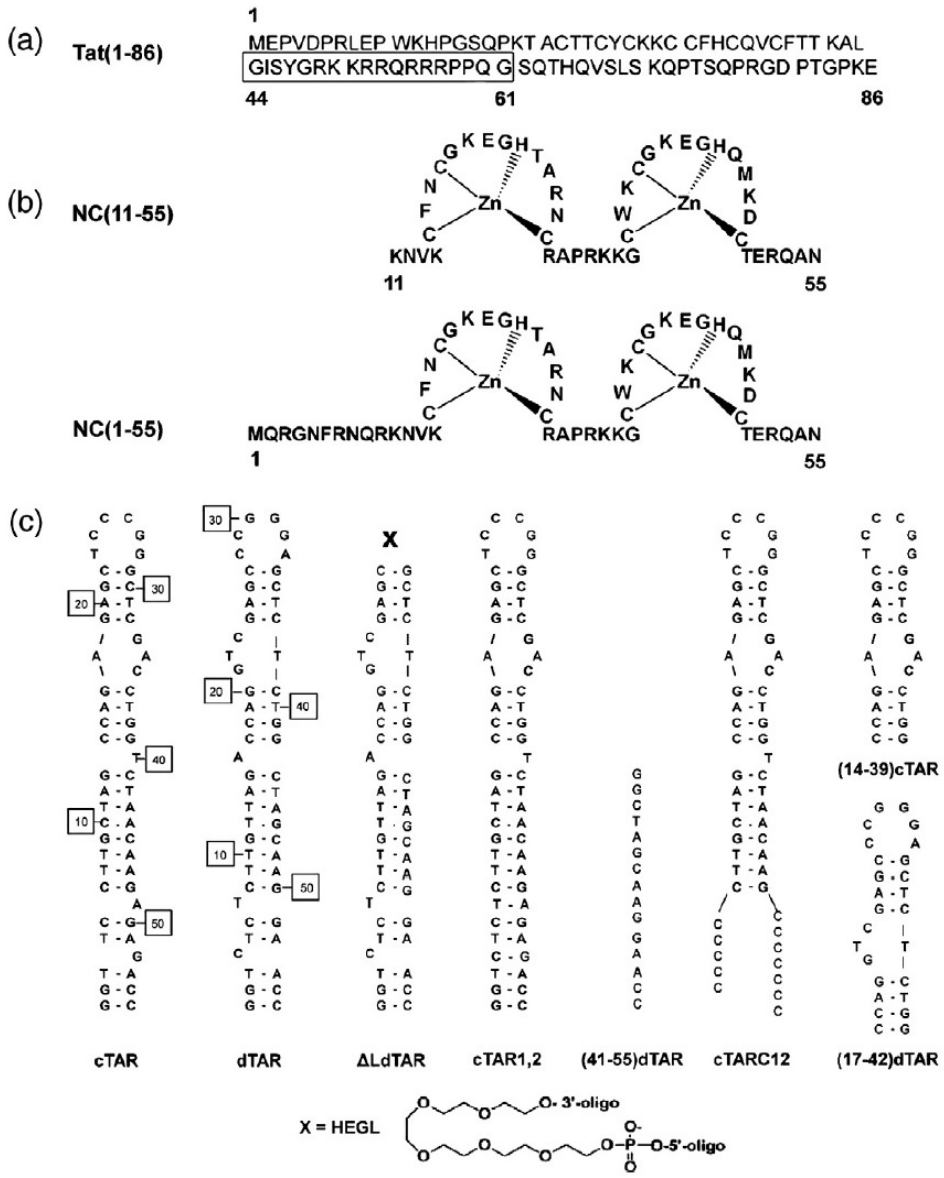


Fig. 1. Sequences of the peptides and TAR oligonucleotides used in this study. (a) Sequence of the Tat(1-86) protein. The boxed sequence corresponds to the basic (44-61) domain. (b) Sequence of the 11-55 and 1-55 peptides of the HIV-1 nucleocapsid protein. (c) Structure of the cTAR and dTAR oligonucleotides used in this study. The dTAR and cTAR DNA sequences are from the HIV-1 MAL strain. The secondary structures of the oligonucleotides were predicted from that of TAR[65] and the mfold program.[66] HEGL: hexaethylenglycol.

in a closed conformation, in which Fl exhibits a poor fluorescence due to its close proximity to TMR, which provides exciton coupling.[41,42] Stem destabilization increases the interchromophore distance and restores the Fl fluorescence.[42] Figure 2 shows that the fluorescence spectra of the doubly labelled cTAR mixed with an equimolar concentration of Tat (1-86) or a three-fold molar excess of Tat(44-61) at 20 °C were nearly superimposable with the emission spectrum of the free doubly labelled cTAR species. This indicates that both peptides were unable to destabilize the secondary structure of cTAR, under the present conditions.

Kinetics of cTAR/dTAR annealing in the presence of Tat(1-86)

In a previous investigation,[30] we described the ability of the full length Tat and of several fragments of the protein to strongly stimulate the annealing of cTAR to dTAR. In order to delineate the mechanism by which Tat(1-86) facilitates this annealing reaction, we reacted 10 nM of TMR-5′-cTAR-3′-Fl under pseudo-first order conditions with dTAR, added in excess, in the presence of holo-Tat(1-86), added at a protein to oligonucleotide molar ratio of 1:1. The annealing reaction was monitored in real time through the Fl

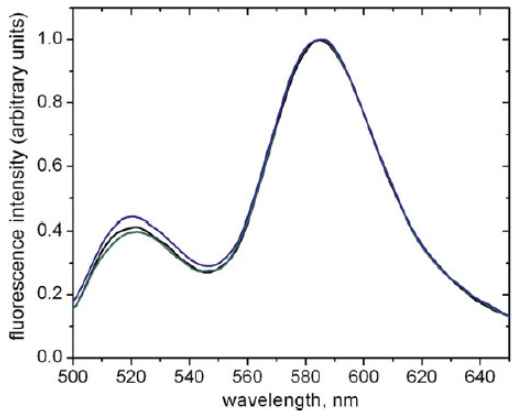


Fig. 2. Absence of destabilization of cTAR secondary structure by Tat(1-86) and Tat(44-61). Fluorescence emission spectra of 10 nM TMR-5′-cTAR-3′-Fl in the absence (blue line) and in the presence of 30 nM Tat(44-61) (green), or 10 nM Tat(1-86) (black), in 25 mM Tris–HCl (pH 7.5), 30 mM NaCl, 0.2 mM $MgCl_2$ at 20 °C. Excitation wavelength was 480 nm.

fluorescence increase resulting from the formation of the extended duplex (ED) (Fig. 3a inset).[30,31]

Figure 3a shows a typical progress curve. The fluorescence intensity at 520 nm markedly increased with time, reaching a plateau value, in about 30 minutes. This plateau value was identical to that obtained by promoting cTAR/dTAR duplex formation by heating, indicating that the reaction went to completion and that Tat(1-86) was unable to dissociate the final ED. The progress curve could not be satisfactorily fitted to a single exponential function, as can be seen from the non-random distribution of the residuals around the 0 value (Fig. 3b). A much better fit was obtained with a double exponential function:

$$I(t) = I_f - (I_f - I_0)\left(ae^{-k_{obs1}(t-to)} + (1-a)e^{-k_{obs2}(t-to)}\right) \quad (1)$$

where $I(t)$ is the actual fluorescence intensity at 520 nm, k_{obs1} and k_{obs2} are the apparent pseudo-first order rate constants, a is the relative amplitude of the fast component, and t_0 is the dead time. I_0 and I_f stand for the fluorescence intensities of TMR-5′-cTAR-3′-Fl in its free state and in the final ED, respectively. I_f, I_0 and t_0 were determined independently. The progress curve in Fig. 3a was adequately fitted to Eq. (1) with k_{obs1}, k_{obs2} and a values of $(2.4\pm0.2)\times10^{-2}$ s^{-1}, $(2.4\pm0.1)\times10^{-3}$ s^{-1} and 0.43, respectively, which provides a nearly random distribution of the residuals (Fig. 3c) and $\chi^2=0.9992$. Such biphasic kinetics has also been reported for the spontaneous and NCp7-facilitated annealing of cTAR to either TAR RNA or dTAR[31,43] as well as for the mini TAR RNA/DNA annealing reaction.[44]

The annealing experiments were then repeated using dTAR concentrations ranging from 100 to 500 nM, keeping constant the ratio of one molecule of protein per molecule of cTAR and dTAR. All the curves were satisfactorily fitted to Eq. (1). We found a linear dependence of both k_{obs1} and k_{obs2} on $[dTAR]$ within the range of concentrations used, suggesting that the two kinetic components participating to the Tat(1-86)-directed cTAR/dTAR annealing conform to simple bimolecular reactions. Thus, for each component, the second order association rate constant k_{ass} and the first order dissociation rate constant k_{diss} can be obtained by:

$$k_{obs1,2} = k_{ass1,2}[dTAR] + k_{diss1,2} \quad (2)$$

From the fits to Eq. 2 (Fig. 3d and 3e), we found $k_{ass1} = (7\pm1)\times10^4$ M^{-1} s^{-1} and $k_{ass2} = (2.1\pm0.4)\times10^4$ $M^{-1}s^{-1}$ (Table 1). These values were 8- and 250-fold higher than those for the corresponding rate constants in the absence of Tat.[31] k_{diss1} was determined to be $1.6\pm0.4\times10^{-2}$ s^{-1}, highlighting the limited stability of the fast reaction product in the presence of Tat(1-86). In contrast, the dissociation rate constant k_{diss2} for the

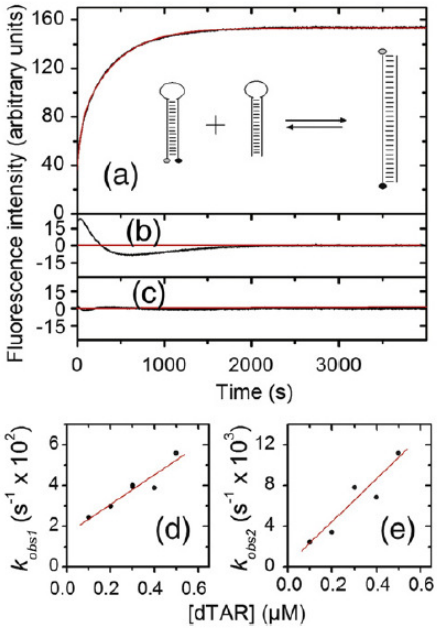


Fig. 3. Promotion of cTAR-dTAR annealing by Tat (1-86). (a) Kinetic trace (black line) for the reaction of 10 nM doubly labelled cTAR with 100 nM dTAR and 110 nM holo-Tat(1-86), in 25 mM Tris–HCl (pH 7.5), 30 mM NaCl, 0.2 mM $MgCl_2$ at 20 °C. Fl emission was monitored at 520 nm (with excitation at 480 nm). The red line represents the best fit of the kinetic data points to Eq. (1) with the k_{obs1}, k_{obs2}, and a values given in the text. Inset: principle of the fluorescence assay used to monitor the annealing kinetics. The low fluorescence of cTAR in the absence of TAR and Tat is related to the close proximity of Fl (open circle) and TMR (filled circle). Formation of the extended duplex strongly increases the interchromophore distance and thus, restores the Fl emission. (b) Weighted residuals for a mono-exponential and (c) for a two-exponential fit of the experimental data. Dependence of the apparent pseudo-first order rate constants k_{obs1} (d) and k_{obs2} (e) on the dTAR concentration. The theoretical lines (red) were calculated with Eq. (2).

Table 1. Kinetic constants for the Tat-promoted annealing of cTAR with dTAR

Labelled sequence	Complementary sequence	Peptide	Peptide to oligonucleotide ratio	Fast kinetic component			Slow kinetic component[g]			
				k_{ass1}(M^{-1} s^{-1})	k_{diss} (s^{-1})	K_a* (M^{-1})[f]	k_3(s^{-1})	k_{ass2}	k_{ass2} or $k_2.K_a$* (M^{-1} s^{-1})
cTAR	dTAR	-	0	$(9.0\pm0.8)\times10^{3}$[a]	-	-	-	-	(85 ± 20)[a,f]
cTAR	dTAR	holo-Tat(1-86)	1	$(7\pm1)\times10^{4}$	$(1.6\pm0.4)\times10^{-2}$	-	-	-	$(2.1\pm0.4)\times10^{4}$[f]
cTAR	dTAR	apo-Tat(1-86)	1	$(7\pm3)\times10^{4}$	$(2.4\pm0.6)\times10^{-2}$	-	-	-	$(2.1\pm0.4)\times10^{4}$[f]
cTAR	dTAR	Tat (44-61)	3	$(1.2\pm0.1)\times10^{5}$[d]	$(1.6\pm0.9)\times10^{-2}$[d]	$(7\pm2)\times10^{3}$[e]	$(1.4\pm0.2)\times10^{-2}$[b]	$(1.3\pm0.3)\times10^{-2}$[e]	$(6\pm3)\times10^{4}$
cTAR	ΔLdTAR	Tat (44-61)	3	$(1.4\pm0.4)\times10^{5}$[d]	$(1.8\pm0.9)\times10^{-2}$[d]	$(1.1\pm0.3)\times10^{2}$[e]	$(1.5\pm0.2)\times10^{-2}$[c]	$(1.2\pm0.2)\times10^{-2}$[e]	$(7\pm3)\times10^{4}$

[a] from Ref. 31.
[b] obtained from k_{obs1} and Eq. (2).
[c] obtained from k_{obs2} and Eq. (3).
[d] obtained by fitting the data to Equations S8 and S9 (Supplementary Material).
[e] obtained by fitting simultaneously all experimental progress curves with the Dynafit software.
[f] for these reaction mixtures, the slow kinetic component was found to follow a bimolecular reaction.
[g] in all cases, the rate constant for the dissociation of the EDs was too low to be determined ($\leq 10^{-5}$ s^{-1}).

slow component could not be determined since the intercept with the Y axis was too close to the origin, indicating that in the presence of Tat(1-86) the slow component leads to a stable extended duplex (ED) as for the spontaneous annealing reaction.[31]

The possible influence of zinc binding to Tat(1-86) on its annealing activity was examined by reacting cTAR, dTAR and the apo–form of Tat(1-86) in the same conditions as above. We found values of k_{ass1} and k_{ass2} not significantly different from those of the annealing kinetics directed by the holo-form of Tat (Table 1). Such a similarity in the DNA annealing reaction for holo- and apo-Tat indicates that this activity does not rely on the binding of zinc to the cysteine-rich region of the protein.

Promotion of cTAR/dTAR annealing by the basic Tat(44-61) peptide

To investigate more in depth the role of Tat on cTAR/dTAR annealing, we substituted the full length protein by the basic peptide 44-61, which binds TAR with an affinity ($K_a = 0.9\times10^8$ M^{-1})[30] similar to that of the full length Tat ($K_a = 2.3\times10^8$ M^{-1}).[30] The reaction of the doubly labelled cTAR with dTAR was investigated in the presence of Tat(44-61) in pseudo-first order conditions as described above, but with a larger dTAR concentration range (50 to 700 nM) and a molar ratio of peptide per oligonucleotide of 3:1. The resulting progress curves were similar to those for the Tat(1-86)-directed ED formation, showing two kinetic components (data not shown). As with Tat(1-86), the final fluorescence plateau was independent of the dTAR concentration, suggesting that Tat(44-61) marginally dissociated the final ED. Interestingly, the fast component, k_{obs1}, linearly varies with [*dTAR*] (Fig. 4), while k_{obs2} shows a hyperbolic dependence on [*dTAR*]

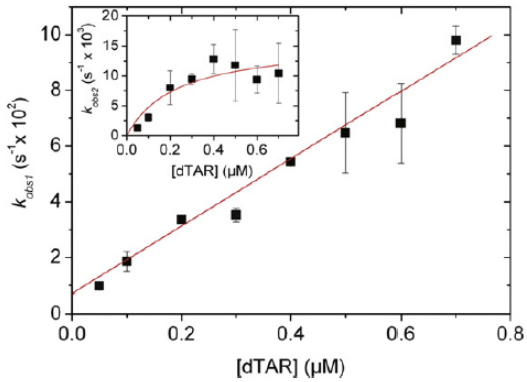


Fig. 4. Kinetics of cTAR/dTAR annealing in the presence of Tat(44-61) at a ratio of three peptides per oligonucleotide. The observed rate constants for the fast kinetic phase (k_{obs1}) and the slow one (k_{obs2}, inset) were plotted as a function of the dTAR concentration. The rate constants were determined as described in Fig. 3. Each data point represents the mean of 3-4 measurements. The red lines were the best fits to Eq. (2) (k_{obs1}) and Eq. (3) (k_{obs2}), using the values of the kinetic parameters reported in Table 1.

(Fig. 4, inset). This suggests that the Tat(44-61)-promoted ED formation probably proceeds *via* a two step-mechanism as shown in Scheme 1.

$$cTAR + dTAR \underset{k_{diss}}{\overset{k_{ass}}{\rightleftharpoons}} ED^* \xrightarrow{k_2} ED$$

Scheme 1.

where a fast pre-equilibrium intermediate ED* precedes the formation of the final stable ED through a monomolecular reaction. The hyperbolic dependence of k_{obs2} on the dTAR concentration may be ascribed to ED* accumulation because of its slow conversion into ED which likely constitutes the rate-limiting step of the Tat-promoted annealing pathway. If the pre-equilibrium is rapidly reached ($k_{ass}.[dTAR]+k_{diss} \gg k_2$), scheme 1 predicts a linear variation of k_{obs1} as described in Eq. (2) and a hyperbolic variation of k_{obs2} as described in Eq. (3):

$$k_{obs2} = \frac{k_2 K_a^*[dTAR]}{1 + K_a^*[dTAR]} \quad (3)$$

where K_a^* is the equilibrium association constant governing the ED* formation ($K_a^* = k_{ass}/k_{diss}$). From the slope and the intercept with the Y axis in Fig. 4, we found that $k_{ass} = (1.20 \pm 0.04) \times 10^5\ M^{-1}s^{-1}$ and $k_{diss} = (7.1 \pm 0.8) \times 10^{-3} s^{-1}$. Non-linear regression of the k_{obs2} data (Fig. 4 inset) with Eq. (3) yielded $k_2 = (1.5 \pm 0.4) \times 10^{-2} s^{-1}$ and $K_a^* = (4 \pm 2) \times 10^6\ M^{-1}$ (Table 1).

In a next step, we checked that fitting the fluorescence data to a double exponential function was consistent with the reaction Scheme 1. To this end, reaction was described in a set of differential equations (integrating all rate constants and species concentrations, see supplementary material) and the resulting calculated concentrations were introduced into Eq. (4) describing the fluorescence intensity at any time t:

$$I_{(t)} = I_0 \frac{[cTAR]}{[cTAR]_0} + I_i \frac{[ED^*]}{[cTAR]_0} + I_f \frac{[ED]}{[cTAR]_0} \quad (4)$$

where $[cTAR]_0$ is the cTAR concentration at time 0, I_0 is the initial fluorescence intensity (due to free TMR-5′-cTAR-3′-Fl), I_i and I_f represent the fluorescence intensity of ED* and ED, respectively.

The final analytical equation describing the fluorescence intensity changes with time (equation S10 in the supplementary material) was fully consistent with the bi-exponential empirical Eq. (1). Moreover, the k_{ass}, k_{diss} and k_2 values determined from the analytical model (using Equations S8 and S9 in the supplementary material) were found to be very close to those determined from the empirical approach (Table 1). In addition, the dependence of the amplitude a on $[dTAR]$ could be adequately fitted to Eq. S16 (Fig. 5), while fixing the values of the kinetic rate constants to those of Table 1. The only variable in this case was I_i, the ED* fluorescence intensity, that was found to correspond to 93±3% of the I_f value. The I_i value was then used to calculate R,

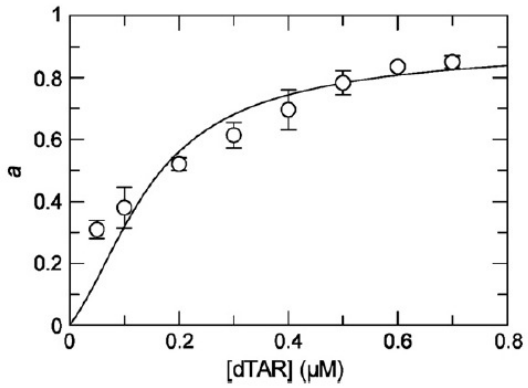


Fig. 5. dTAR concentration dependence of the relative amplitude a of the fast kinetic component for the Tat(44-61)-promoted annealing. The values of a were determined as described in Fig. 3. The dark line represents the fit to the data points using Equation S16 and the value of I_i, the fluorescence intensity of ED*, reported in the text.

the distance separating TMR and Fl in the ED* intermediate:

$$R = R_0(1/E - 1)^{1/6} \text{ and } E = 1 - I_i/I_f \quad (5)$$

where R_0, the Förster distance for the TMR-Fl pair attached to the 5′ and 3′ ends of TAR was determined to be 60 Å,[40] and E is the energy transfer efficiency. We found an interchromophore distance $R = 92$ Å.

To further check the validity of the postulated two-step annealing mechanism, we also used the Dynafit numerical resolution software,[45] which allows to simultaneously fit the experimental progress curves obtained at different dTAR concentrations. The best estimates of the elementary rate constants k_{ass}, k_{diss}

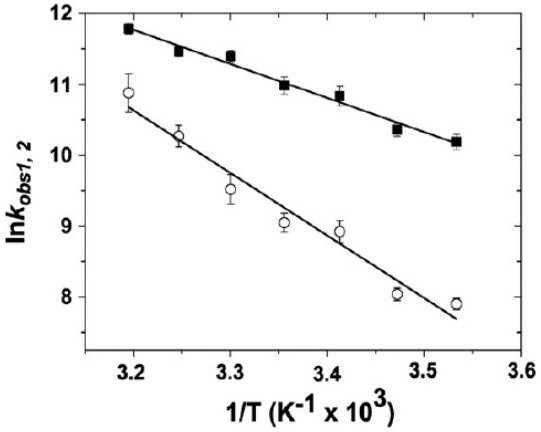


Fig. 6. Temperature dependence of cTAR/dTAR annealing kinetics promoted by Tat(44-61) at a peptide to oligonucleotide ratio of 3:1. 10 nM of doubly-labeled cTAR was reacted with 100 nM dTAR in 25 mM Tris–HCl (pH 7.5), 30 mM NaCl, 0.2 mM $MgCl_2$. The fast (■) and slow (○) kinetic rate constants were fitted according to the Arrhenius equation (solid lines), using $E_{a,i}$ values of 9.6±0.6 kcal/mol and 18±2 kcal/mol, respectively.

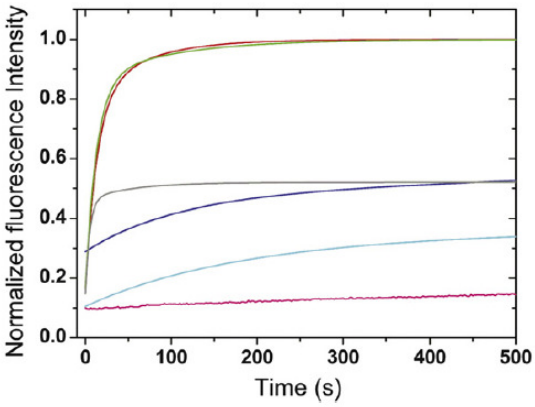


Fig. 7. Kinetic traces for the Tat(44-61)-directed annealing of wild type or mutant cTAR with dTAR derivatives. The concentration of the doubly labelled wild type or mutant cTAR was 10 nM. The concentration of the non-labelled dTAR derivative was 500 nM, except for (41-55) dTAR where it was 300 nM, due to the high reaction speed in this case. Annealing of cTAR to dTAR (red), cTAR to ΔLdTAR (green), cTAR to (41-55)dTAR (grey), cTARC12 to dTAR (blue), cTAR1,2 to dTAR (cyan) and cTAR1,2 to ΔLdTAR (pink).

and k_2 (Table 1) were again in excellent agreement with those found by the empirical and the analytical approaches, further supporting the proposed reaction scheme.

To further characterize the kinetic reaction, the temperature dependence of the k_{obs} values was monitored as a function of the temperature and analyzed using the Arrhenius equation:

$$k_i = A_i exp(-E_{a,i} / RT) \quad (6)$$

where the rate constant k_i is given by k_{obsi}/[dTAR], A_i is the pre-exponential Arrhenius factor, $E_{a,i}$ is the activation energy, R is the universal gas constant, and T is the temperature (in Kelvin). Both reaction rates increased with temperature (Fig. 6), giving positive enthalpy values for the transition state of 9.6±0.6 kcal/mol and 18±1.5 kcal/mol for the fast and slow components, respectively. These values indicate that the Tat-promoted cTAR/dTAR annealing involves premelting of two to three base pairs for the fast component and four to five base pairs for the slow component.[46,47]

Finally, to check that the annealing kinetics were not probe-dependent, we substituted TMR-5′-cTAR-3′-Fl by cTAR covalently linked at its 5′ and 3′ ends by a 6-carboxyrhodamine fluorophore (Rh6G) and a non-fluorescent quencher 4(-4′-dimethylaminophenylazo)benzoic acid (Dabcyl), respectively. The time course of Rh6G-5′-cTAR-3′-Dabcyl annealing with dTAR was again biphasic and well described by Eq. (1), yielding kinetic rate constants similar to those obtained with TMR-5′-cTAR-3′-Fl (data not shown), indicating that the kinetics were only slightly affected by the nature of the probes.

Molecular determinants of cTAR and dTAR annealing by the Tat(44-61) peptide

We used several cTAR and dTAR mutants to characterize the molecular determinants in the annealing reaction promoted by Tat(44-61) (Fig. 1c). To characterize the possible role of the central loops, we used ΔLdTAR, a dTAR mutant where the loop was replaced by a flexible hexaethylenglycol tether. The kinetic traces and their final fluorescence at the plateau were very close to the corresponding ones with dTAR. Furthermore, the values of the fitted parameters (Table 1) were similar to those obtained with the native dTAR, indicating that the loops play a marginal role. Consequently, the annealing process is most probably initiated through the cTAR and dTAR stems.

To confirm the role of the stems in the initiation of the annealing reaction, we substituted cTAR by cTAR1,2 where bases complementary to the bulged residues at position 49 and 52 were added to confer a perfect double stranded structure to the bottom part of cTAR stem (Fig. 1c). The reaction of TMR-5′-cTAR1,2-3′-Fl with dTAR in the presence of Tat(44-61) was found to lead to the same final fluorescence intensity than with cTAR/dTAR, indicating that it yielded the same final ED. The progress curves with cTAR1,2 were also biphasic, but the kinetics were much slower than with cTAR (Fig. 7). From the linear variation of k_{obs1} with dTAR concentration, we found a k_{ass1} value (Table 2) 20-fold lower than for cTAR (Table 1). Due to the very slow kinetics, the nature of the k_{obs2} variation could not be unambiguously determined within the range (0.5-3 μM) of dTAR concentrations used. Therefore, we cannot exclude that cTAR1,2 and cTAR hybridize with a mechanism different from dTAR. Assuming a linear variation of

Table 2. Kinetic rate constants for the annealing of cTAR to dTAR derivatives in the presence of Tat(44-61)[a]

Labelled sequence	Complementary sequence	Fast component		Slow component[b]
		k_{ass1} (M^{-1} s^{-1})	k_{diss} (s^{-1})	k_{ass2} (M^{-1} s^{-1})[c]
cTAR1,2	dTAR	$(6.4\pm2)\times10^3$	$(6\pm3)\times10^{-3}$	2.8×10^2
cTARC12	dTAR	$(1.5\pm0.1)\times10^4$	$(4.5\pm2)\times10^{-3}$	$(8\pm1)\times10^2$
cTAR	(41-55)dTAR	$(7\pm1.5)\times10^5$	$(4\pm1)\times10^{-2}$	$(6\pm2)\times10^4$
(14-39)cTAR	(17-42)dTAR	$(2.1\pm0.1)\times10^5$	$(1.3\pm0.2)\times10^{-2}$	$(4\pm1)\times10^4$
(14-39)cTAR	dTAR	$(3.7\pm0.9)\times10^4$	$(1.4\pm0.4)\times10^{-3}$	$(2.8\pm0.2)\times10^3$

[a] a peptide to oligonucleotide molar ratio of 3 was used.
[b] the dissociation rate constant of the ED was too low to be accurately determined.
[c] determined assuming that the slow kinetic component corresponds to a bimolecular reaction.

k_{obs2}, we obtained a k_{ass2} value (Table 2) only 3-fold higher than that found for cTAR/dTAR hybridization in the absence of Tat. Thus, the stabilization of the bottom part of the cTAR stem strongly reduces the overall rate of the Tat(44-61)-directed annealing reaction by affecting both kinetic components. These results highlight the importance of the bulges at positions 49 and 52, which cooperatively destabilize the bottom part of the cTAR stem.[37] Noticeably, when cTAR1,2 was reacted with ΔLdTAR (Fig. 7), a still slower annealing reaction was observed suggesting a possible contribution of a loop-loop pathway in the Tat(44-61)-facilitated cTAR1,2/dTAR annealing reaction.

The critical role of the bottom part of the stems in the Tat(44-61)-directed annealing was further substantiated by using cTARC12, a doubly labelled variant of cTAR where the 5 and 7 terminal residues of the 5′ and 3′ ends, respectively , are changed to C (Fig. 1c). In this variant, the lower part of the stem is not double-stranded and both 3′ and 5′ terminal portions cannot anneal to the corresponding sequences in dTAR. The rather slow kinetic constants obtained with cTARC12 (Fig. 7 and Table 2) support the importance of the complementarity of the terminal reacting sequences in the Tat-mediated annealing reaction.

To further explore the role of the bottom part of the stems on the annealing kinetics, the doubly labelled cTAR was mixed with (41-55)dTAR, a single-stranded oligonucleotide (Fig. 1) corresponding to the 3′-terminal end of dTAR, and thus complementary to the 5′ terminal 15 nt of cTAR. Mixing of the doubly labelled cTAR with (41-55)dTAR in the presence of Tat(44-61) resulted in a fast fluorescence increase (Fig. 7). Interestingly, the plateau value was about half of that obtained with the native dTAR, irrespective of the (41-55)dTAR concentration used, suggesting that annealing of (41-55)dTAR to the 5′ terminus of cTAR did not cause the melting of the upper portion of the cTAR stem. This is in variance with the complete melting of the stem observed with the HIV-1 NCp7 protein,[31] in line with the stem destabilizing properties exhibited by NCp7 and not by the basic Tat(44-61) peptide. As with wild type dTAR, the progress curves were biphasic. Due to the high reaction speed, the highest (41-55)dTAR concentration compatible with accurate k_{obs1} and k_{obs2} determination using manual mixing was limited to 300 nM. Linear regression analysis of k_{obs1} values yielded $k_{ass1}=7\times10^5$ $M^{-1}s^{-1}$ (Table 2), a value almost 6-fold higher than for the reaction with native dTAR. Because of the restricted range of dTAR concentrations, the variation of k_{obs2} was linear too, yielding $k_{ass2}=6\times10^4$ $M^{-1}.s^{-1}$. The fast hybridization of the (41-55)dTAR sequence to the 5′ end of cTAR confirmed the importance of the complementary reacting strands in the lower half of the cTAR and dTAR stems, in the Tat(44-61)-mediated annealing. Moreover, the fast annealing observed with (41-55)dTAR may also be ascribed to the absence of secondary structure in this sequence.

To further address the importance of the complementarity of reacting sequences at the stem termini, we investigated the Tat(44-61)-promoted annealing of the doubly labelled (14-39)cTAR variant, also called mini-cTAR,[44] composed of the central loop and the upper portion of the stem[48] (Fig. 1c) with its complementary (17-42)dTAR sequence. At the end of the reaction, the fluorescence intensity was about 40% lower than that resulting from cTAR/dTAR annealing, a difference consistent with the smaller distance between the two probes in the (14-39) cTAR/(17-42)dTAR ED. The kinetic rate constants (Table 2) are similar to the ones measured for cTAR/dTAR, indicating that Tat(44-61) can efficiently promote hybridization through the upper part of the stem. Substituting (17-42)dTAR by the native dTAR was found to substantially slow down the reaction rates (Table 2), further indicating that the complementarity between the terminal sequences of the reacting oligonucleotides was critical in the Tat (44-61)-promoted annealing reaction.

Noticeably, the annealing kinetics of all mutants was orders of magnitude slower in the absence than in the presence of Tat (data not shown), indicating that the kinetic rate constants reported in Table 2 were not biased by contributions from spontaneous annealing of the complementary TAR sequences. Taken together, these results suggest that the Tat-promoted annealing of cTAR with dTAR is primarily mediated through their terminal respective complementary sequences.

Comparison of the TAR annealing activity of Tat and NCp7

To further explore the possible role of Tat in the hybridization of complementary TAR sequences involved in RTion, we compared the annealing activities of Tat and NCp7, individually and together. First, we compared the nucleic acid annealing activity of Tat(44-61) and NC(11-55), a truncated NCp7 peptide with an annealing activity similar to that of the full length protein, but with low nucleic acid aggregating properties. The annealing kinetics of doubly labelled cTAR to dTAR (10 and 100 nM, respectively) in the presence of Tat(44-61) at a peptide to oligonucleotide molar ratio of 3:1 was compared to the annealing kinetics with NC(11-55) added at ratios of 4:1, 5:1 and 6:1 (Fig. 8a). It was observed that three molecules of Tat(44-61) exhibited the same annealing activity as five molecules of NC (11-55). This was confirmed by experiments with peptide mixtures, since the kinetic curves with Tat (44-61) added at a ratio of 3:1 and NC(11-55) at a ratio of 3:1 or 5:1 were superimposable to those obtained in the presence of NC(11-55) alone at the ratio of 8:1 and 10:1, respectively (Fig. 8b). Thus, we conclude that the two viral peptides can act in concert to chaperone cTAR/dTAR annealing. These data prompted us to compare the chaperone properties of the native Tat(1-86) and NC(1-55) proteins. To avoid the NC(1-55)-induced aggregation of oligonucleotides, the experiments were performed in 50 mM

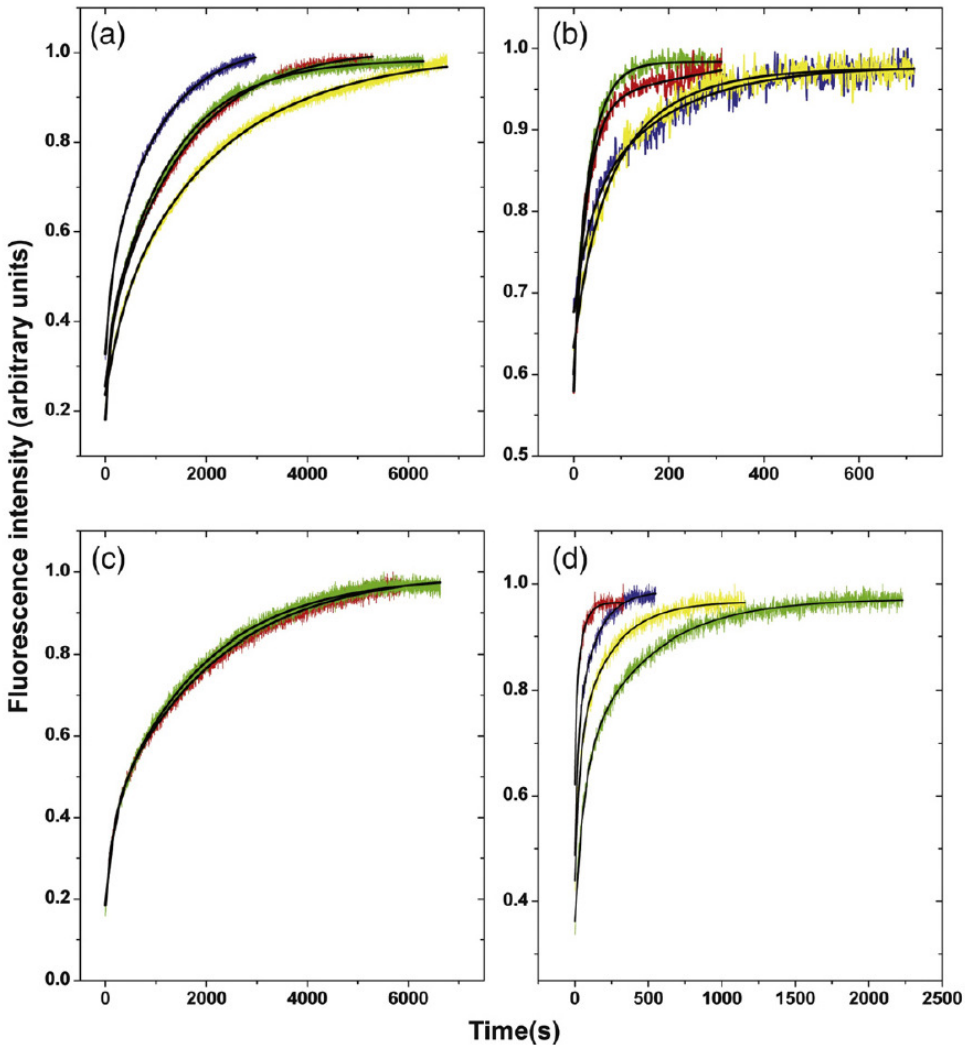


Fig. 8. Comparison of the chaperone activity of the Tat and NC peptides. All progress curves shown describe the peptide-promoted reactions of 10 nM doubly labelled cTAR with 100 nM dTAR. (a) Comparison of the cTAR/dTAR annealing kinetics obtained in the presence of Tat(44-61) at peptide to oligonucleotide molar ratio of 3:1 (green curve) and NC(11-55) at ratio of 4:1 (yellow), 5:1 (red) and 6:1 (blue). (b) cTAR/dTAR annealing kinetics observed in the presence of a mixture of Tat(44-61) and NC(11-55), both added at molar ratio 3:1 (yellow trace) and in the presence of NC(11-55) alone (molar ratio 8:1, blue trace). Similar annealing kinetics were also recorded in the presence of a mixture of Tat(44-61) and NC(11-55) (ratio 3:1 and 5:1, respectively, red trace) and NC(11-55) alone (ratio 10:1, green trace). (c) Comparison of the chaperone properties of Tat(1-86) and NC(1-55). Kinetic traces with Tat(1-86) (green) and NC(1-55) (red), added at a peptide to oligonucleotide molar ratio of 1:1, show that the two proteins exhibit similar chaperone activity for cTAR/dTAR annealing. (d) Cooperative chaperone properties of Tat(1-86) and NC(1-55). Comparison of the kinetic traces shows that cTAR/dTAR annealing is promoted more efficiently by a mixture of Tat(1-86) and NC(1-55) (molar ratio 1:1 and 4:1, respectively, yellow curve) than by NC(1-55) alone (molar ratio 5:1, green curve). Similarly, the annealing reaction proceeds faster in the presence of Tat(1-86) and NC(1-55) added at a molar ratio of 1:1 and 8:1, respectively (red curve) than in the presence of NC(1-55) alone at a ratio of 9:1 (blue curve). All experimental traces were satisfactorily fitted to a double-exponential function (black solid lines) yielding k_{obs1}, k_{obs2} and a. These parameters were used to calculate the time of completion of the reaction, corresponding to the time required to obtain 96.9% of the maximum fluorescence intensity, ie $5\times\ln 2/k_{mean}$, where k_{mean} is the mean rate constant given by $1/k_{mean}=a/k_{obs1}+(1-a)/k_{obs2}$.

Tris-HCl pH 8.0, 50 mM KCl and 10 mM $MgCl_2$ as previously reported.[49] The kinetic traces obtained with either NC(1-55) or Tat(1-86) at a peptide to oligonucleotide ratio of 1:1 showed that both proteins equally promoted the annealing of cTAR to dTAR (Fig. 8c). We then compared the kinetic traces of mixtures of Tat(1-86) at ratio of 1:1 and NC (1-55) at ratios of 4:1 and 8:1 to those obtained in the presence of NC(1-55) alone at ratios of 5:1 and 9:1, respectively. Interestingly, mixing the two proteins promoted a more efficient annealing reaction than their equivalent in NC(1-55) alone as judged by the time of completion (calculated as indicated in the legend of Fig. 8d) which was reduced by 2.2- to 3.8-fold when both proteins act in concert (Fig. 8d). This promoting effect, greater than expected based on the

individual activities of Tat(1-86) and NC(1-55), suggests that the two full length proteins chaperone in a cooperative manner the annealing of complementary oligonucleotides, a property which is not found with the truncated proteins.

Discussion

The HIV-1 Tat protein is essential for virus replication due to its role in the strong transactivation of proviral DNA transcription, which requires the binding of Tat to the TAR hairpin at the 5′ end of the nascent viral RNA.[5] Moreover, Tat is thought to activate RTion [18–21] possibly *via* a direct interaction with reverse transcriptase and through its nucleic acid chaperoning activity.[24,29] In agreement with this, we previously found that Tat, through its basic 44-61 domain, drastically accelerates DNA-DNA annealing and DNA strand exchange, ribozyme-catalyzed cleavage of RNA and RNA transplicing.[30] The present study was aimed at characterizing the mechanism by which Tat and the basic Tat(44-61) peptide accelerate the annealing of the TAR(+) and TAR(-)DNA hairpins, a system already used to study the annealing properties of NCp7.[31,50] Moreover, since Tat may exert its chaperone properties in the presence of NCp7 during RTion, we compared the chaperone properties of the two peptides and examined their possible cooperation in the TAR (+)/TAR(-)DNA annealing reaction. Since Tat is probably present at low concentrations during RTion both in the virus and the cytosol of infected cells, the chaperone properties of Tat were examined at low peptide to TAR DNA molar ratios.

We found that binding of Tat(1-86) or Tat(44-61) to cTAR did not modify the pre-existing equilibrium between the closed and partially melted conformations of cTAR.[37,40] This inability of Tat(1-86) or Tat (44-61) to destabilize the secondary structure of cTAR differs from the NCp7-induced shift of the cTAR equilibrium towards the partly melted conformers that are competent for the fast annealing reaction with the complementary dTAR sequences.[51]

The hybridization of cTAR with dTAR in the presence of Tat(44-61) was found to proceed *via* a reversible intermediate, which rearranges into the final ED. Using series of cTAR and dTAR mutants, the intermediate was shown i) to be initiated through the bottom of the cTAR and dTAR stems, and not through the loops, ii) to rely on the complementary terminal sequences of the reactive species and iii) to be favored by the two terminal bulges in cTAR. The proposed annealing pathway differs from the one in the absence of protein which proceeds through a kissing complex intermediate,[43,52] indicating that Tat modifies the annealing mechanism. Moreover, the FRET signal of the intermediate allowed us to

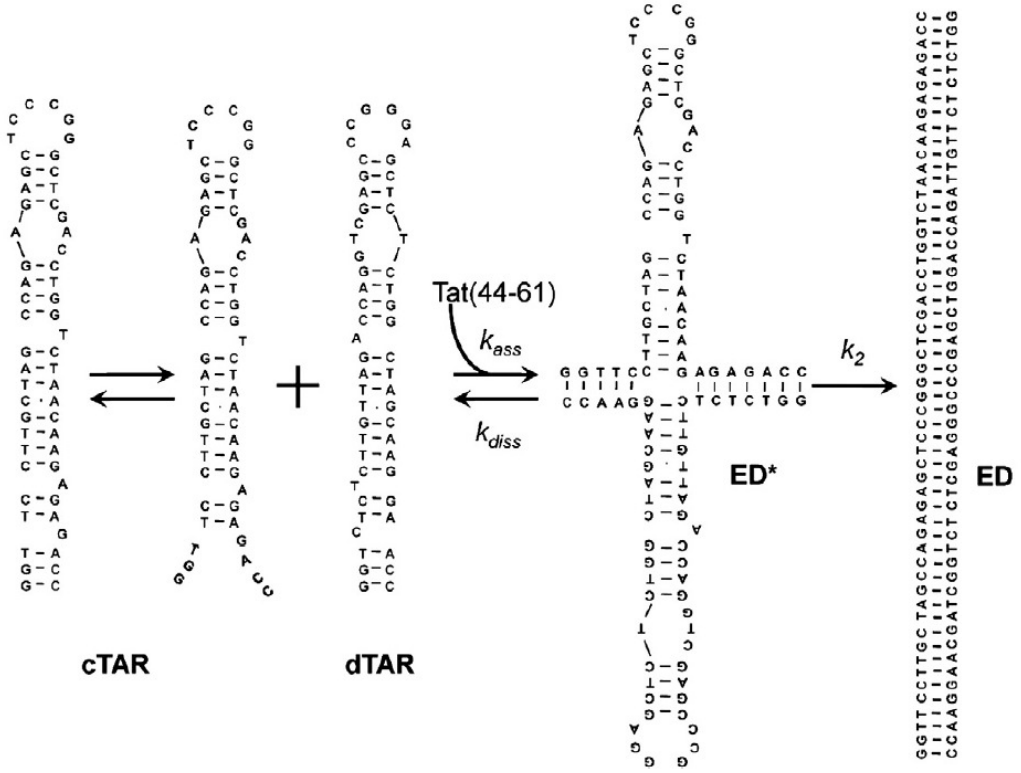


Fig. 9. Proposed reaction scheme for the Tat(44-61)-directed cTAR/dTAR annealing. In the presence of the peptide, the reactive cTAR conformer, spontaneously generated by thermal fraying, anneals to its complementary sequence through their stem termini. This leads to a reversible intermediate ED* involving 12 intermolecular base pairs (see text) which further converts into the final ED, in a step rate-limited by the melting of the more stable upper portion of the stems.

calculate a distance of 92 Å between the fluorophores at the 5′ and 3′ ends of cTAR. Assuming a width of 20 Å for a base pair, a rise per base of 3.4 Å in double strands,[53] and an additional 10 Å distance to take into account the length of each spacer linking the fluorophores to cTAR, the measured interchromophore distance was consistent with the distance of 80 Å, calculated by assuming the formation of 12 intermolecular base pairs, as a result of the annealing of the two terminal double stranded segments and bulges of cTAR and dTAR (Fig. 9). A more extended annealing including the further central double stranded region of cTAR and dTAR, would lead to a distance between the probes of 140 Å, inconsistent with our measurements. Formation of the intermediate requires the premelting of the five base pairs of the two terminal double-stranded segments of cTAR. Since the van't Hoff plot (Fig. 6) suggests a melting of only two to three base pairs for the fast kinetic component and since Tat does not exhibit an extensive nucleic acid destabilizing activity, the reactive cTAR species likely corresponds to species in which the three base pairs of the terminal double stranded segment are already melted. This melting has been shown to occur spontaneously at room temperature as a consequence of the thermally-driven fraying of cTAR termini.[37,40,51] This spontaneous fraying, facilitated by the destabilizing effect provided by the G52 bulge,[37] leads to an equilibrium between the fully closed cTAR species and a few percent of cTAR species where the terminal double strand segment is melted. Thus, formation of the intermediate only requires the additional melting of the penultimate double stranded segment of cTAR. In addition to this temperature-induced melting component, the other major contribution to the formation of the intermediate is probably provided by the Tat(44-61)-induced attraction between the Tat (44-61)/oligonucleotide complexes, which facilitates the diffusional search for the complementary sequences.[54] In a next step, the ED* intermediate is further converted to the ED, which appears to be rate-limited by the melting and associated conformational changes of the upper part of both TAR species. Comparison of the E_a values for the fast and slow kinetic components further suggests that the melting of two to three base pairs in the intermediate is critical for its conversion to the ED.

Formation of the final ED through rearrangement of a reversible intermediate was also proposed for cTAR/TAR RNA[33,43] and for mini-TAR DNA/RNA annealing[44] by NC protein from HIV and other retroviruses[55] and cTAR/dTAR annealing by the HCV core protein.[56] Regarding the mechanism of the annealing reaction, the only difference between NC and Tat(44-61) is that Tat(44-61) cannot further shift the initial pre-equilibrium between the non reactive closed cTAR species and its partly melted reactive conformer. The similarities in the mechanism of the annealing reaction of Tat(44-61) and NC (11-55) were further substantiated by the additive effect of the two peptides in the annealing reaction (Fig. 8b). From a quantitative point of view, the 6×10^4 $M^{-1}s^{-1}$ value of the rate constant (calculated from the $K_a*.k_2$ product) for the Tat(44-61)-promoted cTAR/dTAR annealing at a fractional oligonucleotide saturation of 0.3 (Table 1) was close to the 1.3×10^5 $M^{-1}s^{-1}$ value of the corresponding rate constant reported for NC(12-55) at a fractional oligonucleotide saturation of 1.[31] Thus, Tat(44-61) can promote cTAR/dTAR annealing as efficiently as NC(12-55), but at much lower saturation levels. Higher rate constants should be observed for Tat(44-61) at higher fractional saturation, but cannot be measured with our techniques, as a consequence of the strong nucleic acid aggregation observed in these conditions. The ability of Tat(44-61) to function at lower saturation level than NCp7 is further strengthened by the two orders of magnitude decrease in the overall rate constant for the NCp7-promoted cTAR/TAR RNA annealing when the fractional oligonucleotide saturation was decreased from 1 to 0.4.[43]

Interestingly, while the NC(12-55)-promoted annealing of cTARC12 to dTAR shows only one kinetic component,[31] two kinetic components were observed with Tat(44-61), suggesting that Tat(44-61) and NC(12-55) exhibit different mechanisms in this case. Since the annealing of the stem ends is impossible with cTARC12, we hypothesized that the biphasic kinetics might be assigned to two slow pathways, one through the upper part of the stems and the other one through the loops.

Tat(44-61) and NC(12-55) also differ in their ability to promote the annealing of (14-39)cTAR and (17-42) dTAR sequences, which correspond to the upper part of the cTAR and dTAR hairpins, respectively. While the NC(12-55)-promoted (14-39)cTAR/(17-42)dTAR annealing shows a two orders of magnitude decrease in the overall annealing rate as compared to cTAR/dTAR,[31] nearly identical overall rate constants were monitored for both systems with Tat(44-61) (Table 1 and 2). This difference could be related to the much higher charge density of Tat(44-61) as compared to NC(12-55), due to its stretch of eight cationic residues interspersed with just one neutral residue. Accordingly, through this positively charged stretch, Tat(44-61) can neutralize more efficiently than NC(12-55) the charges of the DNA phosphates, and thus promote more efficiently the non-specific attraction between the peptide-bound oligonucleotides.[44,54] The difference between the two peptides on the two systems is further amplified by the fact that NC(12-55) destabilizes much less efficiently the (14-39)cTAR hairpin than the cTAR termini,[38] which implies that the NC (12-55)-promoted (14-39)cTAR/(17-42)dTAR annealing[38] relies almost only on its limited nucleic acid aggregating properties.[43]

The kinetics of cTAR/dTAR annealing by Tat (1-86) at a fractional saturation of 0.15 is also biphasic and governed by an overall association rate constant of the same order as that with Tat(44-61), suggesting that both the complete protein and the peptide act according to the same mechanism, confirming that the basic (44-61) domain is the main determinant for the promotion of nucleic acid

annealing.[30] Both the efficient nucleic acid annealing activity of Tat(1-86) and its cooperative effect with NC(1-55) could well explain the stimulatory effect of Tat on RTion.[18-21] This conclusion is further supported by the key role of the basic (49-59) domain and the Tyr47 residue of the core domain[12,19–21] in both the promotion of RTion and nucleic acid annealing. In addition, the ability of Tat(1-86) to efficiently promote cTAR/dTAR annealing at a low fractional saturation is fully consistent with the probable low concentration of Tat in the virus[22] as well as in the cytosol of infected cells, during RTion.[15] The behavior of Tat differs from NCp7, which needs to saturate the oligonucleotides to a greater extent to provide optimal annealing activity[36,43,57–59] but which is present at much higher concentrations during RTion. Interestingly, both the three-fold faster nucleic acid annealing activity of Tat(1-86) as compared to Tat(44-61)[30] and the cooperative effect of Tat(1-86) but not of Tat(44-61) with NCp7 suggest that other domains of Tat play ancillary roles in its nucleic acid chaperone activity. The most likely candidate is the Cys-rich domain, which in contrast to the N-and C-terminal domains of Tat, exhibits significant nucleic acid annealing activity.[30] Our data also suggest that the Cys-rich domain could contribute to the cooperativity of Tat with NCp7 in promoting the cTAR/dTAR annealing, perhaps through a direct interaction with NCp7. This putative function of the Cys-rich domain is in line with its role in RTion stimulation and in the promotion of the first strand transfer in a reconstituted cell-free RTion system.[20,29] In contrast to other functions of Tat,[60–62] the involvement of the Cys-rich domain in the promotion of cTAR/dTAR annealing does not appear to be related to the binding of zinc to the Cys residues (Table 1). This is in line with the Tat $C_{27}S$ mutation that inactivates its transactivation activity but fully supports RTion.[9,20]

In conclusion, Tat was shown to activate through its basic 44-61 sequence and the Cys-rich domain, the annealing of cTAR with dTAR by nucleating an intermediate, which then converts into the final ED in a rate-limiting step. This mechanism was found to be analogous to that of the HIV-1 nucleocapsid protein, except that Tat only marginally destabilizes the TAR oligonucleotides and acts at much lower oligonucleotide fractional saturation. Furthermore, we showed that Tat and NCp7 can cooperatively activate the annealing of cTAR to dTAR. Such a cooperative chaperoning effect of NCp7 and Tat strongly supports an important role of Tat in the stimulation of the RTion reaction. Since the Tat(44-61) peptide mimics the full-length Tat in promoting the annealing of complementary oligonucleotides, this peptide may be used in screening assays to find molecules able to inhibit Tat.

Materials and Methods

HIV-1 Tat (Tat(1-86)), Tat(44-61) (Fig. 1a), NCp7 (NC(1-55)) and NC(11-55) (Fig. 1b) were synthesized and purified as previously described.[63,64] The proteins and peptides were stored lyophilized. Their purity was greater than 98% as judged from the HPLC elution profiles. The zinc-free form of Tat (1-86) (apo-Tat) and its zinc-bound form (holo-Tat) were prepared as previously published.[62] The zinc-bound form of NC(11-55) and NC(1-55) was prepared by reacting each peptide with a three-fold molar excess of zinc sulphate as reported.[49] The concentration of proteins and peptides was measured using an extinction coefficient at 280 nm of 1,300 M^{-1} cm^{-1}(Tat(44-61)) 8,250 M^{-1} cm^{-1} (Tat(1-86)) and 5,700 M^{-1} cm^{-1} (NC(1-55) and NC(11-55)).[30,32] Unless otherwise mentioned, the experiments were done in freshly prepared and degassed 25 mM Tris, pH 7.5 containing 30 mM NaCl and 0.2 mM $MgCl_2$. In the presence of NC(1-55), oligonucleotide annealing experiments were performed in 50 mM Tris-HCl pH 8.0 containing 50 mM KCl and 10 mM $MgCl_2$, as previously described.[49]

The oligodeoxynucleotides corresponding to HIV-1 TAR in the sense and anti-sense orientations and their various mutants (Fig. 1c) were purchased from IBA GmbH Nucleic Acids Product Supply (Göttingen, Germany). Doubly labelled TAR derivatives were prepared *via* an amino-linker with a six carbon spacer arm at their 3′ terminus with Fl or Dabcyl and at their 5′ terminus with TMR or Rh6G.[31,32] The oligonucleotides were purified by reverse-phase HPLC and polyacrylamide gel electrophoresis.

Dynamic light scattering measurements

The aggregation of oligonucleotides upon mixing with Tat(44-61) or Tat(1-86) was monitored by dynamic light scattering with a Zetasizer Nano ZS (Malvern Instruments, UK). A 400-µl aliquot of 2 µM dTAR was mixed with the same volume of a solution containing 0-16 µM of peptide prior to measurement of the particle sizes.

Fluorescence monitoring of the cTAR/dTAR annealing reaction

The kinetics of cTAR/dTAR annealing was investigated at 20 °C under pseudo first-order conditions by reacting 10 nM doubly labelled cTAR with, at least, a ten-fold higher concentration of unlabeled dTAR. In order to investigate their effect on the annealing reaction, Tat(1-86) or Tat(44-61) were added to the reaction mixtures at molar ratios of 1 and 3 peptides per oligonucleotide, respectively. The reaction was triggered by mixing 400 µl of a solution of dTAR and peptide with the same volume of a solution of TMR-5′-cTAR-3′-Fl (or Rh6G-5′-cTAR-3′-Dabcyl) and peptide, a process that avoids aggregation resulting from local high concentration of reagents. The annealing process was followed by continuously monitoring the Fl fluorescence intensity at 520 nm (with excitation at 480 nm) or the Rh6G fluorescence intensity at 555 nm (with excitation at 520 nm). All reported concentrations of the reagents correspond to those after mixing. Emission spectra and kinetics traces were recorded with a FluoroMax 3 spectrofluorimeter (Jobin Yvon Instruments) equipped with a temperature-controlled cell holder. All fluorescence intensities were corrected for buffer emission and lamp fluctuations. Non-linear least square fits were calculated with the Levenberg-Marquardt algorithm using the Origin software (Microcal). Numerical resolution of the kinetic parameters was performed by fitting simultaneously the set of progress curves with the Dynafit software (BioKin Ltd.). Numerical tests (Kolmogorov,

Durbin-Watson, Tukey statistics) and the reduced $\chi 2$ values were used to evaluate goodness of fit.

Acknowledgements

This work, R. S. and K. K. S. were supported by grants from the French Agency for Research on AIDS and viral hepatitis (ANRS).

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2010.05.033

References

1. Rana, T. M. & Jeang, K. T. (1999). Biochemical and functional interactions between HIV-1 Tat protein and TAR RNA. *Arch. Biochem. Biophys.* **365**, 175–185.
2. Weeks, K. M., Ampe, C., Schultz, S. C., Steitz, T. A. & Crothers, D. M. (1990). Fragments of the HIV-1 Tat protein specifically bind TAR RNA. *Science*, **249**, 1281–1285.
3. Cullen, B. R. (1991). Regulation of HIV-1 gene expression. *Faseb J.* **5**, 2361–2368.
4. Jeang, K. T., Xiao, H. & Rich, E. A. (1999). Multifaceted activities of the HIV-1 transactivator of transcription, Tat. *J. Biol. Chem.* **274**, 28837–28840.
5. Gatignol, A. (2007). Transcription of HIV: Tat and cellular chromatin *Advances in Pharmacology* pp. 137–159, second edn. San Diego, CA.
6. Shojania, S. & O'Neil, J. D. (2006). HIV-1 Tat is a natively unfolded protein: the solution conformation and dynamics of reduced HIV-1 Tat-(1-72) by NMR spectroscopy. *J. Biol. Chem.* **281**, 8347–8356.
7. Ivanyi-Nagy, R., Davidovic, L., Khandjian, E. W. & Darlix, J. L. (2005). Disordered RNA chaperone proteins: from functions to disease. *Cell Mol. Life Sci.* **62**, 1409–1417.
8. Veschambre, P., Roisin, A. & Jalinot, P. (1997). Biochemical and functional interaction of the human immunodeficiency virus type 1 Tat transactivator with the general transcription factor TFIIB. *J. Gen. Virol.* **78**(Pt 9), 2235–2245.
9. Kashanchi, F., Piras, G., Radonovich, M. F., Duvall, J. F., Fattaey, A., Chiang, C. M. *et al.* (1994). Direct interaction of human TFIID with the HIV-1 transactivator tat. *Nature*, **367**, 295–299.
10. Marzio, G., Tyagi, M., Gutierrez, M. I. & Giacca, M. (1998). HIV-1 tat transactivator recruits p300 and CREB-binding protein histone acetyltransferases to the viral promoter. *Proc. Natl Acad. Sci. USA*, **95**, 13519–13524.
11. Parada, C. A. & Roeder, R. G. (1996). Enhanced processivity of RNA polymerase II triggered by Tat-induced phosphorylation of its carboxy-terminal domain. *Nature*, **384**, 375–378.
12. Meredith, L. W., Sivakumaran, H., Major, L., Suhrbier, A. & Harrich, D. (2009). Potent inhibition of HIV-1 replication by a Tat mutant. *PLoS One*, **4**, e7769.
13. Chiu, Y. L., Ho, C. K., Saha, N., Schwer, B., Shuman, S. & Rana, T. M. (2002). Tat stimulates cotranscriptional capping of HIV mRNA. *Molec. Cell*, **10**, 585–597.
14. Berro, R., Kehn, K., de la Fuente, C., Pumfery, A., Adair, R., Wade, J. *et al.* (2006). Acetylated Tat regulates human immunodeficiency virus type 1 splicing through its interaction with the splicing regulator p32. *J. Virol.* **80**, 3189–3204.
15. Charnay, N., Ivanyi-Nagy, R., Soto-Rifo, R., Ohlmann, T., Lopez-Lastra, M. & Darlix, J. L. (2009). Mechanism of HIV-1 Tat RNA translation and its activation by the Tat protein. *Retrovirology*, **6**, 74.
16. Bennasser, Y. & Jeang, K. T. (2006). HIV-1 Tat interaction with Dicer: requirement for RNA. *Retrovirology*, **3**, 95.
17. Bennasser, Y., Le, S. Y., Benkirane, M. & Jeang, K. T. (2005). Evidence that HIV-1 encodes an siRNA and a suppressor of RNA silencing. *Immunity*, **22**, 607–619.
18. Harrich, D., Ulich, C., Garcia-Martinez, L. F. & Gaynor, R. B. (1997). Tat is required for efficient HIV-1 reverse transcription. *EMBO J.* **16**, 1224–1235.
19. Ulich, C., Dunne, A., Parry, E., Hooker, C. W., Gaynor, R. B. & Harrich, D. (1999). Functional domains of Tat required for efficient human immunodeficiency virus type 1 reverse transcription. *J. Virol.* **73**, 2499–2508.
20. Apolloni, A., Meredith, L. W., Suhrbier, A., Kiernan, R. & Harrich, D. (2007). The HIV-1 Tat protein stimulates reverse transcription in vitro. *Curr. HIV Res.* **5**, 473–483.
21. Apolloni, A., Hooker, C. W., Mak, J. & Harrich, D. (2003). Human immunodeficiency virus type 1 protease regulation of tat activity is essential for efficient reverse transcription and replication. *J. Virol.* **77**, 9912–9921.
22. Chertova, E., Chertov, O., Coren, L. V., Roser, J. D., Trubey, C. M., Bess, J. W., Jr. *et al.* (2006). Proteomic and biochemical analysis of purified human immunodeficiency virus type 1 produced from infected monocyte-derived macrophages. *J. Virol.* **80**, 9039–9052.
23. Ensoli, B., Buonaguro, L., Barillari, G., Fiorelli, V., Gendelman, R., Morgan, R. A. *et al.* (1993). Release, uptake, and effects of extracellular human immunodeficiency virus type 1 Tat protein on cell growth and viral transactivation. *J. Virol.* **67**, 277–287.
24. Kameoka, M., Morgan, M., Binette, M., Russell, R. S., Rong, L., Guo, X. *et al.* (2002). The Tat protein of human immunodeficiency virus type 1 (HIV-1) can promote placement of tRNA primer onto viral RNA and suppress later DNA polymerization in HIV-1 reverse transcription. *J. Virol.* **76**, 3637–3645.
25. Lori, F., di Marzo Veronese, F., de Vico, A. L., Lusso, P., Reitz, M. S., Jr. & Gallo, R. C. (1992). Viral DNA carried by human immunodeficiency virus type 1 virions. *J. Virol.* **66**, 5067–5074.
26. Zhang, H., Dornadula, G. & Pomerantz, R. J. (1998). Natural endogenous reverse transcription of HIV-1. *J. Reprod. Immunol.* **41**, 255–260.
27. Zhang, H., Dornadula, G. & Pomerantz, R. J. (1996). Endogenous reverse transcription of human immunodeficiency virus type 1 in physiological microenviroments: an important stage for viral infection of nondividing cells. *J. Virol.* **70**, 2809–2824.
28. Kameoka, M., Rong, L., Gotte, M., Liang, C., Russell, R. S. & Wainberg, M. A. (2001). Role for human immunodeficiency virus type 1 Tat protein in suppression of viral reverse transcriptase activity during late stages of viral replication. *J. Virol.* **75**, 2675–2683.
29. Guo, X., Kameoka, M., Wei, X., Roques, B., Gotte, M., Liang, C. *et al.* (2003). Suppression of an intrinsic

strand transfer activity of HIV-1 Tat protein by its second-exon sequences. *Virology*, **307**, 154–163.
30. Kuciak, M., Gabus, C., Ivanyi-Nagy, R., Semrad, K., Storchak, R., Chaloin, O. *et al.* (2008). The HIV-1 transcriptional activator Tat has potent nucleic acid chaperoning activities in vitro. *Nucleic Acids Res.* **36**, 3389–3400.
31. Godet, J., de Rocquigny, H., Raja, C., Glasser, N., Ficheux, D., Darlix, J. L. *et al.* (2006). During the early phase of HIV-1 DNA synthesis, nucleocapsid protein directs hybridization of the TAR complementary sequences via the ends of their double-stranded stem. *J. Mol. Biol.* **356**, 1180–1192.
32. Ramalanjaona, N., de Rocquigny, H., Millet, A., Ficheux, D., Darlix, J. L. & Mely, Y. (2007). Investigating the mechanism of the nucleocapsid protein chaperoning of the second strand transfer during HIV-1 DNA synthesis. *J. Mol. Biol.* **374**, 1041–1053.
33. Liu, H. W., Zeng, Y., Landes, C. F., Kim, Y. J., Zhu, Y., Ma, X. *et al.* (2007). Insights on the role of nucleic acid/protein interactions in chaperoned nucleic acid rearrangements of HIV-1 reverse transcription. *Proc. Natl Acad. Sci. USA*, **104**, 5261–5267.
34. Ivanyi-Nagy, R., Lavergne, J. P., Gabus, C., Ficheux, D. & Darlix, J. L. (2008). RNA chaperoning and intrinsic disorder in the core proteins of Flaviviridae. *Nucleic Acids Res.* **36**, 712–725.
35. Calnan, B. J., Biancalana, S., Hudson, D. & Frankel, A. D. (1991). Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA-protein recognition. *Genes Dev.* **5**, 201–210.
36. Levin, J. G., Guo, J., Rouzina, I. & Musier-Forsyth, K. (2005). Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Prog. Nucleic Acid Res. Mol. Biol.* **80**, 217–286.
37. Beltz, H., Azoulay, J., Bernacchi, S., Clamme, J. P., Ficheux, D., Roques, B. *et al.* (2003). Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the nucleocapsid protein NCp7. *J. Mol. Biol.* **328**, 95–108.
38. Beltz, H., Piemont, E., Schaub, E., Ficheux, D., Roques, B., Darlix, J. L. *et al.* (2004). Role of the structure of the top half of HIV-1 cTAR DNA on the nucleic acid destabilizing activity of the nucleocapsid protein NCp7. *J. Mol. Biol.* **338**, 711–723.
39. Beltz, H., Clauss, C., Piemont, E., Ficheux, D., Gorelick, R. J., Roques, B. *et al.* (2005). Structural determinants of HIV-1 nucleocapsid protein for cTAR DNA binding and destabilization, and correlation with inhibition of self-primed DNA synthesis. *J. Mol. Biol.* **348**, 1113–1126.
40. Bernacchi, S., Stoylov, S., Piemont, E., Ficheux, D., Roques, B. P., Darlix, J. L. *et al.* (2002). HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. *J. Mol. Biol.* **317**, 385–399.
41. Bernacchi, S. & Mely, Y. (2001). Exciton interaction in molecular beacons: a sensitive sensor for short range modifications of the nucleic acid structure. *Nucleic Acids Res.* **29**, E62–62.
42. Bernacchi, S., Piemont, E., Potier, N., Dorsselaer, A. & Mely, Y. (2003). Excitonic heterodimer formation in an HIV-1 oligonucleotide labeled with a donor-acceptor pair used for fluorescence resonance energy transfer. *Biophys. J.* **84**, 643–654.
43. Vo, M. N., Barany, G., Rouzina, I. & Musier-Forsyth, K. (2009). HIV-1 nucleocapsid protein switches the pathway of transactivation response element RNA/DNA annealing from loop-loop "kissing" to "zipper". *J. Mol. Biol.* **386**, 789–801.
44. Vo, M. N., Barany, G., Rouzina, I. & Musier-Forsyth, K. (2006). Mechanistic studies of mini-TAR RNA/DNA annealing in the absence and presence of HIV-1 nucleocapsid protein. *J. Mol. Biol.* **363**, 244–261.
45. Kuzmic, P. (1996). Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.* **237**, 260–273.
46. Rouzina, I. & Bloomfield, V. A. (1999). Heat capacity effects on the melting of DNA. 1. General aspects. *Biophys. J.* **77**, 3242–3251.
47. Cantor, C. & Schimmel, P (1980). *Biophysical Chemistry Part 2: Techniques for the Study of Biological Structure and Function.* Academic Press, NY.
48. Zargarian, L., Kanevsky, I., Bazzi, A., Boynard, J., Chaminade, F., Fosse, P. *et al.* (2009). Structural and dynamic characterization of the upper part of the HIV-1 cTAR DNA hairpin. *Nucleic Acids Res.* **37**, 4043–4054.
49. Grohmann, D., Godet, J., Mely, Y., Darlix, J. L. & Restle, T. (2008). HIV-1 nucleocapsid traps reverse transcriptase on nucleic acid substrates. *Biochemistry*, **47**, 12230–12240.
50. Liu, H. W., Cosa, G., Landes, C. F., Zeng, Y., Kovaleski, B. J., Mullen, D. G. *et al.* (2005). Single-molecule FRET studies of important intermediates in the nucleocapsid-protein-chaperoned minus-strand transfer step in HIV-1 reverse transcription. *Biophys. J.* **89**, 3470–3479.
51. Azoulay, J., Clamme, J. P., Darlix, J. L., Roques, B. P. & Mely, Y. (2003). Destabilization of the HIV-1 complementary sequence of TAR by the nucleocapsid protein through activation of conformational fluctuations. *J. Mol. Biol.* **326**, 691–700.
52. Kanevsky, I., Chaminade, F., Ficheux, D., Moumen, A., Gorelick, R., Negroni, M. *et al.* (2005). Specific interactions between HIV-1 nucleocapsid protein and the TAR element. *J. Mol. Biol.* **348**, 1059–1077.
53. Bloomfield, V. A., Crothers, D. M., Tinoco, I. & Hearst, J. E. (Eds.). (2000). Properties and Functions, Sausalito, CA.
54. Vo, M. N., Barany, G., Rouzina, I. & Musier-Forsyth, K. (2009). Effect of Mg(2+) and Na(+) on the nucleic acid chaperone activity of HIV-1 nucleocapsid protein: implications for reverse transcription. *J. Mol. Biol.* **386**, 773–788.
55. Stewart-Maynard, K. M., Cruceanu, M., Wang, F., Vo, M. N., Gorelick, R. J., Williams, M. C. *et al.* (2008). Retroviral nucleocapsid proteins display nonequivalent levels of nucleic acid chaperone activity. *J. Virol.* **82**, 10129–10142.
56. Sharma K. K., Didier P., Darlix J. L., de Rocquigny H., Bensikkadour H., Lavergne J. P., et al. (2010). Kinetic analysis of the nucleic acid chaperone activity of the Hepatitis C virus core protein. Nucleic acids research. in press.
57. Darlix, J. L., Lapadat-Tapolsky, M., de Rocquigny, H. & Roques, B. P. (1995). First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. *J. Mol. Biol.* **254**, 523–537.
58. Rein, A., Henderson, L. E. & Levin, J. G. (1998). Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem. Sci.* **23**, 297–301.
59. Williams, M. C., Gorelick, R. J. & Musier-Forsyth, K. (2002). Specific zinc-finger architecture required for

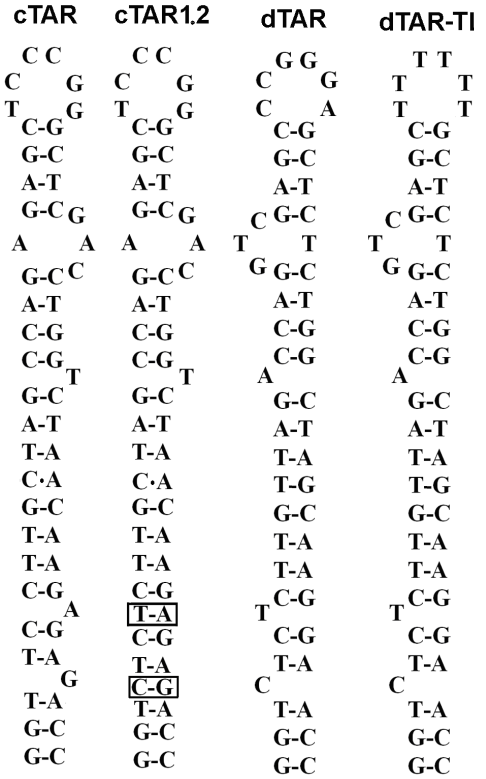
HIV-1 nucleocapsid protein's nucleic acid chaperone function. *Proc. Natl Acad Sci. USA*, **99**, 8614–8619.
60. Garber, M. E., Wei, P., KewalRamani, V. N., Mayall, T. P., Herrmann, C. H., Rice, A. P. *et al.* (1998). The interaction between HIV-1 Tat and human cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein. *Genes Dev.* **12**, 3512–3527.
61. Misumi, S., Takamune, N., Ohtsubo, Y., Wanigushi, K. & Shoji, S. (2004). Zn2+ binding to cystein-rich domain of extracellular human immunodeficiency virus type I tat protein is associated with Tat protein-induced apoptosis. *AIDS Res. Hum. Retroviruses*, **20**, 297–304.
62. Egele, C., Barbier, P., Didier, P., Piemont, E., Allegro, D., Chaloin, O. *et al.* (2008). Modulation of microtubule assembly by the HIV-1 Tat protein is strongly dependent on zinc binding to Tat. *Retrovirology*, **5**, 62–74.
63. Chaloin, O., Peter, J. C., Briand, J. P., Masquida, B., Desgranges, C., Muller, S. *et al.* (2005). The N-terminus of HIV-1 Tat protein is essential for Tat-TAR RNA interaction. *Cell Mol. Life Sci.* **62**, 355–361.
64. de Rocquigny, H., Ficheux, D., Gabus, C., Fournie-Zaluski, M. C., Darlix, J. L. & Roques, B. P. (1991). First large scale chemical synthesis of the 72 amino acid HIV-1 nucleocapsid protein NCp7 in an active form. *Biochem. Biophys. Res. Commun.* **180**, 1010–1018.
65. Baudin, F., Marquet, R., Isel, C., Darlix, J. L., Ehresmann, B. & Ehresmann, C. (1993). Functional sites in the 5′ region of human immunodeficiency virus type 1 RNA form defined structural domains. *J. Mol. Biol.* **229**, 382–397.
66. SantaLucia, J., Jr., Allawi, H. T. & Seneviratne, P. A. (1996). Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, **35**, 3555–3562.

Chapter 4: Kinetic investigation of model DNA annealing in the presence of Bunyamwera Orthobunyavirus Nucleocapsid Protein (BUNV-N)

The *Bunyaviridae* are the largest family of negative-strand viruses and contains more than 350 named members. *Bunyaviruses* are characterized by a tripartite RNA genome. The family is divided into five genera (*Orthobunyavirus*, *Hantavirus*, *Nairovirus*, *Plebovirus* and *Tospovirus*), the members of which are distinguished by biochemical and serological criteria. All viruses share the same four structural proteins, L (*RNA-dependent RNA polymerase*), Gn and Gc (*envelope glycoproteins*), and N (*Nucleocapsid protein*). The N protein varies in size from about 25 kDa (*orthobunyaviruses*) to 50 kDa (*hantaviruses* and *nairoviruses*) (Elliott RM et al., 2005, 1996, Nichol ST et al., 2005, Schmaljohn CS et al., 2001).

Bunyamwera virus (BUNV) is the prototype of both the *Orthobunyavirus* genus and the family as a whole and its N protein (BUNV-N) is consisting of 233 amino acids (aa). BUNV-N is a multifunctional protein that encapsidates negative-sense genome of virus, to form ribonucleoprotein complexes that are the functional templates for viral transcription and replication. In addition, BUNV-N protein molecules interact with themselves to form oligomers, with the viral L (RNA polymerase) protein, with the carboxy-terminal regions of either or both of the virion glycoproteins, and probably also with host cell proteins. However, it has been recently shown that N protein of one of the members of family, named *Hantavirus Sin Nombre* (SNV), shows RNA chaperone activity (Mir and panganiban, 2006). With the analogy within the family, we characterized the chaperone properties of a BUNV-N protein by monitoring the promotion of the annealing of dTAR with cTAR DNA from HIV-1.

(a) Oligonucleotide Sequences **(b) BUNV-N sequence**



M_1IELEFHDVA ANTSSTFDPE VAYANFKRVH
TTGLSYDHIR IFYIKGREIK TSLAKRSEWE
VTLNLGGWK ITVYNTNFP GNRNNPVPDD
GLTLHRLSGF LARYLLEKML KVSEPEKLI
IKSKIINPLA EKNGITWND GEEVYLSFFP
GSEMFLGTFR FYPLAIGIYK VQRKEMEPKY
LEKTMRQRYM GLEAATWTVS LTEVQSALT
VVSSLGWKKT NVSAAARDFL AKFGINM$_{233}$

Figure1. Oligonucleotides and protein used in this study: (a) Structure of the oligonucleotides and (b) amino acid sequence of BUNV-N protein. The cTAR and dTAR sequences are from the HIV-1 MAL strain. The secondary structures of the oligonucleotides were predicted from the structure of TAR and the mfold program (http://mfold.rna.albany.edu/?q=mfold/DNA-Folding-Form).

Destabilisation

To characterize the nucleic acid destabilizing properties of the BUNV-N protein, cTAR labelled at its 5' and 3' ends either by Rh6G or fluorescein and DABCYL or Rhodamine, as fluorophore and quencher respectively were used. The dyes form a nonfluorescent heterodimer when the cTAR stem is closed, while melting of the stem restores the fluorophore fluorescence. Thus, the destabilizing ability of BUNV-N protein can be evaluated from the ratio of the fluorescence intensity in the presence versus the absence of the peptide.

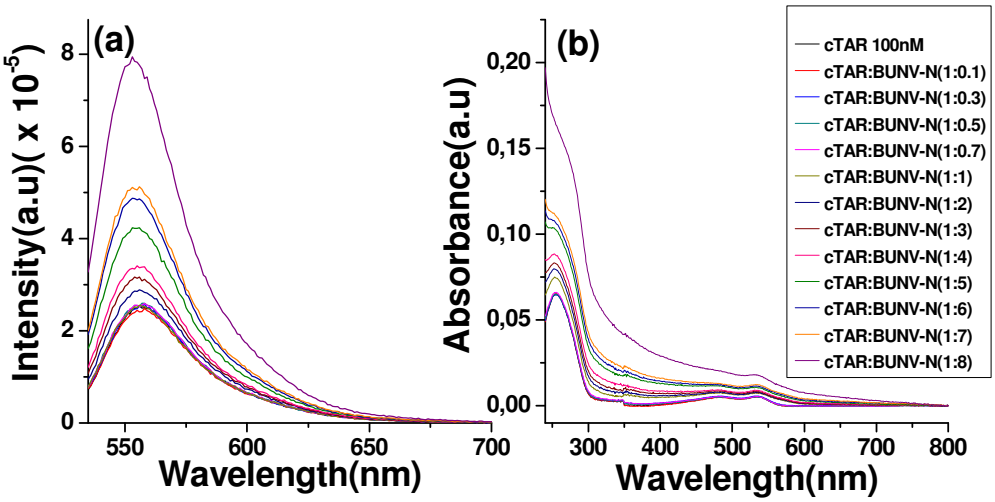


Figure 2: Destabilisation of cTAR secondary structure by BUNV-N protein.
(a) Emission and (b) Absorbance spectra of Rh6g-cTAR-Dabcyl (100 nM) were recorded in the absence (black) and presence of BUNV-N protein at different concentrations in 25 mM Tris, 30 mM NaCl, and 0.2 mM $MgCl_2$, pH 7.5. Excitation wavelength was 520 nm.

A continuous increase in the fluorescence intensity of the Rh6G in the presence of BUNV-N protein was observed (Figure 2a). This increase was visible only when, the presence of BUNV-N was in excess of peptide/cTAR ratio of 1:1. This suggests that there is no melting of cTAR stem in the presence of the peptide till peptide/cTAR ratio of 1:1. However, we observed a systematic increase in fluorescence intensity of Rh6G upon addition of BUNV-N over ratio 1:1, suggesting that BUNV-N protein destabilizes cTAR structure in these conditions. However, the corresponding absorbance spectra showed a sharp increase in light scattering, in line with an BUNV-N protein-induced aggregation of cTAR at ratios >1:1 (Figure 2b).

Though a clear increase in the Rh6G fluorescence suggested a destabilization of cTAR by BUNV-N at peptide/cTAR ratio > 1 (Fig. 2a), this conclusion is questioned by the BUNV-N protein-induced aggregation of cTAR at the same ratios (Fig. 2b). Thus, it may be possible that an opening of cTAR stem only appears in aggregates.

Moreover, due to this BUNV-N-induced aggregation of nucleic acids, it was not possible to determine the binding constant of the peptide to the ODN, by fluorescence techniques.

Investigating the interaction of cTAR DNA with the BUNV-N protein by FCS

To characterize in depth the nucleic acid chaperone properties of the BUNV-N protein, we first had to find experimental conditions where the core peptides did not cause DNA

aggregation in a manner similar to other nucleic acid chaperones (Egele C et al., 2007, Liu HW et al., 2005, Cristofari G et al., 2002, Stoylov SP et al., 1997)), since aggregation can cause strong bias when using fluorescence-based techniques (Godet J et al., 2006, Egele C et al., 2007). Since nucleic acid aggregation by positively charged peptides is concentration dependent (Stoylov SP et al., 1997, Bloomfield VA et al., 1991), we investigated this dependence by means of FCS. Assuming that BUNV-N protein molecules diffuse freely in a Gaussian excitation volume, the normalized autocorrelation function, G(τ), calculated from the fluorescence fluctuations was fitted according to (Thompson NL 1991):

$$G(\tau) = \frac{1}{N}\left(1+\frac{\tau}{\tau_{da}}\right)^{-1}\left(1+\frac{1}{S^2}\frac{\tau}{\tau_{da}}\right)^{\frac{-1}{2}} \quad \textbf{Equation (1)}$$

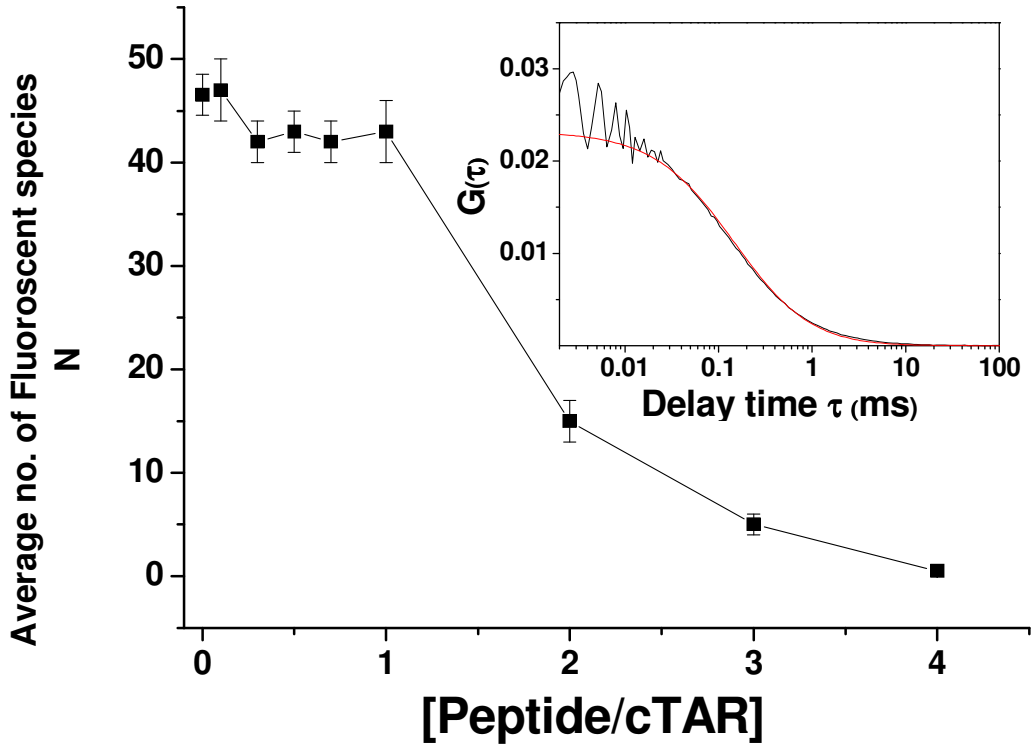


Figure 3: Evidence by fluorescence correlation spectroscopy of BUNV-N-induced aggregation of TMR-5′-cTAR. The average number of fluorescent species, N, for 100 nM (black squares) of TMR-5′-cTAR within the excitation volume was obtained by fitting the autocorrelation curves with Equation (1). (Inset) Autocorrelation curve of 100 nM TMR-5′-cTAR (black line) and its fit (red line) with Equation (1) in the presence of 100 nM of BUNV-N protein.

Aggregation of the TMR-labeled 5′-cTAR molecules by the core peptides is expected to decrease the number of fluorescent species (Sharma KK et al., 2010). In the absence of

peptide, the number of fluorescent TAR molecules in the excitation volume was fully consistent with the theoretical number of molecules expected from their concentration. By adding increasing concentrations of BUNV-N protein, we found no change in the number of fluorescent species up to a peptide/ODN molar ratio of 1:1, indicating that no aggregation occurred under these conditions (Figure 3), in line with the absorption data (Fig. 2b). In contrast, aggregation took place at higher ratios (2:1, 3:1 and 4:1) as evidenced by the sharp drop in the number of fluorescent species.

On the basis of these results, we selected a peptide/ODN molar ratio of 1:1 to characterize the chaperone properties of the core peptides.

Kinetics of cTAR-dTAR annealing in the presence of BUNV-N protein

The real-time annealing kinetics of cTAR with dTAR, was investigated by mixing TMR-5'-cTAR-3'-Fl with an excess of nonlabeled dTAR. Formation of the 55 bp cTAR/dTAR extended duplex (ED) strongly increases the interchromophore distance, leading to a full recovery of Fl emission. In the absence of BUNV-N protein, the annealing of cTAR with dTAR was very slow, involving two distinct kinetic components with second-order rate constants. In the presence of the peptide added at a peptide/ODN ratio of 1/1, the same fluorescence plateau was observed as in the absence of peptide, indicating that ED formation went to completion. Moreover, BUNV-N protein strongly accelerates the annealing reaction, since the reaction was completed in about 30 min (Figure 4), instead of more than 1 day in the absence of peptide. An adequate fit of the annealing kinetic traces was obtained using a bi-exponential function:

$$I(t) = I_f - (I_f - I_{ONC})x\, e^{(-k_{obs1}(t-t0))} - (I_f - I_{ONC})(1-x)e^{(k_{obs2}(t-t0))} \qquad \textbf{Equation (2)}$$

Where, t_0 is the dead time, $k_{obs1,2}$ are the observed kinetic rate constants, x is the amplitude of the fast component, and I_{0NC} and I_f are the fluorescence intensities of the SL and the ED, respectively.

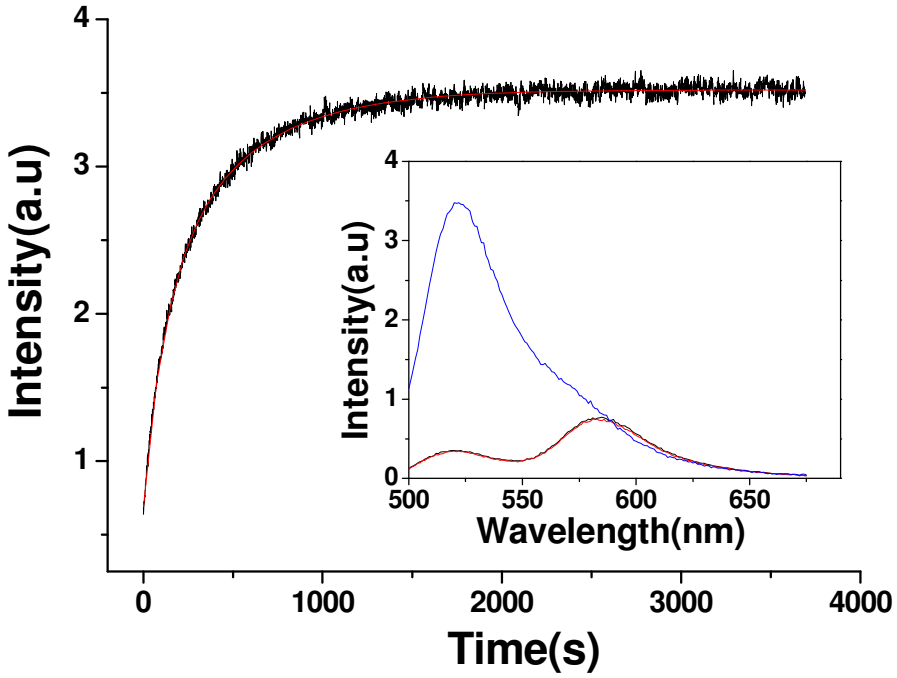


Figure 4: Kinetics of cTAR/dTAR annealing promoted by the BUNV-N protein. Kinetic trace of 10 nM doubly labelled cTAR reacted with 700 nM dTAR in the presence of BUNV-N added at a peptide/ODN molar ratio of 1:1. The continuous red line describes the best fit of the data according to equation (2) with k_{obs1} = 8.3 × 10^{-3} s^{-1}, k_{obs2} = 2.12 × 10^{-3} s^{-1} and a = 0.5. Inset: emission spectra of the doubly labelled cTAR (black line) in the presence of BUNV-N added at a peptide/ODN ratio of 1:1 before (red line) and after completion of the annealing reaction with dTAR (blue).

We already observe full fluorescence increase at the lowest tested oligonucleotide concentrations (10 nM cTAR and 300 nM dTAR), suggesting that the peptide likely binds to the oligonucleotides with high affinity (> 10^7 M^{-1}). Indeed, if the binding constant would be lower than 10^6 M^{-1}, and if we assume that as for NCp7 or other chaperones, BUNV-N does not show any catalytic activity, most of the oligonucleotides would be free and anneal at the low speed of cTAR and dTAR in the absence of protein.

Furthermore, experiments with various dTAR concentrations indicated linear and saturation behaviour for k_{obs1} and k_{obs2} values, respectively (Figure 5). This behaviour of k_{obs} values is consistent with a two-step reaction scheme, as described for the NCp7 and Tat promoted cTAR/dTAR annealing under saturating conditions (Vo et al, 2009; Boudier C et al., 2010), where a fast pre-equilibrium intermediate, IC, precedes the formation of the final stable ED through a monomolecular reaction (Vo, M.N., et al. 2009).

$$\text{cTAR+dTAR} \underset{k_{diss}}{\overset{k_{ass}}{\rightleftharpoons}} \text{IC} \underset{k_b}{\overset{k_f}{\rightleftharpoons}} \text{ED}$$

Scheme (1)

Formation of IC is governed by the second order association rate constant, k_{ass}, and the first order dissociation rate constant, k_{diss}, while the interconversion of IC into ED is governed by the forward and backward interconversion rate constants, k_f and k_b.

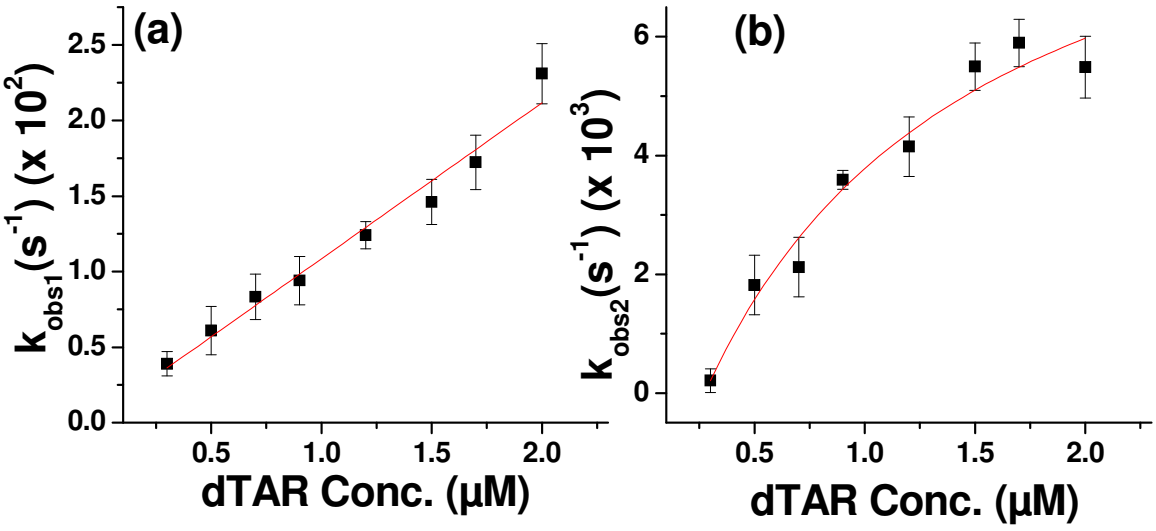


Figure 5: Kinetic parameters of the BUNV-N protein- promoted-cTAR/dTAR annealing. The fast (k_{obs1}) (a) and slow (k_{obs2}) (b) components were determined in pseudo-first order conditions from real-time kinetics. The solid line corresponds to the fit of the k_{obs1} and k_{obs2} data with equation (3) and equation (4) respectively.

The hyperbolic dependence of k_{obs2} may be ascribed to IC accumulation as a consequence of its slow conversion into ED which likely constitutes the rate-limiting step of the annealing pathway.

If the pre-equilibrium is rapidly reached, scheme (1) predicts a linear variation of k_{obs1} as described by equation (3):

$$k_{obs1} = k_{ass}[\text{dTAR}] + k_{diss} \qquad \textbf{Equation (3)}$$

where, k_{ass} is the second order association rate constant and k_{diss} is the first order dissociation rate constant. On the other hand, the hyperbolic variation of k_{obs2} can be described by equation (4).

$$k_{obs} = \frac{k_f K_M [dTAR]}{1 + K_M [dTAR]} + k_b \qquad \textbf{Equation (4)}$$

By fitting data with eqn (3), from the slope and the intercept with the Y axis, we found k_{ass} = 1.03×10^4 $M^{-1}s^{-1}$ and $k_{diss} = 5.3 \times 10^{-3}$ s^{-1} for cTAR/dTAR annealing. Non-linear regression of the k_{obs2} data (Figure3) with eqn (4) yielded $k_f = 1.32 \times 10^{-2}$ s^{-1} and $K_M = 1.02 \times 10^6$ M^{-1}. The value of k_b was very low (<0.005), suggesting that BUNV-N protein was not able to destabilize ED formation.

The binding constant of the intermediate is about two orders of magnitude lower than that with NC(1-55), added at a ratio of 10 peptides per oligonucleotide ($K=10^8$ M^{-1}) (Vo et al, 2009). This difference can be explained on the basis of the ODN coating level by BUNV-N protein. Since the binding constant of the intermediate was shown to be strongly dependent on the level of protein coating in the case of NCp7 (with an about 3 orders of magnitude difference between low and full coating). Interestingly, the comparison of k_f values suggests that the conversion of the intermediate to the final extended duplex is promoted by the BUNV-N protein with similar efficiency than for NCp7.

Furthermore, comparison of K_M and k_f values of the BUNV-N protein-promoted-cTAR/dTAR, shows a resemblance with the kinetic values of the pathways defined for HCV core-promoted-cTAR/dTAR annealing (Sharma KK et al., 2010). In addition, a one order of magnitude difference only in k_{ass} values were observed, when compared to Tat(44-61)-promoted-cTAR/dTAR annealing, added at a ratio of 3:1 (Boudier C et al., 2010). This promptly suggests the resemblance of the annealing reaction mechanism in presence of BUNV-N protein with HCV core and HIV-1 Tat(44-61) protein. Moreover, the fast kinetic pathway values of (23-110)HuPrp-promoted cTAR/dTAR annealing were also similar as compared to those of BUNV-N. Taken together, these data suggest that the kinetics intermediates promoted by HCV Core, HIV-Tat(44-61), (23-110)HuPrp and BUNV-N exhibit similar equilibrium constants and a similar number of base pairs between cTAR and dTAR.

The resemblance of kinetic pathway of BUNV-N promoted cTAR/dTAR annealing to that of other above mentioned proteins, can further be strengthened by the nucleation of annealing reaction through the stem termini of cTAR and dTAR. For this, the effect of the oligonucleotide sequence and their stability on the BUNV-N protein-promoted cTAR/dTAR annealing kinetics was analysed. First, dTAR was substituted by the dTAR T-L mutant where the 6 nucleotides of the loop were changed to T residues, thus preventing its base-pairing with the cTAR loop. These nucleotide substitutions did not significantly change the time course of ED formation (Figure 7), indicating that loop-loop interactions do not play a significant role in the BUNV-N protein-promoted cTAR/dTAR annealing reaction.

Secondly, cTAR was substituted with the cTAR1.2 derivative where bases complementary to the bulged bases at positions 49 and 52 have been introduced in order to stabilize the lower half of the stem. The annealing of this mutant to dTAR in the presence of BUNV-N protein was extremely slow (Figure 7).

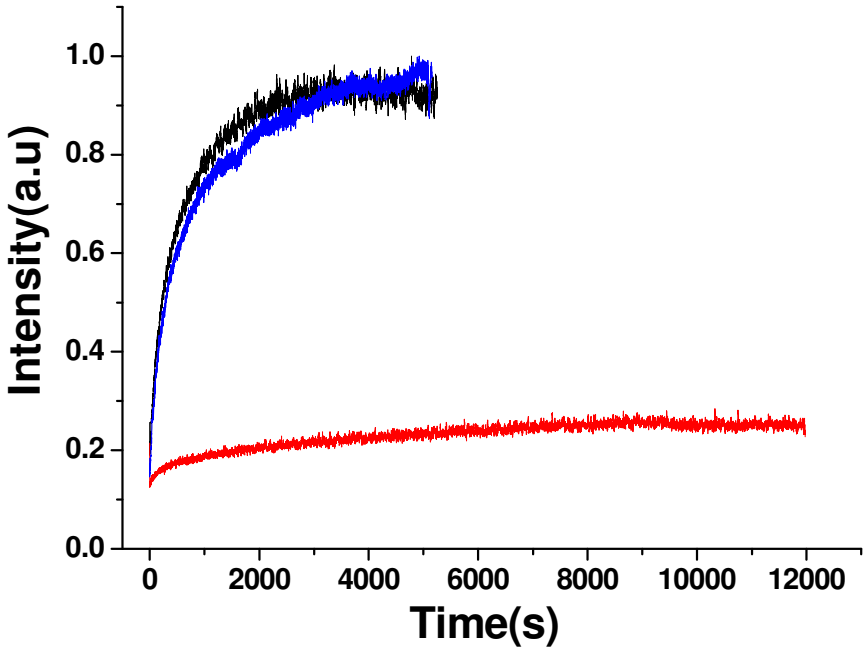


Figure 7: Kinetics of BUNV-N protein-promoted annealing of cTAR derivatives with dTAR derivatives. Kinetic traces of 10 nM doubly labelled TMR-cTAR-Fl derivatives with 500 nM non-labelled dTAR derivatives [cTAR/dTAR (blue), cTAR/dTAR-TL (black)] and 10 nM cTAR1.2 with 500 nM dTAR (red). BUNV-N protein was added at a peptide/oligonucleotide ratio of 1:1 Excitation and emission wavelengths were 480 nm and 520nm respectively.

These results proved that the kinetic pathway of the BUNV-N protein-promoted cTAR/dTAR annealing reaction is nucleated through the stems of cTAR and dTAR.

To get further insight into the reaction pathway, temperature dependence of BUNV-N promoted cTAR/dTAR annealing was performed. The obtained results were fitted by using the Arrhenius equation:

$$k_i = A_i \exp\left(E_{a,i} / RT\right) \quad \textbf{Equation (5)}$$

where, the rate constant k_i is given by k_{obsi}/[dTAR], A_i is the pre-exponential Arrhenius factor, $E_{a,i}$ is the activation energy, R is the universal gas constant, and T is the temperature (in Kelvin).

Both reaction rates increased with increase in temperature (Figure 6). Interestingly, no change was observed for amplitude of the fast component till 35°C (Figure 6). At 40°C, a mono-exponential increase was observed for BUNV-N-promoted cTAR/dTAR annealing.

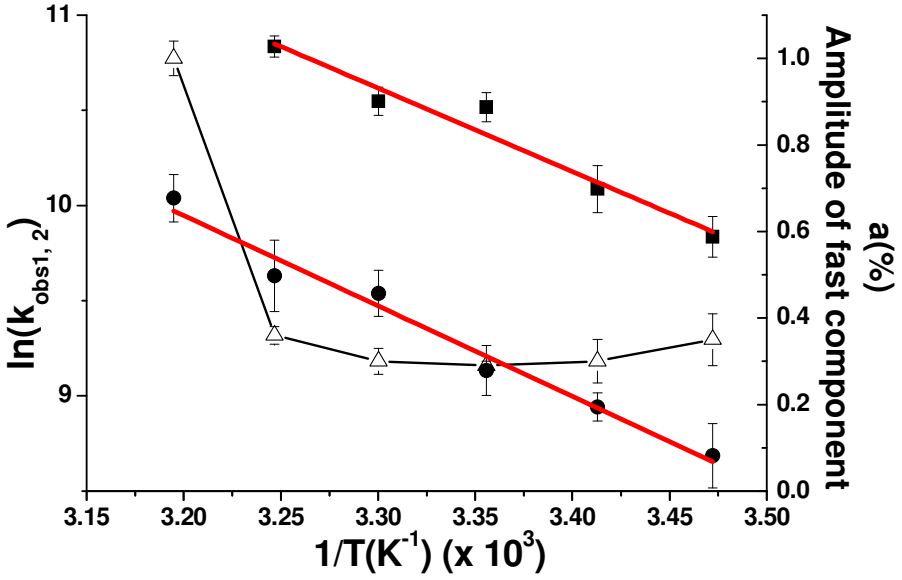


Figure 6: Temperature dependence of cTAR/dTAR annealing kinetics in the presence of BUNV-N protein at ratio 1:1. The reaction was performed with 10 nM doubly labelled cTAR and 500-nM non-labelled dTAR. The natural logarithm of the rate constant values for the fast (black squares) and slow (black circles) components as well as the amplitude of the fast component (open tringles) are indicated at six different temperatures. The solid magenta lines are the best fit to Equation (5) with E_{fast} = 8.8(±1) kcal/mol and Ea_{slow} = 9.5(±0.7) kcal/mol, for the fast and slow components, respectively.

Positive enthalpy values for the transition state of 8.2(±1) kcal/mol and 8.9(±1) kcal/mol for the fast and slow kinetic components respectively, were obtained. These values indicated that cTAR/dTAR annealing promoted by the BUNV-N protein involves premelting of approximately 2 base pairs for the fast as well as for slow component [Rouzina I et al., 1999, Cantor C and Schimmel P, 1980].

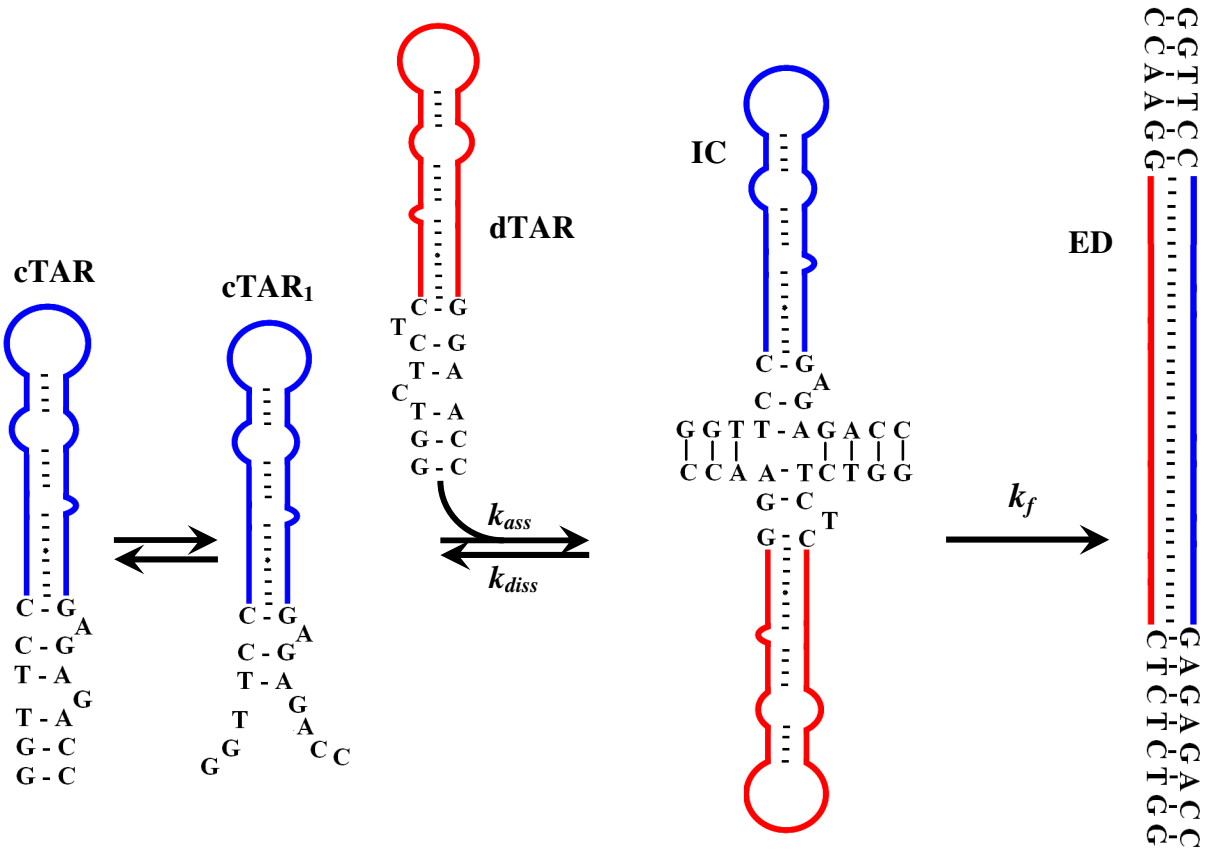


Figure 8: Proposed mechanism for BUNV-N-promoted cTAR/dTAR annealing. The two-step reaction pathway is thought to rely on the thermal fraying of cTAR that leads to a fast equilibrium (µs range) between the premelted reactive $cTAR_1$ species (Azoulay J et al., 2003). The $cTAR_1$ nucleates an intermediate complex (IC) through the stem termini. Then, the ICs are converted into the final extended duplex (ED) by a rate-limiting step.

Finally, on the basis of acquired data, it can be conclude that promotion of cTAR/dTAR annealing by BUNV-N protein starts from the cTAR and dTAR ends and proceed through a single two-step reaction pathway. Due to thermal fraying of cTAR termini at room temperature (Bernacchi S et al., 2002, Azoulay J et al., 2003, Beltz H et al 2003), the pathway is probably associated with a cTAR species where the 3 bp of the terminal stem or original secondary structure of cTAR, are melted (Figure 8). Therefore, IC is thought to be stabilized by 7 intermolecular base pairs, which then interconvert to ED by a rate limiting interconversion step, which most probably relies on the conformational rearrangement and melting of the stable upper part of both TAR species. According to the Ea_{slow} value (9.5(±0.7) kcal/mol), the melting of the 2bp of the penultimate double stranded segment of cTAR plays likely a key role in this interconversion step.

Chapter 5: Transfection ability and intracellular DNA pathway of nanostructured Gene-Delivery Systems

NANO LETTERS

2008
Vol. 8, No. 8
2432-2436

Transfection Ability and Intracellular DNA Pathway of Nanostructured Gene-Delivery Systems

Xin Zhang,[†,‡] Kamal Kant Sharma,[‡] Marcel Boeglin,[§] Joelle Ogier,[†] Didier Mainard,[||] Jean-Claude Voegel,[†] Yves Mély,[*,‡] and Nadia Benkirane-Jessel[*,†]

Institut National de la Santé et de la Recherche Médicale, Unité 595, Faculté de Médecine, 11 Rue Humann, 67085 Strasbourg Cedex, France, and Faculté de Chirurgie Dentaire de l'Université Louis Pasteur (ULP), 67000 Strasbourg Cedex, France, Département de Pharmacologie & Physicochimie, Equipe Photophysique des Interactions Biomoléculaires, Institut Gilbert Laustriat, UMR 7175, Faculté de Pharmacie, Université Louis Pasteur, B.P. 60024, 67401 Illkirch Cedex, France, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Centre National de la Recherche Scientifique (CNRS)/INSERM/ ULP, Collège de France, BP 10142, 67404 Strasbourg, France, and Unite Mixte de Recherches 7561, Centre National de la Recherche Scientifique-Universite Henri Poincare Nancy I, Faculte de Medecine, Vandoeuvre les Nancy, France, and Centre Hospitalier Universtaire de Nancy, Hopital Central, 29 Avenue du Marechal de Lattre de Tassigny, 54000 Nancy, France

Received May 14, 2008; Revised Manuscript Received June 3, 2008

ABSTRACT

Considerable efforts have been devoted to the design of structured materials with functional properties. Polyelectrolyte multilayer films are now a well-established nanostructured concept with numerous potential applications, in particular as biomaterial coatings. This technique allows the preparation of nanostructured architectures exhibiting specific properties for cell-activation control and local drug delivery. In this study, we used a multilayered system made of poly-(L-lysine)/hyaluronic acid (PLL/HA) as a reservoir for active DNA complexes with nonviral gene-delivery vectors, PLL, β-cyclodextrin (CD), and PLL-CD. When embedded into the multilayered films, the transfection efficiencies of the DNA complexes and the cell viability were improved. The highest transfection efficiency was obtained with the PLL-CD/plasmid DNA (pDNA) complexes. We found that this high transfection efficiency was related to an efficient internalization of the complexes in the cell cytoplasm and selected nuclei domains through a nonendocytotic pathway. For the first time, we report the intracellular pathway of the pDNA in complexes incorporated into the multilayered system.

In recent years, considerable efforts have been devoted to the design and controlled fabrication of structured materials with functional properties. The layer-by-layer (LBL) buildup of polyelectrolyte multilayers (PEM) films using oppositely charged polyelectrolytes offers new opportunities for the preparation of functionalized biomaterial coatings.[1,2] This technique allows the preparation of supramolecular nanoarchitectures exhibiting specific properties for cell activation and development of local drug-delivery systems.[3,4] Peptides, protein drugs or DNA, chemically bound to polyelectrolytes, adsorbed or embedded in PEM films, have been shown to retain their biological activities.[5–11]

During the past decade, gene therapy has also become a worldwide research focus and has been advanced considerably. The main objective in gene therapy is constructing an efficient gene-delivery system able to transfer the therapeutic DNA to the targeted tissues and cells. However, although a great deal of effort has been expended, the development of

* Corresponding authors. E-mail: nadia.jessel@medecine.u-strasbg.fr. Tel: 33 390 243 376. Fax: 33 390 243 379. E-mail: mely@pharma.u-strasbg.fr. Tel: 33 390 244 263. Fax: 33 390 244 312.

† Institut National de la Santé et de la Recherche Médicale, Unité 595, Faculté de Médecine, 11 Rue Humann, 67085 Strasbourg Cedex, France, and Faculté de Chirurgie Dentaire de l'Université Louis Pasteur (ULP), 67000 Strasbourg Cedex, France.

‡ Département de Pharmacologie & Physicochimie, Equipe Photophysique des Interactions Biomoléculaires, Institut Gilbert Laustriat, UMR 7175, Faculté de Pharmacie, Université Louis Pasteur, B.P. 60024, 67401 Illkirch Cedex, France.

§ Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Centre National de la Recherche Scientifique (CNRS)/INSERM/ULP, collège de France, BP 10142, 67404 Strasbourg, France.

|| Unite Mixte de Recherches 7561, Centre National de la Recherche Scientifique-Universite Henri Poincare Nancy I, Faculte de Medecine, Vandoeuvre les Nancy, France, and Centre Hospitalier Universtaire de Nancy, Hopital Central, 29 Avenue du Marechal de Lattre de Tassigny, 54000 Nancy, France.

10.1021/nl801379y CCC: $40.75
Published on Web 07/09/2008

an efficient and safe gene-delivery system remains the main challenge for gene therapy.

The vectors in gene therapy can be divided into viral and nonviral vector systems. So far, viral systems have been used most often but the issues of viral immunogenicity, difficulties of large scale industrial production, and possible random genomic integration remain as problems.[12] Therefore, nonviral delivery systems are increasingly investigated as alternative vectors. These systems include complexes of cationic lipids, polymers, or peptides with plasmid DNA. Controlled delivery of DNA from coating, hydrogel, tissue engineering scaffolds and nanoparticles, has also been developed.[13,14] An ideal gene-delivery vector should be efficient, nonimmunogenic, stable, and nontoxic. Although nonviral vectors represent attractive alternative to viral vectors for transfection in vitro and in vivo, they still suffer from relatively low efficiencies.[15,16]

Multilayers containing pDNA are of great interest for applications in sensing,[17] diagnostic,[18] electronics,[19] and gene-delivery.[20–22] Although the incorporation of pDNA into the multilayered films was first reported in 1993 by Lvov et al.,[23] substantive investigation indicated that this method has a significant impact on the development of gene therapies. In the past few years, it has also been shown that PEMs are efficient to deliver pDNA. For instance, Lynn et al. found that multilayered films composed of naked pDNA and a degradable polyamine are able to efficiently release pDNA from the surfaces of model substrates under physiological conditions and subsequent in vitro transfection of adherent cells.[21,22] They suggested that the films present DNA molecules in a condensed form, which could improve the internalization of DNA by cells.[22] Moreover, our group studied the use of PEM as vectors for polymer-precomplexed DNA,[24] and β-cyclodextrin-mediated DNA delivery to perform multiple and time-scheduled cell transfection.[10] Jessel et al. realized the sequential induction of nuclear and cytoplasmic expression products of two different pDNA embedded in PEM. Nevertheless, until now, there has been no report available that shows the intracellular pathway of pDNA in the multilayer system, which is one of the most critical steps in cell transfection. Therefore, to fill in this blank we, unprecedentedly, investigated in this paper the intracellular pathway of the pDNA in complexes in the multilayer system.

Our previous work showed that PEM films onto which cationic CD/pDNA complexes are adsorbed can act as an efficient gene-delivery tool to transfect cells.[10] Moreover, cationic CD derivatives are incorporated into polymer and dendrimer vectors, CDs were shown to increase the cell viability and transfection efficiency.[25,26] Thus, we chose poly(L-lysine) (PLL), a widely used nonviral gene-delivery vector,[27] to incorporate CD. Accordingly, PLL-CD was synthesized at a Lysine and cyclodextrin molar ratio of 2:1 and included as a complex with pDNA in a multilayer system made of the multilayered film PLL/HA. The transfection efficiency and the pDNA intracellular pathway with this PEM were investigated and compared with the same PEM in which pDNA has been complexed with PLL or CD, instead of PLL-CD. Comparison with complexes of pDNA with PLL, CD, or PLL-CD in solution was also performed. The multilayer systems showed much higher transfection efficiencies than the complexes in solution, and the best transfection efficiency was obtained with PLL-CD in the PEM. Using the bis-intercalator YOYO-1 to label the pDNA and an endocytosis marker FM 4-64, [28] we found using confocal microscopy that the enhanced efficiencies with the multilayers could be due to the pDNA intracellular pathway that differs from endocytosis.

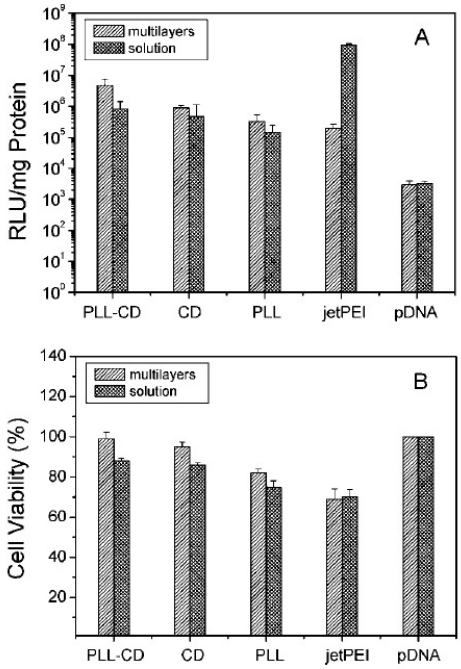


Figure 1. Transfection efficiency (A) and cytotoxicity (B) of the complexes of pDNA with PLL, CD and PLL-CD at a N/P ratio of 3 for *HeLa* cells grown on $(PLL\text{-}HA)_5$−(complexes)−$(PLL\text{-}HA)_5$ films or in the absence of multilayers for 24 h. Cells were lysed for luciferase activity quantification. Gene expression determined from the luciferase assay was expressed as RLU/mg of protein. JetPEI/pDNA complex and naked pDNA were used as controls.

The polyelectrolyte multilayer gene-transfer system we studied was based on pDNA in complexes with PLL, CD, and PLL-CD. The buildup of $(PLL\text{-}HA)_5$−pDNA complexes−$(PLL\text{-}HA)_5$ multilayered films was first monitored by quartz crystal microbalance (QCM). The PLL/pDNA and PLL-CD/pDNA complexes were prepared at N/P = 3, a ratio that is used commonly with PLL for transfection.[29,30] The CD/pDNA complexes were prepared at a CD and pDNA molar ratio of 3, where the amount of CD in complexes with pDNA was the same as that of conjugated CD in PLL−CD/pDNA complexes. The observed increase in $-\Delta f/\nu$ with the number of deposited layers (data not shown) suggested that this multilayer system exhibits a regular film deposition.

In the next step, transfection experiments in the presence of serum were performed with *Hela* cells by using the pCMV-Luc plasmid encoding the luciferase gene (Figure 1). To check the incidence of the multilayer systems on the

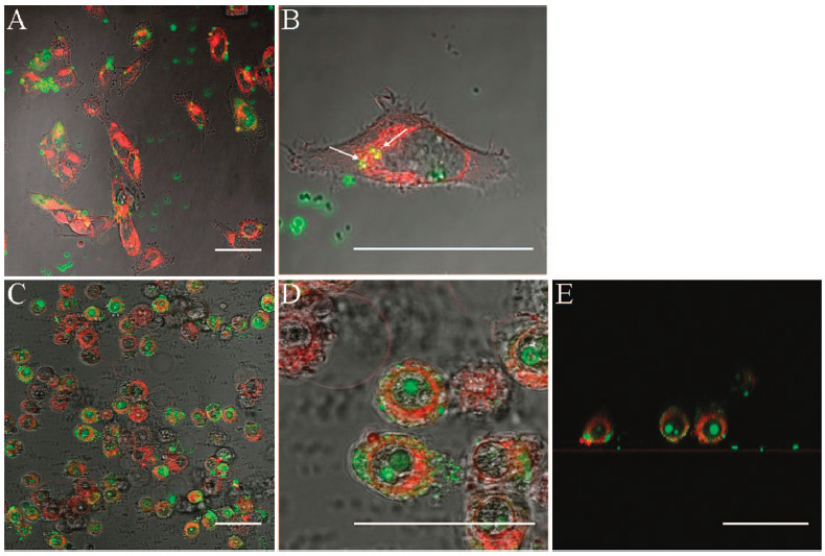


Figure 2. Internalization of the complexes of pDNA with PLL-CD in *HeLa* cells grown on the surface of $(PLL\text{-}HA)_5$−PLL-CD/pDNA complexes−$(PLL\text{-}HA)_5$ multilayered films (C−E) or in the absence of multilayers (A and B). Panels A−D correspond to *x*, *y* sections; Panel E corresponds to *x*, *z* sections. The concentration of pDNA was 60 μM, and the N/P ratio was 3. The images correspond to the overlay of light transmission and fluorescence confocal images and were taken at 24 h post transfection. Yellow spots (arrows) correspond to areas where the complexes and FM4-64 are colocalized. Red and green spots correspond to areas containing only FM4-64 and YOYO-1 labeled complexes, respectively. Images are representative of more than 90% of the observed cells ($n = 3$ specimens). Scale bars = 40 μm.

transfection efficiency of the pDNA complexed with nonviral gene-delivery vectors (PLL, CD, and PLL-CD), we compared the luciferase gene expression of the complexes in solution with those embedded in the multilayers.

In solution, the transfection efficiencies of the complexes of pDNA with PLL-CD, PLL, and CD were lower than that of jetPEI and significantly higher than that of naked pDNA, in line with the literature (Figure 1A). When embedded into the multilayers, the efficiencies of these complexes were increased dramatically and were much higher than that of naked pDNA, even higher than that of jetPEI. Within the tested multilayer systems, the highest efficiency was observed with PLL-CD/DNA complexes, which shows 5- and 14-fold higher transfection efficiency as pDNA complexed with CD and PLL, respectively. Furthermore, the cell viabilities were enhanced significantly when the pDNA complexes with PLL-CD, PLL, and CD were included in the multilayer systems (Figure 1 B). This increase in cell viability may contribute to their improved transfection efficiencies.

To further understand the differences in transfection efficiency, the intracellular trafficking of the complexes of pDNA with PLL-CD, PLL, and CD in solution and embedded into the multilayered system was monitored by confocal microscopy (Figures 2–4). We labeled the pDNA and the endosomes with YOYO-1 and FM4-64, a marker of membrane endocytosis, respectively.[31] These two dyes can be followed simultaneously because the FM4-64 red fluorescence is easily distinguished from the YOYO-1 green fluorescence. In this respect, colocalization of FM4-64 and pDNA complexes will yield yellow spots. The intracellular pathway of the complexes was followed 24 h post transfection of *HeLa* cells.

First, the PLL-CD/pDNA complexes in solution were perceived mainly close to the cell surface (Figure 2A and B). Only a limited number of complexes entered in cells. The complexes entering into the cells provided perinuclear yellow spots that indicated internalization through endocytosis with an accumulation of the complexes in late endosomes and lysosomes (Figure 2B). In sharp contrast to the solution, numerous green spots but no yellow spots were observed in the cytoplasm and the nucleus when PLL-CD/pDNA complexes were delivered from multilayers (Figure 2C−E). The absence of yellow spots and perinulear accumulation of the complexes seemed to indicate that the PLL-CD/pDNA complexes delivered from the multilayer system are not internalized by endocytosis. We suggest that the efficient internalization by the nonendocytic intracellular pathway contributes to the higher transfection efficiency when the complexes were delivered from the multilayer system. This unknown intracellular pathway could facilitate the pDNA in complexes to escape from late endosomes and lysosomes, where pDNA is degraded.[32]

To further confirm these observations above, we also monitored the intracellular trafficking of the complexes of pDNA with PLL and CD in both conditions (Figures 3 and 4). The PLL/pDNA complexes in solution were perceived mainly as large aggregates close to the cell surface (Figure 3A and B). Only a limited number of complexes entered in cells. The complexes entering into the cells were of smaller size and gave yellow spots, suggesting that the complexes were internalized by endocytosis (Figure 3B). Both the limited number of PLL/pDNA complexes that are internalized and their probable accumulation into lysosomes are fully consistent with the observed low transfection efficiency. In

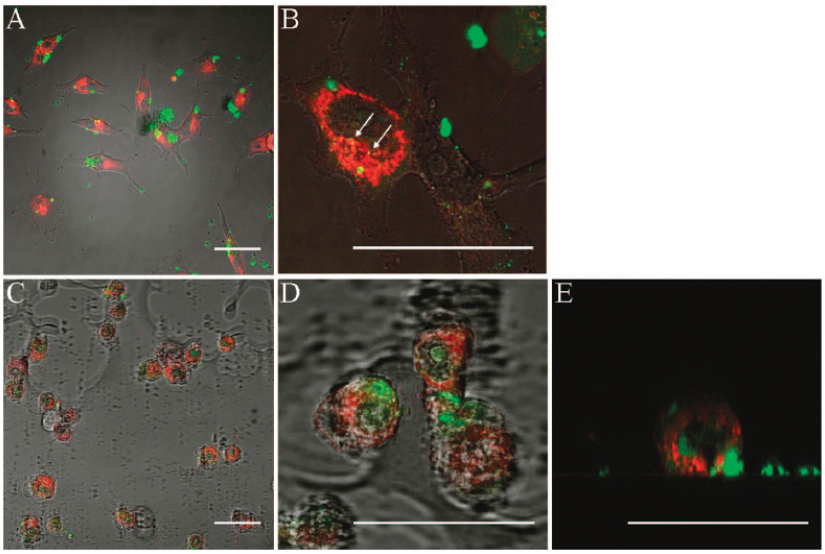


Figure 3. Internalization of the complexes of pDNA with PLL in *HeLa* cells grown on the surface of $(PLL\text{-}HA)_5$−PLL/pDNA complexes−$(PLL\text{-}HA)_5$ multilayered films (C−E) or in the absence of multilayers (A and B). Panels A−D correspond to *x*, *y* sections; Panel E corresponds to *x*, *z* sections. Experimental conditions were the same as those in Figure 2.

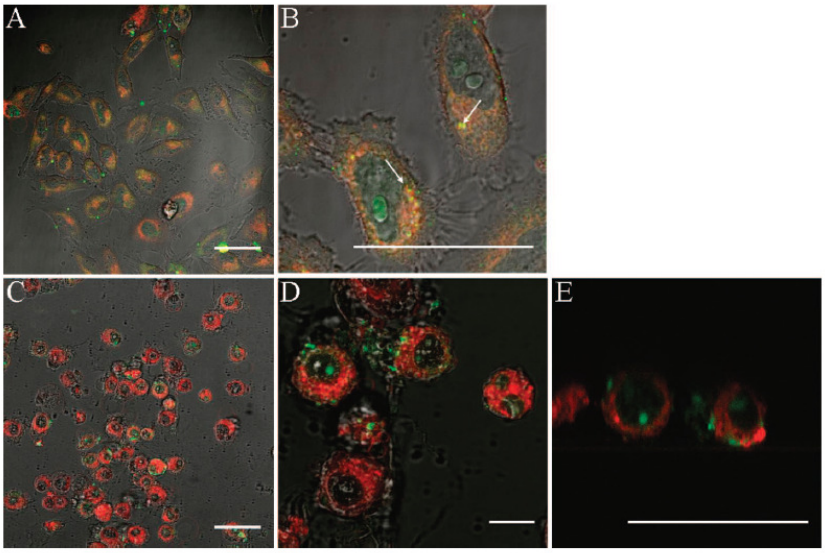


Figure 4. Internalization of the complexes of pDNA with CD in *HeLa* cells grown on the surface of $(PLL\text{-}HA)_5$−CD/pDNA complexes−$(PLL\text{-}HA)_5$ multilayered films (C−E) or in the absence of multilayers (A and B). Panels A−D correspond to *x*, *y* sections; Panel E corresponds to *x*, *z* sections. Experimental conditions were the same as those in Figure 2.

the case of the CD/pDNA complexes, a significant number of small yellow spots could be observed when they were in solution, indicating an internalization by endocytosis in this case also (Figure 4B). This conclusion was further substantiated by the mainly perinuclear localization of the complexes, in line with an accumulation of the complexes in the late endosomes and lysosomes. [31,33]

When the PLL/pDNA and CD/pDNA complexes were released from the multilayers, as the PLL-CD/pDNA complexes, no yellow spot could be perceived, indicating that the complexes are not internalized by endocytosis either. Moreover, the complexes of pDNA with PLL and CD were observed as numerous green spots in the cytoplasm and the nucleus (Figure 3C−E and 4C−E), suggesting that their intracellular trafficking is more efficient than that solution, which gives rise to their higher transfection efficiency.

Taken together, our data suggest that the higher transfection efficiency of the nonviral gene-delivery vectors (PLL-CD, PLL, and CD)/pDNA complexes delivered from the multilayer system than in solution may be related to their efficient internalization in the cytoplasm and the nuclei of the transfected cells. Their internalization clearly differs from

endocytosis by several lines of evidence. First, no colocalization with the endocytosis marker FM4-64 was observed in any transfected cell. Second, no perinuclear accumulation in late endosomes and lysosomes could be observed with the complexes delivered from the multilayers. Third, the complexes were found in a large number of cell nuclei, indicating that the complexes delivered from the multilayers easily cross the nuclear membranes and accumulate in selected regions of the nuclei. A comparison with the same complexes added in solution strongly suggests that escape from late endosomes and lysosomes as well as degradation in the lysosomes may constitute the main bottlenecks for efficient transfection in this case. In addition, complexes embedded in multilayers exhibited low cytotoxicity, indicating that multilayer systems are highly suited for in vivo applications.

Acknowledgment. This work was supported by the project ANR06-BLAN-0197-01/CartilSpray, from the "Agence Nationale de la Recherche", the "Fondation Avenir", the "Ligue contre le Cancer, du haut Rhin, Région Alsace" and "Cancéropôle du Grand Est". X.Z thanks the Faculté de Chirurgie Dentaire of Strasbourg for financial support. N.J is indebted to CHU de Nancy (Contrat d'interface vers l'hôpital).

Supporting Information Available: Experimental procedures. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Decher, G. *Science* **1997**, *277*, 1232–1237.
(2) Wang, D. Y.; Rogach, A. L.; Caruso, F. *Nano. Lett.* **2002**, *2*, 857–861.
(3) Berth, G.; Voigt, A.; Dautzenberg, H.; Donath, E.; Mohwald, H. *Biomacromolecules* **2002**, *3*, 579–90.
(4) Hiller, J.; Mendelsohn, J. D.; Rubner, M. F. *Nat. Mater.* **2002**, *1*, 59–63.
(5) Jessel, N.; Atalar, F.; Lavalle, P.; Mutterer, J.; Decher, G.; Schaaf, P.; Voegel, J. C.; Ogier, J. *Adv. Mater.* **2003**, *15*, 692–695.
(6) Jessel, N.; Schwinté, P.; Falvey, P.; Darcy, R.; Haikel, Y.; Schaaf, P.; Voegel, J. C.; Ogier, J. *Adv. Funct. Mater.* **2004**, *14*, 174–182.
(7) Jessel, N.; Schwinté, P.; Donohue, R.; Lavalle, P.; Boulmedais, F.; Darcy, R.; Szalontai, B.; Voegel, J. C.; Ogier, J. *Adv. Funct. Mater.* **2004**, *14*, 963–969.
(8) Jessel, N.; Lavalle, P.; Meyer, F.; Audouin, F.; Frisch, B.; Schaaf, P.; Ogier, J.; Decher, G.; Voegel, J. C. *Adv. Mater.* **2004**, *16*, 1507–1577.
(9) Jessel, N.; Lavalle, P.; Hubsch, E.; Holl, V.; Senger, B.; Haikel, Y.; Voegel, J. C.; Ogier, J.; Schaaf, P. *Adv. Funct. Mater.* **2005**, *15*, 648–654.
(10) Jessel, N.; Oulad-Abdelghani, M.; Meyer, F.; Lavalle, P.; Haikel, Y.; Schaaf, P.; Voegel, J. C. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8618–21.
(11) Kempf, M.; Mandal, B.; Jilek, S.; Thiele, L.; Voros, J.; Textor, M.; Merkle, H. P.; Walter, E. *J. Drug Target* **2003**, *11*, 11–8.
(12) Khalil, I. A.; Kogure, K.; Akita, H.; Harashima, H. *Pharmacol. Rev.* **2006**, *58*, 32–45.
(13) Garza, J. M.; Jessel, N.; Ladam, G.; Dupray, V.; Muller, S.; Stoltz, J. F.; Schaaf, P.; Voegel, J. C.; Lavalle, P. *Langmuir* **2005**, *21*, 12372–7.
(14) Ren, K.; Wang, Y.; Ji, J.; Lin, Q.; Shen, J. *Colloids Surf., B* **2005**, *46*, 63–9.
(15) Strand, S. P.; Danielsen, S.; Christensen, B. E.; Varum, K. M. *Biomacromolecules* **2005**, *6*, 3357–66.
(16) Zhang, X.; Ercelen, S.; Duportail, G.; Schaub, E.; Tikhonov, V.; Slita, A.; Zarubaev, V.; Babak, V.; Mely, Y. *J. Gene. Med.* **2008**,
(17) Boon, E. M.; Ceres, D. M.; Drummond, T. G.; Hill, M. G.; Barton, J. K. *Nat. Biotechnol.* **2000**, *18*, 1096–100.
(18) Schouten, S.; Stroeve, P.; Longo, M. L. *Langmuir* **1999**, *15*, 8133–8138.
(19) Hartwich, G.; Caruana, D. J.; Lumley-Woodyear, T.; Wu, Y. B.; Campbell, C. N.; Heller, A. *J. Am. Chem. Soc.* **1999**, *121*, 10803–10812.
(20) Jewell, C. M.; Lynn, D. M. *Adv. Drug. Delivery Rev.* **2008**, *60*, 979–99.
(21) Zhang, J.; Chua, L. S.; Lynn, D. M. *Langmuir* **2004**, *20*, 8015–21.
(22) Jewell, C. M.; Zhang, J.; Fredin, N. J.; Lynn, D. M. *J. Controlled Release* **2005**, *106*, 214–23.
(23) Lvov, Y.; Decher, G.; G., S *Macromolecules* **1993**, *26*, 5396–5399.
(24) Meyer, F.; Ball, V.; Schaaf, P.; Voegel, J. C.; Ogier, J. *Biochim. Biophys. Acta* **2006**, *1758*, 419–22.
(25) Gonzalez, H.; Hwang, S. J.; Davis, M. E. *Bioconjugate Chem.* **1999**, *10*, 1068–74.
(26) Arima, H.; Kihara, F.; Hirayama, F.; Uekama, K. *Bioconjugate Chem.* **2001**, *12*, 476–84.
(27) Merdan, T.; Kopecek, J.; Kissel, T. *Adv. Drug Delivery Rev.* **2002**, *54*, 715–58.
(28) Krishnamoorthy, G.; Duportail, G.; Mely, Y. *Biochemistry* **2002**, *41*, 15277–87.
(29) Rimann, M.; Luhmann, T.; Textor, M.; Guerino, B.; Ogier, J.; Hall, H. *Bioconjugate Chem.* **2008**, *19*, 548–57.
(30) Mannisto, M.; Reinisalo, M.; Ruponen, M.; Honkakoski, P.; Tammi, M.; Urtti, A. *J. Gene Med.* **2007**, *9*, 479–87.
(31) Remy-Kristensen, A.; Clamme, J. P.; Vuilleumier, C.; Kuhry, J. G.; Mely, Y. *Biochim. Biophys. Acta* **2001**, *1514*, 21–32.
(32) Akinc, A.; Langer, R. *Biotechnol. Bioeng.* **2002**, *78*, 503–8.
(33) Clamme, J. P.; Krishnamoorthy, G.; Mely, Y. *Biochim. Biophys. Acta* **2003**, *1617*, 52–61.

NL801379Y

Transfection ability and intracellular DNA pathway of nano-structured gene delivery systems

Xin Zhang,[†, ‡] Kamal Kant Sharma,[‡] Marcel Boeglin,[§] Joelle Ogier,[†] Jean-Claude Voegel,[†] Yves Mély,[*, ‡] and Nadia Jessel[*, †]

[†] *Institut National de la Santé et de la Recherche Médicale (INSERM), Unité 595, Faculté de Médecine, 11 Rue Humann, 67085 Strasbourg Cedex, France and Faculté de Chirurgie Dentaire de l'Université Louis Pasteur (ULP), 67000 Strasbourg Cedex, France*

[‡] *Département de Pharmacologie & Physicochimie, Equipe Photophysique des Interactions Biomoléculaires, Institut Gilbert Laustriat, UMR 7175, Faculté de Pharmacie, Université Louis Pasteur, B.P. 60024, 67401 Illkirch Cedex, France*

[§] *Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Centre National de la Recherche Scientifique (CNRS)/INSERM/ULP, collège de France, BP 10142, 67404 Strasbourg, France.*

Corresponding authors:

nadia.jessel@medecine.u-strasbg.fr. Tel: 33 390 243 376, Fax: 33 390 243 379

mely@pharma.u-strasbg.fr Tel: 33 390 244 263, Fax: 33 390 244 312

Materials

Poly-L-lysine hydrobromide (PLL, MW = 30 kDa), hyaluronic acid (HA, MW = 400 kDa), β-Cyclodextrin (β-CD) and YOYO-1 (1,1'-(4,4,8,8-tetramethyl-4,8-diazaundecamethylene)bis[4-{(3-methylbenz-1,3-oxazol-2-yl)methylidine}-1,4-dihydroquinolinium] tetraiodide were purchased from Sigma. The polymer poly-L-lysine-cyclodextrin (PLL-CD) was synthesized by coupling the β-CD to the free ε-amino group of lysine via carboxyl group at a Lysine and cyclodextrin molar ratio of 2:1 by Altergen (Strasbourg, France). The characterization of this coupling was performed by this company also. The pCMV-Luc plasmid (5581 bp) encoding luciferase was grown in E. coli and purified by a Qiagen kit (Qiagen, US). The purity and integrity of the plasmids were assessed by absorption spectroscopy (A_{260}/A_{280} ratio) and electrophoresis on a 1 % agarose gel. The DNA concentration was determined by UV absorbance at 260 nm, using a Cary 400 spectrophotometer. JetPEI was purchased from Polyplus-transfection (Illkirch, France).

Methods

Complex Preparation

Stock solutions of PLL (0.3 mM, 0.5 mg/mL), PLL-CD (1.44 mM) and CD (8.81 mM) were prepared by dissolving them in 0.15 M, pH 7.4 NaCl solution. PLL/pDNA and PLL-CD/pDNA complexes were prepared by adding vectors to the DNA at N/P ratio of 3 followed by intense vortexing for 10 s and 1h incubation at room temperature.[1] The N/P ratio between vectors and pDNA is expressed as the molar ratio between all the protonable amino groups of the vectors and the phosphate groups of the pDNA. The CD/pDNA complexes were prepared at CD and pDNA molar ratio of 3.

Preparation of Polyelectrolyte Multilayers

Polyelectrolyte multilayered films were prepared on glass coverslips pre-treated with 10^{-2} M SDS and 0.12 M HCl for 15 min at 100 °C, and then extensively rinsed with deionized water.[1] Finally, glass slides were placed in 24-well plate. Next, a precursor film constituted by $(PLL/HA)_5$ was built up by alternated immersions for 10 min in polyelectrolyte solutions (300 µL) at the concentration of 0.5 mg/mL for PLL and HA in 0.15 M NaCl solution pH 7.4. After each polyelectrolyte deposition, the cover slips were rinsed three times during 5 min with deionised water. Then, these complexes were added over the multilayers and incubated for 1 h 30 min. After the deposition of complexes over the architecture, another film constituted by $(PLL/HA)_5$ was built on the top of the complexes. PLL was deposited as the last layer, so that cells can adhere to it. After deposition of n bilayers, architectures were dried and sterilised under UV light for 15 min. The obtained architecture was incubated in Cell culture medium (1mL) 24 h and *HeLa* Cells were seeded on it followed by 24 h incubation.

Cell Transfection and Cytotoxicity

Hela human cervix epitheloid carcinoma cells were grown in DMEM (Dulbelcco's modified eagle medium) supplemented with 10 % FBS (CAMBREX, Belgium), 100 units/mL penicillin (Gibco BRL) and 100 µl/mL streptomycin (Gibco BRL). Cells were maintained at 37° C in a 5 % CO_2 humidified atmosphere.[2, 3]

Transfection was done by using polyplexes with 0.6 µg of pDNA (pCMV Luc) and PLL, PLL-CD and CD at N/P = 3 with incubation time of 24 h. After cell lysis, luciferase gene expression was quantified using a commercial kit and a luminometer.[3] Results were expressed as relative light unit (RLU) integrated over 10 s per milligram of a cell protein lysate (RLU/mg of protein)

Cytotoxicity of complexes was evaluated by quantification of the cellular content in proteins. *Hela* cells were seeded at a cell density of 1×10^5 cells/well and incubated for 24 h.

PLL/pDNA, PLL-CD/pDNA and CD/pDNA complexes were added to the cells for 24 h. Cells were then washed with phosphate-buffered saline (PBS), and 100 µL of cell culture lysis reagent (Promega) was added to each well. The cellular proteins were measured using the bicinchoninic acid assay (Interchim, Montlucon, France) following an incubation time of 30 min at 60 °C. The absorbance was read at 562 nm. [4]

Confocal Laser Scanning Microscopy (CLSM)

CLSM observations were carried out with an Zeiss LSM 510 microscope using a X40/1.4 oil immersion objective and with 0.4 µm *z*-section intervals. FITC fluorescence was detected using excitation at 488 nm with a cutoff dichroic mirror 488 nm and an emission band pass filter 505-530 nm (green). Rhodamine fluorescence was detected using excitation at 543nm, a dichroic mirror 543 nm, and an emission long pass filter 585 nm (red).[5] YOYO-1 fluorescence was detected using excitation at 488 nm with a cutoff dichroic mirror 488 nm and an emission band pass filter 506-538 nm (green).[1] FM4-64 fluorescence was detected using excitation at 568 nm and an emission band pass filter 664-696 nm (red). All experiments were performed in liquid conditions.

References

1. Clamme, J. P.; Bernacchi, S.; Vuilleumier, C.; Duportail, G.; Mely, Y. *Biochim Biophys Acta* **2000,** 1467, (2), 347-61.
2. Sato, T.; Ishii, T.; Okahata, Y. *Biomaterials* **2001,** 22, (15), 2075-80.
3. Borchard, G. *Adv Drug Deliv Rev* **2001,** 52, (2), 145-50.
4. Hayek, A.; Ercelen, S.; Zhang, X.; Bolze, F.; Nicoud, J. F.; Schaub, E.; Baldeck, P. L.; Mely, Y. *Bioconjug Chem* **2007,** 18, (3), 844-51.

5. Garza, J. M.; Jessel, N.; Ladam, G.; Dupray, V.; Muller, S.; Stoltz, J. F.; Schaaf, P.; Voegel, J. C.; Lavalle, P. *Langmuir* **2005,** 21, (26), 12372-7.

References

[1] Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. (1989) Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science*; 244:359-62.

[2] Choo QL, Richman KH, Han JH, Berger K, Lee C, Dong C, et al. (1991) Genetic organization and diversity of the hepatitis C virus. *Proc Natl Acad Sci U S A*; 88:2451-5.

[3] Lauer GM, Walker BD. (2001) Hepatitis C virus infection. *N Engl J Med*; 345:41-52.

[4] Giannini C, Brechot C. (2003) Hepatitis C virus biology. *Cell Death Differ*; 10 Suppl 1:S27-38.

[5] Eyster ME, Alter HJ, Aledort LM, Quan S, Hatzakis A, Goedert JJ. (1991) Heterosexual co-transmission of hepatitis C virus (HCV) and human immunodeficiency virus (HIV). *Ann Intern Med*; 115:764-8.

[6] Wyld R, Robertson JR, Brettle RP, Mellor J, Prescott L, Simmonds P. (1997) Absence of hepatitis C virus transmission but frequent transmission of HIV-1 from sexual contact with doubly-infected individuals. *J Infect*; 35:163-6.

[7] Clarke A, Kulasegaram R. (2006) Hepatitis C transmission -- where are we now? *Int J STD AIDS*; 17:74-80; quiz

[8] Serpaggi J, Chaix ML, Batisse D, Dupont C, Vallet-Pichard A, Fontaine H, et al. (2006) Sexually transmitted acute infection with a clustered genotype 4 hepatitis C virus in HIV-1-infected men and inefficacy of early antiviral therapy. *Aids*; 20:233-40.

[9] Stauber RE, Stadlbauer V. (2006) Novel approaches for therapy of chronic hepatitis C. *J Clin Virol*; 36:87-94.

[10] Honda M, Beard MR, Ping LH, Lemon SM. (1999b) A phylogenetically conserved stem-loop structure at the 5' border of the internal ribosome entry site of hepatitis C virus is required for cap-independent viral translation. *J Virol*; 73:1165-74.

[11] Blight KJ, Rice CM. (1997) Secondary structure determination of the conserved 98-base sequence at the 3' terminus of hepatitis C virus genome RNA. *J Virol*; 71:7345-52.

[12] Kolykhalov AA, Feinstone SM, Rice CM. (1996) Identification of a highly conserved sequence element at the 3' terminus of hepatitis C virus genome RNA. *J Virol*; 70:3363-71.

[13] Hijikata M, Kato N, Ootsuyama Y, Nakagawa M, Shimotohno K. (1991) Gene mapping of the putative structural region of the hepatitis C virus genome by in vitro processing analysis. *Proc Natl Acad Sci U S A*; 88:5547-51.

[14] Lemberg MK, Martoglio B. (2002) Requirements for signal peptide peptidase-catalyzed intramembrane proteolysis. *Mol Cell*; 10:735-44.

[15] Santolini E, Migliaccio G, La Monica N. (1994) Biosynthesis and biochemical properties of the hepatitis C virus core protein. *J Virol*; 68:3631-41.

[16] Barba G, Harper F, Harada T, Kohara M, Goulinet S, Matsuura Y, et al. (1997) Hepatitis C virus core protein shows a cytoplasmic localization and associates to cellular lipid storage droplets. *Proc Natl Acad Sci U S A*; 94:1200-5.

[17] McLauchlan J. (2000) Properties of the hepatitis C virus core protein: a structural protein that modulates cellular processes. *J Viral Hepat*; 7:2-14.

[18] Pietschmann T, Lohmann V, Kaul A, Krieger N, Rinck G, Rutter G, et al. (2002) Persistent and transient replication of full-length hepatitis C virus genomes in cell culture. *J Virol*; 76:4008-21.

[19] Schwer B, Ren S, Pietschmann T, Kartenbeck J, Kaehlcke K, Bartenschlager R, et al. (2004) Targeting of hepatitis C virus core protein to mitochondria through a novel C-terminal localization motif. *J Virol*; 78:7958-68.

[20] Ray RB, Ray R. (2001) Hepatitis C virus core protein: intriguing properties and functional relevance. *FEMS Microbiol Lett*; 202:149-56.

[21] Lerat H, Honda M, Beard MR, Loesch K, Sun J, Yang Y, et al. (2002) Steatosis and liver cancer in transgenic mice expressing the structural and nonstructural proteins of hepatitis C virus. *Gastroenterology*; 122:352-65.

[22] Boni S, Lavergne JP, Boulant S, Cahour A. (2005) Hepatitis C virus core protein acts as a trans-modulating factor on internal translation initiation of the viral RNA. *J Biol Chem*; 280:17737-48.

[23] Cimarelli A, Luban J. (1999) Translation elongation factor 1-alpha interacts specifically with the human immunodeficiency virus type 1 Gag polyprotein. *J Virol*; 73:5388-401.

[24] Ivanyi-Nagy R, Kanevsky I, Gabus C, Lavergne JP, Ficheux D, Penin F, et al. (2006) Analysis of hepatitis C virus RNA dimerization and core-RNA interactions. *Nucleic Acids Res*; 34:2618-33.

[25] Herschlag D. (1995) RNA chaperones and the RNA folding problem. *J Biol Chem*; 270:20871-4.

[26] Cristofari G, Ivanyi-Nagy R, Gabus C, Boulant S, Lavergne JP, Penin F, et al. (2004) The hepatitis C virus Core protein is a potent nucleic acid chaperone that directs dimerization of the viral (+) strand RNA in vitro. *Nucleic Acids Res*; 32:2623-31.

[27] Lohmann V, Roos A, Korner F, Koch JO, Bartenschlager R. (2000) Biochemical and structural analysis of the NS5B RNA-dependent RNA polymerase of the hepatitis C virus. *J Viral Hepat*; 7:167-74.

[28] Mikkelsen JG, Rasmussen SV, Pedersen FS. (2004) Complementarity-directed RNA dimer-linkage promotes retroviral recombination in vivo. *Nucleic Acids Res*; 32:102-14.

[29] Colina R, Casane D, Vasquez S, Garcia-Aguirre L, Chunga A, Romero H, et al. (2004) Evidence of intratypic recombination in natural populations of hepatitis C virus. *J Gen Virol*; 85:31-7.

[30] Pawlotsky JM. (2003) Use and interpretation of hepatitis C virus diagnostic assays. *Clin Liver Dis*; 7:127-37.

[31] Perrault M, Pecheur EI. (2009) The hepatitis C virus and its hepatic environment: a toxic but finely tuned partnership. *Biochem J*; 423:303-14.

[32] Feinstone SM, Kapikian AZ, Purcell RH, Alter HJ, Holland PV. (1975) Transfusion-associated hepatitis not due to viral hepatitis type A or B. *N Engl J Med*; 292:767-70.

[33] Alter HJ, Purcell RH, Holland PV, Popper H. (1978) Transmissible agent in non-A, non-B hepatitis. *Lancet*; 1:459-63.

[34] Houghton M. (1996) *Fields Virology*. Lippincott-Raven, Philadelphia.

[35] Saito I, Miyamura T, Ohbayashi A, Harada H, Katayama T, Kikuchi S, et al. (1990) Hepatitis C virus infection is associated with the development of hepatocellular carcinoma. *Proc Natl Acad Sci U S A*; 87:6547-9.

[36] Shimotohno K. (2000) Hepatitis C virus and its pathogenesis. *Semin Cancer Biol*; 10:233-40.

[37] Recommendations from the National Institutes of Health consensus development conference statement: management of hepatitis C: 2002. Hepatology, 2002:1039.

[38] Frank C, Mohamed MK, Strickland GT, Lavanchy D, Arthur RR, Magder LS, et al. (2000) The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. *Lancet*; 355:887-91.

[39] Takamizawa A, Mori C, Fuke I, Manabe S, Murakami S, Fujita J, et al. (1991) Structure and organization of the hepatitis C virus genome isolated from human carriers. *J Virol*; 65:1105-13.

[40] Rice CM. (1996) *Fields Virology*. Philadelphia, .

[41] Linnen J, Wages J, Jr., Zhang-Keck ZY, Fry KE, Krawczynski KZ, Alter H, et al. (1996) Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. *Science*; 271:505-8.

[42] Robertson B, Myers G, Howard C, Brettin T, Bukh J, Gaschen B, et al. (1998) Classification, nomenclature, and database development for hepatitis C virus (HCV) and related viruses: proposals for standardization. International Committee on Virus Taxonomy. *Arch Virol*; 143:2493-503.

[43] Simons JN, Pilot-Matias TJ, Leary TP, Dawson GJ, Desai SM, Schlauder GG, et al. (1995) Identification of two flavivirus-like genomes in the GB hepatitis agent. *Proc Natl Acad Sci U S A*; 92:3401-5.

[44] Collett MS, Moennig V, Horzinek MC. (1989) Recent advances in pestivirus research. *J Gen Virol*; 70 (Pt 2):253-66.

[45] Hahn YS, Galler R, Hunkapiller T, Dalrymple JM, Strauss JH, Strauss EG. (1988) Nucleotide sequence of dengue 2 RNA and comparison of the encoded proteins with those of other flaviviruses. *Virology*; 162:167-80.

[46] Miller RH, Purcell RH. (1990) Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. *Proc Natl Acad Sci U S A*; 87:2057-61.

[47] Brown EA, Zhang H, Ping LH, Lemon SM. (1992) Secondary structure of the 5' nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. *Nucleic Acids Res*; 20:5041-5.

[48] Han JH, Shyamala V, Richman KH, Brauer MJ, Irvine B, Urdea MS, et al. (1991) Characterization of the terminal regions of hepatitis C viral RNA: identification of conserved sequences in the 5' untranslated region and poly(A) tails at the 3' end. *Proc Natl Acad Sci U S A*; 88:1711-5.

[49] Simmonds P. (2004) Genetic diversity and evolution of hepatitis C virus--15 years on. *J Gen Virol*; 85:3173-88.

[50] Martell M, Esteban JI, Quer J, Genesca J, Weiner A, Esteban R, et al. (1992) Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. *J Virol*; 66:3225-9.

[51] Zein NN. (2000) Clinical significance of hepatitis C virus genotypes. *Clin Microbiol Rev*; 13:223-35.

[52] Kaito M, Watanabe S, Tsukiyama-Kohara K, Yamaguchi K, Kobayashi Y, Konishi M, et al. (1994) Hepatitis C virus particle detected by immunoelectron microscopic study. *J Gen Virol*; 75 (Pt 7):1755-60.

[53] Andre P, Komurian-Pradel F, Deforges S, Perret M, Berland JL, Sodoyer M, et al. (2002) Characterization of low- and very-low-density hepatitis C virus RNA-containing particles. *J Virol*; 76:6919-28.

[54] Bradley D, McCaustland K, Krawczynski K, Spelbring J, Humphrey C, Cook EH. (1991) Hepatitis C virus: buoyant density of the factor VIII-derived isolate in sucrose. *J Med Virol*; 34:206-8.

[55] Hijikata M, Mizushima H, Akagi T, Mori S, Kakiuchi N, Kato N, et al. (1993) Two distinct proteinase activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus. *J Virol*; 67:4665-75.

[56] Hijikata M, Shimizu YK, Kato H, Iwamoto A, Shih JW, Alter HJ, et al. (1993) Equilibrium centrifugation studies of hepatitis C virus: evidence for circulating immune complexes. *J Virol*; 67:1953-8.

[57] Gastaminza P, Cheng G, Wieland S, Zhong J, Liao W, Chisari FV. (2008) Cellular determinants of hepatitis C virus assembly, maturation, degradation, and secretion. *J Virol*; 82:2120-9.

[58] Huang H, Sun F, Owen DM, Li W, Chen Y, Gale M, Jr., et al. (2007) Hepatitis C virus production by human hepatocytes dependent on assembly and secretion of very low-density lipoproteins. *Proc Natl Acad Sci U S A*; 104:5848-53.

[59] Bukh J, Purcell RH, Miller RH. (1992) Sequence analysis of the 5' noncoding region of hepatitis C virus. *Proc Natl Acad Sci U S A*; 89:4942-6.

[60] Tsukiyama-Kohara K, Iizuka N, Kohara M, Nomoto A. (1992) Internal ribosome entry site within hepatitis C virus RNA. *J Virol*; 66:1476-83.

[61] Lindenbach BD, Rice CM. (2001) *Feilds Virology*. Williams & Wilkins, Philadelphia.

[62] Kato N, Hijikata M, Ootsuyama Y, Nakagawa M, Ohkoshi S, Sugimura T, et al. (1990) Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. *Proc Natl Acad Sci U S A*; 87:9524-8.

[63] Moradpour D, Brass V, Bieck E, Friebe P, Gosert R, Blum HE, et al. (2004) Membrane association of the RNA-dependent RNA polymerase is essential for hepatitis C virus RNA replication. *J Virol*; 78:13278-84.

[64] Shimotohno K, Tanji Y, Hirowatari Y, Komoda Y, Kato N, Hijikata M. (1995) Processing of the hepatitis C virus precursor protein. *J Hepatol*; 22:87-92.

[65] Tanaka T, Kato N, Cho MJ, Shimotohno K. (1995) A novel sequence found at the 3' terminus of hepatitis C virus genome. *Biochem Biophys Res Commun*; 215:744-9.

[66] Friebe P, Bartenschlager R. (2002) Genetic analysis of sequences in the 3' nontranslated region of hepatitis C virus that are important for RNA replication. *J Virol*; 76:5326-38.

[67] Reynolds JE, Kaminski A, Carroll AR, Clarke BE, Rowlands DJ, Jackson RJ. (1996) Internal initiation of translation of hepatitis C virus RNA: the ribosome entry site is at the authentic initiation codon. *RNA*; 2:867-78.

[68] Wang C, Sarnow P, Siddiqui A. (1993) Translation of human hepatitis C virus RNA in cultured cells is mediated by an internal ribosome-binding mechanism. *J Virol*; 67:3338-44.

[69] Poole TL, Wang C, Popp RA, Potgieter LN, Siddiqui A, Collett MS. (1995) Pestivirus translation initiation occurs by internal ribosome entry. *Virology*; 206:750-4.

[70] Appel N, Bartenschlager R. (2006) A novel function for a micro RNA: negative regulators can do positive for the hepatitis C virus. *Hepatology*; 43:612-5.

[71] Astier-Gin T, Bellecave P, Litvak S, Ventura M. (2005) Template requirements and binding of hepatitis C virus NS5B polymerase during in vitro RNA synthesis from the 3'-end of virus minus-strand RNA. *FEBS J*; 272:3872-86.

[72] Friebe P, Lohmann V, Krieger N, Bartenschlager R. (2001) Sequences in the 5' nontranslated region of hepatitis C virus required for RNA replication. *J Virol*; 75:12047-57.

[73] Honda A, Arai Y, Hirota N, Sato T, Ikegaki J, Koizumi T, et al. (1999a) Hepatitis C virus structural proteins induce liver cell injury in transgenic mice. *J Med Virol*; 59:281-9.

[74] Ali N, Siddiqui A. (1995) Interaction of polypyrimidine tract-binding protein with the 5' noncoding region of the hepatitis C virus RNA genome and its functional requirement in internal initiation of translation. *J Virol*; 69:6367-75.

[75] Ali N, Siddiqui A. (1997) The La antigen binds 5' noncoding region of the hepatitis C virus RNA in the context of the initiator AUG codon and stimulates internal ribosome entry site-mediated translation. *Proc Natl Acad Sci U S A*; 94:2249-54.

[76] Spangberg K, Schwartz S. (1999) Poly(C)-binding protein interacts with the hepatitis C virus 5' untranslated region. *J Gen Virol*; 80 (Pt 6):1371-6.

[77] Jopling CL, Yi M, Lancaster AM, Lemon SM, Sarnow P. (2005) Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science*; 309:1577-81.

[78] Penin F, Dubuisson J, Rey FA, Moradpour D, Pawlotsky JM. (2004) Structural biology of hepatitis C virus. *Hepatology*; 39:5-19.

[79] Cocquerel L, Wychowski C, Minner F, Penin F, Dubuisson J. (2000) Charged residues in the transmembrane domains of hepatitis C virus glycoproteins play a major role in the processing, subcellular localization, and assembly of these envelope proteins. *J Virol*; 74:3623-33.

[80] Op De Beeck A, Cocquerel L, Dubuisson J. (2001) Biogenesis of hepatitis C virus envelope glycoproteins. *J Gen Virol*; 82:2589-95.

[81] Op De Beeck A, Montserret R, Duvet S, Cocquerel L, Cacan R, Barberot B, et al. (2000) The transmembrane domains of hepatitis C virus envelope glycoproteins E1 and E2 play a major role in heterodimerization. *J Biol Chem*; 275:31428-37.

[82] Cocquerel L, Duvet S, Meunier JC, Pillez A, Cacan R, Wychowski C, et al. (1999) The transmembrane domain of hepatitis C virus glycoprotein E1 is a signal for static retention in the endoplasmic reticulum. *J Virol*; 73:2641-9.

[83] Cocquerel L, Meunier JC, Pillez A, Wychowski C, Dubuisson J. (1998) A retention signal necessary and sufficient for endoplasmic reticulum localization maps to the transmembrane domain of hepatitis C virus glycoprotein E2. *J Virol*; 72:2183-91.

[84] Weiner AJ, Brauer MJ, Rosenblatt J, Richman KH, Tung J, Crawford K, et al. (1991) Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins. *Virology*; 180:842-8.

[85] von Hahn T, Yoon JC, Alter H, Rice CM, Rehermann B, Balfe P, et al. (2007) Hepatitis C virus continuously escapes from neutralizing antibody and T-cell responses during chronic infection in vivo. *Gastroenterology*; 132:667-78.

[86] Penin F, Combet C, Germanidis G, Frainais PO, Deleage G, Pawlotsky JM. (2001) Conservation of the conformation and positive charges of hepatitis C virus E2 envelope glycoprotein hypervariable region 1 points to a role in cell attachment. *J Virol*; 75:5703-10.

[87] Scarselli E, Ansuini H, Cerino R, Roccasecca RM, Acali S, Filocamo G, et al. (2002) The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *EMBO J*; 21:5017-25.

[88] Bartosch B, Dubuisson J, Cosset FL. (2003) Infectious hepatitis C virus pseudo-particles containing functional E1-E2 envelope protein complexes. *J Exp Med*; 197:633-42.

[89] Roccasecca R, Ansuini H, Vitelli A, Meola A, Scarselli E, Acali S, et al. (2003) Binding of the hepatitis C virus E2 glycoprotein to CD81 is strain specific and is modulated by a complex interplay between hypervariable regions 1 and 2. *J Virol*; 77:1856-67.

[90] Troesch M, Meunier I, Lapierre P, Lapointe N, Alvarez F, Boucher M, et al. (2006) Study of a novel hypervariable region in hepatitis C virus (HCV) E2 envelope glycoprotein. *Virology*; 352:357-67.

[91] Lavillette D, Pecheur EI, Donot P, Fresquet J, Molle J, Corbau R, et al. (2007) Characterization of fusion determinants points to the involvement of three discrete regions of both E1 and E2 glycoproteins in the membrane fusion process of hepatitis C virus. *J Virol*; 81:8752-65.

[92] Lin C, Lindenbach BD, Pragai BM, McCourt DW, Rice CM. (1994) Processing in the hepatitis C virus E2-NS2 region: identification of p7 and two distinct E2-specific products with different C termini. *J Virol*; 68:5063-73.

[93] Carrere-Kremer S, Montpellier-Pala C, Cocquerel L, Wychowski C, Penin F, Dubuisson J. (2002) Subcellular localization and topology of the p7 polypeptide of hepatitis C virus. *J Virol*; 76:3720-30.

[94] Isherwood BJ, Patel AH. (2005) Analysis of the processing and transmembrane topology of the E2p7 protein of hepatitis C virus. *J Gen Virol*; 86:667-76.

[95] Luik P, Chew C, Aittoniemi J, Chang J, Wentworth P, Jr., Dwek RA, et al. (2009) The 3-dimensional structure of a hepatitis C virus p7 ion channel by electron microscopy. *Proc Natl Acad Sci U S A*; 106:12712-6.

[96] Sakai A, Claire MS, Faulk K, Govindarajan S, Emerson SU, Purcell RH, et al. (2003) The p7 polypeptide of hepatitis C virus is critical for infectivity and contains functionally important genotype-specific sequences. *Proc Natl Acad Sci U S A*; 100:11646-51.

[97] Jones CT, Murray CL, Eastman DK, Tassello J, Rice CM. (2007) Hepatitis C virus p7 and NS2 proteins are essential for production of infectious virus. *J Virol*; 81:8374-83.

[98] Steinmann E, Penin F, Kallis S, Patel AH, Bartenschlager R, Pietschmann T. (2007) Hepatitis C virus p7 protein is crucial for assembly and release of infectious virions. *PLoS Pathog*; 3:e103.

[99] Yi M, Ma Y, Yates J, Lemon SM. (2007) Compensatory mutations in E1, p7, NS2, and NS3 enhance yields of cell culture-infectious intergenotypic chimeric hepatitis C virus. *J Virol*; 81:629-38.

[100] Griffin SD, Beales LP, Clarke DS, Worsfold O, Evans SD, Jaeger J, et al. (2003) The p7 protein of hepatitis C virus forms an ion channel that is blocked by the antiviral drug, Amantadine. *FEBS Lett*; 535:34-8.

[101] Griffin SD, Harvey R, Clarke DS, Barclay WS, Harris M, Rowlands DJ. (2004) A conserved basic loop in hepatitis C virus p7 protein is required for amantadine-sensitive ion channel activity in mammalian cells but is dispensable for localization to mitochondria. *J Gen Virol*; 85:451-61.

[102] Santolini E, Pacini L, Fipaldini C, Migliaccio G, Monica N. (1995) The NS2 protein of hepatitis C virus is a transmembrane polypeptide. *J Virol*; 69:7461-71.

[103] Yamaga AK, Ou JH. (2002) Membrane topology of the hepatitis C virus NS2 protein. *J Biol Chem*; 277:33228-34.

[104] Grakoui A, Wychowski C, Lin C, Feinstone SM, Rice CM. (1993c) Expression and identification of hepatitis C virus polyprotein cleavage products. *J Virol*; 67:1385-95.

[105] Lorenz IC, Marcotrigiano J, Dentzer TG, Rice CM. (2006) Structure of the catalytic domain of the hepatitis C virus NS2-3 protease. *Nature*; 442:831-5.

[106] Lohmann V, Korner F, Koch J, Herian U, Theilmann L, Bartenschlager R. (1999) Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science*; 285:110-3.

[107] Liu Q, Bhat RA, Prince AM, Zhang P. (1999) The hepatitis C virus NS2 protein generated by NS2-3 autocleavage is required for NS5A phosphorylation. *Biochem Biophys Res Commun*; 254:572-7.

[108] Pietschmann T, Kaul A, Koutsoudakis G, Shavinskaya A, Kallis S, Steinmann E, et al. (2006) Construction and characterization of infectious intragenotypic and intergenotypic hepatitis C virus chimeras. *Proc Natl Acad Sci U S A*; 103:7408-13.

[109] Jirasko V, Montserret R, Appel N, Janvier A, Eustachi L, Brohm C, et al. (2008) Structural and functional characterization of nonstructural protein 2 for its role in hepatitis C virus assembly. *J Biol Chem*; 283:28546-62.

[110] Dentzer TG, Lorenz IC, Evans MJ, Rice CM. (2009) Determinants of the hepatitis C virus nonstructural protein 2 protease domain required for production of infectious virus. *J Virol*; 83:12702-13.

[111] Bartenschlager R, Ahlborn-Laake L, Mous J, Jacobsen H. (1993) Nonstructural protein 3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions. *J Virol*; 67:3835-44.

[112] Bartenschlager R, Frese M, Pietschmann T. (2004) Novel insights into hepatitis C virus replication and persistence. *Adv Virus Res*; 63:71-180.

[113] Grakoui A, McCourt DW, Wychowski C, Feinstone SM, Rice CM. (1993b) A second hepatitis C virus-encoded proteinase. *Proc Natl Acad Sci U S A*; 90:10583-7.

[114] Tomei L, Failla C, Santolini E, De Francesco R, La Monica N. (1993) NS3 is a serine protease required for processing of hepatitis C virus polyprotein. *J Virol*; 67:4017-26.

[115] Kim DW, Gwack Y, Han JH, Choe J. (1995) C-terminal domain of the hepatitis C virus NS3 protein contains an RNA helicase activity. *Biochem Biophys Res Commun*; 215:160-6.

[116] Suzich JA, Tamura JK, Palmer-Hill F, Warrener P, Grakoui A, Rice CM, et al. (1993) Hepatitis C virus NS3 protein polynucleotide-stimulated nucleoside triphosphatase and comparison with the related pestivirus and flavivirus enzymes. *J Virol*; 67:6152-8.

[117] Wolk B, Sansonno D, Krausslich HG, Dammacco F, Rice CM, Blum HE, et al. (2000) Subcellular localization, stability, and trans-cleavage competence of the hepatitis C virus NS3-NS4A complex expressed in tetracycline-regulated cell lines. *J Virol*; 74:2293-304.

[118] Grakoui A, McCourt DW, Wychowski C, Feinstone SM, Rice CM. (1993) Characterization of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites. *J Virol*; 67:2832-43.

[119] Lin C, Pragai BM, Grakoui A, Xu J, Rice CM. (1994b) Hepatitis C virus NS3 serine proteinase: trans-cleavage requirements and processing kinetics. *J Virol*; 68:8147-57.

[120] Bartenschlager R, Ahlborn-Laake L, Mous J, Jacobsen H. (1994) Kinetic and structural analyses of hepatitis C virus polyprotein processing. *J Virol*; 68:5045-55.

[121] Failla C, Tomei L, De Francesco R. (1994) Both NS3 and NS4A are required for proteolytic processing of hepatitis C virus nonstructural proteins. *J Virol*; 68:3753-60.

[122] Serebrov V, Pyle AM. (2004) Periodic cycles of RNA unwinding and pausing by hepatitis C virus NS3 helicase. *Nature*; 430:476-80.

[123] Banerjee R, Dasgupta A. (2001) Specific interaction of hepatitis C virus protease/helicase NS3 with the 3'-terminal sequences of viral positive- and negative-strand RNA. *J Virol*; 75:1708-21.

[124] Lam AM, Frick DN. (2006) Hepatitis C virus subgenomic replicon requires an active NS3 RNA helicase. *J Virol*; 80:404-11.

[125] Tellinghuisen TL, Rice CM. (2002) Interaction between hepatitis C virus proteins and host cell factors. *Curr Opin Microbiol*; 5:419-27.

[126] Borowski P, Heiland M, Oehlmann K, Becker B, Kornetzky L, Feucht H, et al. (1996) Non-structural protein 3 of hepatitis C virus inhibits phosphorylation mediated by cAMP-dependent protein kinase. *Eur J Biochem*; 237:611-8.

[127] Sakamuro D, Furukawa T, Takegami T. (1995) Hepatitis C virus nonstructural protein NS3 transforms NIH 3T3 cells. *J Virol*; 69:3893-6.

[128] Ma Y, Yates J, Liang Y, Lemon SM, Yi M. (2008) NS3 helicase domains involved in infectious intracellular hepatitis C virus particle assembly. *J Virol*; 82:7624-39.

[129] Lin C, Thomson JA, Rice CM. (1995) A central region in the hepatitis C virus NS4A protein allows formation of an active NS3-NS4A serine proteinase complex in vivo and in vitro. *J Virol*; 69:4373-80.

[130] Tomei L, Failla C, Vitale RL, Bianchi E, De Francesco R. (1996) A central hydrophobic domain of the hepatitis C virus NS4A protein is necessary and sufficient for the activation of the NS3 protease. *J Gen Virol*; 77 (Pt 5):1065-70.

[131] Tanji Y, Hijikata M, Satoh S, Kaneko T, Shimotohno K. (1995) Hepatitis C virus-encoded nonstructural protein NS4A has versatile functions in viral protein processing. *J Virol*; 69:1575-81.

[132] Florese RH, Nagano-Fujii M, Iwanaga Y, Hidajat R, Hotta H. (2002) Inhibition of protein synthesis by the nonstructural proteins NS4A and NS4B of hepatitis C virus. *Virus Res*; 90:119-31.

[133] Kato J, Kato N, Yoshida H, Ono-Nita SK, Shiratori Y, Omata M. (2002) Hepatitis C virus NS4A and NS4B proteins suppress translation in vivo. *J Med Virol*; 66:187-99.

[134] Doedens JR, Kirkegaard K. (1995) Inhibition of cellular protein secretion by poliovirus proteins 2B and 3A. *EMBO J*; 14:894-907.

[135] Konan KV, Giddings TH, Jr., Ikeda M, Li K, Lemon SM, Kirkegaard K. (2003) Nonstructural protein precursor NS4A/B from hepatitis C virus alters function and ultrastructure of host secretory apparatus. *J Virol*; 77:7843-55.

[136] Jourdan N, Maurice M, Delautier D, Quero AM, Servin AL, Trugnan G. (1997) Rotavirus is released from the apical surface of cultured human intestinal cells through nonconventional vesicular transport that bypasses the Golgi apparatus. *J Virol*; 71:8268-78.

[137] Kaneko T, Tanji Y, Satoh S, Hijikata M, Asabe S, Kimura K, et al. (1994) Production of two phosphoproteins from the NS5A region of the hepatitis C viral genome. *Biochem Biophys Res Commun*; 205:320-6.

[138] Reed KE, Xu J, Rice CM. (1997) Phosphorylation of the hepatitis C virus NS5A protein in vitro and in vivo: properties of the NS5A-associated kinase. *J Virol*; 71:7187-97.

[139] Tanji Y, Kaneko T, Satoh S, Shimotohno K. (1995b) Phosphorylation of hepatitis C virus-encoded nonstructural protein NS5A. *J Virol*; 69:3980-6.

[140] Hugle T, Fehrmann F, Bieck E, Kohara M, Krausslich HG, Rice CM, et al. (2001) The hepatitis C virus nonstructural protein 4B is an integral endoplasmic reticulum membrane protein. *Virology*; 284:70-81.

[141] Park JS, Yang JM, Min MK. (2000) Hepatitis C virus nonstructural protein NS4B transforms NIH3T3 cells in cooperation with the Ha-ras oncogene. *Biochem Biophys Res Commun*; 267:581-7.

[142] Egger D, Wolk B, Gosert R, Bianchi L, Blum HE, Moradpour D, et al. (2002) Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex. *J Virol*; 76:5974-84.

[143] Gouttenoire J, Castet V, Montserret R, Arora N, Raussens V, Ruysschaert JM, et al. (2009) Identification of a novel determinant for membrane association in hepatitis C virus nonstructural protein 4B. *J Virol*; 83:6257-68.

[144] Gouttenoire J, Montserret R, Kennel A, Penin F, Moradpour D. (2009b) An amphipathic alpha-helix at the C terminus of hepatitis C virus nonstructural protein 4B mediates membrane association. *J Virol*; 83:11378-84.

[145] Lundin M, Lindstrom H, Gronwall C, Persson MA. (2006) Dual topology of the processed hepatitis C virus protein NS4B is influenced by the NS5A protein. *J Gen Virol*; 87:3263-72.

[146] Yu GY, Lee KJ, Gao L, Lai MM. (2006) Palmitoylation and polymerization of hepatitis C virus NS4B protein. *J Virol*; 80:6013-23.

[147] Einav S, Elazar M, Danieli T, Glenn JS. (2004) A nucleotide binding motif in hepatitis C virus (HCV) NS4B mediates HCV RNA replication. *J Virol*; 78:11288-95.

[148] Brass V, Bieck E, Montserret R, Wolk B, Hellings JA, Blum HE, et al. (2002) An amino-terminal amphipathic alpha-helix mediates membrane association of the hepatitis C virus nonstructural protein 5A. *J Biol Chem*; 277:8130-9.

[149] Tellinghuisen TL, Marcotrigiano J, Gorbalenya AE, Rice CM. (2004) The NS5A protein of hepatitis C virus is a zinc metalloprotein. *J Biol Chem*; 279:48576-87.

[150] Tellinghuisen TL, Marcotrigiano J, Rice CM. (2005) Structure of the zinc-binding domain of an essential component of the hepatitis C virus replicase. *Nature*; 435:374-9.

[151] Miyanari Y, Atsuzawa K, Usuda N, Watashi K, Hishiki T, Zayas M, et al. (2007) The lipid droplet is an important organelle for hepatitis C virus production. *Nat Cell Biol*; 9:1089-97.

[152] Huang L, Hwang J, Sharma SD, Hargittai MR, Chen Y, Arnold JJ, et al. (2005) Hepatitis C virus nonstructural protein 5A (NS5A) is an RNA-binding protein. *J Biol Chem*; 280:36417-28.

[153] Gale M, Jr., Blakely CM, Kwieciszewski B, Tan SL, Dossett M, Tang NM, et al. (1998) Control of PKR protein kinase by hepatitis C virus nonstructural 5A protein: molecular mechanisms of kinase regulation. *Mol Cell Biol*; 18:5208-18.

[154] Tellinghuisen TL, Foss KL, Treadaway JC, Rice CM. (2008b) Identification of residues required for RNA replication in domains II and III of the hepatitis C virus NS5A protein. *J Virol*; 82:1073-83.

[155] Lan KH, Lan KL, Lee WP, Sheu ML, Chen MY, Lee YL, et al. (2007) HCV NS5A inhibits interferon-alpha signaling through suppression of STAT1 phosphorylation in hepatocyte-derived cell lines. *J Hepatol*; 46:759-67.

[156] Appel N, Pietschmann T, Bartenschlager R. (2005b) Mutational analysis of hepatitis C virus nonstructural protein 5A: potential role of differential phosphorylation in RNA replication and identification of a genetically flexible domain. *J Virol*; 79:3187-94.

[157] Appel N, Zayas M, Miller S, Krijnse-Locker J, Schaller T, Friebe P, et al. (2008) Essential role of domain III of nonstructural protein 5A for hepatitis C virus infectious particle assembly. *PLoS Pathog*; 4:e1000035.

[158] Evans MJ, Rice CM, Goff SP. (2004) Phosphorylation of hepatitis C virus nonstructural protein 5A modulates its protein interactions and viral RNA replication. *Proc Natl Acad Sci U S A*; 101:13038-43.

[159] Tellinghuisen TL, Foss KL, Treadaway J. (2008) Regulation of hepatitis C virion production via phosphorylation of the NS5A protein. *PLoS Pathog*; 4:e1000032.

[160] Wang C, Gale M, Jr., Keller BC, Huang H, Brown MS, Goldstein JL, et al. (2005) Identification of FBL2 as a geranylgeranylated cellular protein required for hepatitis C virus RNA replication. *Mol Cell*; 18:425-34.

[161] Ye J, Wang C, Sumpter R, Jr., Brown MS, Goldstein JL, Gale M, Jr. (2003) Disruption of hepatitis C virus RNA replication through inhibition of host protein geranylgeranylation. *Proc Natl Acad Sci U S A*; 100:15865-70.

[162] Macdonald A, Harris M. (2004) Hepatitis C virus NS5A: tales of a promiscuous protein. *J Gen Virol*; 85:2485-502.

[163] Behrens SE, Tomei L, De Francesco R. (1996) Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus. *EMBO J*; 15:12-22.

[164] Lohmann V, Korner F, Herian U, Bartenschlager R. (1997) Biochemical properties of hepatitis C virus NS5B RNA-dependent RNA polymerase and identification of amino acid sequence motifs essential for enzymatic activity. *J Virol*; 71:8416-28.

[165] Poch O, Sauvaget I, Delarue M, Tordo N. (1989) Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J*; 8:3867-74.

[166] Ago H, Adachi T, Yoshida A, Yamamoto M, Habuka N, Yatsunami K, et al. (1999) Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Structure*; 7:1417-26.

[167] Bressanelli S, Tomei L, Roussel A, Incitti I, Vitale RL, Mathieu M, et al. (1999) Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Proc Natl Acad Sci U S A*; 96:13034-9.

[168] Lesburg CA, Cable MB, Ferrari E, Hong Z, Mannarino AF, Weber PC. (1999) Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nat Struct Biol*; 6:937-43.

[169] Bressanelli S, Tomei L, Rey FA, De Francesco R. (2002) Structural analysis of the hepatitis C virus RNA polymerase in complex with ribonucleotides. *J Virol*; 76:3482-92.

[170] Hong Z, Cameron CE, Walker MP, Castro C, Yao N, Lau JY, et al. (2001) A novel mechanism to ensure terminal initiation by hepatitis C virus NS5B polymerase. *Virology*; 285:6-11.

[171] Ivashkina N, Wolk B, Lohmann V, Bartenschlager R, Blum HE, Penin F, et al. (2002) The hepatitis C virus RNA-dependent RNA polymerase membrane insertion sequence is a transmembrane segment. *J Virol*; 76:13088-93.

[172] Schmidt-Mende J, Bieck E, Hugle T, Penin F, Rice CM, Blum HE, et al. (2001) Determinants for membrane association of the hepatitis C virus RNA-dependent RNA polymerase. *J Biol Chem*; 276:44052-63.

[173] Oh JW, Ito T, Lai MM. (1999) A recombinant hepatitis C virus RNA-dependent RNA polymerase capable of copying the full-length viral RNA. *J Virol*; 73:7694-702.

[174] De Francesco R, Migliaccio G. (2005) Challenges and successes in developing new therapies for hepatitis C. *Nature*; 436:953-60.

[175] Huang Z, Murray MG, Secrist JA, 3rd. (2006) Recent development of therapeutics for chronic HCV infection. *Antiviral Res*; 71:351-62.

[176] Tanaka T, Kato N, Cho MJ, Sugiyama K, Shimotohno K. (1996) Structure of the 3' terminus of the hepatitis C virus genome. *J Virol*; 70:3307-12.

[177] Ito T, Lai MM. (1997) Determination of the secondary structure of and cellular protein binding to the 3'-untranslated region of the hepatitis C virus RNA genome. *J Virol*; 71:8698-706.

[178] Kolykhalov AA, Mihalik K, Feinstone SM, Rice CM. (2000) Hepatitis C virus-encoded enzymatic activities and conserved RNA elements in the 3' nontranslated region are essential for virus replication in vivo. *J Virol*; 74:2046-51.

[179] Yanagi M, St Claire M, Emerson SU, Purcell RH, Bukh J. (1999) In vivo analysis of the 3' untranslated region of the hepatitis C virus after in vitro mutagenesis of an infectious cDNA clone. *Proc Natl Acad Sci U S A*; 96:2291-5.

[180] Yi M, Lemon SM. (2003) 3' nontranslated RNA signals required for replication of hepatitis C virus RNA. *J Virol*; 77:3557-68.

[181] Yi M, Lemon SM. (2003) Structure-function analysis of the 3' stem-loop of hepatitis C virus genomic RNA and its role in viral RNA replication. *RNA*; 9:331-45.

[182] Ito T, Tahara SM, Lai MM. (1998) The 3'-untranslated region of hepatitis C virus RNA enhances translation from an internal ribosomal entry site. *J Virol*; 72:8789-96.

[183] Song M, Balakrishnan M, Chen Y, Roques BP, Bambara RA. (2006) Stimulation of HIV-1 minus strand strong stop DNA transfer by genomic sequences 3' of the primer binding site. *J Biol Chem*; 281:24227-35.

[184] Gontarek RR, Gutshall LL, Herold KM, Tsai J, Sathe GM, Mao J, et al. (1999) hnRNP C and polypyrimidine tract-binding protein specifically interact with the pyrimidine-rich region within the 3'NTR of the HCV RNA genome. *Nucleic Acids Res*; 27:1457-63.

[185] Luo G. (1999) Cellular proteins bind to the poly(U) tract of the 3' untranslated region of hepatitis C virus RNA genome. *Virology*; 256:105-18.

[186] Tsuchihara K, Tanaka T, Hijikata M, Kuge S, Toyoda H, Nomoto A, et al. (1997) Specific interaction of polypyrimidine tract-binding protein with the extreme 3'-terminal structure of the hepatitis C virus genome, the 3'X. *J Virol*; 71:6720-6.

[187] Petrik J, Parker H, Alexander GJ. (1999) Human hepatic glyceraldehyde-3-phosphate dehydrogenase binds to the poly(U) tract of the 3' non-coding region of hepatitis C virus genomic RNA. *J Gen Virol*; 80 (Pt 12):3109-13.

[188] Wood J, Frederickson RM, Fields S, Patel AH. (2001) Hepatitis C virus 3'X region interacts with human ribosomal proteins. *J Virol*; 75:1348-58.

[189] Cheng JC, Chang MF, Chang SC. (1999) Specific interaction between the hepatitis C virus NS5B RNA polymerase and the 3' end of the viral RNA. *J Virol*; 73:7044-9.

[190] Wakita T, Pietschmann T, Kato T, Date T, Miyamoto M, Zhao Z, et al. (2005) Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat Med*; 11:791-6.

[191] Buckwold VE, Beer BE, Donis RO. (2003) Bovine viral diarrhea virus as a surrogate model of hepatitis C virus for the evaluation of antiviral agents. *Antiviral Res*; 60:1-15.

[192] Meyers G, Tautz N, Becher P, Thiel HJ, Kummerer BM. (1996) Recovery of cytopathogenic and noncytopathogenic bovine viral diarrhea viruses from cDNA constructs. *J Virol*; 70:8606-13.

[193] Moormann RJ, van Gennip HG, Miedema GK, Hulst MM, van Rijn PA. (1996) Infectious RNA transcribed from an engineered full-length cDNA template of the genome of a pestivirus. *J Virol*; 70:763-70.

[194] Beames B, Chavez D, Lanford RE. (2001) GB virus B as a model for hepatitis C virus. *ILAR J*; 42:152-60.

[195] Muerhoff AS, Leary TP, Simons JN, Pilot-Matias TJ, Dawson GJ, Erker JC, et al. (1995) Genomic organization of GB viruses A and B: two new members of the Flaviviridae associated with GB agent hepatitis. *J Virol*; 69:5621-30.

[196] Hope RG, Murphy DJ, McLauchlan J. (2002) The domains required to direct core proteins of hepatitis C virus and GB virus-B to lipid droplets share common features with plant oleosin proteins. *J Biol Chem*; 277:4261-70.

[197] Grakoui A, Hanson HL, Rice CM. (2001) Bad time for Bonzo? Experimental models of hepatitis C virus infection, replication, and pathogenesis. *Hepatology*; 33:489-95.

[198] Farci P, Purcell RH. (1998) *Viral Hepatitis*. Churchill Livingstone, Edinburgh.

[199] Yanagi M, Purcell RH, Emerson SU, Bukh J. (1997) Transcripts from a single full-length cDNA clone of hepatitis C virus are infectious when directly transfected into the liver of a chimpanzee. *Proc Natl Acad Sci U S A*; 94:8738-43.

[200] Bradley DW. (2000) Studies of non-A, non-B hepatitis and characterization of the hepatitis C virus in chimpanzees. *Curr Top Microbiol Immunol*; 242:1-23.

[201] Bassett SE, Brasky KM, Lanford RE. (1998) Analysis of hepatitis C virus-inoculated chimpanzees reveals unexpected clinical profiles. *J Virol*; 72:2589-99.

[202] Mercer DF, Schiller DE, Elliott JF, Douglas DN, Hao C, Rinfret A, et al. (2001) Hepatitis C virus replication in mice with chimeric human livers. *Nat Med*; 7:927-33.

[203] Turrini P, Sasso R, Germoni S, Marcucci I, Celluci A, Di Marco A, et al. (2006) Development of humanized mice for the study of hepatitis C virus infection. *Transplant Proc*; 38:1181-4.

[204] Meuleman P, Libbrecht L, De Vos R, de Hemptinne B, Gevaert K, Vandekerckhove J, et al. (2005) Morphological and biochemical characterization of a human liver in a uPA-SCID mouse chimera. *Hepatology*; 41:847-56.

[205] Moriya K, Fujie H, Shintani Y, Yotsuyanagi H, Tsutsumi T, Ishibashi K, et al. (1998) The core protein of hepatitis C virus induces hepatocellular carcinoma in transgenic mice. *Nat Med*; 4:1065-7.

[206] Moriya K, Yotsuyanagi H, Shintani Y, Fujie H, Ishibashi K, Matsuura Y, et al. (1997) Hepatitis C virus core protein induces hepatic steatosis in transgenic mice. *J Gen Virol*; 78 (Pt 7):1527-31.

[207] Hiraga N, Imamura M, Tsuge M, Noguchi C, Takahashi S, Iwao E, et al. (2007) Infection of human hepatocyte chimeric mouse with genetically engineered hepatitis C virus and its susceptibility to interferon. *FEBS Lett*; 581:1983-7.

[208] Kimura T, Imamura M, Hiraga N, Hatakeyama T, Miki D, Noguchi C, et al. (2008) Establishment of an infectious genotype 1b hepatitis C virus clone in human hepatocyte chimeric mice. *J Gen Virol*; 89:2108-13.

[209] Feinstone SM, Alter HJ, Dienes HP, Shimizu Y, Popper H, Blackmore D, et al. (1981) Non-A, non-B hepatitis in chimpanzees and marmosets. *J Infect Dis*; 144:588-98.

[210] Karayiannis P, Scheuer PJ, Bamber M, Cohn D, Hurn BA, Thomas HC. (1983) Experimental infection of Tamarins with human non-A, non-B hepatitis virus. *J Med Virol*; 11:251-6.

[211] Xie ZC, Riezu-Boj JI, Lasarte JJ, Guillen J, Su JH, Civeira MP, et al. (1998) Transmission of hepatitis C virus infection to tree shrews. *Virology*; 244:513-20.

[212] Zhao X, Tang ZY, Klumpp B, Wolff-Vorbeck G, Barth H, Levy S, et al. (2002) Primary hepatocytes of Tupaia belangeri as a potential model for hepatitis C virus infection. *J Clin Invest*; 109:221-32.

[213] Michalak JP, Wychowski C, Choukhi A, Meunier JC, Ung S, Rice CM, et al. (1997) Characterization of truncated forms of hepatitis C virus glycoproteins. *J Gen Virol*; 78 (Pt 9):2299-306.

[214] Flint M, Maidens C, Loomis-Price LD, Shotton C, Dubuisson J, Monk P, et al. (1999) Characterization of hepatitis C virus E2 glycoprotein interaction with a putative cellular receptor, CD81. *J Virol*; 73:6235-44.

[215] Pileri P, Uematsu Y, Campagnoli S, Galli G, Falugi F, Petracca R, et al. (1998) Binding of hepatitis C virus to CD81. *Science*; 282:938-41.

[216] Barth H, Schafer C, Adah MI, Zhang F, Linhardt RJ, Toyoda H, et al. (2003) Cellular binding of hepatitis C virus envelope glycoprotein E2 requires cell surface heparan sulfate. *J Biol Chem*; 278:41003-12.

[217] Baumert TF, Ito S, Wong DT, Liang TJ. (1998) Hepatitis C virus structural proteins assemble into viruslike particles in insect cells. *J Virol*; 72:3827-36.

[218] Baumert TF, Vergalla J, Satoi J, Thomson M, Lechmann M, Herion D, et al. (1999) Hepatitis C virus-like particles synthesized in insect cells as a potential vaccine candidate. *Gastroenterology*; 117:1397-407.

[219] Triyatni M, Saunier B, Maruvada P, Davis AR, Ulianich L, Heller T, et al. (2002) Interaction of hepatitis C virus-like particles and cells: a model system for studying viral binding and entry. *J Virol*; 76:9335-44.

[220] Steinmann D, Barth H, Gissler B, Schurmann P, Adah MI, Gerlach JT, et al. (2004) Inhibition of hepatitis C virus-like particle binding to target cells by antiviral antibodies in acute and chronic hepatitis C. *J Virol*; 78:9030-40.

[221] Bartosch B, Bukh J, Meunier JC, Granier C, Engle RE, Blackwelder WC, et al. (2003b) In vitro assay for neutralizing antibody to hepatitis C virus: evidence for broadly conserved neutralization epitopes. *Proc Natl Acad Sci U S A*; 100:14199-204.

[222] Hsu M, Zhang J, Flint M, Logvinoff C, Cheng-Mayer C, Rice CM, et al. (2003) Hepatitis C virus glycoproteins mediate pH-dependent cell entry of pseudotyped retroviral particles. *Proc Natl Acad Sci U S A*; 100:7271-6.

[223] Voisset C, Dubuisson J. (2004) Functional hepatitis C virus envelope glycoproteins. *Biol Cell*; 96:413-20.

[224] Op De Beeck A, Voisset C, Bartosch B, Ciczora Y, Cocquerel L, Keck Z, et al. (2004) Characterization of functional hepatitis C virus envelope glycoproteins. *J Virol*; 78:2994-3002.

[225] Regeard M, Lepere C, Trotard M, Gripon P, Le Seyec J. (2007) Recent contributions of in vitro models to our understanding of hepatitis C virus life cycle. *FEBS J*; 274:4705-18.

[226] Bartenschlager R, Lohmann V. (2001) Novel cell culture systems for the hepatitis C virus. *Antiviral Res*; 52:1-17.

[227] Blight KJ, McKeating JA, Rice CM. (2002) Highly permissive cell lines for subgenomic and genomic hepatitis C virus RNA replication. *J Virol*; 76:13001-14.

[228] Blight KJ, Kolykhalov AA, Rice CM. (2000) Efficient initiation of HCV RNA replication in cell culture. *Science*; 290:1972-4.

[229] Krieger M. (2001) Scavenger receptor class B type I is a multiligand HDL receptor that influences diverse physiologic systems. *J Clin Invest*; 108:793-7.

[230] Lohmann V, Korner F, Dobierzewska A, Bartenschlager R. (2001) Mutations in hepatitis C virus RNAs conferring cell culture adaptation. *J Virol*; 75:1437-49.

[231] Kato T, Furusaka A, Miyamoto M, Date T, Yasui K, Hiramoto J, et al. (2001) Sequence analysis of hepatitis C virus isolated from a fulminant hepatitis patient. *J Med Virol*; 64:334-9.

[232] Lindenbach BD, Evans MJ, Syder AJ, Wolk B, Tellinghuisen TL, Liu CC, et al. (2005) Complete replication of hepatitis C virus in cell culture. *Science*; 309:623-6.

[233] Zhong J, Gastaminza P, Cheng G, Kapadia S, Kato T, Burton DR, et al. (2005) Robust hepatitis C virus infection in vitro. *Proc Natl Acad Sci U S A*; 102:9294-9.

[234] Kato T, Date T, Miyamoto M, Furusaka A, Tokushige K, Mizokami M, et al. (2003) Efficient replication of the genotype 2a hepatitis C virus subgenomic replicon. *Gastroenterology*; 125:1808-17.

[235] Koutsoudakis G, Kaul A, Steinmann E, Kallis S, Lohmann V, Pietschmann T, et al. (2006) Characterization of the early steps of hepatitis C virus infection by using luciferase reporter viruses. *J Virol*; 80:5308-20.

[236] Zeisel MB, Koutsoudakis G, Schnober EK, Haberstroh A, Blum HE, Cosset FL, et al. (2007) Scavenger receptor class B type I is a key host factor for hepatitis C virus infection required for an entry step closely linked to CD81. *Hepatology*; 46:1722-31.

[237] Dubuisson J. (2007) Hepatitis C virus proteins. *World J Gastroenterol*; 13:2406-15.

[238] Evans MJ, von Hahn T, Tscherne DM, Syder AJ, Panis M, Wolk B, et al. (2007) Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry. *Nature*; 446:801-5.

[239] Meertens L, Bertaux C, Cukierman L, Cormier E, Lavillette D, Cosset FL, et al. (2008) The tight junction proteins claudin-1, -6, and -9 are entry cofactors for hepatitis C virus. *J Virol*; 82:3555-60.

[240] Zheng A, Yuan F, Li Y, Zhu F, Hou P, Li J, et al. (2007) Claudin-6 and claudin-9 function as additional coreceptors for hepatitis C virus. *J Virol*; 81:12465-71.

[241] Liu S, Yang W, Shen L, Turner JR, Coyne CB, Wang T. (2009b) Tight junction proteins claudin-1 and occludin control hepatitis C virus entry and are downregulated during infection to prevent superinfection. *J Virol*; 83:2011-4.

[242] Ploss A, Evans MJ, Gaysinskaya VA, Panis M, You H, de Jong YP, et al. (2009) Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. *Nature*; 457:882-6.

[243] Barth H, Schnober EK, Zhang F, Linhardt RJ, Depla E, Boson B, et al. (2006) Viral and cellular determinants of the hepatitis C virus envelope-heparan sulfate interaction. *J Virol*; 80:10579-90.

[244] Callens N, Ciczora Y, Bartosch B, Vu-Dac N, Cosset FL, Pawlotsky JM, et al. (2005) Basic residues in hypervariable region 1 of hepatitis C virus envelope glycoprotein e2 contribute to virus entry. *J Virol*; 79:15331-41.

[245] Andreo U, Maillard P, Kalinina O, Walic M, Meurs E, Martinot M, et al. (2007) Lipoprotein lipase mediates hepatitis C virus (HCV) cell entry and inhibits HCV infection. *Cell Microbiol*; 9:2445-56.

[246] Agnello V, Abel G, Elfahal M, Knight GB, Zhang QX. (1999) Hepatitis C virus and other flaviviridae viruses enter cells via low density lipoprotein receptor. *Proc Natl Acad Sci U S A*; 96:12766-71.

[247] Thomssen R, Bonk S, Propfe C, Heermann KH, Kochel HG, Uy A. (1992) Association of hepatitis C virus in human sera with beta-lipoprotein. *Med Microbiol Immunol*; 181:293-300.

[248] Wunschmann S, Medh JD, Klinzmann D, Schmidt WN, Stapleton JT. (2000) Characterization of hepatitis C virus (HCV) and HCV E2 interactions with CD81 and the low-density lipoprotein receptor. *J Virol*; 74:10055-62.

[249] Saunier B, Triyatni M, Ulianich L, Maruvada P, Yen P, Kohn LD. (2003) Role of the asialoglycoprotein receptor in binding and entry of hepatitis C virus structural proteins in cultured human hepatocytes. *J Virol*; 77:546-59.

[250] Gardner JP, Durso RJ, Arrigale RR, Donovan GP, Maddon PJ, Dragic T, et al. (2003) L-SIGN (CD 209L) is a liver-specific capture receptor for hepatitis C virus. *Proc Natl Acad Sci U S A*; 100:4498-503.

[251] Lozach PY, Lortat-Jacob H, de Lacroix de Lavalette A, Staropoli I, Foung S, Amara A, et al. (2003) DC-SIGN and L-SIGN are high affinity binding receptors for hepatitis C virus glycoprotein E2. *J Biol Chem*; 278:20358-66.

[252] Pohlmann S, Zhang J, Baribaud F, Chen Z, Leslie GJ, Lin G, et al. (2003) Hepatitis C virus glycoproteins interact with DC-SIGN and DC-SIGNR. *J Virol*; 77:4070-80.

[253] Geijtenbeek TB, Torensma R, van Vliet SJ, van Duijnhoven GC, Adema GJ, van Kooyk Y, et al. (2000) Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. *Cell*; 100:575-85.

[254] Cormier EG, Durso RJ, Tsamis F, Boussemart L, Manix C, Olson WC, et al. (2004) L-SIGN (CD209L) and DC-SIGN (CD209) mediate transinfection of liver cells by hepatitis C virus. *Proc Natl Acad Sci U S A*; 101:14067-72.

[255] Lozach PY, Amara A, Bartosch B, Virelizier JL, Arenzana-Seisdedos F, Cosset FL, et al. (2004) C-type lectins L-SIGN and DC-SIGN capture and transmit infectious hepatitis C virus pseudotype particles. *J Biol Chem*; 279:32035-45.

[256] Delandre C, Penabaz TR, Passarelli AL, Chapes SK, Clem RJ. (2009) Mutation of juxtamembrane cysteines in the tetraspanin CD81 affects palmitoylation and alters interaction with other proteins at the cell surface. *Exp Cell Res*; 315:1953-63.

[257] Zhang J, Randall G, Higginbottom A, Monk P, Rice CM, McKeating JA. (2004) CD81 is required for hepatitis C virus glycoprotein-mediated viral infection. *J Virol*; 78:1448-55.

[258] Koutsoudakis G, Herrmann E, Kallis S, Bartenschlager R, Pietschmann T. (2007) The level of CD81 cell surface expression is a key determinant for productive entry of hepatitis C virus into host cells. *J Virol*; 81:588-98.

[259] Drummer HE, Wilson KA, Poumbourios P. (2002) Identification of the hepatitis C virus E2 glycoprotein binding site on the large extracellular loop of CD81. *J Virol*; 76:11143-7.

[260] Acton S, Rigotti A, Landschulz KT, Xu S, Hobbs HH, Krieger M. (1996) Identification of scavenger receptor SR-BI as a high density lipoprotein receptor. *Science*; 271:518-20.

[261] Acton SL, Scherer PE, Lodish HF, Krieger M. (1994) Expression cloning of SR-BI, a CD36-related class B scavenger receptor. *J Biol Chem*; 269:21003-9.

[262] Landschulz KT, Pathak RK, Rigotti A, Krieger M, Hobbs HH. (1996) Regulation of scavenger receptor, class B, type I, a high density lipoprotein receptor, in liver and steroidogenic tissues of the rat. *J Clin Invest*; 98:984-95.

[263] Van Eck M, Hoekstra M, Out R, Bos IS, Kruijt JK, Hildebrand RB, et al. (2008) Scavenger receptor BI facilitates the metabolism of VLDL lipoproteins in vivo. *J Lipid Res*; 49:136-46.

[264] Cocquerel L, Voisset C, Dubuisson J. (2006) Hepatitis C virus entry: potential receptors and their biological functions. *J Gen Virol*; 87:1075-84.

[265] Barth H, Cerino R, Arcuri M, Hoffmann M, Schurmann P, Adah MI, et al. (2005) Scavenger receptor class B type I and hepatitis C virus infection of primary tupaia hepatocytes. *J Virol*; 79:5774-85.

[266] Bartosch B, Vitelli A, Granier C, Goujon C, Dubuisson J, Pascale S, et al. (2003c) Cell entry of hepatitis C virus requires a set of co-receptors that include the CD81 tetraspanin and the SR-B1 scavenger receptor. *J Biol Chem*; 278:41624-30.

[267] Schwarz AK, Grove J, Hu K, Mee CJ, Balfe P, McKeating JA. (2009) Hepatoma cell density promotes claudin-1 and scavenger receptor BI expression and hepatitis C virus internalization. *J Virol*; 83:12407-14.

[268] Dreux M, Cosset FL. (2006) HCV and lipoproteins: is oxLDL an Achilles' heel of the Trojan horse? *Hepatology*; 43:903-5.

[269] Voisset C, Callens N, Blanchard E, Op De Beeck A, Dubuisson J, Vu-Dac N. (2005) High density lipoproteins facilitate hepatitis C virus entry through the scavenger receptor class B type I. *J Biol Chem*; 280:7793-9.

[270] Bartosch B, Verney G, Dreux M, Donot P, Morice Y, Penin F, et al. (2005) An interplay between hypervariable region 1 of the hepatitis C virus E2 glycoprotein, the scavenger receptor BI, and high-density lipoprotein promotes both enhancement of infection and protection against neutralizing antibodies. *J Virol*; 79:8217-29.

[271] Voisset C, Op de Beeck A, Horellou P, Dreux M, Gustot T, Duverlie G, et al. (2006) High-density lipoproteins reduce the neutralizing effect of hepatitis C virus (HCV)-infected patient antibodies by promoting HCV entry. *J Gen Virol*; 87:2577-81.

[272] Vishnyakova TG, Bocharov AV, Baranova IN, Chen Z, Remaley AT, Csako G, et al. (2003) Binding and internalization of lipopolysaccharide by Cla-1, a human orthologue of rodent scavenger receptor B1. *J Biol Chem*; 278:22771-80.

[273] Anderson JM, Van Itallie CM. (2008) Tight junctions. *Curr Biol*; 18:R941-3.

[274] Chiba H, Osanai M, Murata M, Kojima T, Sawada N. (2008) Transmembrane proteins of tight junctions. *Biochim Biophys Acta*; 1778:588-600.

[275] Shen L, Weber CR, Turner JR. (2008) The tight junction protein complex undergoes rapid and continuous molecular remodeling at steady state. *The Journal of cell biology*; 181:683-95.

[276] Van Itallie CM, Anderson JM. (2006) Claudins and epithelial paracellular transport. *Annu Rev Physiol*; 68:403-29.

[277] Furuse M, Fujita K, Hiiragi T, Fujimoto K, Tsukita S. (1998) Claudin-1 and -2: novel integral membrane proteins localizing at tight junctions with no sequence similarity to occludin. *J Cell Biol*; 141:1539-50.

[278] Lal-Nag M, Morin PJ. (2009) The claudins. *Genome Biol*; 10:235.

[279] Eyre NS, Baumert TF, Beard MR. (2009) Closing the gap: the tight junction protein occludin and hepatitis C virus entry. *Hepatology*; 49:1770-2.

[280] Cukierman L, Meertens L, Bertaux C, Kajumo F, Dragic T. (2009) Residues in a highly conserved claudin-1 motif are required for hepatitis C virus entry and mediate the formation of cell-cell contacts. *J Virol*; 83:5477-84.

[281] Tellinghuisen TL, Evans MJ, von Hahn T, You S, Rice CM. (2007) Studying hepatitis C virus: making the best of a bad virus. *Journal of virology*; 81:8853-67.

[282] Yang W, Qiu C, Biswas N, Jin J, Watkins SC, Montelaro RC, et al. (2008b) Correlation of the tight junction-like distribution of Claudin-1 to the cellular tropism of hepatitis C virus. *J Biol Chem*; 283:8643-53.

[283] Benedicto I, Molina-Jimenez F, Bartosch B, Cosset FL, Lavillette D, Prieto J, et al. (2009) The tight junction-associated protein occludin is required for a postbinding step in hepatitis C virus entry and infection. *J Virol*; 83:8012-20.

[284] Tsukita S, Furuse M. (1999) Occludin and claudins in tight-junction strands: leading or supporting players? *Trends Cell Biol*; 9:268-73.

[285] Brazzoli M, Bianchi A, Filippini S, Weiner A, Zhu Q, Pizza M, et al. (2008) CD81 is a central regulator of cellular events required for hepatitis C virus infection of human hepatocytes. *J Virol*; 82:8316-29.

[286] Blanchard E, Belouzard S, Goueslain L, Wakita T, Dubuisson J, Wychowski C, et al. (2006) Hepatitis C virus entry depends on clathrin-mediated endocytosis. *J Virol*; 80:6964-72.

[287] Meertens L, Bertaux C, Dragic T. (2006) Hepatitis C virus entry requires a critical postinternalization step and delivery to early endosomes via clathrin-coated vesicles. *J Virol*; 80:11571-8.

[288] Cosa G, Zeng Y, Liu HW, Landes CF, Makarov DE, Musier-Forsyth K, et al. (2006) Evidence for non-two-state kinetics in the nucleocapsid protein chaperoned opening of DNA hairpins. *J Phys Chem B*; 110:2419-26.

[289] Sollner TH. (2004) Intracellular and viral membrane fusion: a uniting mechanism. *Curr Opin Cell Biol*; 16:429-35.

[290] Tscherne DM, Jones CT, Evans MJ, Lindenbach BD, McKeating JA, Rice CM. (2006) Time- and temperature-dependent activation of hepatitis C virus for low-pH-triggered entry. *J Virol*; 80:1734-41.

[291] Boehringer D, Thermann R, Ostareck-Lederer A, Lewis JD, Stark H. (2005) Structure of the hepatitis C virus IRES bound to the human 80S ribosome: remodeling of the HCV IRES. *Structure*; 13:1695-706.

[292] Ji H, Fraser CS, Yu Y, Leary J, Doudna JA. (2004) Coordinated assembly of human translation initiation complexes by the hepatitis C virus internal ribosome entry site RNA. *Proceedings of the National Academy of Sciences of the United States of America*; 101:16990-5.

[293] Fraser CS, Berry KE, Hershey JW, Doudna JA. (2007) eIF3j is located in the decoding center of the human 40S ribosomal subunit. *Mol Cell*; 26:811-9.

[294] Pestova TV, Shatsky IN, Fletcher SP, Jackson RJ, Hellen CU. (1998) A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes & development*; 12:67-83.

[295] El-Hage N, Luo G. (2003) Replication of hepatitis C virus RNA occurs in a membrane-bound replication complex containing nonstructural viral proteins and RNA. *J Gen Virol*; 84:2761-9.

[296] Gosert R, Egger D, Lohmann V, Bartenschlager R, Blum HE, Bienz K, et al. (2003) Identification of the hepatitis C virus RNA replication complex in Huh-7 cells harboring subgenomic replicons. *J Virol*; 77:5487-92.

[297] Shi ST, Lee KJ, Aizaki H, Hwang SB, Lai MM. (2003) Hepatitis C virus RNA replication occurs on a detergent-resistant membrane that cofractionates with caveolin-2. *J Virol*; 77:4160-8.

[298] Welbourn S, Green R, Gamache I, Dandache S, Lohmann V, Bartenschlager R, et al. (2005) Hepatitis C virus NS2/3 processing is required for NS3 stability and viral RNA replication. *J Biol Chem*; 280:29604-11.

[299] Tang H, Grise H. (2009) Cellular and molecular biology of HCV infection and hepatitis. *Clin Sci (Lond)*; 117:49-65.

[300] Miyanari Y, Hijikata M, Yamaji M, Hosaka M, Takahashi H, Shimotohno K. (2003) Hepatitis C virus non-structural proteins in the probable membranous compartment function in viral genome replication. *J Biol Chem*; 278:50301-8.

[301] Quinkert D, Bartenschlager R, Lohmann V. (2005) Quantitative analysis of the hepatitis C virus replication complex. *J Virol*; 79:13594-605.

[302] Wolk B, Buchele B, Moradpour D, Rice CM. (2008) A dynamic view of hepatitis C virus replication complexes. *J Virol*; 82:10519-31.

[303] Randall G, Panis M, Cooper JD, Tellinghuisen TL, Sukhodolets KE, Pfeffer S, et al. (2007) Cellular cofactors affecting hepatitis C virus infection and replication. *Proc Natl Acad Sci U S A*; 104:12884-9.

[304] Moriishi K, Matsuura Y. (2007) Host factors involved in the replication of hepatitis C virus. *Rev Med Virol*; 17:343-54.

[305] Gao L, Aizaki H, He JW, Lai MM. (2004) Interactions between viral nonstructural proteins and host protein hVAP-33 mediate the formation of hepatitis C virus RNA replication complex on lipid raft. *J Virol*; 78:3480-8.

[306] Kapadia SB, Chisari FV. (2005) Hepatitis C virus RNA replication is regulated by host geranylgeranylation and fatty acids. *Proc Natl Acad Sci U S A*; 102:2561-6.

[307] Ikeda M, Abe K, Yamada M, Dansako H, Naka K, Kato N. (2006) Different anti-HCV profiles of statins and their potential for combination therapy with interferon. *Hepatology*; 44:117-25.

[308] Watashi K, Shimotohno K. (2005) [Current approaches for developing new anti-HCV agents and analyses of HCV replication using anti-HCV agents]. *Uirusu*; 55:105-10.

[309] Kaul A, Stauffer S, Berger C, Pertel T, Schmitt J, Kallis S, et al. (2009) Essential role of cyclophilin A for hepatitis C virus replication and virus production and possible link to polyprotein cleavage kinetics. *PLoS Pathog*; 5:e1000546.

[310] Goto K, Watashi K, Murata T, Hishiki T, Hijikata M, Shimotohno K. (2006) Evaluation of the anti-hepatitis C virus effects of cyclophilin inhibitors, cyclosporin A, and NIM811. *Biochem Biophys Res Commun*; 343:879-84.

[311] Paeshuyse J, Kaul A, De Clercq E, Rosenwirth B, Dumont JM, Scalfaro P, et al. (2006) The non-immunosuppressive cyclosporin DEBIO-025 is a potent inhibitor of hepatitis C virus replication in vitro. *Hepatology*; 43:761-70.

[312] Nelson HB, Tang H. (2006) Effect of cell growth on hepatitis C virus (HCV) replication and a mechanism of cell confluence-based inhibition of HCV RNA and protein expression. *J Virol*; 80:1181-90.

[313] Pietschmann T, Lohmann V, Rutter G, Kurpanek K, Bartenschlager R. (2001) Characterization of cell lines carrying self-replicating hepatitis C virus RNAs. *J Virol*; 75:1252-64.

[314] Sainz B, Jr., Chisari FV. (2006) Production of infectious hepatitis C virus by well-differentiated, growth-arrested human hepatoma-derived cells. *J Virol*; 80:10253-7.

[315] Stuyver LJ, McBrayer TR, Tharnish PM, Hassan AE, Chu CK, Pankiewicz KW, et al. (2003) Dynamics of subgenomic hepatitis C virus replicon RNA levels in Huh-7 cells after exposure to nucleoside antimetabolites. *J Virol*; 77:10689-94.

[316] Jirasko V, Montserret R, Lee JY, Gouttenoire J, Moradpour D, Penin F, et al. (2010) Structural and functional studies of nonstructural protein 2 of the hepatitis C virus reveal its key role as organizer of virion assembly. *PLoS Pathog*; 6:e1001233.

[317] Jones DM, Patel AH, Targett-Adams P, McLauchlan J. (2009) The hepatitis C virus NS4B protein can trans-complement viral RNA replication and modulates production of infectious virus. *J Virol*; 83:2163-77.

[318] Ma Y, Anantpadma M, Timpe JM, Shanmugam S, Singh SM, Lemon SM, et al. (2010) Hepatitis C virus NS2 protein serves as a scaffold for virus assembly by interacting with both structural and nonstructural proteins
J Virol; 85:86-97.

[319] Masaki T, Suzuki R, Murakami K, Aizaki H, Ishii K, Murayama A, et al. (2008) Interaction of hepatitis C virus nonstructural protein 5A with core protein is critical for the production of infectious virus particles. *J Virol*; 82:7964-76.

[320] Pyne MT, Konnick EQ, Phansalkar A, Hillyard DR. (2009) Evaluation of the Abbott investigational use only RealTime hepatitis C virus (HCV) assay and comparison to the Roche TaqMan HCV analyte-specific reagent assay. *J Clin Microbiol*; 47:2872-8.

[321] Yi M, Ma Y, Yates J, Lemon SM. (2009) Trans-complementation of an NS2 defect in a late step in hepatitis C virus (HCV) particle assembly and maturation. *PLoS Pathog*; 5:e1000403.

[322] Moradpour D, Englert C, Wakita T, Wands JR. (1996) Characterization of cell lines allowing tightly regulated expression of hepatitis C virus core protein. *Virology*; 222:51-63.

[323] Targett-Adams P, Hope G, Boulant S, McLauchlan J. (2008) Maturation of hepatitis C virus core protein by signal peptide peptidase is required for virus production. *J Biol Chem*; 283:16850-9.

[324] Boulant S, Targett-Adams P, McLauchlan J. (2007) Disrupting the association of hepatitis C virus core protein with lipid droplets correlates with a loss in production of infectious virus. *J Gen Virol*; 88:2204-13.

[325] Shavinskaya A, Boulant S, Penin F, McLauchlan J, Bartenschlager R. (2007) The lipid droplet binding domain of hepatitis C virus core protein is a major determinant for efficient virus assembly. *J Biol Chem*; 282:37158-69.

[326] Benga WJ, Krieger SE, Dimitrova M, Zeisel MB, Parnot M, Lupberger J, et al. (2010) Apolipoprotein E interacts with hepatitis C virus nonstructural protein 5A and determines assembly of infectious particles. *Hepatology*; 51:43-53.

[327] Chang KS, Jiang J, Cai Z, Luo G. (2007) Human apolipoprotein e is required for infectivity and production of hepatitis C virus in cell culture. *J Virol*; 81:13783-93.

[328] Jiang J, Luo G. (2009) Apolipoprotein E but not B is required for the formation of infectious hepatitis C virus particles. *J Virol*; 83:12680-91.

[329] Merz A, Long G, Hiet MS, Brugger B, Chlanda P, Andre P, et al. (2010) Biochemical and morphological properties of hepatitis C virus particles and determination of their lipidome. *J Biol Chem*; 286:3018-32.

[330] Hatters DM, Peters-Libeu CA, Weisgraber KH. (2006) Apolipoprotein E structure: insights into function. *Trends Biochem Sci*; 31:445-54.

[331] Cun W, Jiang J, Luo G. (2010) The C-terminal alpha-helix domain of apolipoprotein E is required for interaction with nonstructural protein 5A and assembly of hepatitis C virus. *J Virol*; 84:11532-41.

[332] Backes P, Quinkert D, Reiss S, Binder M, Zayas M, Rescher U, et al. (2010) Role of annexin A2 in the production of infectious hepatitis C virus particles (2010). *J Virol*; 84:5775-89.

[333] Herker E, Harris C, Hernandez C, Carpentier A, Kaehlcke K, Rosenberg AR, et al. (2010) Efficient hepatitis C virus particle formation requires diacylglycerol acyltransferase-1. *Nat Med*; 16:1295-8.

[334] Roingeard P, Hourioux C. (2008) Hepatitis C virus core protein, lipid droplets and steatosis. *J Viral Hepat*; 15:157-64.

[335] Bartenschlager R, Penin F, Lohmann V, Andre P. (2010) Assembly of infectious hepatitis C virus particles. *Trends Microbiol*; 19:95-103.

[336] Welte MA. (2009) Fat on the move: intracellular motion of lipid droplets. *Biochem Soc Trans*; 37:991-6.

[337] Boulant S, Douglas MW, Moody L, Budkowska A, Targett-Adams P, McLauchlan J. (2008) Hepatitis C virus core protein induces lipid droplet redistribution in a microtubule- and dynein-dependent manner. *Traffic*; 9:1268-82.

[338] Hinson ER, Cresswell P. (2009) The antiviral protein, viperin, localizes to lipid droplets via its N-terminal amphipathic alpha-helix. *Proc Natl Acad Sci U S A*; 106:20452-7.

[339] Dubuisson J, Hsu HH, Cheung RC, Greenberg HB, Russell DG, Rice CM. (1994) Formation and intracellular localization of hepatitis C virus envelope glycoprotein complexes expressed by recombinant vaccinia and Sindbis viruses. *J Virol*; 68:6147-60.

[340] Matto M, Rice CM, Aroeti B, Glenn JS. (2004) Hepatitis C virus core protein associates with detergent-resistant membranes distinct from classical plasma membrane rafts. *J Virol*; 78:12047-53.

[341] Aizaki H, Morikawa K, Fukasawa M, Hara H, Inoue Y, Tani H, et al. (2008) Critical role of virion-associated cholesterol and sphingolipid in hepatitis C virus infection. *J Virol*; 82:5715-24.

[342] Icard V, Diaz O, Scholtes C, Perrin-Cocon L, Ramiere C, Bartenschlager R, et al. (2009) Secretion of hepatitis C virus envelope glycoproteins depends on assembly of apolipoprotein B positive lipoproteins. *PLoS One*; 4:e4233.

[343] Wozniak AL, Griffin S, Rowlands D, Harris M, Yi M, Lemon SM, et al. (2010) Intracellular proton conductance of the hepatitis C virus p7 protein and its contribution to infectious virus production

PLoS Pathog; 6:e1001087.

[344] Olofsson JK, Ronnlund M, Nordin S, Nyberg L, Nilsson LG, Larsson M. (2009) Odor identification deficit as a predictor of five-year global cognitive change: interactive effects with age and ApoE-epsilon4. *Behav Genet*; 39:496-503.

[345] Fields BN. (2001) *Fields Virology*.

[346] Bosshart H, Humphrey J, Deignan E, Davidson J, Drazba J, Yuan L, et al. (1994) The cytoplasmic domain mediates localization of furin to the trans-Golgi network en route to the endosomal/lysosomal system. *J Cell Biol*; 126:1157-72.

[347] Stadler K, Allison SL, Schalich J, Heinz FX. (1997) Proteolytic activation of tick-borne encephalitis virus by furin. *J Virol*; 71:8475-81.

[348] Shepard CW, Finelli L, Alter MJ. (2005) Global epidemiology of hepatitis C virus infection. *Lancet Infect Dis*; 5:558-67.

[349] Huber KR, Sebesta C, Bauer K. (1996) Detection of common hepatitis C virus subtypes with a third-generation enzyme immunoassay. *Hepatology*; 24:471-3.

[350] Kuo G, Choo QL, Alter HJ, Gitnick GL, Redeker AG, Purcell RH, et al. (1989) An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. *Science*; 244:362-4.

[351] Lavanchy D. (2009) The global burden of hepatitis C. *Liver Int*; 29 Suppl 1:74-81.

[352] Thomas SL, Newell ML, Peckham CS, Ades AE, Hall AJ. (1998) A review of hepatitis C virus (HCV) vertical transmission: risks of transmission to infants born to mothers with and without HCV viraemia or human immunodeficiency virus infection. *Int J Epidemiol*; 27:108-17.

[353] Yeung KS, Meanwell NA, Qiu Z, Hernandez D, Zhang S, McPhee F, et al. (2001) Structure-activity relationship studies of a bisbenzimidazole-based, Zn(2+)-dependent inhibitor of HCV NS3 serine protease. *Bioorg Med Chem Lett*; 11:2355-9.

[354] Alter HJ, Purcell RH, Shih JW, Melpolder JC, Houghton M, Choo QL, et al. (1989) Detection of antibody to hepatitis C virus in prospectively followed transfusion recipients with acute and chronic non-A, non-B hepatitis. *N Engl J Med*; 321:1494-500.

[355] Alter HJ, Stramer SL, Dodd RY. (2007) Emerging infectious diseases that threaten the blood supply. *Semin Hematol*; 44:32-41.

[356] Singh R, Kaul R, Kaul A, Khan K. (2007) A comparative review of HLA associations with hepatitis B and C viral infections across global populations. *World J Gastroenterol*; 13:1770-87.

[357] Negro F, Alaei M. (2009) Hepatitis C virus and type 2 diabetes. *World J Gastroenterol*; 15:1537-47.

[358] Hiroishi K, Ito T, Imawari M. (2008) Immune responses in hepatitis C virus infection and mechanisms of hepatitis C virus persistence. *J Gastroenterol Hepatol*; 23:1473-82.

[359] Gale M, Jr., Foy EM. (2005) Evasion of intracellular host defence by hepatitis C virus. *Nature*; 436:939-45.

[360] Sen GC. (2001) Viruses and interferons. *Annu Rev Microbiol*; 55:255-81.

[361] Cook DN, Pisetsky DS, Schwartz DA. (2004) Toll-like receptors in the pathogenesis of human disease. *Nat Immunol*; 5:975-9.

[362] Iwasaki A, Medzhitov R. (2004) Toll-like receptor control of the adaptive immune responses. *Nat Immunol*; 5:987-95.

[363] McCormick CJ, Challinor L, Macdonald A, Rowlands DJ, Harris M. (2004) Introduction of replication-competent hepatitis C virus transcripts using a tetracycline-regulable baculovirus delivery system. *J Gen Virol*; 85:429-39.

[364] Sumpter R, Jr., Loo YM, Foy E, Li K, Yoneyama M, Fujita T, et al. (2005) Regulating intracellular antiviral defense and permissiveness to hepatitis C virus RNA replication through a cellular RNA helicase, RIG-I. *J Virol*; 79:2689-99.

[365] Malmgaard L, Melchjorsen J, Bowie AG, Mogensen SC, Paludan SR. (2004) Viral activation of macrophages through TLR-dependent and -independent pathways. *J Immunol*; 173:6890-8.

[366] Der SD, Zhou A, Williams BR, Silverman RH. (1998) Identification of genes differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays. *Proc Natl Acad Sci U S A*; 95:15623-8.

[367] Foy E, Li K, Sumpter R, Jr., Loo YM, Johnson CL, Wang C, et al. (2005) Control of antiviral defenses through hepatitis C virus disruption of retinoic acid-inducible gene-I signaling. *Proc Natl Acad Sci U S A*; 102:2986-91.

[368] Fitzgerald KA, McWhirter SM, Faia KL, Rowe DC, Latz E, Golenbock DT, et al. (2003) IKKepsilon and TBK1 are essential components of the IRF3 signaling pathway. *Nat Immunol*; 4:491-6.

[369] Pflugheber J, Fredericksen B, Sumpter R, Jr., Wang C, Ware F, Sodora DL, et al. (2002) Regulation of PKR and IRF-1 during hepatitis C virus RNA replication. *Proc Natl Acad Sci U S A*; 99:4650-5.

[370] Sharma S, tenOever BR, Grandvaux N, Zhou GP, Lin R, Hiscott J. (2003) Triggering the interferon antiviral response through an IKK-related pathway. *Science*; 300:1148-51.

[371] Kanazawa N, Kurosaki M, Sakamoto N, Enomoto N, Itsui Y, Yamashiro T, et al. (2004) Regulation of hepatitis C virus replication by interferon regulatory factor 1. *J Virol*; 78:9713-20.

[372] Kanazawa N, Nakagawa M, Watanabe H, Sakamoto N, Enomoto N. (2004) [Virus factors determining the outcome of IFN treatment for chronic hepatitis C]. *Nippon Rinsho*; 62 Suppl 7:474-80.

[373] McWhirter SM, Fitzgerald KA, Rosains J, Rowe DC, Golenbock DT, Maniatis T. (2004) IFN-regulatory factor 3-dependent gene expression is defective in Tbk1-deficient mouse embryonic fibroblasts. *Proc Natl Acad Sci U S A*; 101:233-8.

[374] Bode JG, Brenndorfer ED, Haussinger D. (2007) Subversion of innate host antiviral strategies by the hepatitis C virus. *Arch Biochem Biophys*; 462:254-65.

[375] Bode JG, Ludwig S, Ehrhardt C, Albrecht U, Erhardt A, Schaper F, et al. (2003) IFN-alpha antagonistic activity of HCV core protein involves induction of suppressor of cytokine signaling-3. *FASEB J*; 17:488-90.

[376] Marcello T, Grakoui A, Barba-Spaeth G, Machlin ES, Kotenko SV, MacDonald MR, et al. (2006) Interferons alpha and lambda inhibit hepatitis C virus replication with distinct signal transduction and gene regulation kinetics. *Gastroenterology*; 131:1887-98.

[377] Robek MD, Garcia ML, Boyd BS, Chisari FV. (2007) Role of immunoproteasome catalytic subunits in the immune response to hepatitis B virus. *J Virol*; 81:483-91.

[378] Shackel NA, Bowen DG, McCaughan GW. (2010) Snipping away at hepatitis C. *Hepatology*; 51:703-5.

[379] Muir AJ, Shiffman ML, Zaman A, Yoffe B, de la Torre A, Flamm S, et al. (2010) Phase 1b study of pegylated interferon lambda 1 with or without ribavirin in patients with chronic genotype 1 hepatitis C virus infection
Hepatology
52:822-32.

[380] Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, et al. (2009) Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature*; 461:399-401.

[381] Suppiah V, Moldovan M, Ahlenstiel G, Berg T, Weltman M, Abate ML, et al. (2009) IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. *Nat Genet*; 41:1100-4.

[382] Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, Sakamoto N, et al. (2009) Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. *Nat Genet*; 41:1105-9.

[383] Netski DM, Mosbruger T, Depla E, Maertens G, Ray SC, Hamilton RG, et al. (2005) Humoral immune response in acute hepatitis C virus infection. *Clin Infect Dis*; 41:667-75.

[384] Orland JR, Wright TL, Cooper S. (2001) Acute hepatitis C. *Hepatology*; 33:321-7.

[385] Farci P, Alter HJ, Wong DC, Miller RH, Govindarajan S, Engle R, et al. (1994) Prevention of hepatitis C virus infection in chimpanzees after antibody-mediated in vitro neutralization. *Proc Natl Acad Sci U S A*; 91:7792-6.

[386] Bowen DG, Walker CM. (2005) Adaptive immune responses in acute and chronic hepatitis C virus infection. *Nature*; 436:946-52.

[387] Farci P, Alter HJ, Govindarajan S, Wong DC, Engle R, Lesniewski RR, et al. (1992) Lack of protective immunity against reinfection with hepatitis C virus. *Science*; 258:135-40.

[388] Ishii S, Koziel MJ. (2008) Immune responses during acute and chronic infection with hepatitis C virus. *Clin Immunol*; 128:133-47.

[389] Day CL, Lauer GM, Robbins GK, McGovern B, Wurcel AG, Gandhi RT, et al. (2002) Broad specificity of virus-specific CD4+ T-helper-cell responses in resolved hepatitis C virus infection. *J Virol*; 76:12584-95.

[390] Neumann-Haefelin C, Blum HE, Chisari FV, Thimme R. (2005) T cell response in hepatitis C virus infection. *J Clin Virol*; 32:75-85.

[391] Thimme R, Oldach D, Chang KM, Steiger C, Ray SC, Chisari FV. (2001) Determinants of viral clearance and persistence during acute hepatitis C virus infection. *J Exp Med*; 194:1395-406.

[392] Kaplan DE, Sugimoto K, Newton K, Valiga ME, Ikeda F, Aytaman A, et al. (2007) Discordant role of CD4 T-cell response relative to neutralizing antibody and CD8 T-cell responses in acute hepatitis C. *Gastroenterology*; 132:654-66.

[393] Grakoui A, Shoukry NH, Woollard DJ, Han JH, Hanson HL, Ghrayeb J, et al. (2003) HCV persistence and immune evasion in the absence of memory T cell help. *Science*; 302:659-62.

[394] Shoukry NH, Grakoui A, Houghton M, Chien DY, Ghrayeb J, Reimann KA, et al. (2003) Memory CD8+ T cells are required for protection from persistent hepatitis C virus infection. *J Exp Med*; 197:1645-55.

[395] Wherry EJ, Ahmed R. (2004) Memory CD8 T-cell differentiation during viral infection. *J Virol*; 78:5535-45.

[396] Guidotti LG, Chisari FV. (2001) Noncytolytic control of viral infections by the innate and adaptive immune response. *Annu Rev Immunol*; 19:65-91.

[397] Frese M, Schwarzle V, Barth K, Krieger N, Lohmann V, Mihm S, et al. (2002) Interferon-gamma inhibits replication of subgenomic and genomic hepatitis C virus RNAs. *Hepatology*; 35:694-703.

[398] Shin EC, Protzer U, Untergasser A, Feinstone SM, Rice CM, Hasselschwert D, et al. (2005) Liver-directed gamma interferon gene delivery in chronic hepatitis C. *J Virol*; 79:13412-20.

[399] Ghany MG, Strader DB, Thomas DL, Seeff LB. (2009) Diagnosis, management, and treatment of hepatitis C: an update. *Hepatology*; 49:1335-74.

[400] Sarrazin C, Zeuzem S. (2010) Resistance to direct antiviral agents in patients with hepatitis C virus infection
Gastroenterology; 138:447-62.

[401] Reesink HW, Zeuzem S, Weegink CJ, Forestier N, van Vliet A, van de Wetering de Rooij J, et al. (2006) Rapid decline of viral RNA in hepatitis C patients treated with VX-950: a phase Ib, placebo-controlled, randomized study. *Gastroenterology*; 131:997-1002.

[402] Kuntzen T, Timm J, Berical A, Lennon N, Berlin AM, Young SK, et al. (2008) Naturally occurring dominant resistance mutations to hepatitis C virus protease and polymerase inhibitors in treatment-naive patients. *Hepatology*; 48:1769-78.

[403] Sarrazin C, Kieffer TL, Bartels D, Hanzelka B, Muh U, Welker M, et al. (2007) Dynamic hepatitis C virus genotypic and phenotypic changes in patients treated with the protease inhibitor telaprevir. *Gastroenterology*; 132:1767-77.

[404] Hezode C, Forestier N, Dusheiko G, Ferenci P, Pol S, Goeser T, et al. (2009) Telaprevir and peginterferon with or without ribavirin for chronic HCV infection. *N Engl J Med*; 360:1839-50.

[405] McHutchison JG, Everson GT, Gordon SC, Jacobson IM, Sulkowski M, Kauffman R, et al. (2009) Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection. *N Engl J Med*; 360:1827-38.

[406] McHutchison JG, Lawitz EJ, Shiffman ML, Muir AJ, Galler GW, McCone J, et al. (2009) Peginterferon alfa-2b or alfa-2a with ribavirin for treatment of hepatitis C infection. *N Engl J Med*; 361:580-93.

[407] Pereira AA, Jacobson IM. (2009) New and experimental therapies for HCV. *Nat Rev Gastroenterol Hepatol*; 6:403-11.

[408] Durantel D, Alotte C, Zoulim F. (2007) Glucosidase inhibitors as antiviral agents for hepatitis B and C. *Curr Opin Investig Drugs*; 8:125-9.

[409] Lee J, Wu CC, Lee KJ, Chuang TH, Katakura K, Liu YT, et al. (2006) Activation of anti-hepatitis C virus responses via Toll-like receptor 7. *Proc Natl Acad Sci U S A*; 103:1828-33.

[410] Abel K, Wang Y, Fritts L, Sanchez E, Chung E, Fitzgerald-Bocarsly P, et al. (2005) Deoxycytidyl-deoxyguanosine oligonucleotide classes A, B, and C induce distinct cytokine gene expression patterns in rhesus monkey peripheral blood mononuclear cells and distinct alpha interferon responses in TLR9-expressing rhesus monkey plasmacytoid dendritic cells. *Clin Diagn Lab Immunol*; 12:606-21.

[411] Klade CS, Wedemeyer H, Berg T, Hinrichsen H, Cholewinska G, Zeuzem S, et al. (2008) Therapeutic vaccination of chronic hepatitis C nonresponder patients with the peptide vaccine IC41. *Gastroenterology*; 134:1385-95.

[412] Bain C, Parroche P, Lavergne JP, Duverger B, Vieux C, Dubois V, et al. (2004) Memory T-cell-mediated immune responses specific to an alternative core protein in hepatitis C virus infection. *J Virol*; 78:10460-9.

[413] Elmowalid GA, Qiao M, Jeong SH, Borg BB, Baumert TF, Sapp RK, et al. (2007) Immunization with hepatitis C virus-like particles results in control of hepatitis C virus infection in chimpanzees. *Proc Natl Acad Sci U S A*; 104:8427-32.

[414] Gonzalez-Peralta RP, Fang JW, Davis GL, Gish R, Tsukiyama-Kohara K, Kohara M, et al. (1994) Optimization for the detection of hepatitis C virus antigens in the liver. *J Hepatol*; 20:143-7.

[415] Yap SH, Willems M, Van den Oord J, Habets W, Middeldorp JM, Hellings JA, et al. (1994) Detection of hepatitis C virus antigen by immuno-histochemical staining: a histological marker of hepatitis C virus infection. *J Hepatol*; 20:275-81.

[416] Selby MJ, Choo QL, Berger K, Kuo G, Glazer E, Eckart M, et al. (1993) Expression, identification and subcellular localization of the proteins encoded by the hepatitis C viral genome. *J Gen Virol*; 74 (Pt 6):1103-13.

[417] Boulant S, Montserret R, Hope RG, Ratinier M, Targett-Adams P, Lavergne JP, et al. (2006) Structural determinants that target the hepatitis C virus core protein to lipid droplets. *J Biol Chem*; 281:22236-47.

[418] Hope RG, McLauchlan J. (2000) Sequence motifs required for lipid droplet association and protein stability are unique to the hepatitis C virus core protein. *J Gen Virol*; 81:1913-25.

[419] Lo SY, Masiarz F, Hwang SB, Lai MM, Ou JH. (1995) Differential subcellular localization of hepatitis C virus core gene products. *Virology*; 213:455-61.

[420] Suzuki R, Matsuura Y, Suzuki T, Ando A, Chiba J, Harada S, et al. (1995) Nuclear localization of the truncated hepatitis C virus core protein with its hydrophobic C terminus deleted. *J Gen Virol*; 76 (Pt 1):53-61.

[421] Yasui K, Wakita T, Tsukiyama-Kohara K, Funahashi SI, Ichikawa M, Kajita T, et al. (1998) The native form and maturation process of hepatitis C virus core protein. *J Virol*; 72:6048-55.

[422] McLauchlan J, Lemberg MK, Hope G, Martoglio B. (2002) Intramembrane proteolysis promotes trafficking of hepatitis C virus core protein to lipid droplets. *Embo J*; 21:3980-8.

[423] Boulant S, Vanbelle C, Ebel C, Penin F, Lavergne JP. (2005) Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features. *J Virol*; 79:11353-65.

[424] Jolivet-Reynaud C, Dalbon P, Viola F, Yvon S, Paranhos-Baccala G, Piga N, et al. (1998) HCV core immunodominant region analysis using mouse monoclonal antibodies and human sera: characterization of major epitopes useful for antigen detection. *J Med Virol*; 56:300-9.

[425] Matsumoto M, Hwang SB, Jeng KS, Zhu N, Lai MM. (1996) Homotypic interaction and multimerization of hepatitis C virus core protein. *Virology*; 218:43-51.

[426] Nolandt O, Kern V, Muller H, Pfaff E, Theilmann L, Welker R, et al. (1997) Analysis of hepatitis C virus core protein interaction domains. *J Gen Virol*; 78 (Pt 6):1331-40.

[427] Klein KC, Polyak SJ, Lingappa JR. (2004) Unique features of hepatitis C virus capsid formation revealed by de novo cell-free assembly. *J Virol*; 78:9257-69.

[428] Majeau N, Gagne V, Boivin A, Bolduc M, Majeau JA, Ouellet D, et al. (2004) The N-terminal half of the core protein of hepatitis C virus is sufficient for nucleocapsid formation. *J Gen Virol*; 85:971-81.

[429] Wright PE, Dyson HJ. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol*; 293:321-31.

[430] McLauchlan J. (2009) Lipid droplets and hepatitis C virus infection. *Biochim Biophys Acta*; 1791:552-9.

[431] Murray CL, Jones CT, Tassello J, Rice CM. (2007) Alanine scanning of the hepatitis C virus core protein reveals numerous residues essential for production of infectious virus. *J Virol*; 81:10220-31.

[432] Kunkel M, Lorinczi M, Rijnbrand R, Lemon SM, Watowich SJ. (2001) Self-assembly of nucleocapsid-like particles from recombinant hepatitis C virus core protein. *J Virol*; 75:2119-29.

[433] Shimoike T, Mimori S, Tani H, Matsuura Y, Miyamura T. (1999) Interaction of hepatitis C virus core protein with viral sense RNA and suppression of its translation. *J Virol*; 73:9718-25.

[434] Kang SM, Shin MJ, Kim JH, Oh JW. (2005) Proteomic profiling of cellular proteins interacting with the hepatitis C virus core protein. *Proteomics*; 5:2227-37.

[435] Sato S, Fukasawa M, Yamakawa Y, Natsume T, Suzuki T, Shoji I, et al. (2006) Proteomic profiling of lipid droplet proteins in hepatoma cell lines expressing hepatitis C virus core protein. *J Biochem*; 139:921-30.

[436] Chen CM, You LR, Hwang LH, Lee YH. (1997) Direct interaction of hepatitis C virus core protein with the cellular lymphotoxin-beta receptor modulates the signal pathway of the lymphotoxin-beta receptor. *J Virol*; 71:9417-26.

[437] Matsumoto M, Hsieh TY, Zhu N, VanArsdale T, Hwang SB, Jeng KS, et al. (1997) Hepatitis C virus core protein interacts with the cytoplasmic tail of lymphotoxin-beta receptor. *J Virol*; 71:1301-9.

[438] Zhu N, Khoshnan A, Schneider R, Matsumoto M, Dennert G, Ware C, et al. (1998) Hepatitis C virus core protein binds to the cytoplasmic domain of tumor necrosis factor (TNF) receptor 1 and enhances TNF-induced apoptosis. *J Virol*; 72:3691-7.

[439] Hsieh TY, Matsumoto M, Chou HC, Schneider R, Hwang SB, Lee AS, et al. (1998) Hepatitis C virus core protein interacts with heterogeneous nuclear ribonucleoprotein K. *J Biol Chem*; 273:17651-9.

[440] Sabile A, Perlemuter G, Bono F, Kohara K, Demaugre F, Kohara M, et al. (1999) Hepatitis C virus core protein binds to apolipoprotein AII and its secretion is modulated by fibrates. *Hepatology*; 30:1064-76.

[441] Lu W, Lo SY, Chen M, Wu K, Fung YK, Ou JH. (1999) Activation of p53 tumor suppressor by hepatitis C virus core protein. *Virology*; 264:134-41.

[442] Aoki H, Hayashi J, Moriyama M, Arakawa Y, Hino O. (2000) Hepatitis C virus core protein interacts with 14-3-3 protein and activates the kinase Raf-1. *J Virol*; 74:1736-41.

[443] Ito Y, Sasaki Y, Horimoto M, Wada S, Tanaka Y, Kasahara A, et al. (1998b) Activation of mitogen-activated protein kinases/extracellular signal-regulated kinases in human hepatocellular carcinoma. *Hepatology*; 27:951-8.

[444] Wang F, Yoshida I, Takamatsu M, Ishido S, Fujita T, Oka K, et al. (2000) Complex formation between hepatitis C virus core protein and p21Waf1/Cip1/Sdi1. *Biochem Biophys Res Commun*; 273:479-84.

[445] Jin DY, Wang HL, Zhou Y, Chun AC, Kibler KV, Hou YD, et al. (2000) Hepatitis C virus core protein-induced loss of LZIP function correlates with cellular transformation. *EMBO J*; 19:729-40.

[446] Kittlesen DJ, Chianese-Bullock KA, Yao ZQ, Braciale TJ, Hahn YS. (2000) Interaction between complement receptor gC1qR and hepatitis C virus core protein inhibits T-lymphocyte proliferation. *J Clin Invest*; 106:1239-49.

[447] Ghebrehiwet B, Habicht GS, Beck G. (1990) Interaction of C1q with its receptor on cultured cell lines induces an anti-proliferative response. *Clin Immunol Immunopathol*; 54:148-60.

[448] Kanto T, Hayashi N, Takehara T, Hagiwara H, Mita E, Naito M, et al. (1994) Buoyant density of hepatitis C virus recovered from infected hosts: two different features in sucrose equilibrium density-gradient centrifugation related to degree of liver inflammation. *Hepatology*; 19:296-302.

[449] Alisi A, Giambartolomei S, Cupelli F, Merlo P, Fontemaggi G, Spaziani A, et al. (2003) Physical and functional interaction between HCV core protein and the different p73 isoforms. *Oncogene*; 22:2573-80.

[450] Watashi K, Hijikata M, Tagawa A, Doi T, Marusawa H, Shimotohno K. (2003) Modulation of retinoid signaling by a cytoplasmic viral protein via sequestration of Sp110b, a potent transcriptional corepressor of retinoic acid receptor, from the nucleus. *Mol Cell Biol*; 23:7498-509.

[451] Mamiya N, Worman HJ. (1999) Hepatitis C virus core protein binds to a DEAD box RNA helicase. *J Biol Chem*; 274:15751-6.

[452] Owsianka AM, Patel AH. (1999) Hepatitis C virus core protein interacts with a human DEAD box protein DDX3. *Virology*; 257:330-40.

[453] Li D, Takyar ST, Lott WB, Gowans EJ. (2003) Amino acids 1-20 of the hepatitis C virus (HCV) core protein specifically inhibit HCV IRES-dependent translation in HepG2 cells, and inhibit both HCV IRES- and cap-dependent translation in HuH7 and CV-1 cells. *J Gen Virol*; 84:815-25.

[454] Yan BS, Tam MH, Syu WJ. (1998) Self-association of the C-terminal domain of the hepatitis-C virus core protein. *Eur J Biochem*; 258:100-6.

[455] Kunkel M, Watowich SJ. (2002) Conformational changes accompanying self-assembly of the hepatitis C virus core protein. *Virology*; 294:239-45.

[456] Ruggieri A, Harada T, Matsuura Y, Miyamura T. (1997) Sensitization to Fas-mediated apoptosis by hepatitis C virus core protein. *Virology*; 229:68-76.

[457] Marusawa H, Hijikata M, Chiba T, Shimotohno K. (1999) Hepatitis C virus core protein inhibits Fas- and tumor necrosis factor alpha-mediated apoptosis via NF-kappaB activation. *J Virol*; 73:4713-20.

[458] Ray RB, Meyer K, Steele R, Shrivastava A, Aggarwal BB, Ray R. (1998) Inhibition of tumor necrosis factor (TNF-alpha)-mediated apoptosis by hepatitis C virus core protein. *J Biol Chem*; 273:2256-9.

[459] Zhu N, Ware CF, Lai MM. (2001) Hepatitis C virus core protein enhances FADD-mediated apoptosis and suppresses TRADD signaling of tumor necrosis factor receptor. *Virology*; 283:178-87.

[460] Saito K, Meyer K, Warner R, Basu A, Ray RB, Ray R. (2006) Hepatitis C virus core protein inhibits tumor necrosis factor alpha-mediated apoptosis by a protective effect involving cellular FLICE inhibitory protein. *J Virol*; 80:4372-9.

[461] Ray RB, Lagging LM, Meyer K, Ray R. (1996) Hepatitis C virus core protein cooperates with ras and transforms primary rat embryo fibroblasts to tumorigenic phenotype. *J Virol*; 70:4438-43.

[462] Chang J, Yang SH, Cho YG, Hwang SB, Hahn YS, Sung YC. (1998) Hepatitis C virus core from two different genotypes has an oncogenic potential but is not sufficient for transforming primary rat embryo fibroblasts in cooperation with the H-ras oncogene. *J Virol*; 72:3060-5.

[463] Smirnova IS, Aksenov ND, Kashuba EV, Payakurel P, Grabovetsky VV, Zaberezhny AD, et al. (2006) Hepatitis C virus core protein transforms murine fibroblasts by promoting genomic instability. *Cell Oncol*; 28:177-90.

[464] Yoshida T, Hanada T, Tokuhisa T, Kosai K, Sata M, Kohara M, et al. (2002) Activation of STAT3 by the hepatitis C virus core protein leads to cellular transformation. *J Exp Med*; 196:641-53.

[465] Tsuchihara K, Hijikata M, Fukuda K, Kuroki T, Yamamoto N, Shimotohno K. (1999) Hepatitis C virus core protein regulates cell growth and signal transduction pathway transmitting growth stimuli. *Virology*; 258:100-7.

[466] Shih CM, Lo SJ, Miyamura T, Chen SY, Lee YH. (1993) Suppression of hepatitis B virus expression and replication by hepatitis C virus core protein in HuH-7 cells. *J Virol*; 67:5823-32.

[467] Ray RB, Lagging LM, Meyer K, Steele R, Ray R. (1995) Transcriptional regulation of cellular and viral promoters by the hepatitis C virus core protein. *Virus Res*; 37:209-20.

[468] Ray RB, Steele R, Meyer K, Ray R. (1998b) Hepatitis C virus core protein represses p21WAF1/Cip1/Sid1 promoter activity. *Gene*; 208:331-6.

[469] Kwun HJ, Jang KL. (2003) Dual effects of hepatitis C virus Core protein on the transcription of cyclin-dependent kinase inhibitor p21 gene. *J Viral Hepat*; 10:249-55.

[470] de Lucas S, Bartolome J, Carreno V. (2005) Hepatitis C virus core protein down-regulates transcription of interferon-induced antiviral genes. *J Infect Dis*; 191:93-9.

[471] Lerat H, Rumin S, Habersetzer F, Berby F, Trabaud MA, Trepo C, et al. (1998) In vivo tropism of hepatitis C virus genomic sequences in hematopoietic cells: influence of viral load, viral genotype, and cell phenotype. *Blood*; 91:3841-9.

[472] Hiasa Y, Horiike N, Akbar SM, Saito I, Miyamura T, Matsuura Y, et al. (1998) Low stimulatory capacity of lymphoid dendritic cells expressing hepatitis C virus genes. *Biochem Biophys Res Commun*; 249:90-5.

[473] Sarobe P, Lasarte JJ, Casares N, Lopez-Diaz de Cerio A, Baixeras E, Labarga P, et al. (2002) Abnormal priming of CD4(+) T cells by dendritic cells expressing hepatitis C virus core and E1 proteins. *J Virol*; 76:5062-70.

[474] Large MK, Kittlesen DJ, Hahn YS. (1999) Suppression of host immune response by the core protein of hepatitis C virus: possible implications for hepatitis C virus persistence. *J Immunol*; 162:931-8.

[475] Yao ZQ, Nguyen DT, Hiotellis AI, Hahn YS. (2001) Hepatitis C virus core protein inhibits human T lymphocyte responses by a complement-dependent regulatory pathway. *J Immunol*; 167:5264-72.

[476] Moriya O, Matsui M, Osorio M, Miyazawa H, Rice CM, Feinstone SM, et al. (2001) Induction of hepatitis C virus-specific cytotoxic T lymphocytes in mice by immunization with dendritic cells treated with an anthrax toxin fusion protein. *Vaccine*; 20:789-96.

[477] Perlemuter G, Sabile A, Letteron P, Vona G, Topilco A, Chretien Y, et al. (2002) Hepatitis C virus core protein inhibits microsomal triglyceride transfer protein activity and very low density lipoprotein secretion: a model of viral-related steatosis. *FASEB J*; 16:185-94.

[478] Bernacchi S, Stoylov S, Piemont E, Ficheux D, Roques BP, Darlix JL, et al. (2002) HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. *J Mol Biol*; 317:385-99.

[479] Egele C, Piemont E, Didier P, Ficheux D, Roques B, Darlix JL, et al. (2007) The single-finger nucleocapsid protein of moloney murine leukemia virus binds and destabilizes the TAR sequences of HIV-1 but does not promote efficiently their annealing. *Biochemistry*; 46:14650-62.

[480] Egele C, Schaub E, Piemont E, de Rocquigny H, Mely Y. (2005) Investigation by fluorescence correlation spectroscopy of the chaperoning interactions of HIV-1 nucleocapsid protein with the viral DNA initiation sequences. *C R Biol*; 328:1041-51.

[481] Liu HW, Cosa G, Landes CF, Zeng Y, Kovaleski BJ, Mullen DG, et al. (2005) Single-molecule FRET studies of important intermediates in the nucleocapsid-protein-chaperoned minus-strand transfer step in HIV-1 reverse transcription. *Biophys J*; 89:3470-9.

[482] Stewart-Maynard KM, Cruceanu M, Wang F, Vo MN, Gorelick RJ, Williams MC, et al. (2008) Retroviral nucleocapsid proteins display nonequivalent levels of nucleic acid chaperone activity. *J Virol*; 82:10129-42.

[483] Vo MN, Barany G, Rouzina I, Musier-Forsyth K. (2006) Mechanistic studies of mini-TAR RNA/DNA annealing in the absence and presence of HIV-1 nucleocapsid protein. *J Mol Biol*; 363:244-61.

[484] Vo MN, Barany G, Rouzina I, Musier-Forsyth K. (2009) HIV-1 nucleocapsid protein switches the pathway of transactivation response element RNA/DNA annealing from loop-loop "kissing" to "zipper". *J Mol Biol*; 386:789-801.

[485] Lorsch JR. (2002) RNA chaperones exist and DEAD box proteins get a life. *Cell*; 109:797-800.

[486] Cristofari G, Darlix JL. (2002) The ubiquitous nature of RNA chaperone proteins. *Prog Nucleic Acid Res Mol Biol*; 72:223-68.

[487] Dyson HJ, Wright PE. (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol*; 6:197-208.

[488] Ivanyi-Nagy R, Davidovic L, Khandjian EW, Darlix JL. (2005) Disordered RNA chaperone proteins: from functions to disease. *Cell Mol Life Sci*; 62:1409-17.

[489] Tompa P, Csermely P. (2004) The role of structural disorder in the function of RNA and protein chaperones. *FASEB J*; 18:1169-75.

[490] Rajkowitsch L, Semrad K, Mayer O, Schroeder R. (2005) Assays for the RNA chaperone activity of proteins. *Biochem Soc Trans*; 33:450-6.

[491] Rajkowitsch L, Schroeder R. (2007) Coupling RNA annealing and strand displacement: a FRET-based microplate reader assay for RNA chaperone activity. *Biotechniques*; 43:304, 6, 8 passim.

[492] Godet J, de Rocquigny H, Raja C, Glasser N, Ficheux D, Darlix JL, et al. (2006) During the early phase of HIV-1 DNA synthesis, nucleocapsid protein directs hybridization of the TAR complementary sequences via the ends of their double-stranded stem. *J Mol Biol*; 356:1180-92.

[493] Godet J, Mely Y. (2010) Biophysical studies of the nucleic acid chaperone properties of the HIV-1 nucleocapsid protein. *RNA Biol*; 7:48-60.

[494] Hong MK, Harbron EJ, O'Connor DB, Guo J, Barbara PF, Levin JG, et al. (2003) Nucleic acid conformational changes essential for HIV-1 nucleocapsid protein-mediated inhibition of self-priming in minus-strand transfer. *J Mol Biol*; 325:1-10.

[495] Ramalanjaona N, de Rocquigny H, Millet A, Ficheux D, Darlix JL, Mely Y. (2007) Investigating the mechanism of the nucleocapsid protein chaperoning of the second strand transfer during HIV-1 DNA synthesis. *J Mol Biol*; 374:1041-53.

[496] Tisne C, Roques BP, Dardel F. (2004) The annealing mechanism of HIV-1 reverse transcription primer onto the viral genome. *J Biol Chem*; 279:3588-95.

[497] Zeng Y, Liu HW, Landes CF, Kim YJ, Ma X, Zhu Y, et al. (2007) Probing nucleation, reverse annealing, and chaperone function along the reaction path of HIV-1 single-strand transfer. *Proc Natl Acad Sci U S A*; 104:12651-6.

[498] DeStefano JJ, Titilope O. (2006) Poliovirus protein 3AB displays nucleic acid chaperone and helix-destabilizing activities. *J Virol*; 80:1662-71.

[499] Henriet S, Sinck L, Bec G, Gorelick RJ, Marquet R, Paillart JC. (2007) Vif is a RNA chaperone that could temporally regulate RNA dimerization and the early steps of HIV-1 reverse transcription. *Nucleic Acids Res*; 35:5141-53.

[500] Ivanyi-Nagy R, Lavergne JP, Gabus C, Ficheux D, Darlix JL. (2008) RNA chaperoning and intrinsic disorder in the core proteins of Flaviviridae. *Nucleic Acids Res*; 36:712-25.

[501] Tsuchihashi Z, Brown PO. (1994) DNA strand exchange and selective DNA annealing promoted by the human immunodeficiency virus type 1 nucleocapsid protein. *J Virol*; 68:5863-70.

[502] Zuniga S, Sola I, Moreno JL, Sabella P, Plana-Duran J, Enjuanes L. (2007) Coronavirus nucleocapsid protein is an RNA chaperone. *Virology*; 357:215-27.

[503] Belisova A, Semrad K, Mayer O, Kocian G, Waigmann E, Schroeder R, et al. (2005) RNA chaperone activity of protein components of human Ro RNPs. *RNA*; 11:1084-94.

[504] Bertrand EL, Rossi JJ. (1994) Facilitation of hammerhead ribozyme catalysis by the nucleocapsid protein of HIV-1 and the heterogeneous nuclear ribonucleoprotein A1. *EMBO J*; 13:2904-12.

[505] Coetzee T, Herschlag D, Belfort M. (1994) Escherichia coli proteins, including ribosomal protein S12, facilitate in vitro splicing of phage T4 introns by acting as RNA chaperones. *Genes Dev*; 8:1575-88.

[506] Daros JA, Flores R. (2002) A chloroplast protein binds a viroid RNA in vivo and facilitates its hammerhead-mediated self-cleavage. *EMBO J*; 21:749-59.

[507] Grohman JK, Del Campo M, Bhaskaran H, Tijerina P, Lambowitz AM, Russell R. (2007) Probing the mechanisms of DEAD-box proteins as general RNA chaperones: the C-terminal domain of CYT-19 mediates general recognition of RNA. *Biochemistry*; 46:3013-22.

[508] Mayer MP, Nikolay R, Bukau B. (2002) Aha, another regulator for hsp90 chaperones. *Mol Cell*; 10:1255-6.

[509] Semrad K, Green R, Schroeder R. (2004) RNA chaperone activity of large ribosomal subunit proteins from Escherichia coli. *RNA*; 10:1855-60.

[510] Zhang A, Derbyshire V, Salvo JL, Belfort M. (1995) Escherichia coli protein StpA stimulates self-splicing by promoting RNA assembly in vitro. *RNA*; 1:783-93.

[511] Guo J, Henderson LE, Bess J, Kane B, Levin JG. (1997) Human immunodeficiency virus type 1 nucleocapsid protein promotes efficient strand transfer and specific viral DNA synthesis by inhibiting TAR-dependent self-priming from minus-strand strong-stop DNA. *J Virol*; 71:5178-88.

[512] Guo J, Wu T, Kane BF, Johnson DG, Henderson LE, Gorelick RJ, et al. (2002) Subtle alterations of the native zinc finger structures have dramatic effects on the nucleic acid chaperone activity of human immunodeficiency virus type 1 nucleocapsid protein. *J Virol*; 76:4370-8.

[513] Heilman-Miller SL, Wu T, Levin JG. (2004) Alteration of nucleic acid structure and stability modulates the efficiency of minus-strand transfer mediated by the HIV-1 nucleocapsid protein. *J Biol Chem*; 279:44154-65.

[514] Wu T, Guo J, Bess J, Henderson LE, Levin JG. (1999) Molecular requirements for human immunodeficiency virus type 1 plus-strand transfer: analysis in reconstituted and endogenous reverse transcription systems. *J Virol*; 73:4794-805.

[515] Clodi E, Semrad K, Schroeder R. (1999) Assaying RNA chaperone activity in vivo using a novel RNA folding trap. *EMBO J*; 18:3776-82.

[516] Prenninger S, Schroeder R, Semrad K. (2006) Assaying RNA chaperone activity in vivo in bacteria using a ribozyme folding trap. *Nat Protoc*; 1:1273-7.

[517] Phadtare S, Severinov K, Inouye M. (2003) Assay of transcription antitermination by proteins of the CspA family. *Methods Enzymol*; 371:460-71.

[518] Bae W, Xia B, Inouye M, Severinov K. (2000) Escherichia coli CspA-family RNA chaperones are transcription antiterminators. *Proc Natl Acad Sci U S A*; 97:7784-9.

[519] Mohr S, Stryker JM, Lambowitz AM. (2002) A DEAD-box protein functions as an ATP-dependent RNA chaperone in group I intron splicing. *Cell*; 109:769-79.

[520] Mely Y, Jullian N, Morellet N, De Rocquigny H, Dong CZ, Piemont E, et al. (1994) Spatial proximity of the HIV-1 nucleocapsid protein zinc fingers investigated by time-resolved fluorescence and fluorescence resonance energy transfer. *Biochemistry*; 33:12085-91.

[521] Mely Y, Piemont E, Sorinas-Jimeno M, de Rocquigny H, Jullian N, Morellet N, et al. (1993) Structural and dynamic characterization of the aromatic amino acids of the human immunodeficiency virus type I nucleocapsid protein zinc fingers and their involvement in heterologous tRNA(Phe) binding: a steady-state and time-resolved fluorescence study. *Biophys J*; 65:1513-22.

[522] Baudin F, Marquet R, Isel C, Darlix JL, Ehresmann B, Ehresmann C. (1993) Functional sites in the 5' region of human immunodeficiency virus type 1 RNA form defined structural domains. *J Mol Biol*; 229:382-97.

[523] Valeur B, Weber G. (1977) Resolution of the fluorescence excitation spectrum of indole into the 1La and 1Lb excitation bands. *Photochem Photobiol*; 25:441-4.

[524] Lakowicz JR. (1999) *Fluorescence spectroscopy* Kluwer Academic/Plenum.

[525] Rolinski OJ, Scobie K, Birch DJ. (2009) Protein fluorescence decay: a gamma function description of thermally induced interconversion of amino acid rotamers. *Phys Rev E Stat Nonlin Soft Matter Phys*; 79:050901.

[526] Pigault C, Gerard D. (1984) Influence of the location of tryptophanyl residues in proteins on their photosensitivity. *Photochem Photobiol*; 40:291-7.

[527] Badea MG, Brand L. (1979) Time-resolved fluorescence measurements. *Methods Enzymol*; 61:378-425.

[528] Ware WH, Ashamalla M. (1971) Pulpal response following anterior maxillary osteotomy. *Am J Orthod*; 60:156-64.

[529] Lipari A. (1980) [The use of continuous-infusion ketamine in pediatric anesthesia]. *Minerva Anestesiol*; 46:945-58.

[530] Lipari G, Szabo A. (1980) Effect of librational motion on fluorescence depolarization and nuclear magnetic resonance relaxation in macromolecules and membranes. *Biophys J*; 30:489-506.

[531] Brochon JC. (1994) Maximum entropy method of data analysis in time-resolved spectroscopy. *Methods Enzymol*; 240:262-311.

[532] Livesey AK, Brochon JC. (1987) Analyzing the distribution of decay constants in pulse-fluorimetry using the maximum entropy method. *Biophys J*; 52:693-706.

[533] Tyagi S, Kramer FR. (1996) Molecular beacons: probes that fluoresce upon hybridization. *Nat Biotechnol*; 14:303-8.

[534] Kasha M. (1991) Energy transfer, charge transfer, and proton transfer in molecular composite systems. *Basic Life Sci*; 58:231-51; discussion 51-5.

[535] Bernacchi S, Mely Y. (2001) Exciton interaction in molecular beacons: a sensitive sensor for short range modifications of the nucleic acid structure. *Nucleic Acids Res*; 29:E62-2.

[536] Shetty S, Kim S, Shimakami T, Lemon SM, Mihailescu MR. (2010) Hepatitis C virus genomic RNA dimerization is mediated via a kissing complex intermediate
RNA; 16:913-25.

[537] Stoylov SP, Vuilleumier C, Stoylova E, De Rocquigny H, Roques BP, Gerard D, et al. (1997) Ordered aggregation of ribonucleic acids by the human immunodeficiency virus type 1 nucleocapsid protein. *Biopolymers*; 41:301-12.

[538] Sharma K, Didier P, Darlix JL, de Rocquigny H, Bensikaddour H, Lavergne JP, et al. (2010) Kinetic analysis of the nucleic acid chaperone activity of the hepatitis C virus core protein. *Nucleic Acids Res*; 38:3632-42.

[539] Mir MA, Brown B, Hjelle B, Duran WA, Panganiban AT. (2006) Hantavirus N protein exhibits genus-specific recognition of the viral RNA panhandle. *J Virol*; 80:11283-92.

[540] Mir MA, Panganiban AT. (2006) The bunyavirus nucleocapsid protein is an RNA chaperone: possible roles in viral RNA panhandle formation and genome replication. *RNA*; 12:272-82.

[541] Bloomfield VA, He S, Li AZ, Arscott PB. (1991) Light scattering studies on DNA condensation. *Biochem Soc Trans*; 19:496.

[542] Thompson NL. (1991) *Fluorescence correlation spectroscopy, in topics in fluorescence spectroscopy*. Plenum Publishers, NY.

[543] Boudier C, Storchak R, Sharma KK, Didier P, Follenius-Wund A, Muller S, et al. (2010) The mechanism of HIV-1 Tat-directed nucleic acid annealing supports its role in reverse transcription
J Mol Biol; 400:487-501.

[544] Cantor C, Schimmel P. (1980) *Biophysical chemistry Part 2: Techniques for the study of biological structure and function*. Academic press; NY.

[545] Rouzina I, Bloomfield VA. (1999) Heat capacity effects on the melting of DNA. 1. General aspects. *Biophys J*; 77:3242-51.

[546] Azoulay J, Clamme JP, Darlix JL, Roques BP, Mely Y. (2003) Destabilization of the HIV-1 complementary sequence of TAR by the nucleocapsid protein through activation of conformational fluctuations. *J Mol Biol*; 326:691-700.

[547] Beltz H, Azoulay J, Bernacchi S, Clamme JP, Ficheux D, Roques B, et al. (2003) Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the nucleocapsid protein NCp7. *J Mol Biol*; 328:95-108.
[548] Fofana I, Krieger SE, Grunert F, et al. (2010) Monoclonal anti-claudin 1 antibodies prevent hepatitis C virus infection of primary Human Hepatocytes. *Gastroenterology*; 139:953-964

de la Thèse de Doctorat de l'Université de Strasbourg

de KamalKant SHARMA

" Mécanisme moléculaire des propriétés chaperonnes de la protéine Core du virus de l'hépatite C : étude physicochimique par fluorescence et résonance plasmonique de surface "

Introduction

Le virus de l'hépatite C (HCV) présente un problème majeur en santé humaine car il donne lieu a des hepatites chroniques pouvant conduire à une cirrhose du foie à l'origine du carcinome des hépathocites. Cent soixante dix millions de personnes sont concernées dans le monde même si la prévalence peut varier d'un pays à l'autre. La transmission s'effectue par contact sanguin lors d'une transfusion ainsi que lors des échanges de seringues dans les milieux concernés. Par ailleurs il a été reporté des cas de transmission par voie sexuelle et par passage mère-enfant mais cela semble moins fréquent sauf dans le cas de co-infection avec HIV. D'un point de vue thérapeutique l'utilisation de l'interféron associé à la Ribavirine est un traitement long et couteux entrainant de nombreux effets secondaires. Il est donc nécessaire de poursuivre les études menées dans la compréhension du cycle de réplication d'HCV afin de proposer de nouvelles cibles thérapetiques.

HCV qui fait partie de la famille des *Flaviviridae* est un virus envelopé contenant un ARN(+) simple brin de 9.6 kb. Cet ARN génomique est régulé temporellement au cours du cycle viral. En effet, il sert de matrice pour la synthèse d'ARN(-) lors de la transcription. Il sert également d'ARNm pour la synthèse des protéines virales. Enfin, il est encapsidé de manière spécifique au sein des virions nouvellement formés.

La partie codante de l'ARN génomique est bordée par deux régions non traduites (5'UTR et 3'UTR), fortement conservées et structurées, respectivement de 341 et 230 nucléotides. La région 5'UTR se replie pour former une structure à quatre domaines contenant un IRES (*Internal Ribosome Entry Site*) assurant la traduction coiffe-indépendante d'une ORF de ~3011 acides aminés ainsi que les signaux requis pour la synthèse et la transcription de l'ARN viral. La région 3'UTR est structurée en trois parties : (1) une partie variable comprenant les tiges-boucles VSL1 et VSL2, (2) une séquence poly (U/UC), et (3) une séquence appelée séquence X-tail de 98 nucléotides de long. Cette dernière séquence peut former trois structures en tige-boucle (dénommées SL1, SL2 et SL3). Par ailleurs, elle est hautement conservée, notamment parmi les isolats HCV les plus divergents, ce qui suggère un rôle important dans la réplication du virus.

MSTNPKPQRK TKRNTNRRPQ DVKFPGGGQI VGGVYLLPRR GPRLGVRATR KTSERSQPRG
BD1 **BD2**
RRQPIPKARR PEGRTWAQPG YPWPLYGNEG MGWAGWLLSP RGSRPSWGPT DPRRRSRNLGK
WD **BD3**
VIDTLTCGF ADLMGYIPLV GAPLGGAARA LAHGVRVLED GVNYATGNLP GCSFSIFLLA
LLSCLTIPAS A

Figure 1: séquence d'acides aminés de la protéine de capside du HCV

La protéine core de HCV est l'une des dix protéines codées par l'ORF lors du processus de traduction (Fig 1). C'est une protéine chaperonne multifonctionnelle impliquée dans plusieurs processus viraux comme la prolifération cellulaire, la différentiation, l'encapsidation de l'ARN et la formation de la nucléocapside. Cette protéine de 179 acides aminés comprend un domaine N-terminal de 117 résidus, hydrophile et faiblement replié (Domaine D1) avec 3 sous-domaines basiques (BD1 à BD3) responsables des interactions avec l'ARN, et un domaine C-terminal hydrophobe (Domaine D2) impliqué dans la liaison de la protéine avec les gouttelettes lipidiques. Il a été récemment montré que le domaine N-terminal de la protéine core possède des propriétés chaperonnes pour les acides nucléiques semblables à celles des protéines NCp7 du virus HIV. Il est admis que ces propriétés jouent un rôle important dans plusieurs étapes clefs du cycle viral, telles que la dimérisation de l'ARN génomique de HCV et son encapsidation ainsi que dans la recombinaison entre différents génotypes et sous-types de HCV. Enfin, l'interaction entre la protéine Core et des protéines cellulaires ainsi qu'entre la protéine Core et la partie 5' de l'ARN ont été très étudié. Cependant, l'interacyion entre la protéine Core et la partie 3' de l'ARN génomique d'HCV reste encore mal définie. Afin de mieux comprendre le mécanisme des propriétés de chaperonne de cette protéine au niveau moléculaire, nous avons étudié les effets de ses domaines basiques sur des oligo-nucléotides isolés de la région 3'X-Tail du génome de HCV.

Le but principal de ce travail est donc de caractériser l'interaction entre la protéine core et les acides nucléiques et détailler son activité chaperonne en utilisant les techniques de Fluorescence et de Résonance Plasmonique de Surface. Ceci pourrait se révéler décisif dans la compréhension du rôle de la protéine core lors du cycle viral afin de concevoir de nouvelles approches thérapeutiques. Notre travail de recherche peut se diviser selon trois directions principales.

A) Caractérisation des paramètres de liaison du domaine D1 avec les oligonucléotides.

Pour étudier la liaison aux acides nucléiques et l'activité chaperonne de la protéine core, nous avons utilisé des peptides synthétiques correspondant aux trois sous-domaines BD1+BD2+BD3 (peptide F) ou seulement à deux sous-domaines BD1+BD2 (peptide E) ou au domaine D1 entier.

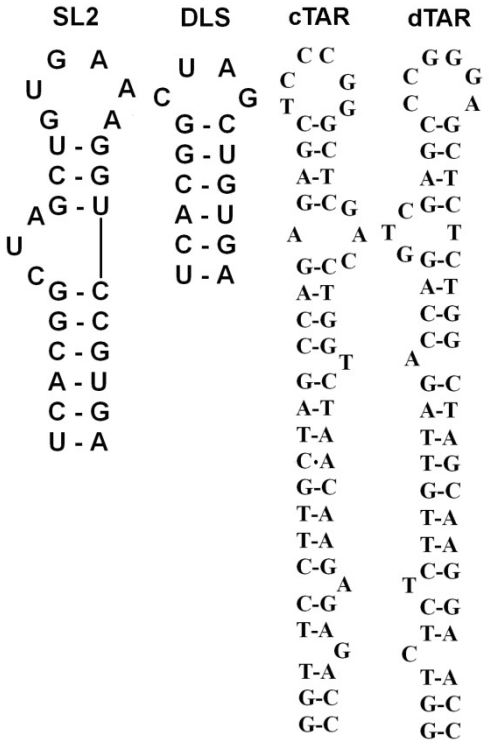


Figure 2 : Structure secondaire des oligonucléotides d'HIV-1 (cTAR and dTAR) et d'CV (SL2 and DLS) utilisés dans cette étude.

Nous avons d'abord étudié les propriétés de liaison du peptide E en utilisant une séquence palindromique de 16-nt (DLS) marquée à son extrémité 3' avec l'Alexa-430. Cette séquence du domaine SL2 de la région 3'X-tail de HCV est responsable de la dimérisation de l'ARN génomique. Lors de la titration de DLS avec le peptide E, on observe un déplacement du pic au maximum d'émission de l'Alexa (Fig 3A). Une représentation de ce déplacement en fonction de la concentration en peptide se caractérise par une courbe sigmoïdale (Fig. 3B). En outre, une forte diffusion entraînant la disparition du pic d'absorption du fluorophore (Fig 3C) a été observée. Ces données suggèrent la formation d'agrégats en présence des peptides E, F et D1, empêchant une mesure quantitative des paramètres de liaison par les techniques de fluorescence.

Afin de surmonter cet obstacle, nous avons repris l'étude de la fixation du peptide E avec les oligonucléotides DLS natifs par Résonance Plasmonique de Surface, en collaboration avec le groupe de D. Altschuh (ESBS). Cette technique permet de s'affranchir des problèmes d'agrégation du fait de l'immobilisation d'un des partenaires de l'interaction sur un sensor. Nous avons immobilisé les oligonucléotides biotinylés sur des sensors streptavidine avant d'injecter le peptide E, utilisé comme analyte. Cette façon d'opérer a donné des sensogrammes reproductibles dans des conditions minorant l'agrégation du peptide (très faible densité d'oligonucléotide (Rmax ≤ 30 RU)) et les interactions non spécifiques (150 mM NaCl), permettant ainsi d'obtenir la valeur de K_D.

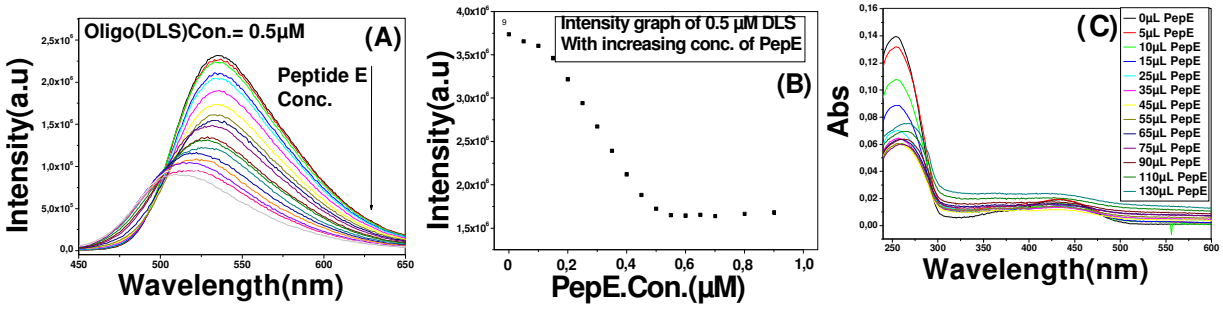
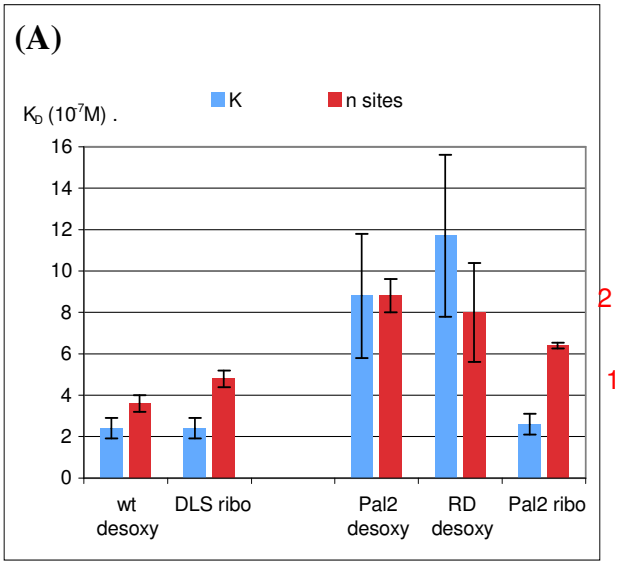


Figure 3 : Fixation du peptide E au DLS. Une concentration de 0.5 µM DLS-Alexa 430 a été titrée par des concentrations croissantes de peptide E. (A) spectres de fluorescence, et courbes de *binding* corespondantes (B) résultant de l'intensité de fluorescence à 430 nm et (C) spectres d'absorption.

Pour souligner la spécificité de l'interaction, d'autres peptides basiques tels que scFvQ34S (pI = 10) ou le lysozyme (pI = 9.2) ont été utilisés à des concentrations identiques à celles du peptide E et aucune réponse en SPR n'a été observée.



(B) The oligonucleotides tested had the following sequences:

1. dDLS : TCA CGG CTA GCT GTG A
2. rDLS : UCA CGG CUA GCU GUG A
3. dPal2 : TGA CGC CTA GCT GTG A
4. rPal2 : UGA CGC CUA GCU GUG A
5. dRD : TGC CGC GTA GAT GTA C

oligonucleotide (16 mer)	K_D (10^{-7} M)	number of sit
wt desoxy	2.4 ± 0.5	0,9 ± 0.1
DLS ribo	2.4 ± 0.5	1.2 ± 0.1
RD desoxy	11.7 ± 3.9	2 ± 0.6
Pal2 desoxy	8.8 ± 3	2.2 ± 0.2
Pal2 ribo	2.6 ± 0.5	1.6 ± 0.04

Figure 4: Liaison du peptide E avec DLS et ses mutants. (A) Constante de dissociation (K_D) et nombre de sites de fixation pour le peptide E avec cinq oligonucléotides. (B) Séquences des oligonucléotides et leurs paramètres de fixation.
r et d symbolisent une structure ribonucléique ou desoxy-ribonucléique

En outre, le nombre de sites occupés par le peptide E dépend des séquences d'oligonucléotide (Fig 4). Un site de fixation unique a été identifié pour des oligonucléotides adoptant une structure en tige-boucle stable, tels rDLS ou dDLS. Ces séquences lient le peptide E avec des affinités similaires (K_D) de 0.24 µM. Au contraire, la liaison du peptide E avec des oligonucléotides non structurés (tels rPal2, dPal2 or dRD) ont montré une

stœchiométrie plus élevée et avec une affinité 4 à 5 fois inférieure. Les paramètres de fixation des domaines basiques de la protéine core avec les oligonucléotides issus de HCV n'avaient jamais été mesurés auparavant par une méthode directe, ce qui donne toute son importance à nos résultats. Ceux-ci suggèrent que le peptide E peut se lier spécifiquement à des oligonucléotides structurés qui sont très abondants dans la région 3'X-Tail du génome de HCV.

B) Cinétique d'hybridation cTAR:dTAR d'HIV par les peptides dérivés de la protéine core d'HCV.

Dans un deuxième temps, nous avons suivi la cinétique d'hybridation des séquences complémentaires en présence des peptides d'intérêts. Pour cela, nous avons utilisé un système modèle, développé au laboratoire, d'hybridation de la séquence dTAR d'HIV avec sa séquence complémentaire cTAR. Ces séquences ont été choisies car elles ont été étudiées avec d'autres protéines chaperonnes (NCp7 et NCp10). Les conditions expérimentales ont été mise au point par FCS (Fluorescence Correlation Spectroscopy) car les peptides E, F et D1 sont des peptides basiques et provoquent une agrégation. Les différents peptides core induisent une hybridation cTAR/dTAR en moins de 30 minutes comparativement aux 24h nécessaires en leur absence. La réaction est biexponentielle. En utilisant des concentrations croissantes de dTAR, nous avons observé une saturation des constantes cinétiques rapide (k_{obs1}) et lente (k_{obs2}) de la réaction en accord avec un mécanisme d'ordre deux et la formation d'un complexe rapide suivi de la formation du complexe étendu plus stable. De plus, l'utilisation de différents mutants de cTAR et dTAR montrent que ces peptides provoquent cette hybridation par l'extrémité de la tige des oligonucléotides. En accord avec les données de cinétiques, un mécanisme réactionnel impliquant deux voies différentes avec deux intermédiaires a été proposé. Dans ce modèle, les voies lente et rapide dépendent du nombre de nucléotides à

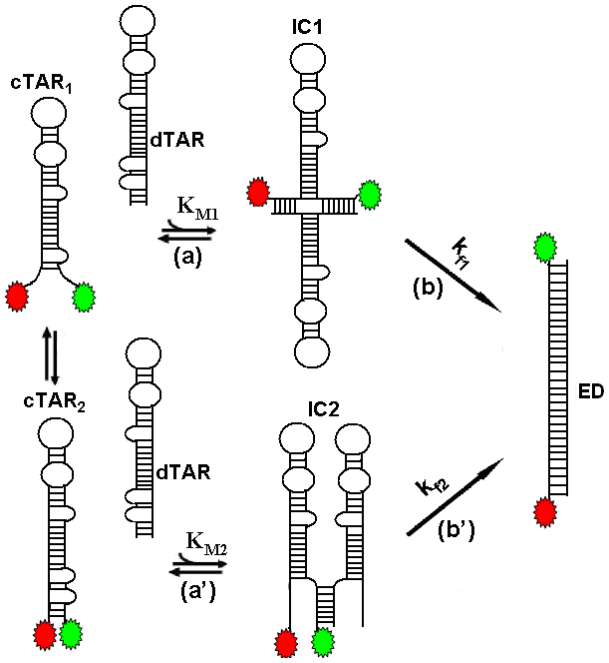


Figure 5: Mécanisme d'hybridation cTAR/dTAR par les protéines core for cTAR/dTAR annealing in presence of core peptides.

déstabiliser. De plus, nous avons montré que cette activité chaperonne dépend essentiellement des acides aminés basiques (Fig 5) (Sharma et al., 2010). Enfin, nous avons comparé ces résultats avec ceux obtenus avec le peptide (11-55) de NCp7 et nous avons montré qu'un seul peptide E possède les mêmes fonctions que quatre peptides (11-55) NCp7. Nous avons également montré que ces deux peptides (11-55 NCp7 et E) peuvent agir de concert pour faciliter les hybridations d'acides nucléiques.

C) Etude cinétique de l'hybridation de la region 3'X-tail d'HCV en présence des peptides core (Peptides E, F et Domaine D1)

Nous avons caractérisé le mécanisme d'hybridation de séquences nucléotidiques de HCV, sélectionnées à partir de la région 55 nt 3'X-tail), en absence et en présence de peptides core de HCV. Nous avons utilisé deux séquences d'ARN : la séquence palindromique 16-nt (DLS) et la tige-boucle II 27-nt (SL2). Ces séquences sont décorées d'un marqueur fluorescent et nous suivons la cinétique d'hybridation avec leur séquence complémentaire non-marquée respective cDLS et cSL2.

En l'absence de protéine, l'hybridation de ces oligonucléotides nécessite huit heures. Comme pour le modèle cTAR/dTAR, on observe une saturation des constantes cinétiques rapides (k_{obs1}) et lentes (k_{obs2}) avec les séquences DLS ou SL2. Ceci est cohérent avec un mécanisme de réaction à deux étapes pour lequel un intermédiaire en pré-équilibre rapide précédé la formation d'un duplex étendu stable via une réaction mono-moléculaire. En outre, différents mutants de DLS/cDLS et de SL2/cSL2 ont suggéré que le processus d'hybridation dans les deux systèmes passe par une étape de nucléation impliquant les boucles complémentaires ('kissing complex').

En présence de chacun des peptides, nous montrons une faible destabilisation des structures en tige-boucle de DLS et de SL2, mais une forte induction (vingt cinq fois plus rapide) de leur hybridation via des 'kissing complex' par le même mécanisme de réaction que celui observé en absence de protéine. Ainsi, les protéines core facilitent fortement l'hybridation des séquences d'oligonucléotides de HCV, mais sans modifier les voies et le mécanisme de réaction.

Sur la base de ces résultats cinétiques, un mécanisme de réaction avec une seule voie cinétique impliquant des complexes intermédiaires de ~12 bp pour SL2/cSL2 et de ~8 bp pour DLS/cDLS a été proposé (Fig 6) (Sharma KK et al, in preparation).

$$\text{DLS + cDLS} \text{ or } \text{SL2 + cSL2} \xrightleftharpoons{K_M} IC \xrightleftharpoons[k_b]{k_f} ED$$

Puisque l'ARN de HCV adopte probablement différentes conformations lors de son cycle viral, la capacité ainsi démontrée de la protéine core de favoriser l'inter-conversion entre différentes structures d'ARN pourrait expliquer comment cette protéine peut réguler les transitions entre traduction et réplication ainsi qu'entre réplication et encapsidation.

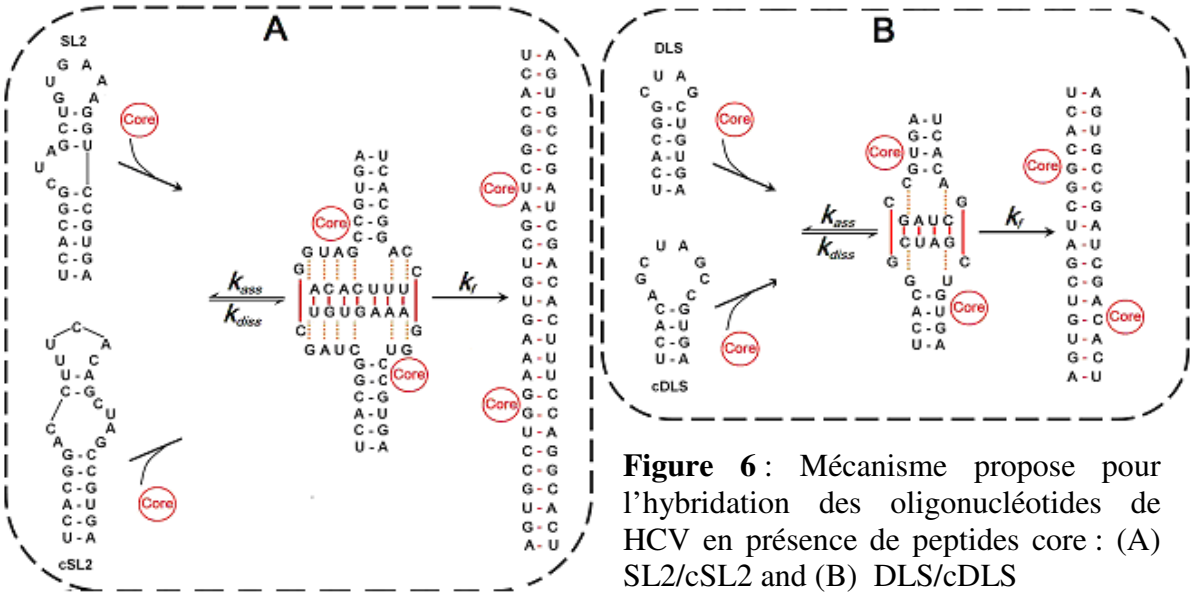


Figure 6 : Mécanisme propose pour l'hybridation des oligonucléotides de HCV en présence de peptides core : (A) SL2/cSL2 and (B) DLS/cDLS

Pour obtenir des details au niveau moléculaire de cette interaction, nous avons utilisé la 2 amino purine (2-Ap). C'est un analogue nucléosidique dont les propriétés de fluorescence varient en fonction de son environnement permettant ainsi une étude dynamique de l'oligonucléotide lors de son interaction avec une protéine. Cette stratégie a déjà été utilisée au laboratoire dans l'étude de l'interaction entre la NCp7 d'HIV et des oligonucléotides (Avilov et al. 2009; Avilov et al. 2008). Ainsi, nous avons fait synthètisé les deux oligonucléotides en plaçant en position 9 de la DLS et dans les positions 9, 16 et 20 de la SL2 (figure 1). Nos résultats indiquent que le peptide E interagit plus particulièrement avec la boucle des deux oligonucléotides.

D) Etude cinétique de l'hybridation de cTAR/dTAR en présence d'autres protéines chaperonnes (Trans-activateur de transcription (Tat) de HIV-1 et Prion humain (HuPrp))

Parallèlement, j'ai aussi participé à l'étude d'autres protéines chaperonnes telles que celle du prion (HuPrp), de la NCp7 et de TAT d'HIV et de la protéine de nuclécapside de Bunyamwera Orthobunyavirus.

- HuPrp présente des propriétés chaperonnes des acides nucléiques similaires à celles des protéines des nucléocapsides de HIV et de HCV. Nous avons utilisé un peptide prion (25-110) HuPrp de 86 acides aminés, contenant 7 tryptophanes et 2 tyrosines ayant la même activité chaperonne que la protéine totale. Ce peptide (25-110)HuPrp augmente fortement la vitesse de réaction de l'hybridation entre cTAR et dTAR. En outre, comme pour l'hybridation de cTAR/dTAR induite par la protéine core de HCV, nous avons montré que la protéine Prp active l'hybridation via un mécanisme de réaction à deux étapes, dans lequel un complexe intermédiaire en prééquilibre rapide précède la formation du duplex étendu final (ED). De même, différents mutants de cTAT et de dTAR ont montré que la nucléation des séquences cTAR et dTAR implique leurs deux extrémités respectives en présence de (25-110)HuPrp (Guichard et al., Submitted to Nucleic Acid Research).

- La principale fonction du trans-activateur de transcription de HIV-1 (la protéine Tat) est de promouvoir la transcription de l'ADN proviral par l'ARN-polymérase de l'hôte pour synthétiser de grandes quantités d'ARN viral non épissé. Nous avons montré que Tat(1-86) était tout aussi efficace que NCp7 pour promouvoir l'hybridation de cTAR/dTAR et pouvait chaperonner cette réaction d'hybridation de manière coopérative avec Ncp7. Ainsi, nos résultats nous permettent d'attribuer à Tat un rôle dans la stimulation des transferts de brin obligatoires lors de la synthèse de l'ADN viral par la transcriptase réverse (Boudier C et al., 2010).

- La transcription de l'ARN d'HIV en ADN proviral nécessite deux transferts de brins qui sont assistés par la NCp7. Lors du second transfert, la NCp7 modifie le mécanisme d'appariement PBS(+)/PBS(-) en favorisant la reconnaissance boucle-boucle. Elle augmente ainsi considérablement la cinétique de cet appariement. Pour comprendre ce mécanisme, les modifications conformationnelles des boucles de PBS en présence de la NCp7 ont été suivies en utilisant la stratégie de la 2Ap. Nous avons pu montrer que la NCp7 pouvait restreindre la dynamique de mouvement des boucles PBS. De plus, nous montrons une stricte corrélation entre cette restriction de mouvement et l'augmentation de la cinétique d'hybridation PBS(+)/PBS(-). Ceci suggère fortement que la NCp7 puisse

sélectionner une des conformations de l'oligonucléotide compétente pour son hybridation (Manuscript submitted to NAR).

- La protein de nucléocapside de Bunyamwera Orthobunyavirus (BUNV-N) est une protéine multifonctionnelle qui permet l'encapsidation des trois brins d'ARN(-) lors de la réplication du virus. De plus, cette protéine N s'oligomèrise, interagit avec la RNA polymerase (protein L) et avec la partie C terminale de la glycoprotéine d'enveloppe ainsi qu'avec des protéines cellulaires. Enfin, il a été montré récemment que la protéine N d'un virus de la même famille appelé Hantavirus Sin Nombre (SNV), présentait des propriété de protéine chaperonne (Mir and panganiban, 2006). Par analogie, nous avons caractérisé l'activité chaperonne de la protéine BUNV-N en suivant l'hybridation des oligonucléotides dTAR et cTAR d'HIV à titre de modèle. Ainsi, la protéine BUNV augmente l'hybridation de ces deux oligonucléotides puisque la réaction est complète en moins de trente minutes à la place d'une journée sans protéine. De même, nous montrons que, comme la protéine TAT d'HIV, le mécanisme de la réaction est composé de deux étapes et nécessite la reconnaissance des deux extrémités 5' et 3' (manuscript in préparation).

Conclusion et Perspectives

Au cours de ce travail, nous avons caractérisé le mécanisme moléculaire de l'activité chaperonne de la protéine core de HCV vis à vis de l'ARN sur des séquences modèles et des oligonucléotides de HCV de la région 3'X-tail. Nous avons ensuite démontré que les déterminants structuraux pour l'activité de la protéine core résidaient essentiellement dans ses trois domaines basiques. A brève échéance, il sera nécessaire de déterminer si, par analogie avec la protéine NCp7 d'HIV, la protéine core d'HCV est capable de "geler" la dynamique, à l'échelle ps-ns, des oligonucléotides de HCV en utilisant des sondes fluorescentes sensibles à l'environnement comme la 2-aminopurine (2Ap). Ensuite, l'activité chaperonne de la protéine core devra être confirmée avec la protéine entière (Domaine 1 + Domaine 2) lorsqu'elle est liée à une membrane ou un substrat lipidique par son domaine hydrophobe D2.

Publications

Publication list (peer-reviewed articles)

1) Zhang X, Sharma KK, Boeglin M, Ogier J, Mainard D, Voegel JC, Mély Y, Benkirane-Jessel N. Transfection ability and intracellular DNA pathway of nanostructured gene-delivery systems. Nano Lett. 2008 Aug;8(8):2432-6. Epub 2008 Jul 9.
2) Kamal kant Sharma, Pascal Didier, Jean Luc Darlix, Hugues de Rocquigny, Hayet Bensikaddour, Jean-Pierre Lavergne, François Pénin, Jean-Marc Lessinger, Yves Mély. Kinetic analysis of the nucleic acid chaperone activity of the Hepatitis C- Virus core protein. Nucleic Acids Research 2010 38(11):3632-3642; doi: 10.1093/nar/gkq094
3) Boudier C, Storchak R, Sharma KK, Didier P, Follenius-Wund A, Muller S, Darlix JL, Mély Y. The Mechanism of HIV-1 Tat-Directed Nucleic Acid Annealing Supports its Role in Reverse Transcription. J Mol Biol. 2010 May 21.PMID: 20493881.
4) Pascal Didier, Kamal Kant Sharma, Yves Mély ; Fluorescence techniques to characterize ligand binding to proteins in Biophysical Approaches Determining Ligand Binding to Biomolecular Targets: Detection, Measurement and Modelling; Edited by Dr A Podjarny, Dr A Dejaegere and Dr B Kieffer; RSC publishing : Thomas Graham House, Cambridge, UK
5) Nick Ramalanjaona*, Julien Godet*, Kamal K Sharma, Ludovic Richert, Etienne Piémont, Guy Duportail, Yves Mély. Effect of NCp7 on the dynamics of the apical loop of the HIV-1 primer binding site probed by 2-aminopurine. (Manuscript submitted to NAR)
6) Kamal kant Sharma, Hugues de Rocquigny, Jean Luc Darlix, Jean-Pierre Lavergne, François Pénin, Jean-Marc Lessinger and Yves Mély. Chaperone activity analysis of the Hepatitis C Virus core protein on the 3' X-Tail region of the Virus genome. (Manuscript in preparation)
7) Cécile Guichard, Roland Ivanyi-Nagy, Kamal Kant Sharma, Caroline Darlix, Daniel Marc, Yves Mély and Jean-Luc Darlix. Analysis of the nucleic acid chaperoning properties of the prion protein and its influence on RNA translation. (Manuscript in preparation)

Results were presented at the following congresses:

1) Nucleic acid chaperone activity of the Hepatitis C- Virus core protein. Kamal kant sharma, Pascal Didier, Jean Luc Darlix, Hugues de Rocquigny, Jean-Marc Lessinger and Yves Mély. "Journées Campus Illkirch-JCI, 2010", Illkirch-Graffenstaden, Strasbourg, France.
2) Mechanism of the chaperone properties of Hepatitis C Virus core protein. Kamal kant sharma, Pascal Didier, Jean Luc Darlix, Hugues de Rocquigny, Jean-Marc Lessinger and Yves Mély. "European Biophysics Congress - Genoa *2009*", Genoa, Italy.
3) Mechanism of the chaperone properties of Hepatitis C Virus core protein. Kamal kant sharma, Pascal Didier, Jean Luc Darlix, Hugues de Rocquigny, Jean-Marc Lessinger and Yves Mély. "Journées Campus Illkirch-JCI, 2009", Illkirch-Graffenstaden, Strasbourg, France.

Abstract:

Open reading frame (ORF) of 9.6kb HCV genomic RNA encodes at least 10 proteins, 4 structural and 6 non-structural, during translation process. The core is one of those 4 structural proteins and considered as a multifunctional chaperone involving in several viral processes like cell proliferation, differentiation, RNA packaging, nucleocapsid formation and recombinant genetic variability. This protein of 179 amino acids consists of a 117 residue long hydrophilic and poorly folded N-terminal domain (Domain D1) with three basic sub-domains (BD1-BD3) responsible for the interactions with RNA and chaperone properties of core, and a hydrophobic C-terminal domain (Domain D2) involved in targeting the protein to lipid droplets. With the virtue of its chaperone properties, Domain D1 dimerises the 3' untranslated region (3' UTR) of the genomic RNA. However, the mechanism of the core chaperone activity in the dimerisation of the genomic RNA and in binding with its target nucleic acids are still unknown and were investigated in this present project. To reach this objective, we used fluorescence and surface plasmon resonance (SPR) techniques. By using the native D1 domain and peptides derived from this domain, we first characterized the binding parameters and the conformational changes associated with the binding of these peptides to the native and mutated sequences from HCV 3' UTR sequences and observed that core binds specifically to the structured 55nt long and fully conserved, sequence of the 3'X-tail region. Next, we investigated the destabilization of model and HCV ODNs secondary structure by the D1 domain and its mutants and found that core peptides only marginally destabilise the secondary structures of ODNs. In a last step, we described the molecular mechanisms of the core chaperone properties based on the hybridization kinetics of various HCV and model oligonucleotides. We found that core anneals HCV ODNs via a two-step reaction mechanism and proceed through loop-loop interactions. These chaperone properties of core are thought to intervene in processes like the encapsidation, the synthesis of the complementary strand of the genomic RNA and the recombination mechanisms participating to the genetic variability of the virus. The understanding of the core chaperone properties in the viral life cycle will enable us to design new therapeutics. Lastly, this project revealed molecular mechanisms and structural determinants of other chaperone proteins viz., the Human prion protein, Tat and NCp7 from HIV-1 virus and Nucleocapsid protein from Bunyamwera Orthobunyavirus.

Résumé:

La protéine core de virus de l'hépatite C (HCV) est l'une des dix protéines codées par l'ARN génomique de 9.6kb du virus. C'est une protéine chaperonne multifonctionnelle impliquée dans plusieurs processus viraux comme la prolifération cellulaire, la différentiation, l'encapsidation de l'ARN, la formation de la nucléocapside, et la variabilité génétique. Cette protéine de 179 acides aminés comprend un domaine N-terminal de 117 résidus, hydrophile et faiblement replié (Domaine D1) avec 3 sous-domaines basiques (BD1 à BD3) responsables des interactions avec l'ARN, et un domaine C-terminal hydrophobe (Domaine D2) impliqué dans la liaison de la protéine avec les gouttelettes lipidiques. Grâce à ses propriétés de chaperonne, le domaine D1 dimérise la région 3' non traduite (3'UTR) de l'ARN génomique. Cependant le mécanisme de cette activité chaperonne ainsi que l'interaction entre la protéine Core et différents oligonucléotides de la partie 3' de l'ARN restent inconnus. Dans ce but, nous avons utilisé différentes approches de fluorescence et la résonance plasmonique de surface. Ainsi, en utilisant le peptide D1, correspondant à un fragment de la protéine core, ainsi que plusieurs dérivés de ce peptide nous avons caractérisé les paramètres de l'interaction et montré que la protéine core se lie spécifiquement aux oligonucléotides en tige-boucle de la partie 3'. Ensuite, nous avons suivi la conformation de ces oligonucléotides et montré que la protéine core n'est pas capable de déstabiliser leur structure secondaire. Enfin, nous avons décrit au niveau moléculaire le mécanisme permettant à la protéine core d'hybrider des oligonucléotides complémentaires et montré que la cinétique d'hybridation repose sur une réaction à deux étapes impliquant un contact boucle-boucle. Ce travail devrait nous permettre de mieux comprendre le rôle de la protéine core lors de l'encapsidation et la transcription de l'ARN et notamment dans les mécanismes de recombinaison expliquant les variations génétiques du virus. En outre, ce travail a permis une meilleure compréhension de l'activité de chaperonne de la protéine core, permettant d'envisager une nouvelle approche dans le développement de molécules douées d'activités antivirales. Enfin, nous avons participé à la caractérisation des mécanismes moléculaires et les déterminants structuraux d'autres protéines chaperonnes comme la protéine humaine du prion et les protéines Tat et NCp7 de VIH-1.