UNIVERSITE DE STRASBOURG

INSTITUT DE SCIENCES ET D'INGENIERIE SUPRAMOLECULAIRES

Quadruplexes de Guanines et Acridines: De la Reconnaissance Moléculaire Au Drug Design

---

Alexis PAUL

Une thèse soumise en vue de l'obtention du titre de Docteur en Sciences Chimiques de l'Université de Strasbourg

Défendue publiquement le 27 Novembre 2009 en présence de la commission d'examen suivante:

Professeur Daniel UGUEN	Président du jury
Professeur Shankar BALASUBRAMANIAN	Rapporteur externe
Docteur Jean-Louis MERGNY	Rapporteur externe
Docteur Sylvain LADAME	Directeur de thèse

"Success is the ability to go from one failure to another with no loss of enthusiasm."

Sir Winston Churchill.

Dedicated to A. Lyczko

UNIVERSITY OF STRASBOURG

INSTITUT DE SCIENCES ET D'INGENIERIE SUPRAMOLECULAIRES

G-Quadruplexes and Acridines: From Molecular Recognition to Drug Design

---

Alexis PAUL

A thesis submitted for the degree of Doctor of Philosophy in the University of Strasbourg

Defended on 27th November 2009

Academic Referees:

Professor Daniel UGUEN	President of jury
Professor Shankar BALASUBRAMANIAN	External Referee
Doctor Jean-Louis MERGNY	External Referee
Doctor Sylvain LADAME	PhD Supervisor

Shortenings

PNA-DNA	Refers to the first chapter of this manuscript
PNA-RNA	Refers to the second chapter of this manuscript
9AA-DCC	Refers to the third chapter of this manuscript
DNA	Deoxyribonucleic Acid
PNA	Peptide Nucleic Acid
RNA	Ribonucleic Acid
CD	Circular Dichroïsm
G4	G-Quadruplexes
A	Adenine
T	Thymine
C	Cytosine
G	Guanine
U	Uracil
“3+1”	Design the assembly of a (3) DNA strand and a (1) PNA strand
ESI-MS	Electro Spray Ionization Mass Spectrometry
UV	Ultraviolet Spectroscopy
9AA	9 Amino Acridine
DCC	Dynamic combinatorial chemistry
ITC	Isothermal Titration Calorimetry
NRAS	Neuroblastoma RAS viral (v-ras) oncogene
UTR	UnTranslated Regions
ATR	Alternatively Translated Regions
C-myc	MYeloCytomatosis proto-oncogene

Résumé long en français :

Quadruplexes de Guanines et Acridines :

De la Reconnaissance Moléculaire au Drug Design.

Alexis PAUL

L'intégrité de la structure de l'acide désoxyribonucléique (ADN), garant de l'information génétique, est assurée par un ensemble de forces intermoléculaires (liaisons de coordination, liaisons hydrogène, forces de Van der Waals…). Les structures d'acides nucléiques sont par conséquent des assemblages supramoléculaires qu'il est possible de cibler et stabiliser de façon spécifique par des molécules pouvant interagir de façon non covalente avec cet édifice.

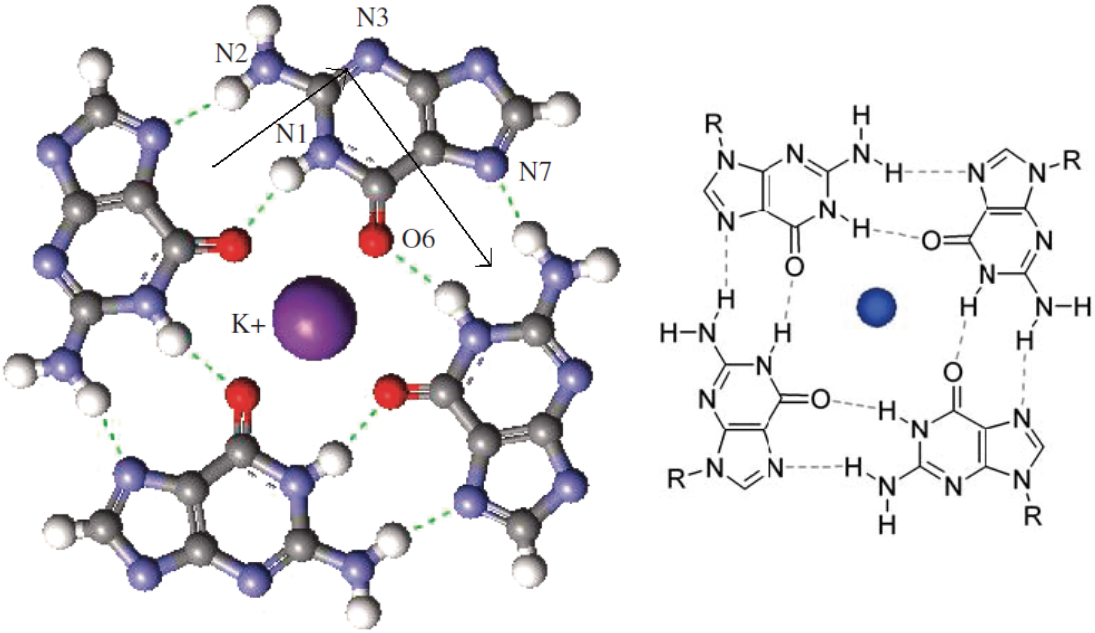


Fig. 1 Le G-quartet

Bien qu'existant très majoritairement sous sa forme double-hélice, l'ADN peut également adopter des structures alternatives stables possédant pour la plupart une fonction biologique établie. Par exemple, il est connu depuis le début des années 1960 que les séquences d'acides nucléiques riches en guanines -**G**- ont la capacité de former *in vitro* des structures supramoléculaires à quatre brins, extrêmement stables, et ceci en présence de cations physiologiques monovalents de type **K+** et Na+ . Dénommées G-quadruplexes (ou **G4**) ces structures résultent de l'assemblage en forme carrée (figure 1), par le biais de liaisons hydrogènes de type Hoogsteen (système accepteur donneur), de quatre guanines puis de l'empilement de ces « tétrades » planes de guanines en présence des cations mentionnés précédemment (figure 2).

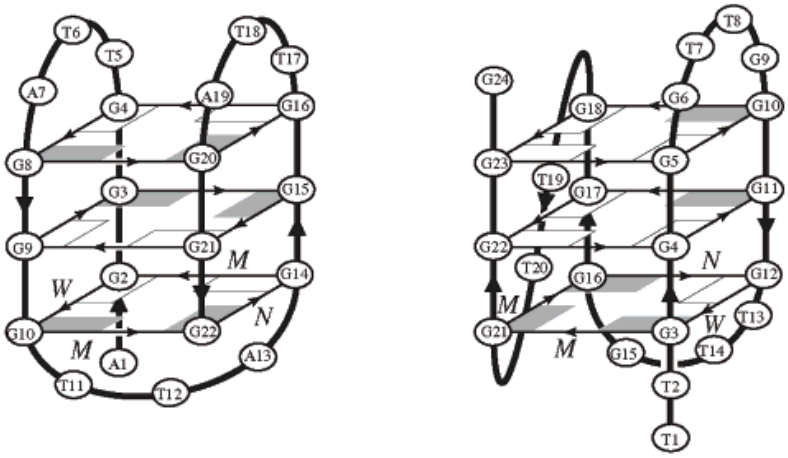


Fig. 2 Deux structures différentes de G-quadruplexes (parmi beaucoup d'autres existantes) sont données ici à titre d'exemple.

L'intérêt croissant pour ces structures secondaires d'acides nucléiques peut s'expliquer par le fait que (1) la formation de tels motifs a été mise en évidence à l'extrémité 3' simple brin de **l'ADN télomérique** (dont la longueur est un témoin du vieillissement cellulaire) et que plus récemment, (2) des études bioinformatiques ont montré que 43 % des gènes, possèdent dans la région de leur promoteur, une séquence pouvant potentiellement former un G4 suggérant ainsi une possible **fonction régulatrice** (figure

3) de ces structures au niveau de l'expression de certains gènes. Pour ces raisons, les quadruplexes ont été établis comme des cibles thérapeutiques de choix pour la conception de ligands à possible activité anti cancéreuse.

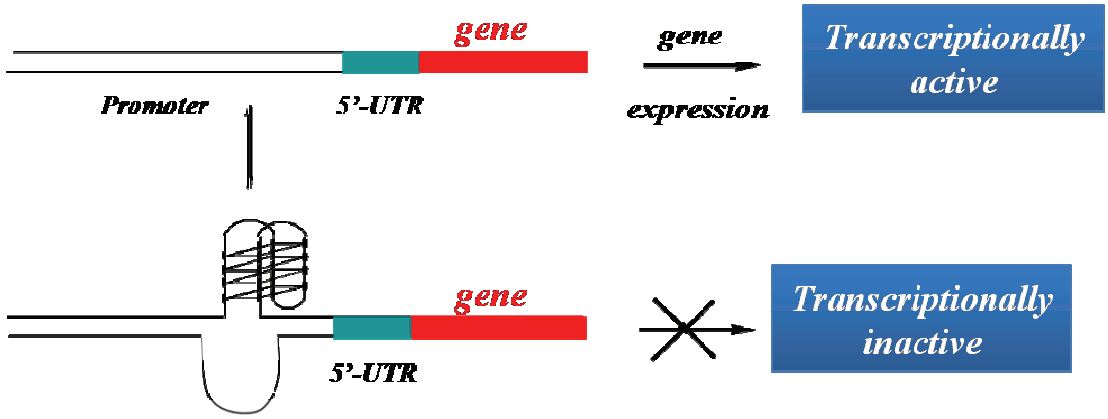


Fig. 3 Possible fonction régulatrice des G-quadruplexes au niveau de la transcription.

Pour accomplir avec succès ce rôle thérapeutique il convient de pouvoir cibler de façon **spécifique** ces structures par l'intermédiaire de petites molécules ou de protéines. Cela implique avant tout la conception de ligands synthétiques capables de discriminer entre l'ADN sous sa forme duplex majoritaire et l'ADN sous sa forme quadruplex. Pour cela, il faut prendre en compte la complexité d'une telle structure : sa formation dans le contexte d'une cellule est toujours sujette à débats (en raison d'une compétition duplex vs quadruplex fortement déplacée vers la forme duplex) mais aussi son polymorphisme structural (une séquence pouvant adopter une multitude de conformations distinctes) rend la conception rationnelle de ligands spécifiques particulièrement complexe. Dès lors, de nombreux groupes se sont intéressés à l'étude de la reconnaissance de G-quadruplexes par des petites molécules dans le but de stabiliser ces structures. Parmi les premiers ligands de G-quadruplexes apparus dans la littérature vers la fin des années 1990 figurent les dérivés hétérocycliques de la famille des « **acridines et acridones** » dont un exemple représentatif est donné Figure 4. Depuis, une grande diversité de

molécules ont été développées qui pour la plupart reconnaissent spécifiquement les quadruplexes (par rapport à l'ADN double brin) en interagissant avec la tétrade de guanines supérieure du quadruplex.

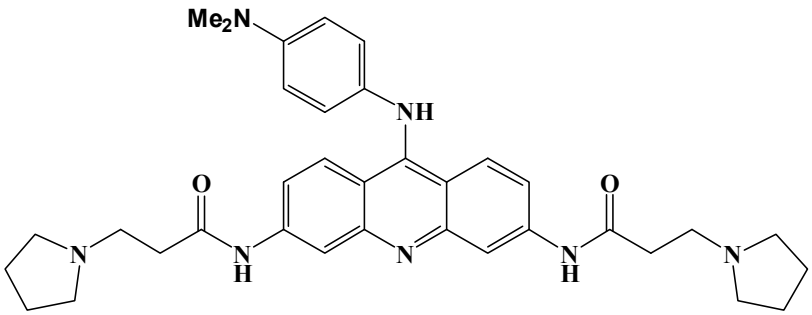


Fig. 4 Exemple de ligand de G-quadruplex : le BRACO 19

Dans ce manuscrit, nous avons développé trois approches originales pour cibler des quadruplexes d'ADN et d'ARN de façon structure et/ou séquence spécifique.

1. Comment cibler les séquences d'ADN contenant seulement trois plages de trois guanines successives ?

Dans une première partie, nous avons développé une nouvelle famille de ligands reposant sur l'association d'une molécule d'acridone et d'un analogue peptidique d'acide nucléique de type PNA (pour *Peptide Nucleic Acid*). En s'inspirant d'une étude récente par résonance magnétique nucléaire (N. Zhang *et al. J. Am. Chem. Soc.*, **2005**, *127*, 17277-17285) ayant démontré qu'un fragment de l'ADN télomérique humain d(GGGTTAGGGTTAGGGT) pouvait adopter en présence de cations sodium un quadruplex bimoléculaire asymétrique unique appelé (3+1), nous avons déduit une nouvelle manière de cibler les G-quadruplexes en mimant cette structure en (3+1). Pour

cela il est utile d'avoir recours à des analogues peptidiques d'acides nucléiques appelés PNAs. Ces analogues stables d'ADN possèdent les mêmes capacités d'hybridation que leur analogues ADN, permettant ainsi de recréer les mêmes structures. Dans le cadre de ce projet, nous avons donc démontré qu'un ligand de type « PNA-acridone » (Lys-**G**-**G**-**G**-Gly-Acridone où **G** est un monomère de PNA) pouvait s'associer avec un fragment d'ADN télomérique pour former, dans des conditions physiologiques, une structure de quadruplex bimoléculaire hybride PNA_1-DNA_1 impliquant notamment la formation de tétrades de guanines hybrides PNA:ADN (Figure 5). Cette structure mime parfaitement la structure mise en évidence par Zhang *et al.* et constitue un nouveau moyen de bloquer l'extrémité 3'-simple brin de l'ADN télomérique dans une conformation de type quadruplex.

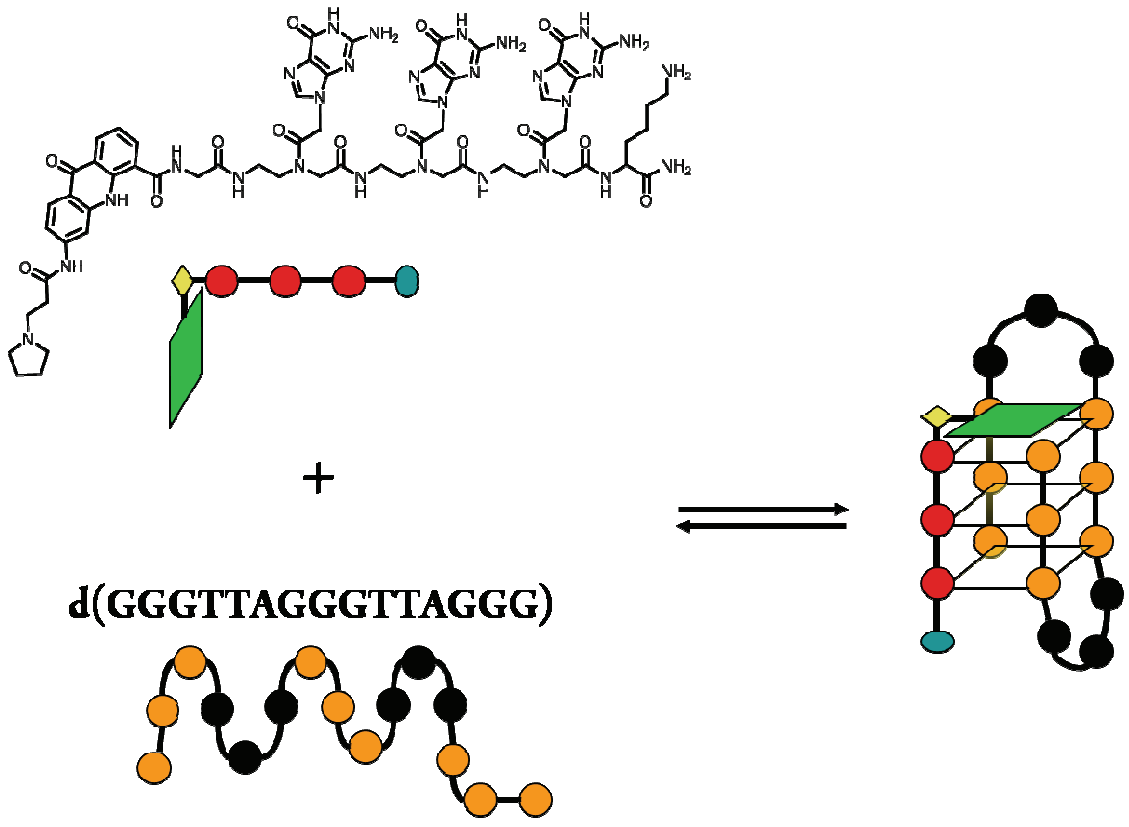


Fig. 5 Formation de quadruplexes bimoléculaires hybrides PNA:ADN [3 + 1]

2. Et comment cibler les quadruplexes d'ARN de façon « séquence+structure » spécifique ?

Dans une deuxième partie, nous nous sommes intéressés à la capacité pour un ligand de type PNA_1-Acridone-PNA_2 à (1) stabiliser de façon spécifique un quadruplex d'ARN d'intérêt, et (2) discriminer entre ce quadruplex et d'autres structures secondaires d'ARN, y compris d'autres quadruplexes d'ARN. Les quadruplexes d'ARN ont à ce jour été beaucoup moins étudiés que leurs analogues ADN. Pourtant ces structures ont statistiquement plus de chances d'exister in vivo dans leur formation n'est pas en compétition avec la formation d'une structure préférentielle de type duplex comme cela est le cas pour l'ADN. Des molécules hybrides de type [hétérocycle + peptide] qui reconnaissent simultanément différentes caractéristiques structurales d'un quadruplex (tétrade supérieure ***et*** boucles) ont déjà été développées qui présentent par exemple une très forte spécificité pour les quadruplexes d'ADN. Dans ce chapitre, nous avons développé **des ligands hautement spécifiques** d'un quadruplex d'ARN en ciblant simultanément la structure du G-quadruplex mais également les deux régions d'ARN simple-brin qui l'entourent. Des molécules hybrides de structure générale acridone-PNA_1-espaceur-PNA_2 ont été préalablement décrites comme capables d'envahir une séquence d'ADN double-brin de façon séquence-spécifique, les deux PNAs contrôlant l'orientation de l'acridone par rapport à l'ADN (Figure 6). Ici, nous avons synthétisé des molécules hybrides de structure générale PNA_1-Hétérocycle-PNA_2 où l'hétérocycle et les deux $PNA_{1,2}$ sont conçus de telle façon qu'ils puissent interagir respectivement avec la structure du G4 et les deux séquences d'ARN adjacentes. Nous envisageons notamment qu'avec de telles molécules, les deux PNAs orienteraient de façon spécifique l'hétérocycle vers le quadruplex d'intérêt, à savoir celui localisé entre les deux séquences complémentaires des PNAs uniquement. Dans ce chapitre, nous avons démontré que ce type de ligand pouvait effectivement discriminer entre deux quadruplexes d'ARN identiques possédant des séquences adjacentes différentes.

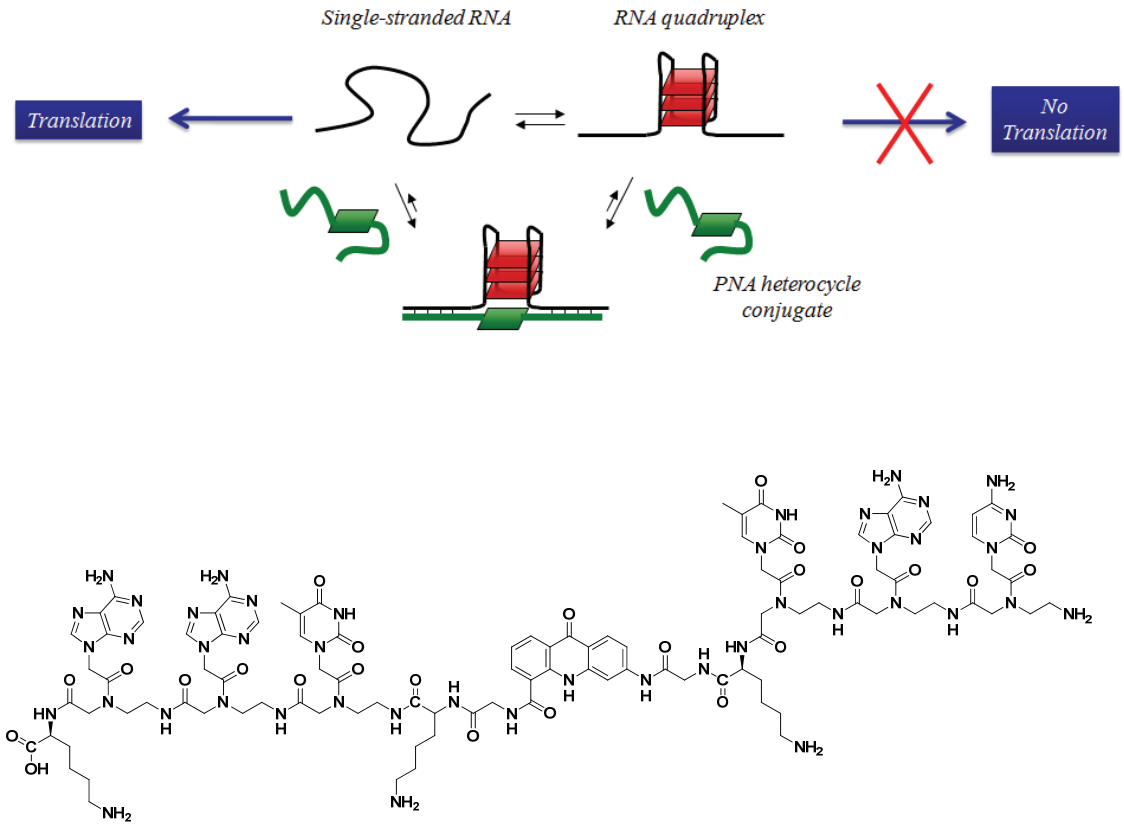


Fig. 6 Stratégie de ciblage des régions adjacentes d'un ARN capable de former un G4 et structure d'un ligand de quadruplex d'ARN de type PNA_1-acridone-PNA_2

3. Lorsque les 9-aminoacridines échangent leurs amines :

Dans une dernière partie, nous avons mis en évidence une propriété chimique encore inconnue des acridines substituées en position 9. En plus de leurs activités antibactériennes et antiparasitaires (trypanocides ou anti-malariennes) les dérivés d'acridine possèdent également des propriétés anticancéreuses. Ces dernières résultent de la capacité pour les dérivés d'acridine à interagir avec les acides nucléiques : soit par intercalation entre les paires de base de l'ADN double-hélice (responsable de l'inhibition

de l'enzyme Topoisomérase II par exemple); soit par la stabilisation de l'ADN sous sa forme G-quadruplex (exemple du ligand BRACO19, Fig7).

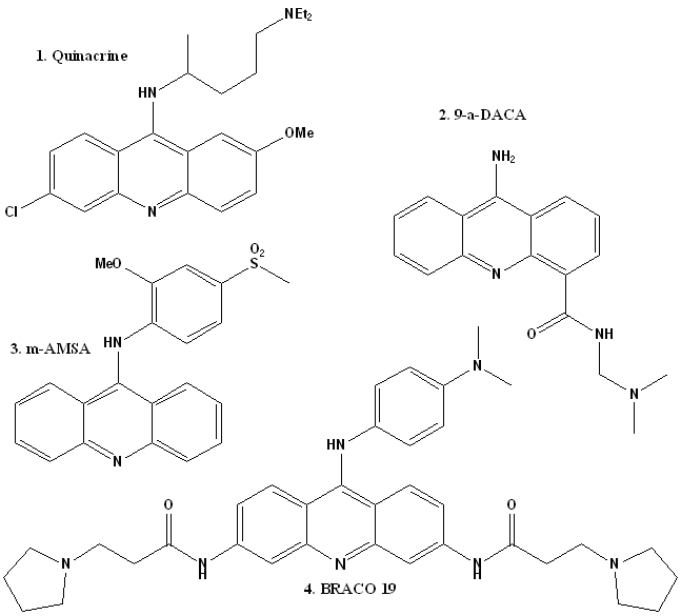


Fig. 7 Quatre exemples de 9-aminoacridines à activité thérapeutique

Or une des caractéristiques communes des dérivés d'acridine est la présence en position 9 (sommet de l'hétérocycle central) d'un substituent de type amine aliphatique ou aromatique dont la nature est fortement liée à l'activité thérapeutique de la molécule. (pour quelques exemples représentatifs, voir Figure 7).

Dans ce dernier chapitre, nous avons démontré que les 9-aminoacridines pouvaient subir une réaction d'échange d'amine en milieu physiologique aqueux et que cette réaction était réversible et sous contrôle thermodynamique. Cette nouvelle propriété chimique des 9-aminoacridines permet d'envisager la sélection par une approche de Chimie Dynamique Combinatoire (DCC), de nouveaux ligands optimisés de quadruplexes de guanines. Cette découverte a également des implications majeures quant à la compréhension du possible mode d'action de ces drogues in vivo, un aspect qui est également discuté dans ce dernier chapitre.

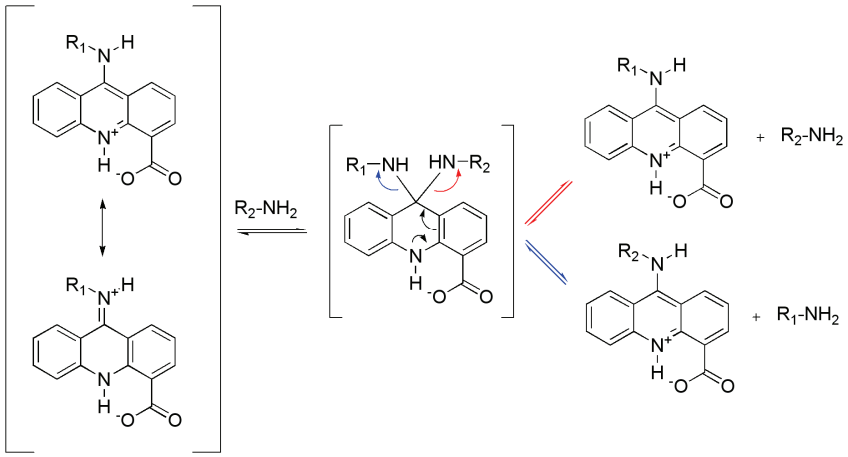


Fig. 9 Mécanisme d'échange en position 9 d'amines

4. Conclusion

En conclusion nous avons mis-au-point trois stratégies différentes de reconnaissance des G4, approches qui dans le futur pourront éventuellement donner naissance à des ligands de G4 hautement spécifiques et biocompatibles pouvant ainsi être utilisés in vivo pour le ciblage de telles structures, aussi bien au niveau de l'ADN que de l'ARN.

Résumé court.

Bien qu'existant très majoritairement sous sa forme double-hélice, l'ADN peut également adopter des structures alternatives stables possédant pour la plupart une fonction biologique établie. Par exemple, il est connu depuis le début des années 60 que les séquences d'acides nucléiques riches en guanines -G- ont la capacité de former *in vitro* et *in vivo* des structures supramoléculaires à quatre brins, extrêmement stables, et ceci en présence de cations physiologiques monovalents. **Les G-quadruplexes (G4)** ont été établis comme des cibles thérapeutiques de choix depuis que des études concordantes tendent à démontrer que ces structures existent in vivo, à l'extrémité des télomères ou dans la région des promoteurs de gènes. Par conséquent, des ligands capables de reconnaître et de stabiliser ces structures possèdent potentiellement une activité anti cancéreuse. Ici, nous avons développé avec succès trois approches originales pour cibler des quadruplexes d'ADN et d'ARN de façon structure et/ou séquence spécifique.

Dans une première partie, nous avons développé une nouvelle famille de ligands reposant sur l'association d'une molécule d'acridone et un analogue peptidique d'acide nucléique (PNA). Nous avons notamment synthétisé une molécule de type « PNA-acridone » et démontré sa capacité à interagir avec un fragment de l'ADN télomérique pour former une structure hybride impliquant notamment la formation de tétrades de guanines hybrides PNA:ADN.

Dans une deuxième partie, nous nous sommes intéressés à la capacité pour un ligand à reconnaître de façon spécifique un quadruplex d'ARN et à discriminer entre ce quadruplex et d'autres structures secondaires d'ARN, y compris d'autres quadruplexes d'ARN. Nous avons synthétisé une molécule de type PNA_1-acridone-PNA_2 capable de cibler la structure même du quadruplex (reconnaissance structurale par le noyau acridone) mais également ses deux ARN simples-brins adjacents (reconnaissance séquence-spécifique par les deux PNAs). Nous avons démontré que ce ligand ne stabilisait fortement que le quadruplex d'ARN dont les séquences simple-brin en amont et aval étaient complémentaires à celles des deux PNA du ligand.

Dans une dernière partie, nous avons mis en évidence une propriété chimique encore inconnue des acridines substituées en position 9. Nous avons alors démontré que les 9-amino acridines pouvaient subir une réaction d'échange d'amine en milieu physiologique aqueux et que cette réaction était réversible. Cette nouvelle propriété chimique des 9-amino acridines permet d'envisager l'utilisation du concept de chimie dynamique combinatoire pour identifier de nouveaux ligands de G4 ou pour optimiser la structure de ligands existants. Enfin, cette découverte a également des implications quant à la compréhension du possible mode d'action de ces drogues in vivo.

Abstract.

Although existing mainly under its double-helix form, DNA can also adopt alternative secondary structures which often prove biologically relevant. For example, it is known since the beginning of 60's that guanine-rich nucleic acid sequences can form highly stable four-stranded supramolecular structures in vitro in the presence of physiological monovalent cations such as K^+ or Na^+. These so-called G-quadruplexes **(G4)** were recently established as highly promising therapeutic targets since there are now increasing evidences that such structures form in vivo, either at the end of telomeres or in the promoter regions of genes. Therefore, ligands that can bind to and stabilize these G4 may potentially exhibit interesting therapeutic (e.g. anti-cancer) properties. Here, we successfully developed three original approaches to target DNA and RNA quadruplexes in a structure and/or specific sequence manner.

In a first chapter, we developed a new family of G4 ligands that combine on a single scaffold an acridone derivative and a nucleic acid peptidic analogue (PNA). We designed and synthesized a "PNA-acridone" conjugate and demonstrated its capacity to interact with telomeric DNA to form a hybrid quadruplex structure implying the formation of PNA: DNA hybrid G-tetrads.

In a second part, we were interested in designing a small molecule that would recognize a unique intramolecular RNA G-quadruplex of interest and would also be capable to discriminate between this quadruplex and other RNA secondary structures, including others RNA G-quadruplexes. We synthesized a ligand of the type PNA_1-acridone-PNA_2 that could target the quadruplex core structure (via the acridone ligand in a structure specific manner) but also its two single stranded flanking regions (via both PNAs, in a sequence specific manner). We demonstrated that this ligand was capable to strongly stabilize the only RNA quadruplex that contains flanking arms complementary in sequence to the ligand PNA sequences.

Finally, we highlighted a still unknown chemical property of 9-amino substituted acridines. We demonstrated that 9-aminoacridines could undergo an amine exchange reaction under aqueous and near physiological conditions and that this reaction was reversible. This new chemical property of the 9-aminoacridines enables to consider using a dynamic combinatorial chemistry approach to (i) identify new G4 ligands or (ii) optimize existing G4 ligands. The discovery of this reversible reaction has also implications on the understanding of the possible mode of action of these drugs in vivo.

Preface

Dear reader,

What you are about to discover is the fruit of 3 years of positive academic research achieved in the laboratory of Chemistry and Recognition of Biomolecules in ISIS Strasbourg. Supervised by Dr. Sylvain LADAME, each chapters of this manuscript are organized with their specific introductions and problematic.

In order to directly dive into the fabulous world of G-quadruplexes a general introduction will detail all the surroundings of this PhD thesis.

The scope of this work is international so it was written in English to bring it to a wider audience. I already apologize to the true defenders of Moliere's language.

I would like to thank Pr. Jean Marie Lehn and ISIS for the funding provided by the French Ministry of Research and Technology.

I would like to offer this little work to all who supported me during this period, and even more to those who made my life a prolific environment to researches.

I would like already to thank Puls Radio, Jack Bauer, Kara Thrace, and David Webb for these fantastic moments of entertainment.

I wish you a pleasant reading,

TABLE OF CONTENTS

Chapter 1
Introduction

1. DNA

a. *Supramolecular Nature and Structure:*

Nucleic acids are vital components of life as we know it. Nucleic acids are made from nucleotides. The combination of a heterocyclic ring (purine or pyrimidine) and a sugar (pentose) gives nucleosides, and their phosphate esters are known as nucleotides. DeoxyRibonucleic Acid (DNA) is a polymer known as "polynucleotide". Each nucleotide consists of a pentose sugar (2-deoxy-D-ribose), a nitrogen-containing base attached to the sugar (adenine A, guanine G, cytosine C or thymine T) and a phosphate group.[1]

DNA is a normally double-stranded macromolecule. Two polynucleotide chains, held together by weak thermodynamic forces, form a DNA molecule. Cohesion of the DNA double helix is linked to Watson-Crick base (nucleotide) pairing (Fig. 1) but other types of nucleic bases associations may also occur (e.g. Hoogsteen type wobble pairs) that will be developed in the following sections.

Within the DNA double-helix, A forms two hydrogen bonds with T on the opposite strand and G forms three hydrogen bonds with C on the opposite strand. Thus, (A – T) and (G – C) are specific Watson-Crick base pairs that allow DNA to carry a unique information. It is noteworthy that

-The ratio (A+T)/(C+G) differs for each species.

-The ratios T/A and C/G in DNA sequences are approximately 1.

- GC pairing is more stable then AT due to a higher number of H-bonds. (3 vs 2) (Fig. 1)

Additional factors also contribute to the B-DNA helix stabilization such as π-π (stacking) interactions between base-pairs, Van der Waals interactions and hydrophobic interactions. Thus, the "supramolecular" nature of DNA arises from those weak but numerous interactions. By convention, DNA sequences are written in the "forward direction", from the 5' to 3' ends. 5' and 3' respectively represent the free carbon at the top of the sugar and the free hydroxyl group at the bottom. This way, the phosphate groups bridge between the C5' and (O3' -1) of each nucleosides. (Fig. 2)

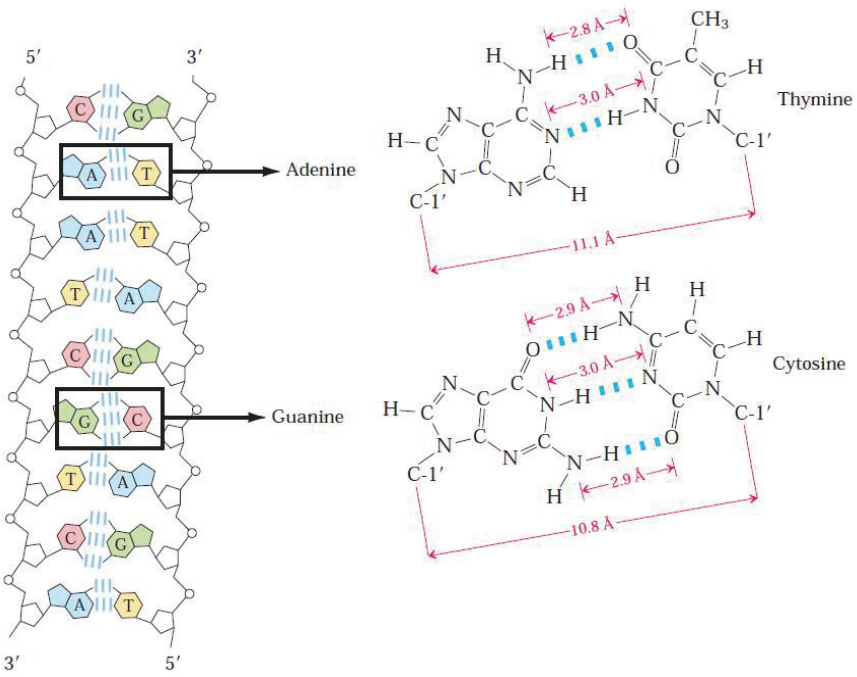


Fig.1 Nucleotides and Watson-Crick base pairing[2]

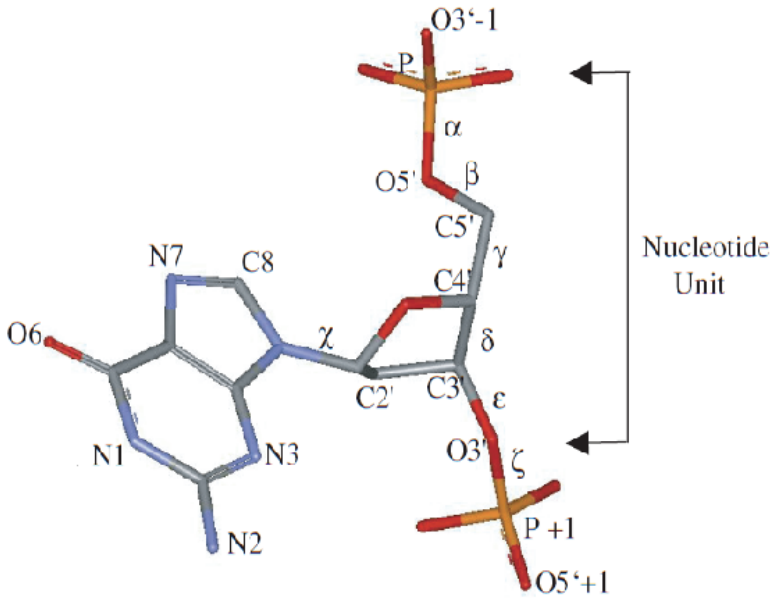


Fig.2 Example of 5'-3' orientation in a guanine nucleotide[3]

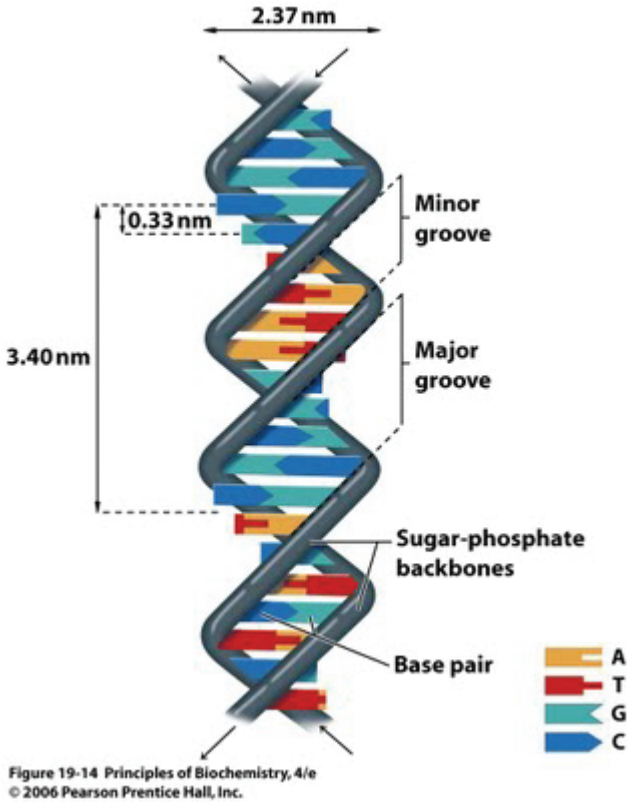


Fig.3 Minor and major grooves of double stranded B-DNA

The structure of DNA was first described in 1953[4] by Watson and Crick aided by the previous work of Chargaff and Franklin (which suggested that DNA should adopt a double helix conformation). We can describe DNA as a string of nucleotides joined by phosphodiester linkages where the unique structure arises from the sequence of bases. The so-called A-DNA and B-DNA are right-handed double helices held together by hydrogen bonding[1, 5]. (Fig. 3)

-A-DNA has 11 residues per turn with a base tilt of 20° to enhance stacking and a distance to central axis of 4.5 Å. This helix is stiff and shows little sequence-dependent variation in structure. The major groove is deep and narrow, whereas minor groove is broad and shallow.

-B-DNA has 10 residues per turn with little tilting in bases. The wide major groove and narrow minor groove are both moderately deep and both are highly solvated by

water molecules. Hence B-DNA is a flexible structure and shows sequence-dependent variation in structure. That's why it is considered as the biologically relevant form of DNA.

b. Genetic information

DNA is the carrier of genetic information. In eukaryotes, DNA is housed within the nucleus. In order for DNA to fit within a cell, long pieces of double-stranded DNA are tightly looped, coiled and folded into structures called chromosomes. Chromosomes form the genome, where specific coding DNA sequences are called genes. In all eukaryotes that engage in sexual reproduction, chromosomes exist in pairs, which means that there are two copies of each chromosome in most cells that compose these organisms' bodies. Humans, for instance, have 23 pairs of chromosomes, for a total of 46 individual chromosomes. As vectors of heritable information they are responsible for enzyme production and therefore their alteration (also known as mutation) may cause an enzymatic response. Likewise, linear sequences of DNA are responsible for primary amino acid sequences in proteins leading to the "one gene – one primary sequence" hypothesis, and their alteration may lead to over or under expression of the associated protein.[1]

c. Transcription and Translation.

Protein synthesis involves the processes of transcription and translation. Transcription is the synthesis of RNA based on a DNA template, and translation is the synthesis of the protein from a mRNA template (Fig. 4). Interferences in these processes may result in a modification of the encoded protein.
Initiation of transcription occurs in the upstream region of genes known as promoters, which are recognized by RNA polymerase and are not gene sequences.
Promoters can be identified via foot printing, which consists of the identification of the non-hydrolyzed parts of DNA protected by RNA polymerase (Gel electrophoresis will show missing bands by comparison with the total DNA sequence where the promoters are located). Once the double helix is open and the promoter recognized by RNA polymerase, the creation of mRNA occurs until the polymerase reaches a termination code written in DNA.

In eukaryotes the newly produced mRNA transcript is modified by adding:
-First a cap to the 5' end
-Second a poly-Adenine tail to the 3' end

The introns (regions not translated into protein) of the precursor mRNA (pre-mRNA) are removed through RNA splicing and the remaining exons thus form the mature mRNA transcript.

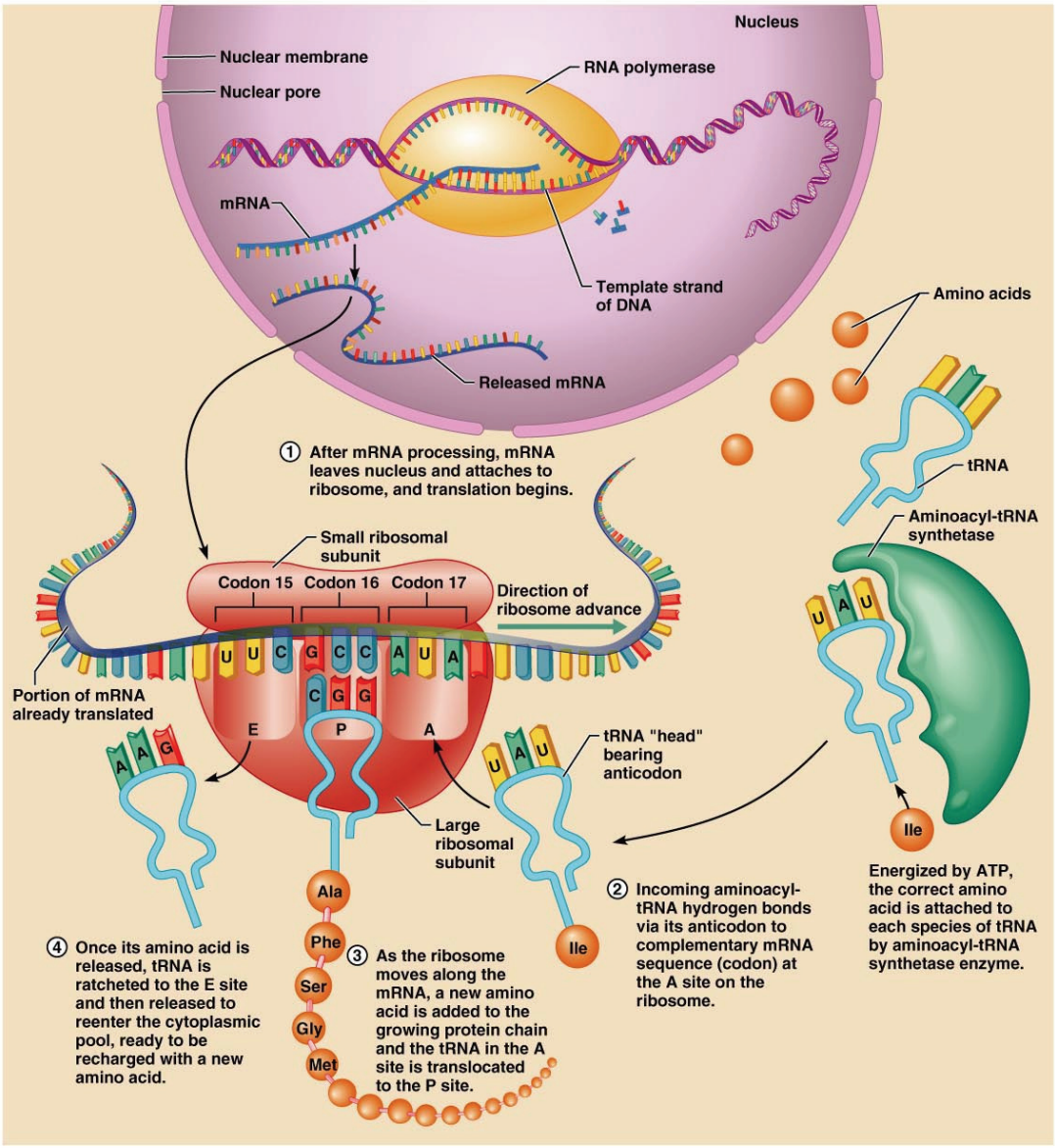


Fig. 4 Transcription and translation – Role of ribosome

Transcription ends with the mRNA transcript leaving the nucleus through its membranes, and translation starts when a ribosome (Fig. 4) attaches to mRNA. Thanks to a ribosome movement on the mRNA transcript, the complementary tRNA sequences elaborate the protein amino acids sequence. The initiation sequence is often the AUG codon (mRNA) where anticodon (tRNA) is linked to a methionine residue. Likewise UAA would signify the termination of the translation and the release of the protein.

d. Telomerase

Chromosomal ends are protected via lasso-like structures called telomeres. They have several functions, such as preventing chromosome recombination or playing a role in DNA replication. Telomeres are consumed during cell division and can potentially be recreated by a reverse transcriptase called telomerase. (Fig. 5)

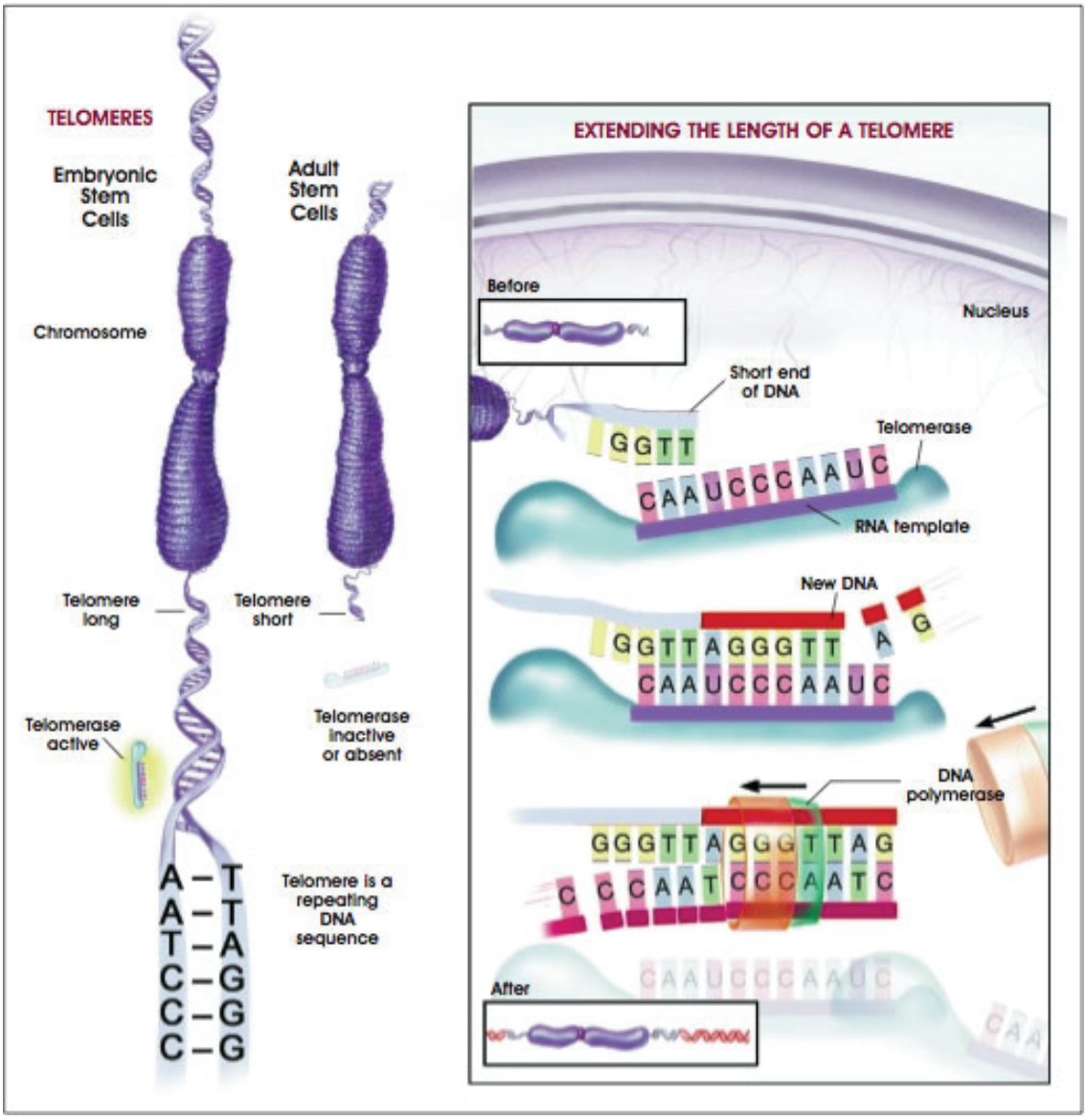


Fig. 5 Structure and extension of telomeres (stemcells.nih.gov)

These structures are necessary to ensure the total replication of the genetic information within the chromosome. In addition, the telomere-associated proteins ensure the chromosome end is not treated as a double-strand DNA break by the cell machinery. This "capping" of the DNA termini along with the replication function are essential for maintaining the integrity of eukaryotic genomic DNA.

Human telomeres are constituted mainly by a double-stranded region of DNA of sequence $(TTAGGG)_n$ hybridized with its complementary sequence and is

terminated at its 3'-end by a single-stranded region of similar sequence $(TTAGGG)_n$ and of c.a. 150-200 nucleotides.
The maintenance of telomeres is essential for the processes of cell proliferation. During each replication cycle the telomeres are shortened due to the incomplete replication of the lagging strand. Hence, telomere length is an indication of the age of a cell. Senescence is triggered when telomeres reach critical length, which can ultimately lead to cell death.
One of the mechanisms available to the cell to combat the erosion of the telomeric ends involves the ribonucleoprotein telomerase, an RNA reverse transcriptase that uses its own internal RNA matrix to elongate the single-stranded 3'-end of telomeric DNA. Interestingly, telomerase is found inactive in most somatic cells, whereas most known cancer cell types are telomerase active. Therefore, the existing link between telomerase and continuous tumor growth, determined telomerase (or its substrate) as a promising therapeutic (anti-cancer) target.

2. RNA

a. Structural differences from DNA

The main differences between RNA and DNA come from the ribose backbone. The absence/presence of a 2'-OH group is the key to explain why DNA exists mainly in cell under its double-stranded form while RNA is mainly present as a single-strand or adopts alternative secondary structures (extremely important information in this dissertation). However, in some RNA virus genomes RNA will also form a double stranded helix. Unlike DNA, RNA forms an A-form double helix. The RNA double helix differs from that of the DNA double helix because of the presence of ribose, rather than deoxyribose, in the sugar phosphate backbone of the molecule. The addition of a hydroxyl group at the C2 position in the ribose sugar is responsible for the A-form geometry in double stranded RNA. .
Interestingly, RNA can form a variety of highly stable secondary and tertiary structures, such as the hairpin loop (Fig. 6), pseudo-knots (Fig. 7), hammerhead structures or the most common "cloverleaf" form (Fig. 8). It is noteworthy that RNA can contain stretches of both structured and unstructured RNA in a single RNA strand.

Also of importance is that the 2'-hydroxy group on the five-membered heterocycle renders RNA prone to hydrolysis and denaturation. Finally, the complementary base to adenine in RNA is uracil, an unmethylated form of thymine.

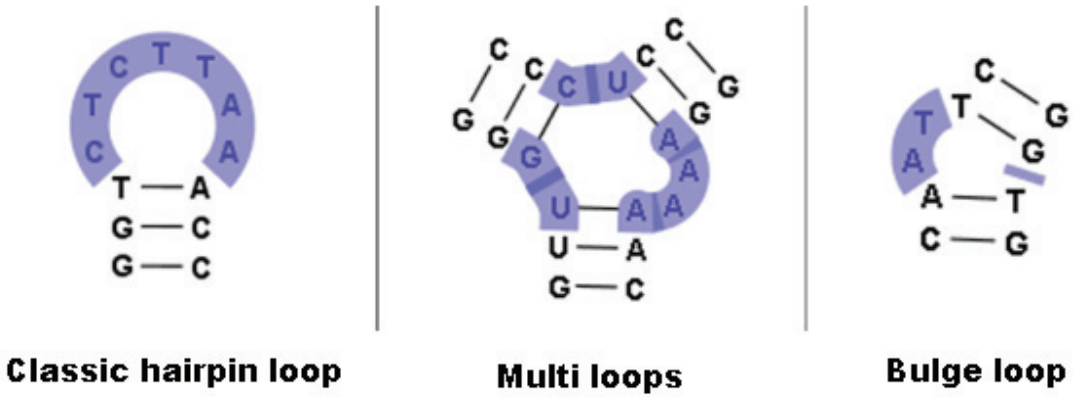


Fig. 6 Loops structures of RNA

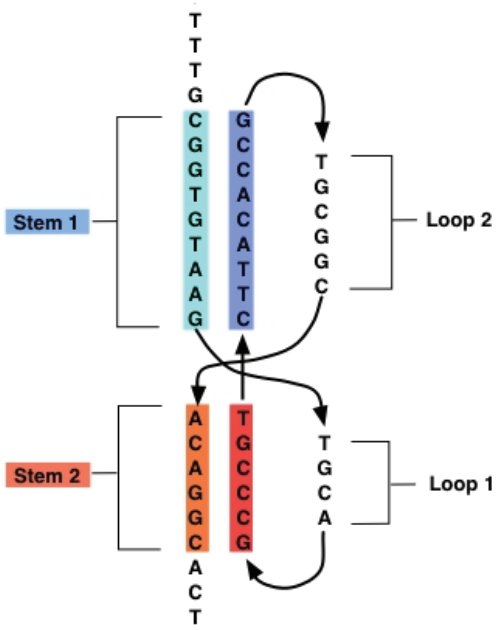


Fig. 7 Pseudo-knot structure

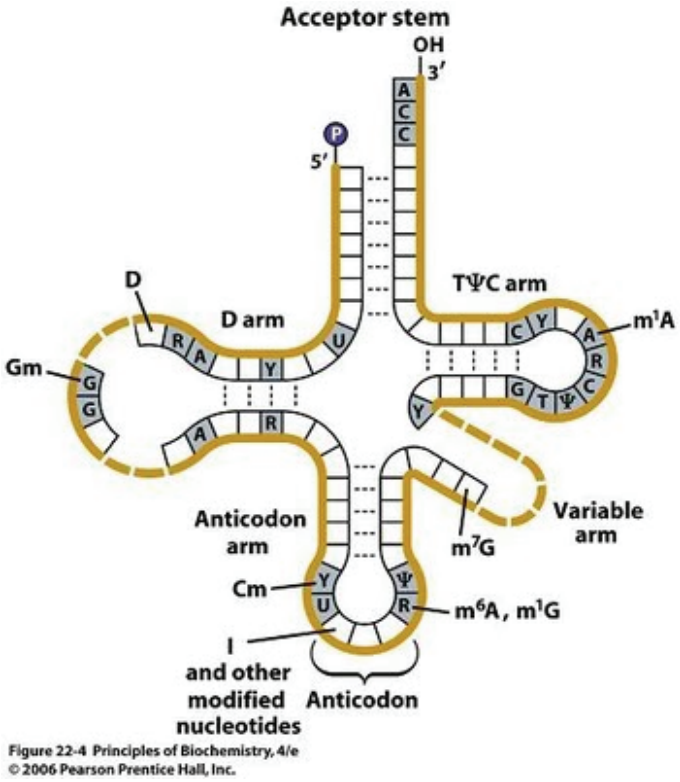


Fig. 8 Cloverleaf structure of RNA

b. <u>Relevance of RNA in genetic information:</u>

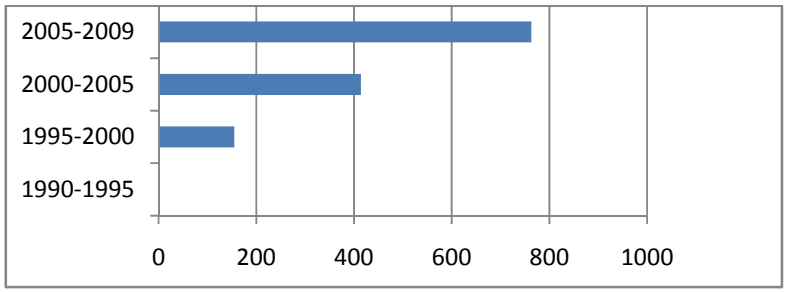
RNA can act as an enzyme alone or by combination with a protein and during translation is responsible for protein formation. The 3 major types of RNA are messenger RNA (mRNA), transfer RNA (tRNA) and the abundant ribosomal RNA (rRNA). Eukaryotic cells contain 4 types of rRNAs, 70 types of tRNAs and thousands of mRNAs.
Comparatively, RNA is the target when interference occurs at the translational level, whereas DNA is targeted when replication or transcription mechanisms are concerned.

We have just seen that nucleic acid oligomers can adopt a large number of secondary and tertiary structures in the context of a cell. In the following section, we will introduce the main subject of this dissertation: the G-quadruplex motif that can be formed by one (or more) DNA or RNA strand(s).

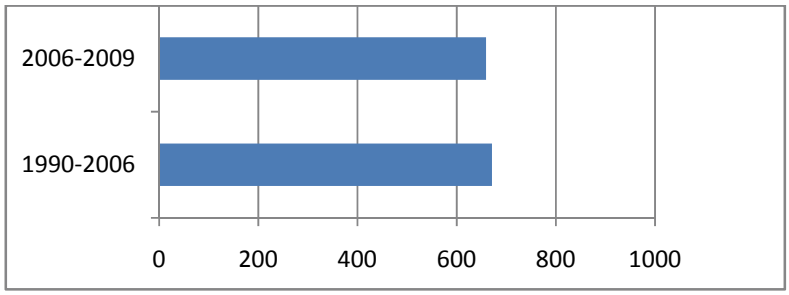
3. G-quadruplexes

a. *Research topic*

We may consider the "explosion" of interest in G-quadruplexes during the last two decades which can be depicted with two histograms:
Using the "PubMed" search engine we can determine the number of 'hits' generated for the word "quadruplex" from the early 90s until today.



Since 1995 the number of hits has doubled every 5 years.
Then from a second search we can plot two periods:



The year 2006 (start of this PhD work) is counted twice, so this method is obviously not entirely precise, but clearly demonstrates the trend. The last three years have proven to be a period of prolific G-quadruplex research.

b. <u>Definition:</u>

"Nucleic acid sequences containing multiple runs of consecutive guanines may fold into a tridimensional structure termed G-quadruplex"

We will try to enhance this description:

Following the base pairing rules established by Watson and Crick the "best" match for guanines are cytosines and the "best" match for adenines are thymines. Over 40 years ago, K. Hoogsteen discovered a novel type of hydrogen bonding between adenine and thymine, termed Hoogsteen base-paring geometry. These pairings require either purines or pyrimidines to interact with the sites on purine bases that are not necessarily involved in Watson-Crick hydrogen bonding (i.e. N7 and O6 for guanines, N7 and N6 for adenines). Since the discovery of Hoogsteen base-pairing, forms of the DNA double helix have been discovered that contain both inter- and intra-strand Hoogsteen base pairs. Following the Hoogsteen type base pairing rules, guanines are also capable to hydrogen bond with themselves to form a tetrad of (four) guanines where each guanine serves as a donor as well as acceptor of two Hoogsteen hydrogen bonds. The first quartet was identified in 1962[6] following the aggregation of 5'-guanosine monophosphate, and it was realized in the late 1980s that G4 resulting from those quartets may play a biological role through DNA.[7-9]

Guanine shows indeed a unique 90° hydrogen bonding orientation of its donor and acceptor sites (Fig. 9 and 10). Therefore the assembly, via supramolecular hydrogen bonding, of four guanosines allows the formation of a planar array of guanines (called G-tetrad or tetrad of guanines or G-quartet). Monovalent cations such as Na^+, K^+ or NH_4^+ are essential for the cohesion of the structure by binding specifically to guanine O6 carbonyl groups between the planes of the G-quartets. Due to the polymeric structure of DNA/RNA more than one tetrad may form and create a non-covalent assembly named ***quadruplex of guanines (also called G-quadruplex or G4)*.**

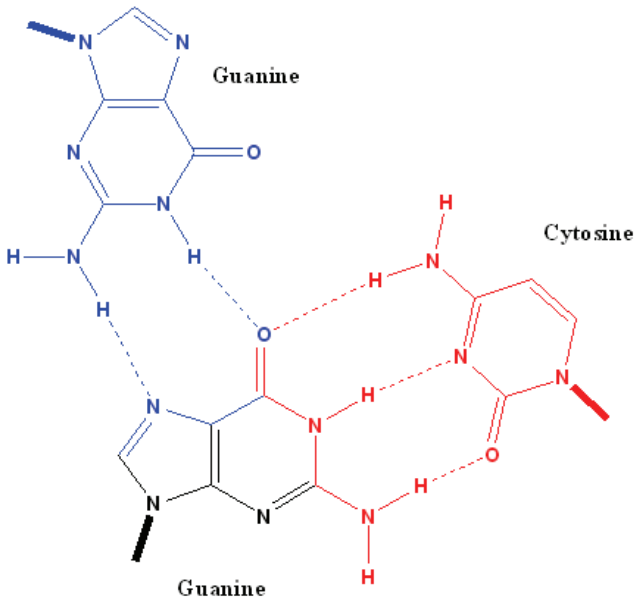


Fig. 9 Hoogsteen pairing (blue) vs. Watson Crick pairing (red)

To clarify, herein the name “quartet” or “tetrad” refers to the assembly of four guanines with or without a metal cation. The term “G-quadruplex” refers to the tertiary structure resulting from the stacking of at least two of those tetrads on top of each other in the presence of cations. The enthalpy gain via hydrophobic interactions is responsible for the π-stacking of the tetrads. It is noteworthy that the center of a G-quadruplex (G4) structure behaves like an ion channel. The flow of cations results from a breathing mechanism. In summary, a G4 is a supramolecular architecture resulting from the self-assembly of nucleic acids via hydrogen bonding, π-stacking and cation recognition.

It has been studied mainly in water near physiological conditions but G-quadruplexes may even exist in organic solvents such as acetonitrile when using lipophilic guanosine derivatives and potassium/cesium picrate.[10]

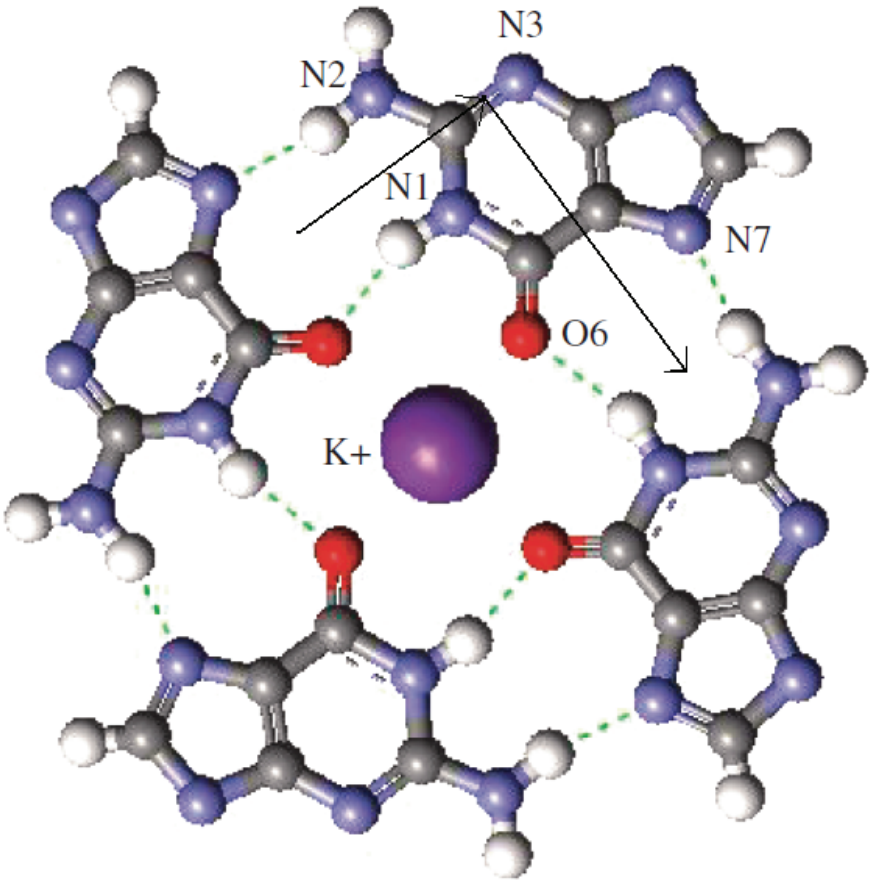


Fig. 10 G-quartet with a central potassium cation[3] and the 90° angle between two consecutive guanines.

Quadruplex stability is closely linked to the nature and concentration of cations coordinating the central oxygen (O6) atoms that form the G-tetrad central cavity (table 1.). Ions that coordinate effectively these 4 oxygen atoms enhance quadruplex stability. Increasing the monovalent cation concentration adds little to the dissociation of the intermolecular quadruplexes but contributes significantly to their rate of refolding. A general trend in alkali ions from the most to the least stabilising is as follows $K^+ > Na^+ > Rb^+ > NH_4^+ > Cs^+ > Li^+$. Possible stabilization of one folded state over another was first described by Sen and Gilbert[11] as a Na^+-K^+ switch. It has biological relevance as the $[K^+] >> [Na^+]$ in the intracellular region, while the situation is reversed for the extracellular region.

Ionic radii of quadruplex stabilizing cations							
Element	K^+	Na^+	NH_4^+	Rb^+	Cs^+	Li^+	Ca^{2+}
Ionic radius	1.52	1.16	1.43	1.66	1.81	0.9	0.99

Table 1 : Different cations that stabilize G-quadruplexes and their ionic radii [3]

Amongst the basic structural information developed later about G-quartet and G-quadruplexes we may already note that they can be either intramolecular (formed from a unique DNA or RNA strand) or intermolecular (formed by the assembly of 2, 3 or 4 different strands) and that the guanines possess two conformations syn/anti (Fig. 11) for each quartet.

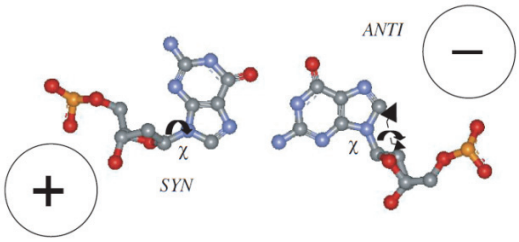


Fig. 11 Guanosines syn/anti orientation and strand polarity [3]

Due to hydrogen bonding (2.8-2.9Å) a quartet alone presents a square surface with sides of 11.6 Å. G-quadruplex dimensions depend on the number of quartets stacked and of the type of central cation. They share similarities with the right handed B-DNA helix: the average height between two quartets is 3.2Å and from quartet to quartet the guanines are tilted by 30°. The groove width is about 11Å and helix width is 23Å.

An extensive hunt began then to identify G-quadruplexes, and now their existence in vitro is proven in DNA as well as in RNA. Moreover there is now increasing evidences that such structures could also form *in vivo*. It was already introduced in 1991[12] and detailed in 2001 with specific antibodies for telomeric G4 DNA in *Stylonychia lemnae* macronuclei[13]: high-affinity single-chain antibody fragment (scFv) probes for the guanine-quadruplex formed by the *Stylonychia* telomeric repeat suggests that guanine-quadruplex is resolved during replication. For recent (2009) review see Lipps and Rhodes[14].

According to the current state-of-the-art, G4 may occur in the following areas (for reviews about biological functions of G-quadruplexes:[15, 16]):

-Telomeres:
The most studied example of biologically relevant G-quadruplex to date is that formed by the single-stranded 3'-end of telomeric DNA. As described previously, telomeres contain a single-stranded 3'-overhang of sequence $(TTAGGG)_n$ or $(TTTTGGG)_n$ in humans and *Oxytricha Nova*, respectively that can form highly stable G-quadruplex structures under near-physiological conditions. Stabilizing or inducing the formation of those structures in vivo has been proposed as an efficient way to inhibit the process of telomere elongation by the enzyme telomerase. This concept will be presented in more details along this manuscript.

-Promoter regions of genes:
Promoters are DNA sequences that bind to the RNA polymerase II enzyme, which is responsible for the generation of mRNA. The promoter region is usually assumed to be the key cis-acting regulatory region that controls the transcription of adjacent coding region(s) into mRNA, which is then directly translated into proteins. DNA sequences within promoters can be identified as binding sites for transcription factors which may cause activation or repression of transcription. It has recently been proposed that formation of quadruplexes in those regions could affect transcription efficiency. Two examples will be developed in the following paragraphs.

-Recombination sites:
Recombination via interstitial pairing may occur in the presence of a protein-mediated bimolecular quadruplex formation [17]. For a recent document dealing with recombinant prone regions, see Mani *et al.*[18]

-5'-UnTranslated Regions (5'-UTRs):
More recently, it has been proposed that formation of a stable RNA quadruplex in the 5'-UnTranslated Region of a gene could represent a new mean to regulate the expression of a gene at the translation level. [19, 20] This concept will be developed further in Chapter 3.

c. *DNA Quadruplex and Gene Regulation*

To better understand the possible roles G4 could play in gene regulatory mechanisms, we will discuss in more details two examples from the Hurley and Balasubramanian's groups:

i. Quadruplex-mediated regulation of the c-myc oncogene

The expression of the c-MYC oncogene is believed to promote the proliferation of cells and has been linked with several malignant cancers. The primary control element is a guanine rich segment of DNA known as the nuclease-hypersensitivity element. The G-rich sequence Pu27 d(TGGGGAGGGTGGGGAGGGTGGGGAAGG) identified in the nuclease-hypersensitivity element (NHE) III_1 was shown to fold into an intramolecular quadruplex *in vitro* under near physiological conditions. In addition, the cytosine rich complementary strand is known to form a i-motif (via protonation of cytosines) structure *in vitro,* although under non physiological conditions. Both of these non-classical structures are believed to play a role in c-MYC transcription. Recent work from Hurley *et al.* has shown that c-MYC expression can be down-regulated in vivo in the presence of the porphyrin TMPyP4, a well-characterized G-quadruplex stabilizing ligand[21]. In addition to that, mutation of guanines into this region to prevent G-quadruplex formation has been shown to lead to an increase in the expression of the c-MYC gene. As a result of this work a quadruplex mediated mechanism for regulation of genes associated with cell proliferation was proposed (see Fig 12a.). Chowdhury *et al.* also demonstrated through competition duplex/G4 that the formation of a G-quadruplex structure was possible despite the unfavorable duplex-quadruplex equilibrium. [22],[23] (For recent review about c-MYC regulation by Laurence Hurley see[24])

ii. Regulation of the c-kit oncogene

The human proto-oncogene c-kit encodes for a 145-160 kDa membrane-bound glycoprotein belonging to a family of growth factor receptors with tyrosine kinase activity. The c-kit gene is critical for the development of mast cells and melanocytes and is an attractive target in the treatment of gastrointestinal tumors (GIST). The c-kit expression levels are also maintained in a number of other types of tumors, such as prostrate and adenocarcinoma lung cancers, while for instance c-kit expression is diminished or absent in breast cancers. Joint work by the Balasubramanian and Neidle groups led to the discovery, in the promoter region of the c-kit gene, of two

G-rich sequences that were shown to form a quadruplex motif in vitro under physiological conditions. A first sequence, located 87 base pairs upstream of the transcription start site of the human c-kit gene and named c-kit87up, was first shown to form a single G-quadruplex species in K^+-containing solution. This 22-nt sequence consists of four runs of three guanine bases, separated by three regions consisting of a single A residue, a CGCT and a AGGA fragment, respectively. A second sequence d(CGGGCGGGCGCGAGGGAGGGG) (c-kit21) located between −140 and −160 bp upstream of the transcription initiation site and close to binding site of the transcription factor Sp1 was found to form a stable quadruplex in vitro. Interestingly, this sequence was also conserved in several organisms (human, rat, chimpanzee, mouse) [25]. An isoalloxazine ligand that binds to and stabilizes the c-kit quadruplex was shown to down-regulate the expression of the c-kit gene in various cell lines, thus reinforcing the hypothesis that quadruplexes in the promoters of genes can act as gene regulatory elements in vivo. [26]

d. Prevalence of G-quadruplexes in the genome:

An attempt has been made to predict which parts of the genome were likely to form G-quadruplexes. With the help of a computer program *quadparser*, about 376000 quadruplexes were found to potentially exist simultaneously in the human genome. [27]

The program is set to evaluate a Possible Quadruplex Sequence (PQS): $X_{3+}N_{1-7}X_{3+}N_{1-7}X_{3+}N_{1-7}X_{3+}$ where X_{3+} is a run of at least three units of G, C, T or A, and N_{1-7} is any type of loop containing the 4 classic nucleic bases up to 7 bases long. Two trends then appeared:

-Strong absence of G-PQS compared to C, A, or T in the exonic regions suggesting the absence of G-quadruplexes in the potential mRNA sequences. (However the first intron of human genes shows clear RNA G4 supremacy but they will not exist in the mature mRNA[28])

-Loop length is an important factor in the existence of G-quadruplexes and the most represented loop length is one nucleotide ($G_{3+}N_1G_{3+}N_1G_{3+}N_1G_{3+}$).

Further studies provided fundamental information about the presence of G4 in gene promoters: amongst 19268 genes studied there are 14769 G-PQS in the promoter regions and at least 42.7% of the gene promoters contain 1 G-PQS (i.e. may form a G-quadruplex.)

(For more information about PQS and the hunt for G-quadruplexes in the genome:[29, 30])

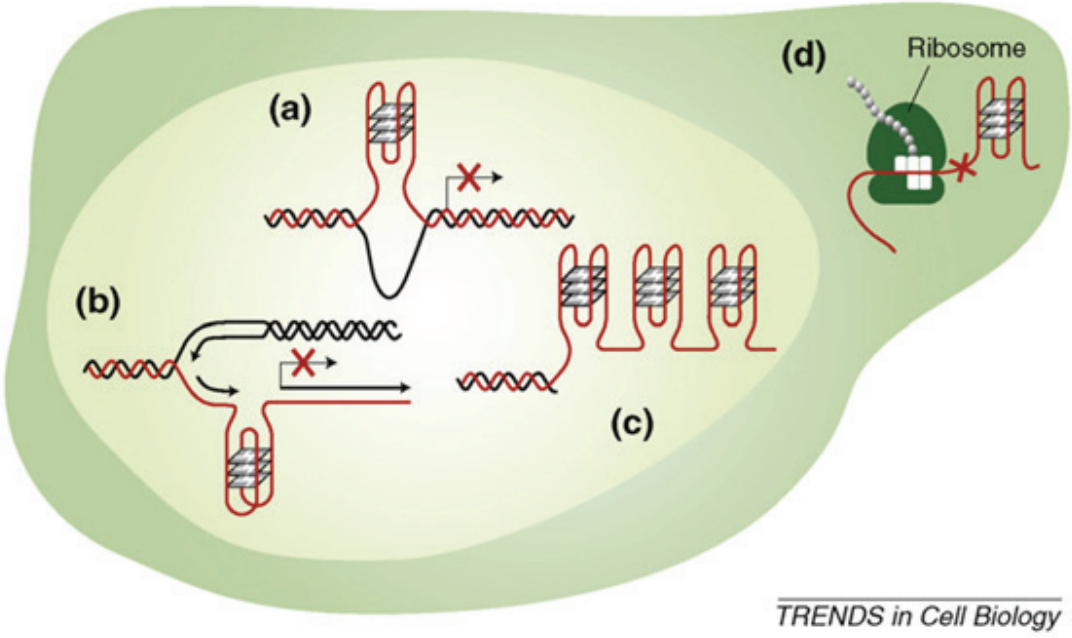


Fig. 12 Possible locations of G-quadruplex structures in cells. Genome-wide analyses have pinpointed the location of G-PQS. Within the nucleus, G-quadruplex formation could occur in G-rich regions of double-stranded DNA, such as at promoter regions of genes (a) and during replication (b) when DNA becomes transiently single stranded. Telomeric G-overhangs are single-stranded, favoring G-quadruplex formation (c). (d) Outside the nucleus, G-quadruplexes could form in mRNA and indeed the 5'-UTRs of mRNAs are enriched in PQS. The red crosses indicate possible impediments caused by G-quadruplexes to replication, transcription and translation.[14]

e. <u>Importance of the loops in quadruplex structure and stability:</u>

The role that the loops play in the existence of G-quadruplexes has been studied, as has been their influence on the quadruplex structure and stability:

With respect to G-quadruplex structures, loops are part of their polymorphism and may be lateral or diagonal when they appear on a quartet surface, and they will be lateral/edgewise (type 1 or 2) or (double chain-strand reversal/propeller/

external) when they are located from top to bottom (or bottom to top) tetrads. Only certain loop combinations are plausible[31] but this will not be discussed in this manuscript. The existence of a strand reversal loop is interesting as it allows the quadruplex to adopt a full parallel structure when at least 2 loops are in such conformation.[32] If loop lengths can strongly influence the topology and stability of intramolecular G-quadruplexes few studies tried to elucidate the effect of loop length on the biophysical properties of G-quadruplexes in a sequence-independent context. In 2008 Bugaut *et al.*[33] in a paper focusing on 1-3 nucleotide loops length emphasized that:

-2 loops may determine the whole conformation of the system especially in the case of 2 nucleotides loops.
-G-quadruplex structures containing short loops are more stable making the existence of this structure more likely, especially given that shorter loop lengths will result in a shorter competing duplex with a correspondingly lower stability. One nucleotide will also favor parallel structure.

f. Quadruplex and Watson-Crick duplex competition in DNA:

If DNA G-quadruplex structures exist *in vivo,* prior to any folding the DNA double strand must unwind and open, i.e. a quadruplex vs duplex competition therefore exists. Indeed we just discussed how to identify the presence of quadruplexes in the genome focusing the "hunt" on guanines, but except for telomeric single stranded DNA a future quadruplex forming sequence may also be found looking for its complementary strand. The Watson-Crick base-pairing rules developed earlier in this general introduction teaches us that guanines are almost exclusively paired with cytosines via three hydrogen bonds within a double helical structure.

To summarize this: Li *et al.* published in 2002[34] (Fig. 13) an interesting analysis on how the duplex formed with $dG_3(T_2AG_3)_3$ and $d(C_3TA_2)_3C_3$ may change its conformation into duplex - quadruplex and I-motif tetraplex:

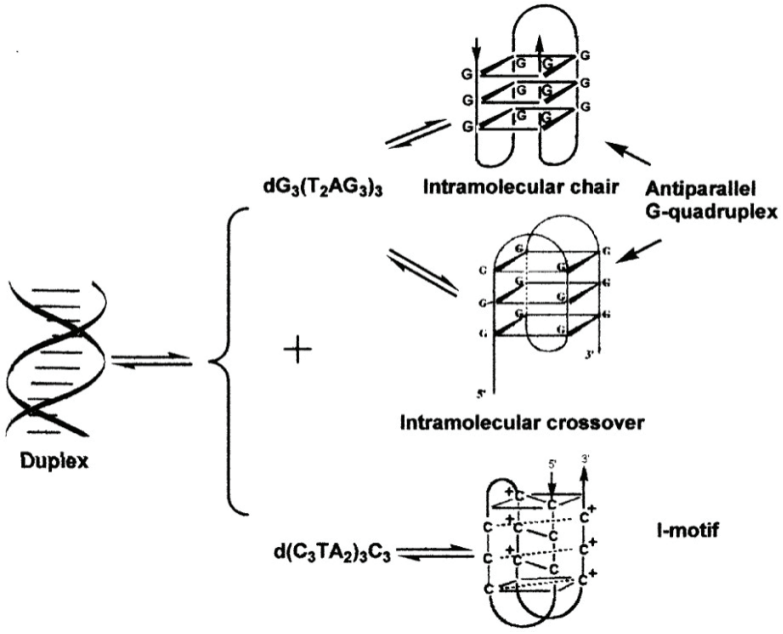


Fig. 13 Competition between duplex and quadruplex /I-motif

A complete study on the stability of intramolecular G4 *vs* their respective duplexes has been reported by Risitano and Fox.[35] taking into account the effect of the DNA sequence, length of guanine runs and metal ions on this equilibrium. As developed earlier, the tetraplex–duplex switching model of c-myc where tetraplex means repression and duplex expression proves that this equilibrium is relevant when examining the possible role of promoter quadruplexes.[21]

Importance of "non-duplexes": G4 can exist in DNA as well as in RNA. But the double helix structure of DNA requires it to open for those structures to exist. This prerequisite doesn't occur in RNA which is mainly present in cells under its single stranded form (hence quadruplex does not need to compete with a highly stable duplex form). (The author acknowledges that the existence of some RNA G4 may require the double helix RNA to open). This important concept will be developed further in the introduction of Chapter 3.

g. Structure and Polymorphism

i. Possible Structures

Depending on the localization and the number of DNA strands involved (dictated by the length of consecutive guanines runs) the resulting G-quadruplex can exist with different stoichiometry. The strand orientation (5'-3') and the stereochemistry of the sugars (syn / anti) attached to guanines also generate specific structures. The DNA double helix is anti parallel and such strand orientation is also observed in quadruplex DNA. However quadruplexes also have the ability to form parallel structures, and whether a quadruplex is anti-parallel or parallel affects several features. Notably the positioning of interconnecting loops in dimeric and monomeric quadruplexes is directly affected by the orientation of the DNA strands. (Fig. 14 and 15)

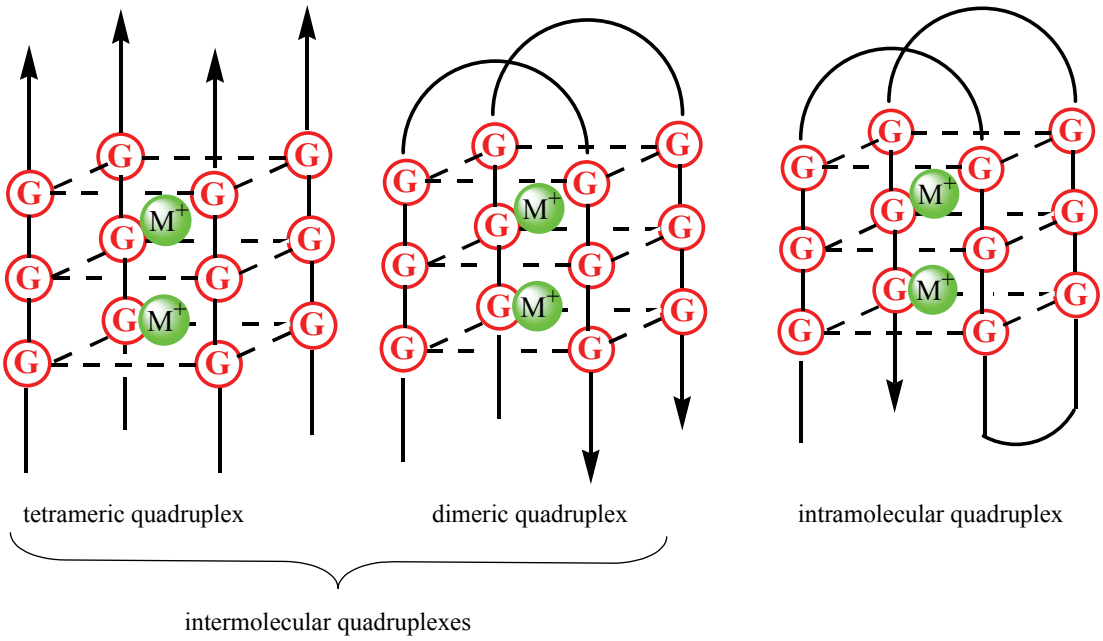


Fig. 14 Three examples of G-quadruplexes containing three G-tetrads and of various stoichiometries (i.e. formed from the association of 1, 2 or 4 strands).

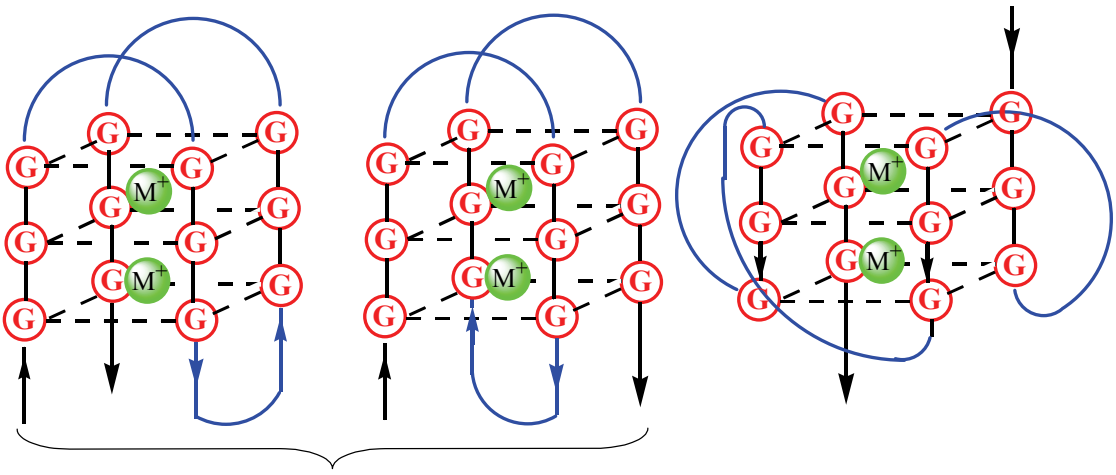


Fig. 15 Three examples of intramolecular G-quadruplexes showing three different loop and strand polarities. (from left to right: 3 lateral loops, two lateral and one diagonal loops, 3 external/chain reversal loops)

ii. The structure of human telomeric quadruplex DNA

There are very few examples in the literature of quadruplex structures resolved by the use of X-ray crystallography or NMR spectroscopy. The intramolecular quadruplex formed by four repeats of human telomeric DNA $(TTAGGG)_4$ is a therapeutic target in telomerase-active cancer cells and has therefore been extensively studied. Two important studies have then been carried out using different techniques and different experimental conditions and ultimately arrived at different conclusions (i.e. significantly different structures), thus highlighting the great polymorphic nature of DNA quadruplexes. As an example, we decided to describe below in more details the first two structures of the intramolecular human telomeric quadruplex solved is sodium and potassium containing solution. More recent structures of quadruplexes formed from human telomeric DNA will be discussed in Chapter 2.

1. Human telomeric quadruplex DNA with sodium[36]

In 1993 the structure of the quadruplex formed by folding of the DNA sequence $d(AGGG(TTAGGG)_3)$ in sodium-containing solution was reported. The structure was

solved using NMR spectroscopy combined with a molecular dynamics approach. In the presence of sodium the quadruplex adopts an antiparallel structure with three stacked tetrads. Of the three interconnecting TTA loops, two are lateral while the other is a diagonal loop. This loop connectivity ensures that any one strand in the structure has a neighboring parallel and antiparallel strand. Three different groove dimensions result from this strand arrangement. The two sets of parallel strands encompass a "medium" groove while the remaining two sets of antiparallel strands line the edges of a "wide" and a "narrow" groove. The presence of the differing groove sizes in the quadruplex is down to the relative orientation of the deoxyribose rings in the neighboring parallel or antiparallel strands (Fig. 16).

2. Human telomeric quadruplex DNA with potassium[32]

The X-ray crystal structure of the same sequence d(AGGG(TTAGGG)$_3$) was solved in 2002 by the group of Stephen Neidle. Here the folding of the intramolecular quadruplex was carried out in the presence of potassium ions. The resulting structure is remarkably different to that solved by Wang and Patel. In this instance all the strands run parallel to each other with each of the three TTA loops running along the sides of the structure in a propeller like arrangement. Each guanine is present in the anti- conformation. Thus the relative orientation of the deoxyribose sugars between adjacent strands remains the same, presenting four grooves of similar dimension, between 9-10 Å in width (Fig. 16).

Differences between these two solution structures highlight the conformational repertoire of the G-quadruplex. Based on the conditions of these two studies, the structure folded in potassium is likely to be a more biologically relevant conformation, as potassium is present in higher concentrations (140 mM) than sodium in the cells. There is however some controversy over the true structure of this quadruplex forming sequence under different physiological conditions. A recent study investigating platinum cross-linked adducts of the same DNA sequence found the predominant species in potassium to be the "basket" form, as described by Wang and Patel. CD studies on human telomeric sequences suggest the presence of two different conformations in the presence of potassium, a phenomenon observed for other quadruplex-forming oligonucleotides. It is possible that under subtly different conditions the "basket" and "propeller-(hybrid)" forms co-exist in different proportions. Additional structures listed below have been recently reported in the literature that will be developed in Chapter 3:
(3+1): A new structure arises for the human telomeric DNA designed as a bimolecular quadruplex 3+1 type.
Hybrid: The human telomeric DNA was shown to adopt in vitro two so-called hybrid

1 and hybrid 2 structures that can be seen as hybrids of the Patel and Neidle structures.

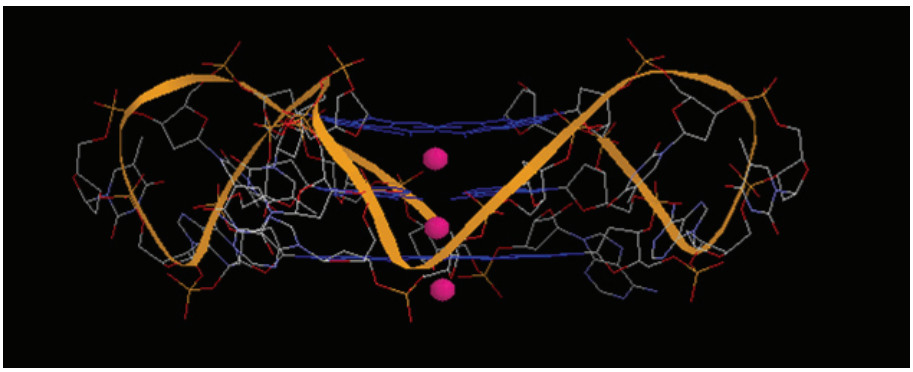
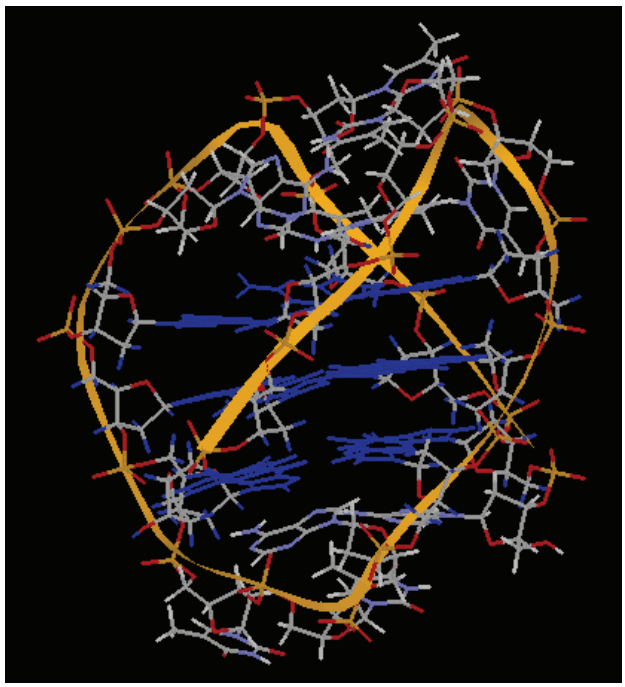


Fig. 16: Top: The crystal structure of the human telomeric sequence quadruplex d[AGGG(TTAGGG)$_3$] obtained in the presence of KCl. All DNA strands were observed as parallel, generating a propeller-like structure (Parkinson *et al.*, 2002).

Below: NMR solution structure of human telomeric sequence quadruplex obtained in the presence of NaCl (Wang and Patel, 1993). This represents an anti parallel strand arrangement generating a basket-type quadruplex structure.



4. Biophysical techniques

A wide range of biophysical techniques are available to investigate nucleic acids folding and their interaction with small molecules and proteins. The following techniques may be used to characterize G-quadruplex formation and their interaction with ligands:

a. Ultraviolet spectroscopy:

One of the first methods that can provide valuable information about the structure of nucleic acids in solution is the Middle UV range (200-300nm) spectroscopy.

i. Nucleic acids Studies:

The pentose and phosphate components of nucleic acids do not exhibit significant UV absorption above 230 nm. However single-stranded polynucleotides (containing π conjugated heterocycles) have a maximum absorption around 260 nm with a molar extinction coefficient close to 10^4. Folding of the polynucleotide into specific secondary and tertiary structures leads to an alteration of the absorption properties, which can be monitored by UV spectroscopy to assess the formation of these structures.

DNA (or RNA) UV melting studies monitored at 260 nm are commonly used for duplex DNA (or RNA) for whom an increase in absorbance of about 25 % is indicative of the melting temperature.
The technique is also used for higher order structure such as triplexes (difficult to interpret due to numerous transitions) or quadruplexes.

ii. G-Quadruplex:

G4 formation/melting is monitored at 295 nm[37] (where the shift in absorbance is higher: 50-80%). Precise determination can be done of not only melting temperature, but also thermodynamic parameters, provided that the profile is reversible. (ΔG=-4.6 kcal/mol for the human telomeric G4 in a pH 7.0 sodium cacodylate buffer containing 0.1 M KCl). The transition at 295 nm is inverted as compared with the melting profile of a classical DNA duplex, i.e. G-quartet association leads to an increase instead of a decrease in absorbance (Fig. 17). The

melting curve shows the second order transition and its derivative minimum indicates approximately melting temperature of the nucleic acid structure being studied.

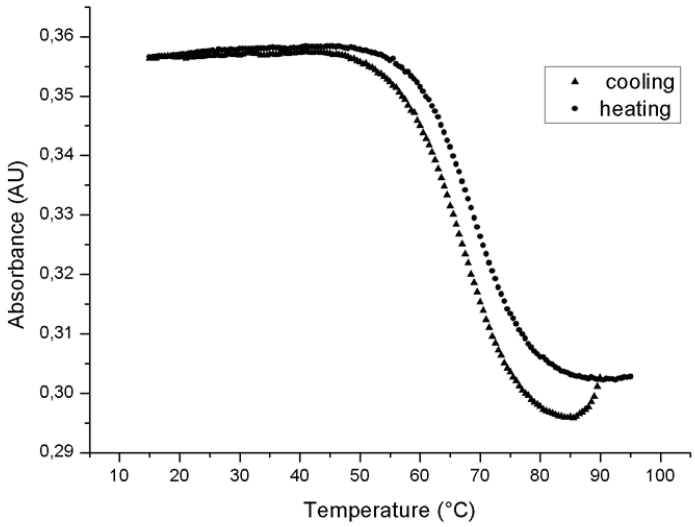


Fig. 17 Example of UV (heating and cooling) denaturation curves.

iii. Ligand stabilization:

UV melting studies can also be used to assess the quadruplex stabilizing potential of a given ligand (small molecule). For such studies, it is mandatory that the ligand itself can endure the melting experiment without decomposition. Stabilizing potential ligands are then expressed as ΔTm, i.e. the difference in melting temperature between a solution of DNA in the presence and in the absence of ligand.

b. Circular Dichroïsm:

Circular Dichroïsm (CD) provides information about the structural complexity of nucleic acid structures. The structural asymmetry induces a change between left-handed polarization and right-handed polarization and the intensity/absorbance difference is measured by CD. CD is reported in ellipticity (deg) or in molar ellipticity (deg.cm^2/dmol), ellipticity corrected to concentration. CD technique can be used in different ways:

-Determination of the folding of one protein or nucleic acid sequence.
-Comparison with data obtained by other means.
-Study of the stability under different stresses like thermal denaturation or pH / salt influence.

In the case of DNA secondary and tertiary structures, quadruplex formation can be detected qualitatively by the existence of maxima (positive bands) at 260 and/or 295 nm, and minima (negative bands) at 240 and/or 260 nm.

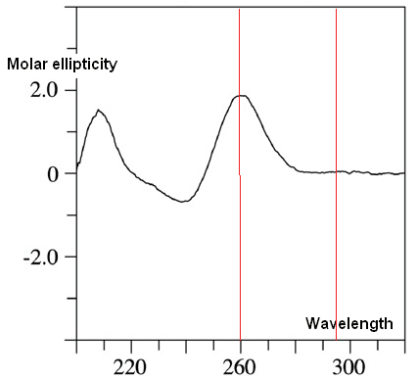


Fig. 18 Example of CD spectrum of a parallel stranded G-quadruplexe

A positive band in the 260 nm region coupled with a negative one in the 295 nm (and 240nm) region indicates a parallel G4 structure (Fig. 18).

A negative band at 260 nm and positive band at 295 nm indicates anti parallel structure or a mixed parallel/antiparallel conformation.
Mixes of positive bands must be studied in more detail and are often due to the structure of external loops and also to the general G4 polymorphism linked to the conditions of the study (cations, pH, and methods).
Like UV, CD can also be used in melting mode to monitor ligand-induced quadruplex stabilization at a fixed wavelength.

Note: CD interpretation must be made carefully, and parallel and anti parallel forms may be at an equilibrium depending for example on the cation used.[38, 39]

c. *Nuclear Magnetic Resonance – NMR*

Solution NMR is a powerful technique to solve quadruplex structures. The prerequisite is to establish the assignment of every imino, amino and aromatic protons through the nucleic acid sequence.[40] NMR requires high concentration of DNA which is a major drawback, besides radioactive enriched oligonucleotides are used to characterize heteronuclear coupling. For the most recent NMR studies of G4, see [41]and[42].

d. Electrospray ionization – mass spectrometry:

ESI method is a useful technique to detect non-covalent interactions between small molecules (or between a small molecule and a macromolecule) and to determine the stoichiometry of the observed stable complexes. However, interpretation must also be made carefully since concentration processes and desalting effects inherent to the technique may alter the complex stoichiometry. It requires small quantities of sample and qualitative results can be easily interpreted. Quantitative analysis remains however difficult due to the necessary evaluation of response factor (comparable to UV molar extinction).[43, 44]
The protocol involves a solution containing the desired complex to be electrosprayed into a Q-TOF mass spectrometer.

Note: All the methods detailed previously (UV, CD, NMR, ESI-MS) can be used to analyze quadruplex conformation and to characterize quadruplex-ligand interactions. The methods below are used only to evaluate the nature of the interaction between a quadruplex target and a putative small molecule ligand.

e. Surface Plasmon resonance-SPR:

SPR is a screening technique showing the binding event (ligand+G4) on a thin metal surface (usually gold). The refractive index of the surface is linked to the number of molecules bound to it. The folded quadruplex is immobilized on the surface and its possible interaction with the ligand will be calculated from the shift in refraction angle (modification of the refractive index).[45],[46]

f. FRET melting assay:

The Förster resonance energy transfer is the fluorescent energy transfer between two chromophores (FRET pair constituted of one acceptor and one donor). Usually they are bound to both ends of the quadruplex forming sequence and the formation

and unfolding of the quadruplex will significantly modify the distance in between both constituents of the FRET pair. The monitoring of the fluorescence emission of the acceptor fluorophore through melting studies in the absence/presence of a ligand allows a rapid and easy evaluation of the quadruplex stabilizing potential of a given ligand. [47] ,[48]

g. G4-FID:

The fluorescent intercalator displacement assay for quadruplexes (G4-FID) consists in monitoring the ability of a given ligand to displace a fluorogenic molecule of thiazole orange (TO) from a DNA G-quadruplex. While TO is non fluorescent when free in solution, it becomes planar upon binding to quadruplexes, thus resulting in an exaltation of its fluorescence properties. Evaluation of the possibility to displace this interaction between G4 and TO by the introduction of a new ligand can therefore be monitored using a spectrofluorimeter (i.e. tight quadruplex ligands will displace TO from the G4, thus leading to a decrease in fluorescence). [49]

5. Ligands for G4:

As described in the previous sections DNA and RNA quadruplexes have recently been considered as promising therapeutic (anticancer) targets whether they form at the end of telomeres or in gene promoter regions. Since the late 90s numerous researchers have designed and engineered small molecules that could target these structures with high affinity and specificity. While those ligands were initially designed to target the human telomeric quadruplex and to prevent elongation of telomeric DNA by telomerase, more recent studies include rational and combinatorial synthesis of ligands. They would bind specific quadruplexes located in gene promoters with a possible application as gene regulators.
In this section, we aim to provide a non-exhaustive list of G4 ligands (proteins and small molecules) that have been engineered in recent years for targeting G-quadruplexes.

a. G-quadruplexes and proteins:

Numerous proteins are known to bind duplex DNA. More recently, natural (and also engineered) proteins that interact with G4 were also identified. These proteins can be classified according to their potential to stabilize, destabilize or promote the formation of G-quadruplexes. While engineering artificial proteins to stabilize quadruplexes proved an interesting alternative to small molecule based approaches, linking G-quadruplexes to natural proteins and natural cellular mechanism would provide indirect proof of the existence of quadruplexes in vivo. The number of natural proteins found to interact with quadruplexes is large: from retroviruses to lower eukaryotes, to yeast and to mammals. A list of those proteins is given below.

Protein	Role - Interaction	Origin	References
Thrombin	Formation promotion, Destabilization, Binding	Human	[50],[51],[52]
MyoD	Binding	Human	[53]
QUAD	Binding	Murine	[54]
Macrophage scavenger receptor	Binding	Bovine	[55]
Tetrahymena binding protein	Binding	*Tetrahymena thermophila*	[56]

G4p1	Binding	*Saccharomyces cerevisiae*	[57]
G4p2/Stm1	Binding	*Saccharomyces cerevisiae*	[58]
quadruplex Telomeric binding protein 42 [CBF-A]	Binding	Murine	[59]
LR1 (nucleolin and hnRNP D)	Binding	Murine	[60],[61]
uqTBP25	Binding	Murine	[62]
Gene 5 protein	Binding	Bacteriophage	[63]
Ku protein	Binding	Human	[64]
Nucleocapsid (Ncp)	Binding , Destabilization	HIV-1	[65]
MutSα	Binding	Human	[66]
Cytoplasmic intermediate filament proteins(cIF)	Binding	Murine	[67]
Pur-1 (aka MAZ and ZF87)	Binding	Rat	[68]
Subunit β of telomere binding protein	Formation promotion	Oxytricha nova	[69]
RAP1	Facilitates formation of G4 in presence of K+	*Saccharomyces cerevisiae*	[70]
DNA topoisomerase I	Formation promotion	Human	[71]
Hop1	Formation promotion	*Saccharomyces cerevisiae*	[17]
Telomere end-binding proteins α and β	Formation promotion	*Stylonychia lemnae*	[72]
MXRN1p	Recognize RNA G4 - catalyses degradation	Mouse	[73]
GQN1 G quartet nuclease	Structure specific cleavage of intermolecular G4	Human	[74]
CArG-box binding protein A (CBF-A)	Destabilization	Murine	[75]
hnRNP A2	Destabilization	Human	[76]
Cdc13p	Destabilization	*Saccharomyces cerevisiae*	[77]
Unwinding Protein UP1	Destabilization	Murine	[78]
Protection of telomere 1 (POT1)	Destabilization	Human	[79]
Replication protein A (RPA)	Destabilization	Human	[80]

Insulin	Binding	Human	[81]
SARS unique domain (SUD)	Binding	Human	[82]
SARS CoV helicase	Binding	Human	[83]
bovine prion protein (bPrP)	Binding	Bovine	[84]
hPARP-1	Protein catalytic activation	Human	[85]
hnRNP A1	Binding, Destabilization	Human	[86]
alpha-hemolysin (alphaHL)	G4 encapsulation	Human	[87]
FANCJ helicase	Unwinding	Baculovirus	[88]
Cytochrome C	Binding	Human	[89]
FMRP	Binding	Human	[90]
Large T-antigen (T-ag)	Unwinding	SV40	[91]
protein arginine methyltransferase 5 (PRMT5)	Protein altered activity and location	Human	[92]
M. jannaschii single-stranded DNA binding protein (SSB)	Binding, Detection	Thermoautotrophic Archeon	[93]
TRF1,2	Binding, Interaction	Human	[94],[95],[96]
eEF1A	Binding	Human	[97]
Zinc finger protein (Gq1)	Binding	Engineered	[98]
EBNA1	Binding	Epstein-Barr virus	[99]
Resolvase (DEHX protein)	Unwinding	Human	[100]
RecQ helicases			
***WRN**	Unwinding	Human	[101]
***BLM**	Unwinding	Human	[102]
***Sgs1**	Unwinding	*Saccharomyces cerevisiae*	[103]

The list above aims mostly to present all currently known quadruplex–protein interactions identified so far.

Some examples can be further developed:

- hPARP1: The human poly (ADP-ribose) Polymerase-1 is abundant in chromatin of eukaryotic cells and is naturally involved in DNA damage response mechanisms but also in transcription regulatory mechanisms. Soldatenkov *et al.* described the first example of a protein which binds to and is catalytically activated upon binding to promoter quadruplexes. (Activity = f[quadruplex formation])

-hRPA: RPA plays essential roles in many aspects of nucleic acid metabolism, including DNA replication, nucleotide excision repair, and homologous recombination. Saintomé *et al.* suggested that hRPA was involved in telomere maintenance through first binding and then unfolding of G-quadruplexes.

-FMRP: This protein is responsible for the mental retardation syndrome. A G-quadruplex mRNA was shown by Darnell *et al.* to be the target of the fragile X syndrome mental retardation protein (FMRP). Mapping experiments revealed that the G-quartet RNA was bound by the FMRP C-terminal RGG box. In a model proposed by Darnell and co-workers, FMRP would recognize G-quartets in untranslated regions of mRNAs to inhibit their translation. This would suggest that (1) RNA quadruplexes form in vivo and are recognized by naturally occurring proteins and (2) quadruplex structures formed by mRNA could potentially be used as translation regulatory elements.

-WRN helicase: Werner syndrome is a genomic instability linked to premature ageing. Fry *et al.* discovered that helicase collaborates with polymerase δ to remove a blocking quadruplex thus allowing the replication process to occur.

A common feature to most of those quadruplex binding proteins is their implication in replication and transcription mechanisms. The same way it was discussed earlier about c-myc and c-kit genes, possible quadruplex formation in the promoter of genes and its subsequent stabilization and/or unwinding by proteins, relate the important link between quadruplex and genetic expression.

Based on this non-exhaustive list we can say that natural proteins are really able to "manipulate" quadruplexes. Although this does not represent a direct and unambiguous proof that quadruplexes exist in vivo and have a biological function, there exist increasing and converging evidences that quadruplexes can play a role in gene regulation and therefore represent a very attractive target for anti-cancer therapeutics.

b. Chemicals

The design and optimization of small molecules that bind to and stabilize G4 structures has been rather prolific during the last decade. These "G4 ligands" are small molecules that are designed to bind quadruplexes as specifically as possible

(G-quadruplex >> DNA duplex) with potential applications as anti-telomerase or anti-cancer therapeutic agents. They usually interact with G4 via π-stacking interactions (essentially with the G-tetrads) and/or by targeting the G4 loops and grooves (e.g. BRACO 19). To illustrate this, a modest attempt has been made to present herein a selection of the most representative examples of G4 ligands. Numerous small molecules have been synthesized to target different features of the G4 structures. In this manuscript 24 have been selected that are represented below plus G4 ligand BRACO 19 that will be the subject of a specific paragraph.
Most G4 binding ligands designed to date are flat aromatic platforms that target the top G-tetrad of the quadruplex ("molecular chaperone"). In order to anchor them onto the quadruplex, those DNA binding platforms are functionalized with "arms" or side chains to create favorable interactions with the loops of the quadruplex. Side chains are usually amines such as pyrrolidines, quinolines or (N,N-dimethyl)-ethylene diamine which are protonated at physiological pH and can therefore be engaged in electrostatic interaction with the negatively charged DNA (or RNA) phosphate backbone. Most of those molecules appear bi-dimensional although they can rearrange when facing a quadruplex. Molecules **1-10** and BRACO 19 belong to this category.
"Arms" may be misused here but we can imagine those structures embracing G-quadruplexes.

-Molecule **2** from the Mergny's group is also called 360A (2,6-N,N′-methyl-quinolinio-3-yl-pyridine dicarboxamide). It is a member of a new family of pyridine derivatives that interact highly selectively in vitro with G-quadruplexes compared with double-stranded DNA and inhibit telomerase activity. This G-quadruplex ligand induces delayed growth arrest followed by massive apoptosis in various immortalized cell lines in direct correlation with telomeric instability.

- Molecule **6** is a fluoroquinolone and represents a close structure (maybe the structure itself) of Cylene pharmaceuticals leading compound and pro-apoptotic antitumor agent CX-3543/Quarfloxin which is undergoing Phase II clinical trials. CX-3543 directly inhibits rRNA biogenesis by targeting a protein (nucleolin?) that is amplified during Pol I transcription of the rRNA genes of cancer cells. This protein is selectively disrupted by CX-3543 and induces apoptotic cell death in cancer cells. CX-3543 demonstrates in vivo efficacy with a broad therapeutic window in xenograft models using multiple types of cancers, and biomarker analyses supported the in vivo anti-tumor effect on rRNA biogenesis leading to apoptosis of tumor cells. Thus, inhibition of rRNA biogenesis is an important new approach in oncology drug discovery and development.

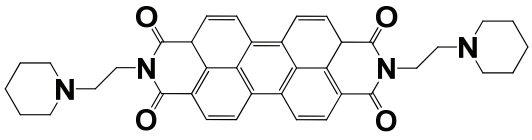
- Molecule **10** is one of Neidle/Thurston's most recent ligands for quadruplexes, showing good stabilizing properties and cytotoxicity.

A second class of ligands is that constituted by a large aromatic platform only with no additional flexible arms. We can include in this category ligands **11-17**. We can find here, the well-known daunamycin **12**, berberines **15**, fluoroquinoanthroxazines **16** or acridines **13**(RHPS4)/**17**.

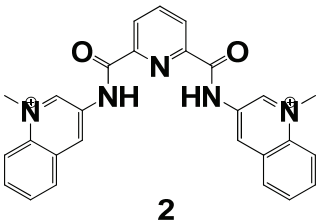
A third family of G4 binders is that of macrocyclic ligands:

- Molecule **19** (or Telomestatin) is a macrocyclic natural product extracted from *Streptomyces anulatus* 3533-SV4. It remains one of the most potent telomerase inhibitor to date.
- Molecule **20** (or TMPyP4) is a cationic porphyrin which has been one the first type of molecules that have been reported to bind G4 with high affinity. Its cyclic shape and π-stacking ability make it an excellent DNA binder although this ligand suffers from a very poor quadruplex/duplex specificity.
- Molecule **23** was selected using a dynamic combinatorial approach and shows a high affinity for the quadruplex c-kit21 (from the promoter of the c-kit gene) and promising selectivity vs other quadruplexes.

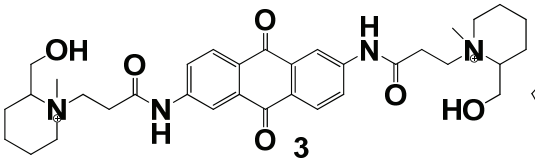
Important considerations when designing some specific ligands is the affinity towards G4 compared to duplex DNA. For example **2** binds 50 times more tightly to G4 than to duplex whereas **15, 22** are not selective at all.
9 and **19** have no detectable affinity for duplex. **13**, **18** and **21** have respective G4 vs duplex specificity of 32, 10 and 15. Usually the range of the binding constant is 10^7 M^{-1} as determinated by SPR, FRET melting or other techniques.
More recently, ligands that can discriminate between one G4 and another for instance isoalloxazines, display selectivity for c-kit2 and c-kit1 G4s over the human telomeric G4[26]. Differential recognition of G4 can also be achieved through architecture control[104].



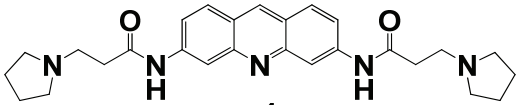
1



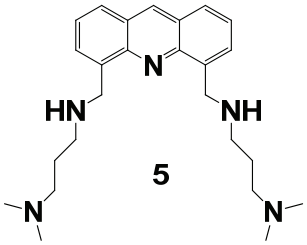
2



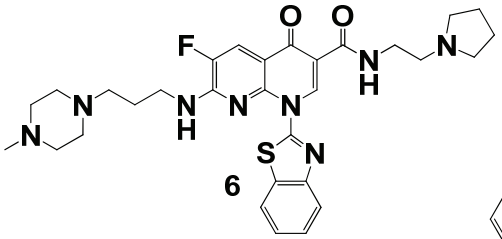
3



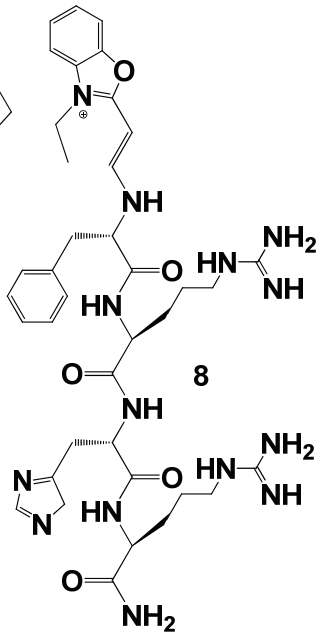
4



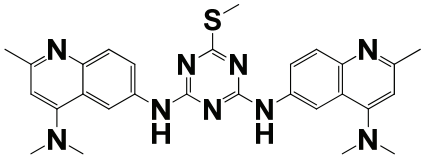
5



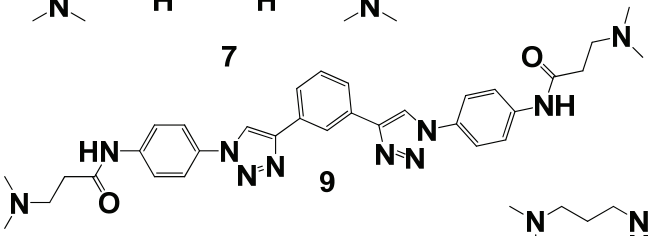
6



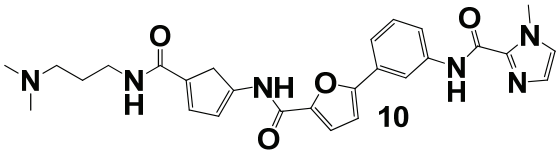
8



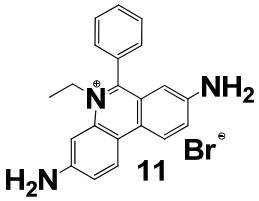
7



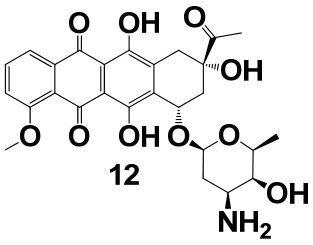
9



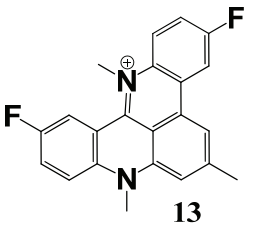
10



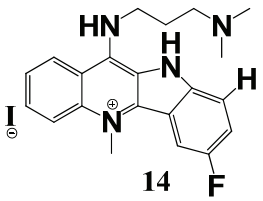
11



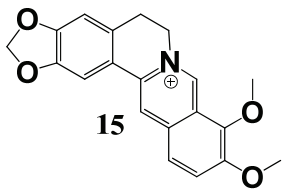
12



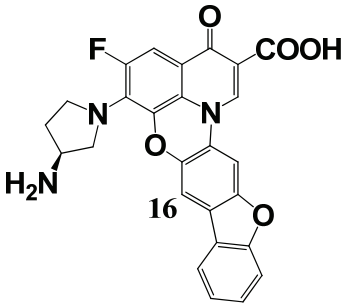
13



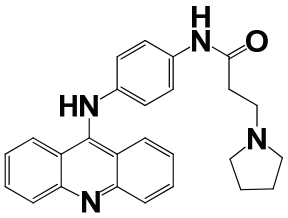
14



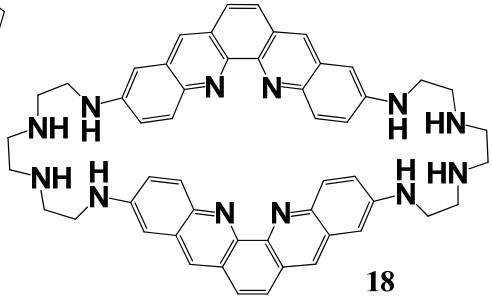
15



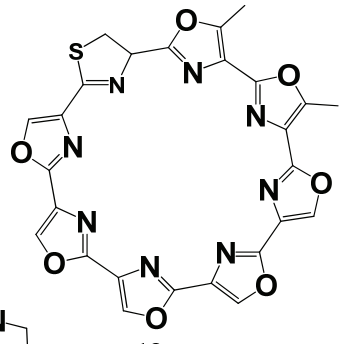
16



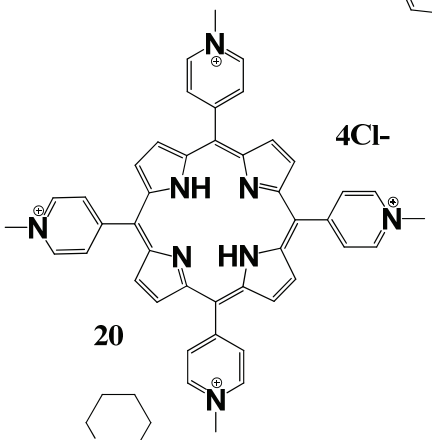
17



18

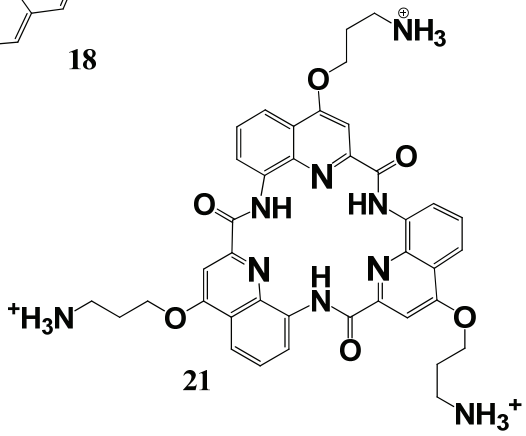


19

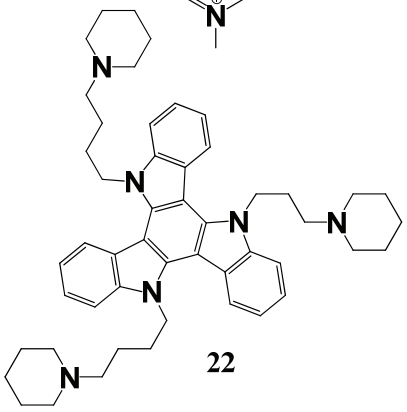


20

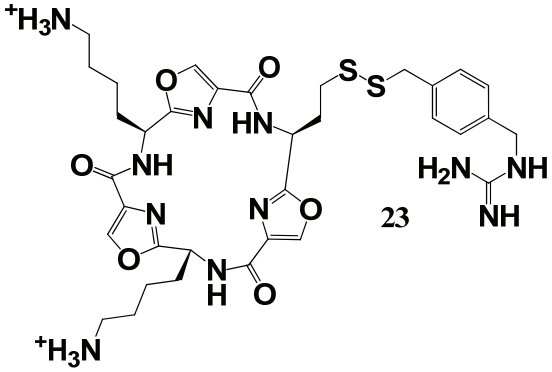
4Cl-



21



22



23

Mol.	1	2	3	4	5	6	7	8	9	10	11	12
Ref.	[105]	[106]	[107]	[108]	[109]	[110]	[111]	[112]	[113]	[114]	[115]	[116]
Mol.	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	
Ref.	[117]	[118]	[119]	[120]	[121]	[45]	[122]	[123]	[26]	[124]	[125]	

c. *What else can bind or induce the formation of a quadruplex?*

An introduction to the quadruplex binding ligands would not be complete without referring to the possible roles played by PNAs. Unlike the ligands listed in the previous section, PNAs are not classic ligands but were reported (mainly by the group of Armitage) to invade quadruplexes. Example of pure PNA, PNA-RNA and PNA-DNA quadruplexes were also reported in the literature and will be developed further in Chapters 2 and 3.

Briefly, Peptide Nucleic Acids (PNAs) are DNA analogues with a peptide backbone instead of the polyphosphate one. (Fig. 19)

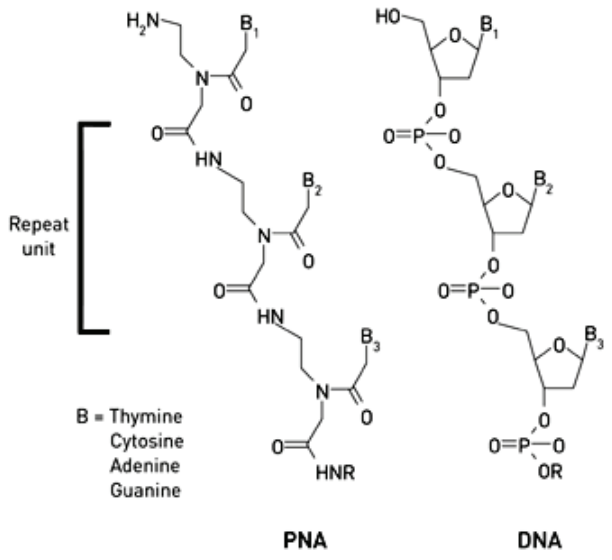
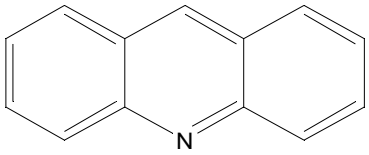


Fig. 19 Structural differences between PNA and DNA (complex.upf.es)

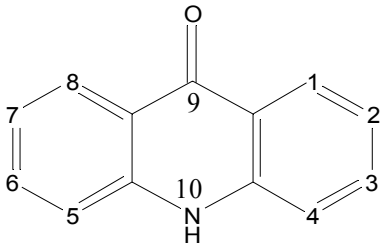
They can hybridize with PNA, DNA or RNA and form anti-parallel double helices via Watson-Crick base pairing. Duplex Stability : PNA:PNA > PNA:DNA > DNA:DNA

6. Acridines

Acridine is a naturally occurring product in coal tar *("être dans le coaltar")* and was isolated in 1871 by Graebe and Caro. It is an anthracene like heterocycle containing a nitrogen atom on its central ring.



Acridone (acridin-9(10H)-one) is the ketone derivative from acridine. In 1880 Graebe and Caro reported oxidation of acridine, which is the first identified synthesis.

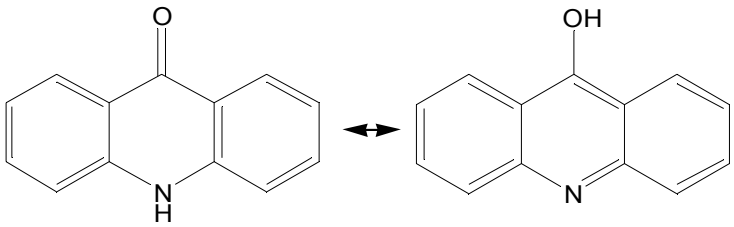


In 1892, Graebe and Lagodzinski reported a similar compound upon treatment of phenylanthranilic acid with concentrated sulfuric acid. This was followed later by Ullman's route to diphenylamines (1905) which facilitated deeper studies of this molecule.

Acridones are planar systems which may interact with biomacromolecules via intercalation or Pi-stacking.

The chemical behavior of acridones depends on the nature and position of their substituents. They are typically solids with high melting points (>300°C) and generally insoluble in ether, water, or methylene chloride.

Noteworthy is the observation that the central ring allows a tautomeric equilibrium which prefers the ketone form (left).



Acridones and acridines are fluorescent molecules (λ_{em} > 400 nm) so they can be detected and quantified using UV or fluorescence spectroscopy. (Quantum yield of acridine orange in basic ethanol is 20% for 400 nm excitation)
Acridines usually bind to B-DNA by intercalation of its polycyclic aromatic ring between adjacent base pairs of the double helix.

(for an excellent review about "acridones/ines" see ref[126])

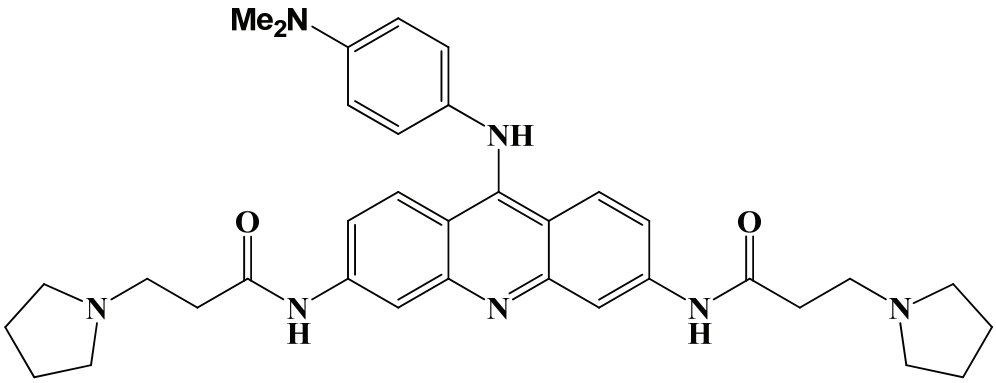
7. Braco19 and Quadruplex recognition

a. *History of BRACO 19*

Introducing the BRACOs family is important for many reasons: first it gives an example about how quadruplex ligands are designed and optimized, secondly because BRACO 19 is one of the best known and most studied G4 ligand nowadays and finally it belongs to the acridine family which is the main focus of this manuscript.

The story of BRACO 19 began in 1999-2001 (start of the patent[127]) after the previous studies of anthraquinone derivatives as therapeutic agents. BRACO molecules are series of 3, 6, 9 tri-substituted acridines. However the development of quadruplex-binding acridines started in 1999 with the study of 3, 6 bi-substituted acridone/acridine derivatives as telomerase inhibitors and G4 ligands. These first generation derivatives proved to be good ligands but with low specificity for quadruplex versus duplex DNA. In order to overcome this specificity issue, trisubstituted acridines were rationally designed by introducing a 9-amino substituent to the previous ligands, which proved a nice and easy way to improve both quadruplex affinity and specificity.

b. *Structure*



BRACO 19

In addition to the acridine core structure, BRACO 19 features 2 lateral tertiary amines optimized in size that are protonated at physiological pH (7,4) and an additional 9 amino derivative to “enhance” binding affinity and G4 specificity over duplex DNA. To be exact BRACO 19 is the tri-protonated (hydrochloride salts) version of the molecule drawn above. Protonation normally occurs on both lateral amines and the central nitrogen atom (10th position). A paradox in the development of BRACO type molecules was to take advantage of π-π stacking of the acridine on the G4 and to improve it with lateral chains; however it turned out that binding is of “three dimensional complexity” and also linked to hydrophobic interactions.
(Fig. 20,21)

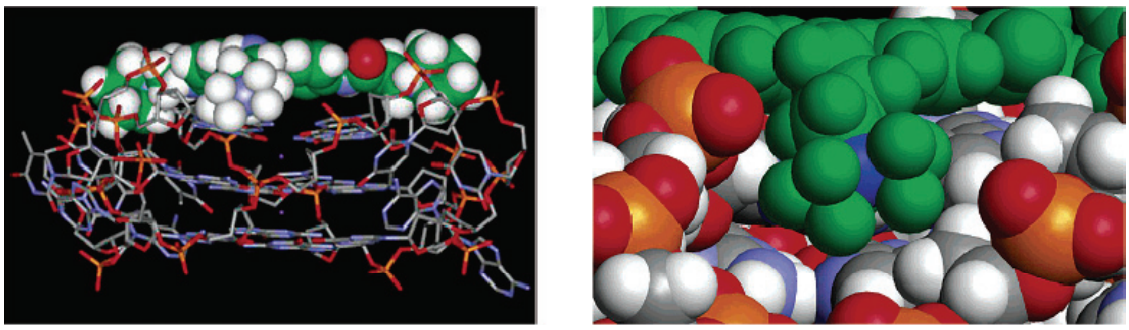


Fig. 20 [128]Left: Computer modelization of BRACO 19 interacting with quadruplex 22AG d[$AG_3(T_2AG_3)_3$]. Right: Localization of positively charged 9-substituent into the quadruplex groove.

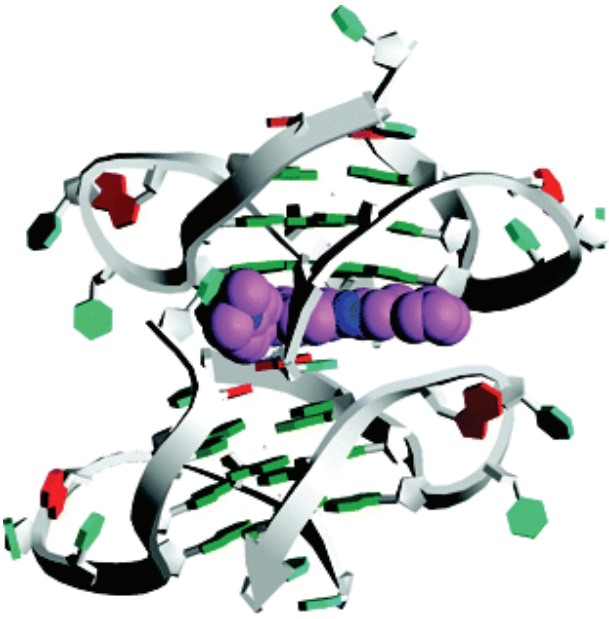


Fig. 21 Crystal structure of BRACO 19 at the interface of 2 quadruplexes [129]

c. *The chemistry of BRACO 19*

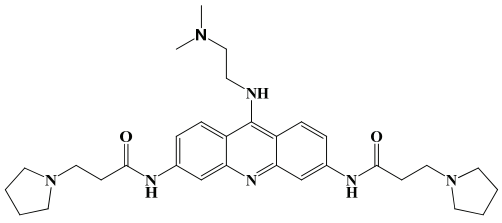
BRACO 19 is a trisubstituted ligand based on an acridine skeleton that is easily accessible from a cheap starting material such as diphenylmethane (2.5 kg=75 €). However major drawbacks are the necessity to use metals (such as Chromium (VI) or Tin (II)), HCl, 3-chloropropionylchloride, pyrollidine and phosphoryl chloride which all represent health hazards. *(The author is fully aware that it is somehow ironic that it is necessary to use dangerous and carcinogenic chemicals listed above in order to synthesize an anti-cancer agent).* However, one benefit of the synthetic route is the fact that it is possible to purify most reaction intermediates and products via precipitation/recrystallization. The synthetic route to acridones and acridines will be developed further in Chapters 2/4 and their reactivity will be the main subject of Chapter 4.

d. *BRACO and discrimination G4 – duplex:*

As evaluated by SPR and TRAP assay, 3,6-disubstituted acridines (e.g. BSU 6039) all proved tight quadruplex binders and potent telomerase inhibitors. However, addition of a 9-substituent resulted in an increased quadruplex affinity and increased telomerase inhibition potency.[128] Of particular interest is the increase in quadruplex vs duplex specificity upon addition of a 9-substituent.

Acridine ligand	$^{tel}EC_{50}$ μM	r(Affinity)
BSU 6039 (see above ligand **4**)	5.2	1
BRACO 19	0.12	40
NH-CH_2-CH_2-NMe_2 (structure below)	0.018	11

r(Affinity) is defined here as the ratio between the affinity constants of each ligand for quadruplex and duplex as evaluated by SPR.
$^{tel}EC_{50}$ μM is defined as the concentration of each ligand required to inhibit 50% of the activity of the enzyme telomerase in a TRAP assay.



NH-CH_2-CH_2-NMe_2 derivative of BRACO 19

From the table above the increase in discrimination factor is obvious: a 40-fold increase in quadruplex specificity due to the 9-substituent arises. Although acridines with different 9-substituents (e.g. *N*,*N*-dimethylethylene diamine) are better telomerase inhibitors, BRACO 19 remains a reference compound, highly specific for the human telomeric quadruplex.
Molecular modeling studies based on the X-ray crystal structure of the 22-mer human telomeric quadruplex have been used to interpret these results and see how BRACO 19 could specifically bind to intramolecular DNA quadruplexes. The external 3' G-quartet was found to be the preferential binding site for the acridine platform (via π-π stacking interactions) then projecting the three side-chains into three different grooves of the quadruplex. It resulted from this model that lateral chains in positions 3 and 6 (i.e. pyrrolidine propionamide) are located inside cavities formed by the TTA loops and that the 9 substituent dock into a cavity where positively charged substituents are favored.

This was completed by the crystalline structure between the BRACO 19 and two 22-mer human telomeric quadruplexes proving that the size and nature of the 3,6 side chains fit into the quadruplex grooves through hydrogen bonding, and explains the role of the 9 substituent as it fits in a narrow planar region of a small pocket big enough for small linear chains and aromatic rings.

e. BRACO 19 in vivo

In 2001 BRACO 19 was studied *in vivo* [130]. It did not cause any nonspecific acute cytotoxicity at similar concentrations to those required to completely inhibit telomerase activity.
Exposure of 21NT human breast cancer cells with 2 μM BRACO 19 showed a valuable reduction of growth after 2 weeks. Intraperitoneal administration of non-toxic doses of BRACO 19 (2mg/kg) to mice bearing advanced stage A431 human vulval carcinoma subcutaneous xenografts also showed similar positive results towards tumor suppression.
Therefore BRACO 19 is one of the most potent cell-free inhibitors of human telomerase (TRAP assay: 50% inhibition concentration at 115 nM, complete inhibition at 500 nM). However, current drawback of BRACO 19 is its relative instability in vivo. Indeed it was shown to partially decompose *in vivo* with the loss by chemical or enzymatic hydrolysis of the 3,6 propionamide side chains, likely due to the action of amidases in vivo. Different aspects of the instability of 9-aminoacridines will be the main subject of Chapter 4.

f. Extension of BRACO 19

Finally a series of "BRACO-analogues" (Fig. 22) with pendant peptides was also prepared (in which the peptidic sequences were selected via an independent combinatorial approach for their capacity to bind preferentially to G4) [131]. This study demonstrates that the nature and orientation of the peptide at position 9 has a dramatic effect on the quadruplex affinity and also on the quadruplex specificity. For instance, selected 9-peptido acridines proved capable to discriminate between one quadruplex and another.[132]

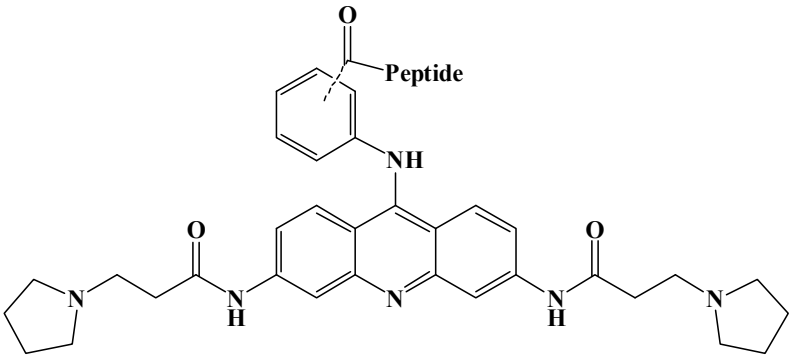


Fig. 22 General Structure of BRACO peptides analogues.

Chapter 2: A Hybrid (3+1) DNA:PNA Quadruplex

"There are no such things as applied sciences, only applications of science."

How do engineered PNA-acridone conjugates interact with human telomeric DNA?

1. Introduction:

a. *History:*

As previously discussed in the main introduction G-quadruplexes are highly polymorphic. This polymorphism can be either *stoechiometric* or *structural*: due to possible variations in the number of DNA strands involved and possible variations in strand orientations. In the case of *stoechiometry*, we have seen in the general introduction that depending on the number of guanine-rich stretches, quadruplexes can be monomolecular[133], bimolecular[134] or tetramolecular[135]. Concerning the *structural* polymorphism, the case of human telomeric quadruplex is the most representative (and most studied example): depending on the conditions it is folded in, a four repeat fragment of human telomeric DNA is capable to form a parallel type[32], a basket type, or the so-called hybrid 1 and hybrid 2 quadruplexes. (Fig. 1 below). The hybrid 1 is obtained from the hybrid 2 by a 180° rotation along the cation axis and inversing the polarity of the guanosines (syn <-> anti).

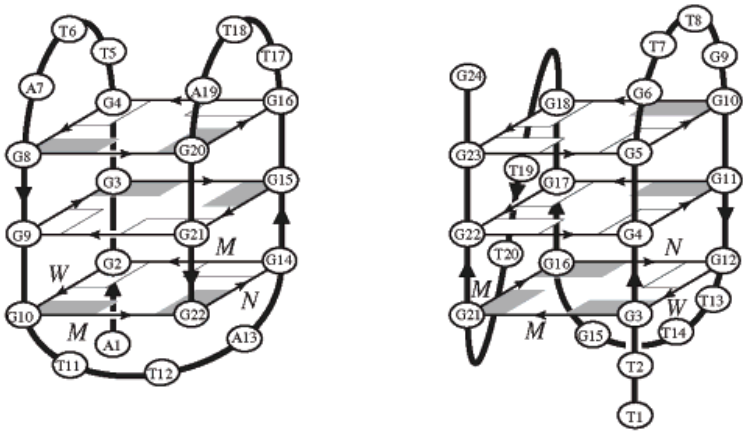


Fig. 1 Examples of intramolecular quadruplexes based on the human telomeric sequence (left: basket type [36], right: hybrid 2 [136])

Telomeric sequences do not only exhibit structural polymorphism but also stoechiometric variations. We have discussed quadruplex architectures composed of 1, 2 or 4 DNA strands: (4), (2+2), (1+1+1+1). However, till recently one

combination was missing, the (3+1) stoechiometry which was solved in 2005 by NMR studies by Zhang and co-workers[137] (Fig. 2). As most of biologically relevant quadruplexes have been found to be intramolecular, this new (3+1) bimolecular telomeric quadruplex is original and may be of particular interest.

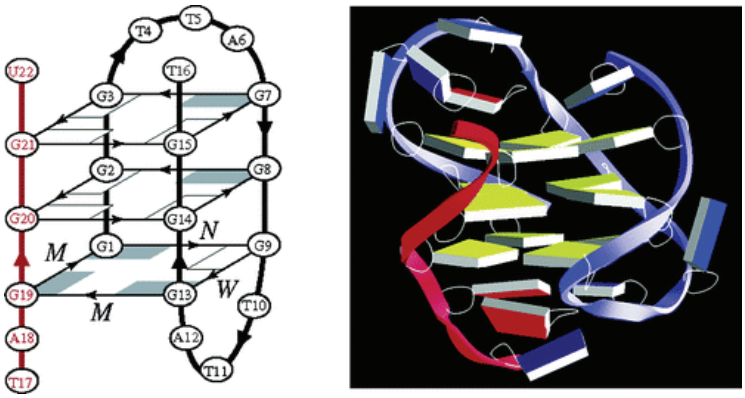


Fig. 2 Two different human telomeric DNA sequences fold into a (3+1) structure. (d(GIGTTAGGGTTAGGGT) + d(TAGGGU)). (Left: syn G are grey, anti G are white)

(NB: This (3+1) notation became confusing through the literature. Here we refer (3+1) as a bimolecular quadruplex and **not** as an intramolecular folding where the 3 strands have the same polarity and the last is in the opposite orientation.)

On Fig. 2 the left picture shows that each tetrad adopts the conformation $(syn)_3$-anti or $(anti)_3$-syn respectively to each guanines. Tetrads are built with 3 guanines from one strand and the 4^{th} from another strand.

The heterodimeric association of two strands of such different lengths generates an asymmetric structure with two lateral loops. This highlights once more the polymorphism [32, 138] of telomeric sequences. This result was linked to the widely studied human telomeric quadruplex sequence [12, 72] and involved two strands of DNA (three runs of 3 consecutive guanines from the first and one run only from the second) in sodium solution.

Another example was published in 2007 by Pedroso and co-workers[139] that involved the formation of such (3+1) assembly. Using the telomeric motif $d(TTAG\mathbf{G}G)_n$ they interchanged the folding stoechiometry (intramolecular <-> intermolecular) by changing a single nucleotide (ex: $d(TTAG\underline{\mathbf{A}}G(TTAGGG)_{n-1}$), resulting in telomeric DNA that would adopt different intermolecular structures (Fig. 3)

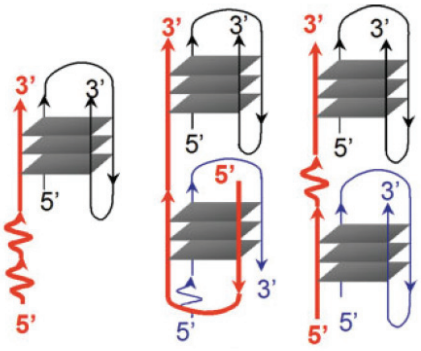


Fig. 3 Intermolecular G-quadruplexes obtained from the assembly of multiple strands of telomeric DNA

b. Concept

While most strategies developed so far for targeting the human telomeric quadruplex consisted in targeting one of the multiple forms of intramolecular quadruplex adopted by a four repeat fragment of human telomeric DNA, we reasoned herein that mimicking this newly reported (3+1) assembly could represent an original way to specifically target DNA sequences that contain 3 stretches of consecutive guanines only (instead of four in the case of intramolecular quadruplexe).
Such strategy would involve the use of ligands that contain 3 consecutive guanines residues in order to mimic the (1) strand of the (3+1) assembly and could be applied to either:

- form bimolecular quadruplexes with DNA sequences that already contain 4 stretches of 3 consecutive guanines but would then have to compete with the formation of intramolecular quadruplexes (as in the case of telomeric DNA).

- induce G4 formation by targeting DNA sequences that contain 3 runs of 3 consecutive guanines only (i.e. induce the formation of G4 where they normally wouldn't form). (Fig. 4)

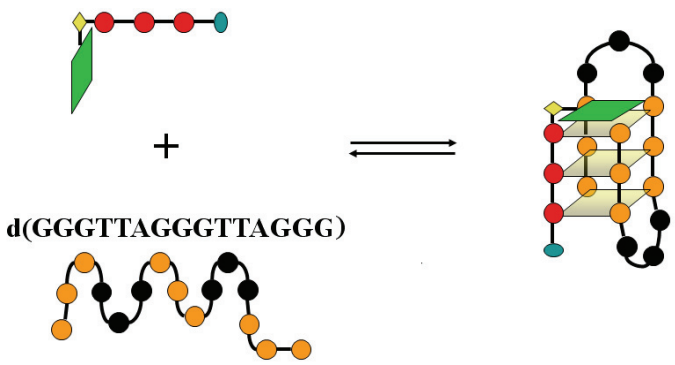


Fig. 4 Formation of a bimolecular G4 from a DNA sequence containing 3 tracts of consecutive guanines and a ligand containing one tract of three consecutive guanines.

What arises from our concept is the necessity to design a synthetic ligand that would contain 3 consecutive guanines capable to interact with the DNA strand and that would all be involved in individual G-tetrad formation. In order to further stabilize the hybrid quadruplex formed we also reasoned that associating this G-strand with a quadruplex binding platform would allow to further stabilize the hybrid quadruplex once formed. Hence our water soluble small molecules designed to interact with telomeric DNA is an example of supramolecular ligand.

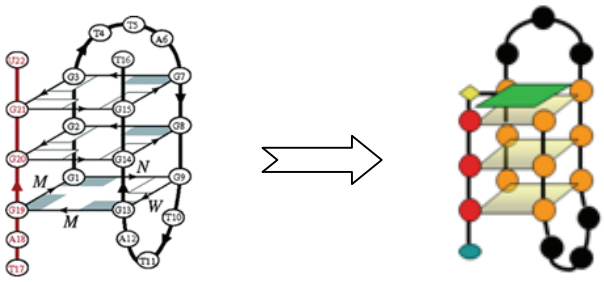


Fig. 5 From left to right: Zhang *et al.* structure compared to our strategy

c. *Prevalence of 3+1:*

From the prevalence of Possible intramolecular Quadruplexes Sequence (PQS) in gene promoters (i.e. 43% of genes contain in their promoter region four consecutive runs of guanines: $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$) [27], we can extrapolate that the number of sequences that contain at least three consecutive runs (i.e. $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$) of guanines is significantly higher than 43% [29]. Therefore a possible strategy to target those sequences would be to introduce the missing "run" in order to form a bimolecular 3+1. We realize that a strategy to target only 3 runs and *not* a G-quadruplex form would be a more general and wider approach than the classic ligand design.

d. *PNA and DNA :*

As it has been described in the general introduction PNAs are synthetic mimics of DNA/RNA and are able to hybridize with high affinity[140] with DNA in a sequence specific manner following the Watson Crick base pairing rules.

Likewise PNAs can also form quadruplexes consisting solely of PNAs or be involved in their structures by association with DNA to form hybrid PNA_n-DNA_x structures [141-144]:

-An example from Armitage's Group in 2003 shows that PNA associates nicely with DNA in a (2PNA + 2DNA) tetramolecular architecture. The figure 6 represents this assembly where the PNA strands are in purple and the DNA strands in blue. The telomeric repeat region of the protozoan Oxytricha nova with the sequence d($G_4T_4G_4$) can fold into a dimeric hairpin quadruplex[134] but in the presence of $G_4T_4G_4$-Lys PNA a hybrid structure appeared in sodium solution. The binding stoichiometry was first supposed to be 1:1 but the CD study suggested the formation of a tetramolecular structure PNA_2-DNA_2 which was further supported by FRET experiments.
Interestingly, whereas the DNA quadruplex stability strongly depends on cations nature and concentration, the hybrid PNA:DNA quadruplex shows less interdependence. It's noteworthy that the resulting hybrid structure plays in favor of an "invasion" of the quadruplex DNA structure by the PNA. [145]

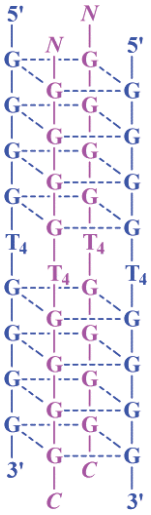


Fig. 6 A PNA_2-DNA_2 assembly

- in 2007 Armitage's group also described a 2 :1 PNA :DNA quadruplex hybrid[144] which is a good example of quadruplex invasion and hybridization monitored by ligand labeled PNA. (Fig.5)

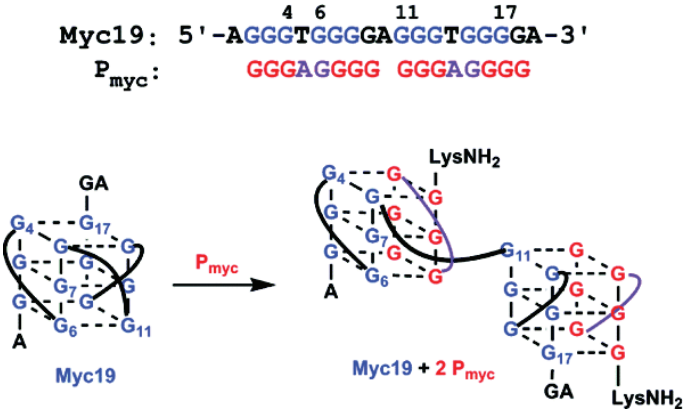


Fig. 7 Structure of the PNA_2:DNA hybrid
(Note that the PNA is either P_{myc} or P_{myc}-Thiazole Orange)

-The Balasubramanian's group succeeded in forming a unique tetramolecular quadruplex with PNAs *without* a central cation[141]. The complex was built by melting and subsequent slow cooling of a Lys-T-G-G-G PNA (Fig. 8). G4 formation was investigated by UV, CD and NMR spectroscopy and it was further confirmed using nano-ESI-MS that respects the fragility of the supramolecular complex. A solution-phase H/D exchange experiment was also carried out to ascertain that the $(PNA)_4$ complex was not the result of a nonspecific interaction due to the electrospray ionization. The conditions of the "hybridization" are described in detail (especially the effect of salt concentration). Interestingly the PNA_4 quadruplex

proved to be equally stable in salt free and K^+ containing solutions (small ДТ=+4°C). This trend has to be kept in mind while analyzing possible hybrids PNA – DNA. (i.e. slow dependence on cations for a PNA-DNA G4)

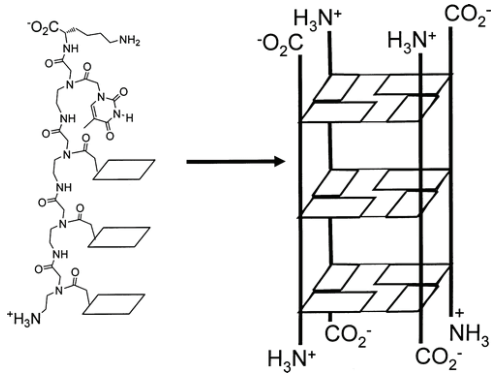


Fig. 8 A unique tetrameric PNA quadruplex

-In addition to this, "Chimeras" – where the PNA backbone alternates with DNA backbone are also capable to fold into a quadruplex.[146] (Fig. 9)

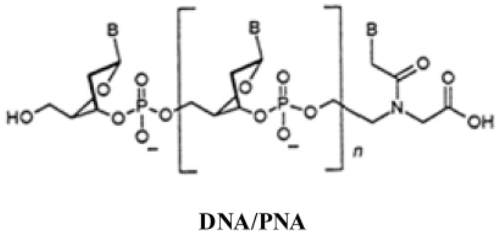


Fig. 9 Design of a DNA/PNA chimera

Petraccone *et al.* determined the thermodynamic constants for the association of chimeric strands into G-quadruplexes by CD and differential scanning calorimetry (DSC). Thermal denaturation indicated higher stabilities for chimeric quadruplexes in comparison with the corresponding unmodified DNA quadruplexes. The authors concluded that the ability of PNA-DNA chimeras to form stable quadruplex structures expands their potential utility as therapeutic agents.

Based on these previous examples, G-rich PNAs (e.g. G-G-G) appeared as a good alternative to DNA for the design of ligands capable of interacting with DNA to form bimolecular complexes mimicking the (3+1) assembly evidenced by Zhang *et al.* Not only the complexes they form with DNA are usually more stable than pure DNA complexes but PNAs are also chemically more stable than DNA.

Indeed all these results suggest that in order to complete the concept of a (3+1) human telomeric DNA + PNA system, a single G3 PNA could be our "missing" run of guanines that would hybridize with DNA to form a hybrid G-quadruplex structure and open a new approach to the design of therapeutic agents.

e. Choosing the ligand:

We have just highlighted examples of the possible quadruplex assemblies in the presence of PNA. However our strategy doesn't concern only the PNA-DNA assembly but also a new design for G-quadruplex ligands. Indeed PNAs are synthesized through solid phase synthesis, a chemical approach that allows the introduction of "non-peptidic" small molecules. Many ligands used for quadruplex stabilization have been shown to display external stacking [147, 148] (Fig. 10) properties for quadruplexes which is supposed to be the main interaction mode. Attaching a ligand at the extremity of our PNA could also prevent the formation of PNA tetramers.
From that we reasoned that introducing a "quadruplex binding ligand" at the end of our PNA could allow for an increased affinity and selectivity for the DNA target. Amongst all the quadruplex binding platforms available in the literature, we decided to take advantage of the properties of acridines/acridones since they are well-known G4 binder (known to bind to the external G-tetrads of the quadruplex as illustrated in Fig. 10) and their fluorescent properties could also allow an easy monitoring of the acridone/G4 interaction by fluorescence spectroscopy.

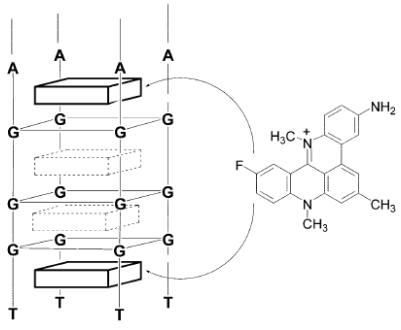


Fig. 10 NMR proven concept of end stacking for RHPS4[148]

(For an overview about G4 ligands please refer to the main introduction in which 24 architectures are examined in further details, see Fig. 20 of the introduction for an illustration of BRACO 19 binding mode)

PNA molecules are commonly synthesized using solid-phase chemistry. A convenient way to introduce a quadruplex binding platform at the end of the PNA would be by amide coupling onto the solid support. In order to do that, we would need to synthesize a molecule (e.g. an acridone as discussed above) with a free carboxylic acid that could be later activated for solid-phase synthesis. The acridone would also possess some "extra" substituents inspired from those of the BRACO 19[128] skeleton, i.e. the 3-(pyrrolidin-1-yl) propionamide that is known to interact with quadruplex loops[129] (Fig. 11).

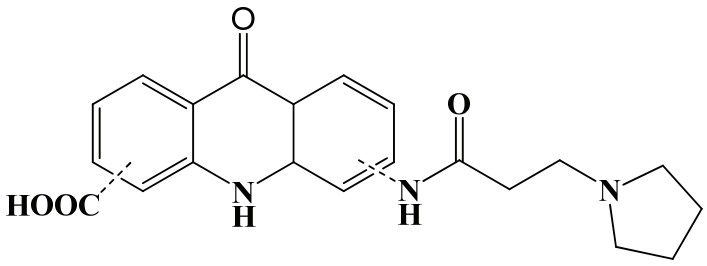


Fig. 11 Possible design of acridone to be incorporated at the end of the G3 PNA backbone

Our strategy involving the use of the solid phase synthesis a general introduction to this particular chemistry is given below.

f. Technical aspects of PNA synthesis:

Solid phase synthesis (SPS) is the convenient way to generate peptides and in particular PNAs. It is a well established multi-step process where the starting point is a small bead of resin.

The size of the spherical bead and the numbers of reactive functions (loading) matters as accessibility is the key to synthetic success. Usually the bead is made

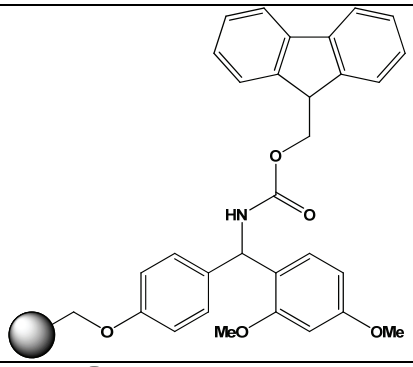
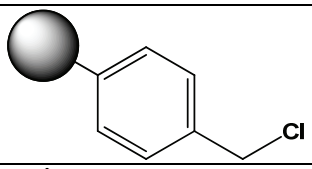
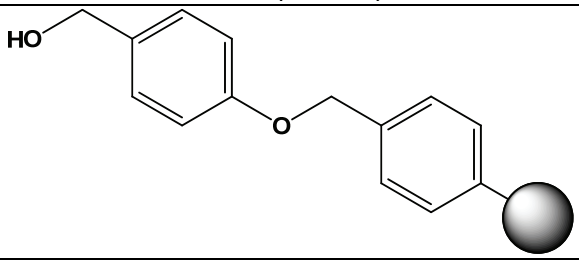
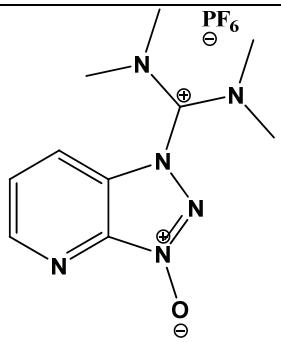
from an inert polymeric material and grafted with the desired reactive functionality, then depending on the bead density in water the choice of the solvent in the synthesis is important for the system not to float and therefore be less reactive.
SPS usually takes place in a specific vessel that can be shaken and where solvents can be sucked away (usually at the bottom) without interfering with the reaction mixture.
Classic solvents for SPS are Dimethylformamide (DMF) or N-methyl 2 pyrrolidone (NMP). High boiling solvents are used because this technique does not require any evaporation. (-> Dimethylsulfoxyde DMSO can also be used for difficult cases)
The desired building blocks typically consist of single molecules bearing a reactive head (or ready to be activated) and a protected tail, where the activated head reacts with the unprotected tail.
Hence the term solid refers to the bead where a head to tail chain construction will take place while every building block will be solubilized using aforementioned solvents.

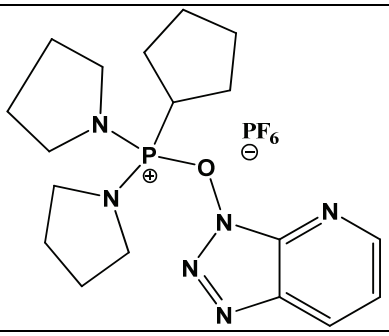
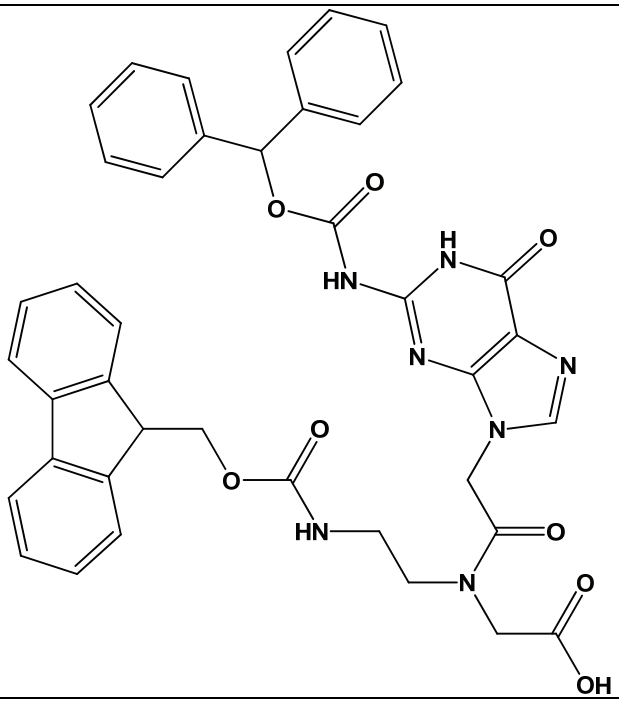
Herein, we would like to highlight important technical considerations:

The first step is cleaning the resin with dichloromethane (DCM) which is then impregnated with the desired solvent so that the reactive functions are wetted, increasing their reactivity. Any protecting groups on the resin are removed (depending on the resin, some come unprotected) so the first block may be added.
The first block is separately solvated and activated, and depending how reactive the tail is a wide choice of activating chemistry may exist. It is mandatory for the activated monomer to be soluble at this point. The next step is the addition of the solvent containing the activated monomer over the accessible beads and stirring/shaking the mixture.
This method requires attention for each step: deprotection steps are often the key steps as it is necessary to have the greatest number of reactive functions. Deprotection releases protecting groups that need to be carefully washed away in order not to spoil the next step. Residual excess of monomers must also be removed in order to avoid undesired double additions.
However a double addition of one specific monomer (without deprotection in between) may be used sometimes to attach difficult chemical components in order to increase the yield of the general procedure.

Once all the addition steps have been performed the desired oligomer is released by cleavage from the resin beads it was attached to. This step usually depends on the nature of the resin.

The table below presents the structures of chemicals (i.e. solid supports, coupling agents, reaction monomers) commonly used for the SPS of PNAs.

Type	Chemical Structure	Supplier
Resins		
Rink amide		EMD chemicals
Merrifield		EMD chemicals
Wang		EMD chemicals
Activating agents		
HATU		EMD chemicals

PyBOP		EMD chemicals
PNA Monomer		
Fmoc-G(Bhoc)-Aeg-OH		ASM research Chemicals

A final purification step is commonly required (usually by HPLC) in order to obtain the oligomer product pure (i.e. to separate it from all the different incomplete/truncated sequences that have been synthesized.
Of course SPS is not always a 99.99 % yield synthesis, but the method definitely helps reaching the desire sequence in high yields and reasonable quantities.

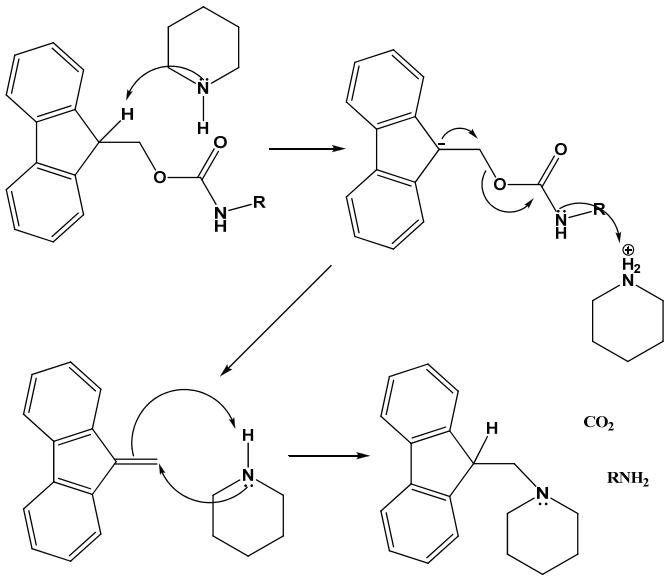
t-Boc solid-phase peptide synthesis:

To remove Boc from a growing peptide chain, acidic conditions are used (usually

neat TFA). Removal of side-chain protecting groups and cleavage of the peptide from the resin is done using anhydrous hydrogen fluoride (dangerous). Although generally safe, and using only small quantities, HF cleavage needs to be done using specialized equipment, so it is generally disfavored.

Fmoc solid-phase peptide synthesis:

To remove an Fmoc protecting group from a growing peptide chain, basic conditions (usually 20% piperidine in DMF) are used. Removal of side-chain protecting groups and peptide from the resin is achieved with a mixture of trifluoroacetic acid (TFA) 99%, triisopropylsilane 0.5% and water 0.5%. Fmoc deprotection (see below) is usually slow because the anionic nitrogen produced at the end is not a particularly favorable product, although the whole process is thermodynamically driven by the evolution of carbon dioxide.



Fmoc cleavage under basic conditions

New trends appeared in the SPS of PNAs: first the possibility of a fully automatic process and more recently the use of a microwave oven. The use of microwave radiation can help (i) enhancing the yield of each step by preventing aggregation of growing chains and helping the coupling reaction to be quantitative, and (ii) reducing significantly the coupling reaction times. However such ovens are still expensive and not widespread.

2. The chemistry:

To successfully obtain a molecule that would fulfill our requirements we used SPS followed by HPLC purification for the sequence Lys-G-G-G-Gly-Acridone. The precise explanation of this chosen sequence is given below:

-we needed a run of consecutive guanines able to participate in the G4 formation.
-we chose to attach a G-tetrad binder to further stabilize the hybrid G4 formation.
-we had to give some mobility to the end stacker hence there was a necessity for a linker/spacer between PNA and acridone.
-most importantly: our PNA had to be water soluble: therefore an extra lysine was mandatory.

All those features can be found on Fig. 12

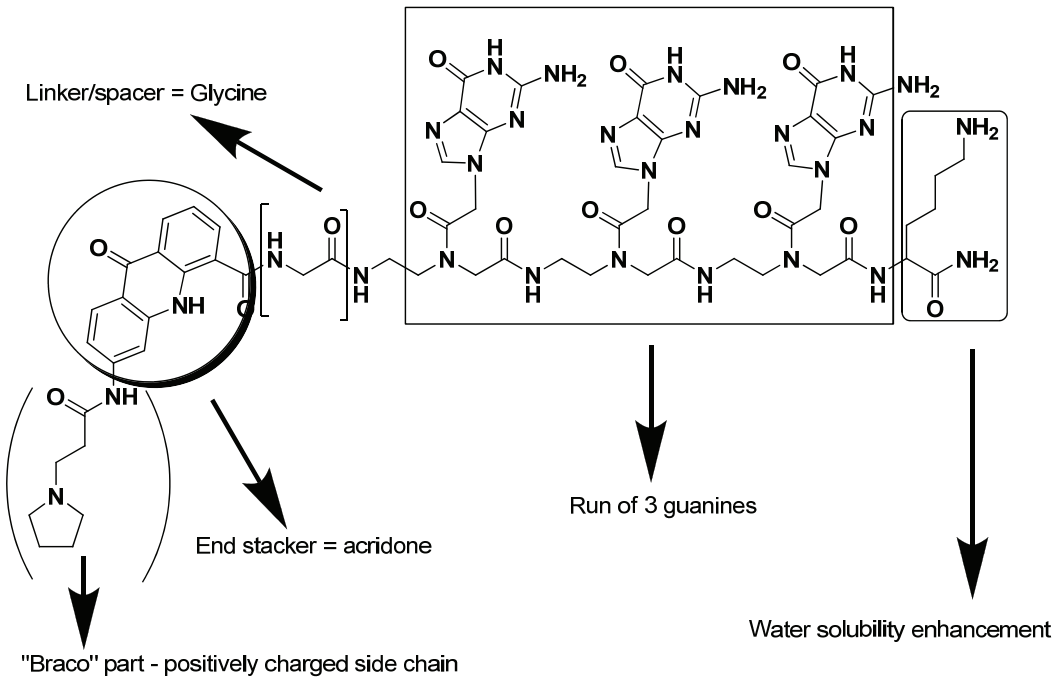
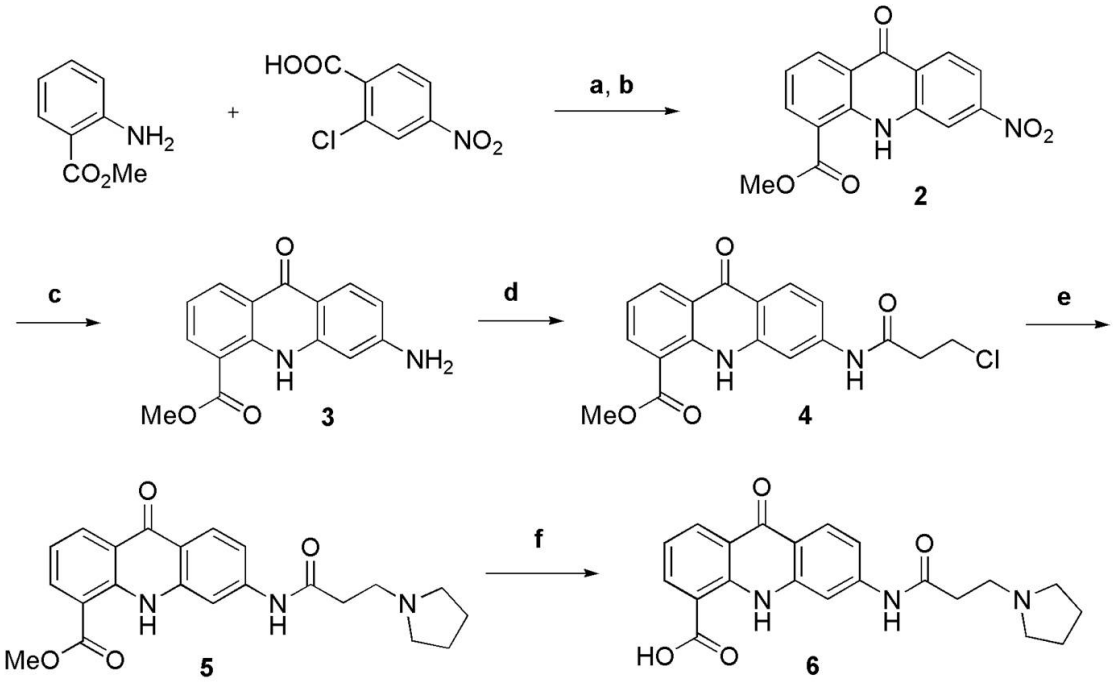


Fig. 12 General Structure of the PNA

We clearly see that our strategy requires commercial monomers of guanines, lysine, glycine and an engineered acridone whose chemical synthetic pathway is described below (Fig. 13):



Reagents and conditions: a) Cu, Cu_2O, $Cu(OAc)_2$, DMF, 160°C, 1h (***50%***); b) PPA, 130°C, 45min (***85%***); c) $SnCl_2.2H_2O$, AcOH, HCl, 15min r.t. then 45min 80°C (***66%***); d) 3-CPC, 60°C, 24h (***85%***); e) pyrrolidine, DMF, 60°C, 30min (***92%***); f) NaOH, H_2O, DMF, 50°C, 40min (***89%***)

Fig. 13 Global synthetic pathway to acridone **6**

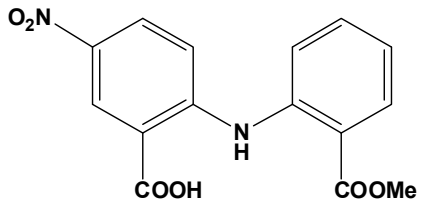
We will now describe in more details this synthetic pathway.

a. Presentation

The desired acridone motif was synthesized via a N,N diphenyl intermediate followed by a cyclization. Another route to synthesize a "BRACO-like" acridone backbone will be described in chapter 3.

Since our target molecule bears both a free carboxylic acid and an amido side-chain then the general synthesis process can be divided in two parts: first the synthesis of acridone amino acid and then its functionalization with a propionamide side-chain.

b. Synthesis of the N-phenylanthranilate



Nowadays there are two main ways to synthesis N-phenylanthranilates, using either the "old" Ullmann coupling reaction or the more recent Buchwald-Hartwig Palladium-mediated coupling.

i. Ullman coupling:

The Ullmann coupling or the Ullmann-Jourdan reaction is a copper mediated reaction between an α-halogenobenzoic acid and a nucleophile (often aromatic: aniline, phenol, thiophenol)
Fritz Ullmann is a famous german chemist that discovered at the start of 20th century one of the first coupling reactions[149, 150]. A coupling reaction is typically the creation of a sigma bond between two carbon atoms. This may also apply to heteroatomic bonds as is the case here.
The first step of the reaction is the copper addition and is the rate determining step. Copper (I) is present in small proportions (2 Cu(I) -> Cu(II) + Cu) in organic solvents and this explains why this reaction requires "harsh" conditions such as extreme heating. Copper (I) undergoes some oxidative addition where the halogen atom is stabilized by the chelating effect of the nearby carboxylate (the chelation process

decreases also the kinetics). Additional copper cations interact with the π orbitals of the benzenic rings modifying the reaction rate dependence in [Cu(I)].
The second step may happen in two different ways:
-the nucleophile (Ph-NH, Ph-OH) may first bind the copper as a new ligand which is followed by a reductive elimination giving the desired product.
-the nucleophile may directly attack via an aromatic nucleophilic substitution on the α position of the carboxylate and "release" the copper.
A possible combination of both pathways can be considered.
Finally the reaction may generate a few by products due to the unstable Cu(III) state that creates radicals and possible dimerization. (Fig. 15)

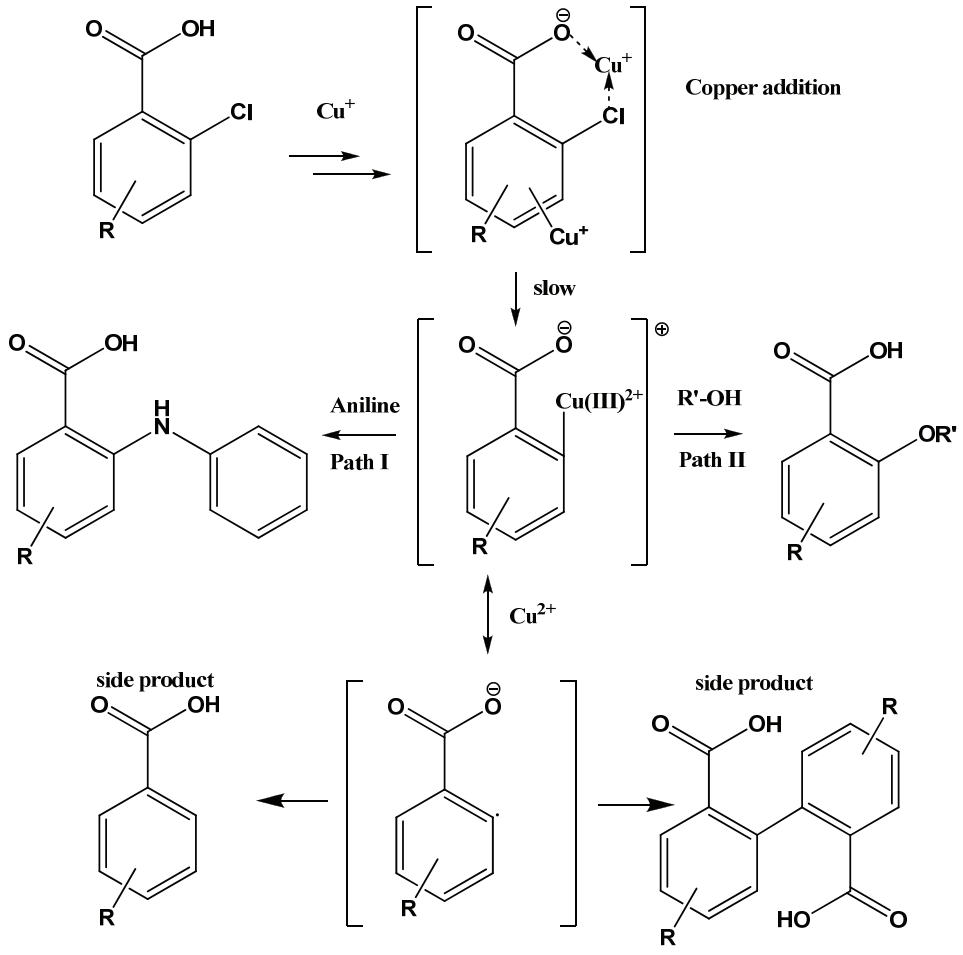


Fig. 15 Probable mechanism of Ullman Jourdan reaction.[151, 152]

This coupling reaction is also obviously influenced by the nature of the halogen atom. The use of α-bromo[153] or α-iodobenzoic[154] acids is generally preferred to the less reactive α-chlorobenzoic acids. Bigger halogens increase the kinetics and the yield of the reaction. Improvements have been made to this synthetic pathway such as the use of activated 2-(phenyliodonio)benzoate or later the use of microwaves chemistry.[155]

ii. Buchwald-Hartwig

We will discuss here about an alternative route to N-phenylanthranilic acid via another famous coupling reaction which emerged in the 90s. [156, 157]
The Buchwald-Hartwig amination in the case of methyl iodobenzoic acid is described on Fig. 16 below[151]:

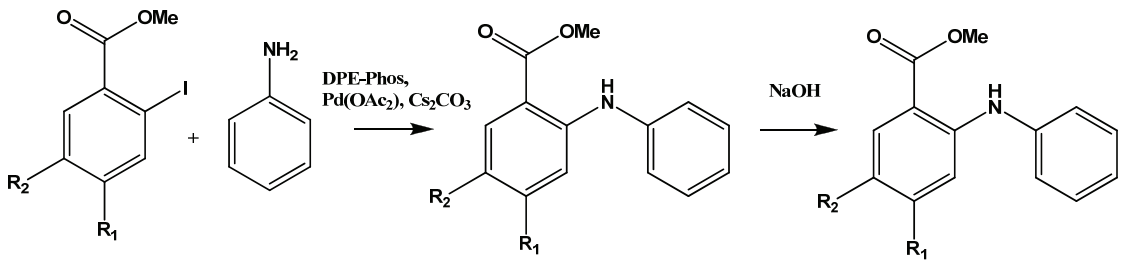


Fig. 16 The Buchwald Hartwig route

This reaction occurs in high yields (75% average) and uses bis(2-diphenylphosphinophenyl) ether and Palladium acetate under basic conditions. It also requires a subsequent hydrolysis step. Indeed the palladium coupling will be affected by the presence of a possible chelating carboxylate (ester necessity). Besides (i) this route requires long reaction times (24-36h) and (ii) high yields are obtained with 2-iodobenzoic acid only.
The shorter reaction times, the fact that the starting monomers are commercially available and much cheaper, the ease of workup (no column purification) led us to choose the Ullman coupling reaction to generate our N,N-diphenyl intermediate of interest.

iii. Consequence on the experimental procedure:

Early attempts to synthesize the desired N-phenyl anthranilate were performed using 4-nitro-2-chlorobenzoic acid[158] and methyl-anthranilate as described in the literature by Baguley and co-workers[159]. This reaction was carried out in N-methylpyrrolidone as solvent and in the presence of N,N-diisopropylethylamine and cupric acetate as a catalyst. Unfortunately in our hands repeated attempts led to partial hydrolysis of the methyl ester, thus resulting in extremely poor yields.
Ullmann type coupling can also be realized in anhydrous 2-ethoxyethanol[160] or DMF [161] in the presence of potassium carbonate. Once again poor yields were obtained due to ester hydrolysis. In summary ester hydrolysis during the Ullamn coupling was identified as the limiting factor to high yields.
Finally the desired methyl N-phenylanthranilate was obtained in good yield when refluxing both reagents in DMF in the presence of a mixture of Copper 0, I and II and in the absence of any base to prevent methyl ester hydrolysis.
Then during the workup careful addition of slightly acidic water allowed our desired product to precipitate selectively.

c. *Cyclization:*

To convert the *N,N*-diphenyl intermediate into the desired acridone **2**, three different routes were possible:

(i) cyclization in sulfuric acid at 110°C
(ii) cyclization in $POCl_3$ at 130°C followed by treatment with diluted HCl
(iii) cyclization in polyphosphoric acid (PPA)/ester (PPE) at 120°C

Here we followed the polyphosphoric acid (PPA) route as described by Slater and co-workers.[162] The cyclization into acridone is a nice reaction that converts N-phenylanthranilic acids into acridone whether the starting molecule is pure or not. Indeed the workup is a precipitation in water removing any traces of possible by-products. Our only concern here was the possibility to keep our methyl ester untouched. This explains the choice of the PPA which is the only reagent from the three above to maintain the methyl ester.

Such cyclization reactions involve two steps:

-formation of a carbonium atom: it is the kinetics determining step and substituents on the aromatic of the carbonium may influence the kinetics. (e.g. a para electrons withdrawing group such as -NO_2 would slow down the reaction by destabilizing the carbonium)
-attack of the carbonium atom on the second aromatic ring: the second step is extremely fast because activated carboniums have short lifetimes due to their high reactivity. This step is an electrophilic addition on an aromatic system taking place so fast that substituents on the attacked ring have almost no effect on the kinetics of the reaction. However, the position of these substituents is critical since different acridones may be obtained depending on the position on which the carbonium reacts as illustrated on Fig. 17:

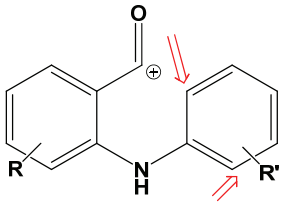


Fig. 17 Possible pathways of the "carbonium reactivity"

In general, and depending on the nature and position of the substituents born by each aromatic ring these acidic cyclization reactions can generate two regioisomers. In the case of our reaction, the possible acridone products are described below (Fig. 18)

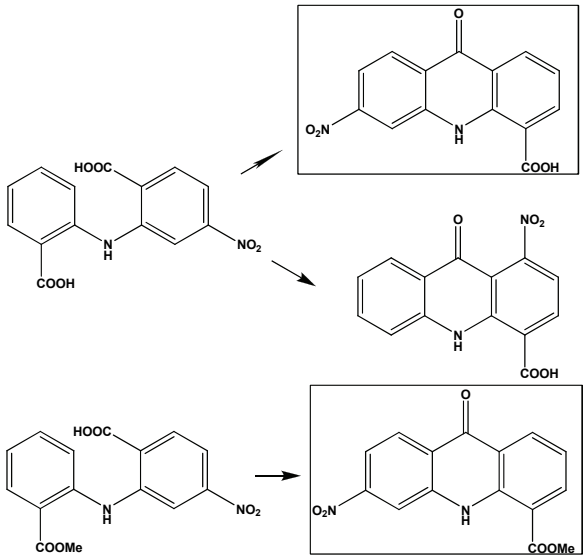


Fig. 18 Necessity to have and keep the methyl ester and possible stereochemistry of the coupling reaction

This figure demonstrates the necessity to keep our methyl ester all along the synthesis of the acridone since its hydrolysis into the corresponding carboxylic acid could lead to the formation of an undesired acridone.

d. Reducing the nitro group on the acridone:

Once more 2 strategies were available to reduce the nitro-acridone **2** into its corresponding amine **3**: tin chloride reduction or palladium-catalyzed hydrogenation.
Successive attempts to reduce the acridone by Pd-catalyzed hydrogenation proved slow and incomplete. Moreover, side reactions such as ester hydrolysis were also observed which led us to change our strategy.
A nice and convenient way to access to the desired amine product was published in 1938 by Matsumura[163] and involves the use of tin II chloride. The tin catalyst is extremely toxic but under acidic conditions and slight heating reduces quantitatively the nitro-acridone **2** without formation of any by-products.

e. Introducing the lateral chain:

3-chloropropionyl chloride (3-CPC) is a toxic lachrymator that reacts violently with water but is however a convenient precusor to add a 3 carbons side chain to our system. (for a reference see the BRACO synthesis in the DCC chapter). The reagent is a powerful chlorinating agent (weaker than thionyl chloride) and the reflux may generate some chloroacridine by-products (discussion about undesired chloroacridines will be developed in the DCC chapter). Therefore moderate heating (60-80°) is advised while using 3-CPC as solvent.
The reaction leading to **4** is a quantitative amidification where an amine readily reacts with an acyl chloride.

f. Addition of pyrrolidine:

The pyrrolidine or tetrahydropyrrole (pK_a = 11.27) readily and easily reacts with the chloropropionamide located at the extremity of the propanamide chain via a SN_2 nucleophilic substitution.
Addition of NaI may increase the rate of the reaction via halogen atom exchange but the reaction is usually fast without the use of sodium iodide.

In the article below we describe the isolation of the acridone **5** by recrystallisation in diethyl ether. It can be added that a nice and convenient alternative way is also to drop the reaction mixture in water and wash it with diethyl ether once dried which also affords the desired product pure.

g. Saponification of the methyl ester:

Ester hydrolysis was carried out at high temperature and in the presence of a strong base to obtain **6**. We would like to highlight that the difficulty here lies once again in the poor solubility of the system in aqueous media. The presence of DMF is a prerequisite for the reaction as well as strongly basic conditions and heating.

3. Publication:

Once synthesized, the ability of our PNA-acridone ligand to interact with telomeric DNA was investigated in vitro using a combination of biophysical techniques (UV, CD, and fluorescence spectroscopy as well as mass spectrometry). All the biophysical data are detailed in the article below entitled **"Combining G-quadruplex targeting motifs on a single peptide nucleic acid scaffold: a hybrid (3+1) PNA-DNA bimolecular quadruplex."** which was published in Chemistry – A European Journal as a full paper in 2008.

The main conclusions of this paper are listed here:

- Our PNA-acridone conjugate can interact with a three-repeat fragment of human telomeric DNA to form a hybrid PNA:DNA (3+1) G4 in vitro
- This quadruplex was characterized by UV, CD and mass spectrometry.
- Unlike pure DNA quadruplexes, this PNA:DNA hybrid quadruplex can form even in neat water without added salt.
- The acridone plays a significant role in the G4 assembly (stabilizing effect)
- No binding is observed to duplex DNA
- Our PNA-acridone conjugate can react with an unfolded single-stranded DNA but cannot invade a prefolded DNA quadruplex.

4. Supplementary results:

As described in the general introduction, Thiazole orange (TO) is a monomethine fluorogenic cyanine dye that is dark when free in solution but becomes highly fluorescent upon binding to duplex or quadruplex DNA. Based on our previous data obtained with our PNA-Acridone conjugate, we reasoned that replacing the acridone with a molecule of TO would provide an easy way to detect the formation of hybrid PNA:DNA quadruplexes by fluorescence spectroscopy. Indeed, one could expect that upon formation of the PNA:DNA quadruplex, TO would stack on top of the quadruplex, thus leading to the appearance of a characteristic fluorescence signal. A PNA-TO conjugate was thus synthesized by solid-phase and purified by HPLC. (Fig. 20)

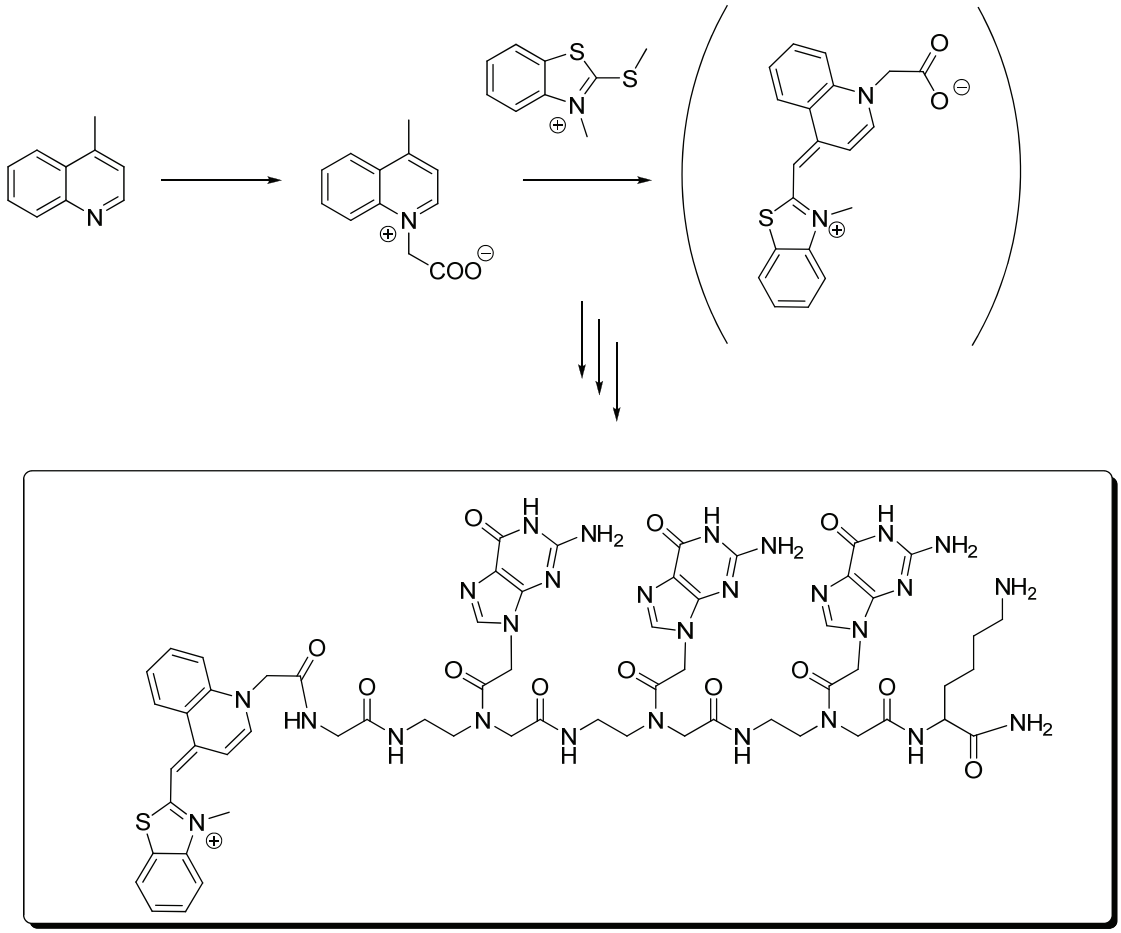


Fig. 20 Thiazole orange version of our "3+1" molecule

The binding of this new ligand with a three repeat fragment of telomeric DNA/RNA was then investigated by fluorescence spectroscopy. As anticipated, the free ligand proved non-fluorescent in solution. Upon binding to the DNA fragment d(GGGTTAGGGTTAGGG), quadruplex was formed which was accompanied with the appearance of a strong emission of fluorescence. As for the PNA-acridone conjugate, a hybrid quadruplex was formed even in salt-free water. However, this new ligand proved less specific than the acridone analogue since an increase in fluorescence was also obtained when incubating our PNA with a short duplex DNA, likely due to intercalation of TO between B-DNA base pairs (Fig. 21). Similar loss of specificity was reported when introducing TO motif on a G4 specific small molecule.[164]

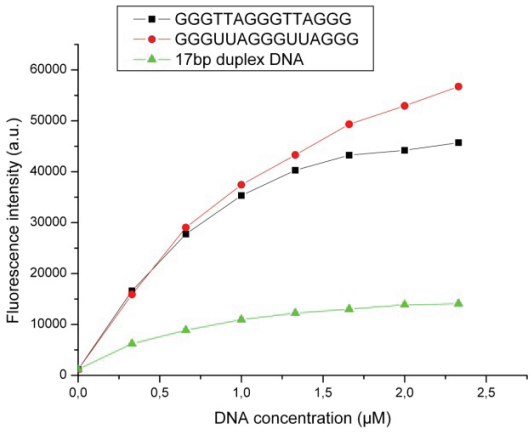


Fig. 21 : Fluorescence titration experiment

Interesting is the fact that our ligand also seems capable to form a (3+1) PNA:RNA G4 therefore requiring further to unambiguously characterize this structure.

In order to further stabilize the PNA:DNA hybrid quadruplexes formed, we reasoned that inserting a quadruplex binding platform at each end of the G3-PNA would lead to a "sandwich"-type PNA:DNA complex, with one acridone stacking with each external G-tetrad of the quadruplex. The Acridone-PNA-acridone (Fig. 22) conjugate below was thus synthesized and its capacity to form a hybrid quadruplex with telomeric DNA was further investigated. Unfortunately the spectroscopic data proved difficult to interpret and the formation of a hybrid 3+1 quadruplex could not be unambiguously characterized.

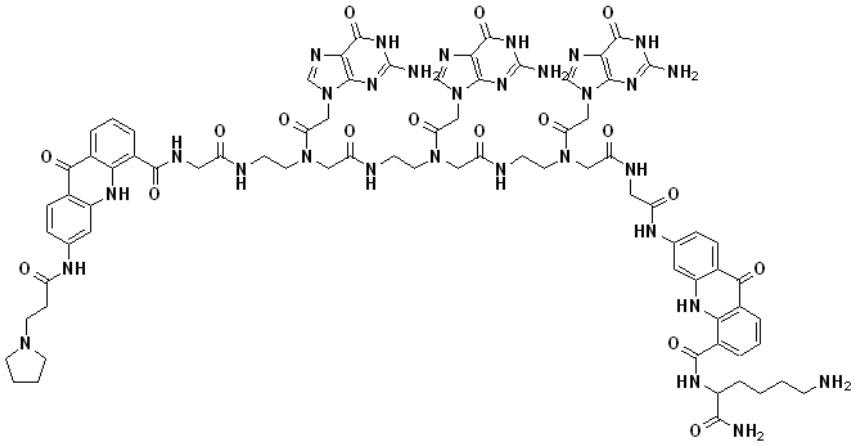


Fig. 22 Synthesized PNA with 2 acridones

In order to obtain a better “sandwich”, some re-design would have to be done to improve the nature and length of the linkers between the PNA scaffold and both quadruplex binding platforms. (Fig. 23)

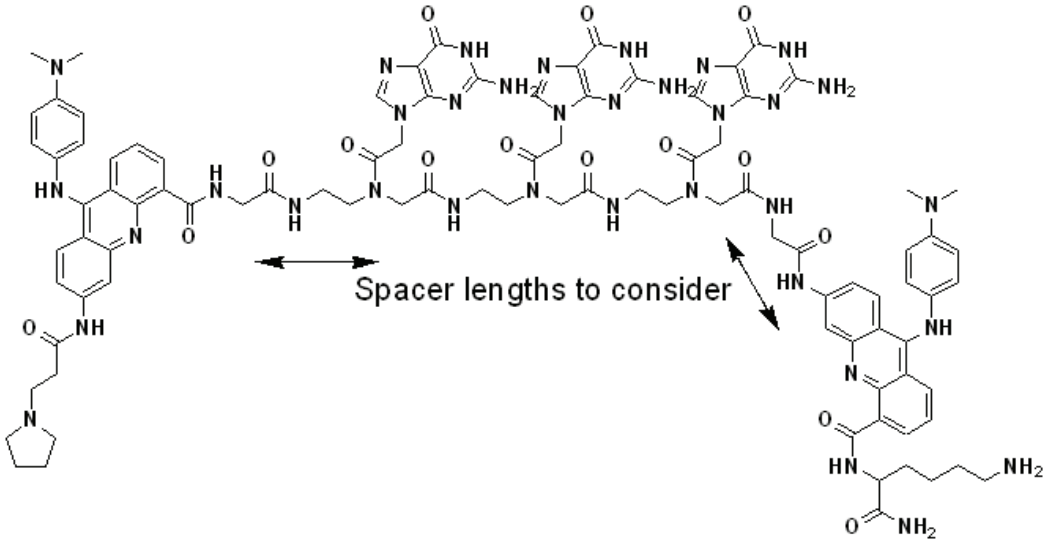


Fig. 23 Improved acridines sandwich and possible spacer extension

5. Openings-Future Work:

Herein, we only investigated the hybridization of a G3 PNA with DNA strand containing 3 runs of 3 consecutive guanines. However, the number of G-tetrad per quadruplex could potentially be decreased to 2. A ligand containing two guanines only (Fig. 24) could therefore be designed that target PQS of the type ($G_{2+}N_{1-7}G_{2+}N_{1-7}G_{2+}$). Indeed, G-quadruplexes are usually studied in the case of runs of 3 consecutive guanines leading to stable 3 tetrads structures. However G4 formed with 2 tetrads may exist with a weaker stability.

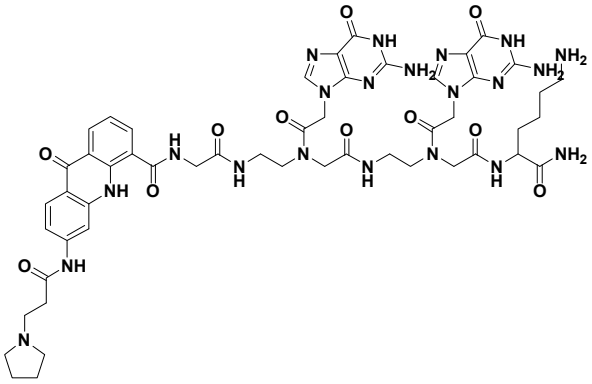


Fig. 24 “Shorter” version of our “3+1” PNA

A natural evolution of this system would simply be to improve the quality of the “terminal G-tetrad ligand”.
Fluoroquinolones (see introduction – ligands) (Fig. 25) are known for their good potential to target quadruplexes. Hence, ligands of the type below could show increased affinity for DNA and form more stable hybrid quadruplexes.

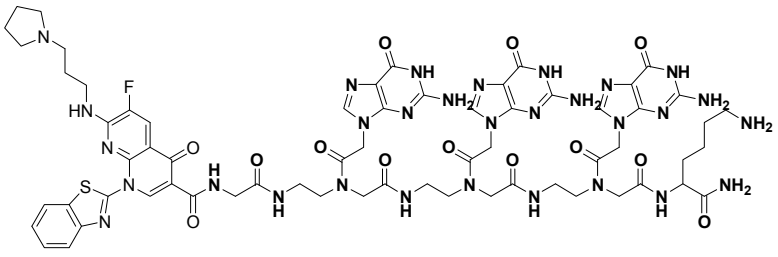


Fig. 25 A “3+1” molecule where a fluoroquinolone replaces the initial acridone

Chapter 3: Targeting RNA quadruplexes

"Dream as if you'll live forever, live as if you'll die today."

How to target a quadruplex without targeting the quadruplex?

1. Introduction:

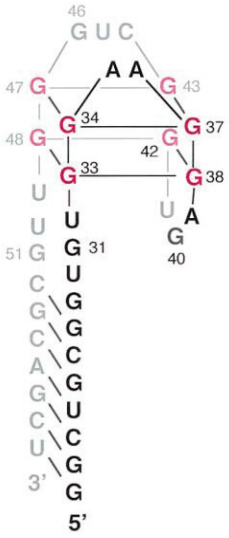
a. *History*

i. The existence and roles of RNA G4s

The first report of RNA G4s appeared in 1991[165]. Compared to DNA G4s, RNA G4s are much more stable[166-168]. Although they have not yet been studied in the same depth as their DNA analogues, their potential implication in natural mechanisms of gene regulation has drawn the attention on them [169, 170]. Here are the most recent examples of biologically relevant RNA quadruplexes. Their location in the genome and their possible biological functions are discussed in the section below:

-A mRNA G4 (Fig. 1 below) was shown to be the target of the fragile X syndrome mental retardation protein (FMRP).[90] The loss of function of a single gene encoding FMRP is responsible for the Fragile-X mental retardation, the most common cause of familial mental retardation. Mapping experiments revealed that the G-quartet RNA was bound by the FMRP C-terminal RGG box. In a model proposed by Darnell and co-workers, FMRP recognizes G-quartets in un-translated regions of mRNAs to inhibit their translation. This would suggest that RNA quadruplexes form *in vivo* and are recognized by naturally occurring proteins (see chapter 1 and [99, 171])

Fig.1. Proposed structure of the FMRP RNA quadruplex as proposed by Darnell *et al.*



-In 2003, Bonnal *et al.*[172] identified an RNA G4 that could drive Fibroblast Growth Factor 2 (FGF2) gene expression. They showed that in the 5'- alternatively translated region (5'-ATR) of the FGF2 mRNA, a specific 176 nucleotides long internal ribosome entry site of FGF2 contains two stem-loops and one G4 (Fig.2. above). Those three structures contribute to regulate translation efficiency by modulating ribosome recruitment.

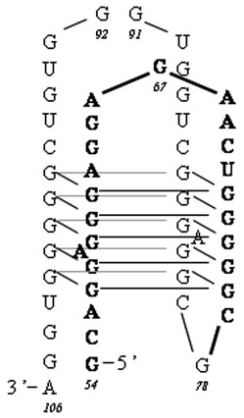


Fig. 2. A "five guanines per run" G4 proposed structure

- In 2006, Wieland *et al.*[173] engineered an artificial hammerhead ribozyme that contains an mRNA quadruplex forming sequence and that could be catalytically activated upon addition of the quadruplex binding ligand TMPyP4. The hammerhead ribozyme (HHR) is a well characterized, small RNA motif that exhibits phosphodiesterase activity and this activity can be inhibited by the small molecule TMPyP4. Formation of stem-loop II is necessary for correct folding of the catalytic core, and hence cleavage activity. A stabilization of stem II would result in the stabilization of the catalytic core, hence leading to enhanced cleavage activity. Exchanging stem-loop II for a G-rich loop destabilized the formation of the catalytically active structure. With increasing TMPyP4 concentrations, the G-rich loop folds into a G4 structure, thus supporting cleavage activity by formation of the catalytic core. A possible quadruplex formation in the HHR altered catalytic core would be responsible for the switching effect of the ligand (Fig. 3.).

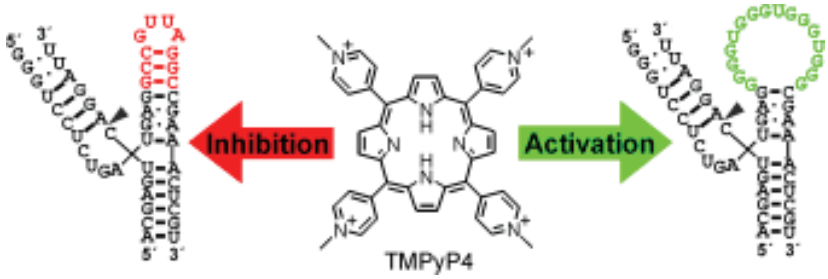


Fig. 3. Effect of TMPyP4 on different altered catalytic core of wt-HHR

- Wieland *et al.* [174] have also prepared a "switchable" system incorporating enhanced Green Fluorescent Protein (eGFP). They designed sequences containing G

rich elements surrounding the “Shine Dalgarno” sequence of a reporter gene coding for eGFP. The designed sequences were able to fold into G-quadruplexes (Fig. 4.). Using this system they demonstrated that quadruplexes could modulate gene expression *in vivo*(Fig. 4.), and that this expression is modulated by temperature as different levels of expression were observed when the temperature varied from 30° to 37°C (i.e. partial unfolding of G4). Although based on artificial (and therefore non biologically relevant) constructs, this proof-of-concept study emphasizes the biological role G-quadruplexes could play in gene regulatory mechanisms and also the possible existence and biological activity of stable two tetrads G-quadruplexes (e.g. G_2U below).

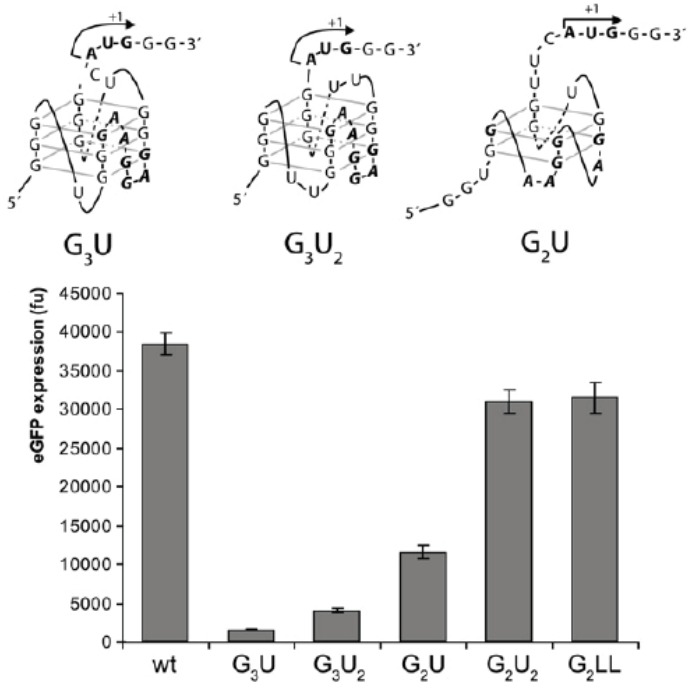


Fig. 4. Gene expression and related (parallel) quadruplex structures.

-The identification of a G4 forming sequence (5′-GGGAGGGGCGGGUCUGGG-3′), in the 5’UTR (un-translated region) of the *NRAS* proto-oncogene and the study of its possible implication in translation regulation mechanisms has been a major breakthrough for the understanding of the role of RNA G4 in the genome.
This 254-nucleotide-long UTR is located 14 nucleotides downstream of the 5′ cap and 222 nucleotides upstream of the translation start site. This sequence forms a highly stable parallel quadruplex in vitro, with a Tm value >> 95°C under near-physiological salt concentrations and can still adopt a quadruplex conformation even in the absence of salt. Kumari *et al.* [19] demonstrated the first example of translational repression by a 5’-UTR RNA G-quadruplex (Fig. 5.). Indeed, deletion of the quadruplex forming sequence or mutations that prevent it to fold into a quadruplex resulted in a 3-4 fold increase in translation efficiency in vitro.

They also pointed out that in the 5'UTR of the *Homo sapiens* genome the quadruplex density (PQS) is 0.3/kb higher than average genome (d=0,06/kb see chapter 1.)

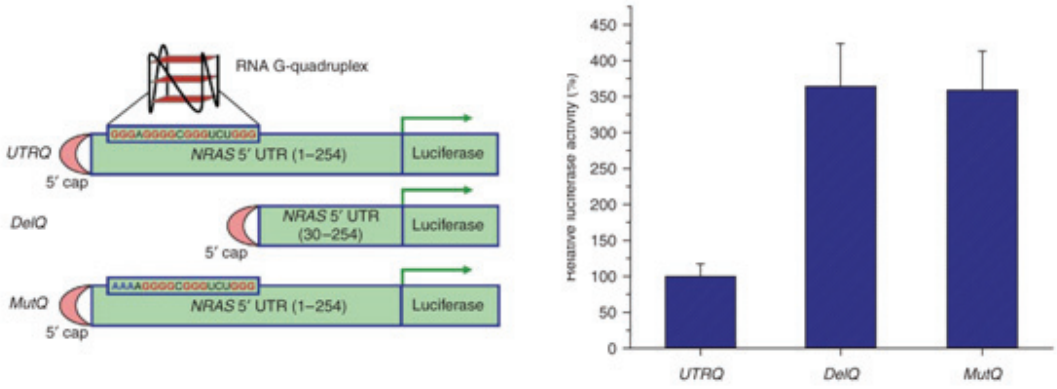


Fig. 5. Representation of mRNA constructs (normal, deleted, mutated) used for in vitro translation assays (left) and corresponding translation efficiency (right)

-More recently, Kumari *et al.* [175] also showed that both the localization of the G4 within the 5'-UTR and the stability of the quadruplex formed play a role in translation efficiency. The "best position" (i.e. that leading to the strongest inhibition) is proximal to the 5′-end, within the first 50 nucleotides. However RNA G4 has no significant effect on translation if located comparatively away from the 5′-end. (e.g. positions +120 or +233 on Fig. 6.) The thermodynamic stability of the G4 at its natural position (within the *NRAS* 5′ UTR) is an important factor contributing to its ability to repress translation: most stable RNA G4s lead to the strongest translation inhibitions.

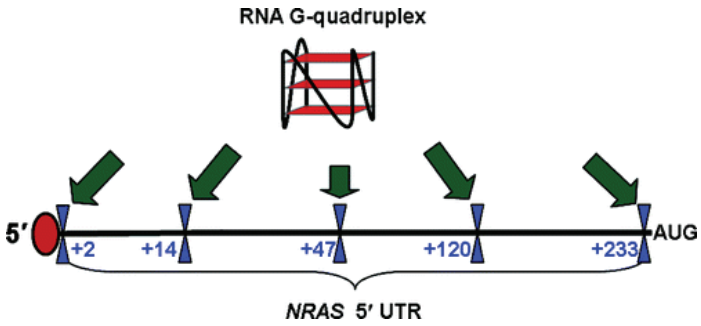


Fig. 6. Possible studied positions of the RNA quadruplex.

- Following the same path to the localization of RNA G4 Huppert *et al.* [20] discussed the incidence of G-quadruplex motifs in 5'-UTRs and in 3'-UTRs. To summarize what has already been mentioned before about translational role of G4 and their existence in UTRs (see Chapter 1: introduction). The hunt of G-quadruplex motifs in and around human genomic UTRs shows significant strand asymmetry and

positional bias, suggestive of functional roles in RNA. Below (Fig. 7) is a schematic representation of the different ways DNA or RNA quadruplex formation would interfere with translation:
Where the DNA sequence in the coding strand (blue) is G-rich (shown as GGG) a DNA G-quadruplex could form in that strand, and is here referred to as a G-PQS. A C-rich region in the coding strand is shown equivalently as CCC, and would allow a G-quadruplex to form on the template strand (red) and is here referred to as a C-PQS. After transcription, G-PQS, but not C-PQS, also results in the formation of a G-quadruplex in the mRNA (green).

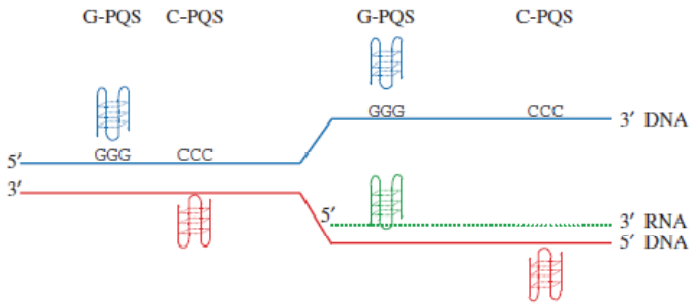


Fig.7 Formation of RNA quadruplexes on the template strand

-A recent finding[176] about direct deregulation of a protein involving RNA has appeared. The human Telomerase RNA Component (hTERC) is the second part of the nucleoprotein complex telomerase. In the telomerase RNA of most vertebrates, the predicted structures possess a helical domain called the P1 helix. This helix may serve as a template boundary element to limit reverse transcription. Gros *et al.* demonstrated first that specific RNA fragments from hTERC could form a stable quadruplex in vitro, and that G4 formation could interfere with the formation of the P1 helix. Moreover, addition of the quadruplex binding ligand 360A[177] (see structure in chapter 1) further inhibits P1 helix assembly. Hence telomerase function/activity might be regulated via formation of this “putative” parallel RNA G4. (Fig. 8.)

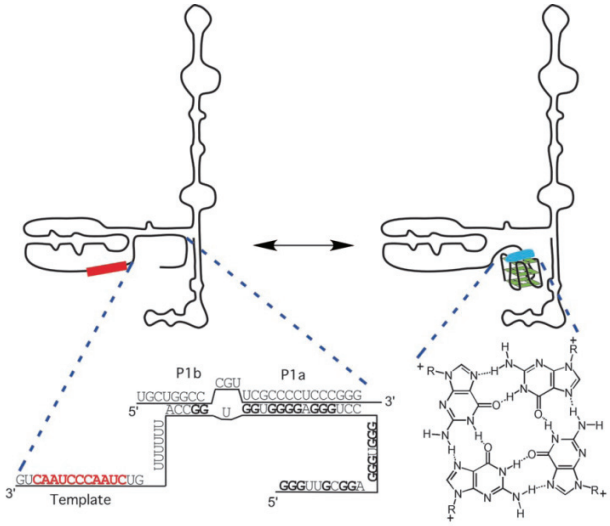


Fig. 8. hTERC quadruplex (right) induction via 360A chaperone, preventing P1 helix (left) formation

-Since then, additional examples have been reported in the literature which demonstrates the inhibition of translation in living eukaryotic cells by RNA G-quadruplex motifs. [178, 179] However, these examples will not be developed in this manuscript.
Having discussed the inhibition of telomerase via the targeting of a quadruplex located in its RNA component hTERC, it is noteworthy that telomerase can also be inhibited by complementary PNA sequences that target its single-stranded mRNA template[180].

ii. PNA recognition of nucleic acids

As highlighted in chapter 2, PNA can hybridize with nucleic acids (DNA and RNA) via sequence specific Watson Crick base pairing [140] . Below are more specific details of these PNA:DNA and PNA:RNA interactions:

-Two types of hydrogen-bonding motifs have been proposed to account for the extraordinary stability of polyamide "peptide" nucleic acid (PNA) hybrids with nucleic acids: the existence of inter-residue and intra-residue hydrogen-bond motifs[181]. (Fig. 9.)

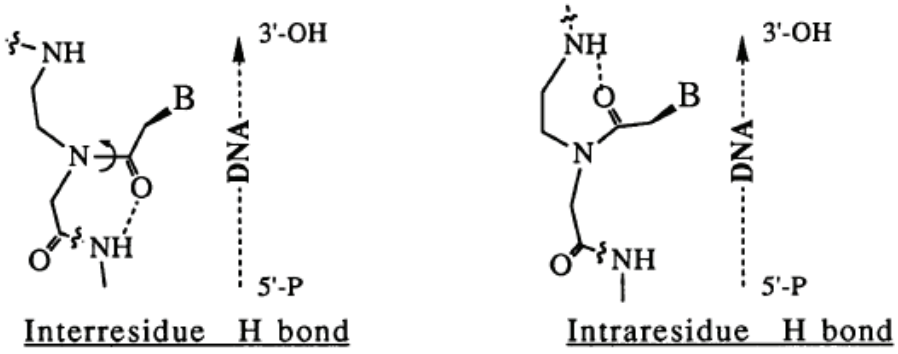
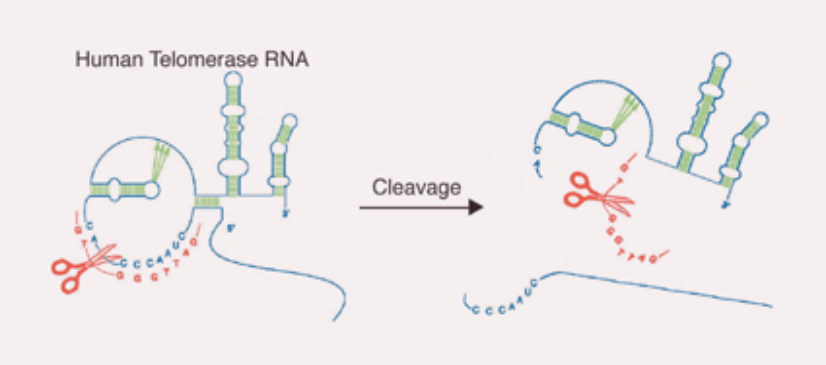


Fig. 9. Hydrogen bond motif for PNA-DNA stability (B is a nucleic base)

Energy minimized structures of Watson-Crick base-paired decameric duplexes of PNA with A-, B-, and Z-DNA and A-RNA polymorphs indicate that the inherent stability of the complementary PNA helical structures is derived from inter-residue, rather than from intra-residue hydrogen bonds. Intra-residue hydrogen-bond lengths are longer than inter-residue hydrogen bonds and helical strand stability with inter-residue hydrogen bond stabilization follows the order[182]:
B-(DNA-PNA) > A-(DNAPNA) = A-(RNA-PNA) > Z-(DNA-PNA).
A preferential binding to DNA over RNA can be arranged by pyrrolidinyl PNAs[183] and in the same manner sequence-specific targeting of nucleic acids can be obtained by locked nucleic acids (LNAs)[184] , but these two examples will not be discussed in details in this manuscript.

-In 2003 Whitney *et al.*[185] designed specific PNA cleavers for human telomerase RNA hTERC. Specific regions of hTERC could be recognized by a complementary peptidic sequence containing a neocuproine·Zn RNA cleaving agent. Ribozyme-mediated cleavage of hTERC had already been reported to inhibit telomerase. The study by Whitney *et al.* used the concept of "recognition and function" and provided the first demonstration of a PNA-based ribonuclease mimic that utilizes a metal-cleaving system. (Fig. 10.)



P_C: K-GATTGGGP*TTGACTC-K
P_N: K-ACAGATTGGGP*
hTR21: 5'UUUGUCUAACCCUAACUGAGA 3'

Fig. 10 top: hTERC cleavage; bottom: Sequence of PNA cleavers and their RNA target. (P* is the neocuproine)

-Since then, antisense PNA sequences have been associated to fluorophores and used as sequence specific probes. Recent examples using PNA-heterocycle conjugates are given below.
(i) Sequence specific intercalation into B-DNA using anthraquinone-PNA conjugates[186],

(ii) PNA encoded substrates using rhodamine fluorescence were designed to profile proteolytic activity from single proteases[187],
(iii) high-affinity peptide nucleic acid (PNA) probes conjugated with a donor (Thiazole orange) and acceptor (Alexa-594) fluorophores were employed for labeling mRNA transcripts [188],
(iv)DNA templated reactions were carried out using two PNA strands, functionalized with either a profluorescent azidorhodamine or a trialkylphosphine, that can react with each other via a Staudinger reaction upon binding to a complementary DNA template only, thus generating a fluorescent azidocoumarin[189].

-Concerning the direct recognition of G4 nucleic acid sequences towards their flanking region a nice example has been provided by Deng and Zhou[190]. Two splits DNA (A and B) can assemble when a third DNA split C contains their complementary sequence. A peroxidase like activity DNAzyme, PS2.M[191], with a special quadruplex structure, was used here (with modifications) to monitor the three splits assembly.
When the split B matches, the result is a hemin containing bimolecular (3+1) quadruplex which upon formation catalyzes H_2O_2-mediated $ABTS^{2-}$ oxidation to obtain $ABTS^{\bullet -}$. The recognition process is highly colorimetric and allows visual detection. (Fig. 11.)

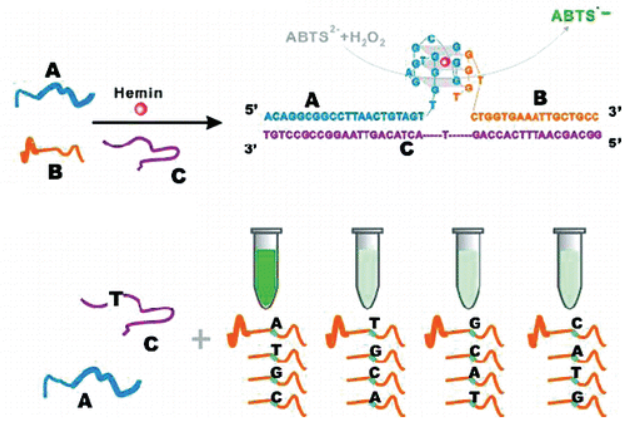


Fig. 11 Model of the recognition process

On the figure we observe that the distance between the quadruplex and the targeted flanking regions is **one** nucleic base (T on each side).

iii. PNA recognition of RNA G4s

A PNA can either interact with the folded target (e.g. interaction with a double-stranded DNA via formation of a triple helix PNA:DNA:DNA) or induce a conformational change by "invading it" (see chapter 2).
Marin and Armitage[192] who already successfully demonstrated that PNA could invade DNA quadruplexes (see chapter 2), designed a PNA probe that would turn an intramolecular RNA quadruplex into 2 linked bimolecular PNA-RNA quadruplexes (Fig. 12.). To prove this hybridization they attached to this short PNA a molecule of Thiazole orange and via fluorescence study, proved that the invasion process happens in only ten minutes at room temperature.

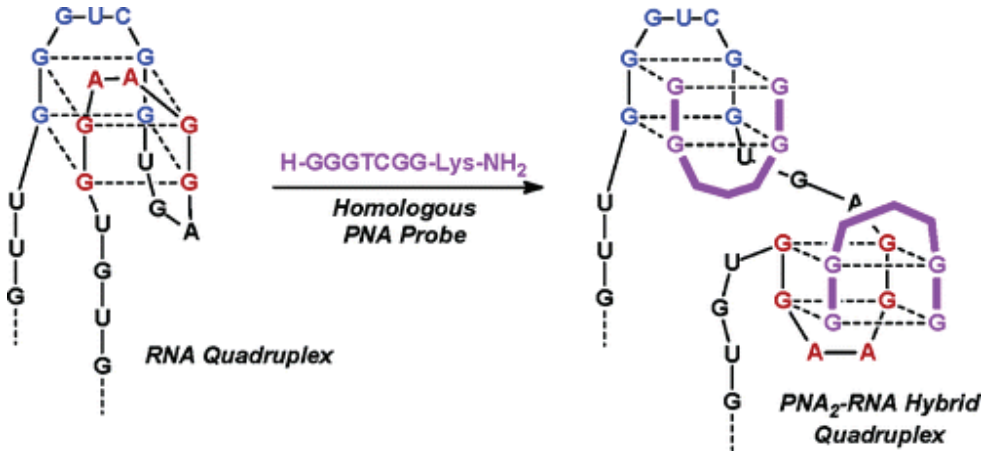


Fig. 12. Description of RNA quadruplex invasion by two PNA strands.

We may note here that a 2-quartet quadruplex re-folds into two 2-quartet quadruplexes.
The driving force of this invasion is still not clear but the number of hydrogen bonding might be involved, therefore the non-duplex structure of RNA would favor such hybridization.

b. Strategy

i. Objectives

Surprisingly, RNA quadruplexes have received so far much less attention than their DNA analogues although there is now growing evidences that RNA quadruplexes are involved in key biological processes. Moreover, RNA quadruplexes are also more likely to form *in vivo* than their DNA analogues given the evidence that RNA is single stranded and thus does not necessarily need to compete with complementary strands to fold into a quadruplex. While recent biological and biochemical studies unambiguously demonstrate the capacity of RNA quadruplexes to interfere with RNA translation for instance, the possibility to regulate such processes by stabilizing or inducing the formation of RNA quadruplexes *in vivo* requires the existence of ligands that can target these structures with both high affinity and specificity.
Very striking is the fact that no synthetic ligands have been designed yet that are directed specifically against RNA quadruplexes and can also discriminate between an RNA quadruplex and another or between an RNA quadruplex and another RNA secondary structure. The main limitation to a therapeutic approach targeting RNA quadruplexes to regulate the expression of specific genes remains the capacity for a small molecule to bind to the quadruplex of interest with a high specificity, and so despite the high number of potential alternative RNA targets *in vivo*.
Therefore, there is an urgent need for highly specific RNA quadruplex targeting ligands in order to validate RNA quadruplexes as a therapeutic target for chemical intervention. In this chapter, we will explore the hypothesis that intramolecular RNA G-quadruplexes represent a new target for chemical intervention to interfere with the transcription or translation of specific genes.

ii. Recognition of flanking bases

Within this chapter, we will explore the possibility of generating the first small molecules (stabilising ligands) that are highly specific for a unique RNA quadruplex versus other RNA secondary structures, including other RNA quadruplexes. The general strategy consists in targeting simultaneously the quadruplex structure itself but also the two single stranded RNA flanking regions in a sequence specific manner. In order to achieve that, we designed hybrid molecules of general structure PNA_1-Heterocycle-PNA_2 where Heterocycle and $PNA_{1,2}$ are predicted to recognize the quadruplex structure and the two flanking single stranded RNAs, respectively

(Fig. 13.). Unlike all other G4 binding ligands described in the literature, these PNA-heterocycle conjugates should exploit the PNA-RNA sequence specific base-pairing association to direct the quadruplex binding platform towards the only quadruplex of interest i.e. that located in between the two single-stranded flanking regions. The probability of finding two RNA quadruplexes with identical flanking sequences being very low, our ligands are very likely to be significantly more specific than any other quadruplex binding ligands reported so far.

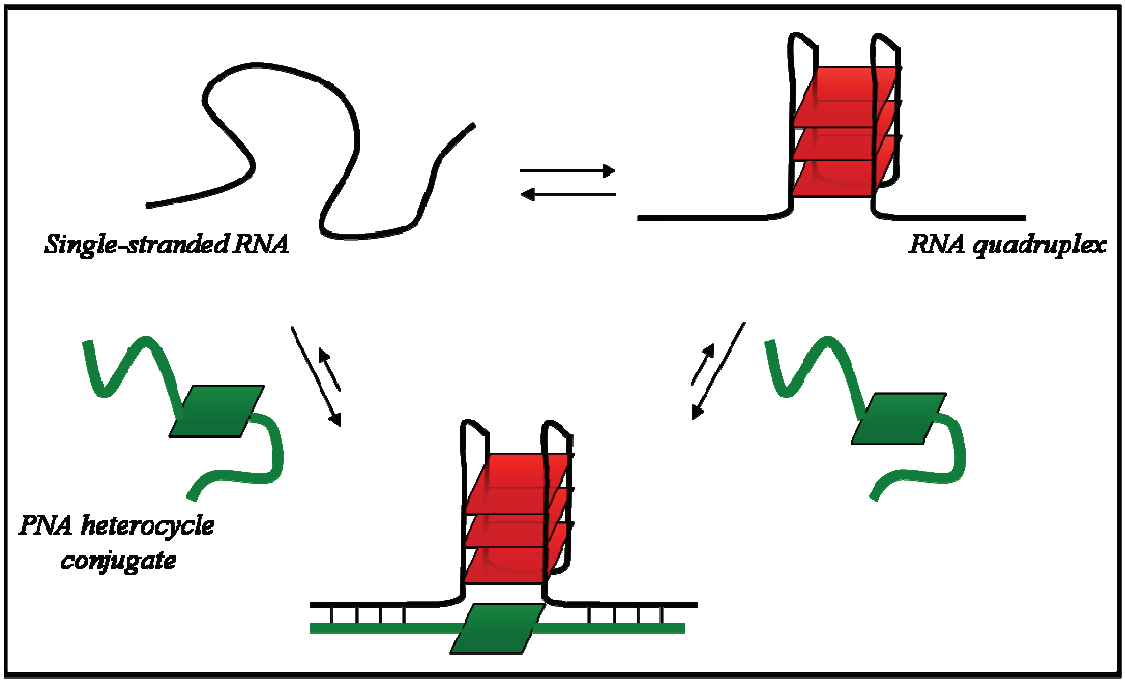
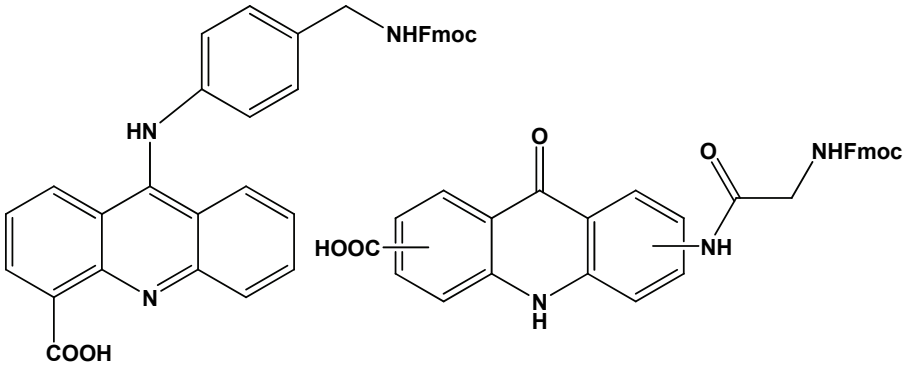


Fig. 13. General Strategy for a "sequence+structure" targeting of RNA quadruplexes

iii. Design of the heterocyclic moiety

Within all the quadruplex binding scaffolds available in the literature, we have chosen to use acridones and acridines since (i) their quadruplex "end-stacking" binding mode has been clearly established, (ii) their fluorescent properties can be exploited to monitor their interaction with RNA, (iii) they can be readily tri-functionalised, thus allowing us to incorporate them into a solid supported PNA synthesis as a Fmoc protected amino acid and to introduce additional recognition motif for an improved RNA affinity.

PNA-acridone conjugates of general structure PNA_1-Acridone-PNA_2 will be synthesized from a carefully designed amino acid acridine and commercially available PNA monomers assembled on solid support, using Fmoc chemistry. It is expected that these ligands will bind to a unique RNA target via a combination of π–π stacking interactions (acridine/G-tetrads), electrostatic interactions (Lysine side-chains/RNA phosphate backbone) and PNA/RNA sequence specific hybridisation (Watson-Crick base-pairing).
This strategy required the synthesis of an Fmoc protected acridone amino acid that could be conveniently incorporated to the PNA strand by solid-phase synthesis. First generation acridine/acridone amino acids envisioned were of general structure **A** or **B**(Fig. 14.). The synthesis of acridine **A** functionalized at positions 4 and 9 has been reported previously by Beal *et al.*[193] and has been used for Fmoc-based solid phase peptide synthesis. However, for an ease of synthesis and for a better stability of the heterocycle (see Chapter 3 for more details about stability of 9-amino acridines) we decided to synthesize an acridone heterocycle of general scaffold B.

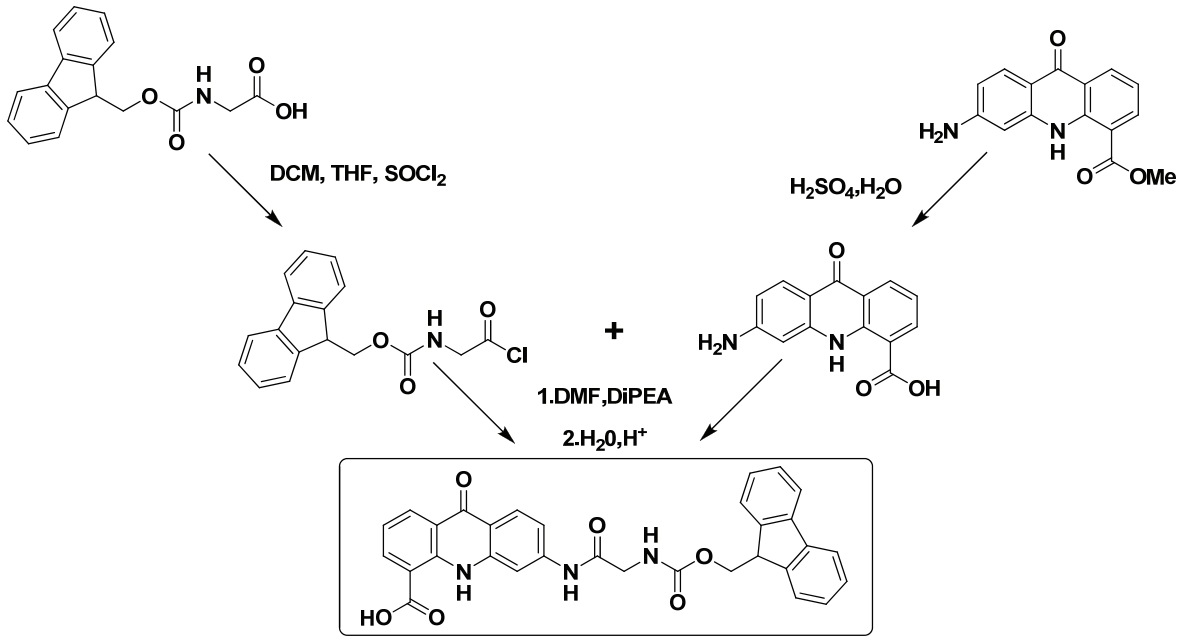


Scaffold A **Scaffold B**

Fig. 14. “Fmoc-Acridones”

2. Chemistry:

General synthetic pathway:



i. Acridone synthesis:

The Fmoc protected acridone of interest was synthesized as described on the scheme above using the same 6-amino-4-methyl ester acridone precursor which was previously used in Chapter 2. The methyl ester was readily hydrolysed in aqueous sulfuric acid to afford quantitatively the corresponding acridone amino acid. The Fmoc protected amine was introduced onto this scaffold by amide coupling between Fmoc-Glycine and the acridone amino acid. This glycine also acts as a linker by introducing some flexibility between the acridone platform and the PNA strand so it will be attached to the acridone. The desired Fmoc protected acridone was obtained by reaction between the amino acid and the acyl chloride of Fmoc-Glycine [194] in the presence of diisopropylethylamine (DIPEA) in DMF. Conveniently, the desired product was isolated by precipitation in DMF.
It was further purified by washes with a refluxing solution of slightly acidic water. Herein we have developed a general strategy to access Fmoc-protected acridone amino acids that can be incorporated directly into the SPS of PNAs.

ii. Choice of the RNA quadruplex target:

According to our strategy, within the PNA_1-Acridone-PNA_2 the acridone could potentially "target" any quadruplex but the lateral sequences would "**match**" in a sequence specific manner the flanking arms of the quadruplex.
When designing our PNA sequences, we chose to target RNA sequences starting <u>one</u> **base** upstream and one base downstream the quadruplex forming sequence, as inspired by the work from Deng *et al.*[190]
For the proof-of-concept experiment, the intramolecular N-RAS RNA quadruplex identified in the Balasubramanian's group was chosen as our target of interest due to the important role it was shown to play in the regulation of the N-RAS m-RNA translation [19]. However, once validated, this general approach could potentially be applied to any quadruplex target, by simply adapting the PNA sequences to the actual target.
As a model system, the following PNA sequence was synthesized which contains (i) a central acridone targeting the terminal G-tetrad of the G4, (ii) two Glycine-Lysine linkers between the acridone and both PNA sequences to confer some flexibility and water solubility to the final ligand, (iii) two 3-bases long PNA sequences targeting the G4 flanking regions and a C-terminal Lysine to make the PNA more water soluble and that can be involved in electrostatic interactions with the negatively charge RNA phosphate backbone:

"Lys-AAT-Lys-Gly-Acridone-Gly-Lys-TAC". See Fig. 15.

According to our design principle, this ligand would therefore be able to target any quadruplex forming sequence of the type:
5'-UUA-B-[$G_nB_{x1}G_nB_{x2}G_nB_{x3}G_n$]-B-AUG-3' where B can be any RNA base (A, U, G or C), x≥1, and n≥2.

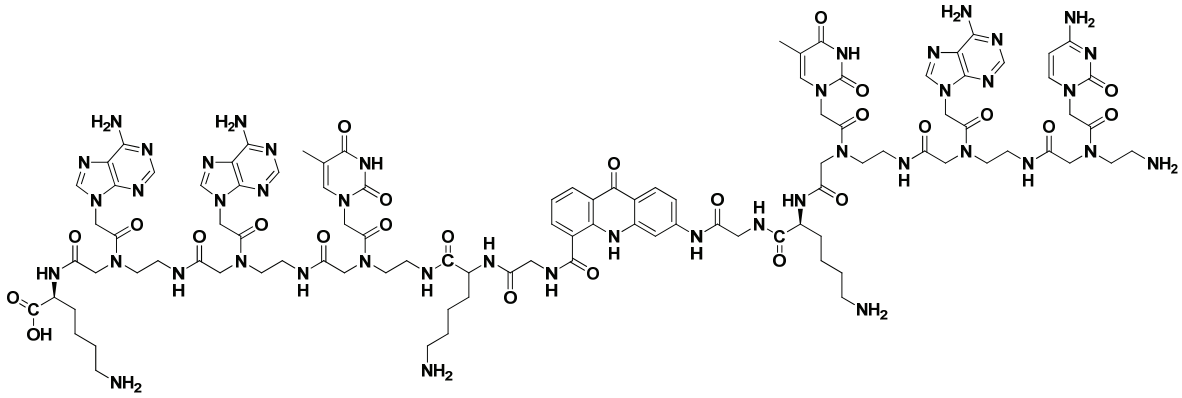


Fig. 15. Structure of the first PNA_1-Acridone-PNA_2 RNA quadruplex ligand. The role of the first lysine is improved water solubility; central (Lys-Gly) motifs improve both water solubility <u>and</u> flexibility.

Once synthesized and purified by HPLC, we investigated the ability of our PNA_1-Acridone-PNA_2 conjugate to interact with the N-RAS RNA quadruplex using UV-melting studies.

3. Results

In order to determine whether our ligand could (i) bind to a RNA quadruplex sequence with two single-stranded flanking regions and (ii) discriminate between two identical quadruplexes bearing different RNA flanking regions, we investigated the interaction between our synthetic ligand and the 5 RNA sequences listed below:
-a "match" sequence with 3 tetrads UUAU**GGG**A**GGGG**C**GGG**UCU**GGG**UAUG (*N-RAS1*)
-a "match" sequence with 2 tetrads UUAU**GG**UA**GGG**UC**GG**UUCU**GG**UAUG (*N-RAS2*)
-a "pseudo-match" sequence with 3 tetrads GCUU**GGG**A**GGGG**C**GGG**UUCU**GGG**UUGC (*N-RAS3*)
-an "empty" sequence with no flanking regions and 3 tetrads **GGG**A**GGGG**C**GGG**UCU**GGG** (*N-RAS0*)
-a "mismatch" sequence with 3 tetrads CUCU**GGG**A**GGGG**C**GGG**UCU**GGG**CUCU (*N-RAS4*)
UV melting studies (monitored at 295 nm) were carried on solution containing 10 µM RNA (*N-RAS 0-4*) in TRIS.HCl buffer (pH 7.4), in the absence and in the presence of a stoichiometric amount of PNA-acridone conjugate, and varying the nature and concentration of added salt (potassium, sodium and lithium). All the melting temperatures (Tm values) obtained are summarized in the table below: (temperatures are given in degrees Celsius)

Melting Temperatures		**NRAS 1 10µM**		**NRAS 2 10µM**		**NRAS 3 10µM**		**NRAS 4 10µM**		**NRAS 0 10µM**	
Cation	Conc. (mM)	No PNA	PNA 10µM	No PNA	PNA 10µM	No PNA	PNA 10µM	No PNA	PNA 10µM	No PNA	PNA 10µM
Na+	0.5	*35°*	*71°*	NC<	43	35°	69°	50°	52°		
	2.5	37°	NC>								
	5	39°	NC>								
	10	43°	NC>								
K+	0.5	51°	NC>			50°	NC>	60°	64°	52°	54°
Li+	50	NC<	52			NC<	50°			42°	48°
	100	NC<	43								
	150	NC<	41								
Without cations		NC<	62°	NC<	40	NC<	63°	NC<	NC<	45°	44°

All the melting temperatures are determined with the minimum of the first derivative of the cooling curve when possible (+/- 1°C). Experiments were carried out in triplicate and the given values are the average of three measurements.
NC < means that the Tm calculation was impossible: if existing the melting temperature was below 25°C
NC > means that the Tm calculation was impossible: if existing the melting temperature was above 85°C

General Analysis of the results:

Salt Dependence and Stability:
We observed the general trend that in the absence of PNA G4 quadruplexes containing solutions in are more stable in K^+ than in Na^+ or Li^+
This is also seen in melting experiments carried out in the presence of PNA:
$K^+ > Na^+ > Li^+$
As already observed by others, increasing concentrations of Li^+ destabilized the RNA quadruplex structure (11°C drop in Tm value when increasing Li^+ concentration from 50 to 150 mM).
In contrary to that, the increasing concentration of either Na^+ or K^+ significantly increased the stability of the every quadruplex (e.g. Tm value increased by 12 °C when increasing the Na^+ concentration from 0.5 to 10 mM).

Effect of flanking regions:
It is generally well accepted that flanking regions (sequence, length) can have a significant effect on RNA quadruplex stability. Our results in salt free conditions are consistent with that observation as we do not observe any stable quadruplex formation for the RNA sequences possessing flanking regions (*NRAS1-4*) whereas the RNA sequence lacking flanking regions (*NRAS0*) forms a stable quadruplex with a measured melting temperature of 42° C.
However the influence of the RNA flanking regions on quadruplex stability decreased when in the presence of K^+. For instance, comparable thermal stabilities are obtained for NRAS1 and NRAS0 in buffer containing 0.5 mM KCl.
As anticipated, the quadruplex formed by folding of the *NRAS2* sequence (i.e. containing 2 tetrads only) proved less stable than the corresponding quadruplex containing three tetrads (i.e. that formed by the sequence *NRAS1*), whatever the conditions (with or without salt, with or without PNA).

Quadruplex induction:
Of particular interest is the capacity of our PNA to induce the formation of an RNA quadruplex under conditions where this RNA naturally does not form any detectable quadruplex. This is illustrated by comparing the UV melting curves of the

NRAS1 RNA sequence in water without added salt in the absence and in the presence of a stoichiometric amount of PNA. (Fig. 16)

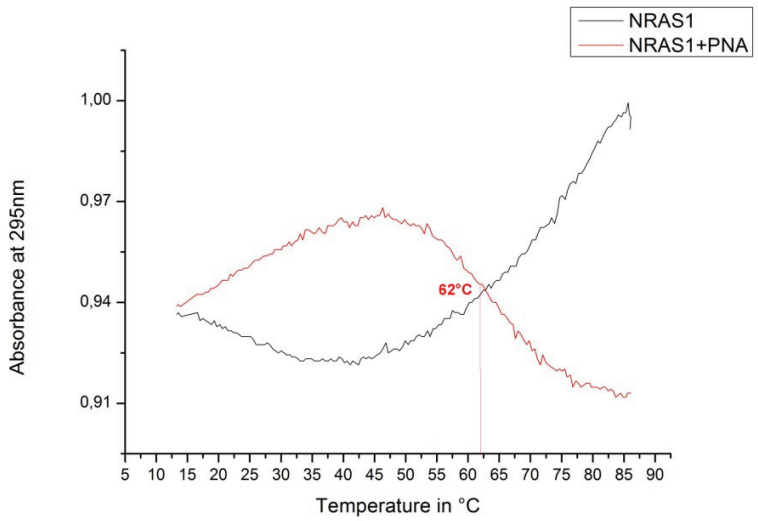


Fig. 16. The sequence match "NRAS1" forms a G4 without added salt and in the presence of PNA whereas no distinctive structure is observed without ligand.

While NRAS1 incubated with our PNA forms a stable quadruplex with a melting temperature of 62 °C (red curve), no real transition is observed in the absence of ligand. A similar trend is observed with the NRAS2 sequence (2-tetrad quadruplex), although the quadruplex formed in the presence of PNA is slightly less stable than that formed from NRAS1 (40 °C vs 62 °C) due to the smaller number of G-tetrads formed.

Our main interest here was to demonstrate the possible ligand-induced stabilization of an intramolecular RNA quadruplex targeting its flanking regions as well as its core structure.
In the absence of added salt, NRAS0 (lacking flanking arms) displays comparable thermal stability with or without PNA. All other RNA sequences tested (NRAS1-4) containing flanking arms (complementary or not to our PNA) proved unable to form a stable quadruplex in the absence of ligand. Interestingly, all the three sequences containing quadruplex flanking arms that are fully (NRAS 1-2) or partially (NRAS3) complementary to the ligand PNA sequence have been proved capable to form a stable quadruplex upon addition of ligand. In contrary to that, NRAS4 with non complementary flanking arms proved unable to form a quadruplex in the absence or presence of ligand.

<u>Quadruplex stabilization:</u>
We have just seen that our PNA ligand is capable to induce the formation of an RNA quadruplex. Herein, we aim to demonstrate that it is also capable to stabilize a quadruplex under conditions where it naturally forms. A major challenge here was

to find conditions of salt where two quadruplexes formed in the absence and in the presence of PNA would both display a melting temperature within the detectable range 20-80°C.
This was achieved when using 0.5 mM Na^+ conditions. While NRAS1 (three tetrad quadruplexes, fully complementary flanking arms) formed a weakly stable quadruplex in the absence of ligand (Tm = 35 °C) a 36 °C (see Fig 17.) increase in the melting temperature was observed in the presence of the PNA. Under similar conditions the mismatch sequence NRAS4 was only poorly stabilized by addition of the ligand (i.e. +2°C). (see Fig. 18)

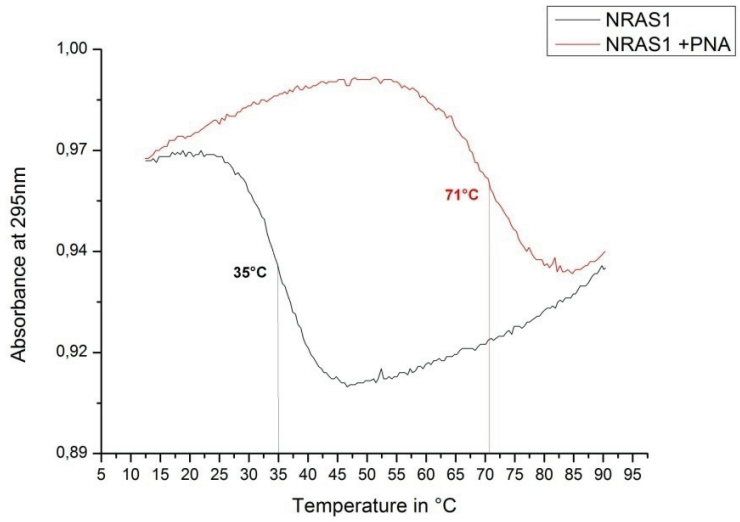


Fig. 17 UV melting curves of the NRAS1 quadruplex in TRIS.HCl buffer (10mM, pH 7.4) also containing 0.5 mM NaCl and in the absence (black curve) or presence (red curve) of a stoichiometric amount of PNA. A ligand-induced +36°C stabilization is observed.

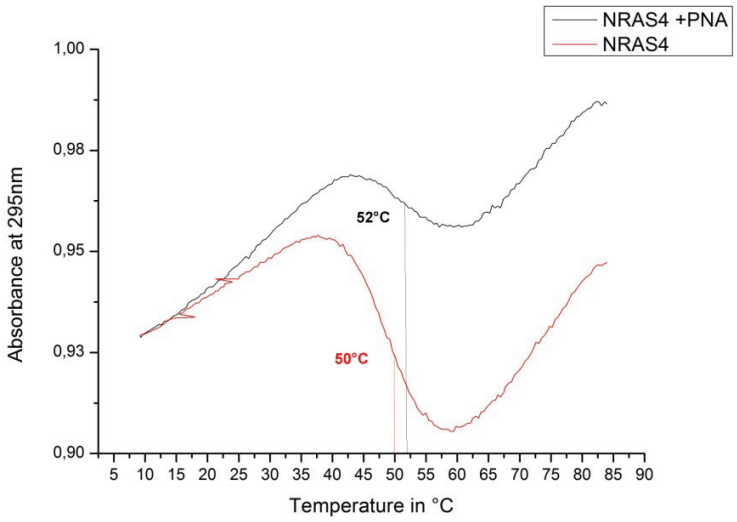


Fig. 18 UV melting curves of the NRAS4 quadruplex in TRIS.HCl buffer (10mM, pH 7.4) also containing 0.5 mM NaCl and in the absence (red curve) or presence (black curve) of a stoichiometric amount of PNA. A ligand-induced +2°C stabilization is observed.
(Note of the author: the "pseudo match" NRAS3 sequence here is brought to the reader as a "careful" result. No interpretation can be reasonably done with those results as we can't define exactly the mode of action of the PNA as long as the "matching" capacity of RNA targets is not optimized)

In potassium containing solutions, the ligand induced stabilization effect proved more difficult to evaluate. Indeed, the RNA alone (i.e. in the absence of ligand) formed a highly stable quadruplex (>50 °C) even with low concentration of KCl (0.5 mM). Therefore under such conditions, addition of our ligand resulted in no thermal transition, which is likely due to the formation of an extremely stable PNA:RNA quadruplex complex displaying a melting temperature higher than 85°C (i.e. not detectable under standard conditions). Therefore, it was not possible to determine a ΔTm under such conditions.

<u>Interpretations of the ligand mode of binding:</u>
We are aware that these results are still preliminary and more work is needed to fully understand the mode of binding of our PNA-acridone conjugate.

The PNA-Acridone conjugate was initially designed to interact with the RNA G4 target by stacking of the acridone onto the terminal G-tetrad of the quadruplex and Watson-Crick base pairing of the PNA sequences with the RNA quadruplex flanking bases. It is very difficult however to differentiate here those two interactions and to determine whether both the PNAs and the acridone interact with the target. In the case of NRAS0 without flanking regions a slight increase in the Tm value upon addition of the ligand in the presence of K^+ (+2°C) and Li^+(+6°C) suggests that the acridone plays a moderate stabilizing role alone. But it still remains to be proven by

synthesizing a PNA ligand lacking the acridone (see section below). A similar effect can be seen for NRAS4: addition of the ligand leads to an increase in Tm value of +4°C and +2°C in the presence of K^+ and Na^+, respectively. This moderate stabilization seen with this "negative control" can be explained by either a stabilizing interaction of the acridone with the RNA quadruplex structure or also by unspecific interaction of the PNA with the non-complementary RNA flanking arms.
The tremendous increase in thermal stability of +37°C observed in 0.5 mM NaCl containing solution is believed to be a combined effect of acridone stacking (resulting in a general hydrophobic effect) and hydrogen bonds through PNA:RNA WC base pairing. Amongst both the WC base-pairing is likely to be the main driving force as we don't see considerable increase in Tm when the PNA cannot match the flanking regions.
This suggests that the contribution of the acridone itself on the ligand-induced G4 stabilization is likely to be only moderate. It would also be interesting to determine if the combined effects described above are synergetic or just complementary.

4. Extension of this work – Openings:

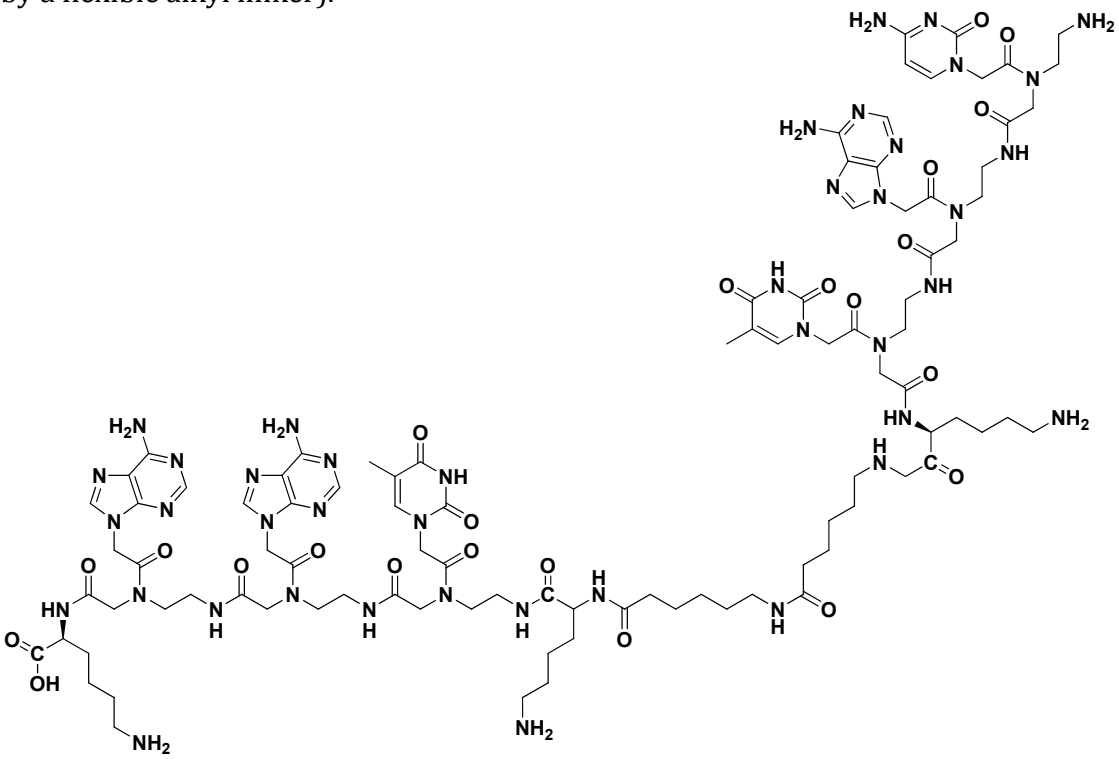
Our results are consistent with a Watson-Crick recognition process of the RNA single stranded flanking regions of the quadruplex by the complementary PNA, thus leading to a specific targeting of a quadruplex of interest (vs other quadruplexes of similar structure but containing different flanking arms), which is, to the best of our knowledge, the first example of that kind. Understanding the precise nature of the PNA:RNA interaction must however still be determined, by X-ray crystallography or NMR spectroscopy.

Although the strong affinity and high specificity of our model-ligand are very promising, these preliminary results also raise a number of questions:

(1) Effect of G4 conformation on the ligand recognition
(2) Effect of the PNA on the G4 conformation
(3) Effect/necessity of the acridone molecule for specific binding
(4) PNA length required to ensure specific recognition
(5) Biological effect of the ligand on RNA translation in vitro and in vivo

Whilst questions 1 and 2 can be addressed with complementary biophysical studies, answers to questions 3 and 4 will require the synthesis of additional ligands (for

instance, see structure below where the “Gly-Acridone-Gly” motif has been replaced by a flexible alkyl linker).



Indeed, hybridization of PNA_1 and PNA_2 to the two flanking regions of the RNA quadruplex may induce a significant “packing” of the RNA strand that could be sufficient to induce the formation of a stable quadruplex.

PNA :

The number of bases that « match » the RNA flanking regions is also an important factor that will need to be investigated. We already have demonstrated that two PNA sequences of 3 bases each were sufficient to induce a sequence specific recognition of RNA quadruplexes. However, longer PNA strands (e.g. 4 or 5 bases long) can also be envisioned to lead to an even better specificity.

CD:

The general conformation of most intramolecular RNA G4 identified to date tending to be parallel[195], one may anticipate that the quadruplex structure remains parallel upon binding of our G4 ligand. However this implies the existence of 3 chain reversal loops and the necessity for our PNA arms to be highly flexible. Alternatively, one may also consider that our ligand induces a G4 conformational switch from parallel to a hybrid parallel/antiparallel structure. Unfortunately, preliminary attempts to use Circular Dichroïsm to determine the effect of PNA binding on the G4 conformation proved difficult to interpret.

If the quadruplex structure is unchanged during the binding process of the PNA which means its structure remains parallel then one can optimize the perfect structure of the PNA with respect of the parallel structure obtained by Parkinson and al. for telomeric G4 (Introduction Fig. 16) . Indeed the 5' and 3' ends of the quadruplex would be on the same side and this particular side would represent an empty groove were the PNA would fit in. One can consider either that 5' and 3' ends are close and design a short link between arms or that molecules with intercalating behaviors like acridines may fit the groove. Those two last approaches have to be investigated to rightfully speak about drug design.

Possible outcome of this project:

In this project, we aimed to develop the first family of highly specific RNA quadruplex stabilizing ligands that can regulate the expression of specific genes at the translation level. Using biophysical approaches, we demonstrate that a hybrid PNA-acridone conjugate ligand could induce a tremendous stabilization of naturally occurring quadruplexes. Proof of concept studies was carried out on the previously characterized NRas RNA quadruplex. However, we have also identified other genes of clinical interest that contain, in their messenger RNA, sequences that could potentially form a quadruplex structure. These include Aurora kinase A and Aurora kinase C, members of the Aurora family of serine/threonine kinases which are essential for mitotic progression (See sequences below).

Putative RNA forming sequences in the mRNA of Aurora kinase A and C
Aurora A: 5'UUUUCGGAGGCCGAGGCAUC**AUG**GACCGA3'
Aurora C: 5'AAAUCGGGCGUCCCCUGGGCAAGGGGAAAUUUGGGAAUGU3'

The Aurora kinase A in particular has attracted intense interest following the discovery that the chromosomal region in which it is located commonly undergoes amplification in epithelial cancers.
It is noteworthy that the quadruplex forming sequence found in Aurora A includes the actual start codon serving for the translation of the Aurora A protein (highlighted in red). Formation of an RNA quadruplex at this particular location could therefore potentially block completely the initiation of the translation process and abolish Aurora A expression. As part of some preliminary work, we have already demonstrated that this sequence could form a quadruplex in vitro. As anticipated for a quadruplex containing two G-tetrads only, this structure has a moderate thermal stability (Tm value of 58 °C as measured in a solution containing 100 mM KCl), which remains relatively stable compared to most DNA quadruplexes but significantly less stable than the N-Ras RNA quadruplex for instance. However, this moderate stability will allow us to better estimate the ability of our ligands to further stabilize this structure, thus leading to a ligand induced biological effect.

Our novel approach for specifically silencing this gene is likely to have great therapeutic potential. Given the large number of putative quadruplex forming sequences that are present in the human genome, our "sequence + structure" specific approach could open the way to a new family of potent anti-cancer agents.

Chapter 4:

When 9-Aminoacridines exchange their amines

"What we think, we become."

Do acridines exist to be used for dynamic combinatorial chemistry?

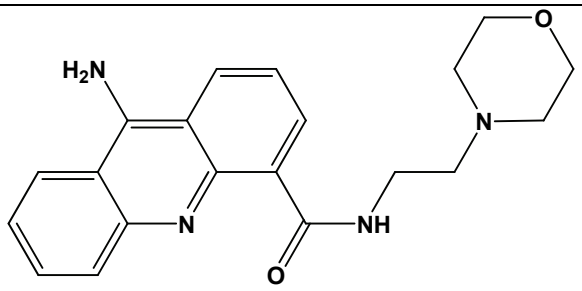
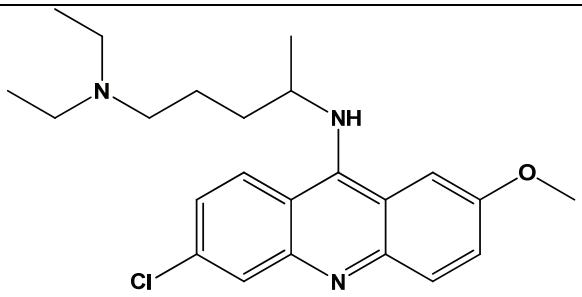
1. History:

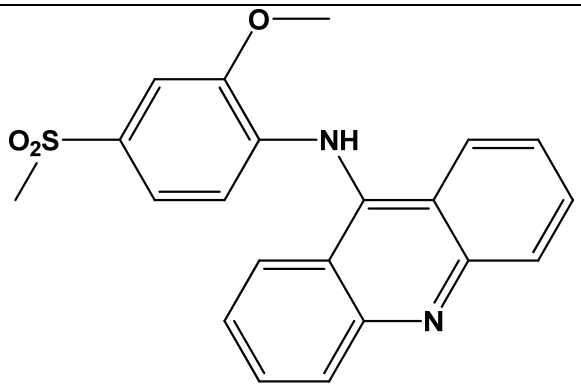
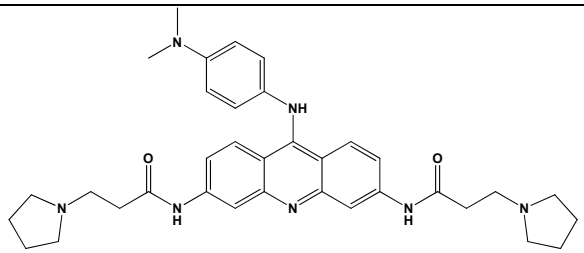
As it was already mentioned in the general introduction acridines are heterocyclic structures that show a broad range of biological properties since they are currently developed as antitumor, antiparasite and antibacterial agents. They have been known since the 19th century as potent antibacterial agents but were more recently shown to exhibit promising anti-cancer activities. Amongst those molecules a specific class can be distinguished: the 9 Amino Acridines (9AA).

a. 9 aminoacridines:

i. Selected examples of biologically active structures:

For a better understanding of those molecules, some representative examples of biologically active 9AA are presented below:

9-a-DACA[196, 197]		Derived from DACA N-[2-(dimethylaminoethyl)] acridine carboxamide family. A well-known Topoisomerase II poison that binds to DNA helices.
Quinacrine		Appeared in 1970 as a fluorescent probe. Originally used for chloroplast membranes but through years has been tested for many other applications such as anti malarial, or antitumor[198] activity or sterilization.

9-AMSA[199]		Intercalative agent m-amsacrine (m-AMSA) is built on the 9-anilino pattern and also inhibits Topoisomerase II by interfering with the covalent enzyme-DNA intermediate. It was initially developed as an anti carcinogenic agent.
BRACO 19		See introduction, BRACO 19 is a 3,6,9 trisubstituted acridine. Its capacity to preferentially bind G-quadruplexes was already detailed. However its synthesis will be shown later in this chapter.

ii. The 9 position:

The 9AAs differ from acridones in structure and in their electronic properties: they owe their structural specificity to the substituent at the C9 position that changes the electronic repartition of the general backbone.

1. Tautomerism:

The electronic specificity of 9 amino acridines is located at their "core".
9-amino acridines exist in solution as an equilibrium between two tautomeric forms. This equilibrium can be compared to an imine-enamine tautomerism as illustrated in Fig. 1

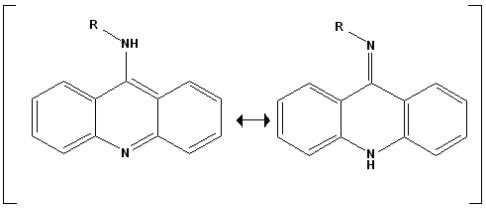


Fig. 1 Tautomerism inside 9AA

This existence of this equilibrium was demonstrated by theoretical analysis but also by chemical investigations (using ^{1}H and ^{13}C NMR).

2. Degradation:

It was first discovered that the 9AA could *irreversibly* hydrolyze into acridones. The meaning of this degradation is that the 9 position is sensitive (and vulnerable) to nucleophilic attacks, whether the nucleophile is ammonium hydroxide, amines, thiols and also water.
Goodell et al.[200] have studied this reaction in details (Fig. 2) and suggested the following mechanism:

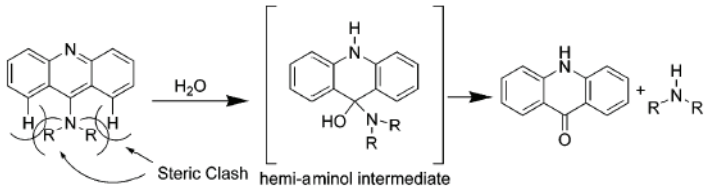


Fig. 2 Steric clashes and proposed hydrolysis mechanism

The general pKa of 9AA is close to 10, and the authors conclude that hydrolysis is more likely to happen under acidic conditions when the central N10 of acridine is protonated which favors electrons delocalization.
Their calculations also showed that steric clashes may favor the process in the case of 9-*N*,*N*-diethylamine-acridines.
Finally they note that hydrolysis may still occur under alkaline conditions, which may reveal the existence of more than one pathway to reformation of acridone from 9AA.

The capacity for hydrolysis of 9AA is therefore an underlying mechanism that has to be taken into account when utilizing such chemistry and as it will be mentioned later, when designing 9AA as therapeutics. It is also the first direct proof of the chemical reactivity of the 9 position of 9AAs.

3. Reactivity:

Understanding the electronic mobility inside the acridine core, Sebestik et al.[201] pointed out an interesting reaction, occurring during the synthesis of some acridine peptide conjugates: they observed that 9AA could undergo an amine transfer process occurring at the 9 position (see Fig. 3) and in organic solvents.

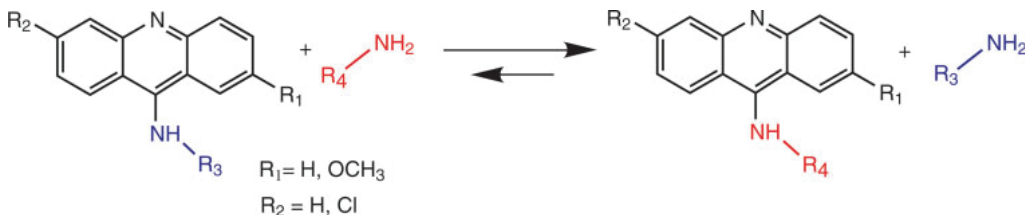


Fig. 3 Amine transfer on a Quinacrine type backbone

The authors conclude that this reaction may explain the reactivity of quinacrines towards several (biological) substrates.

b. *Dynamic combinatorial chemistry:*

i. Notions of thermodynamic templating

In 1997 Brady and Sanders[202], described a new approach to thermodynamic templating.
It starts with the classic description of the enzymatic catalysis when enzyme **E** binds substrate **S** and forms a complex **ES** with binding free energy ΔG_{ES}.

Transition states for the reaction are $\mathbf{TS}_{cat}$ for the catalyzed reaction and $\mathbf{TS}_{uncat}$ for the reaction in the absence of catalyst. The corresponding activation energies for the reaction are $\boldsymbol{\Delta G}_{cat}$ and $\boldsymbol{\Delta G}_{uncat}$. As $\boldsymbol{\Delta G}_{\mathbf{ETS}} > \boldsymbol{\Delta G}_{\mathbf{ES}}$ there is a rate enhancement due to the presence of the enzyme. Fig. 4

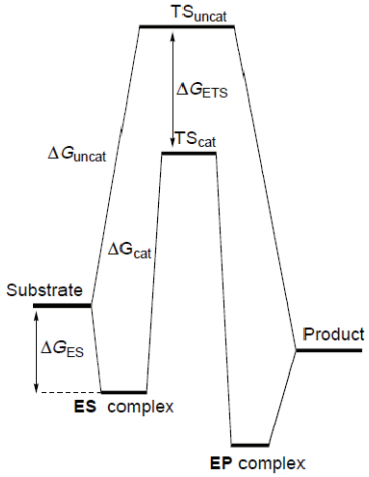


Fig. 4 Free energies profiles of the enzyme catalysis

Then from this "Lock (host) and Key (guest)" principle[203] they suggested a possible combinatorial formation of a host in the presence of a template. Several building blocks would preferentially assemble upon guest binding. Fig. 5

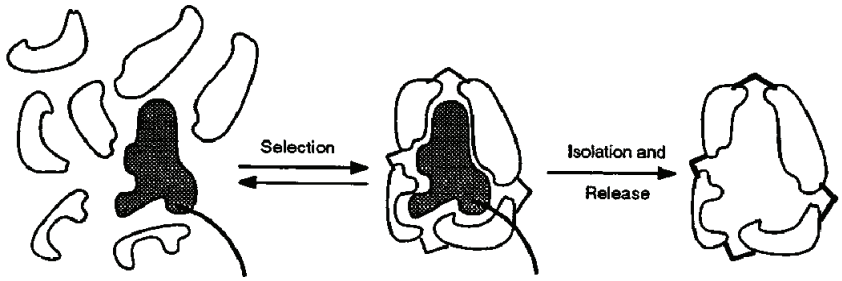
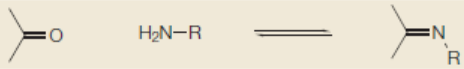
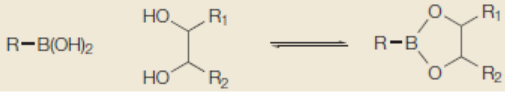
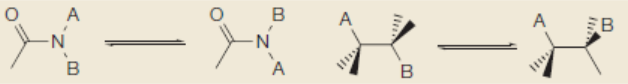
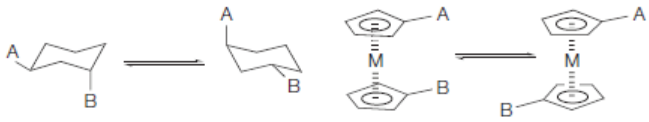
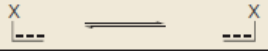


Fig. 5 Host synthesis in the presence of a template

The product should be preferentially formed in a reversible and thermodynamically controlled fashion. They outlined that such a principle would require a fast reaction even in the absence of an excess of reactant and illustrate this with a transesterification reaction[204].
So "thermodynamic templating" can be more generally represented as the possible modification of a reaction's list of products via a thermodynamic controlled reaction.

ii. Existence of reversible reactions

The reversible templating effect mentioned above is possible due to the existence of covalent / non-covalent reversible interactions. A reversible reaction can be defined as a reaction where the interconversion of molecules is still possible via recombination of building blocks in a non-static equilibrium. Below is a non-exhaustive list a such reactions or processes(established in 2001):[205]

Carbonyl reactions	
Imine formation	
Hemiketal formation	
Transacylation	
Aldol formation	
Michael reaction	
Disulphide formation	
Diels–Alder reaction	
Metathesis reaction	
Boronic ester formation	
Reversible interactions	
Metal coordination	
Electrostatic interaction	
Hydrogen bonding	
Donor–acceptor interaction	
Reversible intramolecular processes	
Configurational	
Cis–trans isomerization	
Conformational	
Internal rotation	
Ring inversion	
Structural	
Tautomerism	
Fluxionality	

More recently new reversible reactions have emerged that will not be developed in this manuscript. (For more details, see[206, 207])
From the knowledge of reversible reactions and their possible thermodynamic control arose the concept of dynamic combinatorial chemistry (DCC). The birth of this concept corresponds to simultaneous works from Pr J.K. Sanders and Pr J.M. Lehn.

iii. DCC and virtual combinatorial libraries:

In 1997, Huc and Lehn[208], strengthen the idea of possible "virtual combinatorial libraries" (VCLs). This term is different from "combinatorial libraries" resulting from static reaction products. Indeed, the possibility to generate diversity via molecular or supramolecular self-assembly is shown via the induction of inhibitors of carbonic anhydrase through imines formation. (See Fig. 6) The present case describes a virtual library (imines) of non natural substrates for a biological receptor (CA). Upon addition of the target (i.e. CA) the formation of certain imines is favored, due to stabilizing interaction with the target. A direct comparison of the libraries obtained in the absence and in the presence of the target allows the fast identification of the species that have been amplified upon addition of CA (i.e. the fast identification of the best CA ligands).

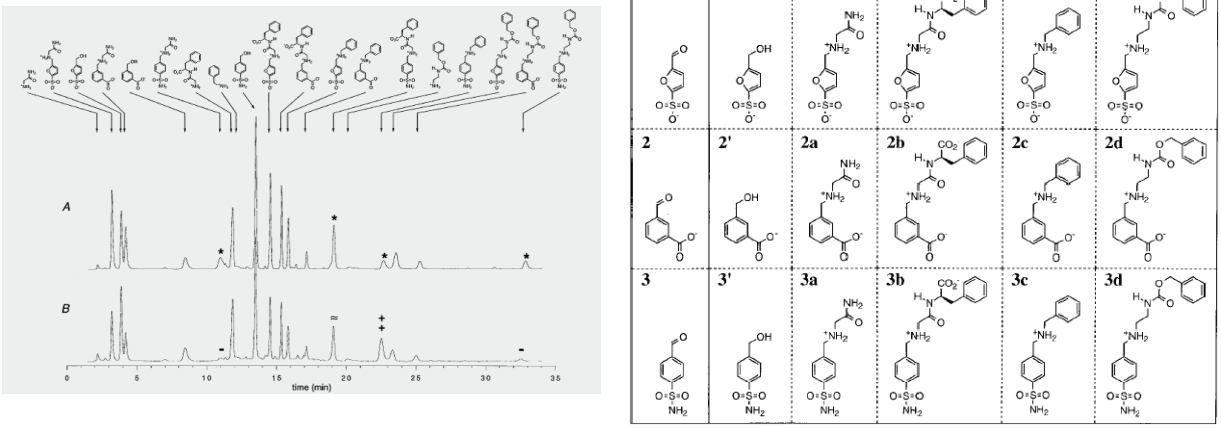


Fig. 6 Left: HPLC traces of dynamic combinatorial libraries without (A) and with (B) carbonic anhydrase.
Right: Possible VCL via imines formation followed by hydride reduction.

This allows us to introduce the so-called concept of DCL (dynamic combinatorial libraries) which highlights the difference between what is virtually accessible and what will happen through dynamic (templated or not) exchange. DCLs are subsets of the VCL and one component may generate many different DCLs depending on its "adaptive" behavior. A DCL is the result of unique thermodynamic control which means that for each possible "template" will match one DCL (a DCL will also exist without templating). Another requirement is that equilibrating constituents of the DCL should be of similar free energy, to really favor one of them.

In 1999, Lehn[209] decrypted the concept of dynamic combinatorial chemistry (DCC) by linking it to 5 essential requirements:

1. The components must possess features for binding to the "target"

2. Reversibility is the key. Components should connect via covalent or non-covalent reversible "interactions"

3. Self-assembled structures of equilibrating constituents must be "locked" to allow analyses

4. VCL is not about screening and retrieval: it can be reduced to few components to identify a high-priority interaction

5. Last and not the least: it must be possible to qualitatively and quantitatively characterize the content of a DCL using techniques like NMR or MS-HPLC.

The DCC is a powerful method that has been developed during the last decade, and which relies on a simple principle of "adaptive behavior". Fig. 7 (For recent reviews:[210, 211]). We will now briefly discuss its possible extension to drug discovery.

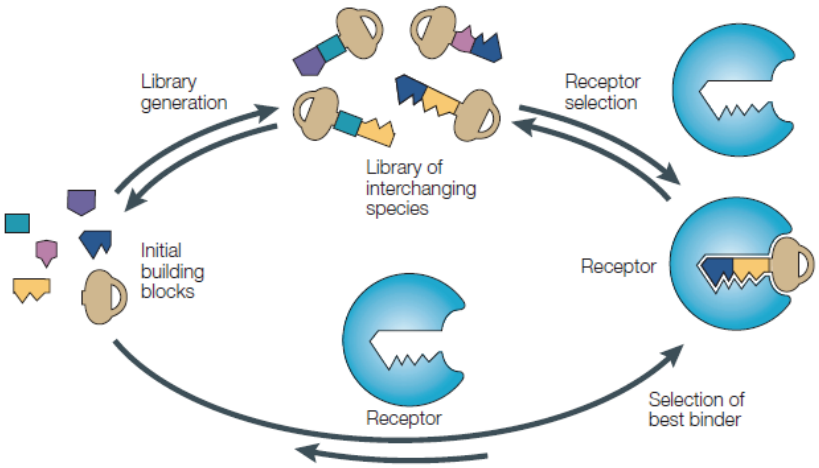


Fig. 7 General scheme of DCC. Through re-equilibration circles in the presence of a receptor, a substrate combination is favored.

iv. Drug discovery

When it comes to drug discovery the current trend is still to rely on the classic combinatorial approach. However DCC represents a very promising alternative to traditional combinatorial chemistry since it enables the identification within a dynamic chemical library of the best ligand without the need to identify and characterize each individual component of this library. Early successful applications of DCC deal with the identification of original DNA binders[212] and enzyme inhibitors (Concanavalin A[213] or Acetylcholinesterase[214]).

2. Objectives:

In this study we would like to focus on:

-demonstrating that 9AA can undergo an amine exchange reaction **in water**

-demonstrating that this reaction is **reversible and under thermodynamic control**

-understanding the possible mode of action of 9AA *in vivo*

-applying this new reversible chemistry to the concept of DCC for the selection of the best 9-amino substituent (hence the best 9AA) for a given target.

3. Strategy:

a. A reversible reaction:

We have seen that 9AA were readily hydrolyzed in water and that they could undergo an amine exchange reaction at the 9 position in organic solvent and in the presence of base.

Therefore we set out to demonstrate unequivocally that the 9-amine exchange of acridines was also working in water and that is was a reversible and thermodynamically controlled process.

This therefore required the synthesis of water soluble 9AAs.

b. Application to biological systems:

We have seen in the Chapter 1 that the anti-cancer activity of 9AAs was highly influenced by the nature of the amine attached onto the carbon C9. For instance, varying the substituent at position 9 of BRACO 19's scaffold significantly influences its affinity for G-quadruplexes. Herein, we aimed to demonstrate that BRACO 19 like drugs could undergo an amine exchange reaction under near-physiological conditions which could help understanding its mode of action in vivo.

Thus this leads to a *major* difficulty of the system: the exchange would have to take place *in water* and close to physiological conditions. It requires therefore the design of water soluble 9AAs which is rather challenging considering that 9-amino acridines can be water sensitive (9-amino acridines are readily hydrolyzed into the corresponding acridones) and are highly hydrophobic by nature.
A general scaffold of water soluble acridine is described in Fig. 8 below:

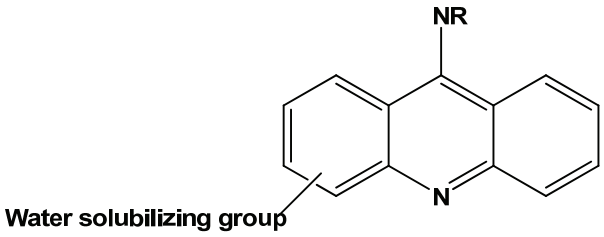


Fig. 8 Scaffold of 9AA where the water solubilizing group must be "effective" at pH 7.4 (-COOH, -NR_2)

4. Chemistry:

As a model system of water soluble 9AA we decided to synthesize a family of 4-carboxylic acid-9-amino-acridines bearing different 9-amino substituents. (See Fig. 9 below)

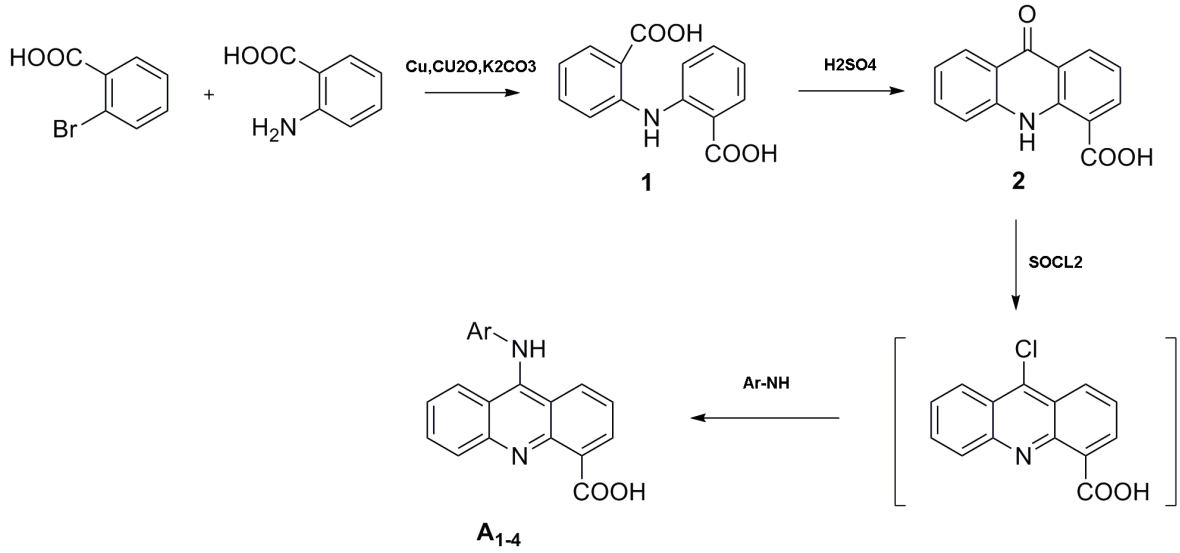


Fig. 9 Overview of the synthetic pathway leading to water soluble 9AAs.

We found that 4-carboxylic acid acridines were the simplest water soluble 9AAs one could design (bearing a water solubilizing group) for our proof-of-concept experiment. They offer the advantage of a high solubility in water at near-physiological pH and present an interesting structural analogy with drugs from the AHMA and DACA families. The general synthetic pathway towards these 4-carboxylic acid-9-aminoacridines is described below.

a. Ullman-coupling

The N-phenylanthranilic acid **1** (Fig.10) is obtained from bromobenzoic acid and anthranilic acid through a Ullman-Jourdan type reaction (see Chapter 2).
Commercially available bromobenzoic acid possesses a bromine atom that will enhance the rate of the reaction leading to higher yield than if using chlorobenzoic acid. Anilines are known to be fairly poor nucleophiles and this trend worsens when the benzene ring possesses an electron withdrawing group. However, anthranilic acid proved a much better reagent than of its methyl ester analogue methyl anthranilate (Fig. 10) (see chapter 2).
As it has been described also in Chapter 1 a good nucleophile is required for the Ullman-Jourdan reaction to occur with good yields.

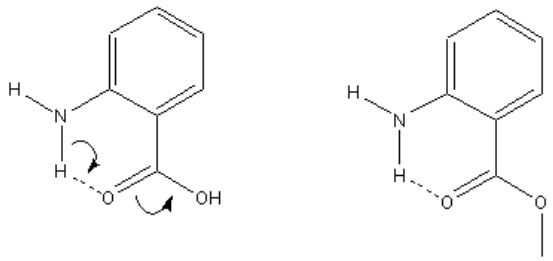


Fig. 10 Left: Anthranilic acid, Right: Methyl Anthranilate

b. *Cyclization*

As the N-phenylanthranilic acid is "somehow" a symmetric molecule the regiochemistry of the final acridone is "locked" (see Chapter 2 - Fig. 11) Cyclization was therefore performed in sulfuric acid leading quantitatively to the desired product.

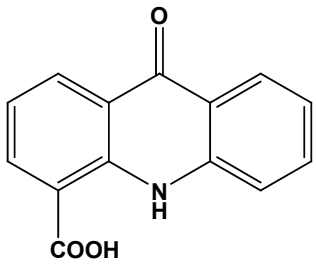


Fig.11 Acridone obtained.

c. *9AA – The 9-chloroacridine pathway*

Historically the first 9AAs were obtained from acridone via formation of a 9-phenoxy intermediate obtained by reaction of phenol onto the highly reactive 9-chloro acridine[215]. What is interesting here is that chemists used first phenols to create an intermediate prone to amination (See Fig. 12).

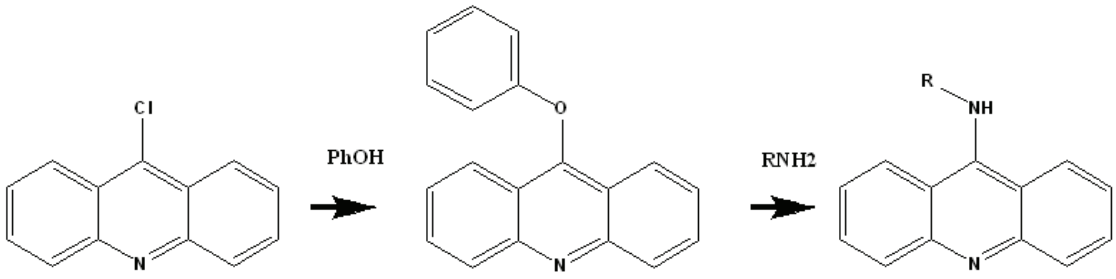


Fig. 12 Synthesis of 9-amino acridines from 9-chloroacridines via a 9-phenoxy acridine intermediate.

This highlights the first example of "nucleophilic exchange" at the 9-position of an acridine scaffold. More recently, the use of phenol was abandoned and 9AAs were readily synthesized by reaction of the appropriate amine onto the highly reactive chloroacridine[216].

The synthetic route to obtain a chloroacridine from the corresponding acridone is by refluxing it in either $POCl_3$ or $SOCl_2$ (sometimes assisted with a drop of DMF). The mechanism of this reaction is depicted on Fig. 13 below.

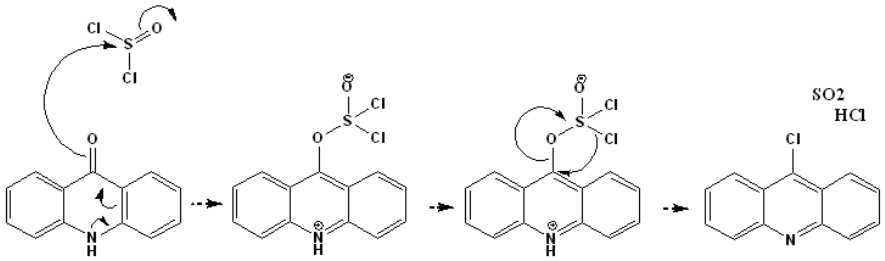


Fig. 13 Proposed mechanism for the chlorination of acridones

Alternatively, 9-chlroroacridine can be directly obtained by reaction of the appropriate N-phenylanthranilic acids with $POCl_3$.

Herein reaction of the 4-carboxylic acid acridone with thionyl chloride not only converts the starting acridone into the desired 9-chloro acridine but also reacts with the carboxylic acid to produce the corresponding acyl chloride. The issue was that both mentioned positions can easily react with amines.

After numerous attempts and assuming that the acyl chloride would react faster with any nucleophile, including water, we decided to hydrolyze it selectively under mild conditions (without touching the 9-chloro position) to generate the desired 9-chloro-4-carboxylic acid acridine.

This selective hydrolysis was carried out in a solvent mixture containing THF and ice cold water. Then the amine is added to substitute the 9 position (Fig. 14):

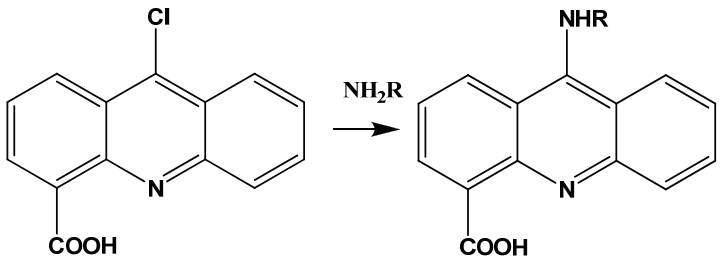


Fig. 14 Conversion of 9-chloroacridines into 9-aminoacridines

After validating the concept of reversible 9-amine exchange reaction in water with the model molecule described above, it was our ambition to demonstrate that this reaction could apply to biologically active 9AAs. Therefore, we decided to synthesize

a 3,6,9-trisubstituted acridine (BRACO 19 analogue bearing a 4-aminobenzoic acid substitutent at the 9-position). This molecule was synthesized following a protocol established by the group of S. Neidle (as described below Fig. 15). Therefore the synthesis of this 9AA will not be described in further details.

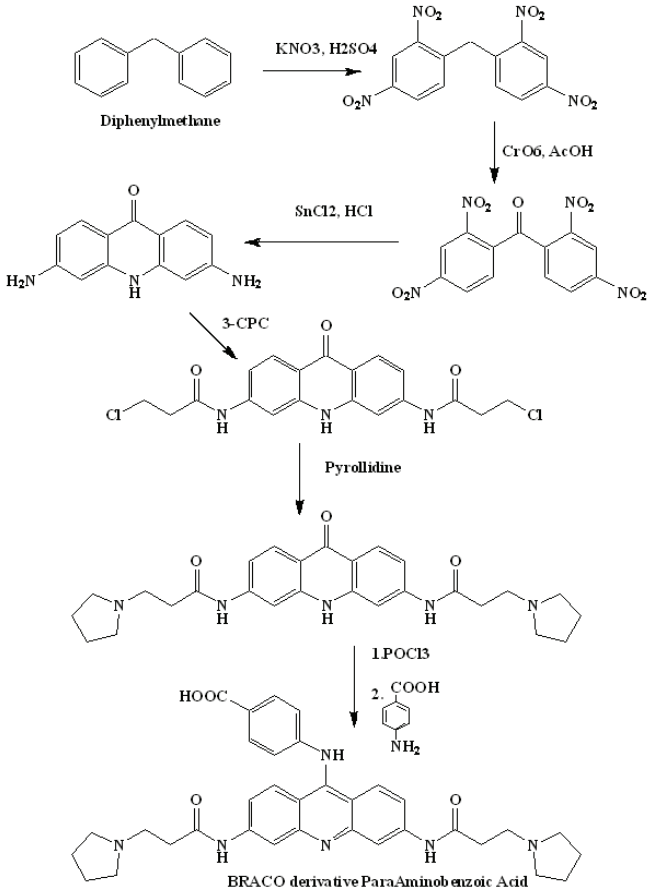


Fig. 15 Chemical pathway to BRACO 19 derivative

5. Results: The reversible thermodynamically driven reaction (RTDR):

We next investigated the ability of water soluble acridine to undergo an amine exchange reaction in water at physiological pH (7.4). The exchange reaction was monitored by LC-MS until thermodynamic equilibrium was reached. Within the scope of this study, we investigated the reactivity of differently substituted 9AAs (DACA and BRACO analogues) with aromatic and aliphatic amines. The results of this work were recently published in Organic Letters as a communication entitled "**9-Amino Acridines Undergo Reversible Amine Exchange Reactions in Water: Implications on Their Mechanism of Action in Vivo**". The main conclusions of this study are listed below.

-we demonstrated that 9AAs could undergo a **reversible** amine exchange reaction at their 9-position and that this reaction was under thermodynamic control.
-we proposed a mechanism for this exchange that proceeds via formation of an unstable hemiaminal intermediate.
-we demonstrated that the amine exchange reaction was independent of the substitution pattern of the acridine.
-we showed that this reaction was compatible with aromatic and aliphatic amines although protonated amines (e.g. most alkylamines at physiological pH) proved unreactive.
-we highlighted the existence of possible steric clashes between the 9-amino substituent and the atoms H(1) and H(8) from the acridine scaffold that could be responsible for the low stability of certain 9AAs.

6. Publication

7. Openings:

The results detailed above represent the first proof-of-concept that 9AA can undergo a reversible amine exchange reaction under near-physiological aqueous conditions. The possible applications of this discovery are multiple.
First, it suggests a new way 9AAs could be metabolized in vivo by reaction with biologically occurring amines.
Second, it represents a new reversible chemistry that is suitable for DCC and will allow the design of DCLs with increased structural diversity. Our proof-of-concept experiment demonstrates that it is possible to create dynamic combinatorial libraries of inter-converting 9AAs.

a. Selection of optimized 9AAs as G4 binders by DCC:

Considering our interest in 9AAs as G-quadruplex specific ligands, we decided to build on this proof-of-concept experiment to demonstrate whether it would be possible to select, by DCC, optimized 9AAs directed against a quadruplex DNA target. Our preliminary results using the BRACO analogue as a starting 9AA and either the intramolecular human telomeric quadruplex or the ckit21T intramolecular quadruplex as templating agents were unfortunately not successful. Briefly, two identical DCLs were equilibrated in the absence and in the presence of a quadruplex target. We were expecting that the quadruplex target would modify the equilibrium by favoring the formation of the tightest 9AA ligands (i.e. those which have the best 9-amino substitutent). However, instead of the expected shift in the position of equilibrium, we observed that the reversible amine exchange was much slower when carried out in the presence of quadruplex. This extremely slow reaction is likely to be the result of a strong interaction of the starting 9AA with the quadruplex target, thus preventing the amine exchange reaction to occur (probably because of a high steric hindrance around the 9-positon after binding to the quadruplex).

To deeply envision DCC at a biological level, there's one close step to link it with our Evolution:
"Generation of the fittest and present adaptation and evolution by spontaneous recombination under changes in the partners or in the environmental conditions" wrote Lehn in 1997 [208]. This "(supra)molecular Darwinism" could have existed in nature long time ago.
Applying DCC to the selection or optimization of G4 binders has already been achieved but this remains one of the possible applications of our system which we

are aiming to explore further. Below are three examples from the Balasubramanian's group that used DCC to target DNA quadruplexes:

-in 2004 Whitney *et al.*[217] demonstrated the first proof of concept that DCC based on reversible thiol-disulfide chemistry could be used to select original quadruplex ligands. Using a mixture of thiol-functionalized acridone **A** and peptide **P** in glutathione containing buffer they demonstrated that **A-P** and **P-P** disulfides were preferentially formed upon addition of G4 while weakly formed in the absence of target. (Fig. 16)

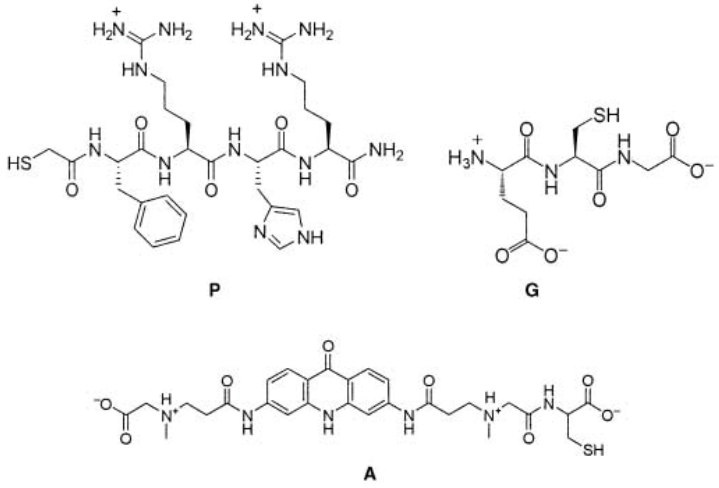


Fig. 16 Structures of molecules **A**, **P** and **G**

-In 2005 small polyamides distamycin-like molecules were selected against[218] two biological targets: an intramolecular human telomeric quadruplex and a duplex located in the promoter region of the oncogene c-kit (Fig. 17). The authors demonstrated the selection (i.e. amplification) of the **P3-P3** homodimer upon addition of either quadruplex or duplex DNA, which indeed proved to be the tightest binder of the DCL for both targets.

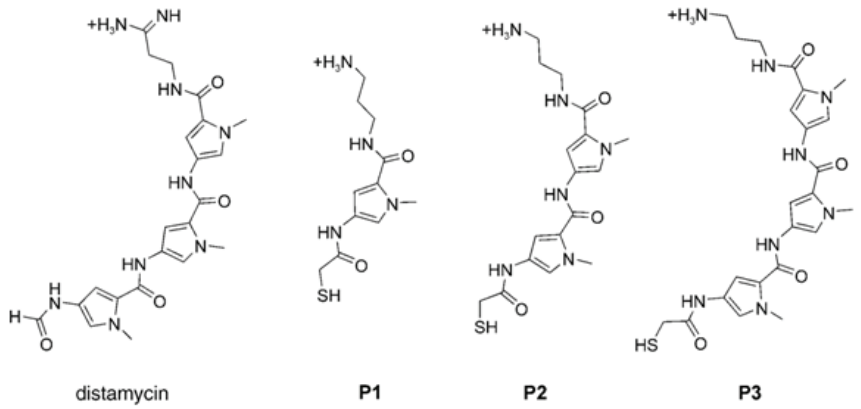


Fig. 17 Structures of thiol containing polyamide building blocks and that of distamycin

-in 2008 Bugaut *et al.*[125] demonstrated that DCC could be applied to the optimization of an existing G4 ligand. Using a thiol-functionalized analogue of a macrocyclic G4 ligand, they used DCC to screen a library of 9 peptidic building blocks based on the *para*-benzylic thiol scaffold and functionalized with either carbohydrates or amines favorable to G-quadruplexes. The main binding platform was an oxazole like ring bearing aliphatic amines for solubility and a free thiol. They showed by choosing two different quadruplexes (c-kit21 and c-myc22) not only that some unique combinations (platform-chain) were amplified upon introduction of a DNA target which were those presenting best affinity for this target but also that the templated species could vary depending on the quadruplex used as a target. This demonstrated the possible use of DCC (i) to select ligands that can discriminate between various quadruplexes and (ii) to optimize a well-known quadruplex binder (1) by appending to it the most appropriate side-chain (L1-L2). (Fig. 18)

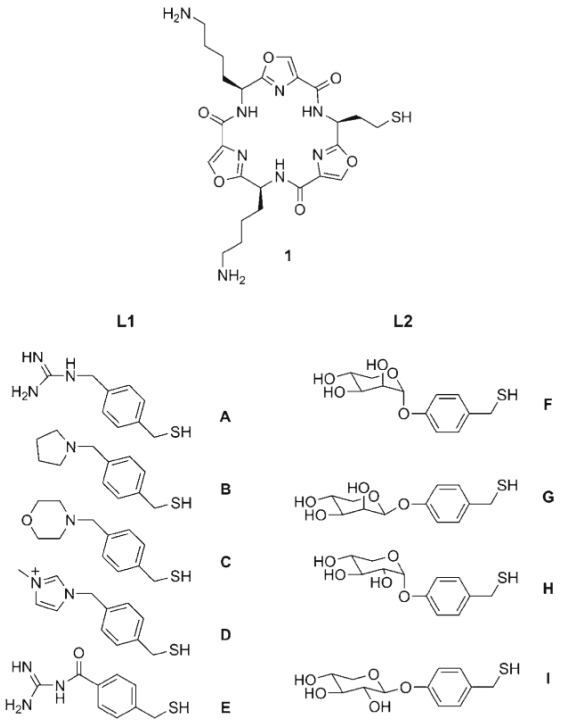


Fig. 18 Binding platform **1** and the two different libraries used here **L1** and **L2**

By attaching the thiol to the macrocycle via a long and flexible linker, the authors ensured that the thiol-disulphide exchange remained fast, even in the presence of quadruplex, which is unfortunately not the case in our system. However, the main drawback of this system is that, once one disulphide-containing ligand has been selected, this labile linker must be replaced by a more stable one before the selected compound can be used as a drug. In our system, and although the exchange conditions clearly require some optimization, the final 9AA drug is obtained directly and does not require any structural post-modification. Further work in this direction is underway in the laboratory.

b. The "thiols track"

9-thioacridines have also been developed that show promising antimicrobial and antibacterial activity. In order to expand on our previous findings, we synthesized a series of 9-thioacridines and demonstrated that they could also undergo a reversible thiol exchange reaction in near-physiological aqueous reaction. Interestingly, the 9-thiol exchange reaction proved significantly faster that the corresponding 9-amine exchange reaction. Also interesting is the fact that 9-thioacridines can be converted

into 9-amino acridines and vice versa (Fig. 19). The design of more complex libraries containing interconverting 9-amino and 9-thio acridines is also underway in the laboratory. Due to the faster kinetics observed for 9-thiol exchange reactions in solution, one may anticipate that the addition of a quadruplex target into the equilibrating mixture will not slow down the reaction too much.

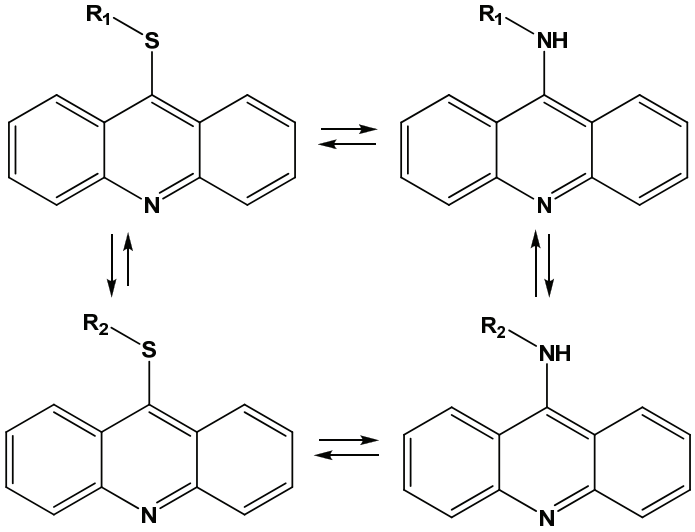


Fig. 19 Possible thiol-amine exchange on a 9 substituted acridine

Conclusion

G-quadruplexes have recently been established as therapeutic targets of choice for the design of ligands with possible anti-cancer activity. Herein, we have been successful in developing three original approaches to target DNA and RNA G4 in a structure and/or specific sequence way.

In a first part, we developed a new family of ligands based on a nucleic acid peptide analogue scaffold (PNA). We demonstrated that such G-rich PNAs could associate with DNA strands that contain only three clusters of three consecutive guanines to form PNA: DNA hybrid (3+1) bimolecular quadruplexes. This original approach is of potential interesting for targeting G-rich sequences in the genome which would normally not be able to adopt of quadruplex conformation because lacking a fourth G-cluster. Therefore our ligands could serve to introduce quadruplexes in the genome where they naturally do not exist, thus creating fully artificial new regulatory elements.

In a second part, we provided proof-of-concept that synthetic small molecules can indeed recognize RNA quadruplexes in a highly specific "sequence+structure" manner by targeting the RNA quadruplex structure in a structure specific way but also its two single-stranded flanking arms in a sequence specific manner. We showed that such ligands (of general structure PNA_1-acridone-PNA_2) are capable to discriminate between two identical RNA quadruplexes that differ only by the sequence of the flanking arms upstream and downstream of the quadruplex. Although this project is still at an early stage, this proof-of-concept is highly promising and complementary experiments will hopefully allow demonstrating the efficiency of those ligands to regulate RNA translation in vitro and maybe also in vivo.

Finally, we highlighted a still unknown chemical property of 9-substituted acridines by demonstrating (i) that 9-aminoacridines (and also 9-thioacridines) can undergo an amine (or thiol) exchange reaction under aqueous and near physiological conditions of salt and pH and (ii) that this reaction is reversible and under thermodynamic control. The discovery of this new property of an old molecule has implications in understanding the mode of action of these molecules in vivo (e.g. via in situ trans-amination reactions) and has therefore implication in the future design of 9-substituted acridine therapeutics.

Hopefully the three projects that have been developed within this manuscript will contribute to a better understanding of how one can target G-quadruplex structures in a highly specific manner, regardless of their polymorphic nature and regardless of the number of quadruplexes that do form simultaneously in the whole genome.

Final words

As far I'm concerned I would never have hoped to be involved in such an interesting project.
Of course having a quick look at what was the initial subject and what it became this demonstrated me the mystical part of researches. The laboratory work combined with the start of the "junior lab" structure was for me the perfect continuation of the engineer / master degree. I started later than expected but during those 3 years a considerable amount of work was achieved and I can only be proud of it. Naturally without the permanent advices of a young open minded supervisor this wouldn't have been possible.

Therefore my greatest consideration goes to the man in ISIS who shared those 3 years with me, and also brought me to this unique healthy lab environment, that I liked while working with him, laughing with him, exchanging ideas with him and sharing more than just his leadership.

I, of course, wish all the best to the ones who will continue my projects. I am actually convinced that many publications will follow ☺ and that I have left many ways to discover new results.

I can encourage them to follow the actual trend about RNA quadruplexes and also to pay attention to the loops.

One of the greatest challenges would be to combine the 2 first chapters and construct a 3 dimensional PNA that will provide 3+1 G4 with recognition of flanking regions!

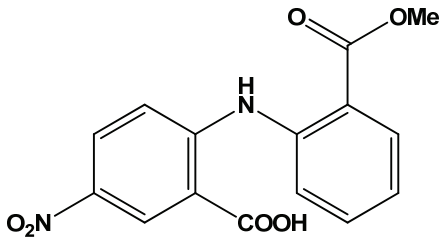
I would like to thank all the people who came to the defense: John, Simon, Capitaine Pitch, Artur, all my co-workers and family!!! I also have to thank the examiners for having been quite nice.

Experimental part

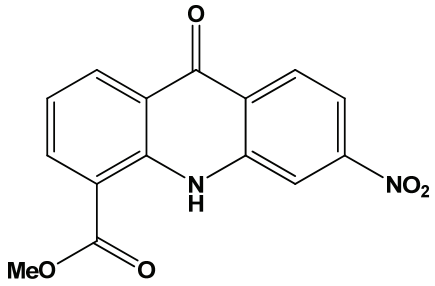
Solvents were of HPLC or reagent quality and purchased commercially. Starting materials were purchased commercially and used without further purification. Compounds were characterized using ^{1}H and ^{13}C NMR that were recorded on a Bruker Avance DRX 400 spectrometer at 400 and 100.6 MHz, respectively. Chemical shifts are reported with reference to the residual solvent peaks.

Chapter 2 : the 3+1 project

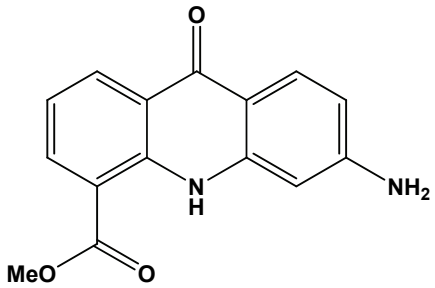
2-[[2-(methoxycarbonyl)phenyl]amino]-para-nitrobenzoic acid:
A mixture of 2-chloro-para-nitrobenzoic acid (5 g, 25 mmol), methyl anthranilate (5 g, 33 mmol),Cu powder (190mg, 3 mmol), Cu_2O (140mg, 1 mmol), $Cu(OAc)_2$ (360mg, 2 mmol) was stirred at 160 °C in DMF (120mL) for 3 hours and then allowed to slowly cool to room temperature. After filtration through a thin layer of silica between two layers of celite, addition of aqueous 0.1M HCl (120 mL) and vigorous stirring led to the formation of an orange precipitate. The desired product was obtained as an orange solid (4g) by filtration and was used for the subsequent step without further purification. Yield = 50%.
^{1}H-NMR (DMSO, 400 MHz): 8.15 (2H, *br* m), 7.98 (1H, d, J = 7.6 Hz), 7.67 (1H, d, J = 7.8 Hz), 7.66 (1H, s), 7.60 (1H, t, J = 7.8 Hz), 7.17 (1H, t, J = 7.6 Hz), 3.88 (3H, s).



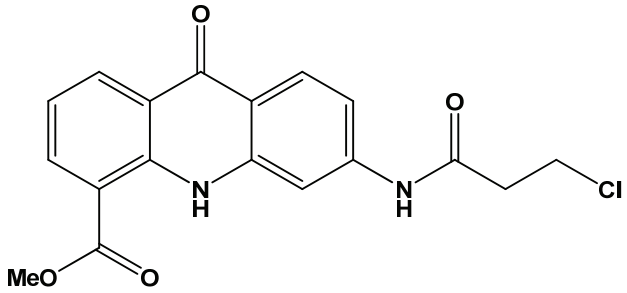
Methyl-3-nitro-acridone-5-carboxylate (2):
A mixture of 2-[[2-(methoxycarbonyl)phenyl]amino]-*p*-nitrobenzoic acid (1g, 3.2 mmol) in polyphosphoric acid (10g) is heated at 130 °C for 45 min. After cooling to room temperature, water was added leading to the formation of a bright yellow solid. The solid was finally filtered off, washed with water and dried under vacuum. The acridone was obtained pure as a yellow solid (0.8g). Yield = 85%.
^{1}H-NMR (DMSO, 400 MHz): 9.02 (1H, s), 8.55 (1H, d, J = 7.2 Hz), 8.47 (1H, d, J = 7.2 Hz), 8.41 (1H, d, J = 8.4 Hz), 8.02 (1H, d, J = 8.4 Hz), 7.44 (1H, t, J = 7.2 Hz), 4.03 (3H, s).



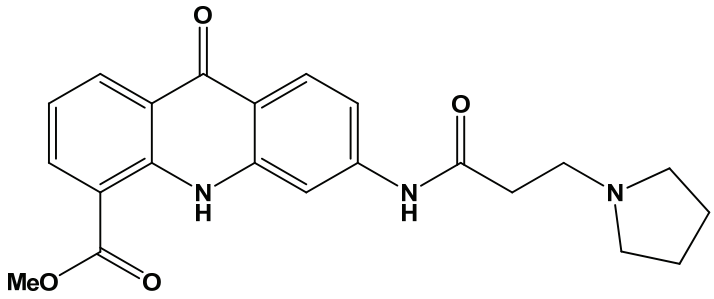
Methyl-3-amino-acridone-5-carboxylate (3):
A mixture of compound **2** (1g, 3.4 mmol), tin(II) chloride (3g, 13.3 mmol) and concentrated HCl (5 mL) was stirred vigorously for 20 min in 20 mL glacial AcOH. The mixture was heated at 80°C for 40 min then cooled to room temperature before water (80 mL) was added to allow the amino acridone to precipitate. After filtration and subsequent washes with AcOEt and diethyl ether, the desired product was obtained pure as a yellow powder (0.6g).Yield = 66%.
^{1}H-NMR (DMSO, 400 MHz): 8.46 (1H, dd, ^{2}J = 8 Hz, ^{3}J = 1.8 Hz), 8.33 (1H, dd, ^{2}J = 8 Hz, ^{3}J = 1.8 Hz), 7.93 (1H, d, J = 8 Hz), 7.28 (1H, t, J = 8 Hz), 6.62 (1H, dd, ^{2}J = 8 Hz, ^{3}J = 1.8 Hz), 6.54 (1H, d, J = 1.8 Hz), 3.98 (3H, s).



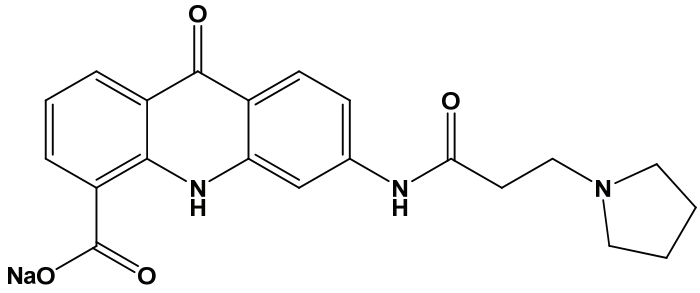
Methyl-3-(3-chloropropionamide)-acridone-5-carboxylate (4):
A suspension of **3** (0.5g, 1.9 mmol) in 3-chloropropionyl chloride (20 mL) was heated at 60°C for 2 days. Product precipitates at room temperature by addition of diethylether and was then filtered off. The resulting dark yellow powder was then washed with AcOEt and diethylether, thus affording the desired product pure as a pale brown solid (0.6 g).Yield = 85%.
^{1}H-NMR (DMSO, 400 MHz): 7.61 (1H, d, J = 8.4 Hz), 7.47 (1H, d, J = 8 Hz), 7.32 (3H, s), 7.26 (1H, d, J = 8.4 Hz), 6.44 (1H, t, J = 8 Hz), 6.40 (1H, d, J = 8 Hz), 3.08 (3H, s), 3.02 (2H, t, J = 6 Hz), 2.03 (2H, t, J = 6 Hz).
^{13}C-NMR (DMSO, 101 MHz): 176, 169.5, 168, 144.5, 141.4, 141.3, 136.5, 133, 127.5, 122, 120.5, 117, 115.2, 114.8, 106.4, 53.2, 41, 24.



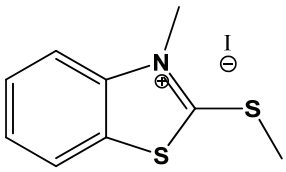
Methyl-3-(3-(pyrrolidin-1-yl)propanamido)-acridone-5-carboxylate (5):
To a solution of **4** (0.5g, 1.4 mmol) in 2.5 ml DMF was added pyrrolidine (500µL, 5.6 mmol). The solution was then stirred at for 30 min until the reaction was complete. After cooling to room temperature, a mixture of diethyl ether containing a trace of methanol was added resulting in the precipitation of a yellow solid. The desired product was obtained pure as a yellow solid (0.4g) by filtration and subsequent washes with AcOEt and diethyl ether. Yield = 73%.
^{1}H-NMR (DMSO, 400 MHz): 8.54 (1H, d, J = 7.8 Hz), 8.42 (1H, d, J = 7.8 Hz), 8.20 (1H, s), 8.20 (1H, d, J = 7.8Hz), 7.39 (1H, t, J = 7.8 Hz), 7.38 (1H, d, J = 7.8 Hz), 4.00 (3H, s), 3.47 (2H, t, J = 6 Hz), 1.85 (2H, t, J = 6 Hz), 3.10 (4H, *br* m), 1.97 (4H, *br* m).
^{13}C-NMR (DMSO, 101 MHz): 176, 169.5, 168, 144.5, 141.4, 141.3, 136.5, 133, 127.5, 122, 120.5, 117, 115.2, 114.8, 106, (53.5, 2C), 53.2, 50, 33, (23, 2C).



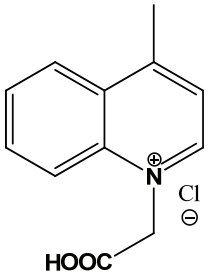
Sodium-3-(3-(pyrrolidin-1-yl)propanamido)-acridone-5-carboxylate (6):
A mixture of **5** (0.25g, 0.64 mmol) and NaOH (50mg, 1.2 mmol) was stirred at 80°C in 20 ml DMF/H2O (1:9) for 40 min. Solvents are the evaporated off and consecutive additions of methanol and ether allow the product to precipitate. The desired acridone is obtained pure as a light yellow solid (0.2g) by filtration and multiple washes with diethyl ether. Yield = 78%.
^{1}H-NMR (D_2O, 400 MHz): 7.88 (1H, d, J = 7.2 Hz), 7.72 (1H, d, J = 7.2 Hz), 7.18 (1H, d, J = 8.8 Hz), 6.85 (1H, t, J = 7.2 Hz), 6.51 (3H, s), 6.17 (1H, d, J = 8.8 Hz), 2.64 (2H, t, J = 7.2 Hz), 2.55 (4H, *br* m), 2.20 (2H, t, J = 7.2 Hz), 1.73 (4H, *br* m).
^{13}C-NMR (D_2O, 133 MHz): 177, 173, 171, 141.5, 140, 139, 136, 129, 126, 120.5, 119.9, 119.8, 115, 113.5, 104, (53.5, 2C), 50, 35, (23, 2C).



3-methyl-2-(methylthio)benzothiazol-3-ium iodide salt:
3-methyl-2-(methylthio)benzothiazol-3-ium was prepared according to the procedure described in: "Cyanine dye conjugates as probes for live cell imaging" Bioorg. Med. Chem. Lett., 2007, 17, pp 5182–5185



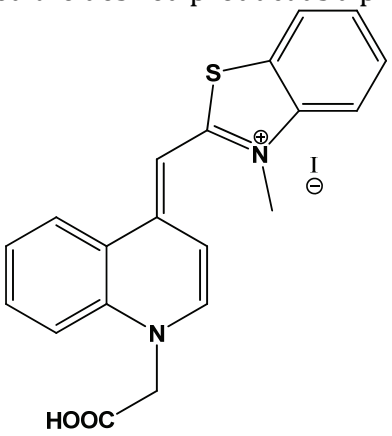
1-(carboxymethyl)-4-methylquinolinium chloride salt:
1-(carboxymethyl)-4-methylquinolinium was prepared according to the procedure described in: Bethge, L., D. V. Jarikote, et al. (2008). "New cyanine dyes as base surrogates in PNA: forced intercalation probes (FIT-probes) for homogeneous SNP detection." Bioorg Med Chem 16(1): 114-25.



Unfortunately the described procedure did not yield any pure product and a mixture was obtained between desired product and bi-substituted lepidine which was however used for the subsequent step without further purification.

Carboxymethylated thiazole orange:

3-methyl-2-(methylthio)benzothiazol-3-ium (145mg,0.45mmol) and previous mixture (140mg) were mixed in the dark in 7.5 ml anhydrous dichloromethane for 16 hours in the presence of 0,1g of triethylamine. The solvent was removed and the solid was suspended in 15 ml anhydrous methanol for 15 min. Upon filtration we obtained the desired product as a pink-red solid (80mg). Yield = 36%.

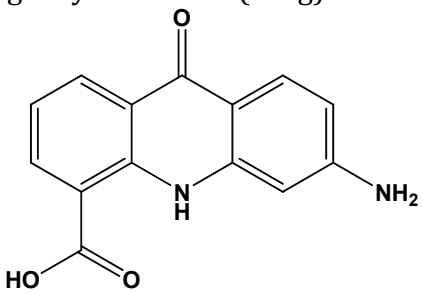


1H-NMR (DMSO/TFA, 400 MHz): 8.75 (1H, d, J = 8.3Hz) 8.49 (1H, d, J = 7.4Hz), 8.03 (1H, d, J = 7.7Hz), 7.92 (1H, t, J = 8.5Hz), 7.83 (1H, d, J = 8.5Hz), 7.77 (1H, d, J = 8.3Hz), 7.69 (1H, t, J = 7.7Hz), 7.60 (1H, t, J = 7.5Hz), 7.42 (1H, t, J = 7.5Hz), 7.31 (1H, d, J = 7.4Hz), 6.90 (1H, s), 5.44 (2H, s),.4.01 (3H, s).

Chapter 3 : The PNA RNA story

3-amino-acridone-5-carboxylic acid:

Methyl 3-amino-acridone-5-carboxylate (1 g, 3.7 mmol) was dissolved in 20 mL of a 1:1 water:sulfuric acid blend and refluxed at 130°C for 2 hours. The mixture was then cooled and poured into cold water (200 mL) which allows slowly the desired product to precipitate as a gold yellow solid (0.5 g). Yield = 53%.



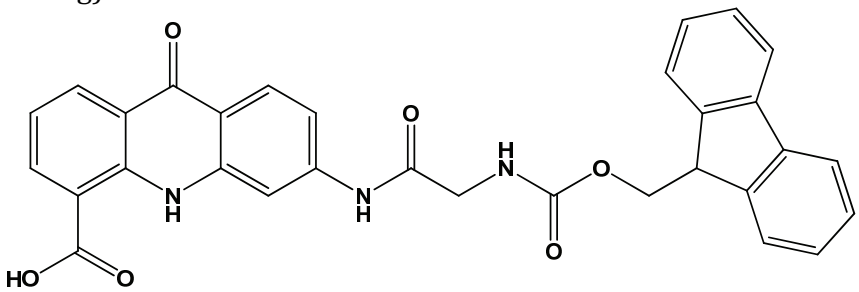
^{1}H-NMR (DMSO, 400 MHz): 8.44 (1H, dd, ^{2}J = 8 Hz, ^{3}J = 1.6 Hz), 8.33 (1H, dd, ^{2}J = 8 Hz, ^{3}J = 1.6 Hz), 7.93 (1H, d, J = 8 Hz), 7.26 (1H, t, J = 8 Hz), 6.62 (1H, dd, ^{2}J = 8 Hz, ^{3}J = 1.6 Hz), 6.50 (1H, d, ^{3}J = 1.6 Hz).

Fmoc-Gly-Cl

Fmoc-Gly-Cl was prepared according to the procedure described in: "Rapid continuous peptide synthesis via FMOC amino acid chloride coupling and 4-(aminomethyl)piperidine deblocking" *J. Org. Chem.*, **1990**, *55*, pp 721–728

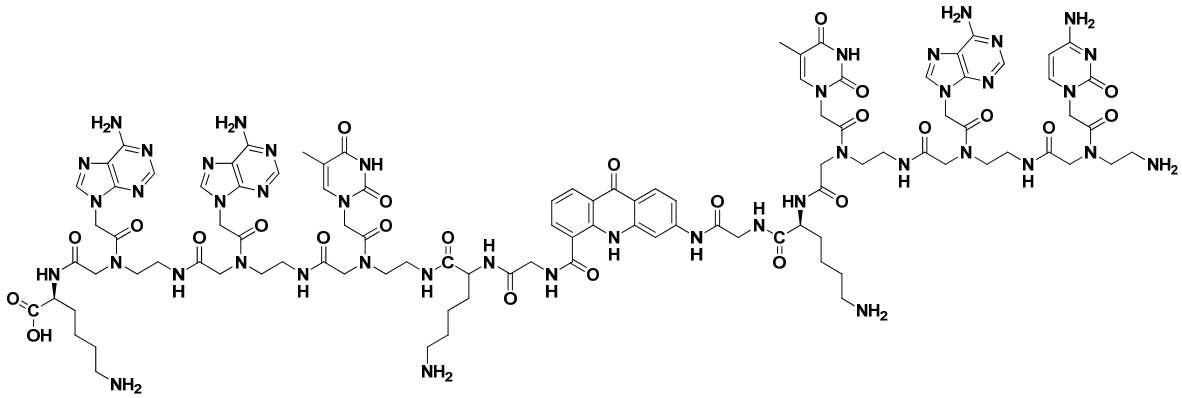
3-(2-(((9H-fluoren-9-yl)methoxy)carbonylamino)acetamido)-acridone-5-carboxylic acid:

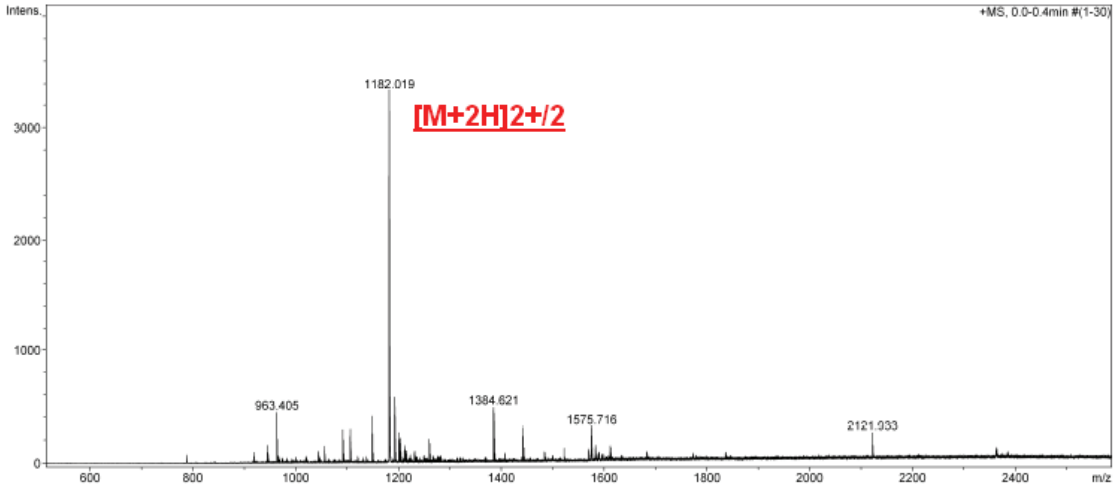
3-amino-acridone-5-carboxylic acid (200 mg, 0.79 mmol) was solubilized into 2.5 mL anhydrous DMF containing 65µL (0.8 mmol) anhydrous pyridine. Then 450 mg (1.4 mmol) of Fmoc-Gly-Cl was added to the solution and after 15 min of stirring the mixture was poured into 200 mL acidic water. The yellow precipitate was filtered off and washed with AcOEt and then diethyl ether. The crude product was refluxed in acidic water for 30 min and hot filtered off to allow the obtention of a clear yellow powder (300 mg).Yield = 71%.



^{1}H-NMR (DMSO, 400 MHz): 9,64 (1H, s), 7.65 (1H, d, J = 8 Hz), 7.57 (1H, d, J = 8 Hz), 7.33 (1H, d, J = 8 Hz), 7.30 (1H, s), 7.06 (2H, d, J = 7.6 Hz), 6.91 (2H, d, J = 7.6 Hz), 6.60 (2H, t, J = 7.6 Hz), 6.52 (2H, t, J = 7.6 Hz), 6.50 (2H, d, J = 8 Hz), 6.49 (2H, t, J = 8 Hz), 3.49 (2H, d, J = 6.8 Hz), 3.41 (1H, t, J = 6.8 Hz), 3.04 (1H, d, J = 5.6 Hz).
^{13}C-NMR (DMSO, 133 MHz): 176, 169.8, 169.5, 157.1, 144.4 (2C), 144.3, 141.9, 141.3, 141.2 (2C), 137, 133, 128.1 (2C), 127.7, 127.5 (2C), 125.7 (2C), 122.2, 120.7, 120.6 (2C), 117, 115.2, 115, 106.1, 66.2, 47.2, 44.7.

ESI-MS HR: synthesized for on a solid support using Fmoc chemistry. The PNA strand was cleaved from the resin with a 95% TFA solution and purified by HPLC. The final compounds were characterized by ESI-HRMS m/z: calcd for $C_{101}H_{132}N_{44}O_{25}$:2362.41; found: 1182.019 $[M+2H]^{2+}/2$



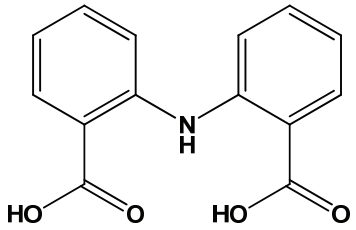

Intens.
+MS, 0.0-0.4min #(1-30)
1182.019
[M+2H]2+/2
963.405
1384.621
1575.716
2121.033
3000
2000
1000
0
600
800
1000
1200
1400
1600
1800
2000
2200
2400
m/z

Chapter 4 : 9AA and their amine exchange

2,2'-azanediyldibenzoic acid

A mixture of 2-bromobenzoic acid (4 g, 20 mmol), anthranilic acid (3.42 g, 25 mmol), anhydrous potassium carbonate (6.17 g, 45 mmol), Cu powder (127 mg, 2 mmol) and Cu_2O (143 mg, 1 mmol) in anhydrous DMF (10 mL) was heated at 130°C for 24 h. The DMF was then evaporated off under reduced pressure and water (50 mL) was added. The mixture was acidified with a 1M aqueous HCl solution until it reached pH 4, thus leading to the formation of a light green precipitate. The desired compound **1** was finally filtered off and used for the subsequent step without further purification (4.6 g, 90%).

^{1}H NMR (400 MHz, DMSO) δ 7.90 (2H, d, *J* = 7.8 Hz), 7.46(2H, d, *J* = 7.8 Hz), 7.42 (2H, t, *J* = 7.8 Hz), 6.95 (2H, t, *J* = 7.8 Hz).

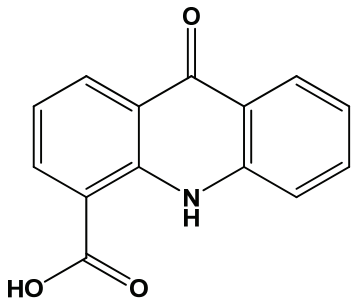


9-oxo-9,10-dihydroacridine-4-carboxylic acid

A mixture of 2,2'-azanediyldibenzoic acid **1** (1 g, 4 mmol) in polyphosphoric acid (10 g) was stirred at 130°C for 30 min. The reaction mixture was then cooled to room temperature and water (100 mL) was added. After triturating the residue in water, a bright yellow solid formed which was then filtered off, washed extensively with water and dried under vacuum. The desired acridone **2** was thus obtained pure as a bright yellow solid (910mg g, 95%).

^{1}H NMR (400 MHz, DMSO) δ 8.52 (1H, d, *J* = 7.8 Hz), 8.42 (1H, d, *J* = 7.8 Hz), 8.32 (1H, d, *J* = 7.8 Hz), 7.75 (2H, m), 7.28 (2H, m).

^{13}C NMR (100 MHz, DMSO) δ 176, 169, 141, 139, 137, 134, 132, 125.5, 121.5, 121, 120, 119.5, 118, 114.5

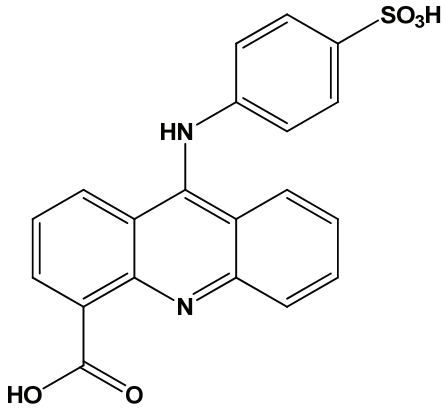


General procedure for synthesis of 9-amino-acridine-4-carboxylic acid

A solution of acridone **2** (200 mg, 0.83 mmol) was heated at reflux in freshly distilled thionyl chloride (3 mL) under Ar atmosphere for 10-15 min. $SOCl_2$ was then removed under reduced pressure and the residual solid was suspended in anhydrous THF (10 mL). The suspension was vigorously stirred at 0°C and 1 mL of a 0.1M aqueous NaOH solution was added. The mixture was stirred for a further 10 min and a solution of the desired amine (1.05 equivalents) in anhydrous THF (1 mL) was added. The mixture was then allowed to return to room temperature and water (4 mL) was added. The resulting precipitate was finally collected by filtration, washed with water and ether and dried under vacuum, affording the desired 9-amino acridine as an orange amorphous solid. Yield: 60-80%.

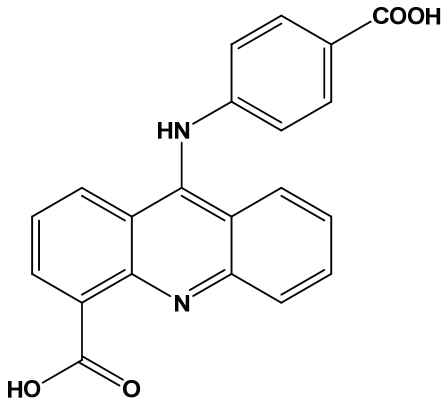
9-(4-sulfophenylamino)acridine-4-carboxylic acid (A_1)

^{1}H NMR (400 MHz, D2O) δ 8.33 (dd, J = 7.2, 0.9 Hz, 1H), 7.83 (t, J = 7.4 Hz, 1H), 7.82 (d, J= 8.6 Hz, 2H), 7.77 (d, J = 7.8 Hz, 1H), 7.56 (d, J = 8.7 Hz, 1H), 7.51 (d, J = 8.5 Hz, 1H), 7.28 (m, 2H), 7.21 (d, J = 8.6 Hz, 2H).
^{13}C NMR (101 MHz, D2O) δ 171.25, 154.15, 142.79, 141.08, 139.16, 138.44, 138.06, 136.15, 128.12, 127.40, 125.13, 124.65, 123.95, 123.74, 121.92, 119.69, 113.54, 112.95.
HRMS (negative electrospray) m/z calcd for $C_{20}H_{13}N_2O_5S$ 393.0540 found 393.0485



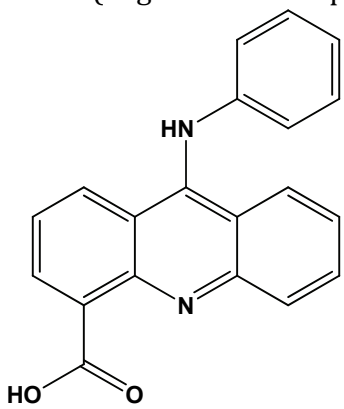
9-(4-carboxyphenylamino)acridine-4-carboxylic acid (A_2)

^{1}H NMR (400 MHz, D2O) δ 8.19 (1H, d, J = 7.6 Hz), 7.70 (2H, d, J = 7.2 Hz), 7.65 (1H, d, J= 7.6 Hz), 7.64 (1H, t, J = 7.6 Hz), 7.42 (1H, d, J = 7.6 Hz), 7.36 (1H, t, J = 7.6 Hz), 7.09 (1H, t, J = 7.6 Hz), 7.07 (1H, t, J = 7.6 Hz), 6.96 (2H, d, J = 7.2 Hz).
^{13}C NMR (101 MHz, D2O) δ 174.70, 172.06, 160.24, 154.42, 139.49, 138.33, 137.63, 135.19, 134.19, 130.63, 128.79, 125.13, 124.07, 122.61, 121.66, 119.21, 114.53, 113.63.
HRMS (negative electrospray) m/z calcd for $C_{21}H_{13}N_2O_4$ 357.0870 found 357.0851



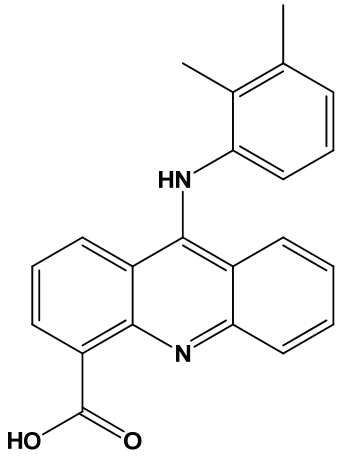
9-(phenylamino)acridine-4-carboxylic acid (A_3)

^{1}H NMR (400 MHz, MeOD) δ 8.74 (dd, J = 7.4, 1.3 Hz, 1H), 8.44 (dd, J = 8.7, 1.2 Hz, 1H), 8.13 (d, J = 8.8 Hz, 1H), 8.10 – 7.73 (m, 2H), 7.62 – 7.16 (m, 7H).
^{13}C NMR (101 MHz, MeOD) δ 169.08, 156.78, 140.30, 140.26, 139.59, 139.22, 136.13, 131.24, 130.22, 128.35, 125.50, 124.75, 124.59, 122.61, 119.81, 116.97, 114.20, 113.60.
HRMS (negative electrospray) m/z calcd for $C_{20}H_{13}N_2O_2$ 313.0972 found 313.0956



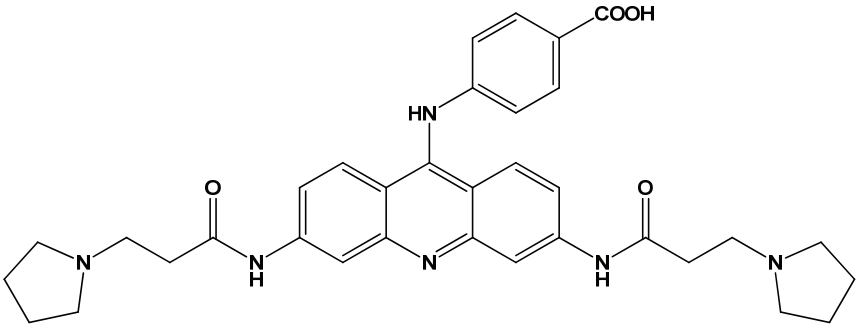
9-(2,3-dimethylphenylamino)acridine-4-carboxylic acid (A_4)

^{1}H NMR (400 MHz, MeOD) δ 8.80 (d, *J* = 6.9 Hz, 3H), 8.39 (d, *J* = 8.8 Hz, 1H), 8.06 (d, *J* = 8.8 Hz, 1H), 8.03 (d, *J* = 3.8 Hz, 1H), 7.54 – 7.48 (m, 1H), 7.48 – 7.41 (m, 2H), 7.33 (t, *J* = 7.7 Hz, 1H), 7.25 (d, *J* = 7.7 Hz, 1H), 2.46 (s, 3H), 2.26 (s, 3H).
^{13}C NMR (101 MHz, MeOD) δ 169.27, 157.06, 139.71, 139.44, 138.98, 138.49, 135.91, 132.96, 130.56, 130.40, 127.27, 124.96, 124.52, 123.63, 122.62, 119.70, 113.68, 112.98, 19.02, 13.06.
HRMS (negative electrospray) *m/z* calcd for $C_{22}H_{17}N_2O_2$ 341.1285 found 341.1267



Synthesis of BRACO 19-derivative:

4-(3,6-bis(3-(pyrrolidin-1-yl)propanamido)-9,10-dihydroacridin-9-ylamino) benzoic acid was prepared accordingly to the procedure described in: "Therapeutic acridone and acridine compounds" Patent n°WO0208193 (A2), 2002



^{1}H NMR (400 MHz, D2O) δ 8.19 (s, 2H), 7.99 (d, J = 8.4 Hz, 2H), 7.85 (d, J = 9.3 Hz, 2H), 7.23 (d, J = 8.4 Hz, 2H), 7.19 (d, J = 9.7 Hz, 2H), 3.76 (bm, 4H), 3.64 (t, J = 6.5 Hz, 4H), 3.21 (bm, 4H), 3.08 (t, J = 6.0 Hz, 4H), 2.21 (bm, 4H), 2.04 (bm, 4H).
^{13}C NMR (101 MHz, D2O) δ 170.73, 169.91, 162.74, 152.65, 145.16, 143.53, 141.39, 131.35, 127.77, 126.60, 122.52, 117.90, 117.78, 114.88, 111.25, 105.78, 54.47, 50.21, 32.33, 22.66.

The end...

References:

[1] B. a. Gait, *Nucleic Acids in Chemistry and Biology*, Oxford University Press, p.
[2] N. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry, 4th edition*, **2004**, p.
[3] S. Neidle and S. Balasubramanian, *Quadruplex Nucleic Acids*, RSC publishing, Cambridge, **2006**, p.
[4] J. D. Watson and F. H. Crick, *Nature* **1953**, *171*, 737-738.
[5] J. D. Watson and F. H. Crick, *Nature* **1953**, *171*, 964-967.
[6] M. Gellert, M. N. Lipsett and D. R. Davies, *Proc Natl Acad Sci U S A* **1962**, *48*, 2013-2018.
[7] E. Henderson, C. C. Hardin, S. K. Walk, I. Tinoco, Jr. and E. H. Blackburn, *Cell* **1987**, *51*, 899-908.
[8] J. R. Williamson, M. K. Raghuraman and T. R. Cech, *Cell* **1989**, *59*, 871-880.
[9] M. Frank-Kamenetskii, *Nature* **1989**, *342*, 737.
[10] S. L. Forman, Fettinger, Pieraccini, Gottarelli and Davis, *J Am Chem Soc* **2000**, *122*, 8.
[11] D. Sen and W. Gilbert, *Nature* **1990**, *344*, 410-414.
[12] A. M. Zahler, J. R. Williamson, T. R. Cech and D. M. Prescott, *Nature* **1991**, *350*, 718-720.
[13] C. Schaffitzel, I. Berger, J. Postberg, J. Hanes, H. J. Lipps and A. Pluckthun, *Proc Natl Acad Sci U S A* **2001**, *98*, 8572-8577.
[14] H. J. Lipps and D. Rhodes, *Trends Cell Biol* **2009**, *19*, 414-422.
[15] R. H. Shafer and I. Smirnov, *Biopolymers* **2000**, *56*, 209-227.
[16] H. Arthanari and P. H. Bolton, *Chem Biol* **2001**, *8*, 221-230.
[17] K. Muniyappa, S. Anuradha and B. Byers, *Mol Cell Biol* **2000**, *20*, 1361-1369.
[18] P. Mani, V. K. Yadav, S. K. Das and S. Chowdhury, *PLoS One* **2009**, *4*, e4399.
[19] S. Kumari, A. Bugaut, J. L. Huppert and S. Balasubramanian, *Nat Chem Biol* **2007**, *3*, 218-221.
[20] J. L. Huppert, A. Bugaut, S. Kumari and S. Balasubramanian, *Nucleic Acids Res* **2008**, *36*, 6260-6268.
[21] A. Siddiqui-Jain, C. L. Grand, D. J. Bearss and L. H. Hurley, *Proc Natl Acad Sci U S A* **2002**, *99*, 11593-11598.
[22] K. Halder, V. Mathur, D. Chugh, A. Verma and S. Chowdhury, *Biochem Biophys Res Commun* **2005**, *327*, 49-56.
[23] K. Halder and S. Chowdhury, *Nucleic Acids Res* **2005**, *33*, 4466-4474.
[24] V. Gonzalez and L. H. Hurley, *Annu Rev Pharmacol Toxicol* **2009**.
[25] H. Fernando, A. P. Reszka, J. Huppert, S. Ladame, S. Rankin, A. R. Venkitaraman, S. Neidle and S. Balasubramanian, *Biochemistry* **2006**, *45*, 7854-7860.
[26] M. Bejugam, S. Sewitz, P. S. Shirude, R. Rodriguez, R. Shahid and S. Balasubramanian, *J Am Chem Soc* **2007**, *129*, 12926-12927.
[27] J. L. Huppert and S. Balasubramanian, *Nucleic Acids Res* **2005**, *33*, 2908-2916.
[28] J. Eddy and N. Maizels, *Nucleic Acids Res* **2008**, *36*, 1321-1333.
[29] J. L. Huppert and S. Balasubramanian, *Nucleic Acids Res* **2007**, *35*, 406-413.
[30] J. L. Huppert, *Biochimie* **2008**, *90*, 1140-1148.

[31] M. Webba da Silva, *Chemistry* **2007**, *13*, 9738-9745.
[32] G. N. Parkinson, M. P. Lee and S. Neidle, *Nature* **2002**, *417*, 876-880.
[33] A. Bugaut and S. Balasubramanian, *Biochemistry* **2008**, *47*, 689-697.
[34] W. Li, P. Wu, T. Ohmichi and N. Sugimoto, *FEBS Lett* **2002**, *526*, 77-81.
[35] A. Risitano and K. R. Fox, *Biochemistry* **2003**, *42*, 6507-6513.
[36] Y. Wang and D. J. Patel, *Structure* **1993**, *1*, 263-282.
[37] J. L. Mergny, A. T. Phan and L. Lacroix, *FEBS Lett* **1998**, *435*, 74-78.
[38] Y. Xu, Y. Noguchi and H. Sugiyama, *Bioorg Med Chem* **2006**, *14*, 5584-5591.
[39] P. A. Rachwal, T. Brown and K. R. Fox, *Biochemistry* **2007**, *46*, 3036-3044.
[40] M. Webba da Silva, *Methods* **2007**, *43*, 264-277.
[41] H. Martadinata and A. T. Phan, *J Am Chem Soc* **2009**, *131*, 2570-2578.
[42] K. W. Lim, S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D. J. Patel, K. N. Luu and A. T. Phan, *J Am Chem Soc* **2009**, *131*, 4301-4309.
[43] V. Gabelica, E. S. Baker, M. P. Teulade-Fichou, E. De Pauw and M. T. Bowers, *J Am Chem Soc* **2007**, *129*, 895-904.
[44] V. Gabelica, F. Rosu and E. De Pauw, *Anal Chem* **2009**.
[45] M. P. Teulade-Fichou, C. Carrasco, L. Guittat, C. Bailly, P. Alberti, J. L. Mergny, A. David, J. M. Lehn and W. D. Wilson, *J Am Chem Soc* **2003**, *125*, 4732-4740.
[46] J. E. Redman, *Methods* **2007**, *43*, 302-312.
[47] C. Allain, D. Monchaud and M. P. Teulade-Fichou, *J Am Chem Soc* **2006**, *128*, 11890-11893.
[48] A. De Cian, L. Guittat, M. Kaiser, B. Sacca, S. Amrane, A. Bourdoncle, P. Alberti, M. P. Teulade-Fichou, L. Lacroix and J. L. Mergny, *Methods* **2007**, *42*, 183-195.
[49] D. Monchaud, C. Allain and M. P. Teulade-Fichou, *Bioorg Med Chem Lett* **2006**, *16*, 4842-4845.
[50] R. F. Macaya, P. Schultze, F. W. Smith, J. A. Roe and J. Feigon, *Proc Natl Acad Sci U S A* **1993**, *90*, 3745-3749.
[51] E. Baldrich and C. K. O'Sullivan, *Anal Biochem* **2005**, *341*, 194-197.
[52] B. Basnar, R. Elnathan and I. Willner, *Anal Chem* **2006**, *78*, 3638-3642.
[53] K. Walsh and A. Gualberto, *J Biol Chem* **1992**, *267*, 13714-13718.
[54] P. Weisman-Shomer and M. Fry, *Biochem Biophys Res Commun* **1994**, *205*, 305-311.
[55] A. M. Pearson, A. Rich and M. Krieger, *J Biol Chem* **1993**, *268*, 3546-3554.
[56] T. Schierer and E. Henderson, *Biochemistry* **1994**, *33*, 2240-2246.
[57] J. D. Frantz and W. Gilbert, *J Biol Chem* **1995**, *270*, 20692-20697.
[58] J. D. Frantz and W. Gilbert, *J Biol Chem* **1995**, *270*, 9413-9419.
[59] G. Sarig, P. Weisman-Shomer, R. Erlitzki and M. Fry, *J Biol Chem* **1997**, *272*, 4474-4482.
[60] L. A. Dempsey, H. Sun, L. A. Hanakahi and N. Maizels, *J Biol Chem* **1999**, *274*, 1066-1071.
[61] V. Gonzalez, K. Guo, L. Hurley and D. Sun, *J Biol Chem* **2009**.
[62] R. Erlitzki and M. Fry, *J Biol Chem* **1997**, *272*, 15881-15890.
[63] A. W. Oliver, I. Bogdarina, E. Schroeder, I. A. Taylor and G. G. Kneale, *J Mol Biol* **2000**, *301*, 575-584.
[64] L. Uliel, P. Weisman-Shomer, H. Oren-Jazan, T. Newcomb, L. A. Loeb and M. Fry, *J Biol Chem* **2000**, *275*, 33134-33141.

[65] S. Lyonnais, R. J. Gorelick, J. L. Mergny, E. Le Cam and G. Mirambeau, *Nucleic Acids Res* **2003**, *31*, 5754-5763.
[66] E. D. Larson, M. L. Duquette, W. J. Cummings, R. J. Streiff and N. Maizels, *Curr Biol* **2005**, *15*, 470-474.
[67] G. V. Tolstonog, G. Li, R. L. Shoeman and P. Traub, *DNA Cell Biol* **2005**, *24*, 85-110.
[68] A. Lew, W. J. Rutter and G. C. Kennedy, *Proc Natl Acad Sci U S A* **2000**, *97*, 12508-12512.
[69] G. Fang and T. R. Cech, *Cell* **1993**, *74*, 875-885.
[70] R. Giraldo and D. Rhodes, *EMBO J* **1994**, *13*, 2411-2420.
[71] P. B. Arimondo, J. F. Riou, J. L. Mergny, J. Tazi, J. S. Sun, T. Garestier and C. Helene, *Nucleic Acids Res* **2000**, *28*, 4832-4838.
[72] K. Paeschke, T. Simonsson, J. Postberg, D. Rhodes and H. J. Lipps, *Nat Struct Mol Biol* **2005**, *12*, 847-854.
[73] V. I. Bashkirov, H. Scherthan, J. A. Solinger, J. M. Buerstedde and W. D. Heyer, *J Cell Biol* **1997**, *136*, 761-773.
[74] H. Sun, A. Yabuki and N. Maizels, *Proc Natl Acad Sci U S A* **2001**, *98*, 12444-12449.
[75] P. Weisman-Shomer, Y. Naot and M. Fry, *J Biol Chem* **2000**, *275*, 2231-2238.
[76] S. Khateb, P. Weisman-Shomer, I. Hershco, L. A. Loeb and M. Fry, *Nucleic Acids Res* **2004**, *32*, 4145-4154.
[77] Y. C. Lin, J. W. Shih, C. L. Hsu and J. J. Lin, *J Biol Chem* **2001**, *276*, 47671-47674.
[78] H. Fukuda, M. Katahira, N. Tsuchiya, Y. Enokizono, T. Sugimura, M. Nagao and H. Nakagama, *Proc Natl Acad Sci U S A* **2002**, *99*, 12685-12690.
[79] A. J. Zaug, E. R. Podell and T. R. Cech, *Proc Natl Acad Sci U S A* **2005**, *102*, 10864-10869.
[80] T. R. Salas, I. Petruseva, O. Lavrik, A. Bourdoncle, J. L. Mergny, A. Favre and C. Saintome, *Nucleic Acids Res* **2006**, *34*, 4857-4865.
[81] A. C. Connor, K. A. Frederick, E. J. Morgan and L. B. McGown, *J Am Chem Soc* **2006**, *128*, 4986-4991.
[82] J. Tan, C. Vonrhein, O. S. Smart, G. Bricogne, M. Bollati, Y. Kusov, G. Hansen, J. R. Mesters, C. L. Schmidt and R. Hilgenfeld, *PLoS Pathog* **2009**, *5*, e1000428.
[83] K. T. Shum and J. A. Tanner, *Chembiochem* **2008**, *9*, 3037-3045.
[84] A. Matsugami, T. Mashima, F. Nishikawa, K. Murakami, S. Nishikawa, K. Noda, T. Yokoyama and M. Katahira, *Nucleic Acids Symp Ser (Oxf)* **2008**, 179-180.
[85] V. A. Soldatenkov, A. A. Vetcher, T. Duka and S. Ladame, *ACS Chem Biol* **2008**, *3*, 214-219.
[86] L. Xodo, M. Paramasivam, A. Membrino and S. Cogoi, *Nucleic Acids Symp Ser (Oxf)* **2008**, 159-160.
[87] J. W. Shim and L. Q. Gu, *J Phys Chem B* **2008**, *112*, 8354-8360.
[88] Y. Wu, K. Shin-ya and R. M. Brosh, Jr., *Mol Cell Biol* **2008**, *28*, 4116-4128.
[89] D. J. Chinnapen and D. Sen, *Biochemistry* **2002**, *41*, 5202-5212.
[90] J. C. Darnell, K. B. Jensen, P. Jin, V. Brown, S. T. Warren and R. B. Darnell, *Cell* **2001**, *107*, 489-499.
[91] N. Baran, L. Pucshansky, Y. Marco, S. Benjamin and H. Manor, *Nucleic Acids Res* **1997**, *25*, 297-303.

[92] Y. Teng, A. C. Girvan, L. K. Casson, W. M. Pierce, Jr., M. Qian, S. D. Thomas and P. J. Bates, *Cancer Res* **2007**, *67*, 10491-10500.
[93] X. Y. Zhuang, J. Tang, Y. H. Hao and Z. Tan, *J Mol Recognit* **2007**, *20*, 386-391.
[94] Y. Xu, Y. Hirao, Y. Nishimura and H. Sugiyama, *Bioorg Med Chem* **2007**, *15*, 1275-1279.
[95] H. Tahara, K. Shin-Ya, H. Seimiya, H. Yamada, T. Tsuruo and T. Ide, *Oncogene* **2006**, *25*, 1955-1966.
[96] I. M. Pedroso, W. Hayward and T. M. Fletcher, *Nucleic Acids Res* **2009**, *37*, 1541-1554.
[97] B. Scaggiante, B. Dapas, G. Grassi and G. Manzini, *FEBS J* **2006**, *273*, 1350-1361.
[98] S. Ladame, J. A. Schouten, J. Roldan, J. E. Redman, S. Neidle and S. Balasubramanian, *Biochemistry* **2006**, *45*, 1393-1399.
[99] J. Norseen, F. B. Johnson and P. M. Lieberman, *J Virol* **2009**.
[100] C. Harrington, Y. Lan and S. A. Akman, *J Biol Chem* **1997**, *272*, 24631-24636.
[101] A. S. Kamath-Loeb, L. A. Loeb, E. Johansson, P. M. Burgers and M. Fry, *J Biol Chem* **2001**, *276*, 16439-16446.
[102] H. Sun, J. K. Karow, I. D. Hickson and N. Maizels, *J Biol Chem* **1998**, *273*, 27587-27592.
[103] H. Han, R. J. Bennett and L. H. Hurley, *Biochemistry* **2000**, *39*, 9311-9316.
[104] S. Muller, G. D. Pantos, R. Rodriguez and S. Balasubramanian, *Chem Commun (Camb)* **2009**, 80-82.
[105] O. Y. Fedoroff, M. Salazar, H. Han, V. V. Chemeris, S. M. Kerwin and L. H. Hurley, *Biochemistry* **1998**, *37*, 12367-12374.
[106] G. Pennarun, C. Granotier, L. R. Gauthier, D. Gomez, F. Hoffschir, E. Mandine, J. F. Riou, J. L. Mergny, P. Mailliet and F. D. Boussin, *Oncogene* **2005**, *24*, 2917-2928.
[107] D. Sun, B. Thompson, B. E. Cathers, M. Salazar, S. M. Kerwin, J. O. Trent, T. C. Jenkins, S. Neidle and L. H. Hurley, *J Med Chem* **1997**, *40*, 2113-2116.
[108] M. A. Read, A. A. Wood, J. R. Harrison, S. M. Gowan, L. R. Kelland, H. S. Dosanjh and S. Neidle, *J Med Chem* **1999**, *42*, 4538-4546.
[109] M. Laronze-Cochard, Y. M. Kim, B. Brassart, J. F. Riou, J. Y. Laronze and J. Sapi, *Eur J Med Chem* **2009**.
[110] J. P. Whitten, M. Schwaebe and T. Moran in *Heterocyclic substituted 1,4-dihydri-4-0x9-1,8-naphthyridine analogs*, *Vol. US2005004160* USA, **2005**.
[111] D. Gomez, T. Lemarteleur, L. Lacroix, P. Mailliet, J. L. Mergny and J. F. Riou, *Nucleic Acids Res* **2004**, *32*, 371-379.
[112] J. J. Green, S. Ladame, L. Ying, D. Klenerman and S. Balasubramanian, *J Am Chem Soc* **2006**, *128*, 9809-9812.
[113] A. D. Moorhouse, A. M. Santos, M. Gunaratnam, M. Moore, S. Neidle and J. E. Moses, *J Am Chem Soc* **2006**, *128*, 15972-15973.
[114] K. M. Rahman, A. P. Reszka, M. Gunaratnam, S. M. Haider, P. W. Howard, K. R. Fox, S. Neidle and D. E. Thurston, *Chem Commun (Camb)* **2009**, 4097-4099.
[115] X. Sun, *Science in China series B: chemistry* **1999**, *42*, 7.
[116] G. R. Clark, P. D. Pytel, C. J. Squire and S. Neidle, *J Am Chem Soc* **2003**, *125*, 4066-4067.
[117] S. M. Gowan, R. Heald, M. F. Stevens and L. R. Kelland, *Mol Pharmacol* **2001**, *60*, 981-988.

[118] Y. J. Lu, T. M. Ou, J. H. Tan, J. Q. Hou, W. Y. Shao, D. Peng, N. Sun, X. D. Wang, W. B. Wu, X. Z. Bu, Z. S. Huang, D. L. Ma, K. Y. Wong and L. Q. Gu, *J Med Chem* **2008**, *51*, 6381-6392.
[119] M. Franceschin, L. Rossetti, A. D'Ambrosio, S. Schirripa, A. Bianco, G. Ortaggi, M. Savino, C. Schultes and S. Neidle, *Bioorg Med Chem Lett* **2006**, *16*, 1707-1711.
[120] M. Y. Kim, Y. Na, H. Vankayalapati, M. Gleason-Guzman and L. H. Hurley, *J Med Chem* **2003**, *46*, 2958-2972.
[121] M. J. Moore, C. M. Schultes, J. Cuesta, F. Cuenca, M. Gunaratnam, F. A. Tanious, W. D. Wilson and S. Neidle, *J Med Chem* **2006**, *49*, 582-599.
[122] M. Y. Kim, M. Gleason-Guzman, E. Izbicka, D. Nishioka and L. H. Hurley, *Cancer Res* **2003**, *63*, 3247-3256.
[123] N. V. Anantha, M. Azam and R. D. Sheardy, *Biochemistry* **1998**, *37*, 2709-2714.
[124] L. Ginnari-Satriani, V. Casagrande, A. Bianco, G. Ortaggi and M. Franceschin, *Org Biomol Chem* **2009**, *7*, 2513-2516.
[125] A. Bugaut, K. Jantos, J. L. Wietor, R. Rodriguez, J. K. Sanders and S. Balasubramanian, *Angew Chem Int Ed Engl* **2008**, *47*, 2677-2680.
[126] P. Belmont, J. Bosson, T. Godet and M. Tiano, *Anticancer Agents Med Chem* **2007**, *7*, 139-169.
[127] S. Neidle, A. P. Reszka and R. J. Harrison in *Therapeutic acridone and acridine compounds Vol. US2006111389* **2006**.
[128] R. J. Harrison, J. Cuesta, G. Chessari, M. A. Read, S. K. Basra, A. P. Reszka, J. Morrell, S. M. Gowan, C. M. Incles, F. A. Tanious, W. D. Wilson, L. R. Kelland and S. Neidle, *J Med Chem* **2003**, *46*, 4463-4476.
[129] N. H. Campbell, G. N. Parkinson, A. P. Reszka and S. Neidle, *J Am Chem Soc* **2008**, *130*, 6722-6724.
[130] S. M. Gowan, J. R. Harrison, L. Patterson, M. Valenti, M. A. Read, S. Neidle and L. R. Kelland, *Mol Pharmacol* **2002**, *61*, 1154-1162.
[131] J. A. Schouten, S. Ladame, S. J. Mason, M. A. Cooper and S. Balasubramanian, *J Am Chem Soc* **2003**, *125*, 5594-5595.
[132] J. E. Redman, J. M. Granadino-Roldan, J. A. Schouten, S. Ladame, A. P. Reszka, S. Neidle and S. Balasubramanian, *Org Biomol Chem* **2009**, *7*, 76-84.
[133] D. J. Patel, A. T. Phan and V. Kuryavyi, *Nucleic Acids Res* **2007**, *35*, 7429-7455.
[134] F. W. Smith and J. Feigon, *Nature* **1992**, *356*, 164-168.
[135] F. M. Chen, *Biochemistry* **1992**, *31*, 3769-3776.
[136] Y. Wang and D. J. Patel, *Structure* **1994**, *2*, 1141-1156.
[137] N. Zhang, A. T. Phan and D. J. Patel, *J Am Chem Soc* **2005**, *127*, 17277-17285.
[138] Y. Wang and D. J. Patel, *J Mol Biol* **1993**, *234*, 1171-1183.
[139] I. M. Pedroso, L. F. Duarte, G. Yanez, K. Burkewitz and T. M. Fletcher, *Biopolymers* **2007**, *87*, 74-84.
[140] M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S. M. Freier, D. A. Driver, R. H. Berg, S. K. Kim, B. Norden and P. E. Nielsen, *Nature* **1993**, *365*, 566-568.
[141] Y. Krishnan-Ghosh, E. Stephens and S. Balasubramanian, *J Am Chem Soc* **2004**, *126*, 5944-5945.
[142] B. Datta, C. Schmitt and B. A. Armitage, *J Am Chem Soc* **2003**, *125*, 4111-4118.
[143] B. Datta, M. E. Bier, S. Roy and B. A. Armitage, *J Am Chem Soc* **2005**, *127*, 4199-4207.

[144] S. Roy, F. A. Tanious, W. D. Wilson, D. H. Ly and B. A. Armitage, *Biochemistry* **2007**, *46*, 10433-10443.
[145] B. Datta and B. A. Armitage, *J Am Chem Soc* **2001**, *123*, 9612-9619.
[146] L. Petraccone, B. Pagano, V. Esposito, A. Randazzo, G. Piccialli, G. Barone, C. A. Mattia and C. Giancola, *J Am Chem Soc* **2005**, *127*, 16215-16223.
[147] S. M. Haider, G. N. Parkinson and S. Neidle, *J Mol Biol* **2003**, *326*, 117-125.
[148] E. Gavathiotis, R. A. Heald, M. F. Stevens and M. S. Searle, *J Mol Biol* **2003**, *334*, 25-36.
[149] F. Ullmann and J. Bielecki, *Chemische Berichte* **1901**, *34*, 11.
[150] Fritz Ullmann and Paul Sponagel, *Berichte der deutschen chemischen Gesellschaft* **1905**, *38*, 2211-2212.
[151] R. Csuk, A. Barthel and C. Raschke, *Tetrahedron* **2004**, *60*, 4.
[152] M. Saphier, A. Masarwa, H. Cohen and D. Meyerstein, *Eur J Inorg Chem* **2002**, 9.
[153] C. Wolf, S. Liu, X. Mei, A. T. August and M. D. Casimir, *J Org Chem* **2006**, *71*, 3270-3273.
[154] G. W. Rewcastle and W. A. Denny, *Synthesis* **1985**, *2*, 4.
[155] R. F. Pellon, M. L. Docampo, Z. Kunakbaeva, V. Gomez and H. Vélez-Castro, *Synth Comm* **2006**, *36*, 5.
[156] A. S. Guram and S. L. Buchwald, *J Am Chem Soc* **1994**, *116*, 2.
[157] F. Paul, J. Patt and J. F. Hartwig, *Journal of the American Chemical Society* **1994**, *116*, 5969-5970.
[158] A. Goldberg and W. Kelly, *J Chem Soc* **1946**, 10.
[159] W. A. Denny, G. J. Atwell and B. C. Baguley, *J Med Chem* **1983**, *26*, 1625-1630.
[160] X. Mei, A. T. August and C. Wolf, *J Org Chem* **2006**, *71*, 142-149.
[161] R. Pellon, R. Carrasco, T. Marquez and T. Manposo, *Tetrahedron Letters* **1997**, *38*, 3.
[162] Y. Bretonniere, M. J. Cann, D. Parker and R. Slater, *Org Biomol Chem* **2004**, *2*, 1624-1632.
[163] K. Matsumura, *J Am Chem Soc* **1938**, *60*, 3.
[164] P. Yang, A. De Cian, M. P. Teulade-Fichou, J. L. Mergny and D. Monchaud, *Angew Chem Int Ed Engl* **2009**, *48*, 2188-2191.
[165] J. Kim, C. Cheong and P. B. Moore, *Nature* **1991**, *351*, 331-332.
[166] J. L. Mergny, A. De Cian, A. Ghelab, B. Sacca and L. Lacroix, *Nucleic Acids Res* **2005**, *33*, 81-94.
[167] A. Arora and S. Maiti, *J Phys Chem B* **2009**.
[168] B. Sacca, L. Lacroix and J. L. Mergny, *Nucleic Acids Res* **2005**, *33*, 1182-1192.
[169] J. Christiansen, M. Kofod and F. C. Nielsen, *Nucleic Acids Res* **1994**, *22*, 5709-5716.
[170] B. C. Horsburgh, H. Kollmus, H. Hauser and D. M. Coen, *Cell* **1996**, *86*, 949-959.
[171] K. Murakami, F. Nishikawa, K. Noda, T. Yokoyama and S. Nishikawa, *Prion* **2008**, *2*, 73-80.
[172] S. Bonnal, C. Schaeffer, L. Creancier, S. Clamens, H. Moine, A. C. Prats and S. Vagner, *J Biol Chem* **2003**, *278*, 39330-39336.
[173] M. Wieland and J. S. Hartig, *Angew Chem Int Ed Engl* **2006**, *45*, 5875-5878.
[174] M. Wieland and J. S. Hartig, *Chem Biol* **2007**, *14*, 757-763.

[175] S. Kumari, A. Bugaut and S. Balasubramanian, *Biochemistry* **2008**, *47*, 12664-12669.
[176] J. Gros, A. Guedin, J. L. Mergny and L. Lacroix, *Chembiochem* **2008**, *9*, 2075-2079.
[177] A. De Cian and J. L. Mergny, *Nucleic Acids Res* **2007**, *35*, 2483-2493.
[178] A. Arora, M. Dutkiewicz, V. Scaria, M. Hariharan, S. Maiti and J. Kurreck, *RNA* **2008**, *14*, 1290-1296.
[179] M. J. Morris and S. Basu, *Biochemistry* **2009**, *48*, 5313-5319.
[180] J. C. Norton, M. A. Piatyszek, W. E. Wright, J. W. Shay and D. R. Corey, *Nat Biotechnol* **1996**, *14*, 615-619.
[181] O. Almarsson, T. C. Bruice, J. Kerr and R. N. Zuckermann, *Proc Natl Acad Sci U S A* **1993**, *90*, 7518-7522.
[182] O. Almarsson and T. C. Bruice, *Proc Natl Acad Sci U S A* **1993**, *90*, 9542-9546.
[183] T. Vilaivan and G. Lowe, *J Am Chem Soc* **2002**, *124*, 9326-9327.
[184] A. N. Elayadi, D. A. Braasch and D. R. Corey, *Biochemistry* **2002**, *41*, 9973-9981.
[185] A. Whitney, G. Gavory and S. Balasubramanian, *Chem Commun (Camb)* **2003**, 36-37.
[186] B. Armitage, T. Koch, H. Frydenlund, H. Orum and G. B. Schuster, *Nucleic Acids Res* **1998**, *26*, 715-720.
[187] N. Winssinger, R. Damoiseaux, D. C. Tully, B. H. Geierstanger, K. Burdick and J. L. Harris, *Chem Biol* **2004**, *11*, 1351-1360.
[188] K. L. Robertson, L. Yu, B. A. Armitage, A. J. Lopez and L. A. Peteanu, *Biochemistry* **2006**, *45*, 6066-6074.
[189] Z. L. Pianowski and N. Winssinger, *Chem Commun (Camb)* **2007**, 3820-3822.
[190] M. Deng, D. Zhang, Y. Zhou and X. Zhou, *J Am Chem Soc* **2008**, *130*, 13095-13102.
[191] P. Travascio, P. K. Witting, A. G. Mauk and D. Sen, *J Am Chem Soc* **2001**, *123*, 1337-1348.
[192] V. L. Marin and B. A. Armitage, *J Am Chem Soc* **2005**, *127*, 8032-8033.
[193] C. B. Carlson and P. A. Beal, *Bioorganic & Medicinal Chemistry Letters* **2000**, *10*, 1979-1982.
[194] M. Beyermann, M. Bienert, H. Niedrich, L. A. Carpino and D. Sadat-Aalaee, *The Journal of Organic Chemistry* **1990**, *55*, 721-728.
[195] A. Joachimi, A. Benz and J. S. Hartig, *Bioorg Med Chem* **2009**, *17*, 6811-6815.
[196] A. Adams, J. M. Guss, W. A. Denny and L. P. Wakelin, *Nucleic Acids Res* **2002**, *30*, 719-725.
[197] A. Adams, J. M. Guss, W. A. Denny and L. P. Wakelin, *Acta Crystallogr D Biol Crystallogr* **2004**, *60*, 823-828.
[198] M. Orzaez, L. Mondragon, A. Garcia-Jareno, S. Mosulen, A. Pineda-Lucena and E. Perez-Paya, *Bioorg Med Chem Lett* **2009**, *19*, 1592-1595.
[199] G. J. Finlay, G. J. Atwell and B. C. Baguley, *Oncol Res* **1999**, *11*, 249-254.
[200] J. R. Goodell, B. Svensson and D. M. Ferguson, *Journal of Chemical Information and Modeling* **2006**, *46*, 876-883.
[201] J. Sebestik, M. Safarik, I. Stibor and J. Hlavacek, *Biopolymers* **2006**, *84*, 605-614.
[202] P. A. Brady and J. K. M. Sanders, *Chem Soc Rev* **1997**, *26*, 10.
[203] E. Fischer, *Chem. Ber.* **1894**, *27*, 9.

[204] P. A. Brady, R. P. BonarLaw, S. J. Rowan, C. J. Suckling and J. K. M. Sanders, *Chemical Communications* **1996**, 319-320.
[205] O. Ramstrom and J. M. Lehn, *Nature Reviews Drug Discovery* **2002**, *1*, 26-36.
[206] N. Giuseppone, J.-L. Schmitt and J.-M. Lehn, *Journal of the American Chemical Society* **2006**, *128*, 16748-16763.
[207] C. Saiz, P. Wipf, E. Manta and G. Mahler, *Organic Letters* **2009**, *11*, 3170-3173.
[208] I. Huc and J. M. Lehn, *Proc Natl Acad Sci U S A* **1997**, *94*, 2106-2110.
[209] J. M. Lehn, *Chem Eur J* **1999**, *5*, 9.
[210] P. T. Corbett, J. Leclaire, L. Vial, K. R. West, J. L. Wietor, J. K. M. Sanders and S. Otto, *Chemical Reviews* **2006**, *106*, 3652-3711.
[211] S. Ladame, *Org Biomol Chem* **2008**, *6*, 219-226.
[212] B. Klekota, M. H. Hammond and B. L. Miller, *Tetrahedron Letters* **1997**, *38*, 8639-8642.
[213] O. Ramstrom and J. M. Lehn, *Chembiochem* **2000**, *1*, 41-48.
[214] T. Bunyapaiboonsri, O. Ramstrom, S. Lohmann, J. M. Lehn, L. Peng and M. Goeldner, *Chembiochem* **2001**, *2*, 438-444.
[215] A. Albert and W. Gledhill, *Journal of the Society of Chemical Industry, London* **1945**, *64*, 169-172.
[216] B. F. Cain, R. N. Seelye and G. J. Atwell, *J Med Chem* **1974**, *17*, 922-930.
[217] A. M. Whitney, S. Ladame and S. Balasubramanian, *Angew Chem Int Ed Engl* **2004**, *43*, 1143-1146.
[218] S. Ladame, A. M. Whitney and S. Balasubramanian, *Angew Chem Int Ed Engl* **2005**, *44*, 5736-5739.