UNIVERSITÉ DE STRASBOURG

Thèse présentée pour obtenir le grade
de Docteur de l'Université de Strasbourg

Dicipline: Aspects moléculaires et cellulaires de la Biologie

Par Emiko Uchikawa

A structural approach of RNA-protein recognition and kinetics of binding in two examples: tRNA aminoacylation by arginyl-tRNA synthetase and 7SK stabilization by LaRP7

Soutenue publiquement le 1 Mars 2011

Membres du Jury

Dr. Akio TAKENAKA	Rapporteur externe
Dr. Kei YURA	Rapporteur externe
Dr. Catherine FLORENTZ	Examinateur
Dr. Keiko TAKANO	Examinateur
Dr. Anne-Catherine DOCK-BREGEON	Directeur de Thése
Dr. Michiko KONNO	Codirecteur deThése

Acknowledgement

First I would like to express my gratitude to Prof. KONNO Michiko and Dr. DOCK BREGEON Anne Catherine I was strongly supported by them. They always have passion and supervised me. I was able to learn a lot of things from them. I am deeply grateful to them.

In Japan

I would like to thank people who worked with me during my PhD time at KONNO laboratory.
I was able to learn a lot of experimental skills at Dr. YOKOYAMA Shigeyuki laboratory.
I appreciate for his acceptance and his laboratory members.
I appreciate Prof. TAKANO Keiko to give me a lot of opportunities.
I also would like to thank all teachers and students who help and cheer me up.

I really appreciate JSPS (Japan society for the promotion of science) and Ochanomizu University. I was supported 11 months by ITP(International Training Program) from JSPS and Ochanomizu University.
I really appreciate Bourses du government français, Foundation of Dr. YUASA Toshiko, Japan-France Science technology foundation. I was supported for 6 months from these organizations.

In France

I am grateful to Prof. Dino Moras for acceptance to his laboratory.
I would like to thank my laboratory members especially MARTINEZ ZAPIEN Denise, DURAND Alexandre and Structural Biology and Genomics technology platform.
I thank all staffs and students who worked with me at IGBMC.

I appreciate all Juries
Prof. TAKENAKA Akio
Prof. OGAWA Haruko
Prof. TAKANO Keiko
Prof. YURA Kei
Prof. FLORENTZ Catherine

Most of all I would like to thank my parents and sisters for their kind supports.

Presentation of the manuscript

This manuscript describes my PhD work, which was done in two different laboratories (countries). In a fist 1.5-year period, I worked under the supervision of Prof. Michiko KONNO, at Ochanomizu University, and investigated Arginyl-tRNA syntetase (ArgRS). 1.5-year was spent at Illkirch (France), where I investigated Threonyl-tRNA synthetase and LaRP7 under the supervision of Dr. Anne-Catherine DOCK-BREGEON. This special organization for the PhD course was an opportunity given by Ochanomizu University, which I was very glad to accept. It gave me the occasion to stay in a new country, improve my English, learn many experiments, and pursue my approach of RNA-protein interactions.

During the PhD my interest was interaction RNA and protein interactions, which play key roles in the cell. I studied these interactions from structural aspects, with the hope of solving a 3D structure by crystallography. A lot of difficulties exist to solve the structure of RNA-protein complex, especially complexes derived from RNA and protein from human have been reported in only a few case at atomic resolution. RNA preparing has a lot of difficulties (i.e. transcription, degradation, conformation and cost), also control of the expression and stability of protein is difficult. In crystallization step, especially at the time of making a crystal of protein and RNA complex, the parameters for crystallization condition that need to be fixed is expanded. However the analysis of protein and RNA gives us critically interesting information and 3D structure is really powerful information for understanding the RNA and protein interaction. In my thesis, I always dealt with RNA to understand the feature of protein. This thesis is written about my endeavor and contribution for understanding RNA and protein binding.

There are two projects, one related to translation and the other is transcription. The manuscript is organized in four chapters, as follows. A general introduction to RNA-protein interactions highlights their biological importance and describes shortly the structures of some remarkable complexes. A special attention is given to two types of complexes. The anticodon binding domains of aminoacyl-tRNA synthetases on one hand and the RNA Recognition Motif, on the other hand, are of direct interest to my personal work. The structural point of view is favored throughout. Chapter II comprises the introduction to aminoacyl-tRNA synthetases (aaRS), their function and structural organization. A first sub-chapter gives a special focus on class I enzymes, ArgRS. Our project, and results, on the subject of ArgRS is then described. The second part of Chapter II is devoted to a class II synthetase, ThrRS, and a project on which my small contribution allowed me to use new techniques of co-expression.

With Chapter III, I still change the topic to transcription. The introduction gives information on transcription elongation, the 7SK riboregulator, and the role of LaRP7 as a chaperone for 7SK. My work on LaRP7, and my results are then described. Chapter IV describes the protocols, recipes, and gives useful information about the techniques that I used. The manuscript finishes with a short personal conclusion, and the bibliography.

Summary of the manuscript

In order to help my referees to review my PhD work, here I describe what forms the foundation of my interest through these studies, and briefly mention the purpose and result of each chapter. My Thesis is organized into four sections.

Chapter I: is a general introduction to RNA recognition motifs, written from a structural point of view. In the cell, we can observe a lot of examples of RNA-protein interactions, which are involved in important biological roles. I worked on two different projects with two supervisors, but a common aim is to investigate protein: RNA bindings by aspect of structural biology. During my stay in Ochanomizu University under Prof. Michiko KONNO, I studied the interaction of Arginyl-tRNA synthetase and tRNA (a subject related to translation), and in Strasbourg University under Dr. Anne-Catherine DOCK-BREGEON, I studied the interaction of LaRP7 protein and 7SK RNA (a project related to Transcription). Both proteins bind to RNA, but the way and function are different. Therefore in this introductory chapter, I described general knowledge about Protein: RNA binding modes related to these two projects. Thus, ***Chap I-3*** introduces to the binding between anticodon binding domains of aminoacyl-tRNA synthetases and the anticodon of tRNAs, and ***Chap I-4*** gives an introduction to the binding between the RNA recognition motif (RRM), a motif present in LaRP7, and RNA.

Chapter II is a study of aminoacyl-tRNA synthetases (aaRS), comprising structural and kinetics analysis. ***Chap II-2*** describes my investigation about Arginyl-tRNA synthetase (supervised by Prof. KONNO). The aim of the project on ArgRS was to understand the mechanism of ArgRS and tRNA binding and aminoacylation reaction using the system of *P.horikoshii*, for which a crystal structure of ArgRS-$tRNA^{Arg}$ complex was solved in Prof. KONNO's laboratory. During my PhD I (1) participated in the structural analysis of the ArgRS-$tRNA^{Arg}$ complex and the project to purpose an idea of aminoacylation reaction in

P.horikoshii ArgRS. I also investigated (2) the contribution of the N-terminal domain of ArgRS to the aminoacylation reaction, using the techniques of kinetic measurement (enzymatics) and crystallization. My results showed that the N-terminal domain contribution was not so important for tRNA binding. I (3) aimed for further understanding of the binding mechanism between anticodon of tRNA and anticodon binding domain of ArgRS, by mutating of $tRNA^{Arg}$ and by crystallization and got some crystals.

Chap II-3 is about Threonyl-tRNA (supervised by Dr. Anne-Catherine DOCK-BREGEON).
The aim of this project was to understand the ThrRS and tRNA interaction mechanism in a special ThrRS present in Crenarchaeal, a subset of archaeas. These organisms are subset is different from other organisms and has two ThrRS, one assuming the synthetic, aminoacylation function and the other assuming an editing function. We wanted to understand if these two functions were physically linked. Our experiments showed that the editing domain of *S.tokodai* ThrRS-2 has a strong affinity for tRNA, when compared with the aminoacylating domain of *S.tokodai* ThrRS-1. We could obtain crystals of *S.tokodai* ThrRS-2 and *E.coli* $tRNA^{Thr}$ complex but has not been solved the structure yet.

Chapter III describes our studies of LaRP7 in the 7SK snRNP. (All this part was investigated with Dr. Anne-Catherine DOCK-BREGEON).
The aim of this project is to investigate the binding mechanism between LaRP7 and 7SK, in the scope of a study of the role of 7SK in the regulation mechanism of the elongation in transcription. My work included (1) the trial and error for preparation of crystallizable protein (focused on the La-domain of LaRP7) and RNA (a domain of 7SK) by studying the factors for crystallization (stability, affinity, polydispersity, etc); finally some crystals were obtained.
I performed in parallel (2) the analysis of binding affinity between N-terminal region of LaRP7 and 3' end region of 7SK using different techniques. As a result, the *K*d between LaRP7 or the N-terminal region of LaRP7 and HP4U were obtained, and we could detect some idiosyncratic feature for binding between LaRP7 and HP4U.

Chapter IV describes the material and methods used for my experiments. During my PhD I learned a lot of techniques and procedures. Reproducing crystal needs very accurate experimental condition therefore it is important to describe the details of experiments.

Finally I mentioned in a brief "Conclusion and perspectives" what I learned and did during my PhD struggle to have results, and indicate what should be done the next step in these inquiries.

Abbreviations

aaRS		Aminoacyl-tRNA synthetase
tRNA		Transfer ribonucleic Acid
AMP-PNP		adenosine 5'-(β,γ-imido) triphosphate
A	Ala	alanine
C	Cys	cysteine
D	Asp	aspartic acid
E	Glu	glutamic acid
F	Phe	phenylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
K	Lys	lysine
L	Leu	leucine
M	Met	methionine
N	Asn	asparagines
P	Pro	praline
Q	Gln	glutamine
R	Ar	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophan
Y	Tyr	tyrosine
IPTG		isopropyl-β-D-thiogalactopyranoside
β-Me		2-merchaptoethanol
EtOH		ethanol
DTT		dithiothreitol
TCEP		tris(2-carboxyethyl)phosphine

SDS	Sodium dodecylsurfate
EMSA	Electrophoresis Mobility Shift Assay
ITC	isothermal titration calorimetry
ATP	Adenine triphosphate
CTP	Ctosine triphosphate
UTP	Uridine triphosphate
GTP	Guanosine triphosphate
GMP	Guanosine monophosphate
RNP	Ribonucleo protein
mRNA	Messenger ribonucleic acid
snRNP	Small nuclear ribonucleo protein
P-TEFb	Positive Transcription Elongation Factor b
RNAPI	RNA polymerase I
RNAPII	RNA polymerase II
RNAPIII	RNA polymerase III
CTD	Carboxy terminal domain
DSIF	DRB sensitivity inducing factor
NELF	Negative elongation factor
FACT	Facilitates chromatin transcription factor
HEPES	N-{2-hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid]
Tris	tris[hydroxymethyl]aminomethane
PEG	polyethylene glycol

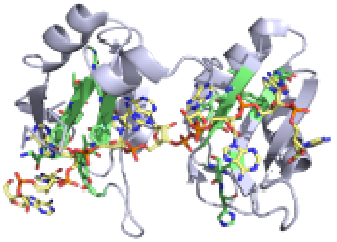
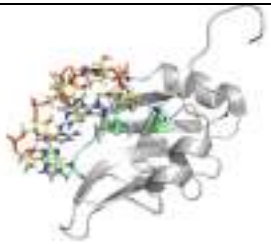
Chapter I A general introduction to RNA-protein complexes

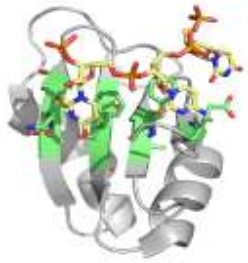
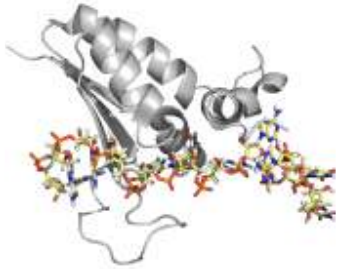
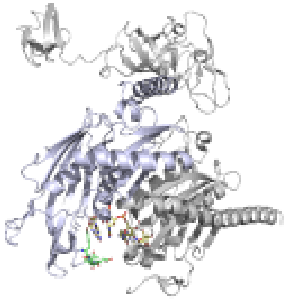
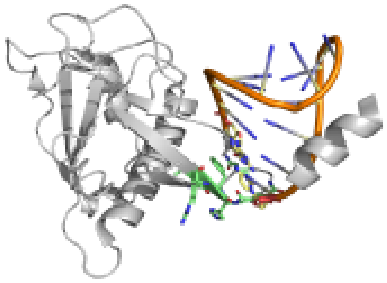
1. RNA binding proteins

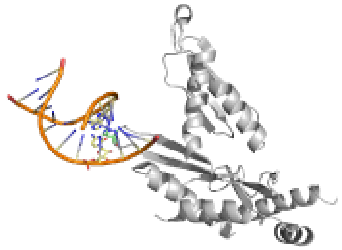
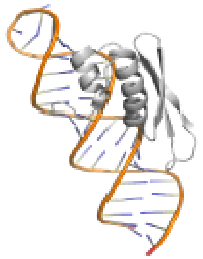
RNA binding proteins are involved in a multitude of RNA-protein interactions that are used widely in the cell to regulate gene expression. In eukaryotes, protein synthesis does not occur until the newly transcribed RNA is extensively modified and exported into the cytoplasm. In the cytoplasm, varieties of RNA-protein interactions are employed by the cell and regulate the levels of protein expression. RNA binding proteins were early revealed to play a fundamental role in translation, where the deciphering of the genetic message embedded in the succession of codons of the messenger RNA (mRNA) relies upon the precise recognition of the mediator tRNA by the specific aminoacyl-tRNA synthetase (aaRS) protein. Accurate RNA recognition also directs other aspects of translation, the ribosome architecture being of major importance, as is the complex orchestration of translation initiation. Research has now unveiled many complex recognition events in the processing of the primary transcripts into mRNA, which includes many specific steps such as polyadenylation, splicing and alternative splicing, as well as editing, nuclear export and cytoplasmic targeting. Many interesting networks of interactions between RNA and protein occur also at the level of mRNA turnover and quality control. In addition to these post-transcriptional processes, RNA-protein interactions play a key role in transcriptions illustrated by the life cycle of retroviruses. These are exciting area of research, with numerous impacts on human health. Today many RNA-protein complex 3D structures have been published to describe at atomic resolution. For example the RNA recognition by aaRSs, endonucleases, RNA helicases, isomerases, or cytosolic sensors. Some interesting examples, chosen in the various processes of the cell life mentioned above, are listed in Table I-1-1. Unlike DNA, the structure of RNA is highly variable and has flexible conformation, thus creating a number of unique binding sites for RNA binding proteins. The RNA binding proteins contain structural motifs, as for example RRM (RNA Recognition Motif), dsRBM (double strand RNA Binding Motif), RecA related domain, KH domain, PAZ domain, Sm domain or OB fold (Chapter I-2). Interactions of aaRS with tRNAs have been thoroughly studied by crystallography, probably because of the relative stability of tRNAs, and have provided a wealth of data that we will introduce in the next section (Chapter I-3). Another interesting recognition mode, still mysterious, relies upon the RRM, the most frequent fold observed in proteins. Since RRMs are frequently found in the field of transcription and mRNA maturation, we shall give a structural introduction to RRMs in the next (Chapter I-4) section.

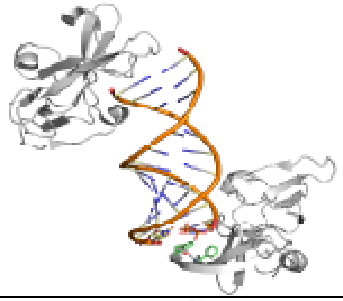
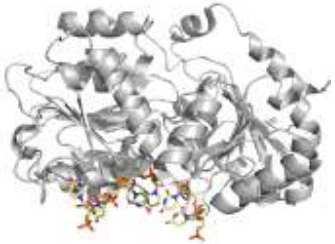
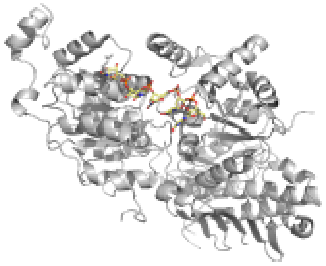
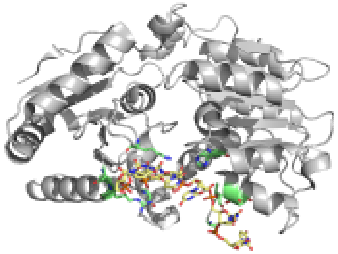
Table. I-1-1 RNA-Protein complex structures

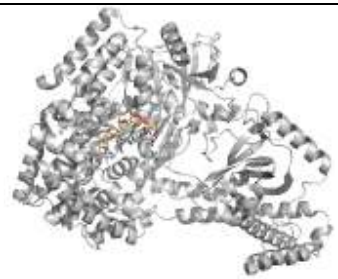
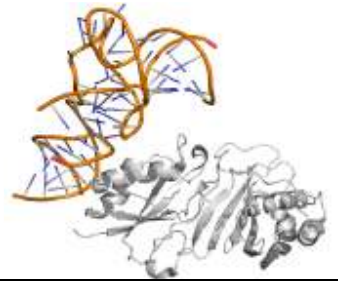
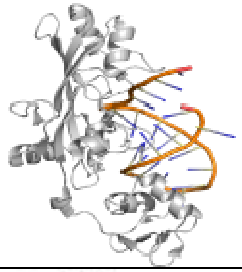
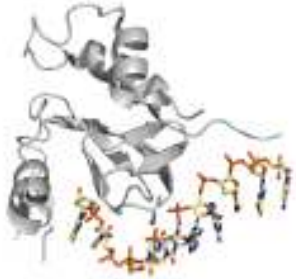
Several RNA: Protein complex structures are selected from Protein Data Bank (PDB) and classified by its RNA binding motif. The motif characteristics are described in Chap. I.2.

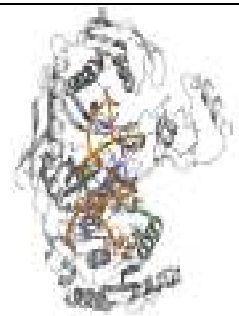
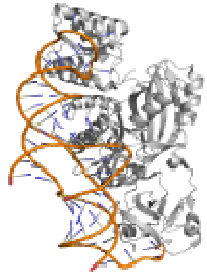
	RNA binding motif	RNA	Protein	PDB ID	Structure
Processing	RRM 1, 2 (RNA Recognition Motif)	mRNA poly(A) tail	Poly(A)-binding protein (PABP)	1CVJ	
Regulation mRNA metabolism	RRM	G-tract RNA	hnRNP (heterogeneous nuclear ribonucleo protein)	2KFI	
Gene expression	RRM	UAUAU	Rna15 and Hrp1 (A cleavage factor I component)	2KM8	

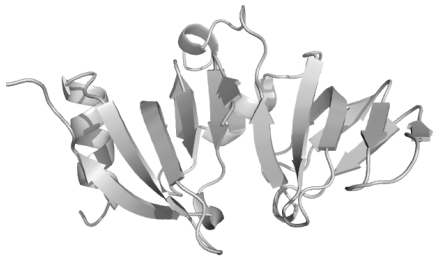
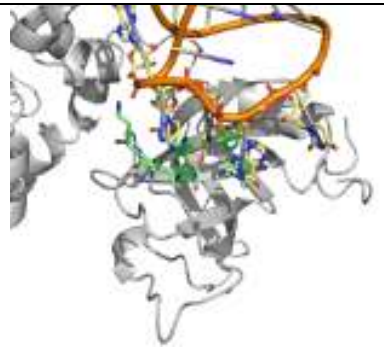
Transcription termination	RRM	snRNA UCUU	Nrd1 Nab3	2XNR	
Splicing Factor	KH (K homology) domain αβββα	branch point sequence (BPS) in the pre-mRNA transcripts	SF1 (splicing factor 1)	1K1G	
Exosome	KH like motif	6bp RNA	Rrp41,Rrp42 (A RNase PH-like proteins an 3'-5' exoribonuclease and nucleotidyltransferase)	3M7N	
Endonuclease	Ferredoxin-like domain (KH) (RRM β-sheet)	16-nucleotideminimal RNA substrate CRISPR transcript (pre-crRNA) processing	Csy4 (Endoribonuclease)	2XLI	

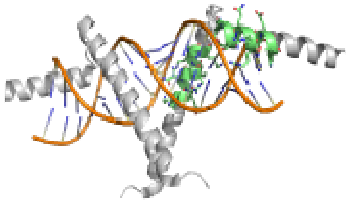
Polyadenylation specificity factor	KH like motif	6 bp uridine	CPSF (Cleavage and polyadenylation specificity factor)	3AF6	
RNA processing factor	KH like motif	3′ end sequence of 16S rRNA	eIF2 (Eukaryotic Initiation Factor 2)	3AVE	
Processing	dsRBD (RNA binding domain) αββββα	dsRNA	HYL1 (Hyponastic Leaves1) processing of primary miRNAs.	3ADI	
Isomerase	dsRBD αββββα	Stem-loop pre-mRNA encoding the GluR-2 R/G RNA.	ADARs (Adenosine deaminases that act on RNA)	2L2K	

Immune response	RIG-I αββββα	14 bp 5'ppp dsRNA	RIG-I (retinoic acid-inducible gene-I)	3LRN	
RNA helicase	RecA like domain βαβαββαβαβαβ	10bp uridine mRNA mimic	DDX19 (DEXD/H-box RNA helicase)	3G0H	
Splicing	RecA like domain	A poly uracil oligonucleotide mRNA mimic	A multiprotein exon junction complex	3EX7	
Export protein	RecA like domains	6sU	Dbp5 (DEAD-box protein essential for mRNA export from the nucleus)	3FHT	

RNA processing	RecA	10s A	Mtr4 (a conserved RNA helicase that functions together with the nuclear exosome)	2XGJ	
Isomerase	RecA like domain (TruA)	tRNA	TruA (pseudouridine synthase)	2NQP	
Isomerase	RecA like domain (Rluf)	23S RNA stem-loop fragments	Rlu (pseudouridine synthase)	3DH3	
RNAs mediate gene silencing	PAZ domain (PIWI, Argonaut, Zwille) αααββαβββ	siRNA-like duplex	PAZ domain	1SI2	

Gene silencing	PIWI domain ββαβαββαβ PAZ domain	Guid RNA and DNA of mRNA	Argonaut (catalytic component of RNA induced silencing complex)	3HJF	
Isomerase	Cbf5	snoRNA H/ACA	RNP pseudouridine synthase	3HJW	
RNA chaperone	Hfq domain αββββββ	AU-rich sequences	Hfq (A small, highly abundant hexameric protein)	3GIB	
Transcription repression	A five-stranded anti-parallel β sheet and two helices βββββαα	Cognate hairpin RNA	PP7 (Ser/Thr phosphatase of the PPP family)	2QUX	

Splicing Processing	Sm protein	mRNA	U1 small nuclear ribonucleoprotein particles (snRNPs)	1B34 3CW1	
Ligase	OB fold (Oligonucleotide binding fold)	$tRNA^{Asp}$	Aspalanyl-tRNA synthetase	1ASY	
Splicing	Nop domain αβαβαβαβ–αααα	snRNA	snRNP	2OZB	

Gene expression	α-helical repeat	8 bp RNA	PUF proteins (regulate gene expression post-transcriptionally by binding to sequences in the 3’ untranslated region of target transcripts)	3BX2	
Binds misfolded noncoding RNAs	vWFA(von Willebrand factor A) αα–αα–αα–αα	pre-5S ribosomal RNA	Ro auto antigen (bind to misfolded RNA)	2I91	
RNA silencing suppression	α-helix backbone	siRNA duplex	TAV2b (Tomato aspermy virus 2b)	2ZI0	

2. RNA binding motif

So far a lot of RNA-protein structures were solved. The RNA-protein structures solved by X-ray crystallography and NMR were picked up in the Table I-1-1. RNA-protein structure has wide varieties, thus complicated RNA protein structures were not picked up (i.e. ribosome). These RNA-protein binding ways are expanded and some proteins has common motif named RRM, KH (K homology) domain, dsRBD (double strand RNA binding domain), RecA like domain, among these motifs and if motif is not named, some of them have similar structure and characteristics. However, the way of bindings show highly variability. Hence it is difficult to explain RNA-Protein binding regularity or mechanism only from the structural point of view. From structural aspect, it is important to know the structural dynamics between RNA and protein.

Anticodon-binding domain of aminoacyl-tRNA synthetase and RRM will be focused on Chapter I-3 and Chapter I-4.

2.1. KH domain (K homology domain) (Fig. I-2-1)

K homology domain is a widely used RNA binding motif that has been detected by sequence similarity searched in such proteins as heterogeneous nuclear ribonucleoprotein K (hnRNP K) and ribosomal protein S3. KH domain share significant sequence similarity and possess a compact structure. KH domain include βααβ unit KH domain can be described as α+β two-layer sandwich with α–β plate topology this topology is also replicated by as the ferredoxin-like protein.

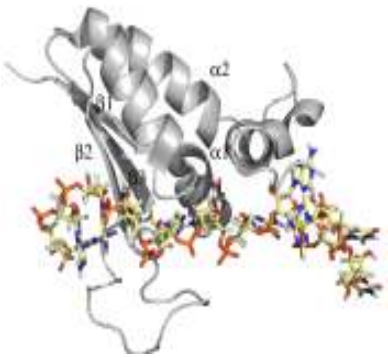


Fig. I-2-1 KH domain (PDBID : 1K1G)

RNA interaction region often placed at surface of β-sheet and loop 1 loop2 (between β1 and α1, β2 and β3).

2.2. dsRBD (Double strand RNA binding Domain) (Fig. I-2-2)

DsRBD is αββββα fold with a well-characterized function to bind structured RNA molecules. DsRBD containing proteins are involved in processes ranging from RNA editing to protein phosphorylation in translational control. And they contain a variable number of dsRBD domains. The RNA duplex contacts residues in loop 2 and 4 (between β1 and β2, β3 and α2) in dsRBD. DsRBDs are also able to recognize non-RNA targets and can act in combination with other dsRBDs and non-dsRBM.

Fig. I-2-2 DsRDB (PDBID: 2L2K)

2.3. OB fold (Oligonucleotide binding fold) (Fig. I-2-3)

Oligonucleotide binding (OB) fold motifs has a five-stranded mixed β barrel. In the majority of cases, the barrel has a Greek key topology and one end of the barrel is capped by an α helix. This is also thought as a DNA/RNA binding motif. OB-fold protein use β-sheet surface of β2, β3 and loop region of loop2 and loop3 (between β2 and β3 and between β3 and β4) to bind to RNA (the anticodon-binding domains of LysRS, AspRS), and ssDNA.

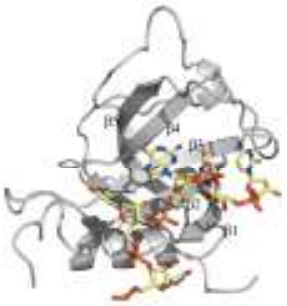


Fig. I-2-3 OB fold (PDBID: 1ASY)

2.4. RecA like fold (Fig. I-2-4)

Originally, RecA is known as the protein is essential for the repair and maintenance of DNA. RecA is also the RNA and nucleotide binding core of all RNA helicases. RecA has fold interface involves the packing of α helix and β sheet in the central of the domain. The interaction region with RNA is often found in a loop region.

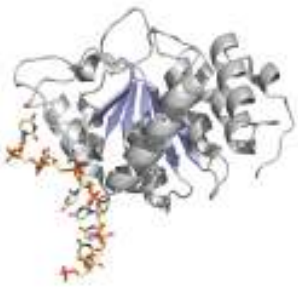


Fig. I-2-4 RecA like fold (PDBID: 3G0H)

2.5. Zinc finger (Fig. I-2-5)

Zinc fingers are small protein structural motifs that can coordinate one or more zinc ions to help stabilize their folds. They can be classified into several different structural families and typically function as interaction modules that bind DNA, RNA, proteins, or small molecules. Zinc fingers that interact with RNA was found among the structures of members from the Gag knuckle. The structure of this motif has been reported from the retroviral nucleocapsid (NC) protein from HIV and other related viruses. The Gag knuckle binds to single-stranded RNA and is involved in recognizing specific sequences of RNA needed for viral packaging.

Fig. I-2-5 Zinc finger (PDBID: 1TF6)

2.6. Others

Based on the classification in the Table. I-1-1, PAZ domain (Piwi Argonaut Zwille), RIG-I, Cbf5, Hfq, PP7, Sm domain has similar structure with RRM, KH and dsRDB. They always have parallel β-sheets in or out side of protein. In addition RNA interaction site was often placed at loop region and surface of β-sheets.

On the other hand, vWFA, TAV2b, PUF proteins have only α-helixes in their RNA binding region. TAV2b binds RNA directly α helix surface. In vWFA case, binding region is placed at the loop region between two α helices.

2. Anticodon binding domains of aminoacyl-tRNA synthetases and tRNA anticodon binding

Aminoacyl-tRNA synthetases (aaRS) must recognize and aminoacylate without errors their cognate tRNAs. Most of these enzymes (but there are notable exception, like seryl-tRNA synthetase) possess an anticodon binding domain, which binds the anticodon loop of the cognate tRNAs. Several tRNAs can be bound to one aaRS, because of the degeneracy of the genetic code (Fig. I-3-1). This explains why the first base of the anticodon triplet (base 34) is of less importance in direct recognition. However, its structural confinement by the particular structure of the anticodon loop plays a role in the orientation of the determinant bases.

The thorough studies of aaRSs, started in the 50's, and the way they recognize their cognate tRNAs have shown that these proteins use several different folds to recognize the anticodon loop of tRNAs, even though the anticodon loops have similar structures in the free (unbound) form. Anticodon binding domain is about 100 to 200 aminoacids and varieties of structures were described. In all cases when bases from the anticodon are bound to the aaRS anticodon binding domain, a patchwork of hydrogen bonds and stacking interaction leads to the specific recognition, and the exclusion of non-cognate molecules. Table I-3-1 recalls the partition of the 20 aaRSs in two classes, corresponding to two different ancestral catalytic domains for the recognition of ATP, the Rossman fold (six parallel β-strands linked to two pairs of a-helices) for class I (Rossmann et.al 1974), and the aaRS class II particular fold (antiparallel β-strands) (Ruff et al.1991, Cusack et al. 1990).

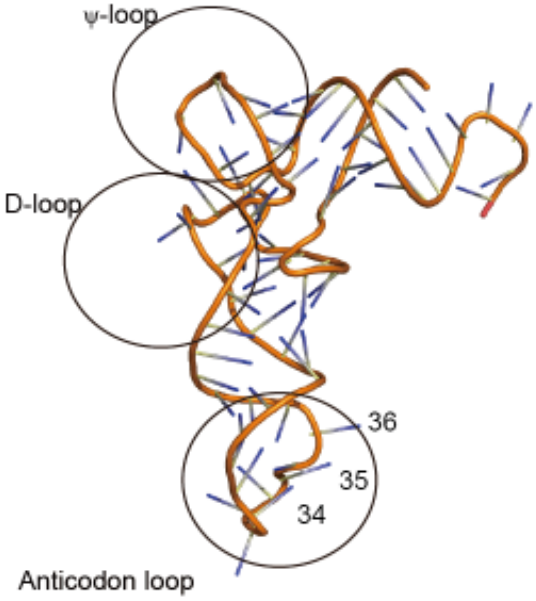


Fig I-3-1 *S.cereviseae* **tRNA^{Arg} structure**
34,35,36 is anticodon of tRNA

Table I-3-1

aaRS classification by catalytic domain
Class I aaRSs have Rossman fold,
Class II aaRSs have antiparallel β-sheet fold

ClassI	Class II
Class Ia	**Class IIa**
MetRS ArgRS ValRS IleRS CysRS LysRS LeuRS	GlyRS HisRS ProRS ThrRS SerRS
Class Ib	**Class IIb**
GlnRS GluRS	AsnRS AspRS LysRS
Class Ic	**Class IIc**
TrpRS TyrRS	AlaRS GlyRS PheRS

During evolution, other domains were added, among which the anticodon binding domains. This results in the present classification, with sub-classes of more closely related aaRS, which generally possess similar anticodon binding domains in each group (here again, SerRS is special as it belongs to the same subgroup to ThrRS, from sequence similarity, although it has no anticodon binding domain). In the abundant sub class Ia, comprising aaRSs specific for methionine, isoleucine, valine, and arginine, the anticodon binding domains form structurally similar helical bundle domains. Five complex structures with the tRNAs have been solved so far (Table. I-3-2). In the helix bundle domain of class Ia aaRSs, anticodons binds between two helixes. Analysis of the sequences shows that apart from some key aminoacids, the conservation of residues is rather poor. Several residues in Helix I and Helix II were conserved but not always participated in tRNA binding. Especially the conserved residues in Helix II were thought to maintain quite similar local conformation in the helix bundle structure, important for anticodon binding (Fig. I-3-2). The reported structures showed no conserved amino acids that appear to be directly interacting with the tRNA.

Table I-3-2 aaRS-tRNA complex 3D structures of Claas Ia

Anticodon binding domain is helical bundle structure	
MetRS	*Aquifex aeolicus* Methionyl -tRNA synthetase and $tRNA^{Met}$ (2CSX)
IleRS	*Staphylococcus aureus* Isoleucyl-tRNA synthetase and $tRNA^{Ile}$ (1FFY)
ValRS	*Thermus thermophilus* Valyl-tRNA synthetase and $tRNA^{Val}$ (1GAX)
ArgRS	*Saccharomyces cerevisiae* Arginyl-tRNA synthetase and $tRNA^{Arg}$ (1F7U)
ArgRS	*Pyrococcus horikoshii* Arginyl-tRNA synthetase and $tRNA^{Arg}$ (2ZUE)
Cys	*Escherichia Coli* Cysteinyl-tRNA synthetase and $tRNA^{cys}$ (1UDB)
Leu	*Pyrococcus horikoshii* Leucyl-tRNa synthetase and $tRNA^{Leu}$ (1WZ2)
Leu	*Thermus thermophilu* Leucyl-tRNa synthetase and $tRNA^{Leu}$ (2VOG)

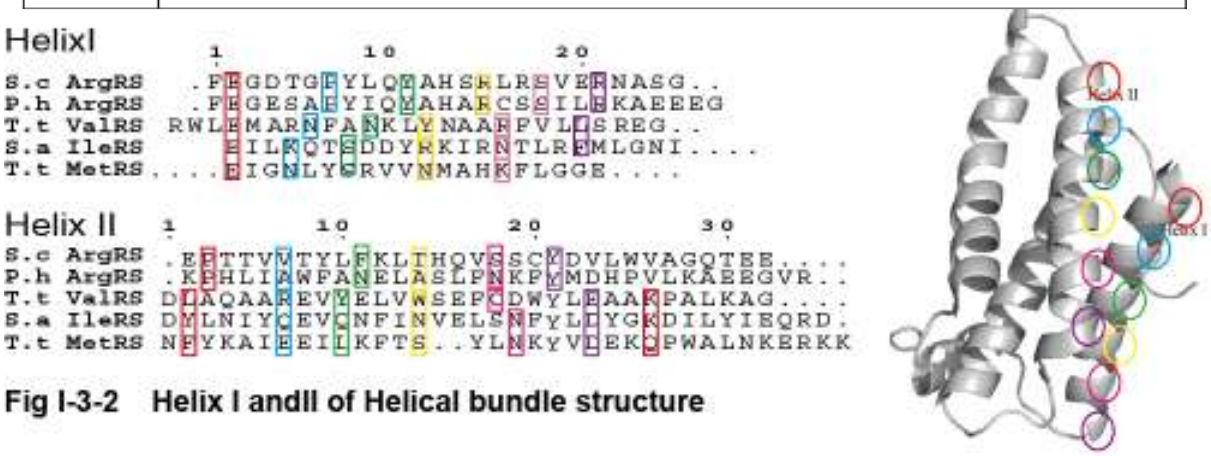


Fig I-3-2 Helix I andII of Helical bundle structure

The comparison of these structures shows interesting differences in the way that two anticodon binding domains of two different, but closely evolved aaRS such as MetRS and ArgRS, bind the corresponding anticodon binding domain (Fig. I-3-3).

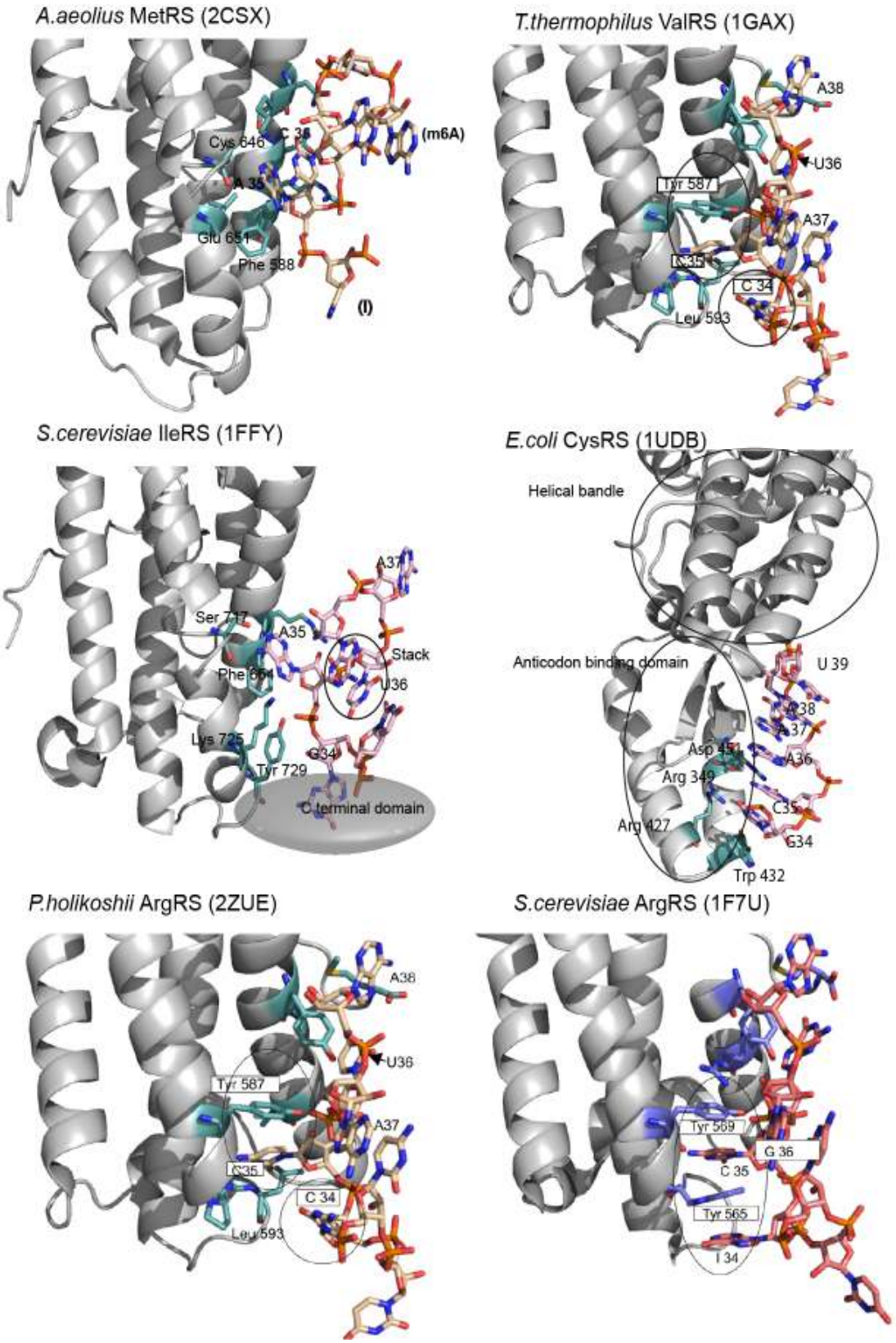


Fig. I-3-3 The comparison of interactions between anticodon binding domain and tRNA of MetRS, IleRS, ValRS and ArgRS that has a helix bundle structure.

Sub-class Ib contains GluRS and GlnRS. These overall structures are similar with each other but anticodon binding domain has different fold, GluRS anticodon binding domain is composed of only α-helix and anticodon bases were stacked with each other (Sekine, et al. 2001). On the other hand, the GlnRS anticodon binding domain is composed only β-strands and the anticodon bases of $tRNA^{Gln}$ are not stacked on each other but trapped in different isolated pockets on GlnRS (Fig. I-3-4). TrpRS and TyrRS have anticodon binding domain with a similar structure (Fig. I-3-4).

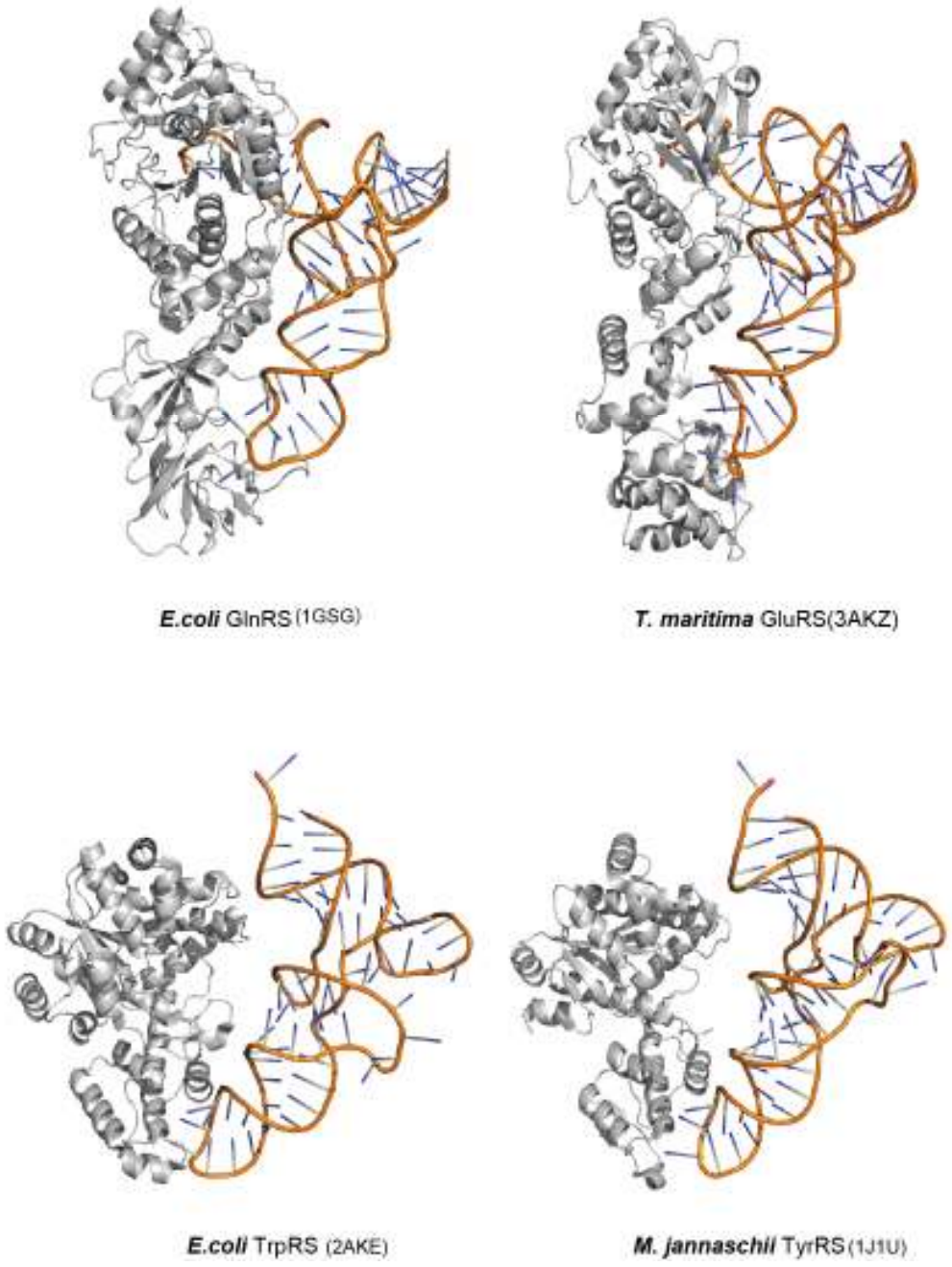


Fig. I-3-4 The structure comparison of GluRS and GlnRS, TrpRS and TyrRS.

Leucyl-tRNA synthetase (LeuRS) and Cysteinyl-tRNA (CysRS) synthetase are also belonging to Class Ia. The tRNA complexes of *P.horikoshii* and *T.thermophilus* LeuRS structures have been solved. These structures showed no interaction between the anticodon binding domain and anticodon. In leucine system, the tRNA comprises an additional arm, the variable arm, which is involved in the recognition (Fig.I-3-5). In CysRS-tRNA complex structure, interactions is observed between the anticodon and a helix bundle structure, but this helix bundle domain is not the same domain of that conserved in Class Ia aaRSs. The conserved helix bundle domain in Class Ia in CysRS has not contact with tRNA.

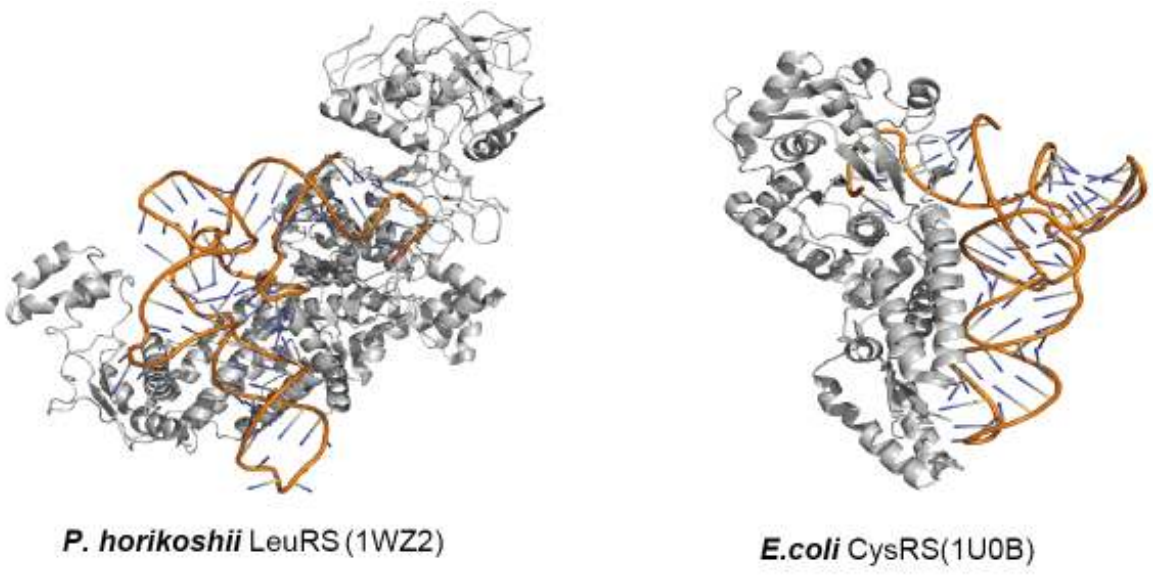


Fig. I-3-5 The structure of LeuRS and CysRS

Fig. I-3-6 shows other examples of such variability in class II, AsnRS, LysRS, AspRS have OB fold in their anticodon binding domain. GlyRS, HisRS, ProRS, ThrRS have a similar domain in its anticodon binding domain with an α/β fold specific for sub class IIa (Eriani et al. 1990, Chuck et al. 1998 Structure)(Fig I-3-7)

Belonging to the same sub class does not imply that the anticodon-binding domain is the same as mentioned before. For example ThrRS and SerRS, SerRS do not possess an anticodon binding domain as it does not interact with the anticodon, they have a long coiled coil extension at its amino terminus, which is used to recognize the large variable arm, a structural feature particular to $tRNA^{ser}$.

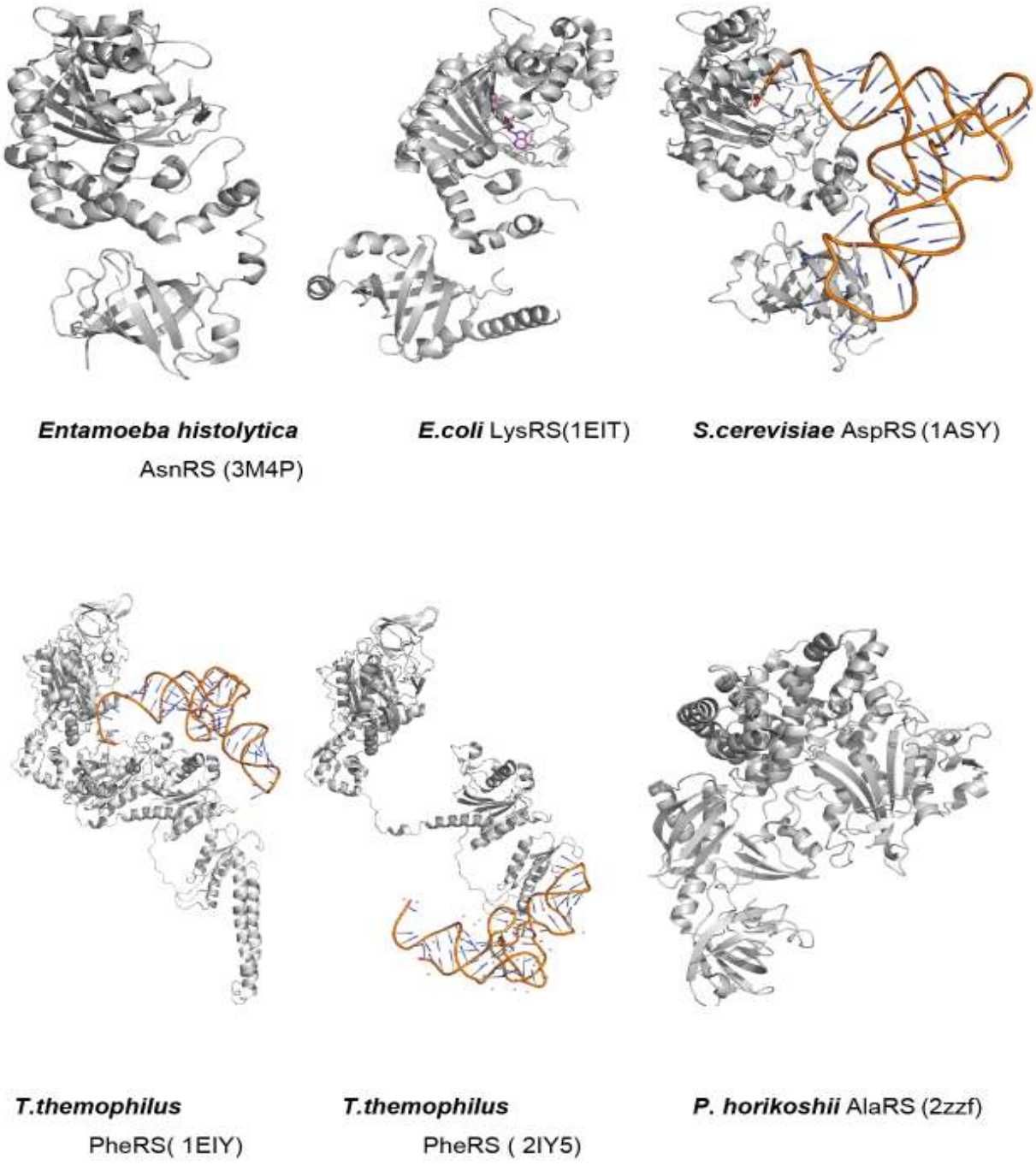


Fig. I-3-6 The structure of Class IIb and Class IIc aaRSs

AsnRS, AspRS, LysRS, have OB fold anticodon binding domain and belong to Class IIb, PheRS and AlaRS is belong to Class IIc

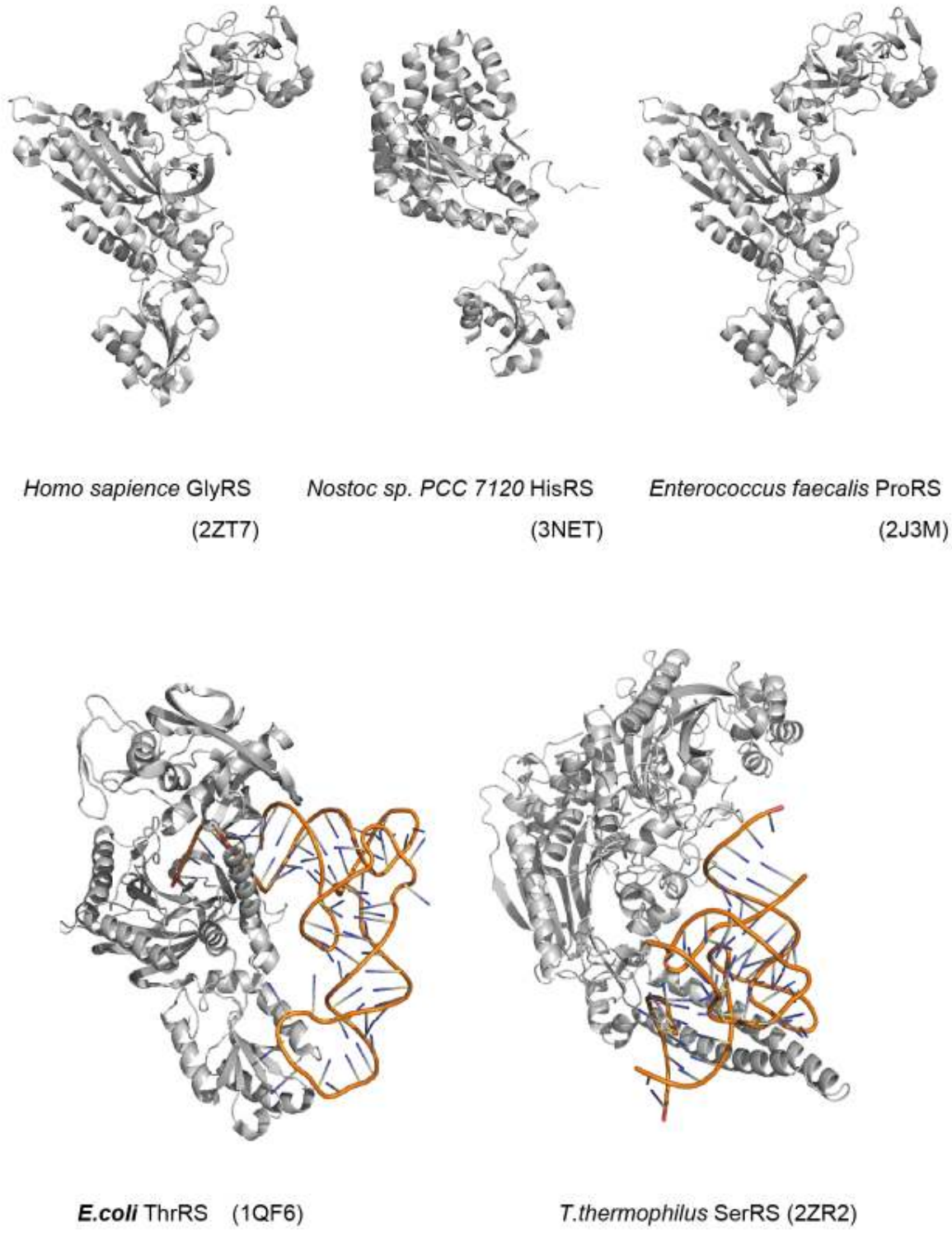


Fig. I-3-7 The structure of Claa IIa aaRSs

As I noted, so far many aaRSs have the interaction between anticodon of tRNA and anticodon binding domain. Recognition of tRNA is not only between anticodon and anticodon binding domain but additional parts of the protein recognize RNA. Important interactions with other regions in tRNA such as the extra arm, the inside of the L-shaped tRNA, and the

phosphate backbone have also been observed. For example, AlaRS contains a domain which recognizes the discriminate base-pair G:U in tRNA acceptor stem (Gabriel et al. 1996). Interestingly, this domain shows high similarity to the N-terminal editing domain of ThrRS, which also participates in $tRNA^{Thr}$ recognition (Fig. I-3-6). This domain also shows some similarity with the N-terminal domain of class I ArgRS which participates in $tRNA^{Arg}$ discrimination, with recognition at the level of the D-loop (detailed in Chapter II).

Deciphering of aaRSs 3D structures have highlighted the modular nature of these proteins and suggested that present proteins were built, during evolution, from various building blocks. Rossman fold, the catalytic center and primary building block of class I aaRSs, found in many proteins binding ATP, like deshydrogenases (Rao et al. 1973). Similarly, the OB fold, used as anticodon binding module such as AspRS and LysRS was previously identified in Oligonucleotide Binding domains, such as RNA polymerase subunit RBP8, the molybdate-binding protein (MOP)-like superfamily (Arcus 2002).

The 3D structures and structure-function relationships in aaRSs will be developed later in this manuscript. Class I aaRSs will be described in the Chapter II class II aaRSs will be described with the example of ThrRS (Chapter III). Before going further, the following point needs to be emphasized. This is, that from the crystal structures, we can apprehend only one picture, only one step of the RNA-protein encounter. This is particularly restrictive when dealing with transient interactions like tRNA and aaRS interaction. Indeed, aaRS bind and recognize tRNA, but immediately after the reaction has occurred in the catalytic domain, the tRNA unbinds and released. It can be expected that the aaRS and tRNA structures have a potential to change their conformation in each steps, in order to switch from one step to the next. To fully understand the binding process, greater number of snapshots, in particular those of native protein and its complex with tRNA, are required, but this could be obtained by crystallography in very few cases only such as AspRS and its complex with tRNA (Cavarelli et al. 1994). Alternative structural tool, which is able to investigate the molecular dynamics of biomolecules in solution is NMR, and another tool, which is able to investigate structural variety of biomolecules in freezed state is cryo-EM (when classes of images can be classified in alternative conformations). This is why other approaches, devoted to analyze the dynamics of binding are always an important asset in the functional investigation.

Before going into further description of our contribution to the understanding of the aaRS world, I will now briefly describe another RNA-binding module, the RRM, which is relevant to the other field approached in this manuscript.

4. Proteins with RRM domains

The RNA recognition motif (RRM), also known as RNA-binding domain (RBD) or ribonucleoprotein domain (RNP) is one of the most abundant protein domains in eukaryotes (this motif present about 0.5-1% of human genes). RRM containing proteins are involved in most post-transcriptional gene expression processing. These RNA recognition motifs were first identified in the late 1980s when it was demonstrated that mRNA precursors (pre-mRNA) and heterogeneous nuclear RNAs (hnRNAs) are always found in complex with proteins. Over the past 10 years biochemical and structural studies have shown that this domain in not only involved in RNA/DNA recognition but also in protein-protein interaction (Lunde et al. 2010). Most RRM proteins contain multiple RRMs that are thought intuitively to help achieving higher affinity and specificity considering the weak binding affinity of most RRMs in isolation for their RNA targets. RRM are able to bind RNA of different length with weak or high affinity and specificity and to contact several partners at the same time. Biochemical pathways in the cell often involve a highly sophisticated network of intermolecular interactions. The fact that RRMs can interact not only with nucleic acid but also with themselves and with other proteins could explain why RRM proteins are so numerous and achieve a central function in the cell. Structural data also show that RRMs have the capacity to interact using all the elements composing their structure (i.e. β-strands, loops, α-helices) explaining certainly the high conservation of this motif during evolution. Moreover, the small size of this domain also facilitates its incorporation into protein genes that could evolve toward a new and more sophisticated function (Lunde et al. 2010; Cléry et al. 2008).

Biochemical characterizations of the mRNA polyadenylate binding protein (PABP), which binds to the polyA tail of mRNAs and the hnRNP protein C shed light on a consensus RNA-binding domain of approximately 90 amino acids long, with a typical β1α1β2β3α2β4 RRM topology that forms a four stranded β-sheet packed against two α-helices (Fig. I-4-1(PABP full structure is in Fig I-4-2D)).

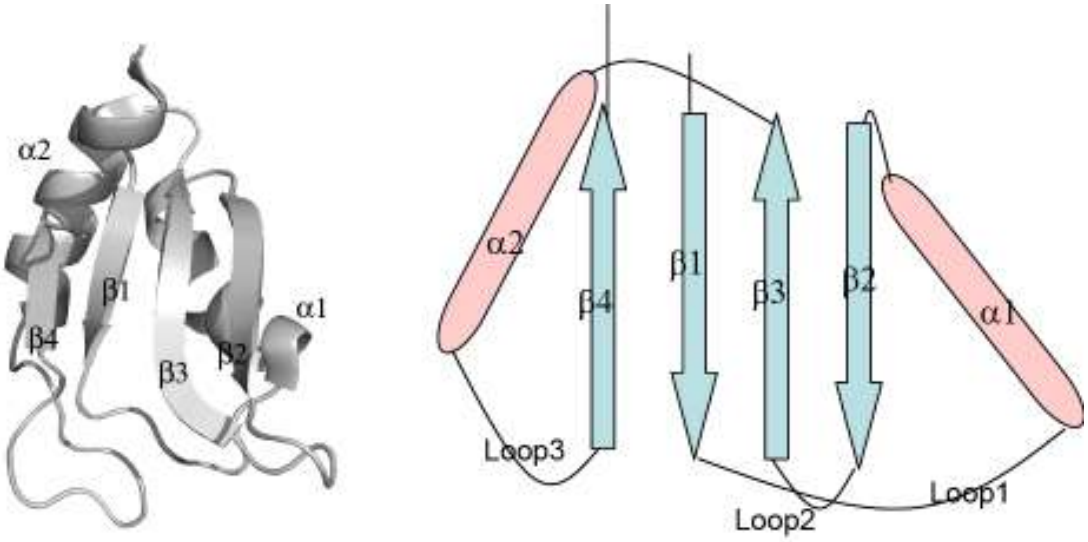


Fig. I-4-1 The typical RRM structure and topology

RRMs are containing some conserved residues that are mainly aromatic and positively charged in β-sheet 1, β-sheet 3, loop 1 and loop 2. These residues most commonly used to interact with single stranded RNA. Alignment is shown in Fig I-4-2. The examples of a complex with one RRM and RNA and two RRM and RNA were shown in Fig.I-4-2. Often multiple RRM (mostly 2 consecutive RRMs) are found in proteins. They combine together to increase the size of the binding surface, and to obtain new specificity. While the structure of RRM domain is well characterized, the mode of protein and RNA recognition by RRM is not always predictable and shows high variability. In Chapter III I will describe the part of my PhD project related to the RRM-containing protein LaRP7. In Fig. I-4-2.A and B. the sequence of LaRP7 RRM1 and RRM2 are aligned with other typical RRMs.

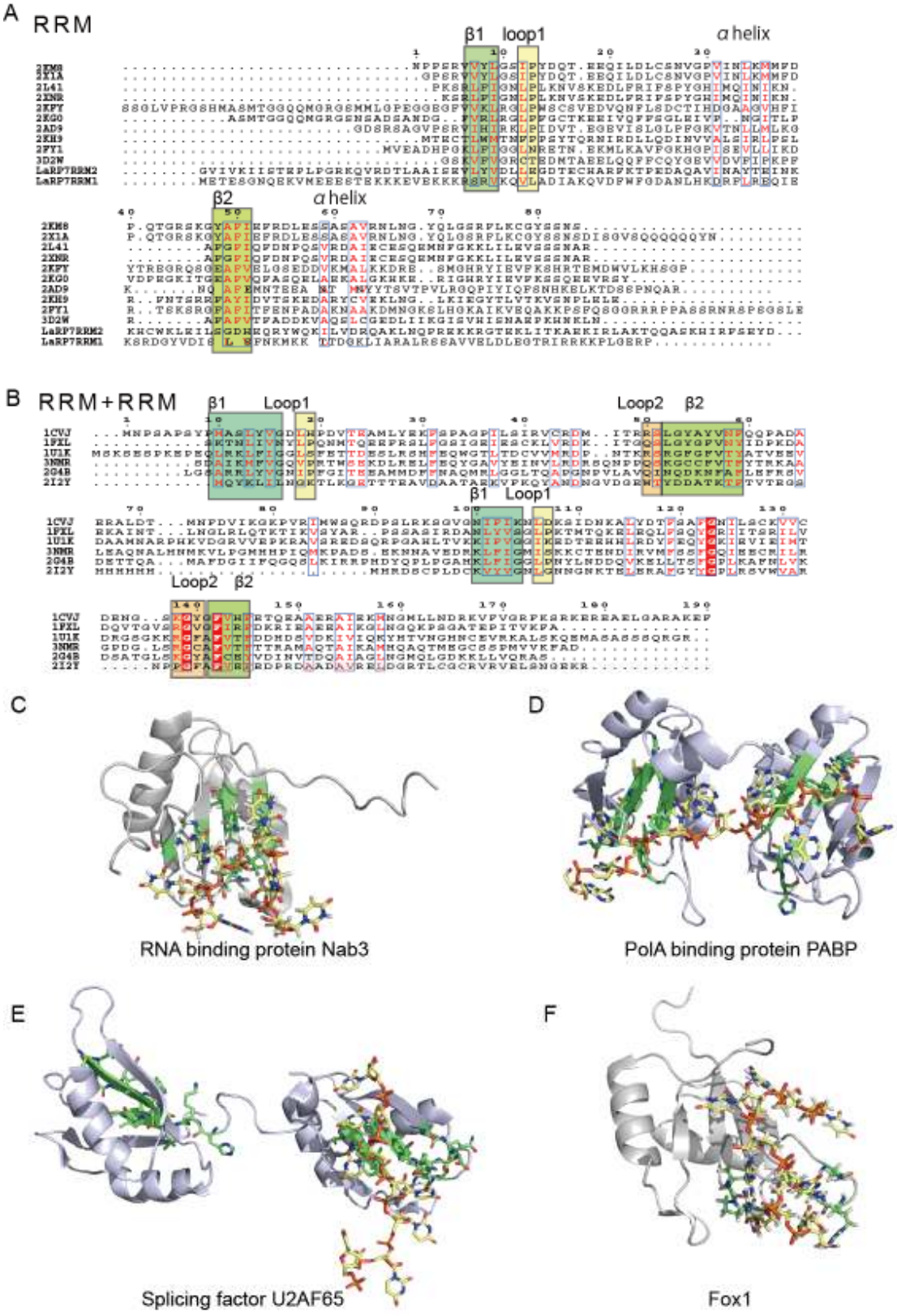


Fig. I-4-2 A,B Sequence alignment of RRM C,D,E,F is 3D structure of RRMs.

A) Sequence alignment of protein that has 1 RRM. B) Sequence alignment of protein that has two RRMs. C) Nab3 that has 1 RRM. D) PABP that has 2 RRMs. E) U2AF65 that has two RRMs and loop contributes binding. F) Fox1 that has one RRM and loop contributes binding.

Chapter II Aminoacyl-tRNA synthetases Structural and Kinetic analysis

1. Introduction

1.1. Aminoacyl-tRNA synthetase

The aminoacyl-tRNA synthetases (aaRSs) have fascinated biologists for a long time. They are the key of translation, the link between the worlds of protein and nucleic acid. Their structures and functions, which have both practical and basic significance, are deserving of and have received much attention. However, it is not only the structure-function aspect of these enzymes that has captured the biologist's imagination; it is also the possibility that they could tell us the secrets of the genetic code.

Aminoacyl-tRNA synthetases catalyze the first step of protein synthesis in all organisms. They are responsible for the precise attachment of amino acids to their cognate transfer RNAs (tRNA). Once an aminoacyl-tRNA has been synthesized, amino acid part makes no contribution to accurate translation of the mRNA. The amino acid is passively driven by its tRNA and inserted into a growing peptide chain following codon-anticodon recognition between the mRNA and tRNA. There are twenty different types of aaRSs, unique to each amino acid (two additional aminoacyl-tRNA synthetases; Selenocysteinyl-tRNA synthetase and Pyrrolidyl-tRNA synthetase)

		Second			
		U	C	A	G
First	U	UUU (Phe/F) UUC (Phe/F) UUA (Leu/L) UUG (Leu/L)	UCU (Ser/S) UCC (Ser/S) UCA (Ser/S) UCG (Ser/S)	UAU (Tyr/Y) UAC (Tyr/Y) UAA Ochre UAG Amber	UGU (Cys/C) UGC (Cys/C) UGA Opal UGG (Trp/W)
	C	CUU (Leu/L) CUC (Leu/L) CUA (Leu/L) CUG (Leu/L)	CCU (Pro/P) CCC (Pro/P) CCA (Pro/P) CCG (Pro/P)	CAU (His/H) CAC (His/H) CAA (Gln/Q) CAG (Gln/Q)	CGU (Arg/R) CGC (Arg/R) CGA (Arg/R) CGG (Arg/R)
	A	AUU (Ile/I) AUC (Ile/I) AUA (Ile/I) AUG (Met/M)	ACU (Thr/T) ACC (Thr/T) ACA (Thr/T) ACG (Thr/T)	AAU (Asn/N) AAC (Asn/N) AAA (Lys/K) AAG (Lys/K)	AGU (Ser/S) AGC (Ser/S) AGA (Arg/R) AGG (Arg/R)
	G	GUU (Val/V) GUC (Val/V) GUA (Val/V) GUG (Val/V)	GCU (Ala/A) GCC (Ala/A) GCA (Ala/A) GCG (Ala/A)	GAU (Asp/D) GAC (Asp/D) GAA (Glu/E) GAG (Glu/E)	GGU (Gly/G) GGC (Gly/G) GGA (Gly/G) GGG (Gly/G)

Fig. II-1-1 Genetic code of amino acid

These aaRSs have been divided into two classes depending on their catalytic domain. Class I has Rossmann fold whose active sites contain a canonical dinucleotide-binding fold with four or five stranded parallel β-sheet, display two signature amino acid sequences

‘HIGH’ and ‘KMSKS’ and Class II aaRSs are built around an antiparallel β-sheet partly closed by helices and containing three characteristic motifs. The topology of each active site is the structural constraint governing the conformation of the ATP molecule and conferring a region specificity to the second step of the reaction. Class I aaRSs aminoacylate the 2′-OH of the 3′- terminal adenosine of the tRNA, while class II aaRSs aminoacylate the 3′-OH (Eriani et al. 1993).

AaRSs catalyze the aminoacylation reaction. Now this reaction is thought to occur in two steps (Amez et al. 1997). This reaction requires amino acid activation through condensation of the amino acid with ATP to form an aminoacyl-adenylate. Then, the activated amino acid is transferred to the ribose at the 3′-end CCA of the cognate tRNA to form aminoacyl-tRNA.

The amino acid (aa) is first condensed with ATP to form a tightly bound aminoacyl-adenylate (aa–AMP) in the presence of Mg^{2+}, and inorganic pyrophosphate (PP) is released: [aa + ATP + aaRS → aaRS (aa–AMP) + PP (Reaction I)]. The activated aa-AMP is then transferred from the adenylate to the 3’end of the tRNA to form aminoacyl-tRNA [aaRS + aa-AMP + tRNA →aa-tRNA+ AMP + aaRS (Reaction II)] (Fig. II-1-2)

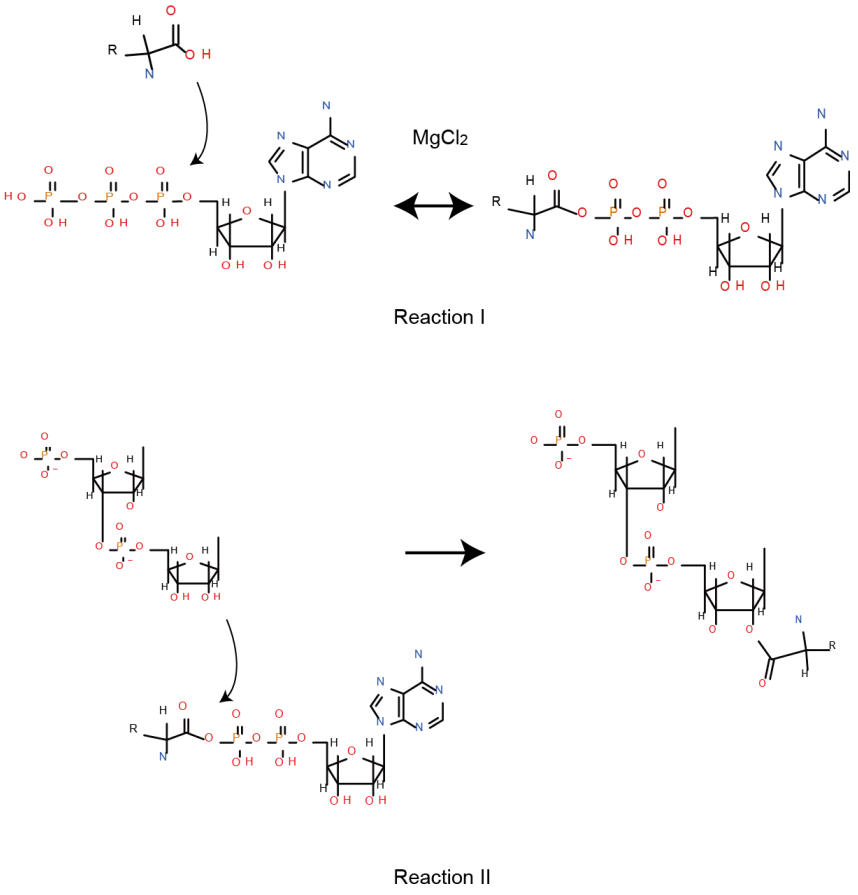


Fig. II-1-2 Two steps aminoacylation reaction

For most aaRSs, the formation of aminoacyl-AMP does not require tRNA. On the other hand, for Arginyl-tRNA synthetase, and Glutaminyl-tRNA synthetase from all species investigated, and Glutamyl-tRNA synthetase from *Escherichia coli*, the ATP-PPi exchange reaction corresponding to the formation of aminoacyl-AMP and its reverse reaction, has never been observed without tRNA. In the presence of cognate tRNA, the ATP-PPi exchange reaction was observed for ArgRS, GlnRS, and GluRS. The error rate in tRNA aminoacylation is extremely low and even for related amino acids it does not appear to exceed 1 per 10000 under normal condition.

The accurate flow of genetic information during DNA replication and protein biosynthesis cannot be explained in terms of conventional specificity-related mechanisms. Therefore, additional proofreading mechanisms operating at several levels are required.

About half of the aaRSs added an editing function, which enables removal of the wrong amino acid from its cognate tRNA.

To discriminate strictly each similar amino acid (i.e. Ile Val, some aaRS possess a special error-correcting mechanism that hydrolyzes the misacylated tRNA to yields free amino acid and deacylated tRNA referred to as post-transfer editing. The hydrolysis of aminoacyl-AMP referred as pre-transfer editing has also been hypothesized.

There are recently accumulated numbers of reports suggesting that aaRSs are capable of functioning in a broad repertoire of other cellular activities than aminoacylation, including tRNA processing, RNA splicing, RNA trafficking, apoptosis and transcriptional and translational regulation (Guo et al. 2010, Szymański et al. 2000; Woese et al. 2000)

Before going into details about ArgRS and ThrRS, all aminoacyl-tRNA structures with and without ligands were picked up from PDB in Table II-1-1 (until 2010). Many structures are reported so far, but still structural information of complex structure with tRNA is required.

Table. II-1-1 Structures of Aminoacyl-tRNA synthetase with or without ligands

Class I	*Origin*	*Ligand*	*PDB ID*
MetRS	*Aquifex aeolicus*	tRNA tRNA, methionyl-adenyate	**2CSX** **2CT8**
	Thermus Thermophilus	---- PEG6000	**2D54** **2D5B**
	Leishmania major	Methionyl-adenylate-phyrophosphate	**3KFL**
	Escherichia coli	Methionine Azidonorleucine	**3H99** **3H9B**
	Pyrococcus abyssi	-----	**1ROG**
	Mycobacterium smegmatis	Methionine Adenosine	**2XIL**
ArgRS	*Saccharomyces cerevisiae*	tRNA Arg -----	**1F7U** **1BS2**
	Pyrococcus horikoshii	tRNA AMP-PNP	**2ZUE**
	Thermus thermophilus	------	**1IQ0**
	Campylobacter jejuni	(putative)	**3FNR**
IleRS	*Staphylococcus areus*	tRNA,Mupirocine Editing substrate analog Ile adenocine analog Mupirocine	**1FFY,1QU2** **1OBH** **1JZQ** **1JZS**
	Thermus thermophilus	tRNA ------	**1ILE** **1GAX**
ValRS	*Thermus Thermophilus*	tRNA Valyl adenylate analog -----	**1GAX,1IVS** **1IYW**
LeuRS	*Pyrococcus horikoshii*	tRNA ------	**1WZ2** **1WKB**
	Thermus thermophilus	tRNA Sulphamoyl analogue of Leu adenylate -----	**2VOG** **2VOC** **1OBC**
CysRS	*Escherichia coli*	tRNA -----	**1UDB** **1LJ5**
	Giandia lamblia	Poly adenylate	**3IAL**

GlnRS	*Escherichia coli*	tRNA	**1GTS**
		tRNA Gln	**1QTQ**
		tRNA ATP	**1GSG**
		------	**1NYL**
	Deinococcus radiodurance	------	**2HZ7**
GluRS	*Thermus thermophilus*	tRNA	**2DXI**
		Glu ATP	**2CUZ**
		Glutamate	**1GLN**
	Thermotoga maritime	tRNA Glutamyl-AMP	**3AKZ**
		------	**2O5R**
	Mycobacterium tuberculosis		**2JA2**
	Methanothermobacter Tthermautrophicus		**3AII**
TrpRS	*Homo sapiens*	tRNA Tryptophamid ATP	**2QUI**
		Trp AMP	**2QUJ**
	Yersinia pestis CO92	------	**3N9I**
	Campylobacter jejuni	-------	**3M5W**
	Aeropyrum pernix K1	-------	**3AO4**
		Trp	**3AO5**
	Pyrococcus horikoshii	Trp AMP	**3JXE**
	Mycoplasma pneumonia	Trp AMP	**3FI0**
		------	**2YY5**
	Thermotoga maritime	------	**3FHJ**
	Geobacillus stearothermophilus	------	**1I6M**
	Saccharomyces cerevisiae	Trp AMP	**3KT3**
		Trp	**3KT8**
		-------	**3KT0**
Tyr	*Methanocaldococcus janaschii*	tRNA, Tyrosine	**1J1U**
		-----	**1V7D**
	Thermus thermophilus	tRN, ATP, Tyr	**1H3E**
	Saccharomyces cerevisiae	tRNA	**2DLC**
	Escherichia coli	Tyr, AMS	**1VBM**
		Tyrosine	**1WQ4**
		tRNA, Tyr	**1X8X**
	Human mitochondrial	------	**2PID**
	Staphylococcus areus	SB-219383	**1JII**
	Homo sapiens	-----	**1Q11**

	Mycobacuterium tuberculosis	------	**2JAN**
	Pyrococcus horikoshii	------	**2CYC**
	Archaeoglobus fulgidus	------	**2CYB**
	Aeropyrum pernix	------	**2CYA**
	Staphylococcus phagetwort	Intron I RNA	**2RKJ**
	Acanthamebapolyphagemimivirus	-----	**2J5B**
Class II			
SerRS	*Thermus thermophilus*	tRNA	**1SER**
	Mammarian mitocondorial	Seryl adenilate	**1WLE**
	Typanosoma brucei	ATP -----	**3LSS** **3LSQ**
	Pyrococcus horikoshii	ATP	**2ZR2**
	Methanocarcina barkeri	ATP -----	**2CIM** **2CJA**
ProRS	*Giardia lamblia*	-----	**3IAL**
	Enterococcus faecalis	Prorinal	**2J3M**
	Methanothermobacter	tRNA Proline sulfamoyl adenilate	**1NJ2** **1NJ5**
	Thermantotrophicus	tRNA -----	**1H4Q** **1H4T**
	Rhodosendomonas palustris	Pro AMS Cys AMS ATP	**2I4M** **2I4N** **2I4O**
	Thermus thermophilus	tRNA ATP Prolinol tRNA Prolyl adenilate	**1H4L** **1H4S**
HisRS	*Nostoc sp pcc 7120*	-----	**3NET**
	Trypanosoma brusei	-----	**3HRI**
	Escherichia coli	His adenylate	**2EL9**
	Thermoplasma acidophilum	-----	**1WU7**
	Thermus thermophilus	-----	**1H4V**
	Trypanosoma cruzi	His adenylate	**3HRK**
GlyRS	*Homo sapiens*	Ap4A AMP-CPP Gly ATP	**2ZT5** **2ZT6** **2ZT7**

		Gly-adenylate	**2XT8**
	Thermus thermophilus	----- Gly-adenyte	**1ATI** **1GCM**
ThrRS	*Escherichia coli*	tRNA mRNA operator	**1QF6** **1KOG**
	Staphylococcus aureus	ATP	**1NYR**
	Aerophyrus pernix	ThrRS-1	**3A31**
PheRS	*Staphylococcus haemolyticus*	---	**2RHQ**
	Homo sapiens	----	**3L4G**
	Porphyromonas gingivalis	----	
	Human mitocondorial	m-tyrocine	**3HFV**
	Thermus thermophilus	m-tyrocine tRNA Sulphamoyl adenosine	**3HFZ** **2RHQ** **2ALY**
	Bacteroides flagilis(putative)	----	**3IG2**
Ala	*Pyrococcus horikoshii*	(without origomelic domain)	**2ZZE**
	Aquifex aeolicus	(without origomelic domain)	**2ZTG**
LysRS	*Bacillis stearothermophilus*	AP4A Lysil sulphamoyl adenosine	**3A74** **3E9H**
	Homo sapiens	-----	**3BJU**
	Pyrococcus horikoshii	-----	**1IRX**
	Escherichia coli	-----	**1BBU**
AspRS	*Saccharomyces cerevisiae*	tRNA -----	**1ASY** **1EOV**
	Escherichia coli	tRNA(yeast) -----	**1IL2** **1EQR**
	Thermococcus kodakaraensis	-----	**3NEN**
	Thermus thermophilus	-----	**1LOW**
	Thermus thermophilus	tRNA(E.coli)	**1EFW**
AsnRS	*Entamoeba histolytica*	Asparaginyl-adenilate	**3N4P**
	Pyrococcus horikoshii	Asparaginyl-adenilate	**1X54**

1.2. Arginyl-tRNA synthetase (ArgRS)

ArgRS is belonging to Class Ia. The 3-dimensional structure analysis of ArgRS in the form of tertiary complex ArgRS-$tRNA^{Arg}$(ICG) Arginine from *Saccharomyces cerevisiae* was reported by Cavarelli's group (Delagoutte et al. 2000). The structure of ArgRS can be divided into five domains; catalytic domain which contains Rossmann fold, helical bundle anticodon binding domain, additional N terminal domain, inserted domain 1 (N terminal side of catalytic domain) and inserted domain 2 (C-terminal side of catalytic domain)(Fig. II-1-1). In addition to Rossmann fold domain and helix-bundl anticodon binding domain which are characteristic structure of class Ia enzymes, ArgRS has un unique domain at its N-terminus. ArgRS and MetRS do not have editing domain (Cavarelli et al. 1998).

1.2.1. Aminoacylation reaction on ArgRS

It is intriguing that the three Class I synthetases ArgRS, GlnRS, and GluRS, as well as the exceptional class I LysRS, do require tRNA to accomplish the formation reaction of aminoacyl-AMP. The absolute requirement for tRNA in the ATP-PPi exchange reaction has been described for ArgRS from different sources (from the bacteria *Escherichia coli*, *Bacillus stearothermophilus* or *Mycobacterium smegmatis*, from the lower eukaryotes *Saccharomyces cerevisiae* or *Neurospora crassa* and from mammals) and should be a general feature of this enzyme (Mitra et al. 1966, Sequential 1974, Gangloff et al. 1976, Godeau 1980, Lazard et al. 2000).

Previous studies have stressed the importance of the terminal adenosine in the ATP-PPi exchange reaction of *E. coli* ArgRS (Alan H M. et al. 1967). A mutant tRNA with a deletion of A76, or a complete tRNA molecule with A76 displaying a 2', 3' cyclic phosphate group generated after cleavage by the HDV ribozyme, was not able to activate Arginine. These data strongly suggest that the terminal adenosine is essential to convert Arginine into its activated mode (i.e. Arg-AMP). Also in the presence of the tRNA treated with periodate, which oxidizes the 2'-OH and 3'-OH groups of the ribose of A at the 3'-end to convert them into dialdehyde groups, the ArgRS, GlnRS and GluRS enzymes were incapable of catalyzing the ATP-PPi exchange reaction. The hydroxyl group of the ribose of the 3'-terminal A76 of tRNA is essential for the ATP–PPi exchange reaction on ArgRS, GlnRS, and GluRS. In addition, in human ArgRS, ATP-PPi exchange reaction and aminoacylation reaction was measured with tRNA mutants and mini-helices(acceptor stem part of tRNA). The mini-helices were unable to initiate the amino acid activation in human ArgRS (Guigou et al.

2005). *E. coli* GlnRS is able to aminoacylate a seven-base pair RNA micro helix mimicking the acceptor stem of $tRNA^{Gln}$ (Wrigh et al. 1993). Hence arginine-activation requires the full length tRNA molecule.

Requirement of $tRNA^{Arg}$ for the ATP-PPi exchange reaction on ArgRS was found in 1960s and reasonable explanation about why $tRNA^{Arg}$ is required for this reaction has not been yet given. This thesis gives the most reasonable answer to the remaining question, namely how $tRNA^{Arg}$ is involved in this reaction. In the chemical point of views, this thesis deals with molecular aspects of the reaction on the surface of the protein and chemical forces inducing the reaction on the basis of structural analysis.

1.2.2. tRNA recognition of ArgRS

The bases at positions 35^{th} and 36^{th} of tRNA (2^{th} and 3^{th} of the anticodon) are major identity elements in *E. coli, S. cerevisiae* and *Thermus thermophilus*. Adenosine at position 20^{th} of the D-loop in the corner of the convex side of $tRNA^{Arg}$ was reported to contribute to recognition in most organisms with the noticeable exception of lower eukaryotes and of the mitochondria of higher eukaryotes that have Dhu in place of Ade at this position. The aminoacylation experiment using mutant tRNA from *E. coli* suggested that Ade20 of D-loop of $tRNA^{Arg}$ binds to the N-terminal domain as well as ArgRS has uniquely additional N-terminal domain. The interaction between D-loop and N-terminal domain ofArgRS may be involved in recognition of $tRNA^{Arg}$ (McClain et al. 1988, Schulman and Pelka 1989a, Tamura et al. 1992, Saks 1998, McClain et al. 1990).

Indeed the crystal structure of yeast ArgRS-$tRNA^{Arg}$ revealed interactions between the side chains of Asn106 and Gln111 in the N-terminal domain and the base of Dhu20 of the D-loop of $tRNA^{Arg}$(ICG) which is specific for *S.cerevisiae*. On the other hand, it was reported that *K*m and *k*cat values in the aminoacylation reaction for tRNA of Asn106Ala and Gln111Ala mutant proteins of *S.cerevisiae* ArgRS were the same as those of the wild type ArgRS (Cavarelli 2003) indicating that this mechanism may have only weak contribution to specificity.

1.2.3. Modification of tRNA on recognition of tRNA

Modifications of nucleoside are important for recognition by proteins. Indeed, natural tRNAs with posttranscriptional modification at conserved positions, particularly, at wobble positions 34^{th} and 37^{th}, have been reported (Agris 2004).

For example m^2A37 is important for recognition of $tRNA^{Gln}$ and the modification lysidine, k^2C34, of $tRNA^{Ile}$ with methionine anticodon CAU plays important role in recognition by IleRS.

A slipped recognition mechanism of ArgRS recognizes $tRNA^{Asp}$(GUC) transcript principally via C36 and G37, while $tRNA^{Arg}$(UCU), is recognized via C35 and U36. This is dependent on the methylation of G37 to methylguanosine, which prevents aminoacylation of $tRNA^{Asp}$(GUC) by ArgRS (Pütz et al. 2007, Sissler et al. 1996). This alternative recognition of C36 and unmodified G37 in $tRNA^{Asp}$ by ArgRS is strictly dependent on the sequence context of these nucleotides (Sissler et al. 1998). *S.cerevisiae* tRNA was modified at first position of anticodon; 34 C to I, and this modified nucleotides of $tRNA^{Arg}$ also contribute to the capacity of $tRNA^{Arg}$ to activate (the ATP-PPi exchange reaction) and perform the catalysis (Perret et al. 1990).

1.2.4. Other aspects

Although the mode of $tRNA^{Arg}$ binding seems to be well conserved from prokaryotes to eukaryotes, one of the major differences characterizes aminoacyl-tRNA synthetases from higher eukaryotes (from Drosophila to human), is the presence of polypeptide chain extensions that may serve as auxiliary tRNA-interacting factors.

Human cytoplasmic ArgRS is a component of a macromolecular complex consisting of at least nine aminoacyl-tRNA synthetases (IleRS, LeuRS, MetRS, GlnRS, ArgRS, LysRS, AspRS, GluRS, ProRS). and three auxiliary proteins (Quevillon et al. 1999). In mammalian cells, one genome is used to produce two forms of ArgRSs; free ArgRS enzyme and N-terminal extension containing isoform. N-terminal eukaryotic specific polypeptide extension is required for assembly of ArgRS within the multi-synthetase complex. N-terminal moiety of the auxiliary protein p43 associates with the N-terminal extension of ArgRS (Zheng et al. 2006). P43 is specifically bound to ArgRS and is homologous to Arc1p. Arc1p is a yeast general RNA-binding protein that associates with MetRS and GluRS and has the role of tRNA-binding cofactor (Simos et al. 1998). On the other hand, p43 may not modulate tRNA aminoacylation properties (Guigou et al. 2004).

In *S.cerevisi*ae, mutated ArgRS was identified in which Ile is bound at position 570 in place of Val. In the crystal structure of the complex of wild type ArgRS with $tRNA^{Arg}$(ICG)and Arg, the amide nitrogen of Val 570 likely forms a hydrogen bond with Oxygen of C35 in the anticodon loop. The aminoacylation efficiency of this mutated ArgRS

declined largely for $tRNA^{Arg}$(CCG), compared with that for wild type ArgRS. Interestingly the La protein orthologue Lhp1p, partially rescues the mutants from the severe declination of aminoacylation for $tRNA^{Arg}$(CCG). This suggests that the synthetase functions redundantly with Lhp1p to stabilize the fragile anticodon stem of $tRNA^{Arg}$(CCG) in the correct conformation (Copela et al. 2006).

2. *P.horikoshii* ArgRS Structural and Kinetic analysis

2.1. Structure of ArgRS

The structures of the ternary complex (*P.horikoshii* ArgRS, $tRNA^{Arg}$(CCU), AMP-PNP) and the binary complex (*P.horikoshii* ArgRS and $tRNA^{Arg}$(CCU)) were obtained with R_{factor} = 0.213 at 2.0 Å resolution and R_{factor} = 0.201 at 2.3 Å, respectively (PDBID: 2ZUE, 2ZUF). The structure of newly obtained ternary complex of *P.horikoshii* ArgRS is compared with *S.cerevisiae* ArgRS bound arginine and $tRNA^{Arg}$(ICG) (Protein Data Bank ID:1F7V) in order to clarify interaction between ArgRS protein and tRNA, that is, between N terminal domain of ArgRS and D loop of tRNA, between the anticodon binding domain of ArgRS and the bases of anticodon of tRNA and between the catalytic domain of ArgRS and bases of G73C74C75A76 of tRNA (Fig. II-2-1, Fig. II-2-2).

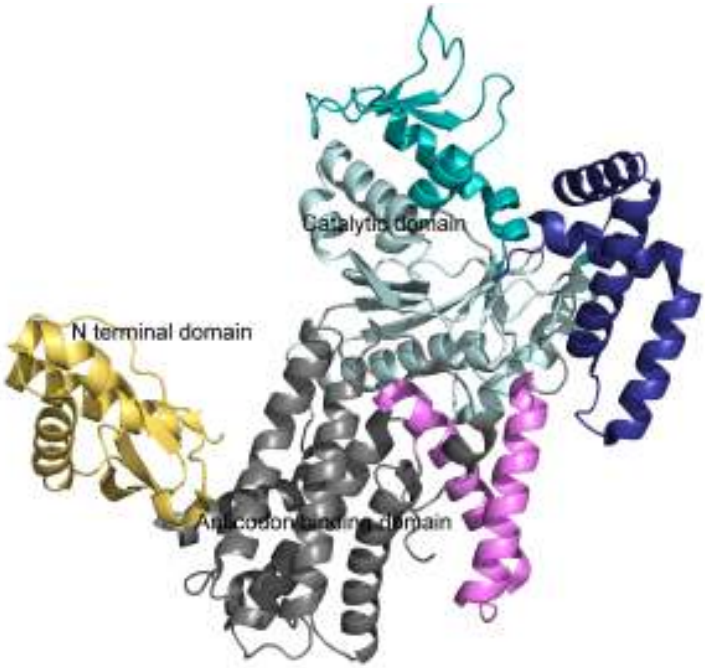


Fig. II-2-1 Organization of Arginyl-tRNA synthetase from *P.horikoshii* (2ZUE)

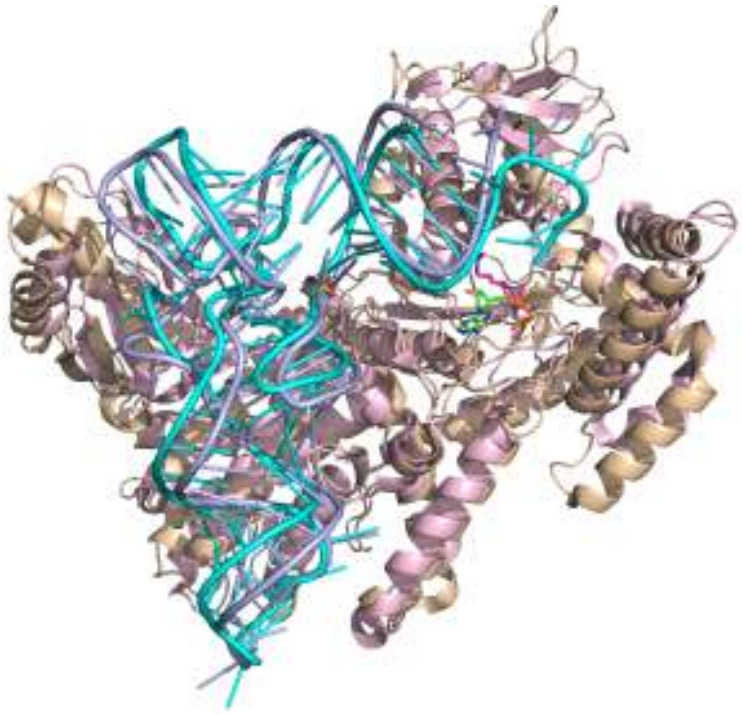


Fig. II-2-2 Superimposition of *S.cerevisiae* ArgRS (1F7U) (Pink:ArgRS, Cyan $tRNA^{Arg}$(ICG)) and *P.horikoshii* ArgRS (2ZUE)(Salmon Pink :ArgRS. Slate $tRNA^{Arg}$(CCU))

2.1.1. N-terminal domain

Adenosine at position 20 of the D-loop of $tRNA^{Arg}$ was reported to contribute to the recognition in most organisms with the exception of lower eukaryotes and of the mitochondria of higher eukaryotes that have D or C.

Crystal structures of binary and ternary complexes of ArgRS and $tRNA^{Arg}$(ICG) from *S.cerevisiae* and Arg revealed that base D20 in the D-loop, which is specific to *S.cerevisiae* $tRNA^{Arg}$(ICG), is positioned in close proximity to the side chains of Asn106, Phe109, and Gln111, which are included in the characteristic N-terminal domain of ArgRS.

In *P.horikoshii* ArgRS, the base of G19 interacts with the hydrophobic side chains of Pro44 and Phe47 in the N-terminal domain. The base of A20 is packed into the hydrophobic space surrounded by the side chains of Val82 and Tyr85 in the turn (Val82-Asn83- Gly84-Tyr85) and the hydrophobic side chains of Pro34 and Leu38. The base of A20 lies close to the side chain of Asn87. The plane of the base of A20 and the end plane of the carbamoyl group of Asn87 are out of the coplanar orientation. In particular, carbamoyl group of Asn87 is positioned far out of the base plane. Large values B-factors of residues in the N-terminal domain (average B-factors of residues 2–118 in the N-terminal domain and residues 119–629 in other domains are 49.9 $Å^2$ and 29.5 $Å^2$, respectively) indicate that the

D-loop does not make a stable contact with the N-terminal domain. Both A20 of $tRNA^{Arg}$ and D 20 of $tRNA^{Arg}$ interact with the hydrophobic area of N-terminal domain. (Fig. II-2-3)

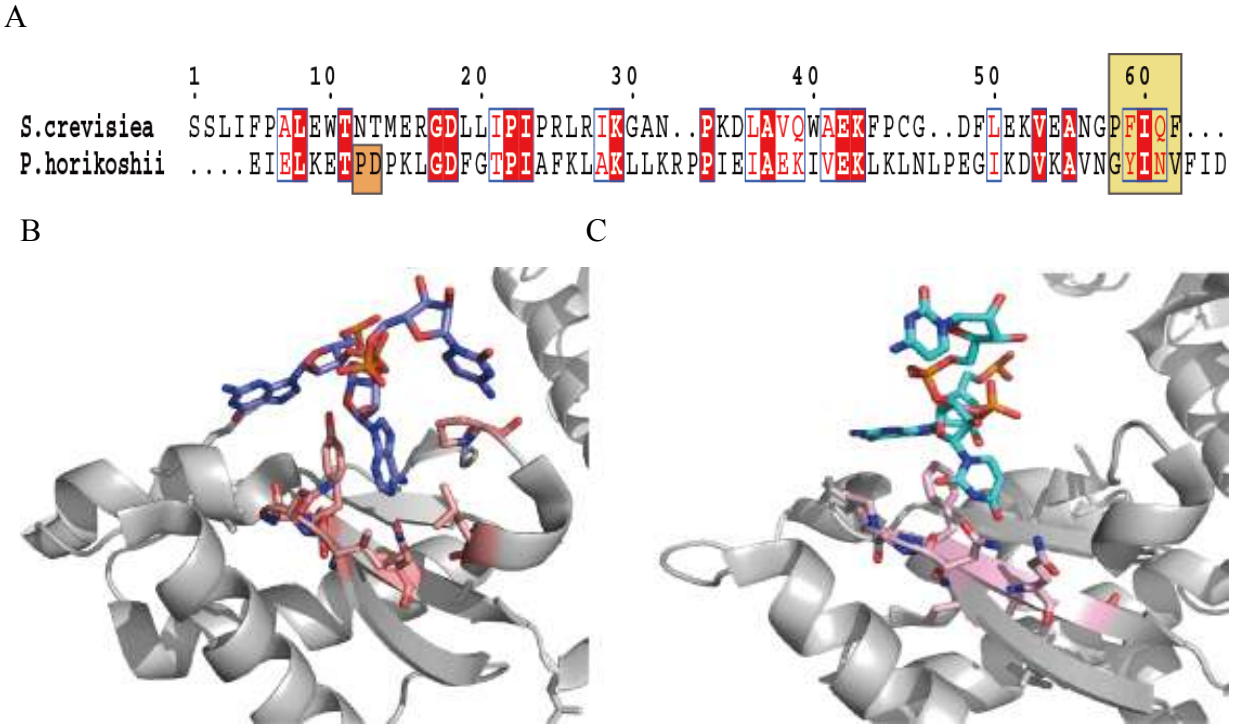


Fig. II-2-3 Interaction between D-loop of tRNA and N-terminal domain of ArgRS.

A Sequence alignment of N-terminal domain of *S.cerevisiae* ArgRS and *P.horikoshii* ArgRS.
B Interaction between A20 of tRNA and N-terminal domain of *P.horikoshii* ArgRS.
C Interaction between Dhu20 of tRNA and N-terminal domain of *S.cerevisiae* ArgRS.

N-terminal domain characteristic for ArgRS contains a core structure consisting of the β-sheet of four antiparallel β-strands and three helices on the N-terminal side and a long helix and a loop continuing to the catalytic domain of ArgRS. Its mode reminds of RNA binding on RNA-recognition motif (RRM) with in the hydrophobic surface of the β-sheet consisting of antiparallel four β-strands. The structure of this motif is stabilized by two helices on one side of β-sheet. N-terminal domain of ArgRS is also stabilized by two helices on one side. The mechanism used for stabilizing β-sheet of N-terminal domain is very similar to that of RRM. The side interacting with the base Dhe20/A20 contains aromatic residue of F/Y and carbamoyl-containing residue Q/N. Q/N residue is on β4 strand. The position of β4 strand in the β-sheet of N terminal domain corresponds to position of β3 strand of RRM. The aromatic residues are on β3 strand. In *P.horikoshii* ArgRS, Tyr85 corresponds to the aromatic residue (Fig. II-2-4).

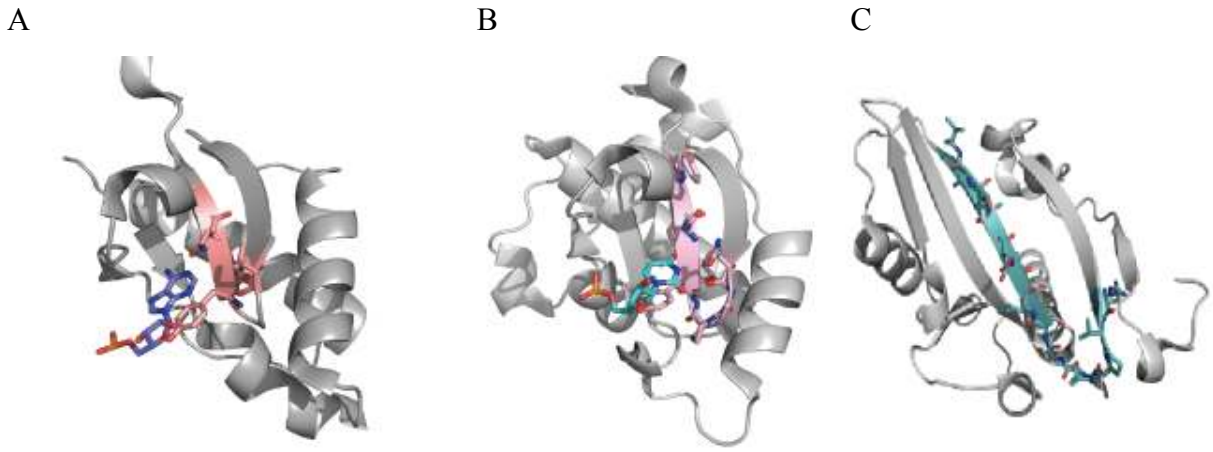


Fig. II-2-4 Comparison of N-terminal domains

A N-terminal domain of *P.horikoshii* ArgRS., B N-terminal domain of *S. cerevisiae* ArgRS., C C-terminal domain of GAPDH

From the analysis of the N-terminal domain, the Cavarelli group (Cavarelli et al. 1998) shows structural analogy between N-terminal domain of ArgRS and C-terminal domain of GAPDH (Glyceraldehyde 3-phosphate dehydrogenase). C-terminal region of GAPDH is also stabilized by two helices. This structure ascribes thermodynamic stability and is commonly used by several other proteins including muramoyl-pentapeptide carboxypeptidase (Protein Data Bank ID: 1LBU) and the oligomerization and L-arginine-binding domain of arginine receptor (Protein Data Bank ID: 1XXA).

S.cerevisiae ArgRS has an eukaryote specific extension of 30 residues at N-terminus making one α-helix. The mammalian ArgRS has extension of 73 residues, which may contain two α-helices. Mammalian ArgRS interacts with p43 RNA-binding protein and associates with the multisynthetase complex (Quevillon et a.l 1999, Rho et al. 1999). This complex form is made through this extension region(Guigou et al 2004)

It was reported recently that RRM motif participates in protein-protein interaction (Cléry et al. 2008). When thinking about the role of the core structure of β-sheet of this N-terminal domain of mammalian ArgRS, the hydrophobic surface of β-sheet might help assembly with other proteins as well as recognition of nucleotide in the RNA.

2.1.2. tRNA and anticodon binding domain

In the complex of *P.horikoshii* ArgRS, the base of C35 of $tRNA^{Arg}$(CCU) is located in the hydrophobic pocket formed by the aromatic ring of Tyr587 at the C-terminal end and the hydrophobic side chains of Ile517 of and Pro591, Val592 and Leu593. C35 is found within the distance of the hydrogen bonds with the main chain of Tyr587 and with the main chain of Leu593 of the turn of the loop. The base of C34 undergoes no interaction with the protein. The base of U36 is surrounded by the side chains of Tyr509, Ala512, Ser516, and Met629 in the C-terminal end, U36 is in close proximity to Met629. The base of A37 is stacked on a C31:G39 base pair, and the base of C32 is stacked on the base of A37. The base of A38 lies among the hydrophobic side chains of Leu451, Lys455 and Val471 in the 'stem contact fold' domain and the side chains of Pro505 and Met629. In *S.cerevisiae,* the base of C35 of $tRNA^{Arg}$(ICG) is also located in the hydrophobic pocket formed by the aromatic ring of the conserved Tyr565 and C36 is also stacked to Tyr565 therefore C35,Tyr565 and C36 are stacked to each other (Fig. II-2-5).

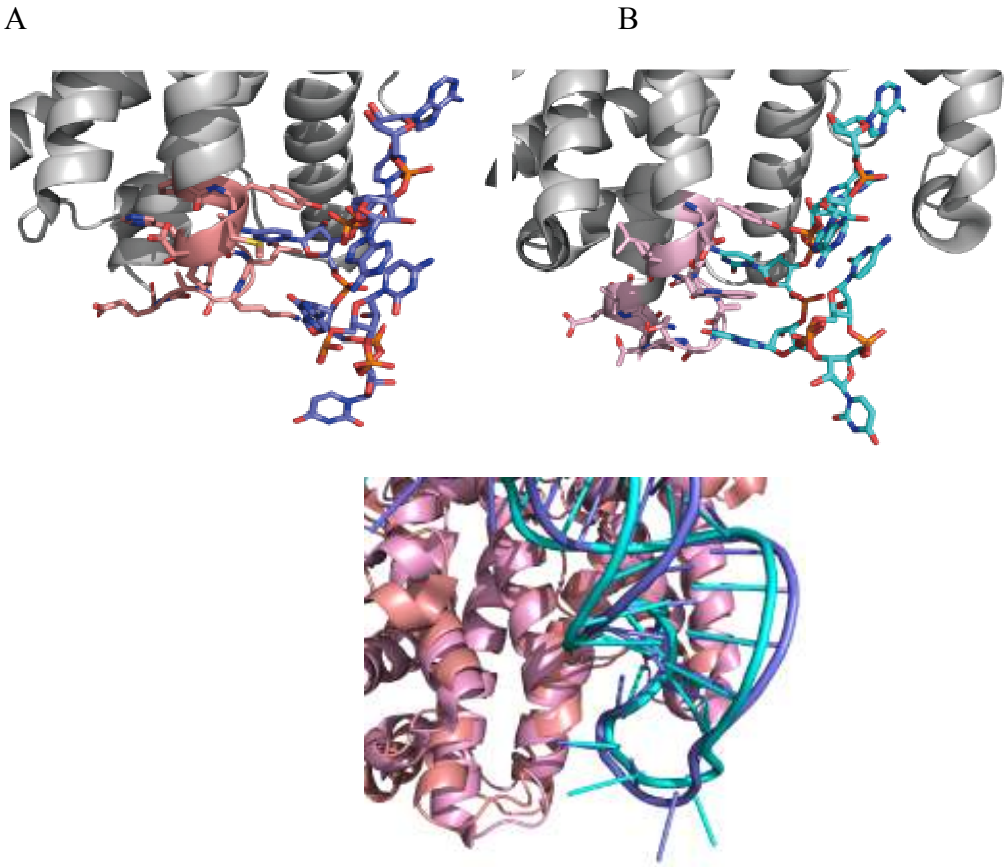


Fig. II-2-5 Interaction between anticodon binding domain and anticodon of tRNA
A Interaction between anticodon binding domain of *P.horikoshii* ArgRS and $tRNA^{Arg}$(CCU)
C35 is stacking with Y587 but C34 is flipped out from hydrophobic
B Interaction between anticodon binding domain of *S.cerevisiae* ArgRS and $tRNA^{Arg}$(ICG)
C35 stacked with Y565 and W569 C34 is also stacked with W569

2.1.3. CCA end of tRNA and catalytic domain

In the 3'-terminal G73–C74–C75–A76 sequence of $tRNA^{Arg}$(CCU), two transient forms were observed in the ternary complex as well as in the binary complex, depending on crystallization conditions(PDBID: 2ZUE and 2ZUF).

In one conformation (II-2-6 A) the base of G73 is stacked upon a G1-C72 base pair. This structure was first observed by NMR analysis in the $tRNA^{Ala}$ acceptor end micro-helix (Ramos et al. 1997). The C74–C75–A76 sequence is invisible in the electron density map, which indicates increased conformational flexibility around G73.

In another conformation (II-2-6 B) of the 3'-terminal end of $tRNA^{Arg}$(CCU) the base of G73 is not stacked upon a G1-C72 base pair, and the conformation of the phosphodiester bridge of C5'–O–P–O–C3' between C72– G73 and G73–C74 is not of the normal helix type (Fig. II-2-7).This local conformation of C72–G73–C74 is similar to the conformation observed for $tRNA^{Arg}$(ICG) bound to *S cerevisiae* ArgRS in the tertiary complex. The ribose of G73 and the bases of C75–A76 are invisible in the electron density map.

A B

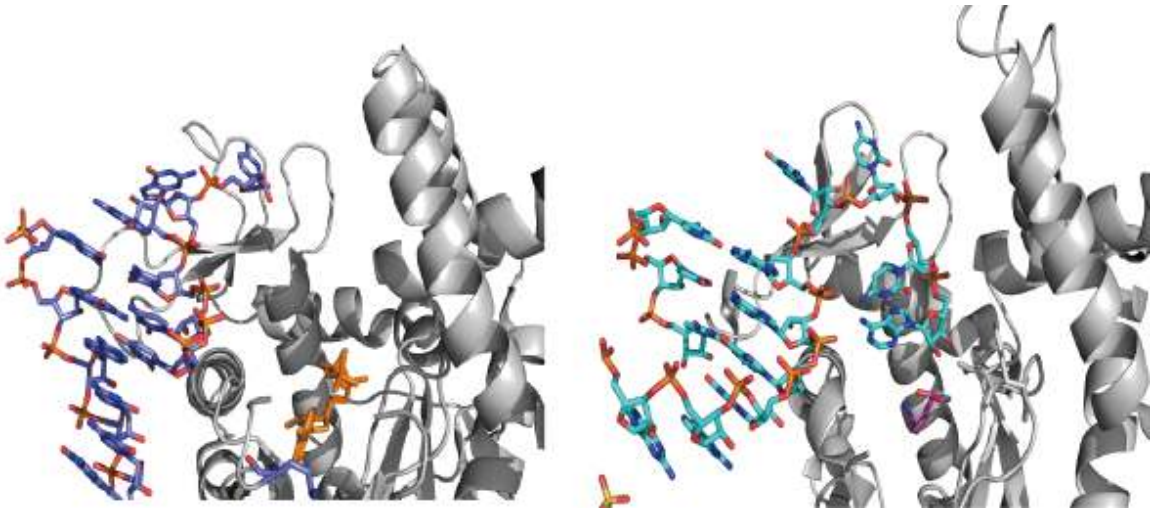


Fig. II-2-6 CCA end of tRNA of *P.horikoshii* ArgRS and *S.cerevisiae* ArgRS

A is CCA end of $tRNA^{Arg}$(CCU) B is the CCA end of $tRNA^{Arg}$(ICG)

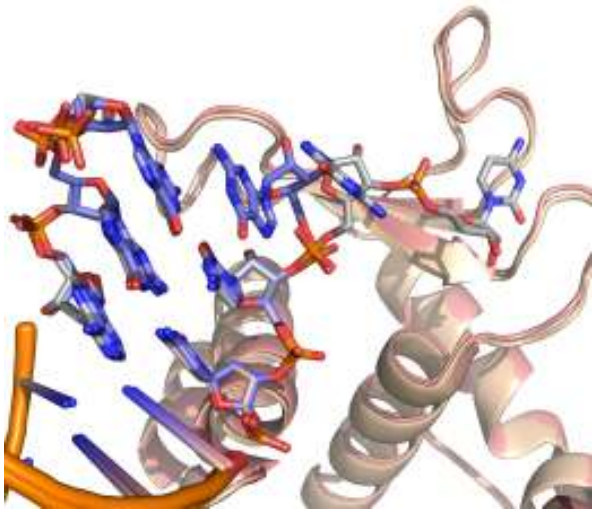


Fig. II-2-7 Two conformations of CCA end of *P.horikoshii* ArgRS

Blue is the conformation that 73 stacked on 1:72 base pair.
Gray is the conformation that 73 does not stacked on 1:72 base pair

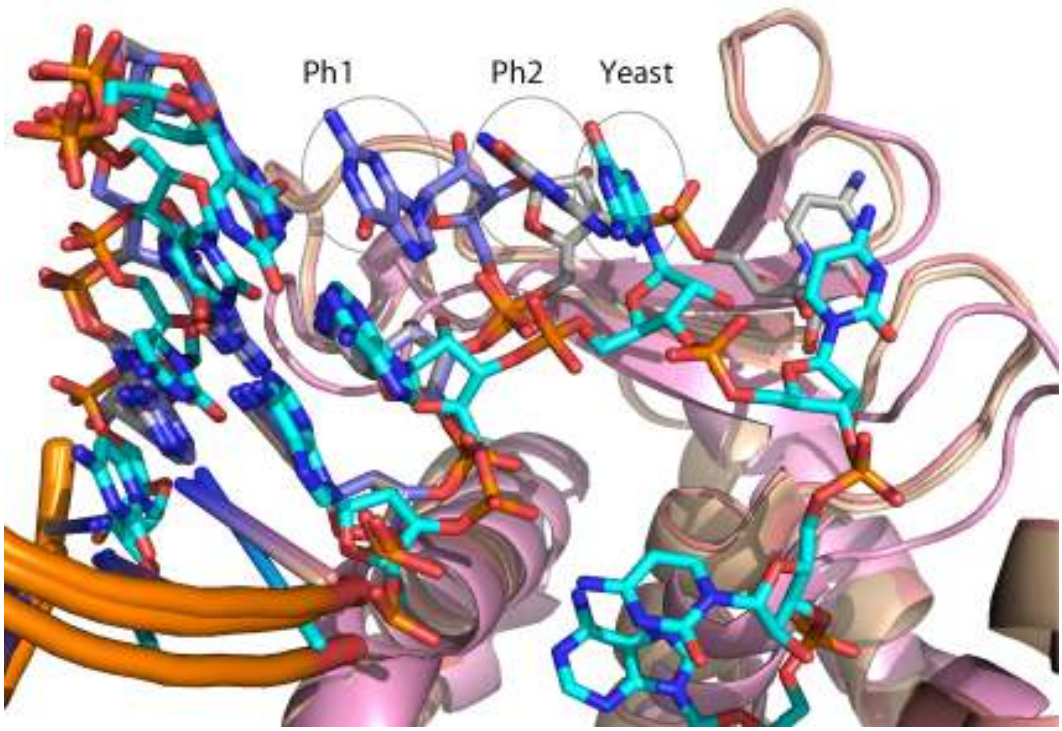


Fig. II-2-8 Superimposition of *P.horikoshii* ArgRS two conformations (2ZUE,2ZUF)and *S.cerevisiae* ArgRS (1F7U)

Ph1 is one conformation of *P.horikoshii* ArgRS and Ph2 is second conformation of *P.horikoshii* ArgRS

This newly observed transient form is the intermediate form, through which the conformation of the 3'-terminal end changes from the first stage to the final stage. The base of C74 is found near the surface of the connective polypeptide domain, which is in a transient position. A hydrophobic cleft is constructed by the side chains of Tyr300, Ala303, Val321, Arg324 and Ser325 in the connective polypeptide domain. The relative orientation of G73 and C74 to the connective polypeptide domain is similar to that observed in $tRNA^{Arg}$ (ICG) bound to *S.cerevisiae* ArgRS in the tertiary complex. It is predicted that the conformational change from the first stage to the final stage takes place in the absence of Arg, and the hydrophobic environment changes the hydration state around the phosphodiester bridges in C72–G73–C74–C75–A76. In *T.thermophilus* Val-tRNA synthetase bound to $tRNA^{Val}$ (Protein Data Bank ID: 1GAX) (Fukai et al. 2003) and *T. thermophilus* Leu-tRNA synthetase bound to $tRNA^{Leu}$ (Protein Data Bank ID: 2BYT) (Fukunaga et al. 2005), tRNA is left in the first stage, where the base of A73 is stacked on a G1-C72 base pair. Moreover, in *Aquifex aeolicus* Met-tRNA synthetase bound to $tRNA^{Met}$ (Protein Data Bank ID: 2CSX) (Nakanishi et al. 2005), the base of A73 is still stacked on a G1-C72 base pair; which means the change of the conformation of the 3'terminal end of $tRNA^{Met}$ is not in progress. On the other hand, in *E. coli* Cys-tRNA synthetase bound to $tRNA^{Cys}$ (Protein Data Bank ID: 1U0B), the structure shows that the base of U73 is no longer stacked on a G1-C72 base pair which allows the 3'terminal CCA end to enter into the active site (Fig II-2-7). In addition, it is interesting to observe that $tRNA^{Pro}$ is aminoacylated by AlaRS in vitro only when the 3-70 pair is mutated to G3·U70 and the end of the stem is converted to A73:G1·C72 (Ramos et al. 1997).

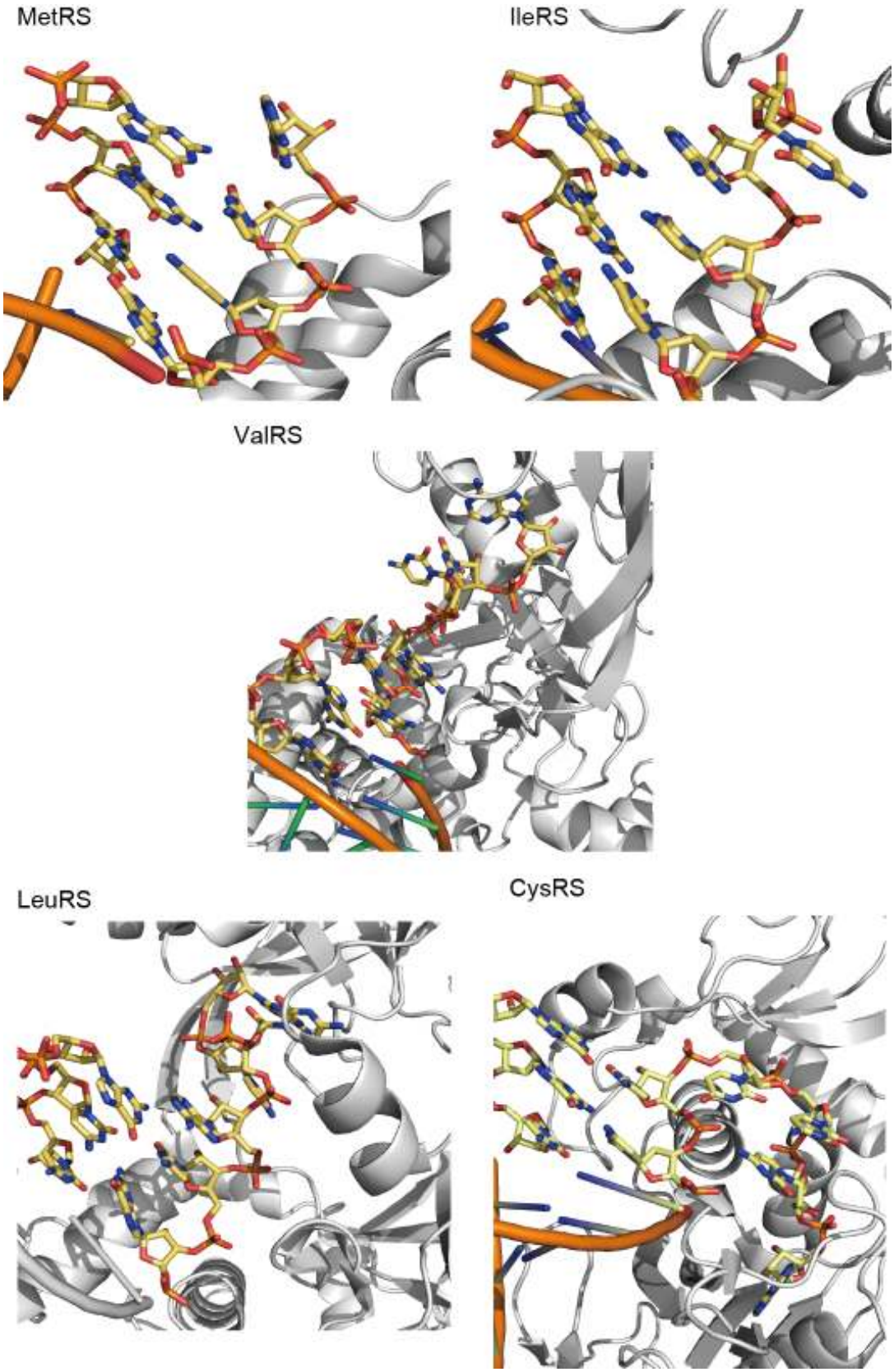


Fig. II-2-9 CCA end conformation of ClassI aaRSs

$tRNA^{Met}$ and $tRNA^{Ile}$ were not visible its CCA end and 73 stacked to 1:72
$tRNA^{val}$ and $tRNA^{Leu}$ were visible its CCA end but oriented to editing domain and 73 is not stacked to 1:72. $tRNA^{cys}$ ^{was} visible its CCA end and oriented to catalytic domain and 73 is not stacked to 1:72

2.1.4. AMP-PNP binding

In *P. horikoshii* ArgRS, the ATP analog (AMP-PNP) molecule was clearly found in the active site. *P.horikoshii* ArgRS has His 417 in the active site the same position of *S.cerevisiae* ArgRS is Met. This His 417 stacks to the AMP-PNP ribose. Because of this the AMP-PNP remained in the catalytic core of *P.horikoshii* ArgRS. In *P.horikoshii* ArgRS, the 2'-OH of the ribose of AMP-PNP is in close proximity to Nitrogens of Gly384 and Glu386 (Gly384 –Ala385 –Glu386 –Gln387 turn), the distances being 2.7 Å and 2.8 Å, respectively. The distance of Lys132 and AMP-PNP is 2.8Å (Fig. II-2-10). The adenosine moiety is fitted into this hydrophobic groove. The Ala372– Ser373–Gln374–Gln375 turn in *S.cerevisiae* ArgRS suggested that this backbone formation is quite similar.

The side chain of Lys132, located three residues upstream from the signature sequence motif 'HIGH', is close to AMP-PNP but there is no interaction. In *P.horikoshii* ArgRS, the 'KFSG' portion does not contribute to interaction with AMP-PNP (Fig. II-2-10).

In aaRSs belonging to class I, the turn that interact with AMP-PNP is almost conserved, as Gly/Ala-Xaa-Asp/Glu-Xaa (Xaa stands for any amino acid).The residue corresponding to Val418 is directed inside. In *P.horikoshii* the distance between Cα of Glu386, the third residue, and Cα of Val418 is 12.8Å, and in free *E. coli* Met-tRNA synthetase (Protein Data Bank ID: 1QQT) (Nakanishi et al. 2005), the distance between Cα of the third residue, Asp296, in the S14–H14 turn and Cα of Val326 is 12.7 Å, in free *T.thermophilus* Ile-tRNA synthetase (Protein Data Bank ID: 1ILE), the distance between the Cα atoms of the corresponding Asp553 and Ile584 is 13.4 Å, in free *P.horikoshii* Leu-tRNA synthetase (Protein Data Bank ID: 1WKB) (Tukalo et al. 2005), the distance between Asp612 and Gly644 is 12.5 Å, and in free *T.thermophilus* Val-tRNA synthetase (Protein Data Bank ID: 1IYW) (Fukai et al. 2003), the distance between Asp490 and Val521 is 11.9 Å. These distances within 0.9 Å of the distance of 12.8 Å in ArgRS indicate that, in these aaRSs, this space is the binding site of the adenosine moiety of ATP, and in *E. coli* Cys-tRNA synthetase bound to $tRNA^{Cys}$ (Protein Data Bank ID: 1U0B) (Hauenstein et al. 2004), the distance between Asp229 and Val260 is 11.4 Å. AMP weakly inhibits the binding of ATP in a competitive manner in the aminoacylation reaction.

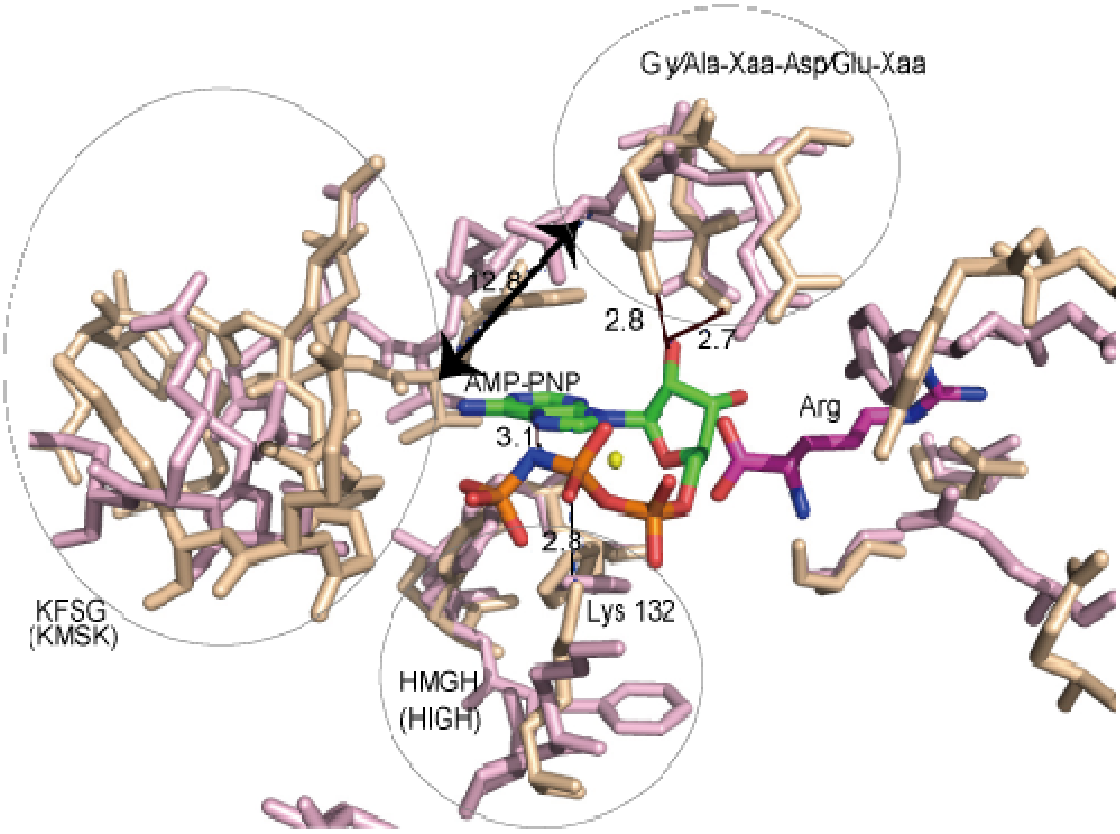


Fig. II-2-10 Catalytic core of ArgRS

Light orange: *P.horikoshii* ArgRS, Plink: *S.cerevisiae* ArgRS

2.1.5. tRNA

Distances between C1 of C72 of the 1-72 pair in the acceptor stem, C1 of G18 forming a hydrogen bond to U55 of the T-loop and C1 of C35 of the anticodon loop of tRNA were compared with corresponding distances in *S.cerevisiae* $tRNA^{Arg}$ (ICG) to confirm the similarity of the three dimensional structures of tRNAs. The distance difference was within 1 Å. These facts indicate that the framework of $tRNA^{Arg}$ of the L-shape is conserved in these two cases (Fig. II-2-11).

tRNA		
	Ph $tRNA^{Arg}$ (Å)	Yeast $tRNA^{Arg}$ (Å)
Cyt/Ade72C1---Gua18C1	40.1	39.4
Cyt/Ade72C1---Cyt35 C1	60.7	60.8
Gua18C1---Cyt35C1	42.1	42.9
Ade/Dhu 20C1---Cyt72C1	37.8	40.9
Ade/Dhu 20C1---Cyt35C1	45	43.8

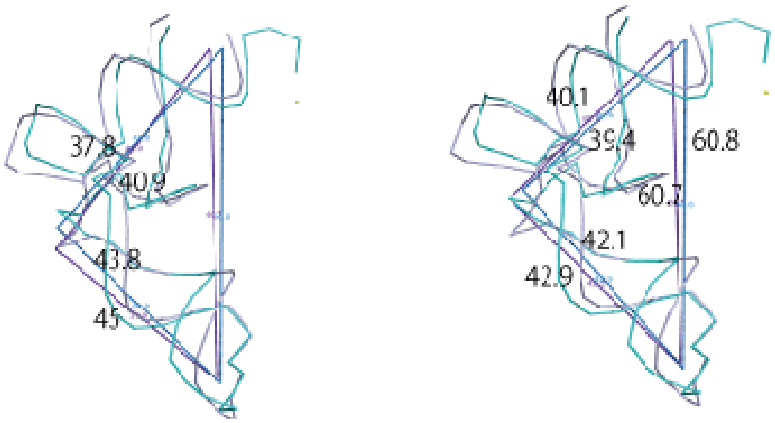


Fig. II-2-11 Comparison of backbone of $tRNA^{Arg}$(CCU) and $tRNA^{Arg}$ (ICG)

2.2. The aminoacylation for $tRNA^{Arg}$(CCU) with wild type ArgRS or lacking the N-terminal domain ArgRS (ΔN ArgRS)

To clarify whether the binding of the D-loop of $tRNA^{Arg}$ (CCU) to the N-terminal domain contributes to the activation effect of tRNA or not. I measured tRNA assisted Arg-AMP formation reaction or the aminoacylation reaction, I constructed *P.horikoshii* ArgRS (residues 92–629; ΔN ArgRS) lacking the core region of the N-terminal domain (residue 1 to residue 91) in order to completely eliminate interactions between the N-terminal domain and the D-loop of $tRNA^{Arg}$ and compared the kinetic parameters of the aminoacylation reaction for wild-type ArgRS and ΔNArgRS (Fig. II-2-12).

The gene encoding a ΔNArgRS (92-629) was subcloned into pET28c and transformation of BL21(DE3) codon plus. *E.coli* were grown with LB at 37 degree.

The ArgRS full length expression was induced with IPTG but induction of IPTG was not necessary because ΔNArgRS was sufficiently expressed. The cultured *E.coli* was suspended with lysis buffer and cell-breakage was done with a sonicator. The crude proteins underwent thermal treatment at 80 degrees and proteins originated from *E.coli* were eliminated. The remaining proteins were followed by a purified on an Ni-chelating column and further purified by FPLC with an Hitrap heparin, Resource Q column. The crude protein of ΔN ArgRS was purified in the same way as the wild-type protein (Fig. II-2-13).

2.2.1. $tRNA^{Arg}$(CCU)

There are four species of *P.horikoshii* $tRNA^{Arg}$. In *P.horikoshii*, codon usages for AGA and AGG codons are used 19 and 34 times, respectively, and they amount to 98% among six codons for Arg (Fig. II-2-13). The D-loops of isoacceptor $tRNA^{Arg}$(UCU) and $tRNA^{Arg}$(CCU) contain nine (AGCAGGAC20aA) or ten nucleotides (AGCCA17a GGAC20aA), respectively. The anticodon CCU of $tRNA^{Arg}$(CCU) is different at position 35 from CAU anticodon for methionine $tRNA^{Met}$(CAU). From those facts, we choose to use the $tRNA^{Arg}$ (CCU). The *P.horikoshii* $tRNA^{Arg}$ was transcribed *in vitro* with T7 RNA polymerase, and the crude transcript was extracted with phenol-chloroform and charged on a Mono Q column (Fig. II-2-14, II-2-15).

2.2.2. Measurement of the aminoacylation reaction for $tRNA^{Arg}$ (CCU)

For wild-type ArgRS and ΔNArgRS, the *K*m values for $tRNA^{Arg}$(CCU) were 2.6 μM and 3.8 μM and the measured ratio of the V-value of ΔNArgRS to that of wild-type ArgRS was $[(8 \pm 2) \times 10^2]$. This indicates that the binding of the D-loop of $tRNA^{Arg}$(CCU) with the N-terminal domain makes a minor contribution to the aminoacylation reaction, and that ΔN ArgRS performs the aminoacylation reaction of $tRNA^{Arg}$(CCU) well enough. In particular, the proper acceptance of C35 and U36 of $tRNA^{Arg}$(CCU) on the plausible structures are the predominant contributors to the aminoacylation reaction of specificity *P.horikoshii* ArgRS.

Reaction activity on *P.horikoshii* ArgRS was low in the conditions of reaction used for *S.cerevisiae* and *T.thermophilus* ArgRS. The low *K*cat value of *P.horikoshii* ArgRS may be related to the stable binding between N-terminal domain of ArgRS and $tRNA^{Arg}$(CCU) observed in the crystal structure.

2.3. Crystallization of ΔN ArgRS and $tRNA^{Arg}$(CCU)

After a lot of crystallization ways, a crystal ΔNArgRS and $tRNA^{Arg}$(CCU) finally grew in 2.0 M $(NH_4)_2SO_4$, 2% PEG400, 0.2 M Hepes-NaOH pH 7.0 at 24 degree. However, it was difficult to reproduce and the crystals were sensitive to temperature. Therefore crystals strong enough for X-ray measurement were not obtained.

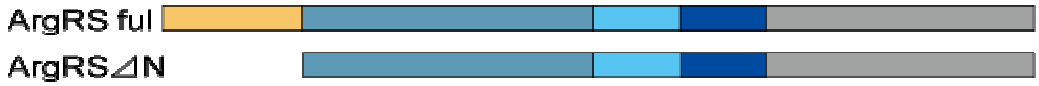


Fig. II-2-12 Construct of ArgRS

Orange: N-terminal domain. Slate: Catalytic domain Sky blue: Insertion doamin1 Dark blue: Insertion domain 2, Gray: anticodon binding domain

codon	E.c. usage	anticodon	T.t usage	anticodon	S.c. usage	anticodon	P.h. usage	anticodon
CGU	28		1		6		1	
CGC	21	ICG	3	UCG	3	ICG	1	
CGA	3		0		3		1	
CGG	4		33	CCG	2	CCG	0	
AGA	1		1		21	UCU	19	UCU
AGG	1	CCU	14	CCU	9	CCU	34	CCU

Fig. II-2-13 Codon usage of *P.horikoshii* ArgRS

A B

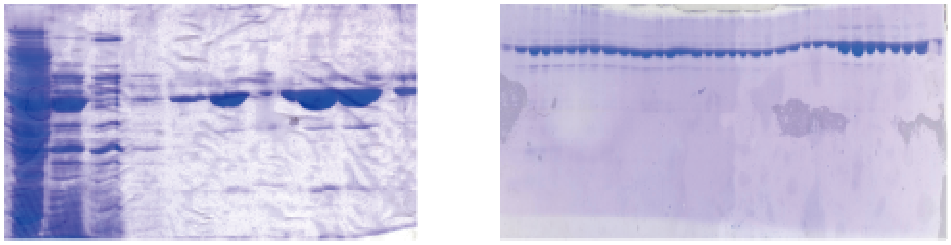


Fig. II-2-14 *P.horikoshii* ΔNArgRS purification

A cell extract to Heparin column B Resource Q column

A B

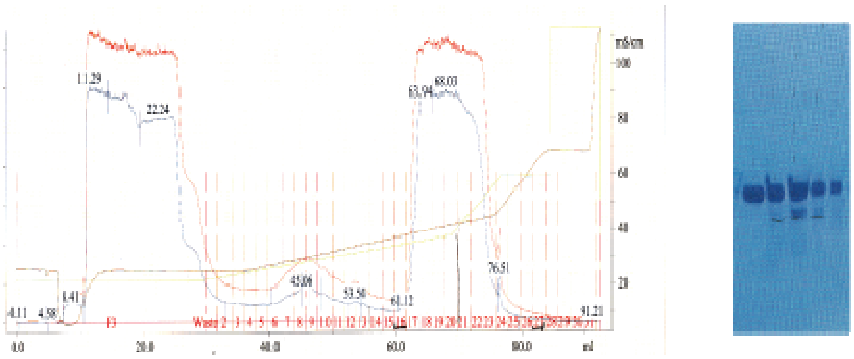


Fig. II-2-15 *P.horikoshii* tRNA purification

A. chromatogram of tRNA purification B. Acrylamide gel analysis of tRNA

2.4. Discussion

As described above, in the case of ArgRS the tRNA-assisted ATP–PPi exchange reaction has been observed, but tRNA-independent ATP–PPi exchange reaction has never been observed. Furthermore, in the presence of the tRNA treated with periodate, which oxidizes the 2'-OH and 3'-OH groups of the ribose of A at the 3'-end to convert them into dialdehyde groups, the ArgRS, GlnRS and GluRS enzymes were incapable of catalyzing the ATP-PPi exchange reaction. In the case of GluRS from *E. coli*, *S.cerevisiae*, porcine liver, and rat liver, the tRNA-independent ATP–PPi exchange reaction was observed at much higher concentrations of Glu, whereas the tRNA-assisted ATP–PPi exchange reaction was observed at lower concentrations of Glu. The *K*m value for Glu measured in the tRNA assisted ATP–PPi exchange reaction decreases significantly by 10^2-10^3-fold in comparison with that in the tRNA-independent ATP–PPi exchange reaction. In the absence of tRNA, Arg-AMP and Gln-AMP were not detectable as intermediates formed by ArgRS and GlnRS, respectively. As mentioned above, the hydroxyl group of the ribose of the 3' terminal A76 of tRNA is essential for the ATP–PPi exchange reaction on ArgRS, GlnRS, and GluRS. The cognate tRNA is also necessary for the ArgRS catalyzed pyrophosphorolysis of chemically synthesized Arg-AMP in the presence of PPi and Mg^{2+} (Lui et al. 1978). Furthermore, the pyrophosphorolysis of Arg-AMP and the ATP–PPi exchange reaction catalyzed by ArgRS in the presence of tRNA have pH optima of 6.2 and 6.5, respectively.

This indicates that the 2'-OH group of tRNA interacts with Glu directly. As the 2'-OH group of tRNA is close to the α-carboxyl group of Glu, an interaction between the 2'-OH group and the α-carboxyl group occurs, with a preference to bind with this Oxygen atom rather than the Carbon atom. On the other hand, the interaction with Carbon atom results in the aminoacylation reaction. It is expected that the direct interaction between the the 2'-OH group of tRNA and the α-carboxyl group of Arg/Gln will also take place in the ATP–PPi exchange reaction on ArgRS and GlnRS. The interaction between the 2'-OH group and the Oxygen atom of the α-carboxyl group of Arg is thought to accelerate the pyrophosphorolysis reaction on ArgRS. In a ternary complex of *S.cerevisiae* ArgRS (PDB ID: 1F7V), the Oxygen atom of the 2'-OH group in the 3'end A76 of $tRNA^{Arg}$ (ICG) is in the distance of 3.18 Å, 3.71 Å and 3.57 Å from the Carbon atom, O1, and O2 (the Oxygen contacting Pα of ATP is defined as O1 and the other as O2), respectively, of the α carboxyl group of Arg (Fig. II-2-16).

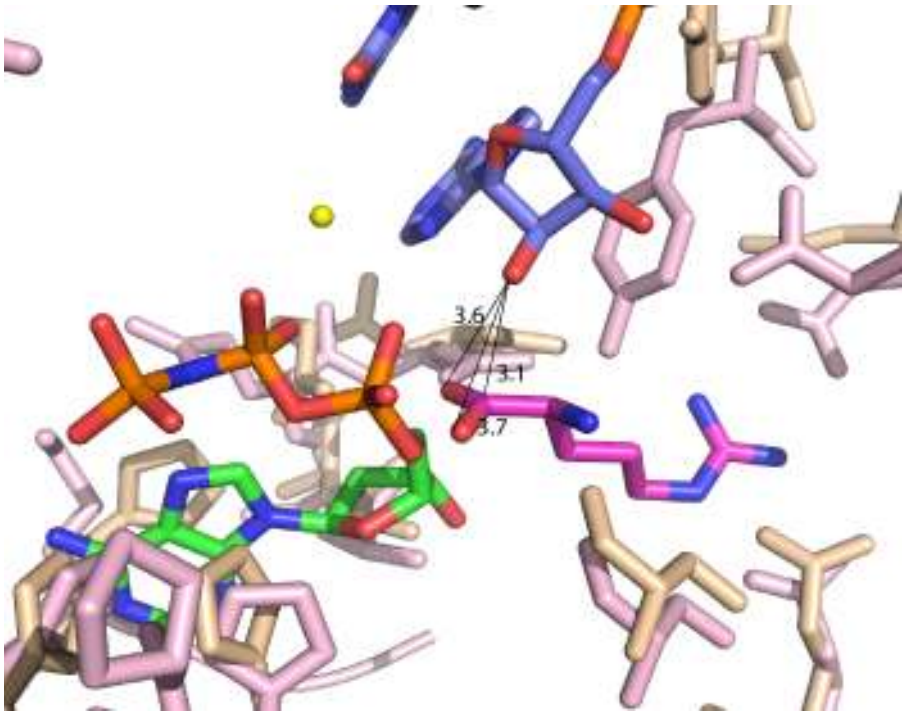


Fig. II-2-16 The distance between 2'-OH of A76 and Oxygen atom, Carbon atom of Arginine

The movement of O2 by a rotation of 45 degree around the Cα–C bond causes a decreased distance to 2.77Å between O2 and the Oxygen atom of the 2'-OH group, which is a suitable distance to form a hydrogen bond to form between them, note that the Carbon is not moved by this rotation. Such a slight rotation induces the formation of a hydrogen bond between O2 and the 2'-OH group of A76 of tRNA, and thereby the anionic form of the C-1 is converted into the C=O1 form. Hence, the intermolecular rearrangement reaction for Arg-AMP formation is considered to take place through a pathway such as the one shown in Fig. II-2-16. The O1 orbital of the double bond C=O1 changes initially from sp^2 to sp^3 hybrid orbital, and a bond is formed between O1 and Pα of ATP through formation of trigonal bipyramid coordination around Pα. At the same time, the H-O2 bond of H-O2-C of the α-carboxyl group of arginine is transferred to the O2-C bond to form the double bond of O2=C, and a proton is transferred to a water molecule. Successively, the bond of C=O1 is transferred to the O1-Pα bond, the Pα-O bond is transferred to the O-Pβ bond, and thereby the O-Pβ bond is converted into the form of O=Pβ. Finally, PPi is released in the form of Mg-PPi. Because O1 of the amino acid binds to Pα and PPi is released from Pα in the reverse side, this reaction of the Arg-AMP formation is a S_N2 reaction.

The modeling of Arg, ATP and A76 of tRNA on *P.horikoshii* ArgRS for the Arg-AMP formation reaction indicates that when the straight side chain of an Arg molecule is inserted into the hydrophobic pocket as in the Arg molecule bound to *S.cerevisiae* ArgRS, its carboxyl group can locate in close proximity to Pα of ATP

The α-carboxyl group of the Arg molecule can assume such a conformation that O2 forms a hydrogen bond with the 2'-OH group of the ribose moiety of A76 of $tRNA^{Arg}$ by rotation around Cα–C. When Pα of ATP gains access to O1 of C=O1 of the α-carboxyl group of Arg, and two oxygen atoms of Pβ and Pγ are coordinated by Mg^{2+} then, the intermolecular rearrangement reaction occurs(Fig. II-2-17, II-2-18).

In the pyrophosphorolysis reaction of synthetic Arg-AMP in the presence of PPi and Mg^{2+}, the 2'-OH group of $tRNA^{Arg}$ can not be placed in the proper configuration for the aminoacylation reaction without conformational rearrangement but should be in much preferable configuration to form a hydrogen bond to C=O2 of the Cα-(C=O2)-O1-Pα moiety. Even though the pyrophosphorolysis reaction is also S_N2 reaction, the hydrogen bond between 2′-OH of tRNA and C=O2 is also required for the pyrophosphorolysis reaction. The step of the cleavage of the bond between CO1 and Pα of Arg-AMP is accelerated by this hydrogen bond. The formation of this hydrogen bond leads double bond between Carbonyl and O1 and the intermolecular rearrangement reaction proceeds in the reverse direction of the formation reaction of Arg-AMP shown in Fig. II-2-17 and Fig. II-2-18. The amount of neutral PPi, in which the lone pairs of double bonds of Pβ=O and Pγ=O are coordinated by Mg^{2+}, increases at lower pH in the hydrophobic circumstance of the reaction region of ArgRS. PPi is involved in the pyrophosphorolysis reaction in the form of Mg-PPi. This fact explains why the pyrophosphorolysis reaction of synthetic Arg-AMP has an optimum at lower pH 6.2 in the presence of tRNA and is a little lower than that for the ATP–PPi exchange reaction (Lui et al. 1978)

The observation that the ATP–PPi exchange reaction requires tRNA in ArgRS, GlnRS and GluRS can be explained by the mechanism that the formation of the hydrogen bond between O2 of the carboxyl group of the cognate amino acid and the 2'-OH group of the ribose moiety of A76 to maintain the appropriate orientation of the carboxyl group of the amino acid for Pα.

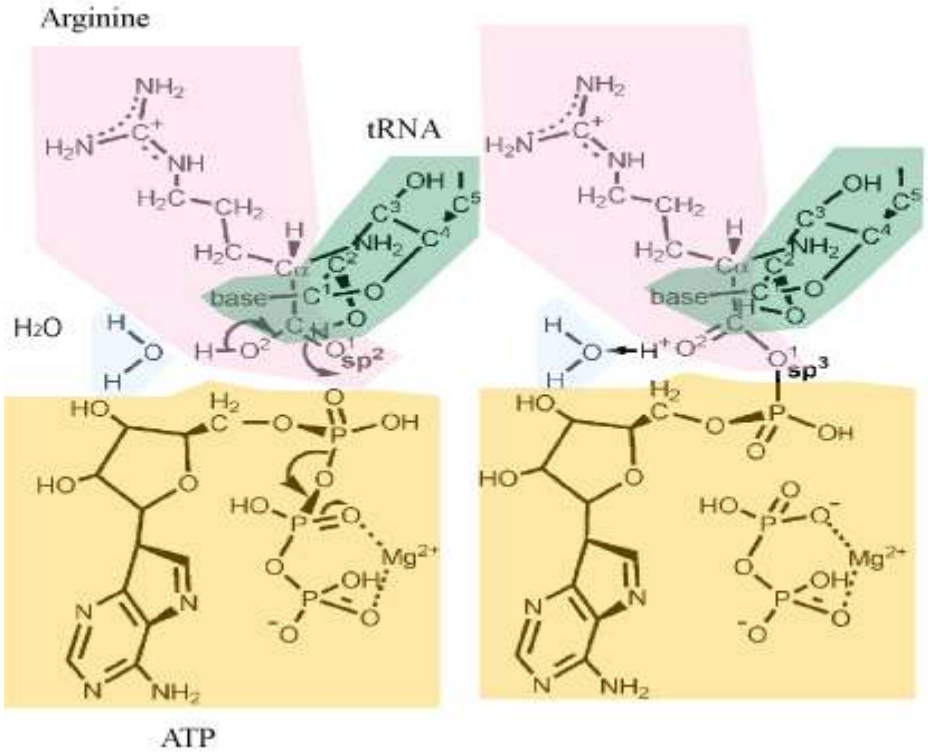


Fig. II-2-17 The scheme of 2'OH of A76 contributed aminoacylation reaction.

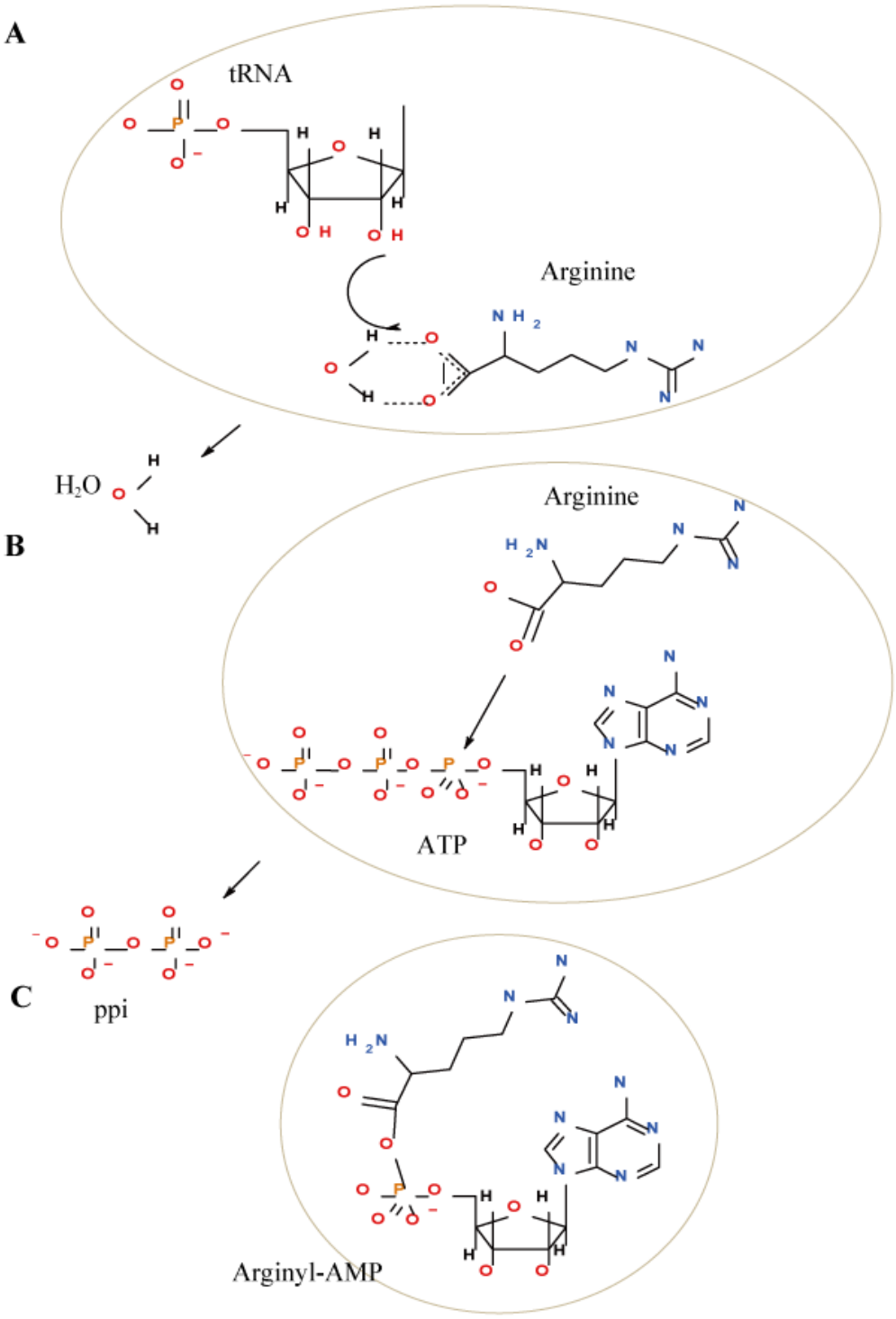


Fig. II-2-18 The scheme of 2'OH of A76 contributed aminoacylation reaction.

A) 2'-OH of tRNA interact with carbonyl group of Arginine therefore the water molecule dissociate from Arginine.
B) Carbonyl group of Arginine can attack to a phosphate of ATP and ppi is released.
C) Forming of Arginyl-AMP

2.5. Arginyl-tRNA synthetase and $tRNA^{Asp}$

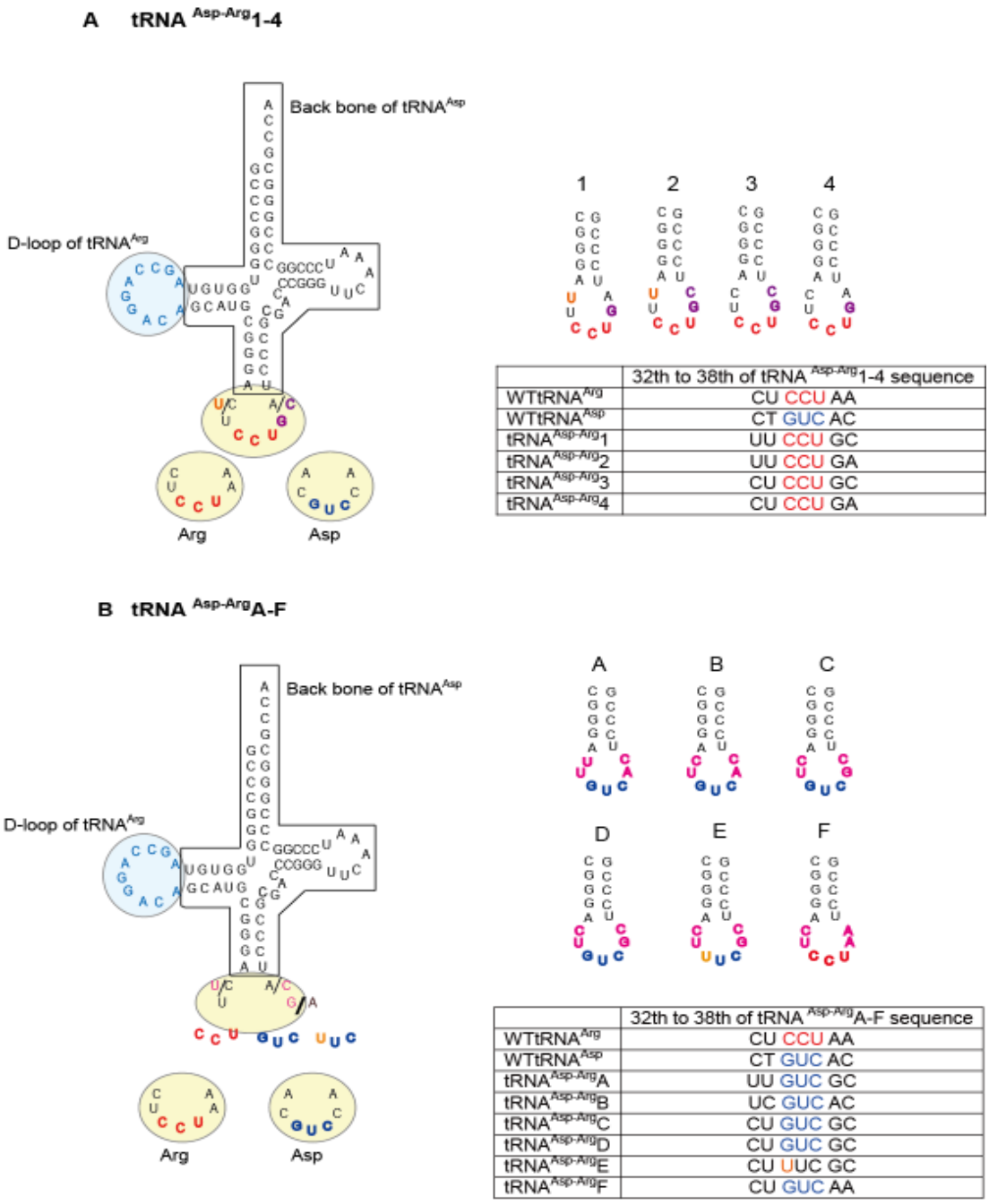
In *S.cerevisiae*, ArgRS arginylates a non-cognate $tRNA^{Asp}$(**GUC**) transcript with about 10 fold less efficiency than a cognate $tRNA^{Arg}$(**UCU**) transcript. In addition, the *k*cat/*K*m value of $tRNA^{Asp}$ **G**34U mutant is the same as that of $tRNA^{Arg}$. These facts suggest that in the anticodon-binding domain of ArgRS, the base **U**34 of $tRNA^{Asp}$(**UUC**) mutant binds at the site accepting **U**34 of $tRNA^{Arg}$(**UCU**) (Sissler et al. 1996). The fact that $tRNA^{Met}$(**CAU**) is also arginylated by *E. coli* ArgRS (Schulman and Pelka 1989b) indicates that the bases of the anticodon **C**34, **A**35, **U**36 interact with sites accepting **C**34, **C**35 and **G**36 bases of $tRNA^{Arg}$, respectively. Taking these reports into consideration, the bases of **U**35 and **C**36 of $tRNA^{Asp}$(**UUC**) mutants are expected to also interact with sites accepting 35^{th} and 36^{th} bases, respectively. A 100% charging level is observed for wild type $tRNA^{Arg}$ and wild type $tRNA^{Asp}$ transcripts, whereas the arginylation plateaus $tRNA^{Asp}$U34 mutant, which corresponds to optimal charging level, decreases by 15% compared with that of wild type $tRNA^{Asp}$, $tRNA^{Asp}$(**GUC**G) is not common to **C**35 and **G**36/**U**36 of the anticodon of $tRNA^{Arg}$ but proposed 'slipped' recognition set hypothesis that in *S.cerevisiae* $tRNA^{Asp}$(**GUC**G), **C**36-G37 may be used in place of **C**35-**G**36 of $tRNA^{Arg}$ (Sissler et al. 1998).

Compared the structure of $tRNA^{Asp}$(**GUC**G) (Ruff et al. 1991) and the structures of complexes of $tRNA^{Arg}$(**ICG**) of *S.cerevisiae* ArgRS (Delagoutte et al. 2000). The backbone of the anticodon loop of $tRNA^{Arg}$(**ICG**) assume the same structure. But local conformation of anticodon region is different. **U**35 and **C**36 are inserted into hydrophobic pockets of ArgRS. The relative orientation of **G**37 base for **C**36 base in $tRNA^{Asp}$(**GUC**G) is quite different from the relative orientation of **U**36 base for **C**35 base in $tRNA^{Arg}$(**CCU**).

To investigate the 'slipped' recognition set hypothesis and observed anticodon loop movement, making tRNA mutants. The $tRNA^{Asp}$ mutants were made used the backbone of $tRNA^{Asp}$(GUC) but D-loop and anticodon loop were changed. D loop of $tRNA^{Asp}$ was replaced to that of $tRNA^{Arg}$, and anticodon loop was changed to the anticodon of **UUC**, **GUC**, **CCU** (Fig. II-2-19 ($tRNA^{Asp\text{-}ArgA\text{-}F}$))**.**

The 32^{th} and 37^{th} bases in the anticodon loop of $tRNA^{Asp}$GUC (C32-U33-G34-U35-C36-A37-C38) from *P.horikoshii* are different from those of $tRNA^{Asp}$GUC (U32-U33-G34-U35-C36-G37-C38) from *S.cerevisiae*. Since the conformation of the anticodon loop may be affected also by 32^{th} and 37^{th} bases, several mutants with anticodon of **GUC** and **UUC** but different 32^{th} and 37^{th} bases were constructed (Fig. II-2-19 $tRNA^{Asp\text{-}Arg1\text{-}4}$, $tRNA^{Asp\text{-}ArgA\text{-}F}$).

Using these mutant tRNA$^{Asp\text{-}Arg(1\text{-}4\ or\ A\text{-}F)}$ and *P.horikoshii* ArgRS, crystallization was done The rystals grew in 0.1 mM Hepes-NaOH pH 7.5, 30 % PEG 8000, 0.2 M ammonium sulfate, but crystals were grew only the drop that combination of *P.horikoshii* ArgRS and tRNA$^{Asp\text{-}Arg2}$, *P.horikoshii* ArgRS and tRNA$^{Asp\text{-}Arg4}$. The crystal were measured at PF (Photon Factory in Tsukuba), the diffraction was 4 Å. The crystals were quite thin and could not get the crystals strong enough for the measurement.



	32th to 38th of tRNA $^{Asp\text{-}Arg}$1-4 sequence
WTtRNA^{Arg}	CU CCU AA
WTtRNA^{Asp}	CT GUC AC
tRNA$^{Asp\text{-}Arg}$1	UU CCU GC
tRNA$^{Asp\text{-}Arg}$2	UU CCU GA
tRNA$^{Asp\text{-}Arg}$3	CU CCU GC
tRNA$^{Asp\text{-}Arg}$4	CU CCU GA

	32th to 38th of tRNA $^{Asp\text{-}Arg}$A-F sequence
WTtRNA^{Arg}	CU CCU AA
WTtRNA^{Asp}	CT GUC AC
tRNA$^{Asp\text{-}Arg}$A	UU GUC GC
tRNA$^{Asp\text{-}Arg}$B	UC GUC AC
tRNA$^{Asp\text{-}Arg}$C	CU GUC GC
tRNA$^{Asp\text{-}Arg}$D	CU GUC GC
tRNA$^{Asp\text{-}Arg}$E	CU UUC GC
tRNA$^{Asp\text{-}Arg}$F	CU GUC AA

Fig. II-2-19 tRNA mutations
A) tRNA $^{Arg\text{-}Asp}$ 1-4 mutations B) tRNA $^{Asp\text{-}Arg}$ A-F mutations

3. Threonyl-tRNA synthetase

3.1. Threonyl-tRNA synthetase (ThrRS)

As I described in introduction about aaRS, aaRSs are divided into two groups ClassI and Class II. Threonyl -tRNA synthetase belongs to Class II and has an editing domain. aaRSs have evolved a correction mechanism called editing, which eliminates the mis-acylated amino acid from the tRNA.

ThrRS possesses post transfer editing activities (Sankaranarayanan et al. 1999). The crystal structure of ThrRS from *E. coli* (PDB ID 1QF6 Fig II-3-1) allowed delineation of the mechanisms this enzyme employs to ensure the fidelity of Thr-tRNA^{Thr} synthesis (R Sankaranarayanan et al. 1999). The active site efficiently discerns between threonine and valine by using a zinc ion as a cofactor; the zinc coordinates both the amino group and the side-chain hydroxyl group of threonine, whereas it rejects valine because of steric hindrance with one of its methyl groups Serin is nor sticked in the catalytic site, but the seryl-tRNA is hydrolyzed in the editing site (Dock-Bregeon et al. 2000).

3.2. Crenarchaeal Threonyl-tRNA synthetases

ThrRS generally consists of three domains namely a N-terminal editing domain, a catalytic domain and a C-terminal anticodon-binding domain (Sankaranarayanan et al. 1999). Although the editing domain of ThrRS highlited in *E. coli* ThrRS is always found in ThrRS proteins from all bacteria and eukaryotes, a homologous domain could not be identified in most archaeal ThrRSs. In archaeas, the editing domain is different, with a fold similar to D-amino acids deacylases (Hussain et al. 2006). Moreover several crenarchaeal species have separate individual genes encoding the catalytic (ThrRS-1) and editing (ThrRS-2) synthetases. Interestingly, these two proteins both comprise the anticodon binding domain of ThrRS.

ThrRS-2 act as a trans-editing enzyme in order to hydrolyze specifically the misacylated Ser-tRNA^{Thr}, or alternatively, it could be complex with ThrRS-1 to provide the editing function to the catalytic enzyme. Deletion analyses indicate that ThrRS-2 is dispensable for growth of *S.solfataricus* (Korencic et al. 2004). The sequences of ThrRS-1 and ThrRS-2 from *A.pernix, from S.tokodaii* and from other organisms in the Sulfolobaceae family (*S.solfataricus*, *S.acidocaldarius and Metallosphaera sedula*) were analyzed in terms of structural modules (Shimizu,et al. 2009).

All ThrRS-1 have catalytic domains and anticodon binding domains, while ThrRS-2

has anticodon binding domains and editing domains of almost the same length and very similar sequence. An exception is StThrRS-2, which is truncated in its N-terminal region.

Among these ThrRS, the 3D structure of ApThrRS-1 was solved (Shimizu, et al. 2009). But the details of the functional interaction with tRNA are missing. To investigate the ThrRS-1 and ThrRS-2 interaction with tRNA, we performed binding assays and tried to make crystals.

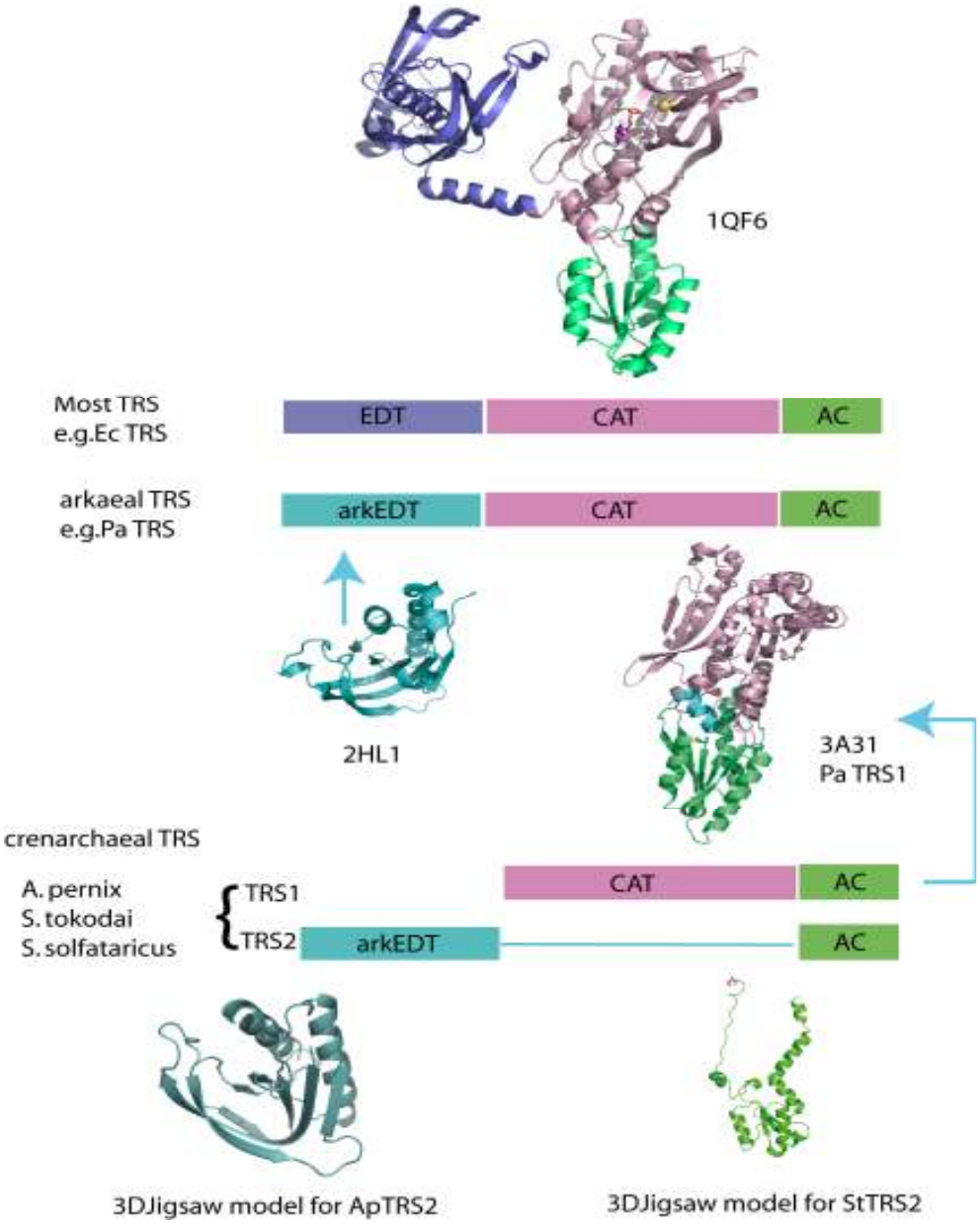


Fig. II-3-1 Comparison of *E.coli* ThrRS and Crenarchaeal ThrRS

Most ThrRS has three domains Catalytic domain (pink), Anticodon Binding domain (green), Editing domain (blue). Crenarchaeal ThrRS-1 has Catalytic domain and Anticodon Binding domain, ThrRS-2 has Anticodon Binding domain and Editing domain. ApThrRS-1 and St ThrRS-2 was predicted by 3Djigsaw.

3.3. Cloning of ThrRS-1 and ThrRS-2

Sequences of *Aeropyrum pernix* (Ap), and *Sulfolobus tokodaii* (St), *Sulfolobus solfataricus* (Ss) each with two ThrRS genes of different origin (Fig. III-3-2) were identified. They form 2 groups of sequences for ThrRS-1 (ApThrRS-1, StThrRS-1, SsThrRS-1) and ThrRS-2 (ApThrRS-2, StThrRS-2, SsThrRS-2) (Table III-3-1).

To investigate if each couple of ThrRS-1 and ThrRS-2 makes a complex or not, all ThrRS-1 and ThrRS-2 were subcloned into pnEA and pnCS vectors, which are compatible for co-expression in *E.coli*. pnEA provides for a His-Tag at the N-terminal end of the recombinant, and ampicilline resistance. pnCS doesn't provide for a Tag and confers spectinomycin resistance. The absence of tag allows producing a fully native protein. For each species, each protein was cloned into each vector, in order to eliminate possible effects of tags. I performed double digestion with restriction enzyme. The template was treated with restriction enzyme and the insert was purified with agarose gel, taking care of product when the DNA fragment was sealed from UV, all cloning succeeded.

3.4. Co-expression of ThrRS-1 and ThrRS-2

For each species, the constructs were co-expressed and purified by tag-affinity to investigate a possible complexation of ThrRS-1 and ThrRS-2. In parallel, the individual ThrRS-1 and ThrRS-2 constructs were transformed and expressed at the same time (Fig. III-3-3). The expressed proteins were purified with Ni-chelating column. The combination used is shown in Table 3-1-1.

The BL21(DE3) pRARE bacteria were grown in autoinduction medium. The result of purification with Ni-chelating column is showed in Fig. III-3-4.

We observed that the ApThrRS-1 (pnCS) and ApThrRS-2 (pnEA) were co-expressed and could be co-purified, but the expression ratio of these two proteins was not equal. For StThrRS-1-StThrRS-2 and SsThrRS-1-SsThrRS-2, the result was not clear, but there seems to be no complexation. Different lysis buffers were then assayed (changing ionic strength) but the complexation was not identified clearly.

This test is very dependant on the level of expression of each construct. Indeed, if one protein is in limited amount, compared to the other, the complexation can be difficult to see. Since individual expression of each ThrRS was very variable, we decided to continue with a more direct method.

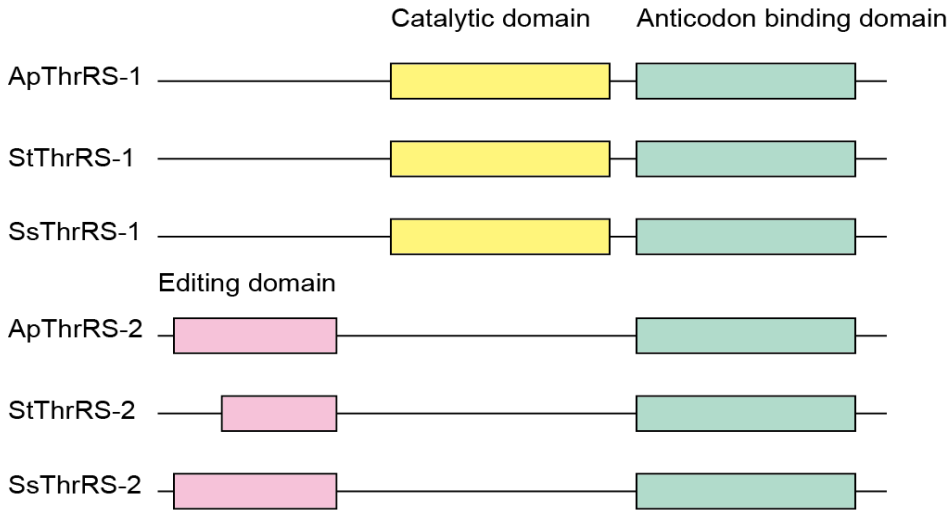


Fig. II-3-2 Domain composition of ThrRS-1 and ThrRS-2

ThrRS-1 has Catalytic domain and Anticodon binding domain, ThrRS-2 has Editing domain and Anticodon binding domain.

Table. II-3-1 The combination of expression test of ThrRS-1 and ThrRS-2 used in the co-expression test

	pnEA (with His-Tag)	pnCS (without His-Tag)
1	ApThrRS-1	ApThrRS-2
2	StThrRS-1	StThrRS-2
3	StThrRS-1	StThrRS-2
4	ApThrRS-2	ApThrRS-1
5	StThrRS-2	StThrRS-1
6	StThrRS-2	StThrRS-1

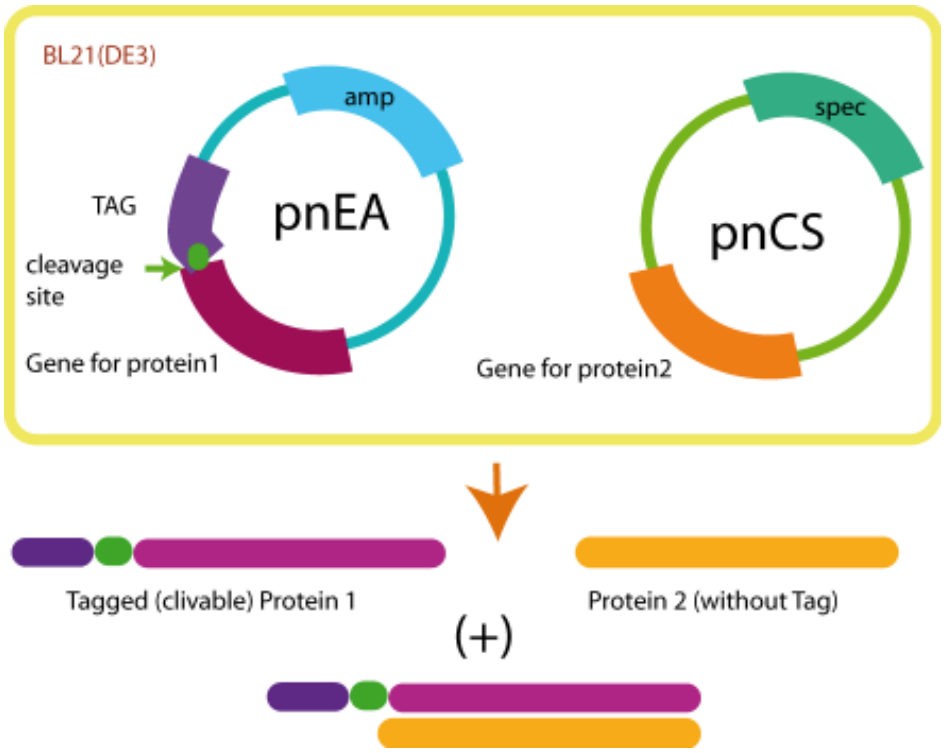


Fig. II-3-3 Co-expression system

The pnEA contains His-Tag at its N-terminal end and ampicilline resistance, pnCS doesn't have Tag and has spectinomycin resistant. Two plasmids are transformed together in BL21(DE3). This allows proteins to co-express and if the proteins make a complex, the complex should be purified by Ni-chelating column, thanks to the His-tag present on one of the protein (the construct on pnEA), which is verified with a denaturing gel (PAGE). The proteins are expressed in each plasmid, to eliminate possible effect of tag (the tag may prevent complex formation). In parallel, the individual proteins are also expressed, to detect possible problems with the level of expression.

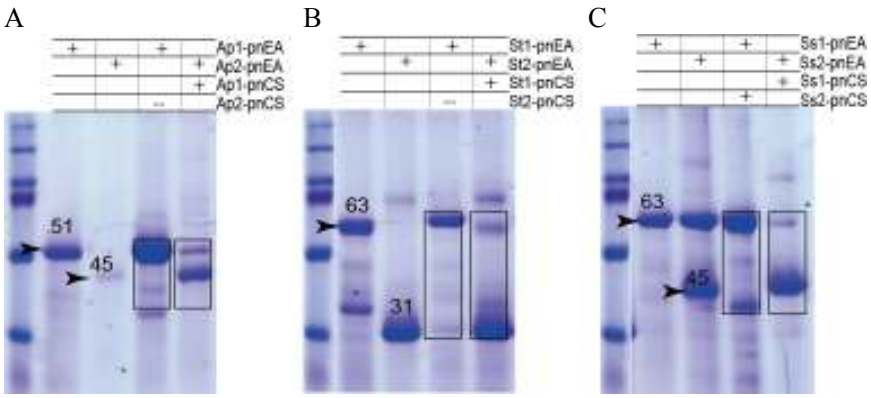


Fig. II-3-4 SDS-Page analysis of co-expression

All samples were migrated after Ni-chelating column purification.
A) ApThrRS expression test B) StThrRS expression test C) SsThrRS expression test

3.5. Complexation assay with gel filtration

All ThrRS-1 and ThrRS-2 in pnEA (His-tagged constructs) were expressed separately and purified. After Ni-chelating column, the proteins were mixed and loaded on a gel filtration column to analyze the complex formation.

We choose the system of StThrRS-1 and StThrRS-2 because the StThrRS-2 was truncated its N-terminal region (Fig III-3-2). Because of this truncation, the molecular weight is much different between StThrRS-1 and StThrRS-2, therefore it is a suitable combination to analysis by GF column.

The purified proteins StThrRS-1 and StThrRS-2 were incubated together, then subjected to molecular sieving. Fig. III-3-5 shows that complexation was not observed in our conditions.

Next, we thought that the tRNA could mediate the complex formation of ThrRS-1 and ThrRS-2. The proteins StThrRS-1, StThrRS-2 and the tRNA were incubated together and loaded on the GF column (Fig. III-3-6). We used tRNA of *E.coli* (*Ec*tRNA) that was purified from *E.coli* (fully modified). When the protein and tRNA samples are loaded at the same concentration, the absorption of the tRNA is much larger than protein's one. In addition, for StThrRS-1, StThrRS-2 and *Ec*tRNA, the elution volume from Superdex75 column was similar. As a result, it was difficult to see the complex. To clarify whether the protein and *Ec*tRNA are making complex or not, the fractions were loaded on a native gel. But the complex wasn't observed (Fig. III-3-6).

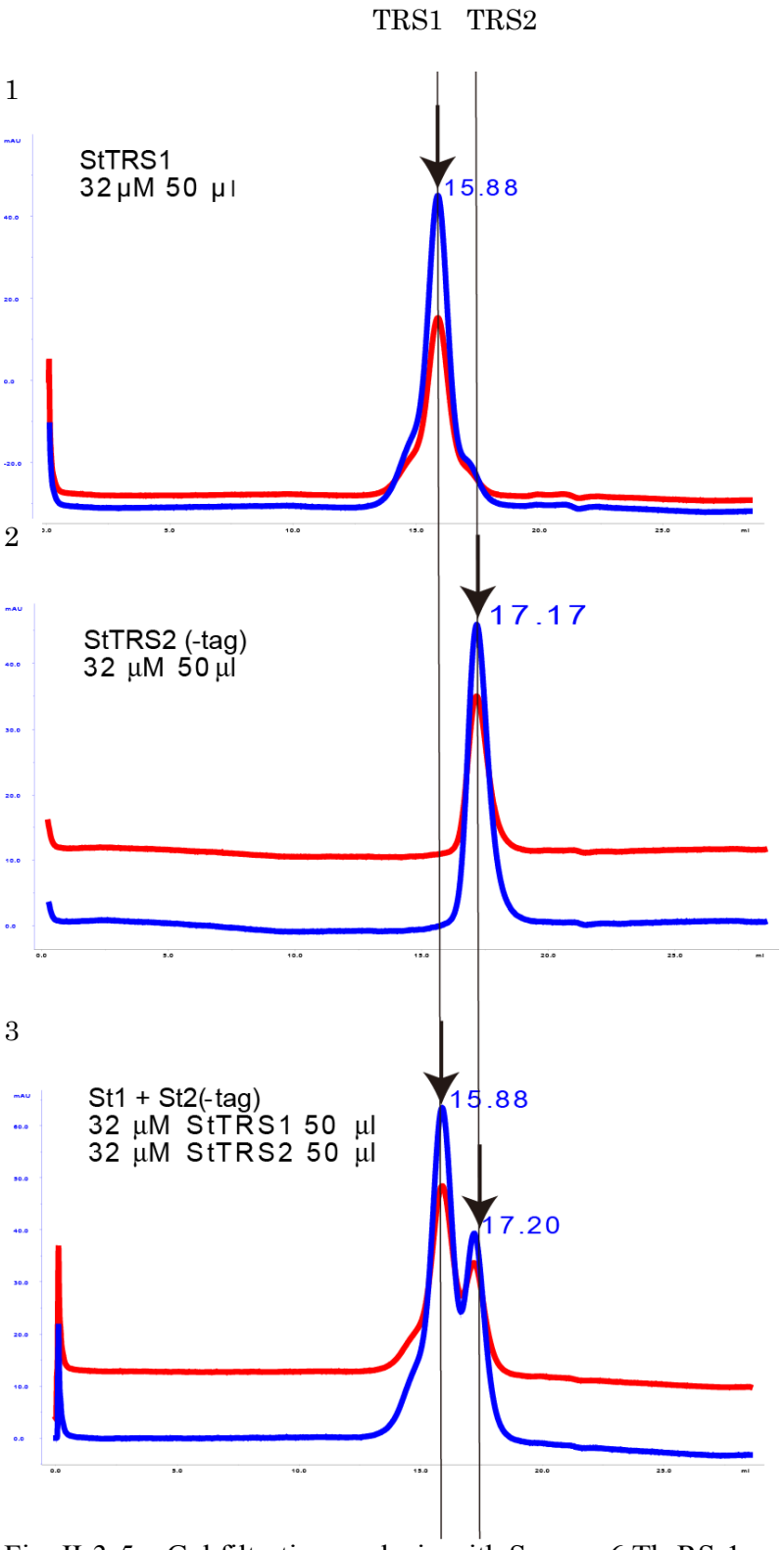


Fig. II-3-5 Gel filtration analysis with Sperose6 ThrRS-1 and ThrRS-2

Sample was incubated in 20 mM Hepes - NaOH pH 7.6, 100 mM NaCl, 5 mM $MgCl_2$ at 20 degree for 20 minutes and loaded to Sperose 6
1. 32 µM StThrRS-1,
2. 32 µM StTheRS-2
3. 32 µM StThrRS-1 and StTheRS-2

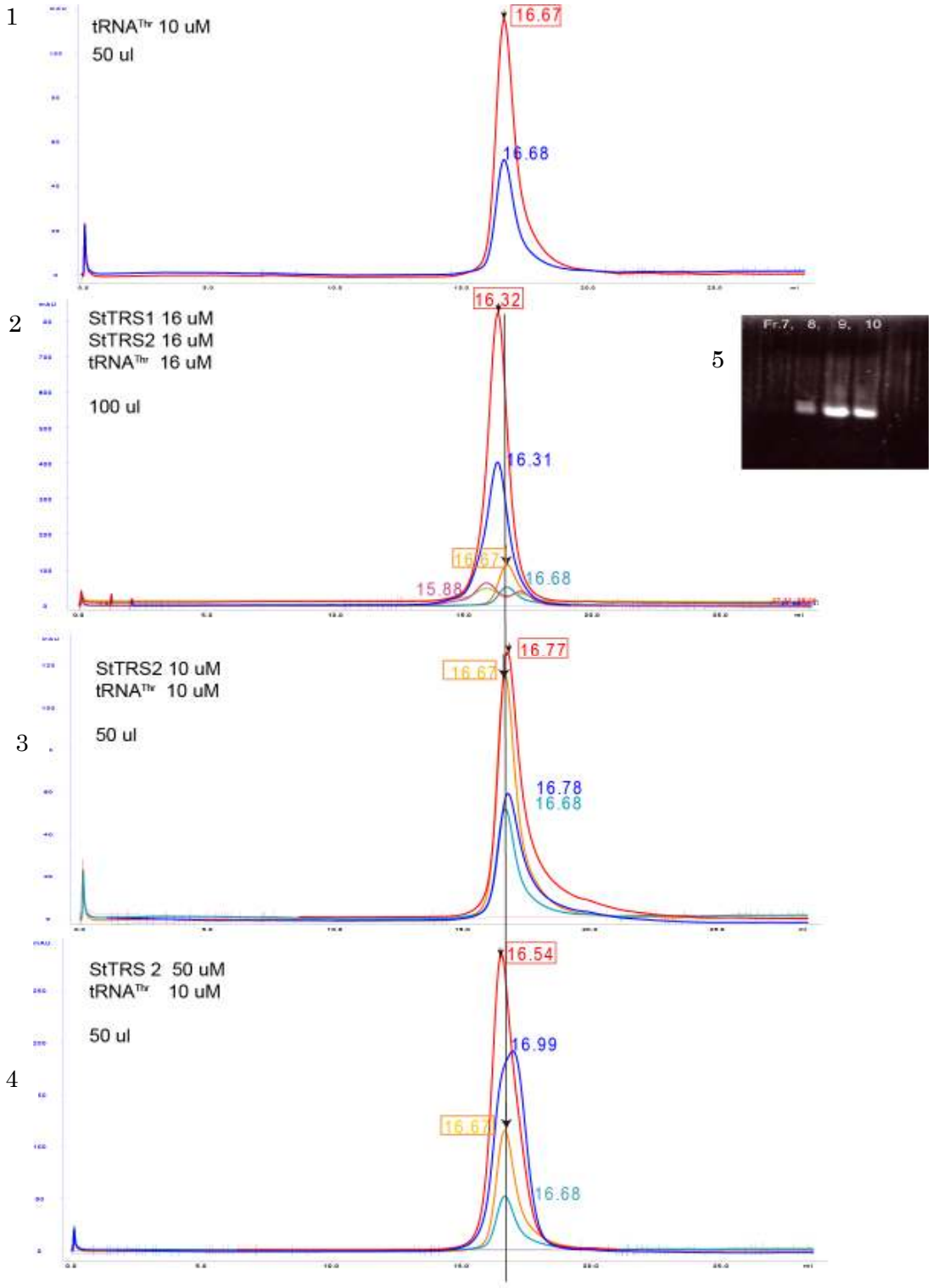


Fig. II-3-6 Gel filtration analysis with Sperose6 ThrRS-1, ThrRS-2 and tRNA

Sample was incubated in 20 mM Hepes-NaOH pH 7.6, 100 mM NaCl, 5 mM $MgCl_2$ at 20 degree for 20 minutes incubate and loaded to Sperose 6

1. 16 µM *Ec*tRNA, RED: OD=260, BLUE; OD=280
2. 16 µM StThrRS-1(Magenta :OD=260): StTheRS-2: (Orange: OD=260) *Ec*tRNA(RED: OD=260, BLUE; OD=280)
3. 10 µM StThrRS-2 (Orange; OD=260): *Ec*tRNA(RED;OD=260)
4. 50 µM StThrRS-2(Orange; OD=260): 10 µM *Ec*tRNA(RED;OD=260).
5. Peak fraction analyze by agarose gel.

3.6. Complexation assay with native gel electrophoresis

We then used a native agarose gel, and performed an EMSA binding test. In this case, the protein is in excess with respect to the tRNA. In contrast to GF column analysis, the EMSA showed a shift, indicating a complex, between protein and tRNA. StThrRS-2 bound clearly *Ec*tRNA. StThrRS-1 bound *Ec*tRNA only at high concentration. We then mixed StThrRS-1, StThrRS-2 and *Ec*tRNA, but were unable to detect a super shift which would indicate a ternary complex. The shift was almost the same as with StThrRS-2- *Ec*tRNA , therefore StThrRS-1 does not increase StThrRS-2 binding (Fig. III-3-7, III-3-8).

From this gel shift analysis, we conclude that StThrRS-2 binds *Ec*tRNA^{Thr} more strongly than StThrRS-1. Since StThrRS-1 has the anticodon binding domain and the catalytic domain, and StThrRS-2 has anticodon binding domain and editing domain, this means that the editing domain contributes to bind *Ec*tRNA. It will be interesting in the future to assess which residues are responsible for this contribution.

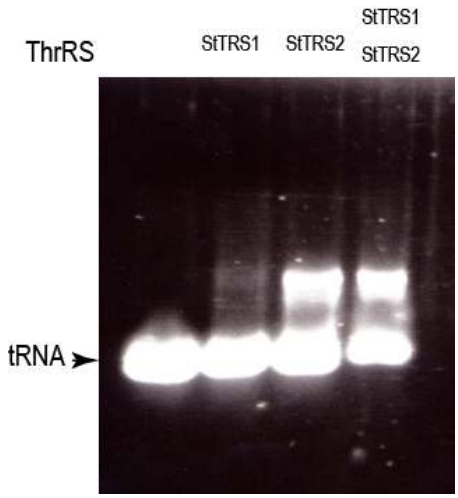


Fig. II-3-7 Native gel analysis

The protein (2μM) and tRNA(2μM) mixture incubated in 100 mM NaCl, 20 mM Tris-HCl pH 7.8 at 20 degree for 20 minutes. The agarose gel was dyed with ethidium bromide and exposed with UV. StThrRS-1-tRNA didn't shift and StTRS-2-tRNA shifted.StThrRS-1StTRS-2-tRNA shifted but there was not difference between StTRS-2-tRNA and StThrRS-1StTRS-2-tRNA.

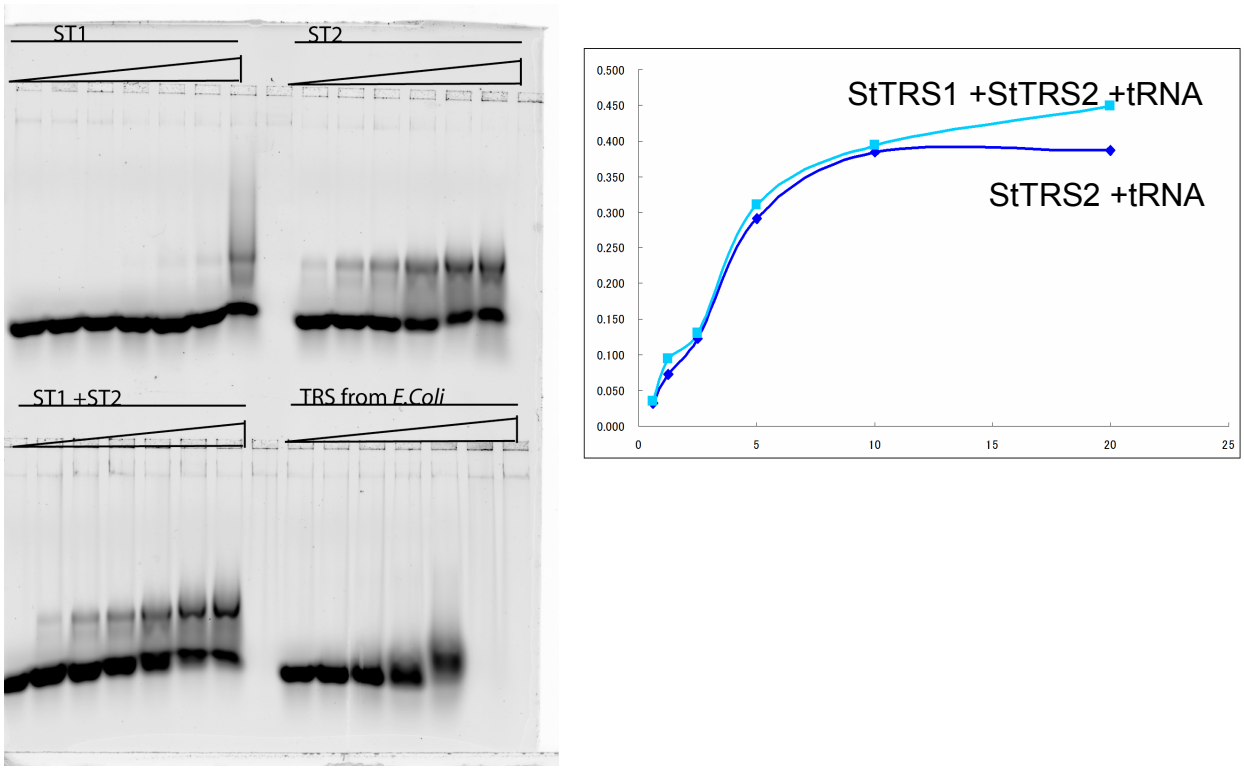


Fig. II-3-8 Native gel analysis

*Ec*tRNA(2μM) and StTRS-1, StTRS-2, StTRS-1-StTRS-2,(0.6 μM to 20 μM) incubate in 100 mM NaCl, 20 mM Tris-HCl pH 7.6, at 20°C for 20 minutes. The binding curve of StThrRS-2-tRNA and StThrRS-1-StThrRS-2-tRNA was almost same. StThrRS-1 was not increase binding ability for tRNA.

3.7. Crystallization

In order to understand the details of this increased binding, the crystallization of the binary StThrRS-2 and *Ec*tRNA complex was tried. The initial screening was done using Crystal screen I, II, JCSG+ at 4 degree.

Some crystals grew in 0.2 M $(NH4)_2So_4$, 0.1 M NaOAc 4.6, 25% PEG4000), but the crystal tested showed no diffraction.

So far this study indicated an interesting contribution of the crenarchaea editing domain to the interaction with tRNA. We will now try to improve the crystals, in order to solve the 3D structure of the binary complex, as it will help to understand the editing mechanism in crenarchaea.

Chapter III The function of LaRP7 in the 7SK snRNP

1. Introduction to 7SKsnRNP

1.1. Transcription

Transcription is the process of creating an equivalent RNA copy from a DNA sequence.

In eukaryotes, DNA is assembled into chromatin, which maintains genes in an inactive state restricting access to transcription machinery and its accessory factors. Chromatin is composed of histone, which form a structure called a nucleosome. Transcription of DNA is carried out by RNA polymerases. Whereas a single enzyme is responsible for this job in eubacteria and archaea, eukaryotes have three nuclear RNA polymerases to share the task of transcribing the nuclear genes RNA polymerase I (RNAPI), RNA polymerase II (RNAPII) and RNA polymerase III (RNAPIII), RNA polymerase IV(RNAP IV) is found only in plant. An additional RNA polymerase is found in mitochondria and chloroplasts, which carry a small DNA molecule of its own.

Three RNA polymerases are necessary because of the far greater complexity of most eukaryotic genomes. Within the nucleus, RNAP I is responsible for synthesizing most of the large rRNA, RNAP II synthesizes mRNAs and most of snRNA. RNAP III synthesizes a variety of structural and kinetic RNAs including tRNA ,7SK, 5S rRNA and U6 snRNA.

RNAP II transcription cycle can be divided into several distinct steps. First, RNAPII is recruited to the promoter of a gene, where it forms a pre-initiation complex with the general transcription factors. Initiation ensues, and RNAPII leaves the promoter behind in a process termed promoter clearance. Next, RNAPII enters prossessive elongation stage, which ends when the gene has been completely transcribed (Fig. III-1-1).

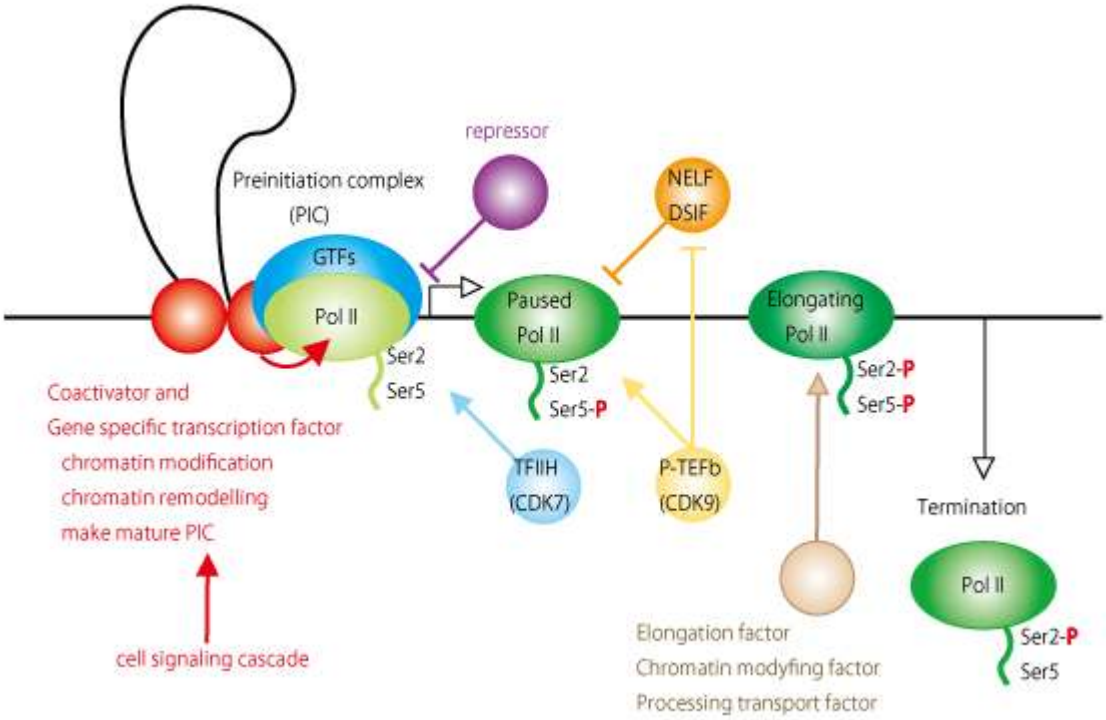


Fig. III-1-1 Transcriptional regulation in coding gene

Various environmental responses activate transcription factors recruitment to their response element. Distal factors interact with more proximal factors through DNA-looping. These gene-specific factors interact with coactivator and cooperatively act on the chromatin such as histone modifications and chromatin remodeling resulting in the basal transcription machinery assembly at the initiation site. Large protein network including general transcription factors (GTFs) as well as RNA polymerase II (Pol II) make the preinitiation complex (PIC) which can be modulated by some repressors. On initiation of transcription, Pol II carboxyl-terminal domain (CTD) is phosphorylated at Ser 5 by cyclin-dependent kinase 7 (CDK7) in TFIIH trigger to elongation. In higher eukaryotes Pol II arrests immediately after the initiation by the action of negative elongation factor (NELF) and DRB sensitivity inducing factor (DSIF) to generate the paused Pol II. CDK9 in P-TEF-b phosphorylates the Pol II CTD at Ser 2, as well as at NELF and DSIF, resulting in resumption of elongation. Ser 2 and 5 phosphorylated form of Pol II can recruit the various factors that are involved in chromatin modification, maintain the elongation, RNA processing, and termination.

All three eukaryotic RNA polymerases are very similar in structure and subunit configuration, RNAP II uniquely possesses an extra C-terminal domain (CTD) on its largest subunit, Rpb1. Many functions have been proposed for this domain, including interactions with nucleic acids and displacement of nucleosomes. However, most evidences indicate that the CTD primarily functions as a binding platform for other proteins involved in transcription, mRNA processing, and histone modifications. The CTD consists of multiple repeats of the heptamer sequence Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7, and as expected from this sequence, this region is highly phosphorylated in transcribing RNAPII. Serine 2 and 5 (Ser2 and Ser5) were identified as major phosphorylation sites (Buratowski 2009), and multiple functions for these modifications have been elucidated. More recently, phosphorylation of serine 7 (Ser7) and other covalent modifications have been described (Boeing et al. 2010).

1.2. Elongation

The elongation cycle comprises three basic steps (a) binding of a template-complementary nucleoside triphosphate (NTP) into the active site; (b) chemical reaction of the RNA chain 3'-OH with the NTP α-PO4, catalyzed by a pair of bound Mg2+ ions, resulting in one NMP addition to the RNA and liberation of pyrophosphate; and (c) translocation of the nucleic acid assembly to place the next template base in the active center. The elongation complex of RNAP is stabilized by several sets of interactions among protein and nucleic acid

After recruitment to the promoter, the polymerase simply behaves like a machine, quickly reading the gene. However, over the past decade a revolution in this thinking has occurred, culminating in the idea that transcript elongation is extremely complex and highly regulated and, moreover, that this process significantly affects both the organization and integrity of the genome.

1.3. Pausing mechanism (Fig. III-1-2)

For numerous genes it has been known that RNAP II pauses soon after promoter escape. This promoter proximal pausing was first described for the Drosophila heat-shock genes (Rougvie and Lis, 1988). Under non-inducing conditions, RNAP II fully occupies the HSP70 promoter and transcribes to between +20 and +40 where it pauses. RNAP II only resumes transcription under inducing conditions, such as heat shock. This facilitates rapid induction of the gene's expression. Similar mechanisms have been described for a handful of

viral and mammalian genes (Saunders et al. 2006). This indicates that postrecruitment regulation occurs much more often than was previously assumed. Promoter-proximal pausing is mediated by the action of pause factors, these include 5,6- dichloro-1-b-D- ribofuranosyl benzimidazole (DRB) sensitivity-inducing factor (DSIF) and negative elongation factor (NELF). The factors that alleviate promoter-proximal pausing by DSIF and NELF include TFIIS and the positive transcription- elongation factor-b complex (P-TEFb) (Peterlin et al. 2006).

Transcription elongation is positively regulated by the P-TEFb. P-TEFb is a heterodimer consisting of cyclin-dependent kinase 9 (Cdk9) and one of the cyclins T1, T2a, or T2b.

Cdk9 is the kinase inhibited by DRB. Cdk9 phosphorylates serine 2 within the heptapeptide repeats of the carboxyl-terminal domain of the largest subunit of RNAP II. Cdk9 also phosphorylates negative elongation factors NELF and DSIF, leading to dissociation of NELF from paused RNAP II and conversion of DSIF into a positive elongation factor. P-TEFb releases NELF from the elongation complex, either directly or indirectly, which leads to release of pausing factors and association of factors that promote productive elongation. This is a pivotal point of regulation in RNAP II transcription (Margaritis et al. 2008).

The amount and availability of active P-TEFb in the cell is controlled. 7SK small nuclear RNA (7SK snRNA) binds to P-TEFb and inhibits its kinase activity. Due to its abundance, 7SK has been known since the 1960s. Its function as a transcriptional regulator, however, has only recently been discovered. The high conservation and abundance of 7SK suggest an important physiological function of this RNA. In Hela cells, about half of P-TEFb is kinase inactive and is associated to the 7SK snRNA and to the hexamethylene bisacetamide (HMBA) inducible protein HEXIM1 and less after HEXIM2. Human 7SK RNA is an abundant 332 nucleotides nuclear transcript generated by RNA polymerase III. 7SK RNA has been evolutionary conserved in vertebrates and homologues are found in annelid, mollusc and insect genomes. 7SK RNA folds into several hairpins that serve as specific platforms for binding proteins (structure details in III-2-4). 7SK RNA is a paradigm for non-coding RNAs regulating transcription. P-TEFb and 7SK associate in a specific and reversible manner (Yang et al. 2001; Nguyen et al. 2001). 7SK also associates with its specific methyl-phosphate capping enzyme (MePCE) and a La-Related Protein 7 (LaRP7). These two proteins form a core complex with 7SK (Xue et al. 2010, Diribarne et al. 2009, Krueger et al.2008). When the cell is exposed to inhibitors of transcription such as actinomycin D or DRB, P-TEFb dissociates from 7SK and HEXIM1/2 and become free active P-TEFb. The released 7SK

RNA associates with a subset of heterogeneous nuclear ribonucleoproteins (hnRNP).

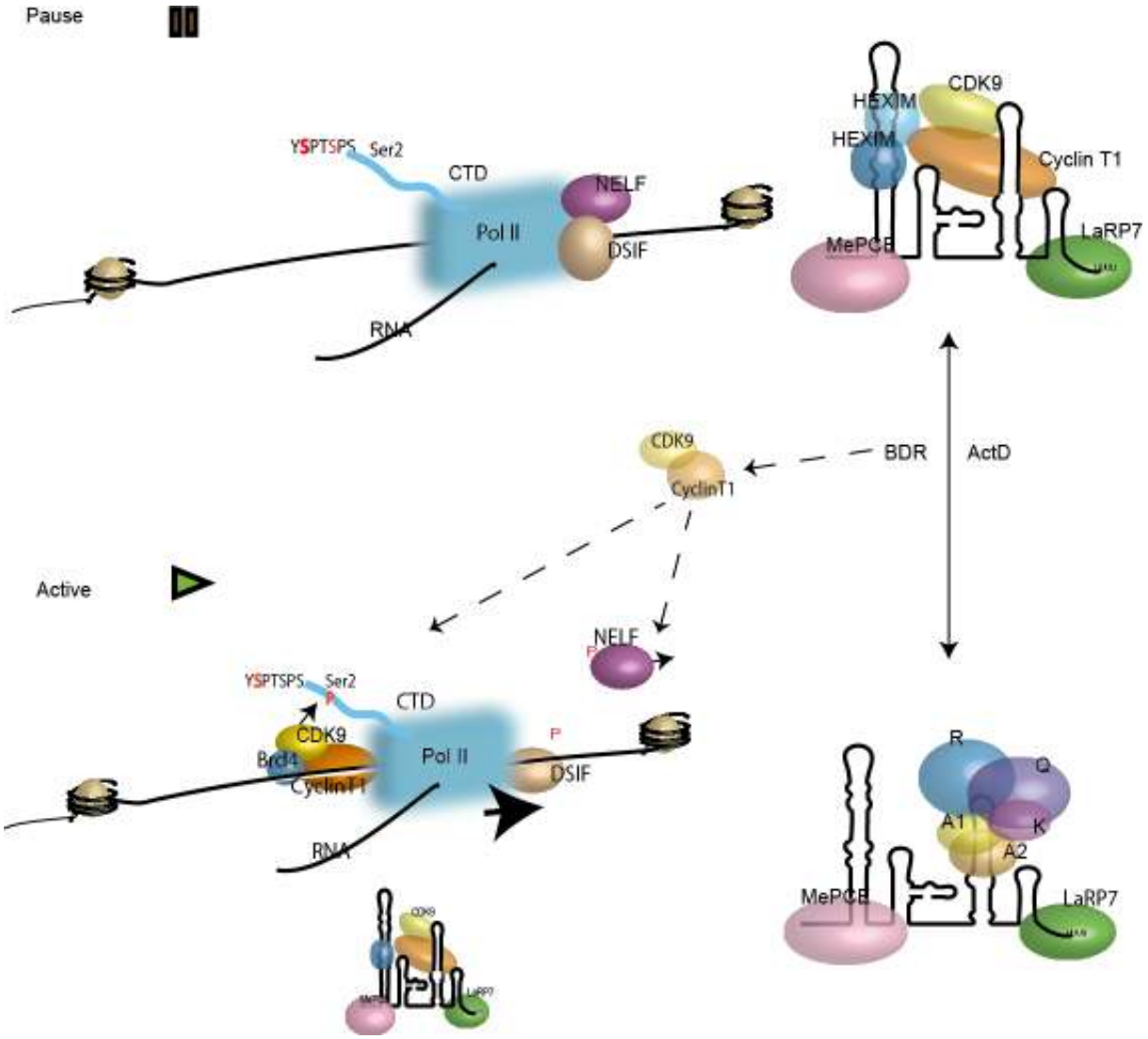


Fig. III-1-2 Pausing mechanism of Transcription

1.4. LaRP7

LaRP7 is an RNA binding protein with high homology to human La protein (III-1.5). LaRP7 binds to and enhances the stability of the 7SK snRNA along with MeCPE, in a core 7SK snRNP). LarRP7 is bound with 7SK even under condition of stress (actinomycine D, DRB) and high salt (0.8M KCl) (He et al. 2008).

From a sequence analysis, LaRP7 comprises the same domain as La protein previously implicated in specific UUU-3′OH recognition, the LAM and RRM in its N-terminal region. There is an additional RRM2 at C terminal region (Fig III-1-3, Fig III-1-4). C terminal RRM2 is frequently deleted in human tumor. As 7SK RNA is a RNAP III transcript ending in UUU-3′OH, it seems likely that LaRP7 would use a binding mode similar to La for the engagement of the 7SK 3′ end. In support of this, EMSA experiments with LaRP7 deprived of the N-terminal LAM motif or 7SK lacking its 3'UUUU end shown that their interaction decreased. On the other hand, the Y127D mutation in Larp7 located in the RRM1 of the La domain, significantly reduces the affinity for 7SK RNA, while mutation of the equivalent residue in human La (Y114A) has little effect on pre-tRNA binding interaction (He et al. 2008). Replacement of the LAM (where the majority of UUU-3′OH specific contacts are made) from LaRP7 with that of La results in decrease in 7SK RNA binding in vivo (He et al. 2008). Consistent with this, other regions provide RNA substrate specificity and these differ for La and LaRP7. More significantly, and consistent with a 7SK specificity determinant residing in the RRM of the La domain of LaRP7, replacing the LaRP7 RRM1 with the La RRM1 led to significant decreases in 7SK RNA binding and 7SK snRNP formation in vivo (He et al. 2008). While the loop-3 of the RRM1 of La provides significant affinity for pre-tRNA, it remains unclear if this region of Larp7 is important for 7SK RNA binding. Because of the basic nature of loop-3 of La RRM1, it is tempting to speculate that the La RRM1 exhibits general affinity for RNA, perhaps recognizing RNA backbone or other features of structure with little requirement for sequence specificity, consistent with its ability to engage a wide variety of different pre-tRNA sequences as well as other RNAs, whereas the RRM1 of LaRP7 is likely to have adopted a sequence-specific interaction with 7SK RNA. As alluded to above, the function of the C-terminus of LaRP7 in RNA recognition appears to be complex, since deletion of the sequence extending from RRM2 to end or only 20 residues from C-terminal (it does not contain RRM2) of the LaRP7 C-terminal region, both results in defects in 7SK RNA binding and 7SK snRNP formation in vivo (He et al. 2008). Other studies suggest more modest defects by the N-terminal La domain 7SK RNA binding assays

in vitro (Markert et al. 2008). In summary, the data suggest that the La domain of both La and LaRP7 provide specific and high-affinity binding to UUU-3′OH, but the LaRP7 C-terminal region seems to enhance binding to 7SK RNA while limiting binding to UUU-3′OH containing RNAs other than 7SK RNA. Apart from the N- and C-terminal regions, the rest of the molecule is not predicted to have a defined structure, but contains stretches of positively charged residues. The current model is that whole LaRP7 structurally stabilize 7SK (Markert et al. 2008, Krueger et al. 2008)

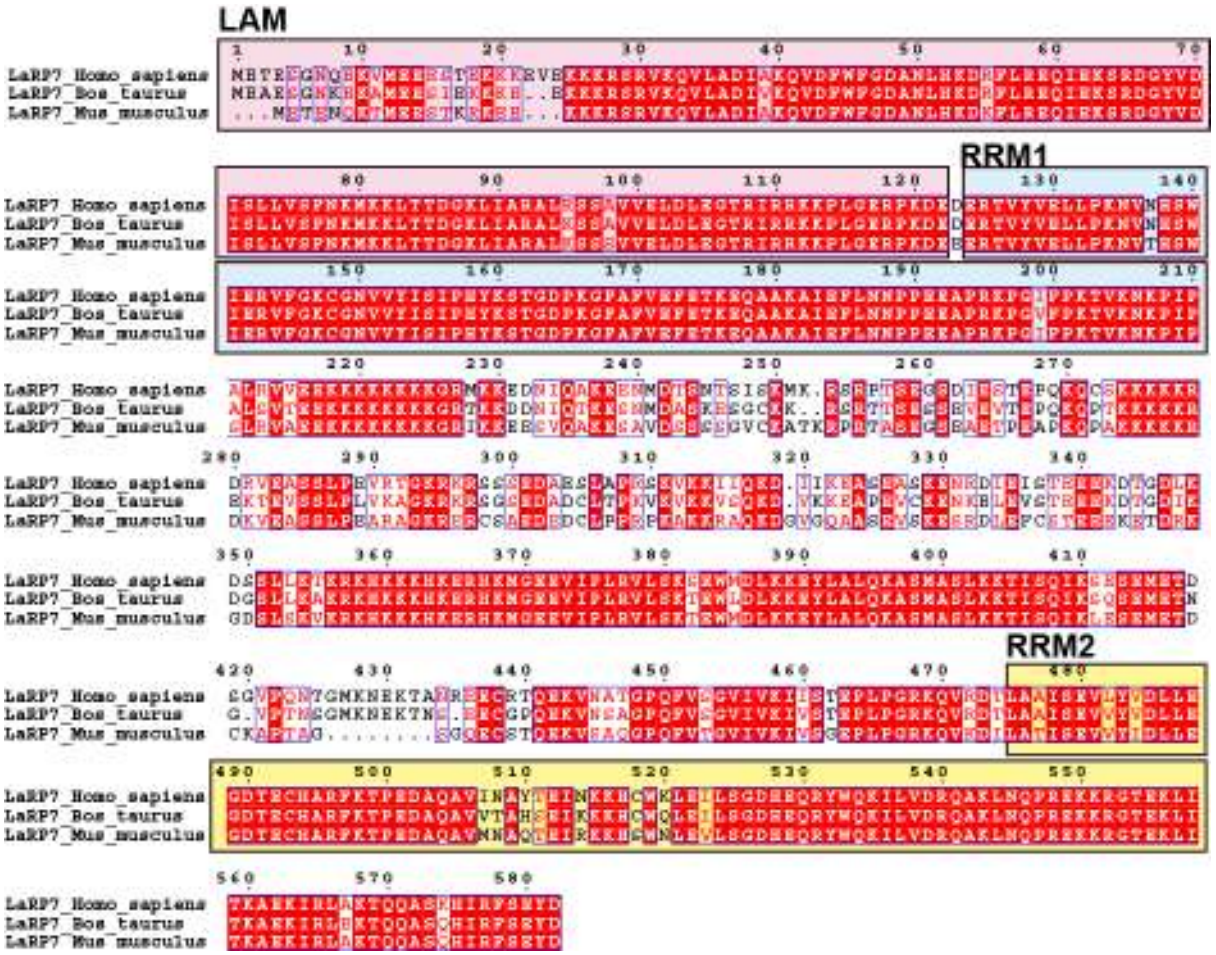


Fig. III-1-3 Sequence alignment of LaRP7 Homo_sapiens, Bus_taurus, Mus_musculus.

Pink and Blue is N-terminal region contains LAM and RRM. Yellow region is C-terminal region contains RRM2. Between Blue and Yellow region the linker region doesn't have structure.

1.5. La protein

. The La protein is a highly abundant nuclear phosphoprotein that is conserved in eukaryotes and was originally identified as an autoantigen associated with rheumatic diseases such as lupus erythematosus and Sjogren syndrome. Eukaryotic La protein recognizes the UUU-3′OH sequences of nascent RNAP III transcripts to assist folding and maturation (Inada et al. 2004, Kufel et al. 2000, Xue et al. 2000, Belisova et al. 2005). Most RNAP III transcripts are bound immediately after synthesis by the La protein. The 3' ends of such RNAs are bound by the N-terminal domain of La protein. In most cases, the binding of the La protein is only transient, the protein being displaced upon excision or modification of the 3'-UUU-OH motif.

In the La protein the N-terminal domain consists of La (LAM and RRM1) whereas the C terminal consists of the RRM2 and unstructured long flexible element. The 3D structure of LAM has been characterized and adopts a winged-helix like fold, preciously found in DNA and RNA binding proteins. RRM is also a very abundant RNA binding motif (described in the General Introduction). The 3D structure of the N-terminal domain of La protein has been solved by NMR, LAM and RRM separately. Several crystal structures have elucidated the interaction with RNAs. One was solved with double stranded RNA, the others were solved with single stranded RNA oligomers (Teplova et al. 2006, Kotik-Kogan et al. 2008, Maraia et al. 2006, Alfano et al. 2004). Fig. III-1-5 shows the RNA-bound La protein structures superimposition. In both structures, RNA binds within the cleft between the La motif and RRM1 domains. The La motif interacts extensively with RNA through only a few non-canonical contacts between the edge of the RRM1 β-sheet.

.

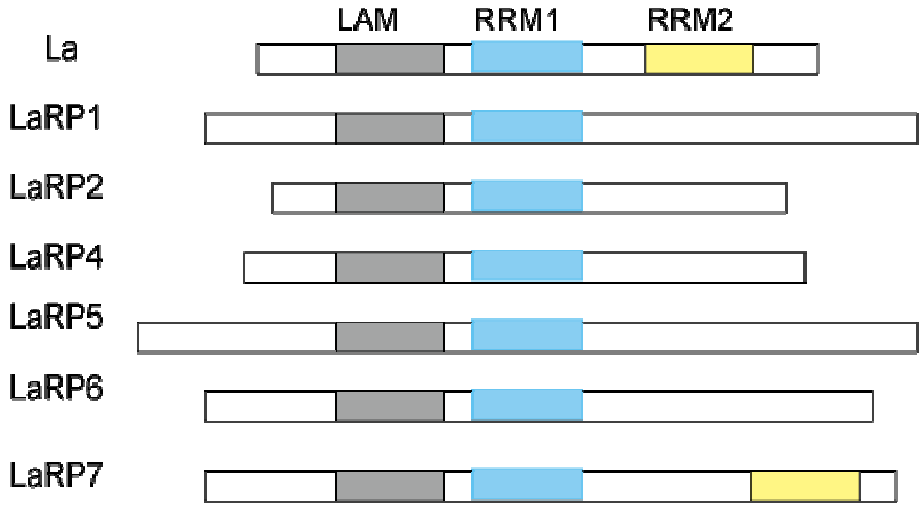


Fig. III-1-4 LaRP families Only La and LaRP7 has RRM2

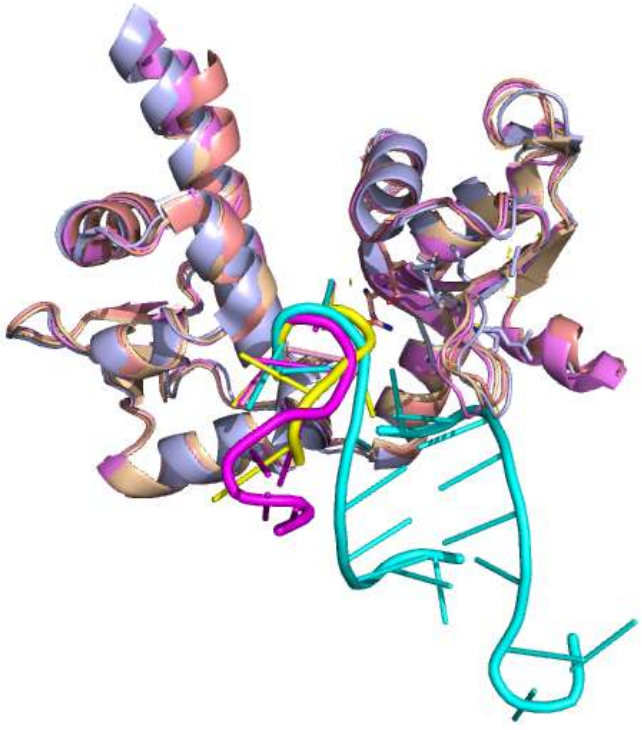


Fig. III-1-5 Superimposition of La proteins
Blue is La protein and dsRNA complex (PDB ID 1ZH5) Magenta and Yellow are (PDB ID 2VOP, 2VOO)

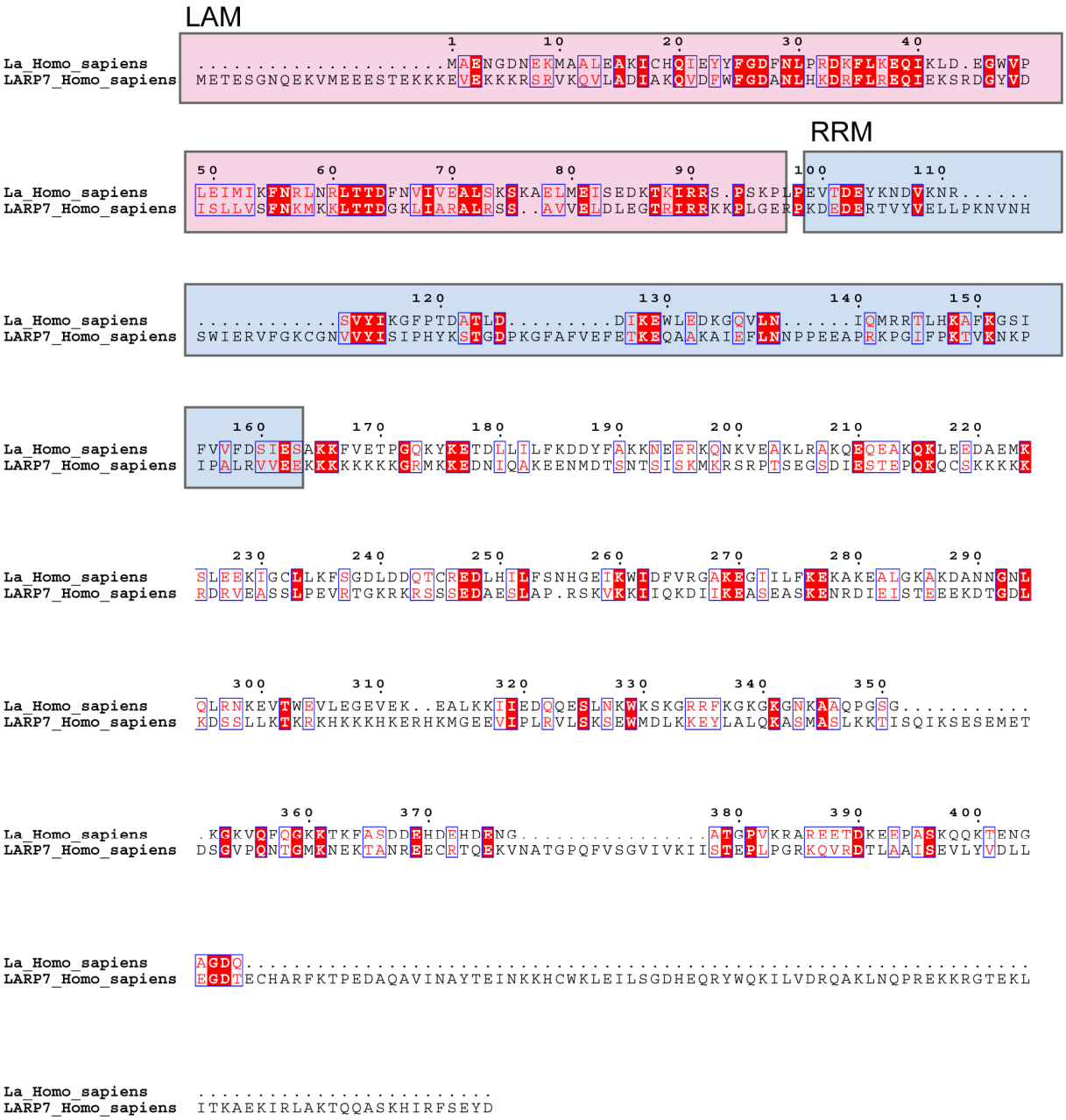


Fig. III-1-6 Sequence Alignment of La protein Homo_sapiens and LaRP7 Homo_sapiens

There are several La protein families. LaRP7 has homology with p43 (Eupliotes aediculatus) and p65 (Tetrahymena thermophila), which for p65 has been shown to assist in the correct folding of the telomerase RNA and hierarchal assembly of the RNP. The role that p65 plays in the assembly is different from that provided by general chaperone activity like La protein (Stone et al. 2007).

LaRP6 family members are characterized by also having a more central, RRM containing La domain and a newly described motif of unknown function at the extreme C-terminus called the LSA motif. LaRP6 interacts with transcription factors, and functions upstream of the transcription factor MyoD, to control muscle development. As a transcription factor-associated protein, LaRP6 may function in the control of mRNA synthesis, and this could conceivably be related to the way LaRP7 controls P-TEFb activity, although no small RNA target (i.e., akin to 7SK snRNA) has been identified for LaRP6

LaRP1 members lack a typical RRM after the LAM. Instead, many LaRP1 family members are predicted to contain an RRM that adopts the RRM fold but lacks the consensus sequences, though some members of this family may have no recognizable RRM adjacent to the LAM. LaRP1 has the ability to bind poly (U) and poly (G). *C.elegans* LaRP1 resulted in an increase in the steady state levels of certain mRNAs. But LaRP1 function and possible mechanisms for target RNA binding and recognition are still unclear. LARP1 was recently identified as a factor important in the replication of human influenza virus.

The LaRP4 family lack of the conserved key side chains in the LAM that are used by La proteins to recognize UUU-3'OH. LaRP4 is also a cytoplasmic, polyribosome-associated protein that interacts with poly-A binding protein (PABP) in an RNase-insensitive manner. LaRP4 over expression stabilizes a luciferase reporter mRNA. These observations suggest that, LaRP4 acts as a positive factor to promote mRNA stability, in contrast to what is suspected for LaRP1. So far LaRP family structure of LaRP–RNA interaction has yet to be solved.

2. Experimental investigation of the function of LaRP7

As I described in La protein part, only La protein 3D structure was solved with X-ray and NMR. Therefore, to understand the role of LaRP7 in the stabilization of 7SK, powerful information would be gained by solving the 3D structure of the complex by X-ray crystallography. Results depend upon obtention of crystals, and it is important to think about crystal packing to design the best protein and RNA candidates to make the crystal. Further, to enhance the chances of crystallization, conditions have to be found where both partners, but especially the protein, are conformationaly as stable as possible, soluble, and monodisperse. This section describes our design of a construct of LaRP7 suitable to crystallization, purification of the protein, and crystallization assays.

2.1. Design of a construct of LaRP7 suitable for crystallization

2.1.1. Sequence analysis

When analyzing the LaRP7 sequence (Fig. II-2-1) with a program predicting secondary structures such as Predict Protein in the Expasy website, potential formation of helices or strands are indicated only in the N-terminal and C-terminal regions. LaRP7 N-terminal region (aa 1-210) is predicted to comprise the La motif (LAM) and subsequent RNA binding motif (RRM), upon which relies the similarity with the La protein that induced its name (La-Related Protein). The C-terminal rejoin contains another, less obvious RRM, RRM2. Between RRM1 and RRM2, there is long linker region, for which no structure is predicted; it could be structurally unfolded.

Functionally, the N-terminal domain 30-210, is indeed able to complement La. Previous studies showed that LAM-RRM region has affinity for U-rich sequences, present in the 3'-end of RNAPIII transcripts like tRNA precursors or RNAs of the splicing machinery. 7SK, also a product of RNAPIII, has likewise a U-rich sequence at its (299-332) 3’ terminal end. Deletion of this sequence has been shown to suppress LaRP7 binding (He et al. 2008). Similarly, deletion of the N-terminal LAM domain abolishes 7SK recognition.

2.1.2. Design of constructs

The 3D structure of the La protein has been solved by NMR (Alfano et al. 2004), and several complexes with a short duplex RNA, or with several 8-mer oligo RNA were solved

with X-ray crystallography (Maraia et al. 2006, Kotik-Kogan et al. 2008). The fact that the La protein has been crystallized prompted us to attempt the crystallization of the N-terminal region of LaRP7, for which a construct had to be designed in the purpose of crystallization.

Sequence analysis of this region of LaRP7 shows that it is highly conserved, but 30 residues longer when compared with hLa protein. Fig. II-2-1 shows that sequence similarity allows to predict residues 30 to 120 to fold as LAM, aa 121 to 208 is the RRM and residues 1 to 29 comprise an extension that could be unfolded. LAM and RRM parts (La domain) superimpose well with the La construct that was crystallized, apart from several notable differences in the RRM region. These are indicated in Fig. II-2-2, where the extension of 6 residues in La protein is in purple and an extension in LaRP7 of 6 residues is represented in red. This analysis suggests that the RRM of La could be better folded than the RRM of LaRP7. The sequence of LaRP following aa 209 is estimated to be unfolded. The sequence was however also analyzed for the distribution of positively charged residues, expected to be involved in the binding of the negatively charged RNA. The N-terminal region 1 to 215 has 19% positive charged residues, and the position of these residues were well conserved between La protein and LaRP7 N-terminal region. These are mainly located at the surface of the protein. Immediately after the La domain is a stretch of positively charged lysines; since these could be functionally important, we chose to construct LaRP7N1_228 (including the stretch of lysines) and LaRP7N1_208 (stops just after the final helix of RRM) as our first candidates for crystallization (Fig III-2-3).

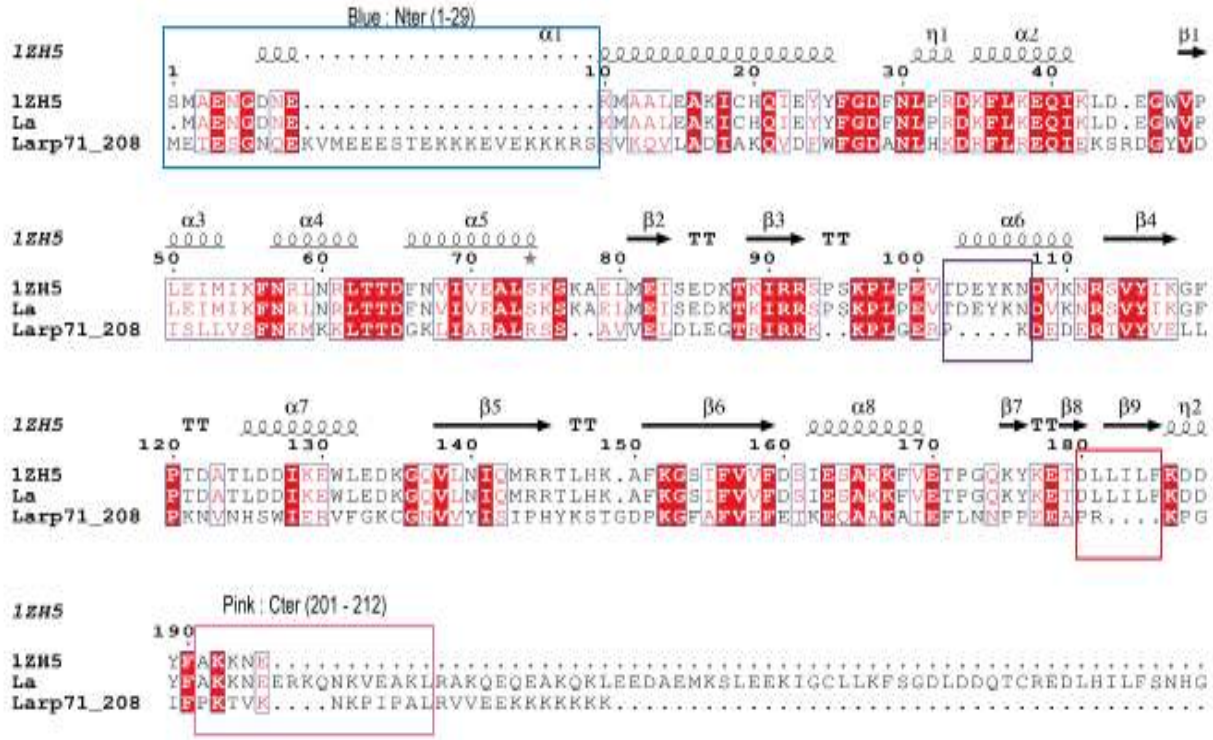


Fig. II-2-1 Sequence alignment of La protein and LaRP7N1_208

1ZH5 is PDB number of La protein structure. Blue area is additional region for LaRP7. Purple is additional α helix for La protein. Red is additional β strand for La protein. Pink is loop region.

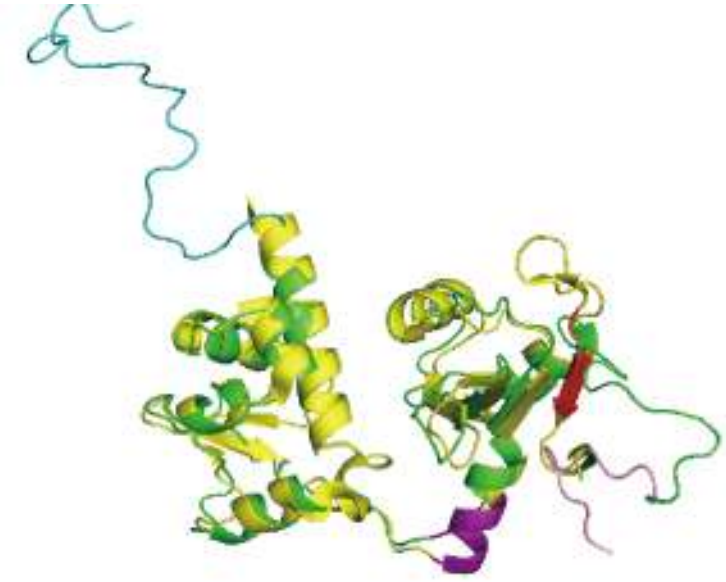


Fig. II-2-2 Superimposition of La protein structure and predicted model of LaRP7N1_208

Yellow is La protein structure (PDB ID 1ZH5) Green is prediction structure of LaRP7 N-terminal region by 3Djigsaw. Blue, Purple, Red, Pink is correspond to Fig II-2-1 alignment

Crystallization is driven by the free energy change from the supersaturated solution of protein to protein crystals in the solvent. Ability to crystallize is associated with surface properties of the proteins. The flexible polar side chains such as those of lysines and glutamates are expected to impede the protein's ability to form stable intermolecular contacts and thus to assemble into a crystal lattice. This prompted us to construct additionally a shorter version, deprived of the potentially unfolded extension at N-terminus, and starting at residue 28, named LaRP7N28_208.

On the other hand, it should not be forgotten that the second RRM and the linker region have functional importance. The linker sequence from 208 to 475 has 27% positively charged amino acid, with almost all positively charged amino acids in a cluster. This suggests that the linker region makes also contact with the 7SK RNA. The full-length protein was also cloned as well as the RRM2-containing fragment (474-582). While the latter was not used during the period allowed for my experimental work, the full-length protein was purified, and used as a control in the binding assays. Fig III-2-4 summarizes all the constructs for LaRP7 used in this work.

A

The sequence of N terminal region of LaRP7

METESGNQEKVMEEESTEKKKEVEKKKRSRVKQVLADIAKQVDFWFGDANLHKDRFLREQ
IEKSRDGYVDISLLVSFNKMKKLTTDGKLIARALRSSAVVELDLEGTRIRRKKPLGERPK
DEDERTVYVELLPKNVNHSWIERVFGKCGNVVYISIPHYKSTGDPKGFAFVEFETKEQAA
KAIEFLNNPPEEAPRKPGIFPKTVKNKPIPALRVV

B

The sequence of Linker region of LaRP7

EEKKKKKKKKGRMKKEDNIQAKEENMDTSNTSISKMKRSRPTSEGSDIESTEPQKQCSKKKKKRD
RVEASSLPEVRTGKRKRSSSEDAESLAPRSKVKKIIQKDIIKEASEASKENRDIEISTEEEKDTG
DLKDSSLLKTKRKHKKKHKERHKMGEEVIPLRVLSKSEWMDLKKEYLALQKASMASLKKTISQIK
SESEMETDSGVPQNTGMKNEKTANREECRTQEKVNATGPQFVSGVIVKIISTEPLPGRKQVRDT

C

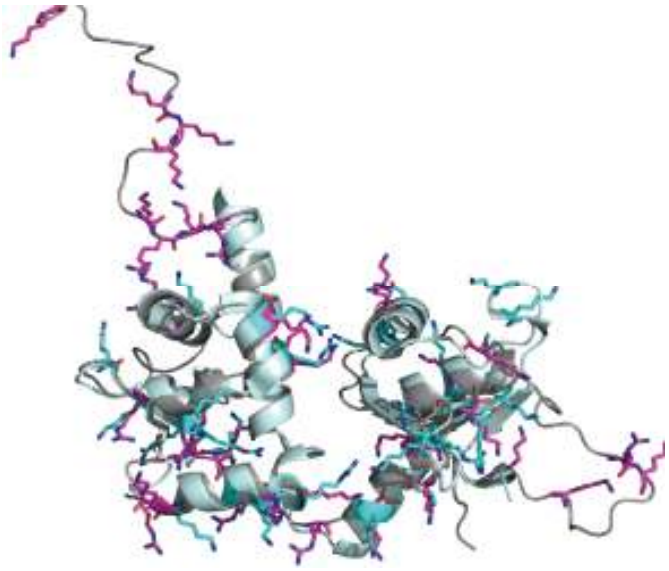


Fig. III-2-3 The Positive and Negative charge of LaRP7

A) The N-terminal region of LaRP7 all positive charge amino acid Lys in Red, Arg in Orange, His in Yellow. B) The sequence of linker region of LaRP7 Red, Orange, Yellow are positive charge amino acid. Blue, Sky blue are negative charge amino acid. C) Blue is structure of La, Gray is prediction structure of LaRP7N1_208. Blue Lys residues are from La protein, Magenta Lys residues are from LaRP7N1_208

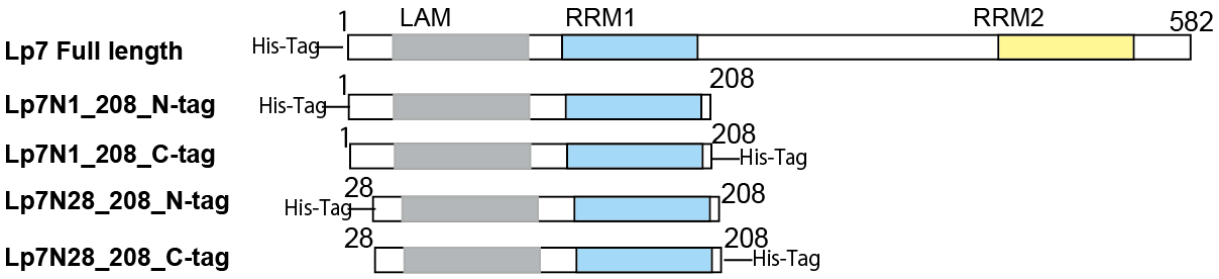


Fig. III-2-4 Summary of the constructs of LaRP7 used in this work

LaRP7N1_208_N-tag is the construct that 1 to 208 residues of LaRP7 and has His-Tag in its N-terminal. Lp7N1_208_C-tag is the construct that 1 to 208 residues of LaRP7 and has His-Tag in its C-terminal. Lp7N28_208_N-tag is the construct that 28 to 208 residues of LaRP7 and has His-Tag in its N-terminal.Lp7N28_208_C-tag is the construct that 28 to 208 residues of LaRP7 and has His-Tag in its C-terminal.

2.2. Purification of LaRP7 N-terminal domain

2.2.1. Expression of the different constructs

Crystallization needs mg amount of pure and homogeneous protein. Among the various expression systems available, *E.coli* system allows to easily and quickly express the intended protein, but does not allow for post-translational modifications, such as phosphorylation for instance. In the absence of known requirement for such modifications of LaRP7, we chose *E.coli* expression system. The designed constructs were cloned in a vector of the pnEA series, engineered by C. Romier in our laboratory. These vectors are derived from the pET15, and provide the expressed protein with a tag (various tags are available, such as His6, GST, etc …), that can be positioned on either side (N-terminal or C-terminal), are cleavable (recognition sequences for thrombin, TEV, or P3C proteases can be chosen). The expression level from pnEA vectors is generally good, and, importantly, these vectors are compatible with another vector type, called pnCS, which opens the possibility for co-expression of several proteins. The vector map and cloning procedure are given in Chap. IV-2.

The different constructs were expressed in *E.coli* strain BL21(DE3) pRARE, the pRARE plasmid providing for several tRNAs able to decode rare codons. Cells were grown in auto-induction medium. As this medium contains a mixture of glucose and lactose, a period of cell growth is followed by spontaneous induction of protein expression without monitoring cell density and without conventional induction with IPTG.

Small cultures (4 ml) were used for assessing the expression level, but generally 1L cultures were done in 5L flasks, under agitation. In order to maximize the production of well-folded, soluble proteins, growth was at low temperature (25 °C) and for long time (15 to 24 hours), aiming for an OD_{600} greater than 6 to maximize expression in the lactose-induced phase.

After harvest, the pellet of bacteria was resuspended in lysis buffer with protease inhibitors. Lysis was performed by short bursts of sonication, with the solution maintained on ice. Cell debris and aggregated proteins were removed by centrifugation at 20000 rpm. To obtain high purity protein, controlling the expression is important. The level of over-expression of soluble protein with respect to *E.coli* proteins, was analyzed on denaturing gels, and compared with the cellular extract obtained after sonication (before centrifugation).

The LaRP7N1_228 protein was poorly soluble, and change of lysis buffer (for the various lysis buffer which were currently tried, see Chap. IV) did not improve the solubility. We turned to LaRP7N1_208, which showed enough expression of soluble protein to proceed to purification.

2.2.2. Purification of the N-terminal domain of LaRP7: General process

High level of purification is very important part to grow crystals. Purification should eliminate proteins that coexist with LaRP7 in the soluble fraction. At the same time the protein has to be maintained in the correct conformation and aggregation prevented. During purification, the protein should be kept in optimal condition. To reach the quality of purification which gives crystals, a lot of trials have been done. The factors currently varied are pH, temperature, salt concentration and additives. During the purification, the protein ,sensitive to temperature, should be maintained at 4°C.

First we searched the best lysis buffer. We considered several buffers, including Tris-HCl pH 8.0, Tris-HCl pH 7.4, K-Phosphate buffer pH 8.0, but observed no obvious improvement of solubility in these buffer, and decided to continue with Tris-HCl pH 8.0 buffer, compatible with the next step. Protein was then purified by Ni-chelating affinity chromatography, followed by gel-filtration (GF) on a Superdex 75 16/60 column. This step is important to eliminate aggregates. We always observed that in the chromatogram of the GF column, the absorbance at 260 nm (OD_{260}) was higher than at 280 nm (OD_{280})), an indication that nucleic acids were not eliminated enough. We could improve the elimination of spuriously bound nucleic acids at the Ni-affinity step, by washing the beads with high salt buffer, after the protein was bound to Ni-beads. Gel-analysis of the protein after two purification steps showed that some contaminants were still present. To increase the purity, the protein was loaded on an ionic-exchange SP column. The protein was eluted by a gradient of salt concentration, around 600 mM NaCl. After SP chromatography, the protein was dialyzed and concentrated (Fig. II-2-5).

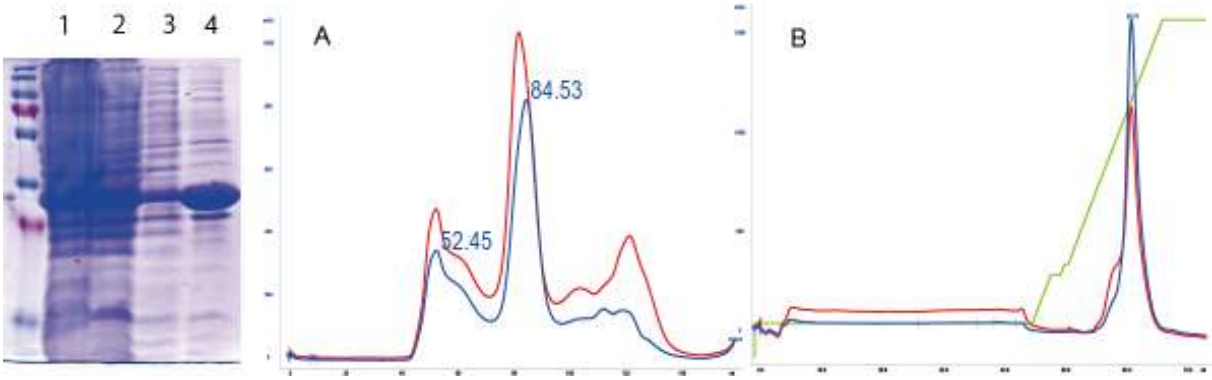


Fig. III-2-5 Purification of LaRP7N1_208

1 Protein debris protein total, 2. Supernatant of cell debris, 3. wash solution of Ni-chelating column, 4. elution of Ni-chelating column.
A. purification by GF column, B. purification by SP column

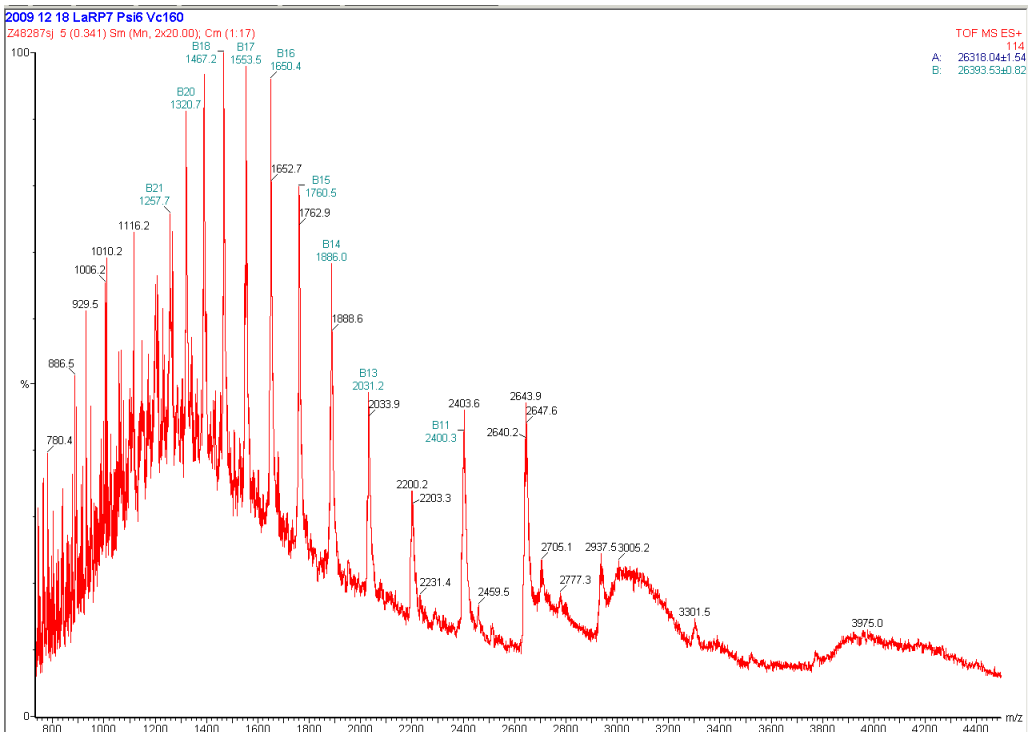
Concentration of LaRP7N1_208 at 10 mg/ml could be attained, but a DLS analysis showed that the protein was aggregated. We tried to optimize the storage buffer by varying KCl, detergent and $MgCl_2$ concentrations, while keeping 20 mM Hepes-KCl pH 7.6 (Phosphate was avoided, because it crystallizes easily, as was Tris, since its pH is very sensitive to temperature changes). The KCl concentration was changed from 50 mM to 100 mM, 200 mM, or 250 mM. The detergents tried were 5 mM CHAPS or 0.1% Triton-X. The $MgCl_2$ concentrations tried were 2 mM, 5 mM, 10 mM. Salt concentration influenced the aggregation of protein. When protein was concentrated to about 10 mg/ml, it was aggregated in 100 mM KCl but not in 200 mM or 250 mM KCl. There was not big difference with and without detergent. $MgCl_2$ somehow played a role. Some precipitation could be observed when protein was dialyzed with $MgCl_2$ after Ni-affinity purification, but $MgCl_2$ effect was not clear because when we purified in parallel, for comparison, LaRP7N1_208 with and without $MgCl_2$, the chromatograms were not changed. Functional assays (gel shift assay or ITC experiments, see section II-2-5) gave similar results when done with 2 or 5 mM $MgCl_2$. Since there was no indication that $MgCl_2$ is necessary for this protein, we finally decided to avoid $MgCl_2$ in the buffer for protein purification.

Finally, to keep good condition of protein until just before crystallization, the GF column should be done at a last step and loaded with high concentration protein to eliminate aggregation. The optimized purification sequence was Ni → SP→ GF.

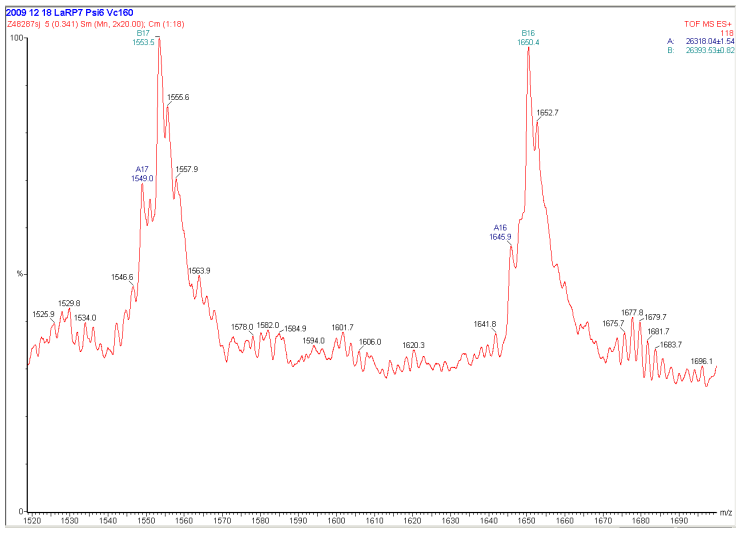
As a final check of purity, the protein was measured with Mass Spectrometry. The mass in native condition was also measured. It indicated that our protein preparation contains CHAPS and K+. The presence of numerous peaks (Fig. III-2-6) corresponding to numerous m/z

charged species, is an indication of an unfolded protein. Better results were obtained when the salt was changed from KCl to NaCl and purified without detergent. The conformation problem suggests that the protein would not be crystallized by itself.

A



B



A) = LaRP7 without Met_{ini}
B) = A + 2 with K^+

Fig. III-2-6 Mass spectrometry result of LaRP7N1_208

The LaRP7N1_208 was measured in AcNH4 250mM pH7.2 at native condition
A is zoom from 800-4400 to 1500-1700 B is zoom from 1500 to 1700 A) is LaRP7N1_208 without initial Met, B) is LaRP7 without initial Met with K+

2.2.3. Purification of the N-terminal domain of LaRP7: Tag handling.

The process described above was applied first with the N-terminal domain of LaRP7 with an N-terminal His_6-Tag. It is well-known, however, that the presence of a tag, even small as the His_6 could prevent (or favor) crystal packing. Importantly, tags can also influence the function of a protein. For example, in ArgRS, a C-terminal tag would most probably affect the recognition of tRNA, since the -COOH group is involved it the tRNA binding. Often, the N-terminal of protein has a function and sometimes the tag blocks up the interaction surfaces. In order to check the function of LaRP7, and also to test for crystallization in both configurations, we decided to cut the His-Tag.

2.2.3.1. Cleavage of tag in LaRP7N1_208 with thrombin

This was possible with thrombin thanks to the thrombin cleavage site introduced by the cloning in pnEA-Ht (Chap IV-2) at the sequence Leu Val Pro Arg ↓ Gly Ser in recombinant fusion proteins. Many popular expression vectors encode this thrombin recognition sequence. The cleavage site is usually located between the cloning sites and other vector-encoded fusion peptide sequences, which allows proteolyric removal of the fused peptide from the expressed target protein. For setting-up the conditions of thrombin cleavage, phosphate buffer was used as suggested by the fabricant. The soluble fraction of the cell lysate was divided in two parts, to try cutting directly on the beads, after binding the protein by Ni-affinity, or after elution. In this case, thrombin was added to the protein eluted with imidazole, and dialyzed together. This method allows to get rid of the imidazole during dialysis. It is a way to circumvent possible inhibition of the cleavage due to the imidazole (a problem encountered with TEV protease) and to gain time. “On beads” assays of cleavage were very disappointing; since after 12 h incubation at 4 degree in the presence of thrombin, almost no protein was released “During dialysis”, assays of cleavage were also disappointing, since two bands arose during the cleavage (Fig III-2-7).

1 2

Fig. III-2-7 Analyzed with 15% SDS-page the protein after thrombin treatment.

1. Elution of Ni-chelating column nothing was eluted from beads. 2. Ni beads almost all protein remains on beads.

To improve cleavage efficiency of His-Tag, several buffer conditions were tested, several additives tried. Each reaction was incubated at 4 degree for different incubation time (5 to 20 hours) then analyzed on SDS-PAGE. Table III-2-1, Fig. III-2-8 shows that the His-tag of LarRP7N1_208 was not cut efficiently with thrombin. An explanation of the fact that thrombin cleavage efficiency was low can be that the thrombin recognition sequence was hidden by protein conformation. To increase cutting efficiency, we tried to add detergent in the cleavage buffer. However, even in the presence of the detergent, the efficiency of thrombin cutting was not much improved. Lastly, in an attempt to separate His-tagged and cleaved protein, we tried chromatography on the SP column, but even after SP and GF column the protein with His-Tag and without His-Tag were not separated (Fig. III-2-9, Fig. III-2-10). We concluded that thrombin was not effective to eliminate His-tag of LarRP7N1_208.

Table. III-2-1 20 μl scale Thrombin cleavage test

	1	2	3	4	5	6	7	8
Protein(5mg/ml)	2	2	2	2	2	2	2	2
EDTA (50 mM)	2				2			
CaCl2(25 mM)		2				2		
Chaps (5 mM)			2				2	
Thrombin(1/10)	2	2	2	2				
Thrombin(1/100)					2	2	2	2
Buffer	14	14	14	16	14	14	14	16

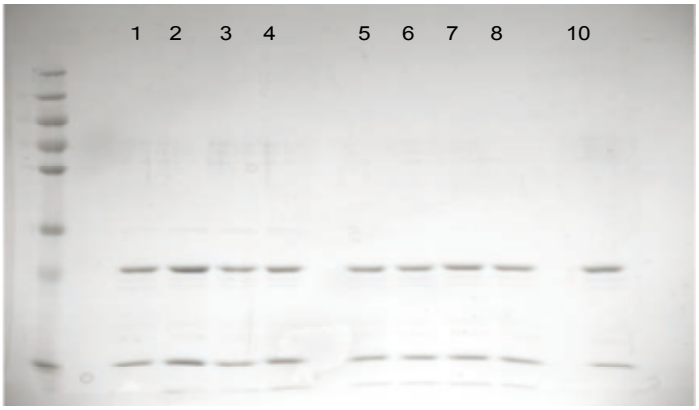


Fig. III-2-8 Thrombin cleavage test analysis of SDS-page

1 to 8 (corresponding to Table III-2-1) are analyzed with varies additions at 2 different thrombin concentrations.

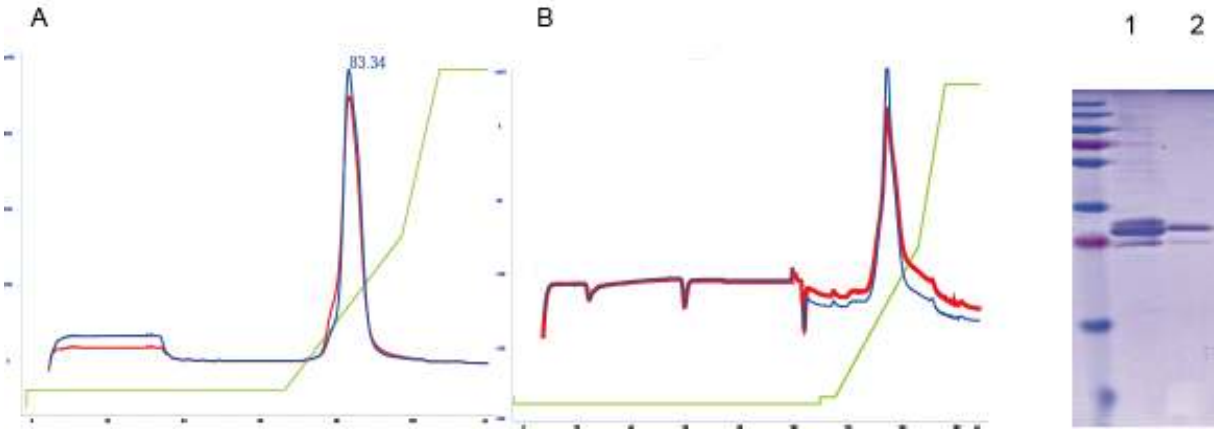


Fig. III-2-9 After SP Hitrap SP FF

A) With thrombin treatment B) Without thrombin treatment C) Analyzed with 15% SDS-PAGE 1. treatment by thrombin. 2. Without treatment.

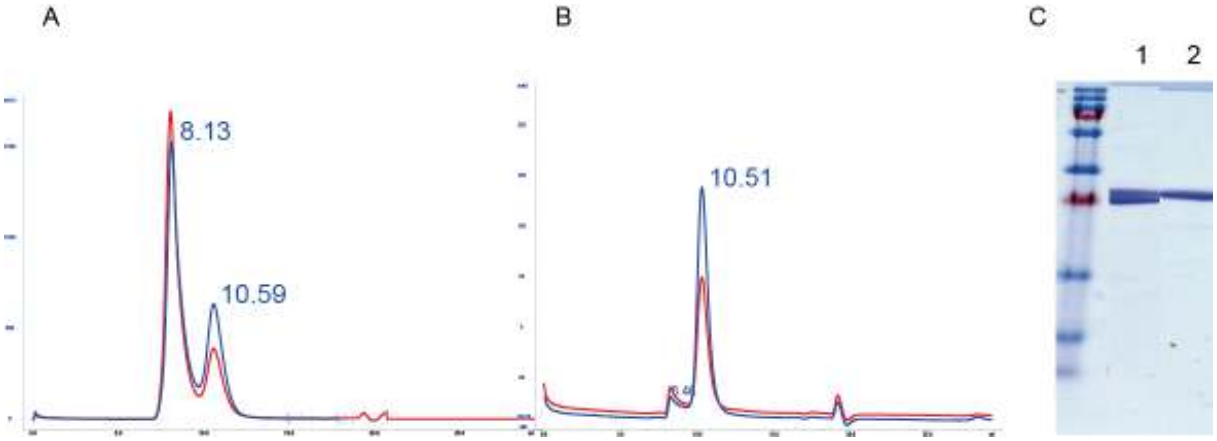


Fig. III-2-10 Gel Filtration Chromatography

A) With thrombin treatment B) Without thrombin treatment. C) Analysis with 15% SDS-PAGE of GF 1. treatment by thrombin. 2. Without treatment.

2.2.3.2. Cleavage of tag in others constructs of LaRP7N_208

To solve the problem, and with the hope to improve the accessibility of the cleavage site, the tag position was changed from N terminal to C terminal of LaRP7N1_208. At the same time, to improve the protein construct for crystallization, we truncated the sequence of the 27 residues at N-terminal end, that did not correspond to the crystallized domain of human La protein, and for which no particular secondary structure were predicted. After PCR-amplification from pnEA-LaRP7, the new constructs were sub-cloned into pnEA vectors, which provided His-tag at N-terminal or C-terminal ends, and thrombin cleavage sites.

The proteins were expressed at a satisfying level. These cell pellets were suspended lysis buffer and recombinants proteins were purified from the soluble fraction by Ni-chelating affinity column, followed by gel-filtration chromatography on Superdex 75 columns. Fig. III-2-11-C shows a superimposition of the chromatograms obtained with LaRP7N1_208, LaRP7C1_208, LaRP7C28-208. The peaks, corresponding to aggregated material eluting first show variable ratio.

LaRP7N28_208 was strongly aggregated indeed, when measured by DLS. The protein was however collected and treated with thrombin, with the idea that tag removal could improve the agregation rate. Thrombin treatment was done at 17 degree for 12 hours. At this temperature, the His_6-tags seemed removed in all proteins but several bands appeared (Fig. III-2-11-E). These bands were interpreted as degradation, favored by the thrombin treatment at 17 degree.

From the GF purification, DLS measurement and thrombin treatment at 17 degree, LaRP7C1_208 and LaRP7C28_208 proteins seemed to better. These two proteins were purified with Ni-chelating affinity column. Tag was removed with thrombin during dialysis at 4 degree. In LaRP7C1_208 and LaRP7C28_208, thrombin cleavage efficiency (Fig. III-2-12) was much better than LaRP7N1_208 but uncleaved protein remained, and several degradation bands arose. Purification was continued with SP-chromatography and Superdex 75 gel-filtration. Finally the degradation bands were not separated, and the protein was rather further degraded during GF.

The LaRP7 sequence was checked for the thrombin recognition site but it was not found, and we could not explain why several bands appeared after treatment with thrombin (such as the band at 13kDa seen in Fig. III-2-8), but for a low specificity of the commercial thrombin stock. Other colleagues had observed similar lack of specificity.

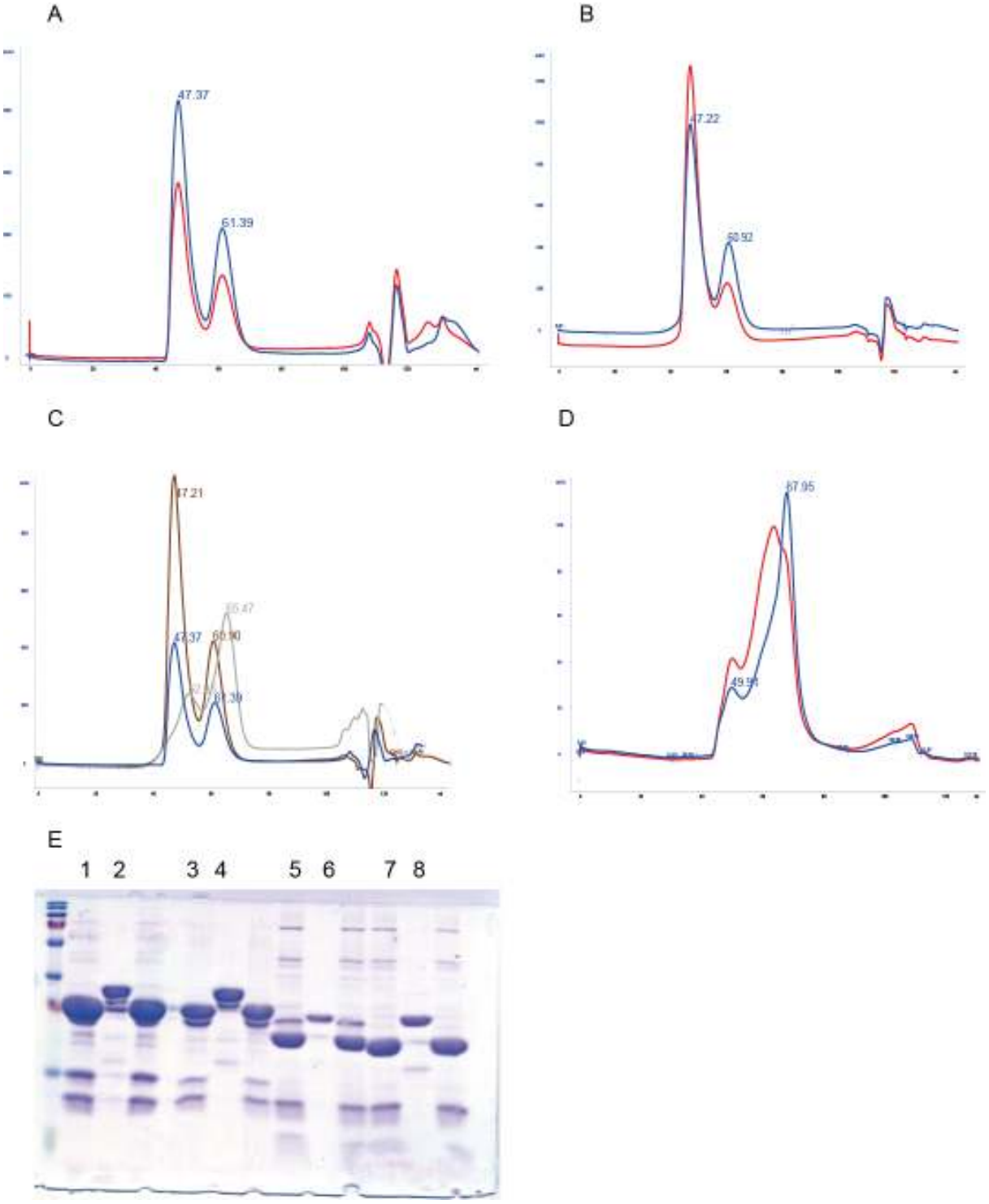


Fig. III-2-11 Chromatogram for purification by GF column

A) LaRP7N1_208 GF superdex75 16/60, B) LaRP7C1_208 GF superdex75 30/100, C) LaRP7N28-208 GF superdex75 30/100, D) LaRP7C28-208 GF superdex75 30/100, E) 15% SDS-PAGE. 2,4,6,8,are before treatment 1,3,5,7 are after treatment. 1,2 is La7N1_208 3,4 is LaRP7C1_208, 5,6 is La7N28_208, 7,8 LaRP7C28_208.

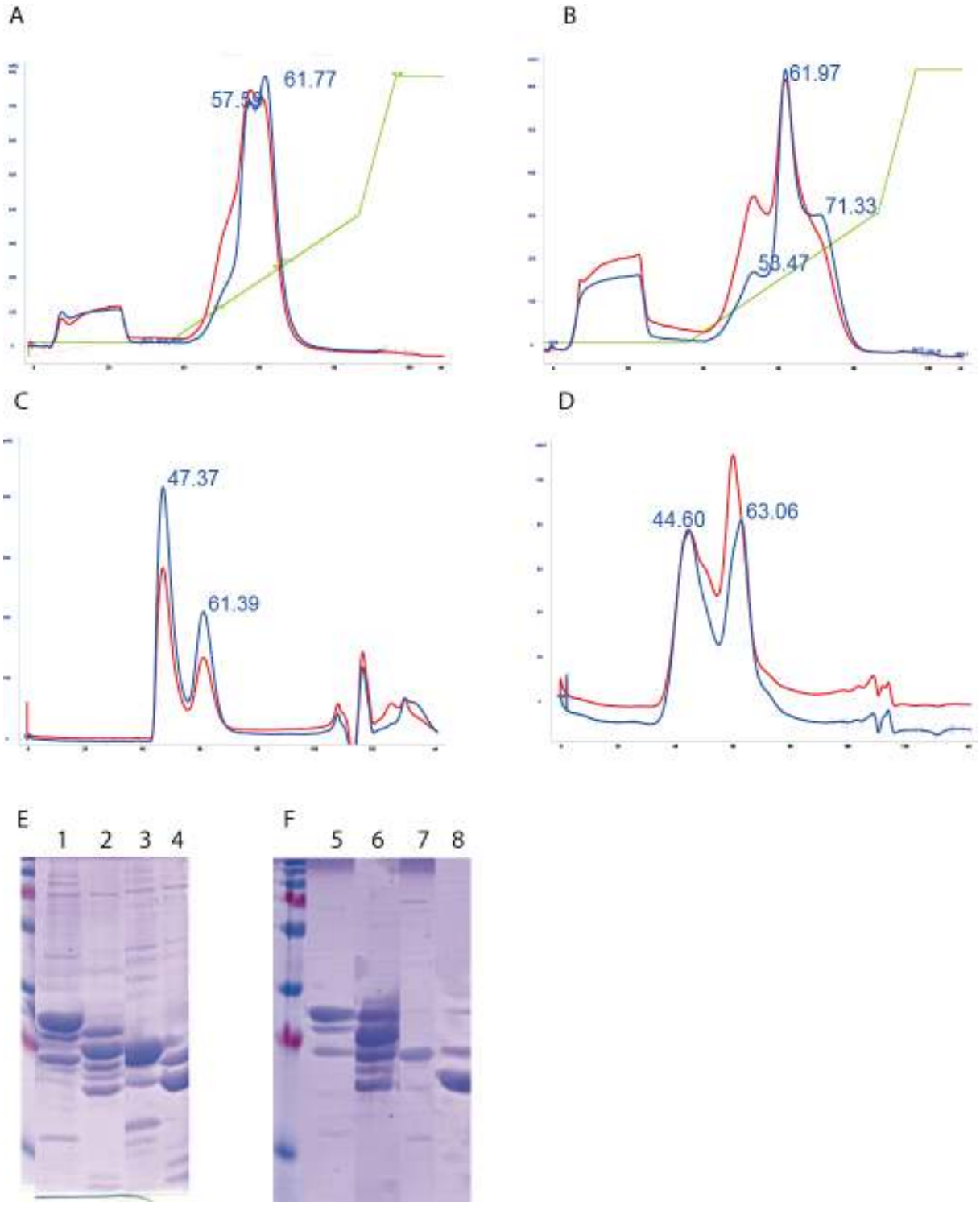


Fig. III-2-12 LaRP7C1_208 LaRP7C28_208 purification

A) LaRP7C1_208 SP column after thrombin treatment, B) LaRP7C28_208 SP column after thrombin treatment. C) LaRP7C1_208 GF column after SP column D) LaRP7C28_208 GF column after SP column) E) 15 % Acryl amide SDS-PAGE after SP column 1,2, are LaRP7C1_208 3,4 are Lp7C28_208. 1,3 are before thrombin treatment, 2,4 are after SP after thrombin treatment. F) 15 % Acryl amide SDS-PAGE after GF column are LaRP7C1_208 7,8, are LaRP7C28_208. 5,7 are before thrombin treatment 6,8 are after GF after thrombin treatment.

2.2.3.3. Cleavage with protease 3C.

Since thrombin introduced degradation of LaRP7, we decided to change the protease recognition site, for a more specific protease, the 3C (also named Precision). The sequence from 1 to 208 was sub-cloned into the vector pnEA-3H, that has the 3C protease recognition sequence, and gives an N-terminal His_6-tag (In our laboratory N-terminal His-Tag vector only available detail in Chap. IV-2). This time, the protein sequence was also chosen from 1 to 208.

The recombinant proteins were purified as usual from the soluble fraction by Ni-chelating affinity column, from which it was eluted by imidazole. A His-tagged version of the 3C protease (produced in the laboratory, see Material & Methods) was added during dialysis, overnight at 4 degree. After treatment, the protein solution was adsobed to Ni-NTA beads again for 2 h, to eliminate 3C protease and protein that still had the tag. With 3C cleavage , the N-terminal His-tag of LaRP7N1_208 was completely removed at 4 degree.

Purification was achieved with SP-chromatography and Superdex 75 gel-filtration. There were two peaks in SP chromatgram. These were collected separately and loaded on the gel-filtration column. First peak contains more aggregated material than the second peak. The peaks of GF column that correspond to non-aggregated protein were collected. A gel-analysis of the proteins originating from each SP-peak shows that they are the same. Finally the non-tagged LaRP7N1_208 protein was obtained as single band and purity > 90%. The protein was concentrated around 10 mg/ml. The protein condition was followed with DLS, which showed that at this concentration, the polydispersity was under 30%, thus this protein was used for crystallization (Fig. III-2-13). Fig. III-2-14 shows purification that gave crystals.

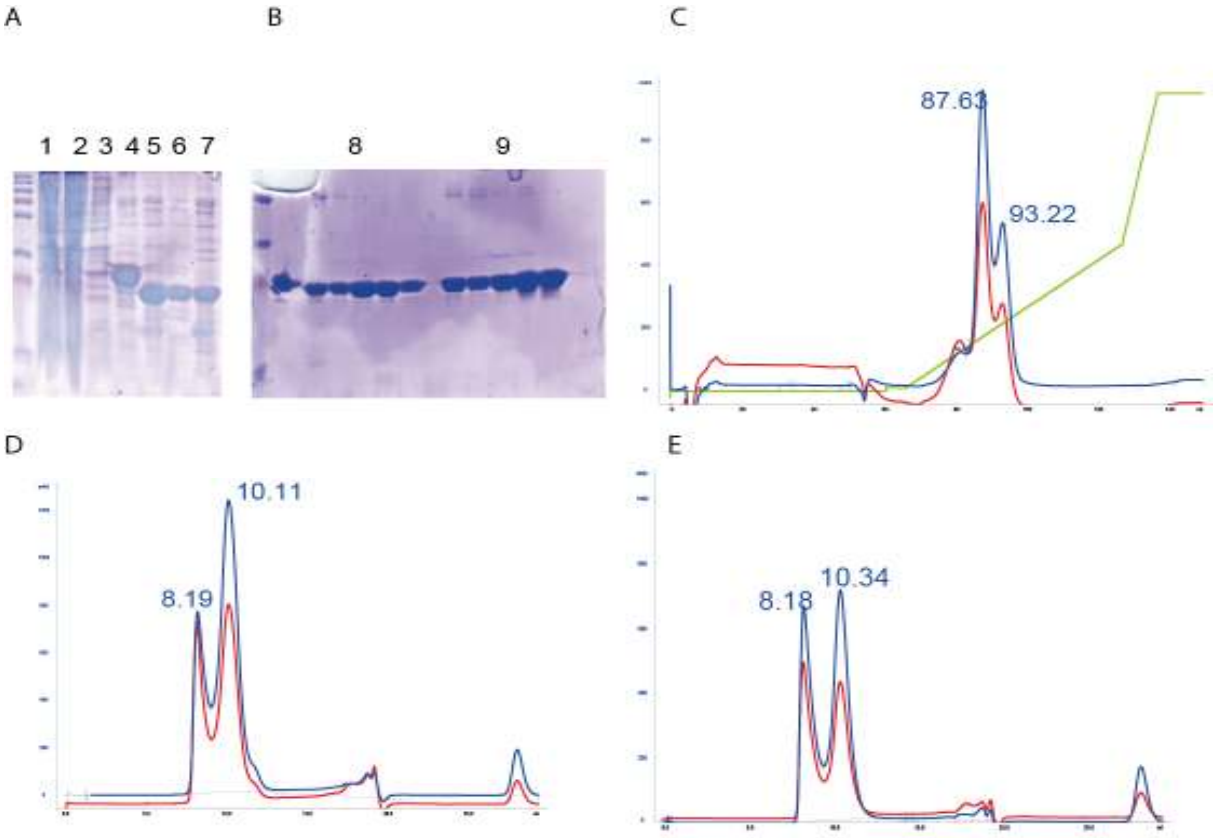


Fig. III-2-13 Chromatograms of purification of LaRP7N1_208

A)1 soluble protein 2, flow through 3, wash 4, before 3c treatment, 5, after 3c treatment , 6, elution of second Ni-chelating purification, 7, Ni beads B) 8 are fractions of D (8 are peaks of GF of first peak of SP), 9 are fractions of E (9 are peaks of GF of second peak of SP), C) Purification of SP after thrombin treatment D) GF purification of first peak of SP D) GF purification of second peak of SP

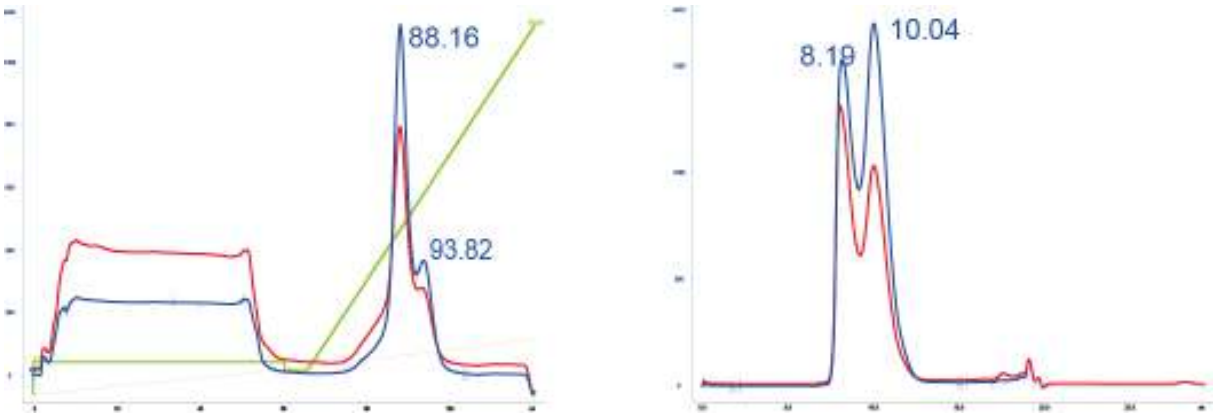


Fig. III-2-14 Protein of this purification gave crystals

A) SP column chromatogram B) GF column chromatogram

2.3. Design of an RNA substrate for crystallization of LaRP7 N-terminal domain

To investigate the binding of 7SK and LaRP7, we designed HP4U, HP4 and Oligomer(Fig. III-2-15). As I described in III-1-4, LaRP7 is expected to recognize the 3'UUUU-OH sequence of 7SK. The secondary structure of 7SK has 4 hairpins (332 nucleotides, 110kD that is Fig. III-15-A). HP4U placed in 3' end and end with the UUUU sequences. 7SK is such a big RNA difficult to obtain in a big amount. It is easily degraded therefore this RNA is not suitable for crystallization and it is adviced to design more stable RNA. Furthermore it is reported that LaRP7 recognize 3'UUUU sequence but the binding affinity was always analyzed with 7SK full length. We wished to investigate whether other parts of 7SK contribute to binding with LaRP7 as well as the oligomer contained. As a first step we designed with (HP4U) and without (HP4) the 3'-UUUU sequence and only U-rich sequence(UUUUCUUU). The sequences are described in Fg.III-2-16.

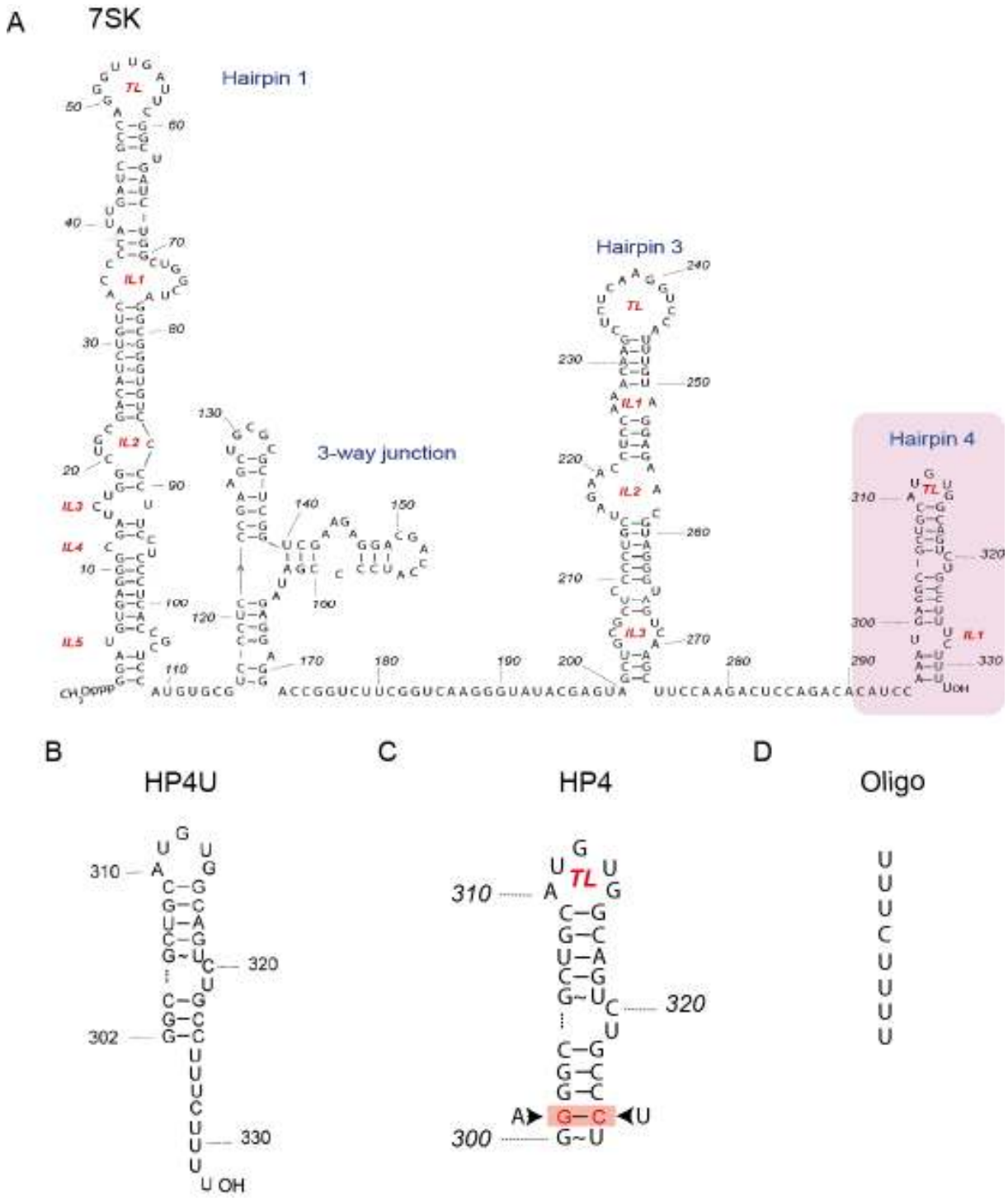


Fig. III-2-15 7SK RNA

A) 7SK RNA full length B) HP4U C) HP4 The mutation A 301 G and U 325 C mutation of 7SK were introduced to allow t7 polymerase efficient D) Oligomer 3' of 7SK 325 to 332

2.4. Preparation of pure HP4U

Milligram quantities of RNA are required for structural studies. This can be easily prepared by *in vitro* transcription from DNA template using T7 RNA polymerase.

2.4.1. Templates for HP4U transcription

For transcription of HP4U, two types of templates were used. The first (oligo template) result from the hybridization of two DNA oligomers, namely the T7 promoter, and the complementary sequence of HP4U with upstream T7 promoter. This is simple to use, and produces large quantities of transcripts. The products lack homogeneity, however, since the T7 polymerase has a tendency to add several nucleotides (non-template) at the 3'end of transcripts. The second (plasmid template), is a plasmid, pHDV_HP4U, linearized with XbaI. The plasmid encodes, downstream of the sequence of interest, for a ribozyme sequence (HDV hammerhead type). This ribozyme forms and cleaves when incubated with magnesium. The transcription buffer contains enough magnesium to induce the cleavage (16 mM), which can however be improved by adding more magnesium (to 40 mM) and performing a thermal treatment to help the ribozyme to fold properly (see Chap. V for details). The advantage of the transcription with the ribozyme is that the non-template additional residues at the 3' end are cleaved-off with the ribozyme. This leads to clean 3' ends, often advantageous for crystal contacts. This process is however much longer, since the template preparation requires two steps before the transcription, the plasmid preparation, currently done by maxi-preparations from 250 mL cultures, and linearization, by overnight incubation with XbaI. The purification of the final product can also be complicated by the presence of the HDV ribozyme. We use the plasmids engineered by Greame Conn, allowing to choose between two HDV variants, of 68 or 84 nucleotides. HP4U (32 nucleotides), was cloned in the pHDV vector, with the 68 residues ribozyme, which is the most efficient.

The choice of the restriction enzyme used for linearization was optimized. We tried the different restriction enzyme cutting downstream of the HDV ribozyme in the plasmid: XbaI, XhoI, and EcoRV, but could see no obvious differences. Fig. III-2-16 shows small-scale transcriptions for the differently linearized plasmids, together with the oligo-templated transcription. The yield are similar, but the products more pure in the plasmid-template case, as indicated by the clear bands seen in the corresponding lanes.

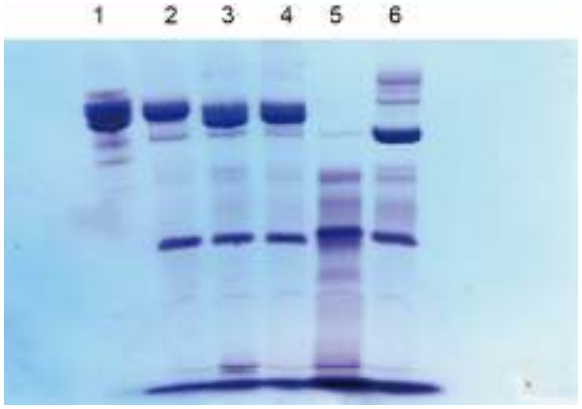


Fig. III-2-16 Transcription test of difference restriction enzyme and difference template

1. Migration reference (tRNA migration reference) 2. XbaI, 3. XhoI, 4. EcoRV, 5. oligo-template , 6 PCR template

2.4.2. HP4U purification

Transcription with T7 polymerase produces in a few hours large quantities of RNA (typically, 2.5 mg of RNA for a 5 mL transcription batch), but purification of final RNA product still remains rather time-consuming. Preparative denaturing polyacrylamide gel electrophoresis is still most commonly used to purify transcribed RNAs. Although this method provides nucleotide resolution for RNAs up to 100 nucleotides long, it is both time consuming and denatures the RNA. This can lead to misfold or aggregated RNA species after elution from the denaturing gel. First, we tried HP4U purification without using denaturing gel.

To avoid conformation problem, HP4U purification was tried with only chromatography. RNA purification was done under with and without DNase, G25 following G2000, or Mono Q (Fig. III-2-17).

Finally, HP4U was not purified satisfactorily with columns and we turned to preparative denaturing poly acryl amide gel electrophoresis. Preparative gel allows separating unreacted rNTPs, degradation and template DNA. The HP4U band was cut and eluted. To ascertain the best quality of purification required by crystallization, and in particular, to try to purify one conformation of the product, an additional step of ionic-exchange chromatography was performed on Mono Q. The RNA, eluted by a gradient of NaCl, was precipitated with EtOH and stored.

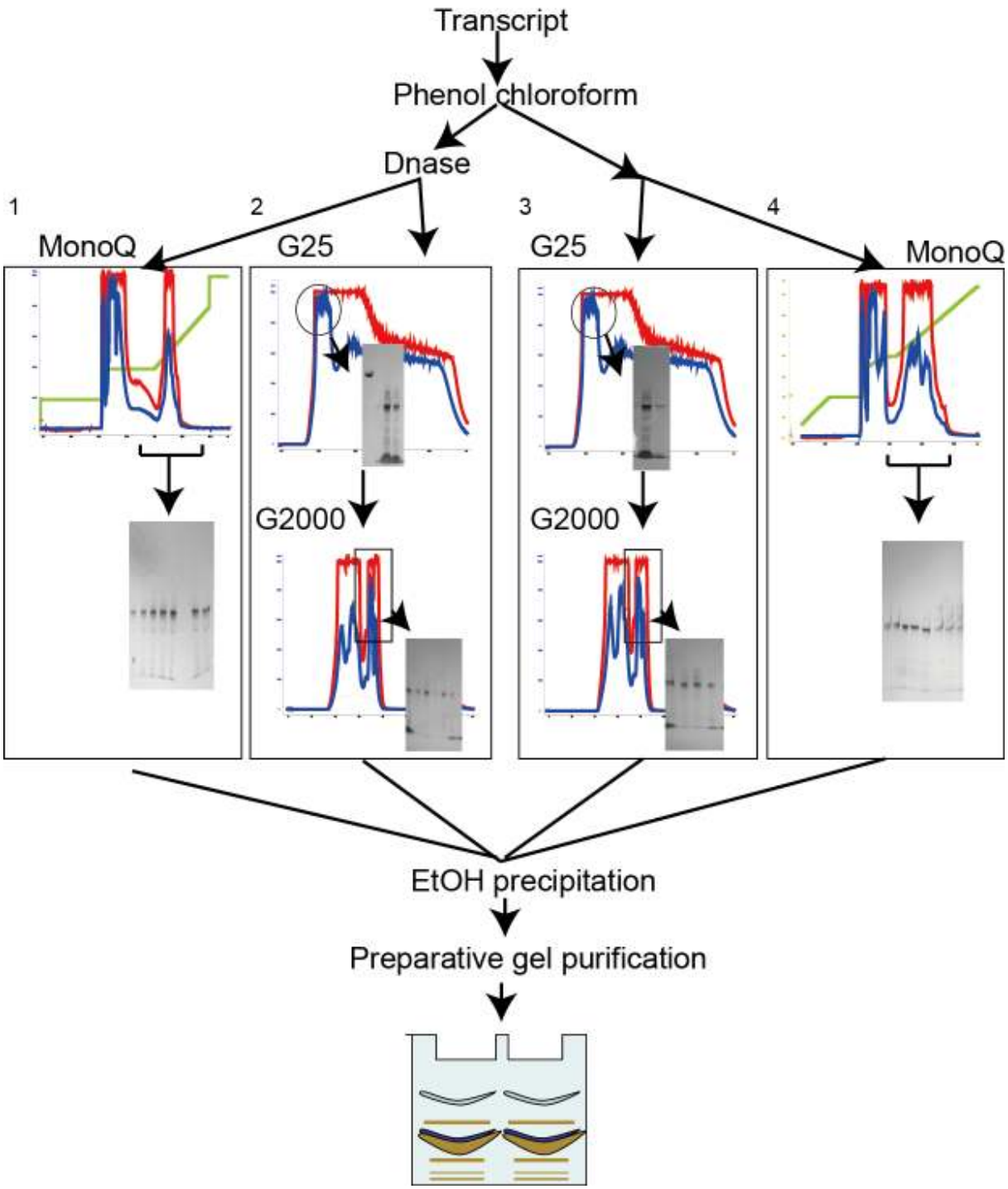


Fig. III-2-17 Trial of HP4U purification

2.4.3. Comparison of HP4U from oligo-template or plasmid-template

The HP4U transcribed from oligo-template was measured by mass spectrometry. There were indeed a lot of variations of mass, corresponding to polymerase “stammering” at the 3’ terminal UUUU (Fig.III-2-18). To make a pure product, we decided to transcribe HP4U from pHDV. pHDV vector allows for transcription with a 3' hammerhead ribozyme therefore the transcript has a unique cyclic phosphate at 3’-end.

We checked the produced HP4U from oligo template and pHDV linearized template with and without thermal treatment on native gel. The HP4U from oligo template shows several bands, including degradation (or abortive transcription) bands. Stable HP4U was then prepared from pHDV plasmid. Fig. III-2-19 shows that a thermal treatment favors the obtention of a single species (single band in panel B).

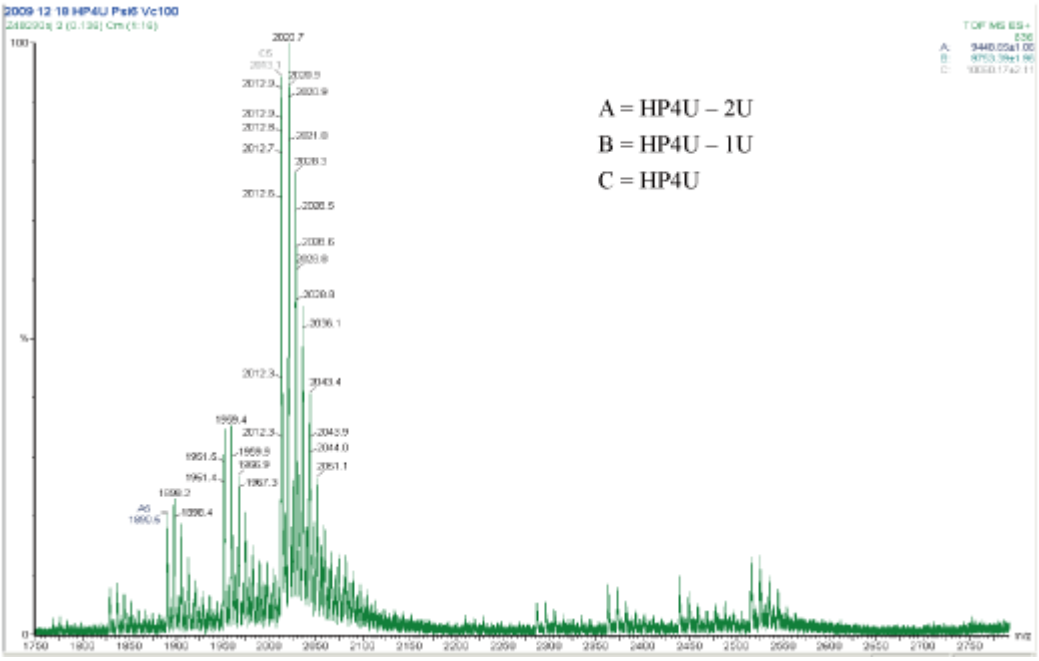


Fig. III-2-18 Mass spectrometry data of HP4U

HP4U was measured in $AcNH_4$ 250mM pH7.2 in native condition. A is lacking 2U. B is lacking 1U, C is HP4U

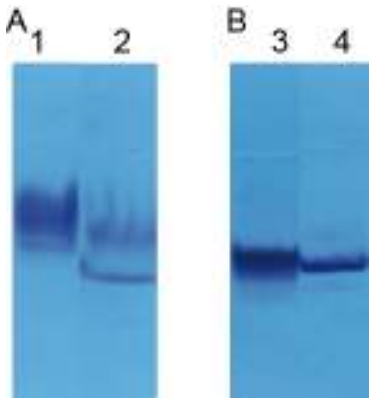


Fig. III-2-19 Acryl amide native gel of HP4U

A)Without thermal treatment 1 HP4U from oligo. 2 HP4U from pHDV.
B)Treated 85 degree 5 minutes 3 HP4U from oligo.4 HP4U from pHDV.

2.4.4. Homogeneity of the 5'-end. Transcription with GMP

The mass analysis showed also heterogeneity at the level of the 5'-end, with products corresponding to the tri-, di- and mono-phosphate species. Indeed, during transcription, with reaction buffer containing the four rNTPs ATP, UTP, CTP, GTP, the initial 5'-nucleotide is GTP. This is easily degraded. Since we want to maximize the homogeneity of the product, we introduced GMP in the reaction medium, in excess compared to the rNTPs, so as to force the polymerase to start with GMP (the GMP cannot be incorporated during the elongation phase). The chromatogram of the last, Mono Q step of purification showed that HP4U produced with or without GMP was of the same quality (Fig. III-2-20).

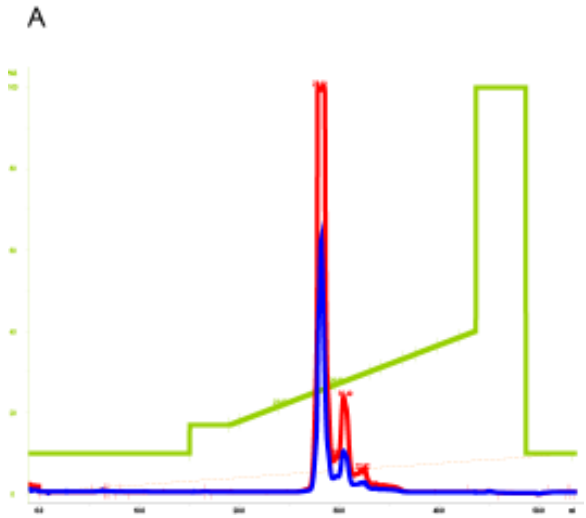


Fig. III-2-20 Chromatogram of HP4U with GMP

2.4.5. HP4 purification

For binding assays, in order to control the importance of the 3'-end sequence UUUU, we also produced the final 3'-end hairpin of 7SK, represented in Fig. III-2-15. HP4 transcripts from oligo template was of better quality than for HP4U, and didn't show degradation bands so it was not necessary to clone HP4 and transcribe from pHDV plasmid. A typical production of HP4 is shown in Fig. III-2-21.

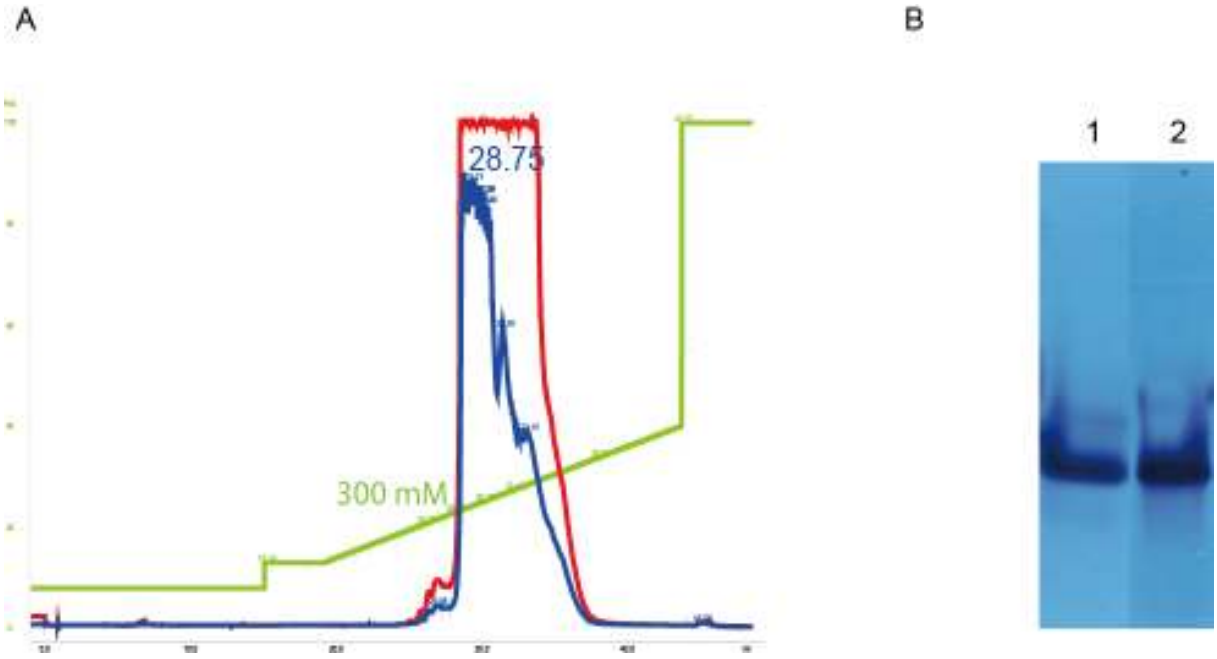


Fig. III-2-21 HP4 purification by Mono Q and gel analysis

A) Chromatogram of HP4. B) 15% acryl amide native gel analysis 1 is without thermal treatment, 2 is with thermal treatment

2.5. Binding assays

2.5.1. Purpose of the binding experiments

When LaRP7 was identified as a constant partner of 7SK(He et al. 2008) the fact that LaRP7 possesses a La domain (LAM and RRM) at its N-terminal region, prompted the hypothesis that LaRP7 binding ability to 7SK relies upon binding to the 3'-UUUU sequence. The importance of the La-domain and RNA 3'-end for binding was indeed probed in the original paper, but this interaction does not fully explain the high stability, and the specificity, of the interaction between LaRP7 and 7SK. We expect that the interaction between LaRP7 and 7SK goes beyond the binding of La domain and 3'UUUU-OH of 7SK. In order to decipher what is the basis for the specificity of recognition, and understand the function of LaRP7, we decided to focus first on the La-domain (LaRP7N1_208, comprising LAM and RRM of LaRP7) interaction, establish protocols for measurements and compare bindings with the full-length protein and the full-length RNA. The decision to tackle first the N-terminal domain was strengthened by the reported NMR structure of the parent La protein, and the fact that it could be driven to crystallize, in the presence of short RNA oligonucleotides. On the RNA side, we chose first to compare binding of 7SK with a 8-mer oligonucleotide corresponding to residues 295-332. We could not synthesize ourselves this molecule and used a custom-synthesized product (from Dharmacon), which was quite expensive, and quite unstable. We turned then to the final domain of 7SK, comprised of the fourth hairpin (HP4) and the polyU extension (Fig.III-2-15), that we named HP4U, and could produce easily by in vitro transcription. In order to assess the importance of the polyU tail for binding, we also produced a version of HP4 deprived of the polyU. A first step of the project was to find a RNA construct able to bind specifically with the La-domain, to be used in co-crystallization assays.

We did several kinds of binding assays, among those available to investigate the interaction between proteins and RNA. In this thesis, were used Electrophoretic Mobility Shift Assay (EMSA), thermal shift assay, Fluorescence Anisotropy, ITC and Gel-Permeation chromatography.

2.5.2. Electrophoretic Mobility Shift Assay: EMSA

The most common way to see the binding affinity between protein and RNA is EMSA. EMSA is a technique used to characterize protein: RNA interactions and are often

performed concurrently with foot printing and primer extension assays. EMSA is based on the observation that complexes of protein and RNA migrate through a non-denaturing poly acryl amide gel more slowly than free RNA fragments or oligonucleotides. EMSA is performed by incubating a purified protein, or a complex mixture of proteins with a ^{32}P-labelled or end-labeled RNA or fragment containing the putative protein binding site and non specific RNA competitors.

Usually in EMSA, the RNA is radioactively labeled but for technical reasons (difficulties to access a laboratory with authorization for radioactivity handling) we first tried to find a way of binding assays using non labeled RNA. We tried to reveal the RNA band in native acryl amide gels by staining with toluidine blue. Although this staining method is very convenient and sensitive for visualizing RNAs in denaturing gels, both free RNA and shifted bands were not clear in the less-resolving native gels. We tried to optimize both staining method and migration buffers. The stain solution was changed to stains-all (SIGMA). The migrations buffers TBE and TG (see Chap. IV for composition) were compared, and the concentration of RNA increased, but the shifted band intensity did not change dramatically (Fig. III-2-22).

We turned then to 5’ end labeling. HP4U, HP4 and OliU were labeled with fluorescein. RNAs that were obtained by transcription were first dephosphorylated, to get rid of the 5' phosphate group. After dephosphorylation treatment, thiophosphate is transferred from ATPγS to the 5' hydroxyl group by the T4 polynucleotide kinase (PNK). After addition of the thiol functional group, a thiol-reactive label (fluorescein maleimide) is chemically coupled to the 5' end of the nucleic acid. Chemically synthesized oligomer do not have 5'-phosphate group, oligomer was labeled by fluorescein without dephosphorylation. This method is very interesting as it is possible to measure the RNA concentration at 260 nM, without interference from the dye color, and thus monitor the concentrations of the RNA in the assay. The signal is detected easily, either on a traditional light-box or with a phosphor imager, thus permitting data quantization and storage. The method is quite sensitive, and concentrations of RNA in the range of 0.1 - 0.5 μM could be used, which is well in the expected range for correct measurement of LaRP7 *K*d.

After several disappointing trials with acryl amide gels, where bands were blurred at low concentration of acryl amide, or protein could not enter the gel satisfactorily, we tried native agarose gels, in a horizontal set-up, with much better results, even for rather small RNAs such as HP4 (see Fig. III-2-22.B). The EMSA experiments were then performed with

agarose gels (1.5 to 2%). We tested several binding buffers, varying salt concentration, pH and buffer nature. Finally the protein and RNA were incubated in 100 mM K-phosphate pH 8.0, with 500 mM NaCl, 1 mM DTT, 5 mM MgCl2, 0.1 mg/ml BSA. A large amount of tRNA (total) is added as competitor (1 mg/ml, 40 μM), to ensure that the observed shifts do not result from non-specific interactions. These are frequent in mixtures of RNA and positively charged proteins. We noted that often in biochemical or biophysical assays, phosphate buffer improves the binding condition.

Pure proteins were used (see section III-2.2) for the assays. Fig. III-2-22 shows our results when comparing the binding of the N-terminal domain of LaRP7 or the full-length proteins to HP4U, HP4 and OliU. Both LaRP7N1_208 and LaRP7 full-length bound strongly to HP4U. Interestingly, the shift of LaRP7N1_208 showed two bands. This could indicate that LaRP7N1_208 has different conformations or two oligomerization states. With HP4, the EMSA experiment showed no binding of LaRP7 full-length. This highlights the binding dependence on the polyU tail, as was expected from reported data. Surprisingly, LaRP7N1_208 slightly shifted HP4, but the shift was much decreased, compared to HP4U, and could represent a binding site accessible in the N-terminal La-domain, but hidden in the full-length protein. The functional relevance of this binding site has not been assessed yet. Interestingly, no binding shift of either protein was observed with OliU. La-domain (LAM and RRM) has been shown to bind UUUU 3'-OH from EMSA experiments but from my experiments the oligomer wasn't shifted. This seems to be contradictory with experiments with HP4U and HP4, showing that the UUUU 3'-OH was required for the affinity between LaRP7N1_208 and HP4U. When tried without tRNA competitor, a weak shift of OliU was observed, an indication for weak, perhaps non-specific binding. It is also possible that the short OliU binds the tRNAs, and is partly titrated out of the solution. Another possible explanation is that the 8-mer conformation is different in the free state and attached to the hairpin. Apart from the polyU tail, there is also a structural difference between HP4 and HP4U (see Fig. III-2-15), with a longer bottom stem in HP4, stabilized by a sequence change necessary for efficient production. This could reveal influence of the length of the single-stranded moiety. The effect of OligoU must be further examined with competition assays, where the specificity of an observed RNA binding reaction can be evaluated using conditions in which an excess of unlabeled probe is added together with the labeled probe. New molecules will be designed to assess the importance of the length, and insertion context, of the polyU tail.

A

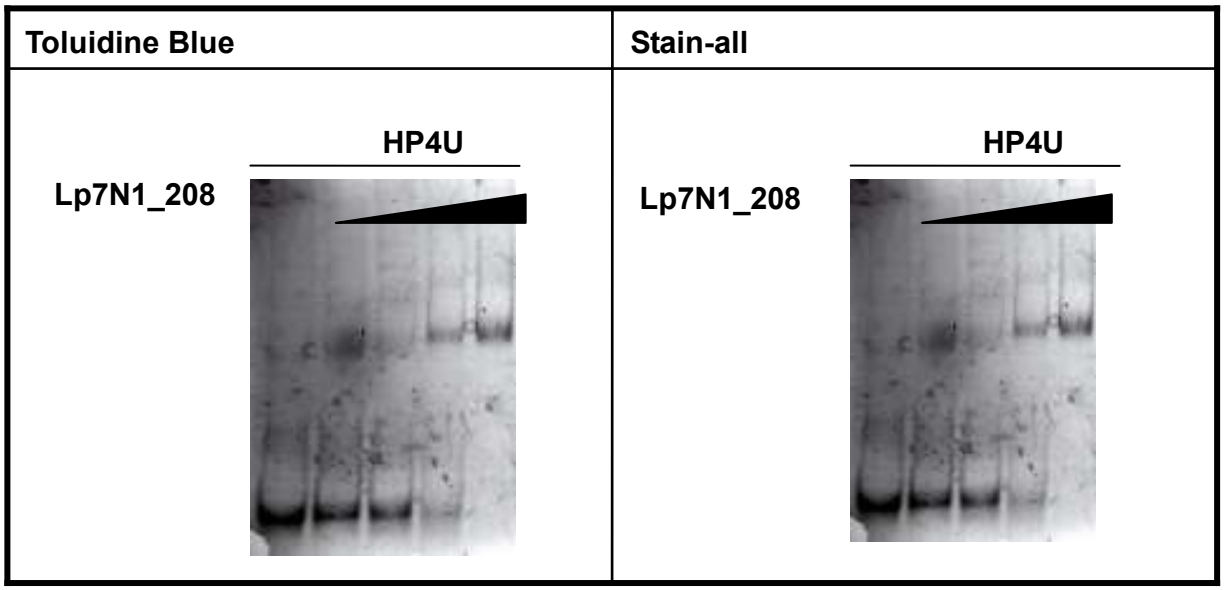


Fig. III-2-22 Comparison of different staining methods in EMSA

EMSA 6 % acryl amide gel
The 8 µl reaction buffer contained 4 µM HP4U and LaRP7N1_208 ranging from 1 µM to 8µM The protein and RNA were mixed in 20 mM Hepes-NaOH pH7.6, 200 mM NaCl, 1 mM DTT and Incubated at room temperature for 20 minutes. These mixtures were migrated on 6 % acryl amid gel at 3W 4 degree.
(a) Stained with toluidine blue. (b) Stained with Stain-all.

B

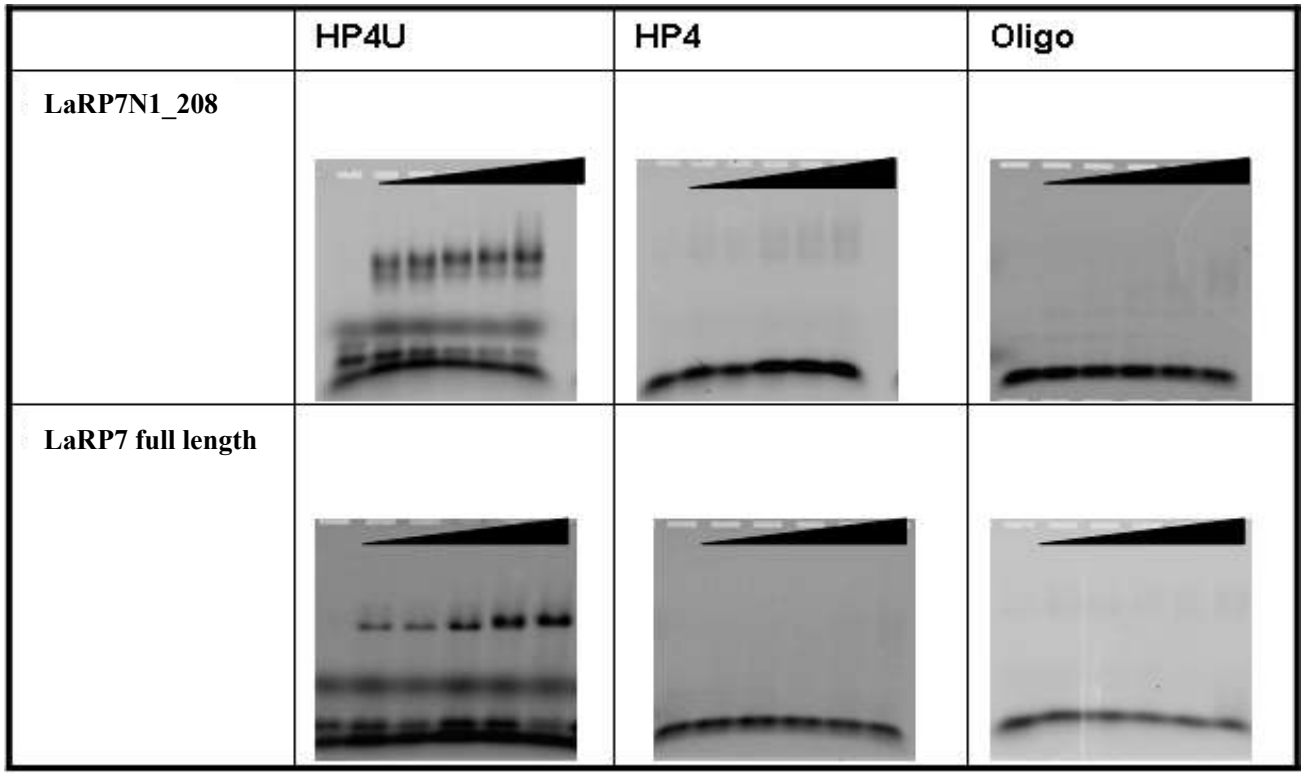


Fig. III-2-22. B

EMSA with 2.5% agarose gel

10 µl reactions containing 2µM of 5'end-labeled HP4U and LaRP7N1_208 ranging from 4 µM to 64 µM concentrations. The protein and RNA were mixed in binding buffer (100mM K-phosphate pH8.0, 500 mM NaCl, 1 mM DTT, 1mg/ml tRNAtotal, 5 mM MgCl2, 0.1 mg/ml BSA,) incubate at room temperature for 20 minutes. These mixtures were migrated on 2.5 % agarose gel with 100 V at room temperature.

(a) LaRP7N1_208 and HP4U. (b) LaRP7N1_208 and HP4. (c) LaRP7N1_208 and Oligomer. (d) LaRP7 full length and HP4U. (e) LaRP7 full length and HP4. (f) LaRP7 full length and Oligomer.

C

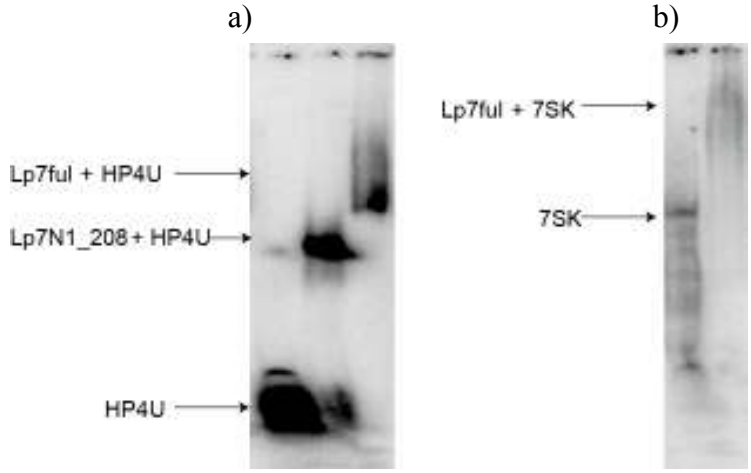


Fig III-2-22. C

EMSA with 6 % acryl amide gel

HP4U was 5' end-labeled with [γ^{-32}P]ATP using T4 polynucleotide kinase (NEB) Protein-RNA binding reactions (10 µl) were incubated for 15 min at room temperature in 50 mM HEPES, 100 mM NaCl, 2 mM MgCl2,1 mM DTT, 0.3 mg/ml tRNA(total) the samples were loaded on a 6 % native poly acryl amide gel at 4 degree in 0.5X TBE (Tris-borate-EDTA) buffer. The gels were run for 1 hr at 100V and dried onto Whatman 3MM chromatography paper. The activity associated with each band was quantified using a phosphoimager.

2.5.3. Thermal Shift Assay

Thermal shift assay is a technique that is generally used to estimate the stability of folded protein. The assay takes advantage of an environmentally sensitive fluorescence dye, such as Sypro Orange, and follows its signal changes while the protein undergoes thermal unfolding. When Sypro Orange is added to a properly folded protein solution, it is exposed in an aqueous environment and its fluorescence signal quenched. As the temperature rises, the protein undergoes thermal unfolding and exposes its hydrophobic core region. Sypro Orange then binds to the hydrophobic regions and becomes unquenched. This will result in the increase of fluorescence signal of Sypro Orange. Since this experiment is quick, and demands only minute amount of protein, it is very convenient to determine good buffer conditions for storage or crystallization of proteins. Ligand binding to a target protein can stabilize a protein's native state, as shown in the increase of the bound protein's melting temperature. The midpoint of the melting curve of a protein will increase in the presence of ligands that bind more tightly to the native state than the unfolded state.

Analyses of melting curves of LaRP7N1_208 were done with the different RNA ligands, HP4U, HP4, OliU and tRNA (to assess for non-specific effects). Fig. III-2-23. shows that the Tm of LaRP7N1_208 mixed with HP4U was 10 degree increased when compared with free protein. Moreover, the Tm with OliU was 17 degree increased. This means the LaRP7N1_208 is much stabilized by binding RNAs. This experiment can help improve crystallization. With the best ligand, chosen as it induces the higher Tm, the possibility of crystallization increase. From this experiment, OliU would be the best choice to stabilize LaRP7N1_208. An explanation for this strong effect is OliU could easily neutralize the numerous positively charged amino acids at the surface of LaRP7N1_208. The effect would be large because of its structural flexibility compared with HP4U and HP4.

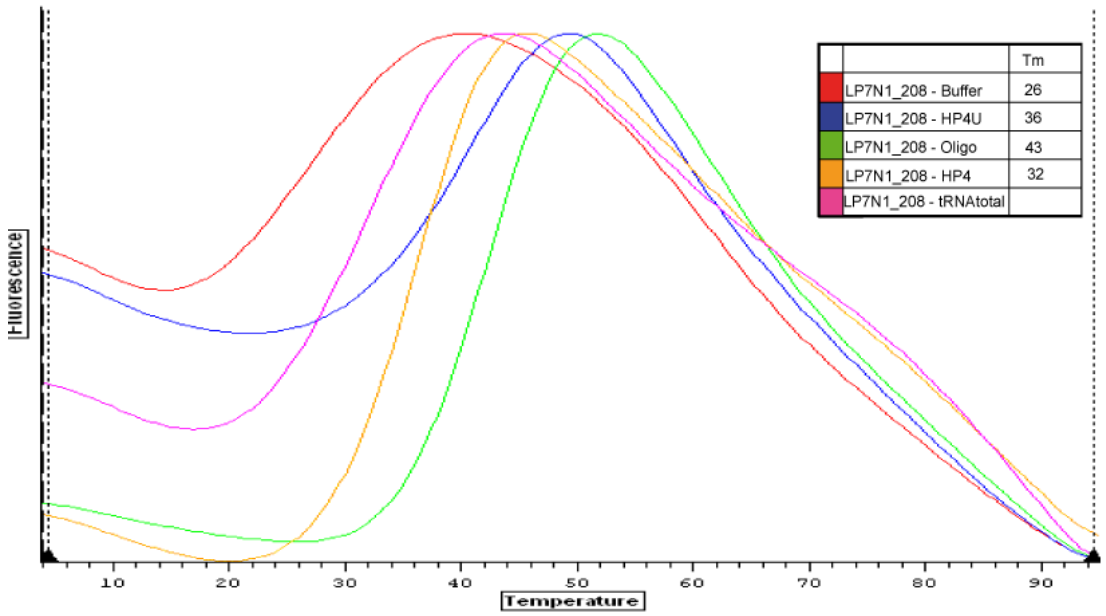


Fig. III-2-23. Thermal shift assay

Thermal shift assays were measured in 20 mM Hepes-NaOH, 100 mM NaCl, from 4degree to 90 degree.
Red : LaRP7N1_208 Blue : LaRP7N1_208 HP4U Green : LaRP7N1_208 Oligomer
Orange : LaRP7N1_208 HP4 Pink : LaRP7N1_208 tRNA

2.5.4. Isothermal titration calorimetry : ITC

Isothermal titration calorimetry (ITC) is a powerful and versatile method to study the physical basis of molecular interactions. ITC can directly measure the binding affinity (*K*a), enthalpy changes (ΔH), and binding stoichiometry (N) of the interaction between two or more molecules in solution. ITC measures the quantity of heat released or absorbed during the titration, in a closed vessel maintained at constant temperature. The measurements are extremely sensitive and reveal any exchange of heat, including buffer modifications, ion binding, or protonations. Therefore, the buffer conditions of the ligand and the protein samples have to be kept strictly the same.

Fig. III-2-24. shows the ITC curves obtained when La7N1_208 was titrated with HP4U. The measured dissociation constant (*K*d) of 183 nM corresponds to quite high affinity. We have now to compare this result with the full-length molecules, LaRP7 and 7SK, before titrating several combinations of proteins and RNA, to estimate the contributions of the different protein domains to RNA binding. A limitation of the method is that, even with the modern, miniaturized instruments, quite high concentrations, and large volumes of molecules are required for most measurements. Indeed, in our example, the concentrations necessary for a good signal with LaRP7N1_208 (25 μM) and HP4U (250 μM) are almost impossible to reach with the full-length molecules.

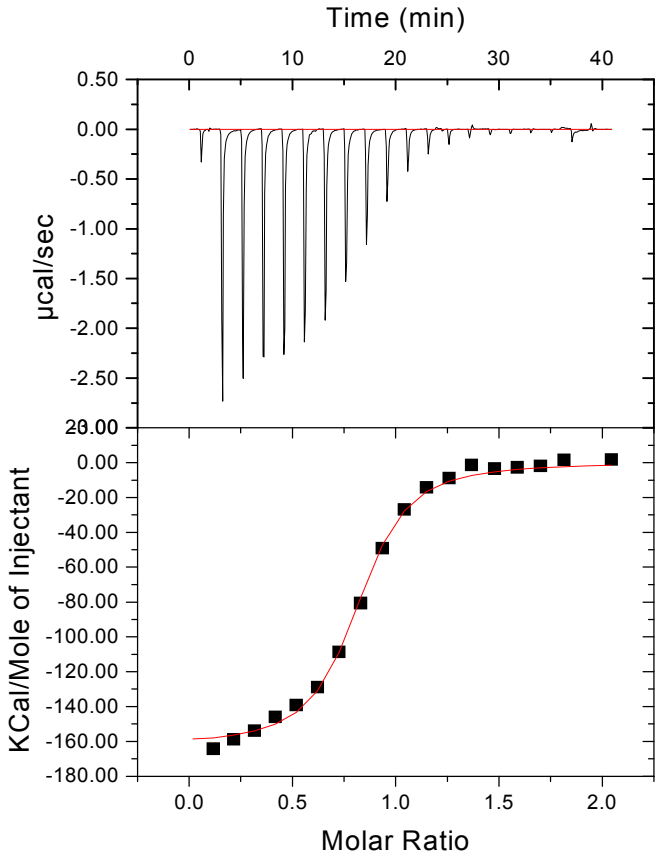


Fig. III-2-24. Isothermal titration calorimetry of LaRP7N1_208 and HP4U.

HP4U (1 mM) was titrated from the syringe into 100 µM LaRP7N1_208 placed in the measurement cell. This revealed a dissociation constant of 183 nM ($K = 5.4 \times 10^6\ M^{-1}$), a molar ratio (N) of 0.8, enthalpy $\Delta H = -1.6 \times 10^5$ cal/mol and entropy $\Delta S = 533$ cal/mol/deg. The experimental buffer condition was 20 mM Hepes-NaOH pH 7.2, 200 mM NaCl, 1 mM TECP, 0.25 mM EDTA, 2 mM $MgCl_2$

2.5.5. Fluorescence Anisotropy

Fluorescence anisotropy can be used for measuring the binding interaction between two molecules, and determine the binding constant (or the inverse, the disassociation constant) for the interaction. The basic phenomenon is that a fluorophore excited by polarized light will also emit polarized light. However, when the molecule is moving in solution, it will tend to "scramble" the polarization of the light by radiating at different directions from the incident light. For a freely moving molecule, the emission of fluorescence will be isotropic. Protein interactions can be detected when one of the interacting partners is fused to a fluorophore: upon binding of the partner molecule a larger, more stable complex is formed which will tumble more slowly. The fluorescence of the resulting complex will show anisotropy. The technique was applied to fluorescein-labeled HP4U, or OliU, that were used at 100 nM concentration. By measuring fluorescence anisotropy during titration with LaRP7N1_208 or LaRP7 full-length, the *K*d of the interactions were measured Fig. III-2-25. shows that with OliU, no plateau could be attained, with either protein. This means that OliU binds multiple sites on the protein, and reflects non-specific interactions. To circumvent that problem, we tried to measure in the presence of tRNA competitor, but could not measure reproducible data in these conditions. The experiment with HP4U was more interesting, with plateau obtained for each protein. The plateau is higher with LaRP7, reflecting the larger complex formed. Analyzing the first part of the curves (panels B and C) allows to estimate the *K*d. Obviously, the affinity for HP4 is better with the full-length protein than with the N-terminal domain. This could indicate a participation of residues outside of the La-domain, even with this small part of 7SK. It will be important in the near future, to see if the second RRM is involved. Unfortunately, 7SK being a very large molecule (110 kD), it was impossible to use this technique to measure its binding (no measurable change of anisotropy upon protein binding).

A

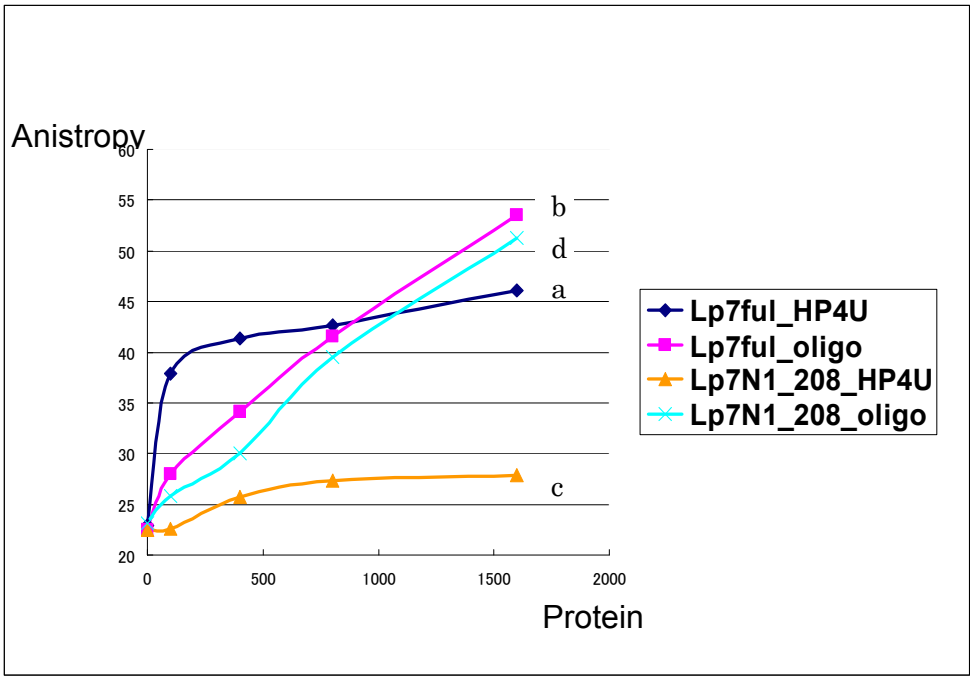


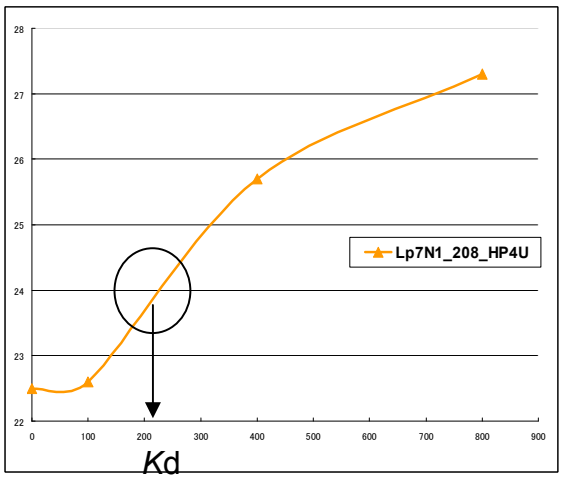
Fig. III-2-25. Fluorescence anisotropy measured with fluorescein-labeled HP4U or OliU

Protein and RNA were mixed in 50 mM K-phosphate at pH 7.6, 250 mM NaCl, 1 mM DTT and incubated 30 minutes at 4°C. Fluorescence was measured in 96-wells box, on 20 µl concentration-points, at 25°C. The RNA concentration was 100 nM

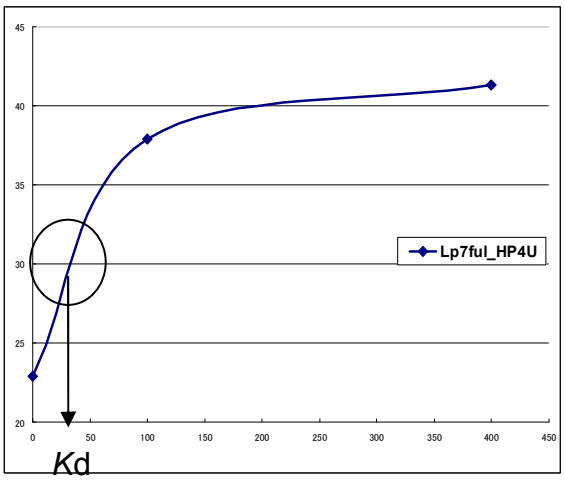
A. Complete titration with concentration of proteins ranging from 0 to 1.6 mM (a; blue) LaRP7 and HP4U (b; magenta) LaRP7 and OliU (c; orange) LaRP7N1_208 and HP4U. (d; cyan) LaRP7N1_208 and OliU

B

a)



b)



B. Close-up of the same experiment, showing the exponential parts of the curves in the titrations of HP4U.

(a) Titration with LaRP7N1_208; the circle point shows the inflexion point of the curve, from which is estimated a *K*d

(b) Titration with LaRP7; the circle point shows the inflexion point of the curve, from which is estimated a *K*d

2.5.6. Gel-filtration chromatography (molecular sieving)

Analytical gel-filtration chromatography assay is using the classical size-exclusion chromatography at an analytical level, in which the molecules in solution separate by their size. During elution, OD measurements at 260 and 280 nM allows to distinguish free protein, free RNA and complexes. Each component is also analyzed independently, in the same buffer condition and loading volume. Fig. III-2-26.

This type of experiment is very direct, and does not require sophisticated apparatus. The buffer can be adapted freely. The complex size can be deduced from a simple calibration. An advantage of this technique is that the eluted complex can be used for crystallization or EM, as it is also a powerful purification. In this experiment, the complex form of LaRP7N1_208 and HP4U was not clear. The complex peak was not clearly separated from LaRP7N1_208 protein peak as it was shifted by only 0.2 ml and the stoichiometry was not 1 : 1 rather 1 : 1/4. An explanation of this lack of clear result is that the complex could be transient. For example, aminoacyl-tRNA synthetase and tRNA complex are undetectable by this assay. This result indicated the LaRP7N1_208 and HP4U make transient complex.

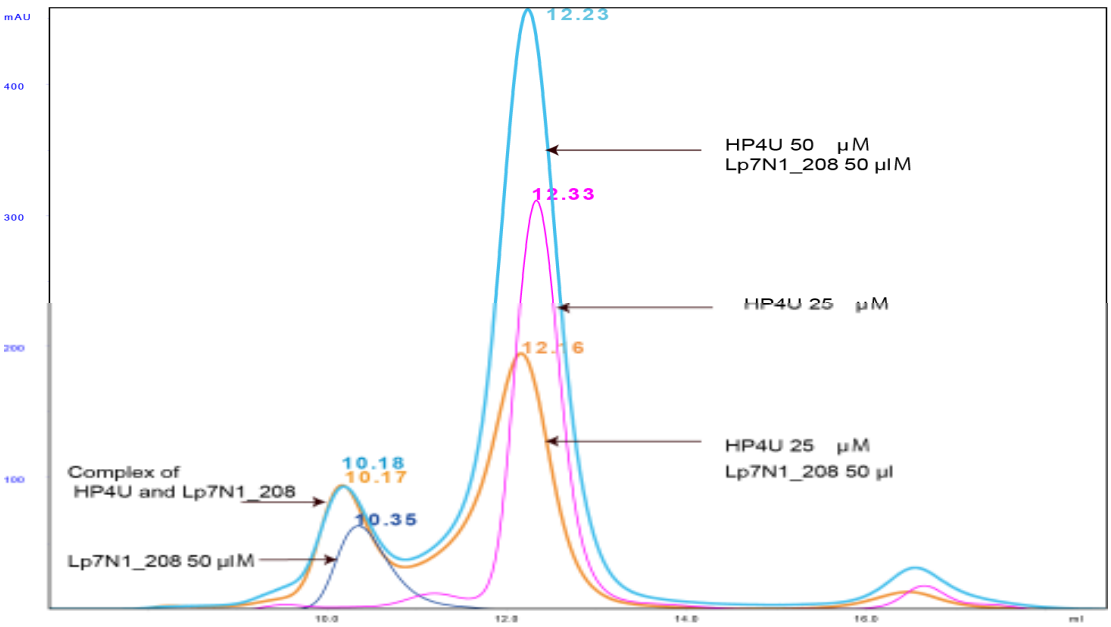


Fig. III-2-26 Analytical gel chromatography

Superposition of the chromatograms obtained with different analytical gel-filtration experiments, performed with a Superdex S75 (10X 300, 25 mL). Samples (100 µL) were injected in buffer containing Na-Hepes (pH 7.2, 20 mM) and 100 mM NaCl. Complexes were incubated 20 min at 4°C before injection.
Pink : HP4U 25 µl; Blue : LaRP7N1_208 50 µl; Orange : LaRP7N1_208 50 µl and HP4U 50 µl; Cyan: LaRP7N1_208 50 µl and HP4U 25 µl

2.5.7. Discussion of Binding assays

From the binding assay experiments, some LaRP7 and 7SK binding characteristic were explained. Interestingly, the thermal shift assay showed that the protein is strongly stabilized by RNA ligands, and especially HP4U. However, the stabilization was not strong enough to allow the use of gel filtration as a binding assay. As I described in III-2-5-6 the gel chromatography assay was unable to detect complexes, indicating that the binding is transient or not strong enough. However, the binding affinity could be analyzed from EMSA, ITC and fluorescence anisotropy. From EMSA, the HP4U bound both LaRP7 full-length and LaRP7N1_208 with a sharper binding shift for LaRP7 full-length compared with La7N1_208, which showed two bands. The ITC and Anisotropy experiments gave the information of *K*d and indicated a stoichiometry of 1:1 for LaRP7 and HP4U. The *K*d was 183 nM from ITC, in accordance with the value of around 200 nM indicated by the Anisotropy experiment. The *K*d of LaRP7 full-length for HP4U was found to be around 50 nM, 4 times better than LaRP7N1_208. The anisotropy data of both proteins for the oligomer showed unusual curves that had no plateau, while the EMSA showed no shifted band. LaRP7 protein is known to bind specifically 3'-UUUU but our results suggest that LaRP7 need the polyU to be bound to HP4. Is the hairpin structure, or individual nucleotides of that sequence, important is still to be investigated.

These binding assays showed the possibility of crystallization of LaRP7N1_208 and HP4U complex, but indicated the importance of regions of LaRP7 other than N1_208. Our working model is now that the La-domain and polyU sequence are not responsible for the specific interaction, but other parts of both RNA and protein are required for specificity. Our next task will be to further delineate the sequences and or structures involved.

2.6. Crystallization

Crystallography is the most powerful method to study the 3D of macromolecules and their complexes at atomic level, which is essential to understand their function. This method is however dependent on the obtention of good-quality crystals. Crystallization of protein is still based on trials and error. The proteins derived from bacteria are comparatively easy to crystallize, when compared with proteins from mammalian origin. In these days, structural research of proteins and RNAs is focused on molecules from higher living cells, which is the case of our targets, LaRP7 and 7SK.

What is important for crystallization is preparing the samples in good condition. This means that proteins and RNAs have to be designed as properly folded and soluble forms. Crystallization of complex of protein and RNA has to take into account conditions that allow for interaction between molecules. A summery of the process of crystallization of complex are shown in Fig. III-2-27.

2.6.1. Factors influencing crystallization

2.6.1.1 Concentration

The concentration of protein is a main factor for crystallization. If the concentration is too dilute, crystal will never arise. On another hand, when the protein is too concentrated, a lot of nucleus or precipitant appear. In a first attempt, it is however advisable to try for the highest possible concentration, in order to maximize the chance of observing a hit.

2.6.1.2 Purity of sample

Purity is also a main factor for crystallization. A purity superior to about 90% is desirable. The aim of purification is not only to eliminate the contaminants proteins but also impurities, which come from the buffers or cellular extract. The state of oligomerization of the target protein should be unified, and aggregation avoided.

2.6.1.3. pH

This factor affects the charge of protein and changes the solubility. The solubility is lowest at the isoelectric point.

2.6.1.4 Temperature

Protein fusion is endothermic reaction therefore to increase temperature the solubility is also increase. However, that depends on surface charge if positive charge the solubility is increased, negative charge the solubility decrease. Usually 20 degree and 4 degree are tried. Temperature influence protein formation solubility but also the physical process of crystallization. Indeed the packing process, dynamics upon molecular interaction can be modified by temperature hydrophobic interaction are maximized at 37 degree. Interaction can be modified by temperature

2.6.1.5. Precipitant

There are several precipitants, including salt, organic solvent, polymers. These precipitants make hydrophobic condition and induce crystallization or give exclusive volume effects and induce assembly of protein.

2.6.2. LaRP7 crystallization

From the La structures, the 3'-UUUU sequence bind between LAM and RRM but HP4U is longer than the RNA that observed crystallized RNA-La protein structures. The LaRP7N1_208 was designed for crystallization as I described before. It is important to follow the protein condition using the technique like DLS until just before crystallization. The sample was prepared the manner that able to duplicate. Until obtain the crystals of LaRP7, many ways for sample preparation were tried (Fig. III-2-28)

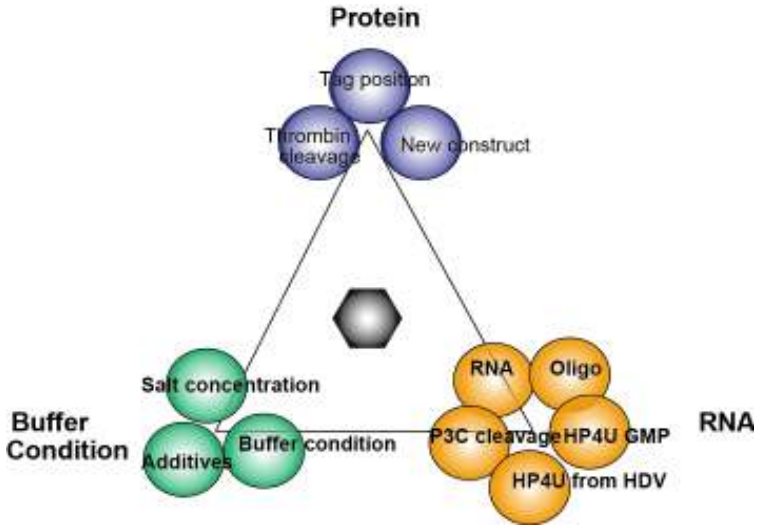


Fig. III-2-27 The factors for crystallization

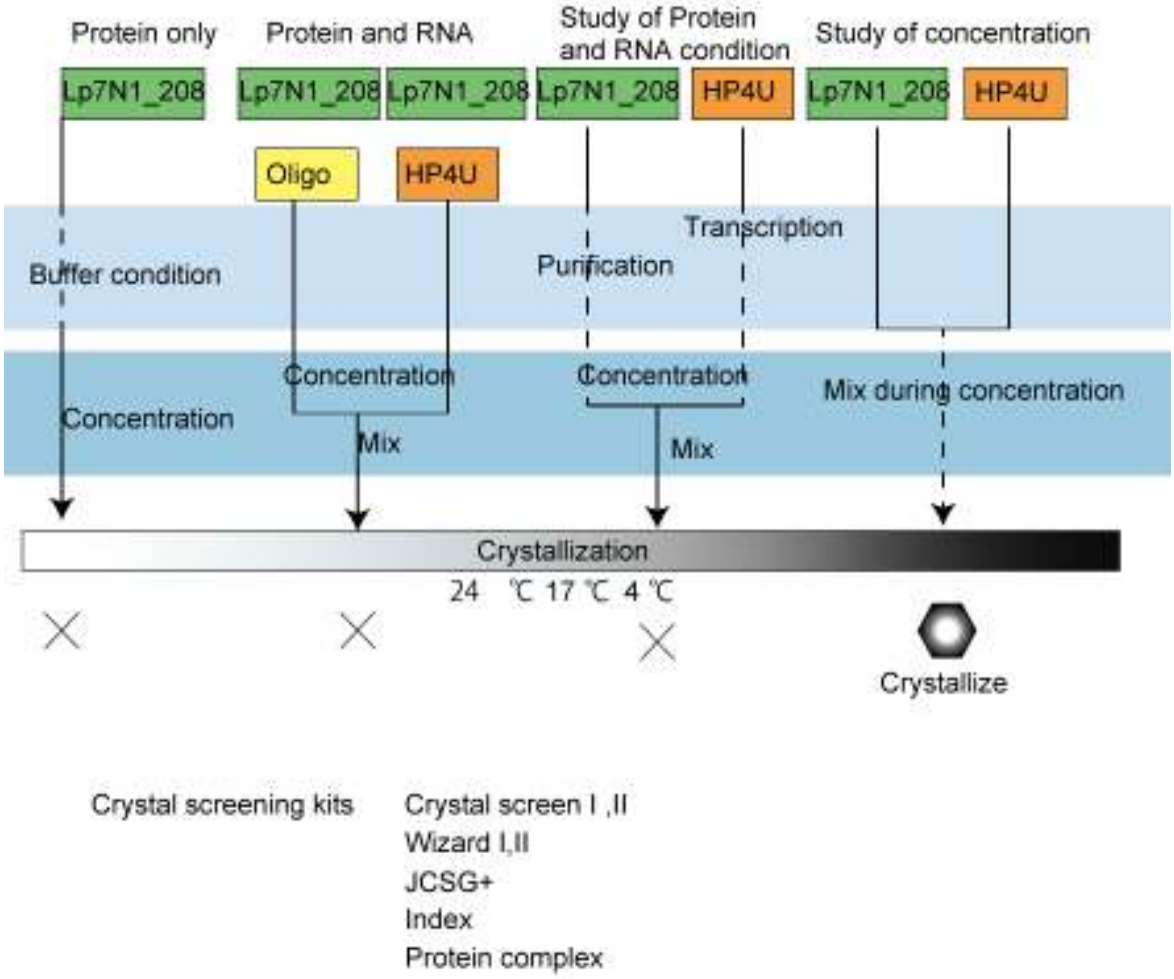


Fig. III-2-28 Sample preparation trials for Crystallization

Sky blue part is trials for purification, buffer condition. Blue part is trials for concentration. **1**. The crystallization "Protein only" was considered buffer condition for crystallization. **2**. Protein and RNA were prepared separately and mixed just before crystallization. **3**. Protein was considered its purification buffer and RNA was considered its transcription. **4**. The concentration way was considered the protein and RNA were mixed during concentration. Only No4 gave crystals.

2.6.2.1. Crystallization of LaRP7N1_208

First we tried to crystallize only LaRP7N1_208. The purification ways were considered. Column order, buffer choice, salt concentration, additives.

A SAXS experiments showed that LaRP7N1_208 didn't have conformational structure. From Mass Spectrometry experiments, K+ was adsorbed to Larp7N1_208 therefore salt in solution was changed KCl to NaCl. Also from Thermal shift assay, the protein became stable when the protein with ligand as oligo and HP4U. Therefore, crystallization was done with ligand.

2.6.2.2. Crystallization of LaRP7N1_208 with RNA oligomer -UUUCUUUU-

Larp7N1_208 was mixed with RNA oligomer at 4 degree at a molecular ratio of 1:1. We observed precipitation in almost all drops. Thin crystals were observed.

2.6.2.3. Crystallization of Larp7N1_208 with HP4U

HP4U transcribed form DNA oligomer and HP4U transcribed from pHDV behaved differently as I described before. HP4U transcribed from pHDV seemed suitable for crystallization. When LaRP7N1_208 and HP4U were mixed after concentration, precipitation occurred. To avoid this precipitation, LaRP7N1_208 and HP4U were mixed during the concentration. Crystals were not obtained.

2.6.2.4. Crystallization of Larp7N1_208 (without tag) with HP4U transcribed with GMP (HP4U(GMP))

After several trials of crystallization, we thought that His-Tag is affecting crystal packing, His-Tag was removed from Larp7N1_208 using protease. At the same time GMP was added to HP4U transcription for preventing heterogeneity of the 3' end. LaRP7N1_208 without Tag and HP4U (GMP) mixed at a molecular ratio of 1:1, then concentrated. This time crystals appeared.

The crystals grew in two weeks in four conditions (Table. II-2-2) after 4 weeks the crystals in the condition B, C, D were melted. Other crystals grew in two conditions E, F.

Condition A, B, E, F was exposed by X-ray. E and F were really thin, the diffraction was not obtained but crystal from condition A, B diffracted to 7 Å.

Table. III-2-2 Buffer condition, size and shape of crystals

	Buffer condition	size	crystals
A	0.1 M Succinic acid pH7.0,15 % PEG 3350	50×100×200 μm	Single, poly
B	0.1M Hepes-NaOH pH 7.5, 15% PEG6000, 0.1M KCl	50×100×200	Single
C	0.1 M Hepes-NaOH pH7.0,15% PEG4000	----------	Poly
D	0.1 M Hepes-NaOH pH7.0,15% PEG4000,0.1M MgCl2	----------	Thin plate
E	0.1 M Hepes-NaOH pH7.5, 12% PEG8000, 0.2M NaCl	100×50×100	Needle
F	0.1 M Tris-HCl pH8.0, 15% PEG 2000 MME, 0.1 M KCl	100×50×50	Plate

To reproduce these crystals, we investigated several variables the buffer pH [Hepes-NaOH, Tris-HCl, Succinic acid 7.0 and 7.5], salt concentration [KCl or NaCl from 0 mM to 200 mM] , precipitant concentration [PEG 4000, 6000, 8000, 2000 MME, from 12% to 20%]. Two kinds of seeding were also tried but crystals not obtained.

Reproduction of crystals was difficult we think that the main reason is variation of the ratio of LaRP7N1_208 and HP4U, and concentration of LaRP7N1_208 and HP4U. Since the LaRP7N1_208 and HP4U were mixed during concentration, it is difficult to duplicate the exactly same concentration that gave crystals. To solve this difficulties, what we can do is to vary the ratio of LaRP7N1_208 and HP4U, measure DLS the sample crystallization (So far for DLS measurement, at least 20 μl sample was needed but new system allow to measure only 5 μl) before crystallization, add some additives to reservoir.

3. Discussion

N-terminal region (LAM and RRM) of LaRP7 is similar to N-terminal region (LAM and RRM) of La protein. Sequence analysis suggested that this region could fold, in a 3D structure similar to La-protein. However, our results suggested that the protein is only folded when RNA was added.

The mass spectrometry experiment showed that the protein seems unfolded. From thermal shift assays, when RNA (HP4U or oligoU) was added to LaRP7N1_208, the Tm was dramatically increased. This is indicating that the stability much increased. SAXS experiments also supported this fact. SAXS experiment revealed that LaRP7N1_208 was not folded by itself. However, LaRP7N1_208 showed a more compact structure upon oligo addition (data is not shown). This explains that so far crystals of La protein were only obtained for protein-RNA complexes (the La-domain structure was solved by NMR) The 3D structure of La protein showed that only the LAM domain of La protein interacts with the RNA molecules. Surprisingly RRM was not participating in the interaction between La protein and RNA. However, it is also reported that the RNA does not bind to isolated LAM domain. This suggests that a competent conformation of LAM could result from RNA binding (induced fit).

The previous studies of La protein showed that La protein is able to distinguish RNAs ending in UUU_{OH} from those ending in UUUp, with a strong requirement for 3'OH. However, our result showed that the situation is different for LaRP7 binding. The affinity of LaRP7N1_208 and HP4U (3' cyclic phosphate) was measured with anisotropy experiment. The *K*d was estimated 200 nM. On the other hand the affinity of LaRP7N1_208 and HP4U (3' OH) was measured with ITC experiment. The *K*d was 183 nM. This indicated that La domain (LaRP7N1_208) of LaRP7 has almost same affinity for HP4U (3' cyclic phosphate) and HP4U (3' OH). However, this result should be confirmed. To know the precise contribution of the 3'OH, the binding of HP4U (cyclic phosphate) to LaRP7N1_208 has now to be measured by ITC.

Our experiments indicated that beside the 3'end, other factors were required for specific binding. Indeed, LaRP7N1_208 bound HP4U more strongly than to Oligo(UUUCUUUU), as shown by the fluorescence anisotropy experiment. This indicates that other determinants exist in the hairpin region. My result has not shown whether specific binding is determined by sequence or structure. When solved, the 3D structure of the complex between LaRP7N1_208 and HP4U will certainly give the necessary information to understand this question. However, a mutation analysis could also give precious information.

A limitation of my studies is that binding of LaRP7 was investigated only with HP4U. The full-length 7SK molecule was difficult to label with fluorescein, preventing the use of EMSA on agarose gel. The fluorescence anisotropy technique could not be used due to the large molecular weight of 7SK. The requirement for large quantities of material prevented the use

of ITC. We have now to turn to radio labeling of 7SK. The binding assessment with other part of 7SK (HP1, HP2, HP3) will then be studied by competitive assays using EMSA.

Finally, our experiments showed that LaRP7 binds HP4U with a better binding affinity than LaRP7N1_208. This suggests that C-terminal of the LaRP7 RRM contribute to the stabilization of HP4U. Beside the large unfolded region in the middle part of LaRP7, which contains some positively charged stretches of residues that need be investigated, an interesting candidate for this RNA-binding sequence is the second RRM in the C terminal region. This domain could participate in HP4U stabilization, and has to be investigated. Comparison of the binding of constructs of LaRP7 mutated in the positive stretches or deleted of the second RRM should give information. The role of this second RRM can additionally be studied directly, if it induces a super-shift in EMSA experiments, for example. Interestingly, this C-terminal RRM, if binding to HP4U, would induce a looping of the protein on one domain of 7SK.

These numerous experiments of bindings were measured by using several techniques. I was able to learn the limitation, advantage and disadvantage. For the measurement of *K*d, ITC give accurate value and the number of binding site. The advantage is the samples were not labeled. However, concentrated material is required and demanding quantities (1 mg to 2 mg) are required for one experiment. For the full-length 7SK and LaRP7 case, it is difficult to prepare samples as such concentrations. Oligo is also difficult to measure because of its cost. In anisotropy experiments, the material needed was only about 1 µg. But the RNA has to be labeled with fluorescein. In some cases, the data at low concentration were difficult to measure (no linearity). EMSA is a simple technique and does not require much (about 1 µg). It indicates quickly if the binding occurs, and allows competition assays, but it is also needed to label RNA and *K*d is not precise. Thermal shift assay provides interesting information on the stability of protein. This technique provides the selection of most stable ligand. Analytical gel chromatography was not suited for protein RNA binding assessment in our case but if the complex could be isolated by this technique, it strongly helps crystallization of protein-RNA complex.

In order to investigate the full-length LaRP7 binding to 7SK, we shall use cryo-EM. Compared with crystallography or SAXS, this technique is not demanding large quantities of material. However, the complexes must be of very good quality. We are now trying to prepare the molecules in this purpose, with the hope that understanding the structure of the LaRP7:7SK complex will help understand the role of LaRP7 in the stabilization of 7SK. This study is done in parallel with investigations of the 7SK:HEXIM interaction (by Denise Martinez-Zapien) and we plan to see whether the HEXIM function is physically, and/or functionally linked to LaRP7.

Chapter IV Experimental procedures

1. ArgRS, $tRNA^{Arg}$

1.1. ArgRS

1.1.1. Cloning ΔN ArgRS in pET28c

The sequence from 92 amino acid of ArgRS (Table IV-1-1) was amplified by PCR using ArgRS full length plasmid with primer contain restriction site Nde I at 5' terminal and Xho I at 3' terminal (Table IV -1-2, IV -1-3)and inserted into TOPO blunt end vector by ligase at 16 degree 15 hours (Table IV -1-4) then treated with restriction enzyme Nde I and BamHI at 37 degree 15 h (Table IV -1-5, IV -1-6) then migrated with 1 % agarose gel then cut the ΔN ArgRS band and purified with Extract II then inserted into pET28c which was also treated with restriction enzyme by ligase.

Table IV-1-1 The protein sequence of P.h ArgRS. Blue part is 1-91 aa from 5' end

Pyrococcus horikoshii ArgRS fullength sequence
5'- MLMEIRESVKERIEEIIKEIAPQWEGEIELKETPDPKLGDFGTPIAFKLAKLLKRPPIEI AEKIVEKLKLNLPEGIKDVKAVNGYINVFIDYPHFARILINDILAKGDRFGSSEIGKGKK VIVEHTSVNPTKPLHMGHARNAILGDVMARILRFLGYEVEVQNYIDDLGIQFAQVYWGYL RLKEEFERIMNELRERGLKDNPIDHALGLLYVEVNRRLEDNPELENEIRDIMKKLESGEL YGRKLAEEVVRAQMVTTYKLGVKYDLLVWESDIVRRKLFEIALELLSKNENFYIPSDGKY RGAFVMDLRKLFPDMKNPILVLRRSDGTATYTGKDIAYHLWKFGKIDVDLLYKEWDSTTW TTAPDGKSMPNKFGNANIVINVIGAEQKHPQLAIKYALQLLGFEDAAANLYHLAYEHVER PEGKFSGRKGTWVGFTVDEVIQEAVKRARELIEEKNPALSDEEKAEVAEKVGIGAIRYNL IKYSPDKKIIFRWEDVLNFEGESAPYIQYAHARCSSILRKAEEEGIKVDPETLFKNADFT KLSERERELVIMLSKFPRIVEQAGKDVKPHLIAWFANELASLFNKFYMDHPVLKAEEGVR EARLLLVMAVEQVLKNALYLMGIEAPERM - 3'

Table IV-1-2 primer for making mutant

Primer name	sequence
Arg274	GGAATTC CATATG TACCCCCACTTCGCAAGGAT

Table IV-1-3 Amplification of the sequence of ΔN ArgRS

Ph⊿N ArgRS PCR reaction	
10×pfu buffer	1×
dNPTs (2.5 mM each)	0.2 mM
primer Arg274 (20 μM)	0.2 μM
T7terminator (20 μM)	0.2 μM
PhWild-Type ArgRS/pET28c	10 ng
Pfu turbo	
MillQ	Up to 50 μl

	Temp	Time	Cycle
1	92 degree	2 min	1 cycle
2	92 degree	30 s	35 cycle
	55 degree	30 s	
	72 degree	2 min	
3	72 degree	5 min	1 cycle

Table IV-1-4 Ligation reaction

Ph⊿N ArgRS ligation for Blunt end vector	
⊿N ArgRS PCR productuion	3 μl
Salt solution	1 μl
Steric Water	1 μl
TOPO vector	1 μl

Table IV-1-5Restriction treatment for Insert

Ph⊿NArgRS PCR product terated with restriction enzyme (30 μ)	
	Final concentration
10×H buffer	1×
Xho I	8-24 U
Nde I	8-24 U
Ph⊿NArgRS/TOPO(0.36μg/μl)	7.2 μg
MillQ	up to 30 μl

Table IV-1-6 Restriction treatment for Vector

pET28c vector treated with restriction enzyme (30 μl)	
	Final concentration
10×H buffer	1×
Xho I	8-24 U
Nde I	8-24 U
pET28c(0.39 μg/μl)	5.85 μg
MillQ	up to 30 μl

1.1.2. Expression

ArgRS and ΔNArgRS sequence was subcloned into pET28c and expressed in *E.coli* strain BL21(DE3). Cells were grow in LB medium at 37 degrees 15 hours after IPTG induction at OD= 0.6. ArgRS and ΔN ArgRS expression was succeeded and to obtain enough soluble protein.

1.1.3. Protein Purification (Fig. IV -1-1)

ArgRS , Δ N ArgRS were purified same way.

The cells were suspended in lysis buffer (50 mM Tris-HCl buffer pH 7.5, 500 mM NaCl, 10% glycerol 10 mM Imidazole 5 mM 2-melcaptoethanol)(Table IV-1-7) sonicated 1 minute 50 % Amplitude and 0.5 cycle (Satrius Labsonnic P) and pause 1minute. A set of this cycle was done 3 times. Cell debris was removed by centrifugation for 30 minute at 10000 rpm, supernatant was heated to 80 degree for 20 minutes and centrifuged again at 10000 rpm for 30 minutes. The soluble fraction was purified with Ni-chelating column the protein was eluted with Elution buffer (50 mM Tris-HCl buffer pH 7.5, 500 mM NaCl, 10% glycerol 250 mM Imidazole 5 mM 2-melcaptoethanol) (Table IV-1-7) and dialyzed against Buffer (20 mM Tris-HCl buffer pH 7.5, 5 mM 2-melcaptoethanol) (Table IV-1-7) further purification was achieved using affinity column Heparin 1ml and anion exchange column. The protein were eluted using salt gradient with Buffer A and B (A 20 mM Tris-HCl buffer pH 7.5, 5 mM 2-melcaptoethanol B 20 mM Tris-HCl buffer pH 7.5, 2 M NaCl, 5 mM 2-melcaptoethanol) (Table IV-1-7, IV-1-8).The fractions containing protein were identified by their absorbance at 280 nm and gel electrophoresis using 12% SDS poly acryl amide gels (SDS-PAGE). These fractions were combined and dialyzed against Buffer (20 mM Hepes-NaOH pH 7.5, 5 mM 2-mercaptoethanol) (Table IV -1-7).

Table IV -1-7 Purification Buffers

Lysis buffer	50 mM Tris-HCl buffer pH 7.5, 500 mM NaCl, 10% grycerol 10 mM Indazole 5 mM 2-melcaptoethanol
Elution buffer	50 mM Tris-HCl buffer pH 7.5, 500 mM NaCl, 10% grycerol 250 mM Indazole 5 mM 2-melcaptoethanol
Dialysis(Heparin and Resourse Q)	20 mM Tris-HCl buffer pH 7.5, 5 mM 2-melcaptoethanol
Heparin column A buffer	20 mM Tris-HCl buffer pH 7.5, 5 mM 2-melcaptoethanol
Heparin column B buffer	20 mM Tris-HCl buffer pH 7.5, 2 M NaCl, 5 mM 2-melcaptoethanol
Resourse Q column A buffer	20 mM Tris-HCl buffer pH 7.5, 5 mM 2-melcaptoethanol
ResourseQ column B buffer	20 mM Tris-HCl buffer pH 7.5, 2 M NaCl, 5 mM 2-melcaptoethanol
Dialysis (crystallization)	20 mM Hepes-NaOH pH 7.5, 5 mM 2-melcaptoethanol

Table IV - 1-8 Columns for Purification

ArgRS	Thermal treatment	
	Ni-NTA	Affinity
	Hitrap heparin	Affinity
	ResourseQ	Anion exchange

1.2. tRNA

1.2.1. $tRNA^{Arg}$(CCU) transcription and purification (Fig. IV-1-1)

The transcription template was amplified by PCR (Table IV-1-9 IV-1-10) and using this template transcribed in the reaction solution (Table IV-1-11). After transcription, phenol extraction and EtOH precipitation was done subsequently the sample was purified with MonoQ and tRNA was eluted using salt gradient, the fractions were precipitated with EtOH and dried. The tRNA dissolved with 20 mM Hepes-NaOH pH 7.5, 10 mM $MgCl_2$ (Table IV-1-12).

Table IV-1-9 Amplification of template

	Final concentration
5×PrimeSTARTM Buffer (Mg2+ plus)	1×
dNPT Mixture (2.5 mM each)	200 μM each
primer M13R	0.2-0.3 μM
primer CCA	0.2-0.3 μM
tRNA ^{Arg}(CCU)/ pUC119	200 ng
Polymerase (2.5U/μl)	1.25 U/50 μl
MillQ	Up to 50 μl

Table IV-1-10 Primer for PCR

primer	seqence
CCA primer	TGGCGGACCGGCGGGGATTTGA (2'O-Me)
M13 Rv primer	CAGGAAACAG CTATGAC

Table IV-1-11 Transcription reaction

	Final concentration
HEPES-NaOH pH 8.0	80 mM
Spermidine	2 mM
DTT	20 mM
ATP	4 mM
CTP	4 mM
GTP	4 mM
UTP	4 mM
GMP	4 mM
KCl	40 mM
$MgCl_2$	20 mM
BSA	0.001%
RNase inhibitor	600 U
TemplateDNA(tRNA^{Arg}(CCU))	
T7 RNA polymerase	
MillQ	Up to 5ml

Table IV-1-12 Purification column for tRNA

MonoQ A buffer	20 mM Tris-HCl pH 7.0, 5 mM $MgCl_2$
MonoQ B buffer	20 mM Tris-HCl pH 7.0, 5 mM $MgCl_2$, 2M NaCl

1.2.2. tRNA^{Arg}(CAU), tRNA$^{Asp\text{-}Arg1\ -4}$, tRNA$^{Asp\text{-}ArgA\text{-}F}$ cloning transcription and purification

The tRNA mutation was made by PCR reaction using primers which has mutation sequence and complementary sequence each other and amplified and made the tRNA template. This template subcloned into pUC119 then obtained the plasmid which contains tRNA mutants sequence.

1.3. Crystallization (Fig. IV-1-2)

1.3.1. ΔN ArgRS and tRNA ^{Arg}(CCU)

The protein ΔN ArgRS was concentrated to 2 mg/ml with (Amincon Ultra-15, Millipore). The protein and tRNA$^{Arg(}$CCU) were mixed molecular ration 1 : 1 and added 1 mM Arg and 1 mM AMP-PNP. The protein ArgRS full length was also concentrated to 2 mg/ml and tRNA$^{Asp\text{-}Arg\ (1\text{-}4,\ A\text{-}F)}$ were added molecular ratio 1 : 1 and added 1 mM Arg and 1 mM AMP-PNP. And the mixture crystallized using several screening kits (Table IV-1-13).

The crystal ΔN ArgRS and tRNA^{Arg}(CCU) grew in 2.0 M $(NH_4)_2SO_4$, 2% PEG400, 0.2 M Hepes-NaOH at 24 degree. The crystal ArgRS full length and tRNA$^{Asp\text{-}Arg2}$, tRNA $^{Asp\text{-}Arg4}$ grew in 30-33% PEG 8000, 0.2 M $(NH_4)_2SO_4$, 0.1 M Hepes-NaOH pH 7.5 (Table IV-1-14).

The crystals were obtained the condition of crystal screen I 39th (0.1 mM Hepes-NaOH pH7.5, 2.0M $(NH4)_2SO_4$, 2% PEG400) at 20 degree. The crystals were plate shape and sensitive to temperature.

Table IV-1-13 Crystalized condition

Sample	Crystallization condition	Temp
⊿N ArgRS tRNA$^{Arg(CCU)}$	2.0 M $(NH_4)_2SO_4$, 2% PEG400, 0.2 M HEPES-NaOH	20
ArgRS tRNA$^{Asp\text{-}Arg2}$	30-33% PEG 8000, 0.2 M $(NH_4)_2SO_4$, 0.1 M Hepes-NaOH pH 7.5	20
ArgRS tRNA$^{Asp\text{-}Arg2}$	30-33% PEG 8000, 0.2 M $(NH_4)_2SO_4$, 0.1 M Hepes-NaOH pH 7.5	20

Table IV-1-14 Screening Kits

Crystallization Screening Kit
Crystal Screen I, II (Hampton)
Index (Hampton)
Natrix (Hampton)
Wizard I, II (Emerald)

1.3.2. ArgRS and tRNA$^{Arg\text{-}Asp2}$, ArgRS and tRNA$^{Arg\text{-}Asp4}$

The protein was concentrated around 2 mg/ml and crystallized same way to ΔN ArgRS and tRNA^{Arg}(CCU) case.

The crystals grew in the condition 0.1 M Tris-HCl 30- 33% PEG 4000, 0.2 M $(NH4)_2SO_4$, and 100 mM Tris-HCl or Hepes-NaOH pH 7.5 or 8.0. Thin, needle-like crystals were obtained and measured at photon factory (Tsukuba) but the crystals were too thin to correct enough diffraction data. The 4 Å diffraction was obtained.

1.4. Aminoacylation reaction

The aminoacylation reaction of tRNA was measured at 65 degree in 100 mM Hepes -NaOH buffer (pH 7.5) containing 30 mM KCl, 0.1 mg/mL BSA, 10 mM $MgCl_2$, 4 mM ATP, 50 mM -[14C]Arg (100 Ci/mL) Moravek, Lane Brea, CA, USA), 0.2 μM wild-type ArgRS or 2 μM ΔN ArgRS, and various concentrations of $tRNA^{Arg}$ (CCU)(0.5 to 16 μM). Aliquots of 9 μl at varying time intervals were quenched with 5 μl of 1% trichloroacetic acid and spotted onto Whatman 3 MM disks (Table IV-1-15). Radioactivity was quantified in a scintillation counter. The kinetic constants were derived from a Line weaver plot.

Table IV-1-15 Km、kcat mesurement of ArgRS against tRNAArg(CCU) (15 μl)

	Final concentration
Hepes-NaOH pH7.5	100 mM
KCl	30 mM
BSA	0.1 ng/ml
MgCl2	10 mM
ATP	20 mM
[14C]Arginine	100 μM
ArgRS full length / ⊿N ArgRS	0.2 mM
$tRNA^{Arg}$ (1μM to32μM)	

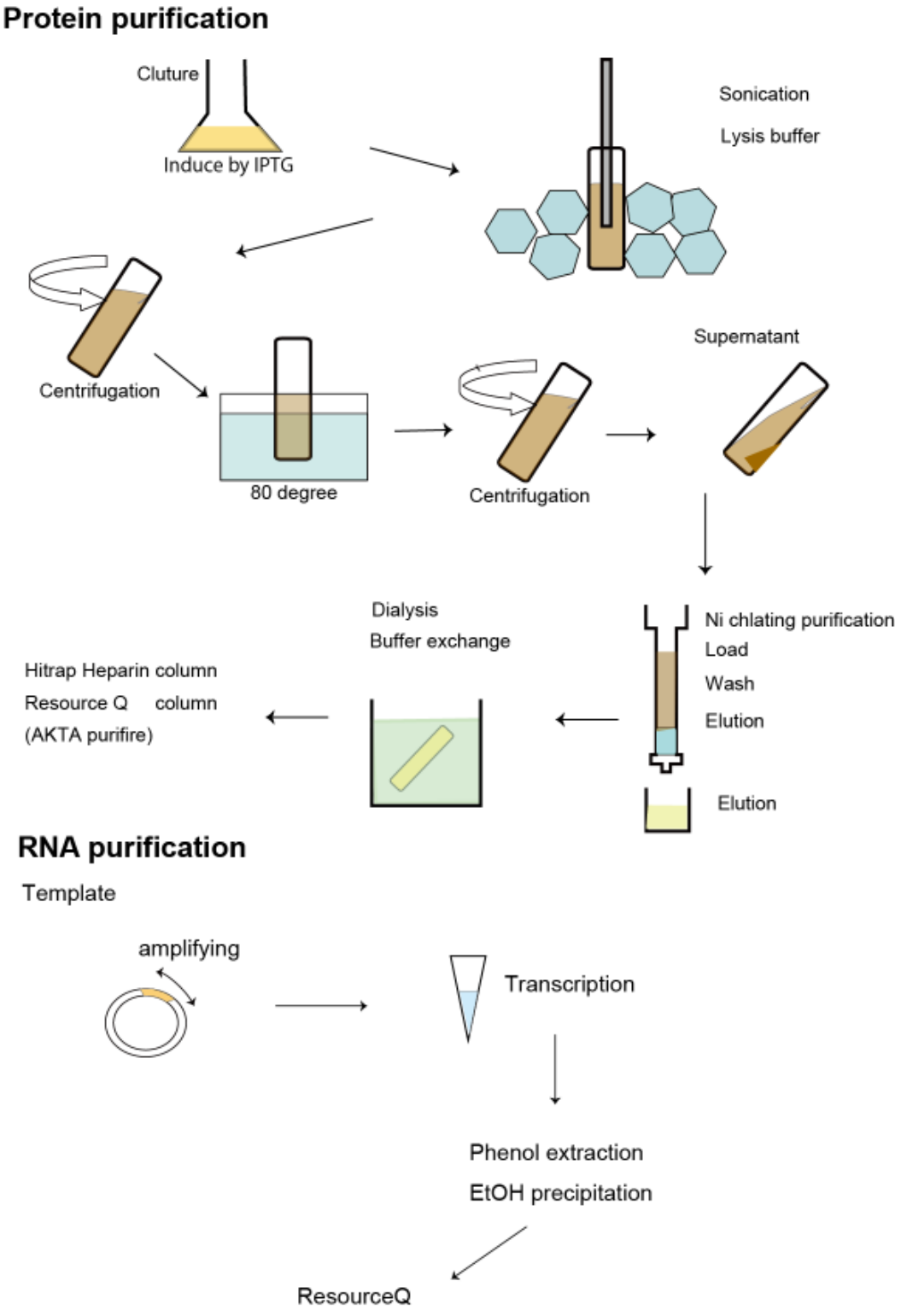


Fig. IV-1-1 The scheme of Protein and RNA purification

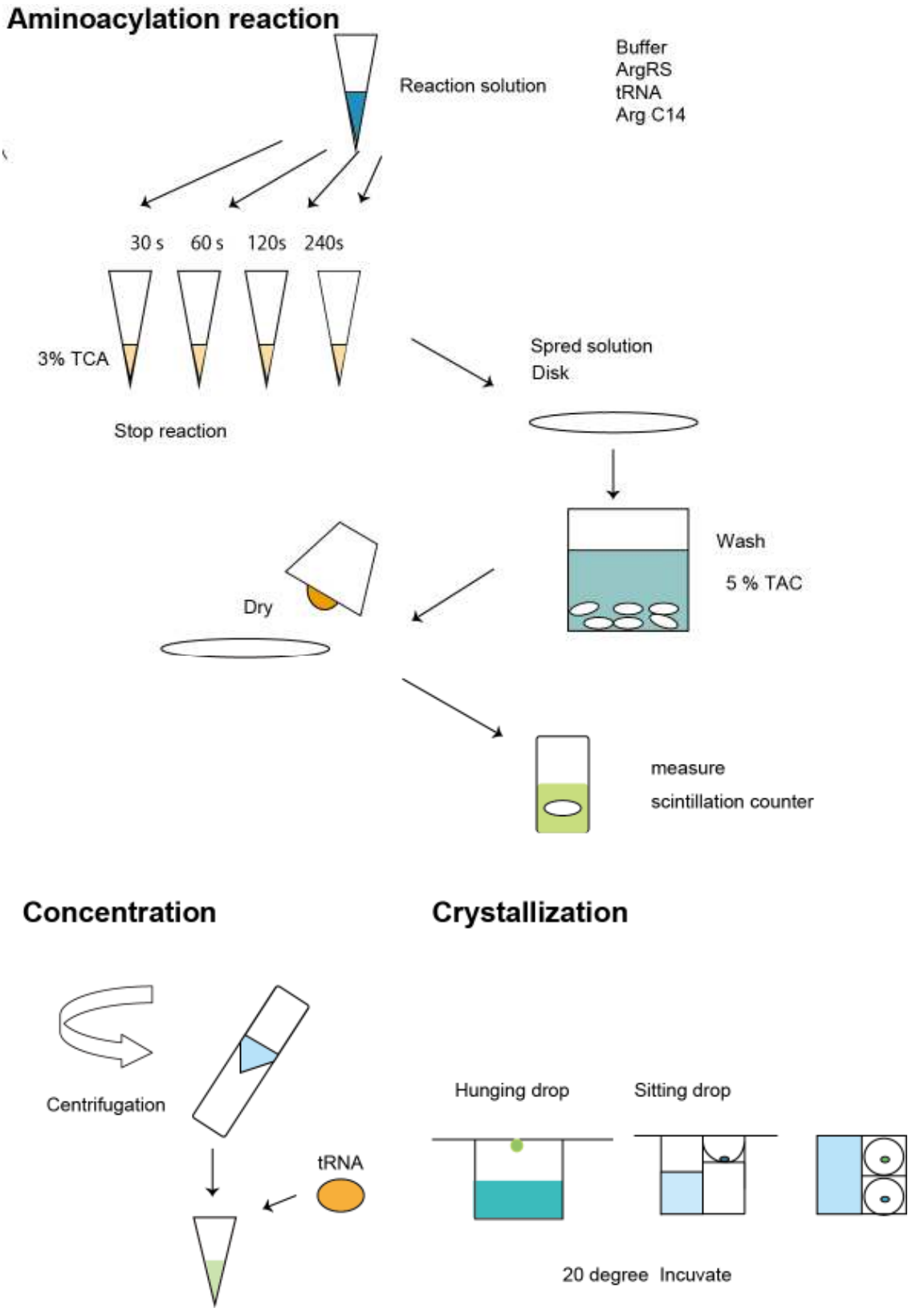


Fig. IV-1-2 The scheme of Aminoacylation and Crystallization

2. LaRP7 7SK

2.1. LaRP7

Plasmid for cloning Fig.IV-2-1, Table IV-2-1. IV2-2

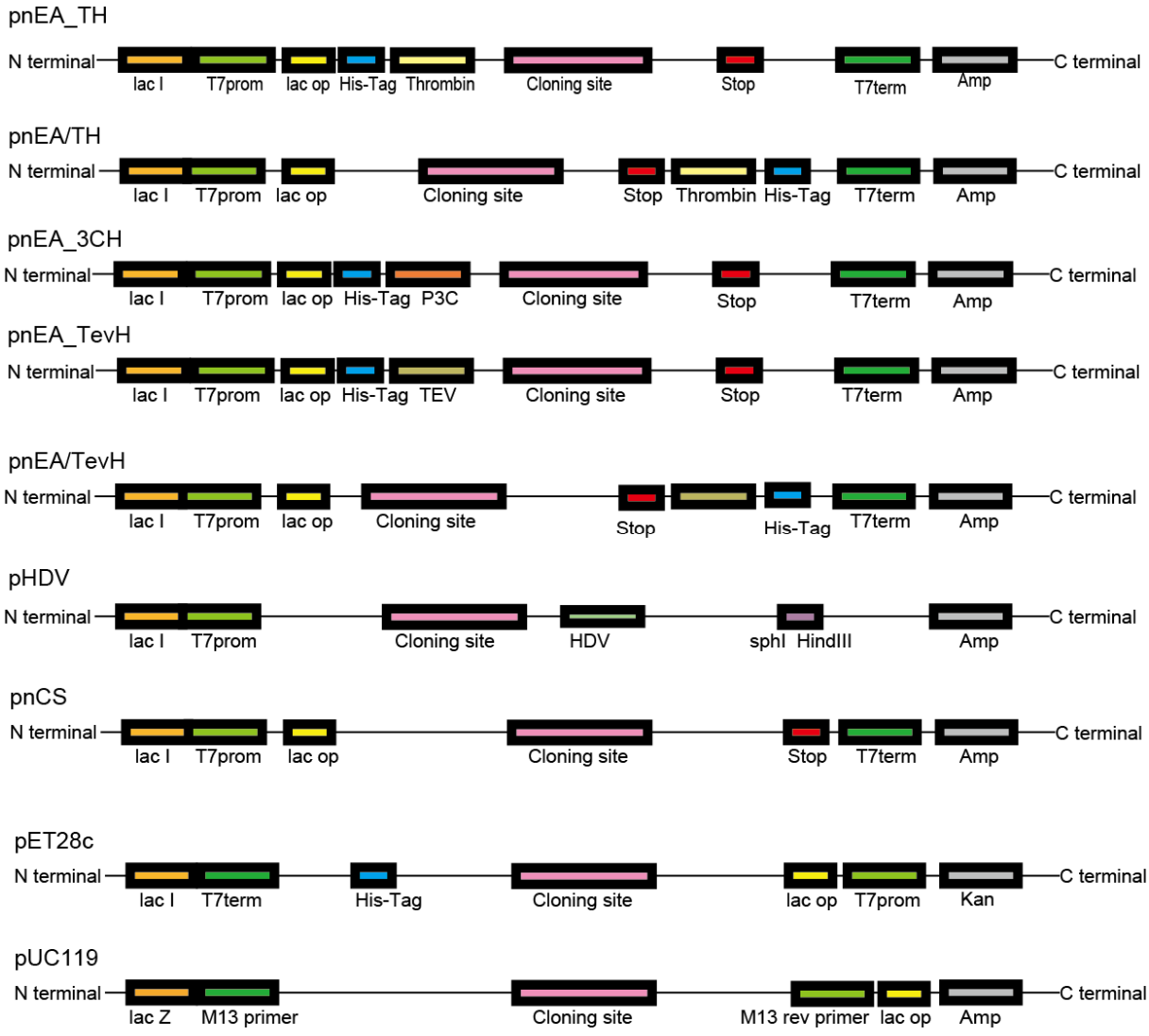


Fig. IV-2-1 Vectors for cloning

Table IV-2-1 Plasmids

	Name	Tag position	Tag	Clevage site		resistance	primer	Restriction
1	PmcnEA _T H	N terminal	His-Tag	Thrombin		Amp	T7	NdeI, BamHI
2	PmcnEA /T H	C terminal	His-Tag	Thrombin		Amp	T7	NdeI, BamHI
3	PmcnEA N 3C H	N terminal	His-Tag	HRV3C		Amp	T7	NdeI, BamHI
4	PmcnEA N TEV H	N terminal	His-Tag	TEV		Amp	T7	NdeI, BamHI
5	pHDV				HDV	Amp	T7	NdeI, ,Nco
6	pnCS					Amp	T7	NdeI, BamHI
7	pET28c	N terminal	His-Tag			Kan	T7	NdeI, XhoI
8	pUC119					Amp	M13	HindIII, BamHI

Table IV-2-2 Site sequence

	Sequence
HisTag	CATCATCATCATCATCAC HisHisHisHisHisHis
Cloning site PnEA	NdeI, XhoI, AflII, MunI, BamHI
Cloning site pHDV	Nde I Nco Dra I, Xho I, Eco RV, Xba I
Cloning site pET28c	Nde I, Nhe I, Bam HI, Eco RI, Sac I, Hind III, Not I, Xho I
Cloning site pUC119	EcoR I, Sac I, Kpn I, Sma I, BamH I, Xba I, Sal I, Pst I, Sph I, Hind III
Thrombin	Leu-Val-Pro-Arg ↓ Gly-Ser
TEV	Glu-Asn-Leu-Tyr-Phe-Gln-Gly ↓ Ser
P3C	Leu-Glu-Val-Leu-Phe-Gln- ↓ -Gly-Pro
T7promotor	TAATACGACTCACTATAGG
T7terminator	GCTAGTTATTGCTCAGCGG
M13(M3)	GTAAA ACGAC GGCCA GT
M13 reverse	CAGGA AACAGCTATG AC

2.1.1. Cloning LaRP7N1_208 in pnEA/TH, LaRP7N28_208 in pnEA-TH, LaRP7N28_208 in pnEA/TH

LaRP7N1_208 in PnEA-TH was already made in our team. The sequence LaRP7N1_208 and LaRP7N28_208 was amplified by PCR (Table IV-2-3) then purified with Extraction II. The amplified sequences and vector pnEA-TH and pnEA/TH were treated with restriction enzyme NdeI and BamHI (Table IV-2.4) then vector was subsequently treated with CIP for dephosphorylation at 37 degree for 1 hour and joined protein sequence and vector by ligase at 16 degree for 15 hours (Table IV-2-5).

Table IV-2-3 Amplification of Lp7N1_208, Lp7N28_208 sequence

	Lp7N1_208		Lp7N28_208	
Template	Lp7FL/pEA-tNH 10μM	0.5 μM	Lp7FL/pEA-tNH 10μM	0.5 μM
Buffer	10 × HF buffer	1×	10 × HF buffer	1×
Primer	BCC 825 (10 μM)	0.2 μM	BDM 256/Larp28 (10 μM)	0.2 μM
Primer	BDM129/Larpb (10 μM)	0.2 μM	BDM129/Larpb (10 μM)	0.2 μM
Nucleotide	dNTP (2mM)	0.2 mM	dNTP (2mM)	0.2 mM
enzyme	Phusion (10 U)	0.2 μl	Phusion (10 U)	0.2 μl
millQ	Up to 20 μl		Up to 20 μl	

	Temp	Sec	Cycle
1	98 degree	1 min	1 cycle
2	98 degree	30 s	30 cycle
	55 degree	7 s	
	72 degree	15 s	
3	73 degree	7 min	1 cycle

Table IV-2-4 Lp7N1_208 treated with restriction enzyme

PCR product	Lp7N1-208	50 μl	Lp7N28-208	50 μl
Buffer 4 10×		1×		1×
BSA 10 mg/ml		0.1 mg/ml		0.1 mg/ml
BamHI(10U)		0.2 U		0.2 U
NdeI(15U)		0.2U		0.2U
millQ		Up to 70 μl		Up to 70 μl

Table IV-2-5 Ligation

Vector pnEA/TH (32 μM)	4 μl	Vector pnEA-TH (64 μM)	1 μl	Vector pnEA/TH (32 μM)	4 μl
Insert 1-208	4 μl	Insert 28-208	7 μl	Insert 28-208	4 μl
ATP 10mM	1 μl	ATP 10mM	1 μl	ATP 10mM	1 μl
Ligasebuffer T4 10×	1 μl	Ligasebuffer T4 10×	1 μl	Ligasebuffer T4 10×	1 μl
Ligase	1 μl	Ligase	1 μl	Ligase	1 μl
MillQ	Up to 10 μl		Up to 10 μl		Up to 10 μl

2.1.2. Cloning LaRP7N1_208 in pnEA-3CH

LaRP7N1_208 in pnEA/TH was amplified in *E.coli* strain DH5α was cultured in 2LB at 37 degree for 15 hours and plasmid was purified with Midi prep. The plasmid was treated with restriction enzyme NdeI and BamHI (Table IV-2-6) then migrated with 1 %

agarose gel and the band of LaRP7N1_208 was cut (without exposed UV) and purified with Extraction II (purification kit) and obtained LaRP7N1_208 insertion. The vector pnEA-3CH was also treated with restriction enzyme NdeI and BamHI then treated with CIP (0.08 U final concentration) dephosphorylation at 37 degree for 30 minutes purified with phenol: chloroform and precipitated with EtOH. The LaRP7N1_208 and pnEA-3CH was joined by ligase at 16 degree for 15 hours (Table IV-2-6).

Table IV-2-6 Lp7N1_208/ pnEA plasmid treated with restriction enzyme and Ligation

Insert		Vector	
Lp71-208/pnEANhistag	OD= 280 ng/µl 100µl 30 µg	pnEA 3C N histag	OD= 290 ng/µl 5µl 1.5 µg
BamHI 15U	3µl 0.3U	BamHI 15U	1µl 0.3U
NdeI 20U	3µl 0.4U	NdeI 20U	1µl 0.4U
10×Reactionbuffer 4	1 ×	10×Reactionbuffer4	1 ×
MillQ	Up to 150	MillQ	Up to 50

Ligation 10µl	
Vector pnEA 3C	3 µl
Insert Lp7N1_208	3 µl
10×buffer	1 ×
ATP 20mM	2 mM
ligase	1µl

2.1.3. Expression

All LaRP7N1_208, LaRP7N28-208 constructs were expressed in *E.coli* strain BL21(DE3) pRARE. Cells are growing in auto induction medium at 37 degree until OD= 0.6 then decrease 25 degree and cultures for 15 hours. All constructs expression were succeeded and to obtain enough soluble protein.

Auto induction medium, a period of cell growth was followed by spontaneous induction of protein expression without monitoring cell density and without conventional induction with IPTG. This method is based on media components that are metabolized differentially to promote growth to high density and automatically induce protein expression from *lac* promoters. Cell mass and target protein yield are often increased several-fold as compared with conventional protocols using induction with IPTG.

2.1.4. Purification (Fig IV-2-2)

Protein purification varies from simple one-step precipitation procedures to large-scale validated production processes. Often more than one purification step is necessary to reach the desired purity. The key of successful and efficient protein purification is to select the most appropriate techniques, optimize their performance to suit the requirements and combine them in a logical way to maximize yield and minimize the number of steps are required.

Most purification schemes involve some form of chromatography. As a result, chromatography has become an essential tool in every laboratory where protein purification is

needed. The availability of different chromatography techniques with different selectivity provides a powerful combination for the purification of any bimolecular.

Recombinant DNA developments over the past decade have revolutionized the production of proteins in large quantities. Proteins can even be produced in forms that facilitate their subsequent chromatographic purification.

Cells were suspended in lysis buffer (100 mM Tris-HCl pH8.0, 500 mM NaCl, 25 mM Imidazole, 15% glycerol, 1.4 mM 2-melcaptethanol, protein inhibitor cocktail) and sonicated 1 minute 50 % Amplitude and 0.5 cycle (Satrius Labsonnic P) and pause 1minute. A set of this cycle was done 3 times. Cell debris was removed by centrifugation for 1 hour at 20000 rpm. The soluble fraction of the cell lysate was purified using Ni-chelating column, Ni-beads were mixed with protein and adsorbing 5 hours at 4 degree. The protein solution pour on column and beads were wash with 10 times volume of lysis buffer and wash buffer (100 mM Tris-HCl pH8.0, 1 M NaCl, 25 mM Imidazole, 15% glycerol, 1.4 mM 2-melcaptethanol) and the protein eluted with elution buffer (100 mM Tris-HCl pH8.0, 500 mM KCl, 250 mM Imidazole, 5% glycerol) with 5 ml. Further purification of the protein achieved using cation exchange column Hitrap SP and the protein was eluted using salt gradient (Buffer A 20 mM Hepes-NaOH pH7.2, 1 mM DTT and Buffer B 20 mM Hepes-NaOH pH7.2, 2M NaCl, 1 mM DTT) subsequently the protein was purified by size-exclusion-chromatography Superdex 75 10/300 with gel filtration buffer (20 mM Hepes-NaOH pH 7.2, 200 mM NaCl, 1 mM DTT), on AKTA system, is used for rapid purification of proteins at low to medium pressures. AKTA purifier is able to run from μg to mg protein. Crystallography is needed mg amount of high purity protein. (Table IV-2-7, IV-2-8)

Table IV-2-7 Purification column

Larp7	Ni-NTA	Affinity column
	Hi trap SP	Cation exchange column
	Superdex 75 10/300	Gel chromatography column

Table IV-2-8 Buffers of purification

Lysis buffer	100 mM Tris-HCl pH8.0, 500 mM NaCl, 25 mM Imdazol, 15% grycerol, 1.4 mM 2-melcaptethanol, 1 protein inhibitor cactail
Wash buffer	100 mM Tris-HCl pH8.0, 1 M NaCl, 25 mM Imdazol, 15% grycerol, 1.4 mM 2-melcaptethanol
Elution buffer	100 mM Tris-HCl pH8.0, 500 mM KCl, 250 mM Imidazome, 5% grycerol
SP Abuffer	20 mM Hepes-NaOH pH 7.2, 200 mM NaCl, 1 mM DTT
SP B buffer	20 mM Hepes-NaOH pH 7.2, 2 M NaCl, 1 mM DTT
GF	20 mM Hepes-NaOH pH 7.2, 200 mM NaCl, 1 mM DTT

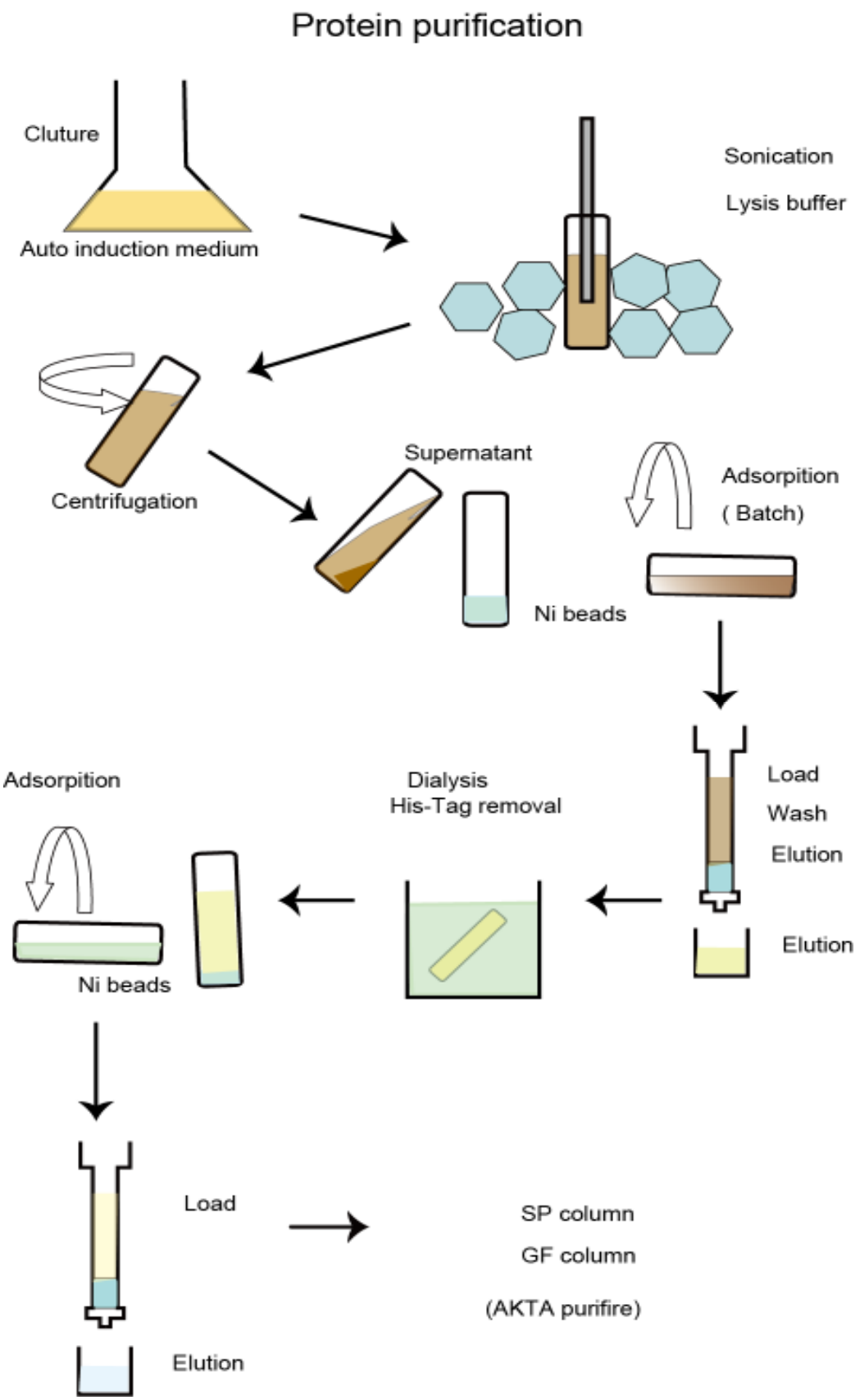

Protein purification
Cluture
Auto induction medium
Sonication
Lysis buffer
Centrifugation
Supernatant
Ni beads
Adsorpition
(Batch)
Load
Wash
Elution
Elution
Dialysis
His-Tag removal
Adsorpition
Ni beads
Load
Elution
SP column
GF column
(AKTA purifire)

Fig. IV-2-2

2.2. 7SK

2.2.1. HP4U and HP4 transcription and purification (Fig. IV-2-3)

To establish more effective transcription protocol, we tried 5 different templates and followed the yield of HP4U. 3 templates were linearized by restriction enzyme, 1 template was amplified by PCR from pHDV plasmid and the other was hybridized of HP4U DNA oligomer and T7 promoter DNA oligomer, (Table IV-2-9, IV-2-10, IV-2-11)

In vitro transcription reaction was done in 4 mM each rNTP, 100 μl templates, 0.1 mg/ml T7 RNA polymerase, 30 mM Tris-HCl pH 8.0, 2 mM Spermidine, 0.01% TritonX 100, 16 mM $MgCl_2$, 5 mM DTT. (Table IV-2-12)

After 4 hours transcription reaction, 40 mM $MgCl_2$ added to the HP4U reaction solution, which is synthesized from linerlization HDV plasmid, and pcr product then incubate at 65 degree for 10 minutes and 37 degree for 1 hour to activate hammerhead ribozyme. Each solution was treated with phenol-chloroform extraction and Ethanol precipitation. Subsequently, four HP4Us purified by Preparative denaturing poly acryl amide gel electrophoresis. HP4U band were cut and extracted with 500 mM $AcNH_4$ pH 7.0, 1 mM EDTA pH8.0, 2 mM MgCl2, 0.1% SDS. HP4U was loaded to Mono Q column (Table IV-2-14, IV-2-15). After Mono Q, HP4U was precipitated with EtOH and diluted with 500 μl. The yield from oligo template showed six times better yield than the others but the HP4U from oligo template band was not clear in acryl amide native gel. This means even after preparative gel, this HP4U is degrading. The molecule has one unique form is important, for crystallization thence HP4U should be transcribed from pHDV template. However, 3'end form is different between HP4U from oligo template and pHDV template. 3' end of HP4U from oligo template has 2'OH on another hand HP4U from pHDV template has cyclic phosphate.

2.2.2. HP4U, HP4 with GMP transcription and purification

10 times GMP added in transcription reaction solution. 4 mM each rNTP (ATP, UTP, CTP, GTP) 40 mM GMP, 0.3~0.5 mg templates, 0.1 mg/ml T7 RNA polymerase, 30 mM Tris-HCl pH 8.0, 2 mM Spermidine, 0.01% TritonX 100, 16 mM MgCl2, 5 mM DTT. Purification was same way of HP4U without GMP (Table IV-2-13).

2.2.3. 7SK purification

For binding assays, in order to compare the binding of LaRP7 to the different RNAs, we used a 7SK stock that was prepared similarly by in vitro transcription from pHDV_7SK, in the laboratory.

Table IV-2-9 Hybridization

Material	Final concentration
Tris-HClpH7.6	100 mM
$MgCl_2$	5 mM
Oligo t7 (MW 5799)	10 μM
Oligo Hp4U (MW 14848)	10 μM
MillQ	Up to 5ml

Table IV-2-10 Linearization

Template Plasmid	HP4U in pHDV	100 μM
Buffer	10 ×Tango buffer	1×
Enzyme	Xba I, Xho I. Eco V	0.2 U

Table IV-2-11 PCR for Template

Template plasmid	HP4U in pHDV	100 ng
Buffer	10 × HF buffer	1 ×
Primer	T7promoter	2 μM
Primer	HDV primer	2 μM
Nucleotide	dNTP 2mM	2 μM
Enzyme	Fusion	1 μl
MillQ	Up to 100 μl	

	Temp	Time	Cycle
1	98 degree	1 min	1 cycle
2	98 degree	30 s	30 cycle
	55 degree	7 s	
	72 degree	15 s	
3	73 degree	7 min	1 cycle

Table IV-2-12 Transcription

Material	Final concentration
PCR product	10 μM
Tris-HCl pH 8.0	30 mM
Spermidine	2 mM
TritonX	0.01%
rNTPs	4 mM
$MgCl_2$	10 mM
DTT	5 mM
T7polymerase	0.1 mg/ml
MillQ	Up to 5 ml

Table IV-2-13 Transcription with GMP

Material	Final concentration
PCR product	10 μM
Tris-HCl pH 8.0	30 mM
Spermidine	2 mM
TritonX	0.01%
rNTPs	4 mM
GMP	20 mM
$MgCl_2$	10 mM
DTT	5 mM
T7polymerase	0.1 mg/ml
MillQ	Up to 5 ml

Table IV-2-14 Purification column

HP4U,HP4	Phenol extraction	Protein and salt eliminate
	Acrylamide gel	Size separation
	Mono Q	anionexchnge

Table IV-2-15 Buffer

RNA2 buffer	10 mM Cacodilate-NaOH, 2 mM MgCl2, 0.25 mM EDTA
Elution buffer	0.5 M NH4Ac pH7.0, 0.1% SDS, 1mM EDTA, 2mM MgCl2
MonoQ A buffer	20mM Bis-tris pH 7.0, 2mM MgCl2, 0.25mM EDTA
MonoQ B buffer	20mM Bis-tris pH 7.0, 2mM MgCl2, 0.25mM EDTA, 2M NaCl

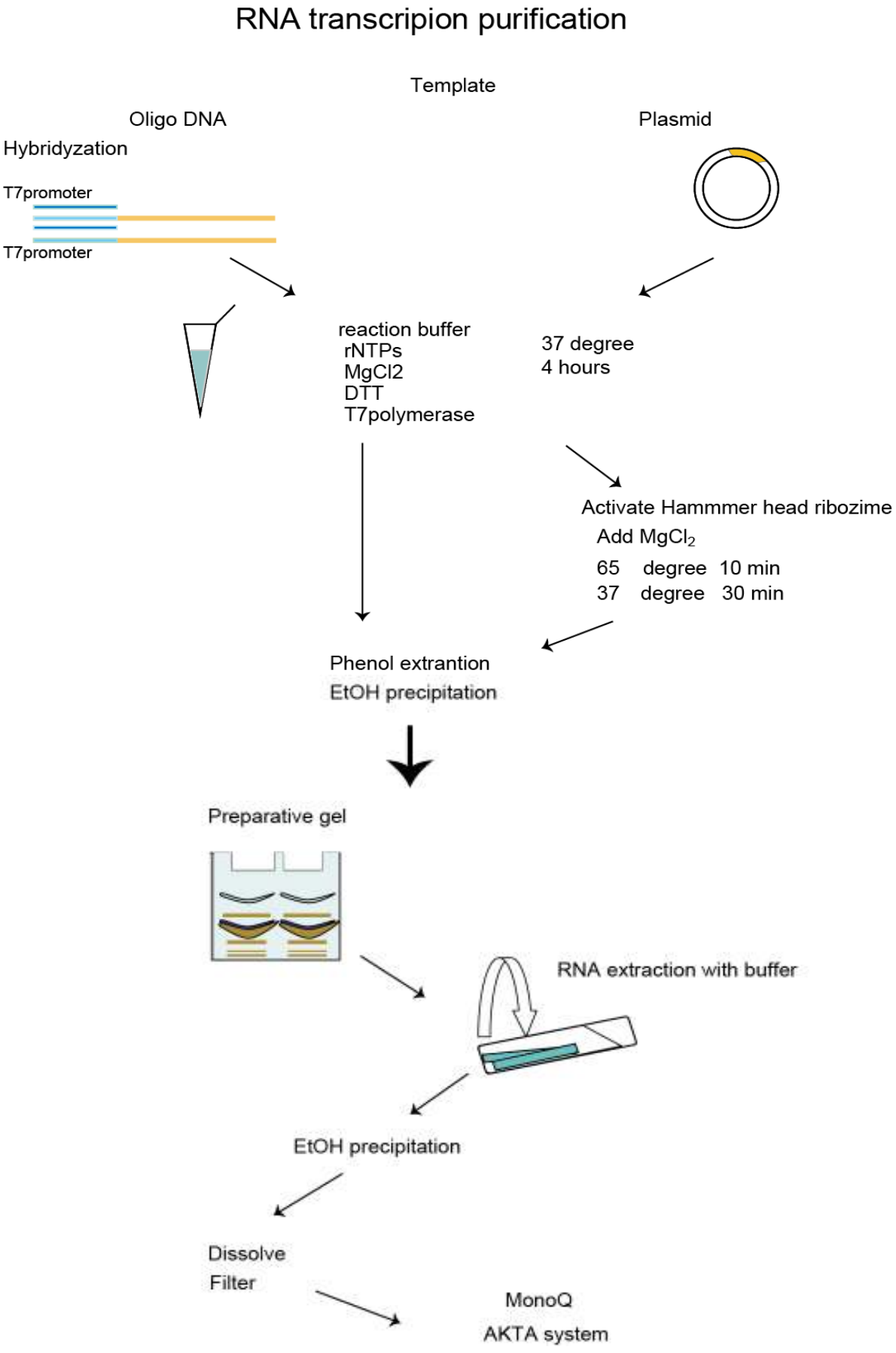

RNA transcripion purification
Template
Oligo DNA
Plasmid
Hybridyzation
T7promoter
T7promoter
reaction buffer
rNTPs
MgCl2
DTT
T7polymerase
37 degree
4 hours
Activate Hammmer head ribozime
Add MgCl2
65 degree 10 min
37 degree 30 min
Phenol extrantion
EtOH precipitation
Preparative gel
RNA extraction with buffer
EtOH precipitation
Dissolve
Filter
MonoQ
AKTA system

Fig. IV-2-3

2.3. Crystallization

X-ray crystallography is the method of choice for determining high-resolution structures of large RNA molecules with protein. The interaction between protein and RNA can only be revealed in detail by X-ray crystal structure determination and can be used to compare related RNAs and identify conformational changes that may accompany biochemical activity.

A crystal is a well-ordered three-dimensional array of molecules held together by non-covalent interactions called “crystal contacts.” In general, crystals of nucleic acids and proteins can be grown by slow, controlled precipitation from aqueous solution under non-denaturing conditions. Crystallization of all combination of Larp7N1_208 and HP4U was using a dispenser machine (Genesis workstation) and a crystallization machine (Cartesian Technology workstation). Sitting drops composed of 200 nil protein solution and 200 nil reservoir of crystallization solution were equilibrated against 50 μl well solution on the round-bottom 96 well plate (The MCR crystallization plate UV polymer). Initial crystallization trials utilized the sitting-drop vapor-diffusion technique at 4 degree, 17 degree and 20 degree. Initial crystallization conditions tested included Classics, JCSG+, Index, ARN (96 conditions made by lab), Wizard I, Wizard II, Protein Complex. The purification ways were considered that column order, buffer choice, salt concentration, additives. During concentration, the protein is followed by DLS.

2.3.1. Crystallization of LaRP7N1_208

For crystallization of Lp7N1_208 several buffer condition were tried. However, crystals did not g row (Table IV -2-12).

2.3.2. Crystallization of LaRP7N1_208 with RNA oligomer -UUUCUUU-, LaRP7N1_208 with RNA (HP4U transcribed by DNA oligomer)(HP4U transcribed by pHDV plasmid)

LaRP7N1_208 in 20 mM Hepes-NaOH pH7.2, 100 mM NaCl, 1 mM TCEP were mixed with RNA oligomer or HP4U in 10 mM Na-cacodylate pH 6.8, 5 mM $MgCl_2$, 0.25 mM EDTA at 24 degree. RNA and protein mixed at a molecular ratio of 1:1 (Table IV -2-17). In almost all drops precipitation observed after several trial of initial screening, in LaRP7N1_208 and oligomer drop, really thin crystals observed in the condition 2.4 M Sodium maronate pH7.0 at 24 degree. Tried to optimized the buffer condition but never grow up again those crystals. When LaRP7N1_208 and HP4U were mixed, precipitant arose. For crystallization, the protein condition is important therefore the precipitant should be avoided.

The concentration way was changed. Protein and RNA were mixed during concentration. In this case precipitant was not observed. The crystallization was continued but crystals were not observed.

2.3.3. Crystallization of LaRP7N1_208 (without tag) with RNA (HP4U transcribed with GMP by pHDV plasmid)

The protein just after purified by GF column (1 mg/ml) LaRP7N1_208 in 20 mM Hepes-NaOH, 200 mM NaCl and same molecular ratio of HP4U in 10 mM Na-cacodylate pH 6.8, 0.25 mM EDTA was mixed and concentrated then crystallized.

The crystals grew in the 5 conditions the condition [0.1 M Succinic acid pH 7.0, 15% PEG 3350], [0.1 M Hepes-NaOH pH 7.0, 15% PEG 4000], [0.1 M Hepes-NaOH pH 7.0, 15% PEG 4000, 0.1 M Magnesium chloride] [0.1 M Hepes-NaOH pH 7.5, 15% PEG 6000,0.1 M Potassium chloride], in two weeks at 4 degree but the crystals grew in the condition [0.1 M Hepes-NaOH pH 7.0, 15% PEG 4000], [0.1 M Hepes-NaOH pH 7.0, 15% PEG 4000, 0.1 M Magnesium chloride], [0.1 M Hepes-NaOH pH 7.5, 15% PEG 6000, 0.1 M Potassium chloride] melted in next 3 weeks. Then 4 weeks after the crystallization, new crystals grew in the condition [0.1 M Hepes-NaOH pH 7.5, 12% PEG 8000, 0.2 M Sodium chloride], [0.1 M Tris-HCl pH 8.0, 12% PEG 2000 MME, M Potassium chloride] but these crystals were thin and like needles(Table IV-2-18).

The crystals were obtained from [0.1 M Hepes-NaOH pH 7.5, 15% PEG 6000, 0.1 M Potassium chloride] and [0.1 M Succinic acid pH 7.0, 15% PEG 3350] was measured by X-ray at SOLEIL (Paris) and obtain 7 Å diffractions. 25% glycerol was used as cryo protectant for measurement.

Table IV-2-16 Protein buffer for Crystallization of Lp7N1_208

Buffer condition	Concentartion
20mM Hepes-NaOH pH7.2, 200 mM NaCl 1.4 mM 2-melcaptethanol	6 mg/ml 8 mg/ml 10 mg/ml 16 mg/ml
20 mM Hepes pH 7.2, 10 mM NaCl	12 mg/ml
20 mM Hepes pH 7.2, 100 mM NaCl	22 mg/ml

Table IV-2-17 Lp7N1_208 and RNA concentration for Crystallization

Lp7N1_208	20 mM Hepe-NaOH pH7.2, 100 mM NaCl, 1 mM TCEP	10 mg/ml (380 μM)
HP4U	10 mM Na-cacodylate pH 6.8, 2 mM $MgCl_2$, 0.25 mM EDTA	500 μM to 1 mM
HP4U	10 mM Na-cacodylate pH 6.8, 0.25 mM EDTA	500 μM to 1 mM
Oligomer	10 mM Na-cacodylate pH 6.8, 0.25 mM EDTA	1 mM

Table IV-2-18 Crystallization reservoir condition

	Buffer condition	size	crystals
A	0.1 M Succinic acid pH7.0,15 % PEG 3350	50×100×200μm	Single, poly
B	0.1M Hepes pH 7.5, 15% PEG6000, 0.1M KCl	50×100×200	Single
C	0.1 M Hepes pH7.0,15% PEG4000	----------	poly
D	0.1 M Hepes pH7.0,15% PEG4000,0.1M MgCl2	----------	Thin plate
E	0.1 M Hepes pH7.5, 12% PEG8000, 0.2M NaCl	100×50×100	needle
F	0.1 M Tris pH8.0, 15% PEG 2000 MME, 0.1 M KCl	100×50×50	plate

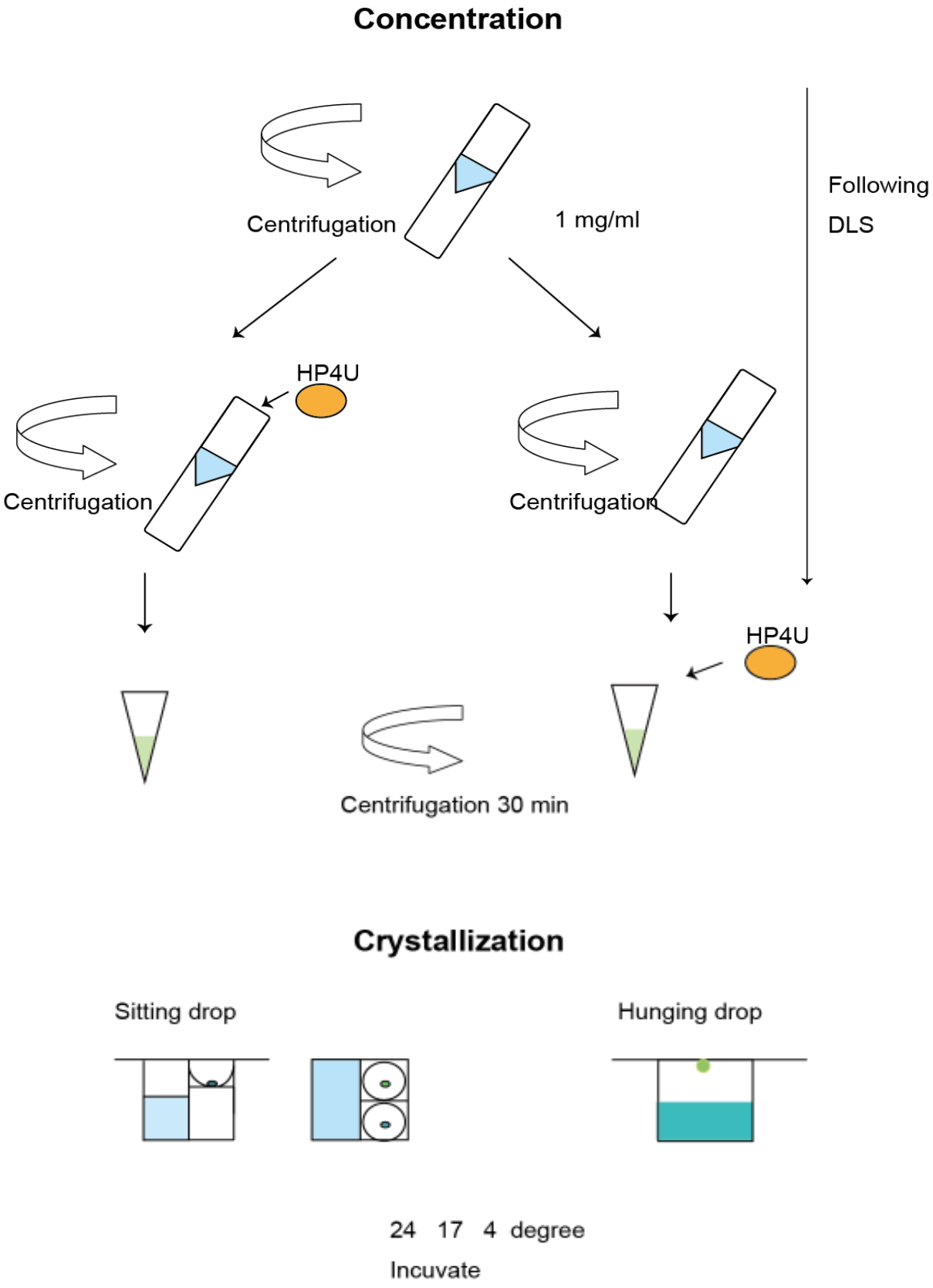

Concentration
Centrifugation
1 mg/ml
Following
DLS
HP4U
Centrifugation
Centrifugation
HP4U
Centrifugation 30 min
Crystallization
Sitting drop
Hunging drop
24 17 4 degree
Incuvate

Fig. IV-2-4

3. Binding analysis

3.1. EMSA (Fig. IV-3-1)

EMSA stands for electrophoretic mobility shift assay. Mobility shift assays are also known as gel shift assay. EMSA is a common technique used to study DNA-protein or RNA-protein interactions. There are in fact two types of EMSAs, including denaturing EMSA and non-denaturing. Denaturing EMSA is conducted under denaturing conditions and functions to determine the number and size of the proteins that are directly binding to and interacting with a DNA or RNA nucleic acid sequence fragment. Denaturing EMSAs are done by UV-cross linking any directly interacting proteins to the DNA or RNA fragment (by making covalent bonds), and then running the complex under denaturing conditions which removes any non-covalent interactions (such as protein-protein interactions). Thus, we could know nucleic acid fragments that are bound only to directly interacting protein factors.

Non-denaturing EMSA functions to determine the different types of complexes that bind to a DNA or RNA nucleic acid sequence fragment. Non-denaturing EMSAs are conducted under non-denaturing conditions.

A mobility shift assay generally involves electrophoretic separation of a protein-DNA/RNA mixture on a poly acryl amide or agarose gel. The speed at which different molecules move through the gel is determined by their size and charge, and to a lesser extent, their shape. The control lane without protein present will contain a single band corresponding to the unbound DNA/RNA fragment. However, assuming the protein is capable of binding to the RNA fragment, the lane with protein present will contain another band that represents the larger complex of RNA bound to protein. From the ratio of bound to unbound RNA, the affinity of the protein to the RNA sequence may be determined.

Often, run with a competitor oligo nucleotide to determine the most favorable binding sequence for the binding protein. The use of different oligo nucleotides of defined sequence allows the identification of the precise binding site by competition. Variants of the competition assay are useful for measuring the specificity of binding and for measurement of association and dissociation kinetics. For visualization, the RMA fragment is usually radioactive or fluorescent label, as standard ethidium bromide staining lacks the sensitivity to detect the relatively small amounts of RNA used in these experiments. 3 types of EMSA were performed in this experiments.

3.1.1. Acryl amide gel and stained with Toluidine Blue and Stain-all

The protein was diluted with protein buffer (20 mM Hepes-NaOH pH 7.5, 100 mM NaCl) and RNA was diluted with RNA buffer (10 mM Na-cacodylate pH 6.8, 2 mM $MgCl_2$, 0.25 mM EDTA) then mixed and incubated for 20 minutes at 4 degree the final concentration of RNA was 4 μM each and protein concentration were 1 μM, 2 μM, 4 μM,8 μM,16 μM. The 10 μl reaction sample was migrated with 6% acryl amide gel then stained toluidine blue or stain-all (Table. IV-3-1). To increase the intensity of RNA two migration buffer were examined TG buffer (Tris-Glycine: 50 mM Tris, 80 mM Glycine) and TBE buffer (Tris-Borate-EDTA 89 mM Tris, 89 mM Borate, 10 mM EDTA) (Table. IV-3-2).

3.1.2. Agarose gel Fluorescein

The protein was diluted with protein buffer (50 mM K-phosphate pH7.6, 250 mM NaCl, 5 mM $MgCl_2$, 0.1 mg/ml BSA) and RNA was diluted with RNA buffer (50 mM K-phosphate pH 7.6, 250 mM NaCl, 2 mM $MgCl_2$, 0.25 mM EDTA, 1 mg/ml tRNAtotal) sample was prepared at 10 μl, 2 μl from RNA, 2 μl from protein, 5 μl from EMSA buffer (100 mM K-phosphate pH 7.6, 500 mM NaCl, 2 mM $MgCl_2$, 0.25 mM EDTA, 1 mg/ml tRNAtotal, 0.1 mg/ml BSA) then mixed and incubated for 20 minutes at 4 degree the final concentration of RNA was 4 μM each and protein concentration were 4 μM, 8 μM, 16 μM, 32 μM, 64 μM. The 10 μl reaction sample was migrated with 1.5% agarose gel then measured fluorescence by Typhoon (Amersham Bioscience) (Table. IV-3-3).

3.1.3. Acryl amide gel P^{-32} label

HP4U was 5' end-labeled with [γ^{-32}P]ATP using T4 polynucleotide kinase (NEB) protein-RNA binding reactions (10 μl) were incubated for 15 min at room temperature in 50 mM HEPES, 100 mM NaCl, 2 mM $MgCl_2$,1 mM DTT, 0.3 mg/ml tRNA (total) after 20 minutes at room temperature the samples were loaded on a 6 % native poly acryl amide gel at 4 degree in 0.5X TBE (Tris-borate-EDTA) buffer. The gels were run for 1 hour at 100V and dried onto Whatman 3MM chromatography paper. The activity associated with each band was quantified using a phosphoimager.

Table IV-3-1 Native EMSA condition

	Concentration
Protein	2 μM – 16 μM
RNA	4 μM
Buffer	20 mM Hpeps-NaOH pH 7.6, 100 mMNaCl,
Temp	Room temperature

Table IV-3-2 Migration buffer

1 X TBE	Tris	89 mM
	Boric Acid	89 mM
	EDTA	10 mM
1X TG	Tris	50 mM
	Glysine	80 mM

Table IV-3-3 Fluoroscence EMSA condition

	Concentration
Protein	2 μM – 64 μM
RNA	2 μM
Buffer	100 mM K-phosphate pH 7.6, 500 mMNaCl, 1 mMDTT, 1 mg/ml tRNAtotal, 5 mM $MgCl_2$, 0.1 mg/ml BSA
Temp	Room temperature

3.2. Thermal shift assay

Ligand binding to a target protein can stabilize a protein's native state, as shown in the increase of the bound protein's melting temperature. The midpoint of the melting curve of a protein will increase in the presence of ligands that bind more tightly to the native state than the unfolded state. The assay takes advantage of an environmentally sensitive fluorescence dye, such as Sypro Orange, and follows its signal changes while the protein undergoes thermal unfolding. When Sypro Orange is added to a properly folded protein solution, it is exposed in an aqueous environment and its fluorescence signal quenched. As the temperature rises, the protein undergoes thermal unfolding and exposes its hydrophobic core region. Sypro Orange then binds to the hydrophobic regions and becomes unquenched. This will result in the increase of fluorescence signal of Sypro Orange.

Protein was used just after GF column in 20 mM Hepes-NaOH pH 7.2, 200 mM NaCl, 1 mM DTT concentration was adjusted around 1 mg/ml (56 μM), RNA was diluted with 10 mM Na-cacodylate, 2 mM $MgCl_2$, 0.25 mM EDTA. The RNA was added the protein solution at molecular ratio 1:1. The samples were generally prepared with the buffer (20 mM Hepes-NaOH pH 7.0, 100 mM NaCl) However, during the "buffer optimization" other buffers were also tested. These included 20 mM Na-cacodylate, 20 mM Tris-HCl, with pH ranging from 7.0 to8.5 and including NaCl concentrations of either 0.1 or 0.25 M. Samples were pipetted by hand into 8 lane tube (maximum 24 samples are measured at the same time) (final volume: 28 μL per well). The final concentration of protein in all experiments was around 6 μM. The fluorescent dye Sypro Orange, used for tracking protein denaturizing was is sold as a stock solution of "5000X" and was added to a final concentration of "6X." RNA was added to a final concentration of 6 μM (Table. IV-3-4). 8 lane tubes were heated in 0.2 degree increments from 4 to 90 °C by Bio Rad Mini Opticon (Fig. IV-3-1).

Table IV-3-4 Thermal Shift Assay condition

		Final concentration
Buffer	20 mM Hepes pH 7.0, 100 mM NaCl	11 mM Hepes pH 7.0, 56 mM NaCl
Cypro-Orabge	200 X diluted	800 X diluted
Protein	1 mg/ml	0.1 mg/ml
RNA	Put morecular ratio 1 :1 against protein	
Temp	4 degree to 90 degree increase by 0.2 degree	

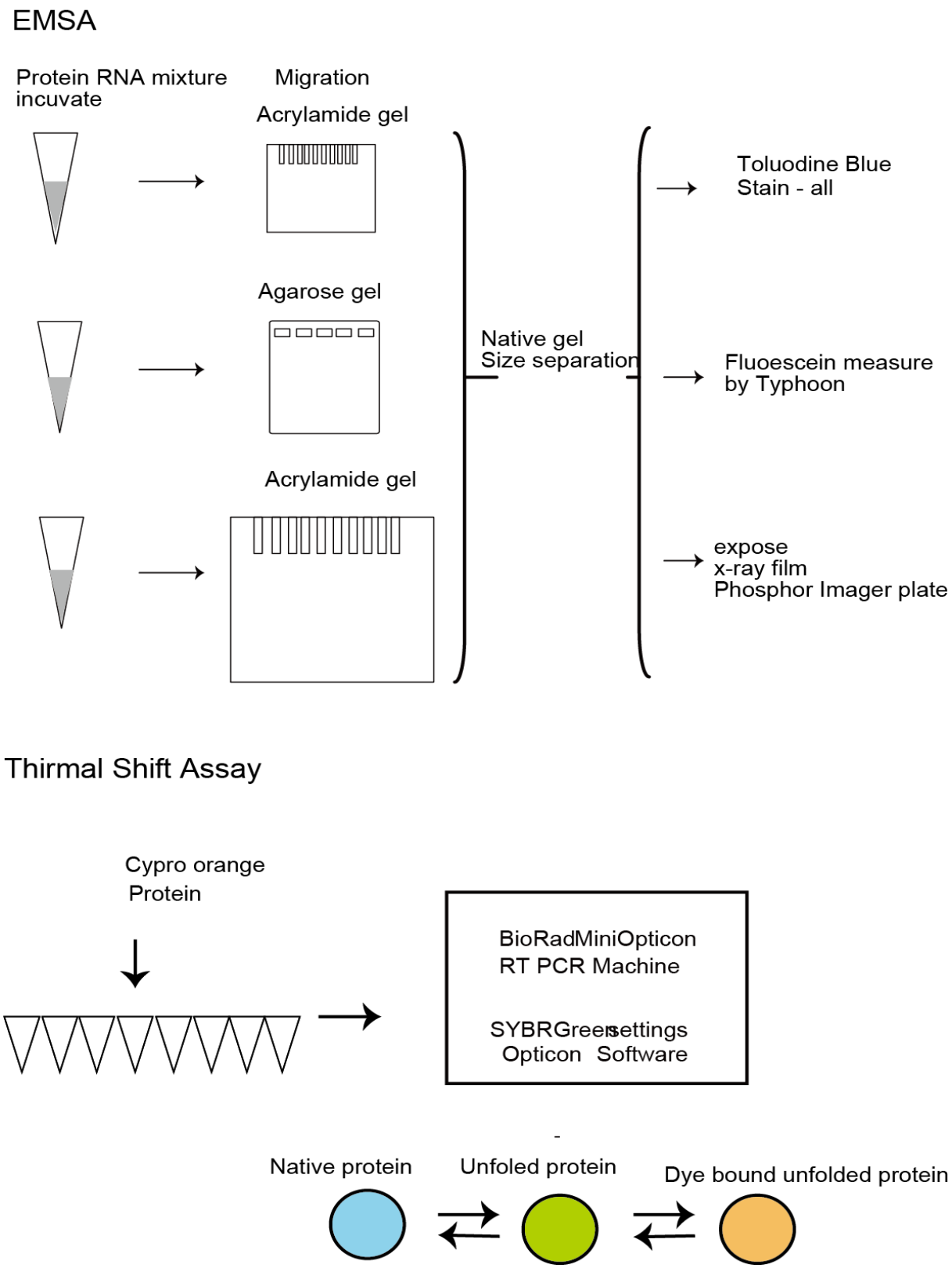

EMSA
Protein RNA mixture
incuvate
Migration
Acrylamide gel
Agarose gel
Acrylamide gel
Native gel
Size separation
Toluodine Blue
Stain - all
Fluoescein measure
by Typhoon
expose
x-ray film
Phosphor Imager plate
Thirmal Shift Assay
Cypro orange
Protein
BioRadMiniOpticon
RT PCR Machine
SYBRGreensettings
Opticon Software
Native protein
Unfoled protein
Dye bound unfolded protein

Fig. IV-3-1

3.3. ITC isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a powerful and versatile method to study the physical basis of molecular interactions. A single well-designed experiment can provide complete thermodynamic characterization of a binding reaction, including Ka, ΔG, ΔH, ΔS and reaction stoichiometry. Repeating the experiment at different temperatures allows determination of the heat capacity change (ΔCP) of the interaction.

If the reaction concentration of protein is [X] and RNA is [M]

Association constant $K = \frac{[XM]}{[X][M]}$

X total = [X] + [XM]

Mtotal= [M] + [XM]

dQ = d[XM]⊿HV

dQ is heat quantity

⊿H is mol enthalpy

V is volume

$r = \frac{1}{KM_{total}}$ $X_r = \frac{X_{total}}{M_{total}}$

⊿G = ⊿H - T⊿S = RTInk

ITC has been applied to protein studies for many years, it is becoming widely applicable in RNA biochemistry as well, especially in studies which involve RNA folding and RNA interactions with small molecules, proteins and with other RNAs.

ITC is used mainly to measure affinity for protein-protein and protein-ligand interactions. In an ITC experiment aliquots of a titrate (protein, peptide or small molecule typically at ≥ 0.5 mM) are injected into the cell containing protein solution (typically 20 to 100μM). Upon each titration the amount of heat released or absorbed is measured. In addition to the equilibrium dissociation constant (*K*d), ITC titration allows to determine the number of binding sites, binding enthalpy and entropy, ITC is not limited by the ligand or protein size. It is relatively artifacts-free, and is not affected by the optical properties of the samples. The only major disadvantage of ITC is that it requires relatively high concentrations of samples.

ITC measurements were carried out by using ITC-200 micro calorimeter (MicroCal, Northampton, MA). The protein and RNA were dialyzed in 20 mM Hepes-NaOH pH7.5, 200 mM NaCl, 1 mM TCEP, 0.25 mM EDTA. The protein was put into sample cell then the

amount of approximately 40 µl was injected into the thermally equilibrated ITC sample cell at 15 degree (Fig. IV-3-2) (Table. IV-3-5). Calculated by the data using Micro Cal Origin software the result showed in Chap II-3-5.

Table IV-3-5 ITC condition

	Concentration
Protein	100 µM
RNA	1000 µM
Buffer	20 mM Hepes-NaOH pH 7.2, 200 mM NaCl, 1 mM TECP, 0.25 mM EDTA, 2 mM $MgCl_2$
Temp	15 degree

3.4. Fluorescein Anisotropy

Fluorescence anisotropy assays the rotational diffusion of a molecule from the decorrelation of polarization in fluorescence, between the exciting and emitted (fluorescent) photons. This decorrelation can measure the "tumbling time" of the molecule as a whole, or of a part of the molecule relative to the whole. From the rotational diffusion constants, one can estimate the rough shape of a macromolecule. Fluorescence anisotropy can be used for measuring the binding interaction between two molecules, to determine the binding constant (or the inverse, the disassociation constant) for the interaction. The basic idea is that a fluorophore excited by polarized light will also emit polarized light. However, if a molecule is moving, it will tend to "scramble" the polarization of the light by radiating at a different direction from the incident light. The "scrambling" effect is greatest with fluorophores freely tumbling in solution and decreases with decreased rates of tumbling. Protein interactions can be detected when one of the interacting partners is fused to a fluorophore: upon binding of the partner molecule a larger, more stable complex is formed which will tumble more slowly (thus, increasing the polarization of the emitted light and reducing the "scrambling" effect). This technique works best if a small molecule is fused to a fluorophore and binds to a larger partner (this maximizes the difference in signal between bound and unbound states). If the fluorophore is attached to the larger protein in a binding pair, the difference in polarization between bound and unbound states will be smaller (because the unbound protein will already be fairly stable and tumble slowly to begin with) and the measurement will be less accurate. By titrating the amount of one of the proteins, a binding curve can be generated (the amount of polarization observed is proportional to the amount of protein complex formed, which is proportional to the concentration of the binding partners in solution). Mathematical models can be applied to this binding curve to determine the binding constant of the protein interaction. In another application of this technique, it is also possible to measure the folding of a protein, since an unfolded peptide chain will tumble differently than a folded one, giving a difference in polarization over time. This provides a measure of the dynamics of how the protein achieves its final, stable 3D shape.

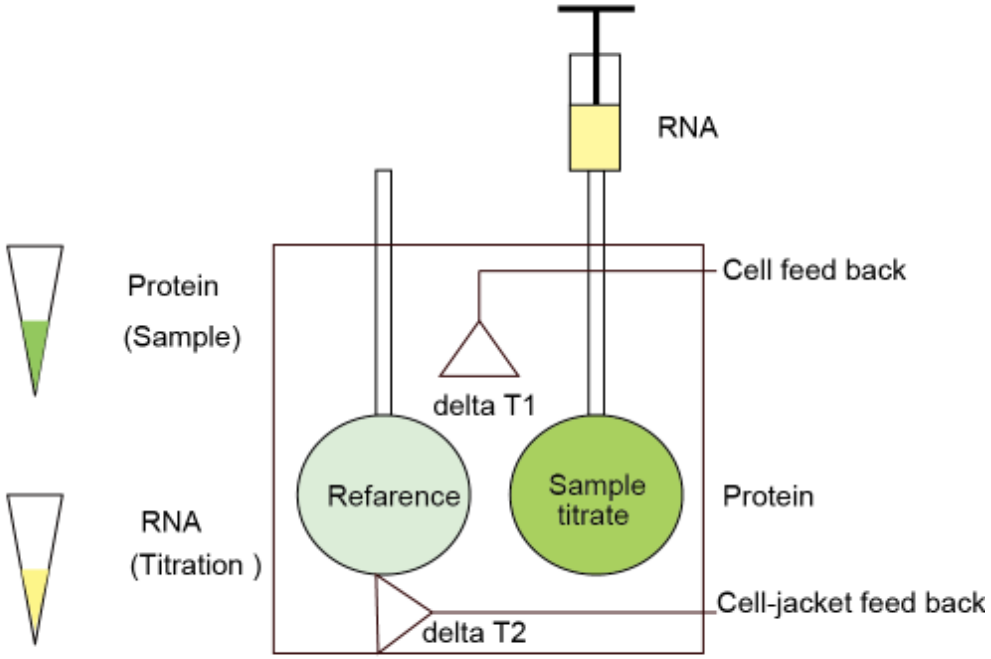
All measurements were conducted in 250 mM NaCl, 50 mM K-phosphate pH 7.5, 1 mM TCEP, 5 mM $MgCl_2$. The protein was diluted with protein buffer (250 mM NaCl, 50 mM K-phosphate pH 7.5, 1 mM TCEP, 5mM $MgCl_2$, 0.1 mM BSA) and 5’ fluorescence labeled HP4U, HP4, Oligo, were diluted with RNA buffer (250 mM NaCl, 50 mM K-phosphate pH 7.5, 1 mM TCEP, 5 mM $MgCl_2$). The final concentration of RNA was 200 nM, the protein

was change the concentrations 0 nM, 50 nM, 100 nM, 200 nM, 400 nM, 800 nM, 1600 nM 3200 nM, 5000 nM. Each protein solution prepared at 15 µl, RNA prepared 150 µl at 4 degree. 12 µl each of them mixed and transferred to 20 µl cuvettes. The measurement was done at 25 degree by PHERA star *Plus* (BMG LAB TECH) (Table. IV-3-6, Fig. IV-3-2). From the measurement, the *K*d value was obtained (See Chap. II-5-6).

Table IV-3-6 Anisotropy condition

	Concentration
HP4U(5' labeled)	Diluted with RNA buffer 200 nM
Lp7N1-208	Dilution with protein buffer (0 – 5000 nM)
RNA buffer	250 mM NaCl, 50 mM K-phosphate pH 7.5, 1 mM TCEP, 5mM $MgCl_2$
Protein buffer	250 mM NaCl, 50 mM K-phosohate pH 7.5, 1 mM TCEP, 5mM $MgCl_2$, 0.1 mM BSA

ITC



Anisotropy

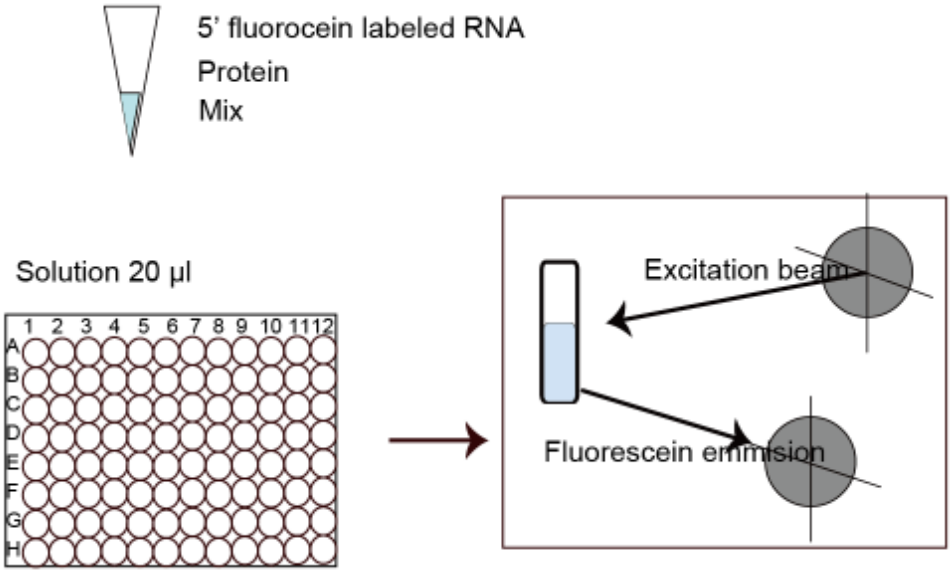


Fig. IV-3-2

3.5. Analytical gel filtration

Gel filtration chromatography separates molecular based on size. Molecules move through a bed of porous beads, diffusing into the beads to greater or lesser degrees. Smaller molecules diffuse further into the pores of the beads and therefore move through the bed more slowly, while larger molecules enter less or not at all and thus move through the bed more quickly. Both molecular weight and three-dimensional shape contribute to the degree of retention. Gel Filtration Chromatography may be used for analysis of molecular size, for separations of components in a mixture. Intended purpose of use Gel Filtration Chromatography for this experiment is to analyze the complex form of Protein and RNA complex. If the protein and RNA form a complex, the complex shows another peak compare with the peaks that separately analyzed. Gel Filtration Chromatography reflects the three dimensional shape, therefore the complex peak could not always estimate from only molecular weight.

The experiments were performed using Gel Filtration column Superdex 75 10/300 by AKTA purifier. LaRP7N1_208 50 μM, HP4U 25 μM, LaRP7N1_208 50 μM and HP4U 25 μM, Lp7N1_208 50 μM and HP4U 50 μM. All samples were prepared 100 μland incubated in 20 mM Hepes-NaOH pH7.2, 200 mM NaCl, 5 mM $MgCl_2$, 0.25 mM EDTA at 4 degree for 20 minutes (Table. IV-3-7). The chromatogram of this experiment was showed in Chap. fuII-5-7.

Table IV-3-7 Gel chromatography buffer condition

	Superdex 75 10/300 global
Column volume	24 ml
Flowrate	0.6 ml/min
Sumple	100μl (program 200μl)
Buffer	200mMNaCl, 20mM Hepes, 5mM MgCl2, 0.25mM EDTA

	Protein μM	RNA(HP4U) μM
1	50	
2		25
3	50	25
4	50	50

Tools

Website		
Sequence analysis	Pipe align	http://bips.u-strasbg.fr/PipeAlign/
Sequence source	NCBI	http://www.ncbi.nlm.nih.gov/
Sequence alignment	Clustal W	http://align.genome.jp/
Sequencing	GACT biotech	http://www.gatc-biotech.com/en/index.html
Protein sequence	NCBI	http://www.ncbi.nlm.nih.gov/
	Peptide cutter	http://expasy.org/tools/peptidecutter/
Restriction site	NEBcutter	http://tools.neb.com/NEBcutter2/
tRNA	tRNA databease	http://gtrnadb.ucsc.edu/
Modification	The RNA modification database	http://rna-mdb.cas.albany.edu/RNAmods/
Secondary structure prediction	ES pript	http://espript.ibcp.fr/ESPript/ESPript/
Prediction of 3D structure	3D-JIGSAW	http://bmm.cancerresearchuk.org/~3djigsaw/
Prediction of 3D structure	Swiss-Model	http://swissmodel.expasy.org/
Surface entropy	Surface Entropy Reduction Prediction Server	http://nihserver.mbi.ucla.edu/SER/
3D structure	Protein Data Bank	http://www.rcsb.org/pdb/home/home.do
References	Pub Med	http://www.ncbi.nlm.nih.gov/pubmed
Software		
Sequence analysis	Bio Edit	
	CLC sequence viewer	
	ClustalX	
	GeneDoc	
3D structure analysis	Jalview	
	Pymol	
	Cimera	
Reference	Mendeley	

Personal Conclusion about my PhD

During my PhD, I studied RNA-Protein binding under Prof. Michiko KONNO and Dr. Ann-Catherine DOCK-BREGEON. I realized to know the RNA-protein interaction is greatly exciting but practically there are a lot of challenges. I always straggled to produce crystals but during struggling, I was able to absorb varieties of skills. At the same time in my PhD, I learned not only research work but went through many experiences. My thesis project's results itself might not be sufficient but I have confidence that what I have learned during my PhD.

At Ochanomizu University, I learned a lot of things, Research work, teaching, applying money for research work and living, as a result, I could know one end of laboratory management and how to form the career as a researcher.

Research work: Through ArgRS project, I learned basic skill of cloning, purification, crystallization, measurement of crystals and principals and methods of analysis of 3D structure. I learned practical procedure of experiments (i.e. cloning, tRNA synthesis, aminoacylation reaction) in Dr. YOKOYAMA laboratory. The laboratory has a lot of solution of experiments from piled-up experiences, therefore I was able to access front-line research environment. When I work with them, I could spend thought-provoking time. From TrpRS project, (I did not write about this story in this thesis), I was able to improve my experiences of measuring crystals and find solutions.

Teaching: I had many chances to teach young students. When I taught new student at our laboratory, I could marshal and garner a lot of knowledge during teaching. With my laboratory members, I dealt with not only ArgRS story but also IleRS, GlyRS, TrpRS, Mov10, I could participate in many projects and gained information through these projects. When I supported lecture of Professors as assistant, I learned how to attract and represent the lectures to the students and I absorbed wide range of knowledge from those lectures.

Application for money: I tried to apply some of scholarships and research money. I had opportunities how to write and represent about my work. Each time I found problems my way of writing and progressing. But even now I indeed feel writing and presentation are needed to make more efforts. At the same time I was given many chances to participate scientific congress. I learned ways of presentation from other researchers and could know a lot of projects. For further development of my skills, I tried to apply to go abroad. I applied ITP (JSPS International Training Program/Ochanomizu University) for 11 months, burse of

Toshiko YUASA and BOURSES DU GOUVERNEMENT FRANCAIS. Fortunately, I became to be supported by these burses therefore, I could stay at France 17 months. As a part of ITP, Ochanomizu University gave me the opportunities to take lecture of English conversation, presentation and cross-cultural understanding. At the same time they introduced to me several foreign students for prepare the English communication.
As I described here, at Ochanomizu University I could study with wide visions.

At Strasbourg University, I could completely devote myself to research work and I could absorb many skills and French ways of approach for research work. This 1 year was fulfilled. Also by changing a country and environment, I was able to see and feel many things.

Research work: IGBMC (Institute of Genetics and Molecular and Cellular Biology) is one of the top research institutes in France. There is a big structure department; a lot of machines are available and easy to access many methods. I learned EMSA, Fluorescence Anisotropy, ITC, Thermal shift assay, RNA preparation, and involved the SAX, Mass and cryo-EM experiments. I was often remained of my lacking of basic techniques and approach. In addition, depends on the laboratory, the way of experiment is different. I always could compare what I have learned and new manners then considered advantages and disadvantages. This experience should help me in the future. In addition, I was able to discuss with expert researchers not only X-ray crystallography but also NMR, cryo-EM we are able to think many possibilities. Once I learned and knew techniques, I could organize and have more choice for research work. In IGBMC, every week I had opportunities to listen seminars by top researchers from the institute or all over the world. From these seminars, I was able to discover new horizon. At Strasbourg University, I could take some lectures especially the course of ‘’effective writing’’ was stimulating I realized my English level was far from average of students and researchers I learned way of writing that also changed my way of reading.

Other experiences: As a matter of course I devote research work, by changing a country, my environment was dramatically changed. I was able to improve my English skills and among people come from different counties, I could know varieties of sense of worth and cultures. I can’t help but feel that changing environment makes us developed.

Fortunately I have chance to continue my thesis project more 10 months (support by JSPS Institutional Program for Young Researcher Overseas Visits/Ochanomizu University) at IGBMC. I would like to give some more contribution for understanding protein: RNA binding.

Future project: What should I challenge to be addressed.

The 3D structure of LaRP7N1_208 HP4U complex has to be solved. For the moment, the crystals were not obtained without fail. It needs to establish the 100% reproducible condition. For this, many problems need to be overcome. Now I planning 1) Change the RNA construct to avoid flexibility from RNA. 2) My concentration way for crystallization was difficult to follow accurately. For this ambiguity, the sample will be measured with new DLS (5 µl for 1 measurement). For the moment, almost all conditions for biochemical experiments were fixed. To detect the interaction part between 7SK and LaRP7 using LaRP7 full length, LaRP7 C terminal region, 7SK full length, each hairpin region of 7SK, HP4U several constructs by competition assay using EMSA and ITC measurement.

References

Agris, P. F. (2004). Decoding the genome: a modified view. Nucleic. Acids. Research ***32***, 223-38.

Alan H Mehler, S. K. M. (1967). The Activation of Arginyl Transfer Synthetase by Transfer Ribonucleic Ribonucleic Acid. J.biol. Chem. ***242,*** 5495-5499

Alfano, C., Sanfelice, D., Babon, Jeff, Kelly, Geoff, Jacks, A., Curry, Stephen, and Conte, M. R. (2004). Structural analysis of cooperative RNA binding by the La motif and central RRM domain of human La protein. Nat. Struct. Mol. Biol. ***11***, 323-9.

Amez, J. G., and Morns, D. (1997). Structural and functional considerations of the aminoacylation reaction. Trends. Biochem. Sci. ***22,***211-16

Arcus, V. (2002). OB-fold domains: a snapshot of the evolution of sequence, structure and function. Curr. Opin. Struct. Biol. ***12***, 794-801.

Belisova, A., Semrad, K., Mayer, O., Kocian, G., Waigmann, E., Schroeder, R., and Steiner, G. (2005). RNA chaperone activity of protein components of human Ro RNPs. RNA ***11***, 1084-94.

Boeing, S., Rigault, C., Heidemann, M., Eick, D., and Meisterernst, M. (2010). RNA polymerase II C-terminal heptarepeat domain Ser-7 phosphorylation is established in a mediator-dependent fashion. J. Biol. Chem. ***285***, 188-96.

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Mol. Cell. ***36***, 541-6.

Cavarelli, J (2003). Pushing Induced Fit to Its LimitstRNA-Dependent Active Site Assembly in Class I Aminoacyl-tRNA Synthetases. Structure ***11***, 484-486.

Cavarelli, J., Eriani, G., Rees, B., Ruff, M., Boeglin, M., Mitschler, A., Martin, F., Gangloff, J., Thierry, J C., Moras, D. (1994). The active site of yeast aspartyl-tRNA synthetase: structural and functional aspects of the aminoacylation reaction. EMBO J. ***13*****,** 327-37.

Cavarelli, J., Eriani, G., and Gangloff, J. (1998). recognition by yeast arginyl-tRNA synthetase L -Arginine. EMBO J. ***17***, 5438-5448.

Cléry, A., Blatter, M., and Allain, F. H.-T. (2008). RNA recognition motifs: boring? Not quite. Current opinion in structural biology ***18***, 290-8.

Copela, L. a, Chakshusmathi, G., Sherrer, R. L., and Wolin, Sandra L (2006). The La protein functions redundantly with tRNA modification enzymes to ensure tRNA structural stability. RNA. ***12***, 644-54.

Cusack S, Yaremchuk A, Tukalo M (2000) The 2Å crystal structure of leucyl-tRNA synthetase and its complex with a leucyl-adeny- late analogue. EMBO J. ***19***, 2351–2361

Dock-Bregeon, A-C., Rees, B., Torres-Larios, A., Bey, G., Cailet, J., Moras, D. (2004) Achieving error-free translation: the mechanism of proof- reading of threonyl-tRNA synthetase at atomic resolution. Mol. Cell. ***16***, 375–386

Dock-Bregeon, A-C., Sankaranarayanan, R., Romby, P., Cailet, J., Springer, M., Rees, B., Francklyn, CS., Ehresmann, C., Moras., D. (2000) Transfer RNA-mediated editing in threonyl-tRNA synthetase: The class II solution to the double discrimination problem. Cell ***103***, 877–884

Delagoutte, B., Moras, D., and Cavarelli, J (2000). tRNA aminoacylation by arginyl-tRNA

synthetase: induced conformations during substrates binding. EMBO J. ***19***, 5599-610.

Eriani, G, Dirheimer, G, and Gangloff, J (1990). Aspartyl-tRNA synthetase from Escherichia coli: cloning and characterisation of the gene, homologies of its translated amino acid sequence with asparaginyl- and lysyl-tRNA synthetases. Nucleic. Acids. Research ***18***, 7109-18.

Eriani, G., Cavarelli, J., Martin, F., Dirheimer, G., Moras, D., Gangloff, J. (1993) Role of dimerization in yeast aspartyl-tRNA synthetase and importance of the class II invariant proline. Proc. Natl. Acad. Sci. USA ***90***, 10816-22.-

Fukai, S., Nureki, O., Sekine, S., Shimada, A. (2003). Mechanism of molecular interactions for tRNA Val recognition by valyl-tRNA synthetase Mechanism of molecular interactions for tRNA Val recognition by valyl-tRNA synthetase. RNA. ***9,*** 100-111.

Fukunaga, R., and Yokoyama, S. (2005). Aminoacylation complex structures of leucyl-tRNA synthetase and tRNALeu reveal two modes of discriminator-base recognition. Nat. Struct. Mol. Biol. ***12***, 915-22.

Gabriel, K., Schneider, J., and McClain, W. H. (1996). Functional Evidence for Indirect Recognition of G:U in trai[1] by Alanyl-tRNA Synthetase. Science ***271***, 195-197.

Gangloff, J., Schutz, A., and Dirheimer, G. (1976). Arginyl-tRNA Synthetase from Baker's Yeast. Purification and Some Properties. European Journal of Biochemistry *65*, 177-182.

Godeau, J. M. (1980). Arginyl-transfer ribonucleic acid synthetase of Bacillus stearothermophilus. Purification and kinetic analysis. European journal of biochemistry FEBS J. ***103***, 169-77.

Guigou, L., and Mirande, M. (2005). Determinants in tRNA for activation of arginyl-tRNA synthetase: evidence that tRNA flexibility is required for the induced-fit mechanism. Biochemistry. ***44***, 16540-8.

Guigou, L., Shalak, V., and Mirande, M. (2004). The tRNA-interacting factor p43 associates with mammalian arginyl-tRNA synthetase but does not modify its tRNA aminoacylation properties. Biochemistry ***43***, 4592-600.

Guo, M., Yang, X.-L., and Schimmel, P. (2010). New functions of aminoacyl-tRNA synthetases beyond translation. Nature reviews. Mol. Cell. Biol. ***11***, 668-674.

Hauenstein, S., Zhang, C.-M., Hou, Y.-M., and Perona, J. J. (2004). Shape-selective RNA recognition by cysteinyl-tRNA synthetase. Nat. Struct. Mol. Biol. ***11***, 1134-41.

He, N., Jahchan, N. S., Hong, E., Li, Q., Bayfield, M. A., Maraia, R. J., Luo, K., and Zhou, Q. (2008). A La-related protein modulates 7SK snRNP integrity to suppress P-TEFb-dependent transcriptional elongation and tumorigenesis. Mol. Cell. ***29***, 588-99.

Hussain, T., Kruparani, S. P., Pal, B., Dock-Bregeon, A.-C., Dwivedi, S., Shekar, M. R., Sureshbabu, K., and Sankaranarayanan, R. (2006). Post-transfer editing mechanism of a D-aminoacyl-tRNA deacylase-like domain in threonyl-tRNA synthetase from archaea. EMBO J. *25*, 4152-62.

Ibba, M, and Soll, D. (2000). Aminoacyl-tRNA synthesis. Annu. Rev. Biochem. ***69***, 617-50.

Inada, M., and Guthrie, C. (2004). Identification of Lhp1p-associated RNAs by microarray analysis in Saccharomyces cerevisiae reveals association with coding and noncoding RNAs. Proc. Natl. Acad. Sci. USA. ***101***, 434-9.

Jacks,A., Babon, J., Kelly, G., Manolaridis, I., Cary, P., Curry, S, and Conte, M. (2003). Structure of the C-Terminal Domain of Human La Protein Reveals a Novel RNA

Recognition Motif Coupled to a Helical Nuclear Retention Element. Structure ***11***, 833-843.

Korencic, D., Ahel, I., Schelert, J., Sacher, M., Ruan, B., Stathopoulos, C., Blum, P., Ibba, Michael, and Söll, Dieter (2004). A freestanding proofreading domain is required for protein synthesis quality control in Archaea. Proc. Natl. Acad. Sci. USA. ***101***, 10260-5.

Kotik-Kogan, O., Valentine, E. R., Sanfelice, D., Conte, M. R., and Curry, Stephen (2008). Structural analysis reveals conformational plasticity in the recognition of RNA 3' ends by the human La protein. Structure. ***16***, 852-62.

Krueger, B. J., Jeronimo C., Roy, B. B., Bouchard, A., Barrandon, C., Byers, S. A., Searcey, C. E., Cooper, J. J., Bensaude, O., Cohen, E. A., Coulombe, B and Price, D. H. (2008). LARP7 is a stable component of the 7SK snRNP while P-TEFb, HEXIM1 and hnRNP A1 are reversibly associated. Nucleic. Acids. Research *36*, 2219-29.

Kufel, J., Allmang, C., Chanfreau, G., Petfalski, E., Lafontaine, D. L., and Tollervey, D. (2000). Precursors to the U3 small nucleolar RNA lack small nucleolar RNP proteins but are stabilized by La binding. Mol. Cell. Boil. ***20***, 5415-24.

Lazard, M., Agou, F., Kerjan, P., and Mirande, M (2000). The tRNA-dependent activation of arginine by arginyl-tRNA synthetase requires inter-domain communication. J. Mol. Biol. ***302***, 991-1004.

Lui, M., Chakraburtty, K., and Mehler, a H. (1978). Partial reactions of aminoacyl-tRNA synthetases as functions of pH. J. Biol. Chem. ***253***, 8061-4.

Lunde, B. M., Hörner, M., and Meinhart, A. (2010). Structural insights into cis element recognition of non-polyadenylated RNAs by the Nab3-RRM. Nucleic. Acids. Research. ***39***, 1-10.

Maraia, R. J., and Bayfield, M. A. (2006). The La protein-RNA complex surfaces. Mol.Cell. ***21***, 149-52.

Markert, A., Grimm, M., Martinez, J., Wiesner, J., Meyerhans, A., Meyuhas, O., Sickmann, A., and Fischer, U. (2008). The La-related protein LARP7 is a component of the 7SK ribonucleoprotein and affects transcription of cellular and viral polymerase II genes. EMBO Rep. ***9***, 569-75.

McClain, W H, Foss, K., Jenkins, R. a, and Schneider, J. (1990). Nucleotides that determine Escherichia coli tRNA(Arg) and tRNA(Lys) acceptor identities revealed by analyses of mutant opal and amber suppressor tRNAs. Proc. Natl. Acad. Sci. USA. ***87***, 9260-4.

Mcclain, W. H., Foss, K., and I, M. A. Y. (1988). Changing the Identity of a tRNA by Introducing a G-U Wobble Pair Near the 3' Acceptor End. Science ***240***, 793-96.

Mitra, K., and Mehler, a H. (1966). The role of transfer ribonucleic acid in the pyrophsphate exchange reaction of arginine-transfer ribonucleic acid synthetase. J. Biol. Chem. ***241***, 5161-2.

Nakanishi, K., Ogiso, Y., Nakama, T., Fukai, S., and Nureki, O. (2005). Structural basis for anticodon recognition by methionyl-tRNA synthetase. Nat. Struct. Mol. Biol. ***12***, 931-2.

Nguyen, V.T., Kiss, T., Michels, A.A., and Bensaude, O. (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature ***414***, 322–325.

Perret,V., Florentz,C., Puglisi,J.D. and Giege,R. (1992) Effect of conformational features on the aminoacylation of tRNAs and consequences on the permutation of tRNA specificities. J. Mol. Biol. ***226***, 323-333

Pütz, J., Dupuis, B., Sissler, Marie, and Florentz, Catherine (2007). Mamit-tRNA, a database of mammalian mitochondrial tRNA primary and secondary structures. RNA. ***13***, 1184-90.

Quevillon, S., Robinson, J. C., Berthonneau, E., Siatecka, M., and Mirande, M (1999). Macromolecular assemblage of aminoacyl-tRNA synthetases: identification of protein-protein interactions and characterization of a core protein. J. Mol. Biol. ***285***, 183-95.

Ramos, A, and Varani, G. (1997). Structure of the acceptor stem of Escherichia coli tRNA Ala: role of the G3.U70 base pair in synthetase recognition. Nucleic Acids Research. ***25***, 2083-90.

Rao, S. T., and Rossmann, M. G. (1973). Comparison of super-secondary structures in proteins. J. Mol. Biol. ***76***, 241-56.

Rho, S. B., Kim, M. J., Lee, J. S., Seol, W. G., Motegi, H., Kim, S., and Shiba, K. (1999) Genetic dissection of protein-protein interactions in multi-tRNA synthetase complex, Proc. Natl. Acad. Sci. U.S.A. ***96***, 4488-4493.

Rossmann, M. G., Moras, D. & Olsen, K. W. (1974).Chemical and biological evolution of a nucleotide-binding protein. Nature, ***250,*** 194-199.

Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C., and Moras, D. (1991). Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). Science ***252***, 1682-1689.

Saks, M. E. (1998). Evolution of a Transfer RNA Gene Through a Point Mutation in the Anticodon. Science ***279***, 1665-1670.

Saunders, A., Core, L.J., and Lis, J.T. (2006). Nat. Rev. Mol. Cell Biol. *7*, 557– 567.

Schulman, L. H., and Pelka, H. (1989a). The anticodon contains a major element of the identity of arginine transfer RNAs. Science. ***246***, 1595-1597.

Sequential, A. (1974). Physical and Kinetic Studies of Arginyl Ribonucleic Acid Ligase of Neurospora. Enzyme. ***249***, 4934-4942.

Sankaranarayanan, R., Dock-Bregeon, A.C., Romby, P., Caillet, J., Springer, M., Rees, B., Ehresmann, C., Ehresmann, B., and Moras, D. (1999). The structure of threonyl-tRNA synthetase- tRNAThr enlightens its repressor activity and reveals an essential zinc ion in the active site. Cell. ***97***, 371–381

Sankaranarayanan R, Dock-Bregeon A-C, Rees B, Bovee M, Cailet J, Romby P, Francklyn CS, Moras D (2000) Zinc ion mediated amino acid discrimination by threonyl-tRNA synthetase. Nat Struct Biol. *7*,461–465

Sekine, S., Dmitry G., Vassylyev, D. G., Nureki, O., Shimada, A. and Yokoyama, S. (2001). Structural basis for anticodon recognition by discriminating glutamyl- tRNA synthetase. Nat Struct Biol. ***8***, 203-206.

Shimizu, S., Margat Juan, E, C., Sato, Y., Miyashita, Y., Hoque, M., Suzuki, K., Sagara T., Tsunoda, M., Sekiguchi T., Dock-Bregeon A-C., Moras, D., Takenaka, A. (2009). Two complementary enzymes for threonylation of tRNA in crenarchaeota: crystal structure of Aeropyrum pernix threonyl-tRNA synthetase lacking a cis-editing domain. J. Mol. Biol. ***394***, 286-296.

Shimizu, S., Ohki, M., Okubo, N., Suzuki, K., Tsunoda, M., Sekiguchi, T., and Takénaka, A. (2009). Crystallization and preliminary crystallographic studies of putative RNA

3'-terminal phosphate cyclase from the crenarchaeota Sulfolobus tokodaii. Acta crystallographica. Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun. ***65***, 565-570.

Simos, G., Sauer, a, Fasiolo, F., and Hurt, E. C. (1998). A conserved domain within Arc1p delivers tRNA to aminoacyl-tRNA synthetases. Mol. Cell. ***1***, 235-42.

Sissler, M, Giegé, R., and Florentz, C (1996). Arginine aminoacylation identity is context-dependent and ensured by alternate recognition sets in the anticodon loop of accepting tRNA transcripts. EMBO J. ***15*****,** 5069-76.

Sissler, M, Giegé, R., and Florentz, C (1998). The RNA sequence context defines the mechanistic routes by which yeast arginyl-tRNA synthetase charges tRNA. RNA. ***4***, 647-657.

Stone, M. D., Mihalusova, M., Oconnor, C. M., Prathapam, R., Collins, K., and Zhuang, X. (2007). Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. Nature ***446***, 458-61.

Szymański, M., Deniziak, M., and Barciszewski, J. (2000). The new aspects of aminoacyl-tRNA synthetases. Acta biochimica Polonica ***47***, 821-34.

Tamura, K., Himeno, H., Asahara, H., Hasegawa, T., and Shimizu, M. (1992). In vitro study of E.coli tRNA(Arg) and tRNA(Lys) identity elements. Nucleic. Acids. Research. *20*, 2335-2339.

Teplova, M., Yuan, Y.R., Phan, A.T., Malinina, L., Ilin, S., Teplov, A., and Patel, D.J. 2006. Structural basis for recognition and seques- tration of UUU(OH) 39 temini of nascent RNA polymerase III transcripts by La, a rheumatic disease autoantigen. Mol. Cell. ***21***: 75–85.

Tukalo, M., Yaremchuk, A., Fukunaga, R., Yokoyama, S., and Cusack, S. (2005). The crystal structure of leucyl-tRNA synthetase complexed with tRNALeu in the post-transfer-editing conformation. Nat. Struct. Mol. Biol. ***12***, 923-30.

Wright.D.J., Martinis,S.A.. Jahn,M., Soll,D. and Schimmel,P. (1993). Acceptor stem and anticodon RNA hairpin helix interactions with glutamine tRNA synthetase. Biochimie. ***75***. 1041-1049.

Woese, C. R., Olsen, G. J., Ibba, M, and Söll, D (2000). Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. Microbiology and molecular biology reviews : MMBR ***64***, 202-36.

Xue, D., Rubinson, D. a, Pannone, B. K., Yoo, C. J., and Wolin, S L (2000). U snRNP assembly in yeast involves the La protein. EMBO J. ***19***, 1650-60.

Xue,Y., Yang,Z., Chen,R., Zhou,Q. (2010) A capping-independent function of MePCE in stabilizing 7SK snRNA and facilitating the assembly of 7SK snRNP, Nucleic. Acids. Res. ***38***,360–369.

Yang, Z., Zhu, Q., Luo, K., and Zhou, Q. (2001). The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. Nature. ***414***, 317–322

Zheng, Y.-G., Wei, H., Ling, C., Xu, M.-G., and Wang, E.-D. (2006). Two forms of human cytoplasmic arginyl-tRNA synthetase produced from two translation initiations by a single mRNA. Biochemistry. ***45***, 1338-1344.

Summary

In the cell, RNA-protein interactions play a key role and are fundamental to many processes in transcription. The high variability and conformational flexibility of RNA structure creates a number of unique binding sites and the potential for complex regulation by RNA binding proteins. This manuscript describes our endeavor to reveal the details of RNA-protein interactions in two different systems transcription and translation.

1 Accurate recognition and aminoacylation of the cognate tRNA by arginyl-tRNA synthetase (ArgRS)

Translation of the genetic code depends on a molecular mediation which establishes the correspondence between codon and amino acid. $tRNA^{Arg}$ recognition of ArgRS the importance of C35 evidenced by aminoacylation experiments. However in *E. coli* and *T.thermophilus* ArgRSs, also evidenced that residue 20, in the D-loop of $tRNA^{Arg}$, A20 in most cases, is also a $tRNA^{Arg}$ determinant. On the 3D structure of *S.cerevisiae* $tRNA^{Arg}$(ICG), Dhu 20 is positioned in close proximity to the side chains included in the characteristic N-terminal domain of ArgRS. However, the aminoacylation reaction values for $tRNA^{Arg}$(ICG) on the mutant of the interacted residues are the same as those on the wild-type ArgRS. In order to clarify whether or not the binding of the D-loop of $tRNA^{Arg}$(CCU) to the N-terminal domain contributes to the activation effect of tRNA on tRNA assisted Arg-AMP formation reaction or the aminoacylation reaction. We constructed *P.horikoshii* ArgRS lacking the core region of the N-terminal domain(□N ArgRS). *K*m measurements indicate that the fixing of the D-loop of $tRNA^{Arg}$ with the N-terminal domain makes a minor contribution to the aminoacylation reaction. The proper acceptance of C35 and U36 of $tRNA^{Arg}$(CCU) on the plausible accepting structures may be predominantly contributory to the aminoacylation reaction of $tRNA^{Arg}$(CCU) on *P.horikoshii* ArgRS. It was reported that the ATP–pyrophosphate exchange reaction catalyzed by ArgRS, GlnRS and GluRS requires the assistance of the cognate tRNA. On the basis of newly obtained structural information (*P.horikoshii* ArgRS, $tRNA^{Arg}$(CCU) and an ATP analog) I constructed a structural model for a mechanism and explained in which the formation of a hydrogen bond between the 2'-OH group o f A76 of tRNA and the carboxyl group of Arg induces both formation of Arg-AMP and pyrophosphorolysis of Arg-AMP.

2 Stabilisation of the human 7SK regulator of transcription elongation by the La-related protein LaRP7

7SK is an abundant and stable RNA found in the nucleus of human cells that was identified as a regulator of transcription. 7SK binds to and inhibit a transcription factor, the Positive Transcriptional Elongation Factor (PTEFb) via a HEXIM protein. PTEFb phosphorylates NELF and DSIF, two factors involved in transcription pausing, and phosphorylates Ser2 of the repeats of the C-terminal domain of the largest subunit of RNA polymerase II (RNAPII), thus switching RNAPII to its possessive elongating form. It has been shown recently that 7SK belongs to different 7SKsnRNPs corresponding to different status of transcription elongation, with-7SK bound to proteins of the hnRNP family when not bound to PTEFb. Whichever its functional state 7SK binds LaRP7 and MePCE, its capping enzyme. LaRP7 is thought to maintain steady-state level of 7SK, by an unknown mechanism. LaRP7 shows homology to La protein as it possesses a LAM and a RRM domain at its N-terminal region, in addition to a second RRM in the C-terminal region. La proteins are involved in protecting the U-rich 3'end of RNAs produced by RNA polymerase III. However, the documented specificity of LaRP7 for 7SK cannot be explained by the sole binding to the 3'UUUU sequence of 7SK. With the aim to clarify at the molecular level how 7SK is stabilized by LaRP7, we studied the specificity and affinity of LaRP7 for 7SK using truncated 7SK and LaRP7 (HP4U and LaRP7N1_208). The binding affinity of full-length LaRP7 is stronger than LaRP7N1_208, which indicates that the second RRM at the C-terminal region also contribute to bind 7SK. On the other hand, I could show that if the binding affinity of LaRP7 and LaRP7N1_208 for HP4 is significantly reduced compared with HP4U, as expected from the hypothesis that LaRP7 has a similar recognition mode than La for the 3'Us, the affinity for UUUCUUUU is much reduced, thus indicating that some determinant for LaRP7 lies in the hairpin 4 of 7SK. In order to clarify the details of the binding, we attempted to crystallize the complex of the LaRP7N1_208 and HP4U. Crystals were obtained in several conditions but had a tendency to melt. They showed encouraging diffraction to 7.0 Å, but need optimization.