Contributions to the study of the architecture and evolution of ribozymes

Mélanie Meyer

UNIVERSITÉ DE STRASBOURG

Ecole Doctorale des Sciences de la Vie et de la Santé

THÈSE

présentée par :

Mélanie MEYER

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline : Sciences du Vivant

Spécialité : Biochimie, Biologie Moléculaire et Structurale

CONTRIBUTIONS TO THE STUDY OF THE ARCHITECTURE AND EVOLUTION OF RIBOZYMES

Soutenue le **13 Septembre 2013** devant la commission d'examen :

Dr. MASQUIDA Benoît	Directeur de thèse
Pr. THORE Stephane	Rapporteur externe
Dr YOSHIZAWA Satoko	Rapporteur externe
Pr CAVARELLI Jean	Examinateur
Dr. SARGUEIL Bruno	Examinateur
Pr. WESTHOF Eric	Examinateur

*First and foremost, I want to thank all the members of my jury, **Satoko Yoshizawa**, **Stéphane Thore**, **Bruno Sargueil**, **Jean Cavarelli** and **Eric Westhof**, who agreed to judge my thesis.*

*Tout d'abord je souhaite remercier les membres de mon jury, **Satoko Yoshizawa**, **Stéphane Thore**, **Bruno Sargueil**, **Jean Cavarelli** et **Eric Westhof**, qui ont accepté de juger mes travaux de thèse.*

*Mes sincères remerciements sont naturellement adressés à **Benoît Masquida**, mon directeur de thèse, qui a su en trois années me transmettre son intérêt pour la recherche, son expérience du microcosme de la science et bien sur ses connaissances en cristallographie et analyse des structures tridimensionnelles d'ARN. Mes meilleurs souvenirs seront sans aucun doute nos expéditions aux synchrotrons. Merci beaucoup !*

*Je tiens à remercier **Eric Westhof** pour m'avoir accueilli au sein de son équipe et nous avoir permis de poursuivre le projet du Lariat-Capping ribozyme mais aussi pour m'avoir permis de participer à deux éditions du RNA meeting.*

*Je remercie également **Ivan Tarassov** et **Nina Entelis** qui nous ont très chaleureusement accueilli et très vite intégré à la vie de leur laboratoire au sein duquel j'ai rafraîchi mes connaissances sur la mitochondrie et je me suis initiée au russe.*

*Merci à **Hubert Becker**, pour ses conseils avisés avant et pendant ma thèse mais aussi pour les pauses café ! Merci à toute son équipe pour les discussions enrichissantes que nous avons partagés et tous les p'tits coups de pouces logistiques.*

*Merci à **Catherine Florentz** qui à su me guider au moment de choisir ma thèse et m'a motivée pour prendre part à la vie de l'école doctorale des sciences de la vie et de la santé. Cette expérience a été très enrichissante.*

*Merci à **Laurence Drouard** et **Michel Labouesse** pour m'avoir offert la possibilité de participer à la mission de diffusion de l'information scientifique et technique « OpenLAB » deux années consécutives. Merci à **Joern Pütz** de m'avoir permis de faire un monitorat en Biochimie. Merci aux « Openlabistes » et moniteurs qui ont rendus ces deux expériences formidables.*

*Merci à **Vincent Olieric** pour les nombreux créneaux sur la ligne PXIII au SLS durant lesquels ils nous a accompagnés et a partagé son savoir. Merci à **Pierre Roblin** pour son accompagnement et son expertise lors des expériences de SAXS sur la ligne SWING de SOLEIL. Merci **à Claude Sauter**, **Pierre Legrand**, **Andrew Thompson**, **Alastair McEwen**, **Pierre Poussin-Courmontagne** et **Jean Cavarelli.** Tous ont contribué aux résultats présentés dans cette thèse.*

*Merci à **Stéphane Bellemin-Laponaz** pour la synthèse et le dont d'iridium hexamine ainsi qu'à **Marat Yusupov** pour l'osmium hexamine qu'il nous a offert.*

*Merci à notre collaborateur, **Henrik Nielsen**, qui travaille depuis le début sur le projet du Lariat-Capping ribozyme et qui a produit les données de mutagénèse évoquées dans ce manuscrit. Merci aussi à **Steinar Johansen** qui, le premier, a observé le Lariat-Capping ribozyme.*

*Merci à **Pascale Legault** qui a encadré mon stage de Master 2 à l'université de Montréal lors duquel je me suis initiée aux techniques de production et de purification d'ARN à grande échelle et ai appris les rudiments de la RMN.*

*A l'IBMC je souhaite remercier toutes les personnes que j'ai côtoyées et avec qui j'ai travaillées pendant ces trois années. Merci pour vos conseils, vos coups de main, votre soutien. Merci à **Laure** et **Delphine** pour leur aide technique.*

A l'IPCB je souhaite remercier toutes les personnes du 4ème étage pour les moments de détente et de convivialité que vous avez organisé. Merci aussi pour votre aide lors du déménagement. Je n'ai passé que quelques mois avec vous mais j'en garderai un très bon souvenir.

*Un grand merci à mes collègues **doctorants** avec qui j'ai partagé les hauts et les bas des manips, les congrès, les pique-niques, la rédaction…*

*Je souhaite remercier mes amis, **Anne**, **Benoît**, **Cécile**, **Delphine**, **Emilie**, **Franzi**, **Gilles**, **Ludovic**, **Marie**, **Morgan**, **Vincent** vous avez tous, chacun à votre manière contribués au bon déroulement de cette thèse.*

*Enfin, vous arrivez en dernier, mais nous n'êtes pas les moins important, merci à tous les membres de ma famille. Plus particulièrement à **mes parents** qui m'ont soutenus tout au long de ses dix années d'études supérieures et m'ont permis de partir à Leeds et à Montréal ! Merci à toi **Anne-Laure** ma petite sœur pour tous les moments de complicité que nous partageons. Merci à toi **Alexandre** qui contribue quotidiennement à mon épanouissement.*

J'espère n'avoir oublié personne et si tel n'est pas le cas alors veuillez m'en excusez, mon étourderie est a blâmer.

TABLE DES MATIERE

ABBREVIATIONS

2D: Two-dimensional Structure
3D: Three-dimensional Structure
3WJ: Three-way junction
ACE: 2′-bis(2-Acetoxyethoxy)methyl
AlLCrz: Lariat Capping ribozyme from *Allovahlkampfia sp*
ATP: Adenosine 5'-triphosphate
BP: Binding Protein
BSA: Bovine Serum Albumin
cp: Circularly permutated
CTP: Cytidine 5'-triphosphate
DiLCrz: Lariat Capping ribozyme from *Didymium iridis*
DMSO: Dimethyl Sulfoxide
DNA: Deoxyribonucleic Acid
dsDNA: double stranded Deoxyribonucleic Acid
DTT: Dithiothreitol
EDTA: Ethylenediaminetetraacetic Acid
FLC: Full Length Circle
FPLC: Fast Protein Liquid Chromatography
FRET: Förster Resonance Energy Transfer
GIR1: Group I-like Ribozyme 1
GIR2: Group I-like Ribozyme 2
GTP: Guanosine 5'-triphosphate
HDV: Hepatitis Delta Virus
HE: Homing **E**ndonuclease
HEG: Homing Endonuclease Gene
HMH: Hammerhead
HPLC: High-Performance Liquid Chromatography
IEP: Intron Encoded Protein
IGS: Internal Guide Sequence
IPS: Internal Processing Site
kb: kilobase
LB: Luria Bertani broth
LCrz: Lariat Capping ribozyme
LUCA: Last Universal Common Ancestor
LW: Leontis-Westhof
MAD: Multi-wavelength anomalous dispersion
MR: Molecular replacement
mRNA: messenger Ribonucleic Acid
mtRNA: mitochondrial Ribonucleic Acid

NaLCrz: Lariat Capping ribozyme from *Naegleria sp*
ncRNA: non-coding Ribonucleic Acid
NDB: Nucleic acid Database
nt: nucleotide
OD: Optical Density
ORF: Open Reading Frame
PAGE: Polyacrylamide Gel Electrophoresis
PDB: Protein Data Bank
PEG: Polyethylene Glycol
PEGMME: Polyethylene Glycol Monomethyl Ether
PEI: Polyethyleneimine
PMSF: Phenylmethanesulfonyl fluoride
pre-mRNA: pre-messenger Ribonucleic Acid
pre-rRNA: pre-ribosomal Ribonucleic Acid
pre-tRNA: pre-transfert Ribonucleic Acid
PRORP: Proteinaceous Ribonuclease P
RNA: Ribonucleic Acid
RNase P: Ribonuclease P
RNP: Ribonucleoprotein
rDNA: Ribosomal Deoxyribonucleic Acid
rRNA: Ribosomal Ribonucleic Acid
RT: Reverse Transcription
SDS: Sodium Dodecyl Sulfate
SELEX: Systematic Evolution of Ligands by Exponential enrichment
siRNA: Small Interfering Ribonucleic Acid
S_N2: Bimolecular Substitution Nucleophilic
snoRNA: Small Nucleolar Ribonucleic Acid
SS: Splice Site
SSU: Small Sub-unit
TBE: Tris-Borate-EDTA
TEMED: N,N,N',N'-Tetramethylethylenediamin
TL: Tetraloop
TLR: Tetraloop and Receptor
tmRNA: Tranfert Messenger Ribonucleic Acid
TMSDT: Tris, Magnesium, Spermidine, Dithiothreitol, Triton
TR: Tetraloop Receptor
tRNA: Tranfert Ribonucleic Acid
TTP: Thymidine 5'-triphosphate
UTP: Uridine 5'-triphosphate
UTR: Untranslated Region
VS: Varuk Satelite
WC: Watson-Crick

INTRODUCTION

I. INTRODUCTION

Ribonucleic acid is a large family of biological molecules essential to most processes of life. RNA is with DNA and protein one prop on the central dogma of molecular biology. The continuously increasing number of data on genomics, transcriptomics and proteomics indicates that if most eukaryotic genomes are nearly 100% transcribed only about 2 to 10% is translated into protein (Jackson et al, 2000). Hence, most of the RNA in cells is non-coding RNA. This ncRNA has been considered as junk for a long time but now we know that by various mechanisms it can subtly tune gene expression.

In 1989 the Nobel Price in Chemistry was awarded to Thomas R Cech and Sydney Altman for their discovery of the catalytic properties of RNA (Guerrier-Takada et al, 1983; Kruger et al, 1982). These catalytic RNA are called ribozymes (ribonucleic acid enzymes) and among all non-coding RNA, ribozymes are the ones I have been interested in for my thesis work.

Naturally occurring ribozymes are divided into two sub-groups. The small ribozymes group includes the hammerhead, hairpin, Varkud satellite, hepatitis delta virus and *glmS* ribozymes. The second group of ribozymes encompasses group I and II introns, ribonuclease P, the ribosome and the Lariat-Capping ribozyme (LCrz). All, but the VS ribozyme and LCrz crystal structure have been solved in the last ten years. Nicolai Suslov from the team of Jo Piccirilli, presented the VS ribozyme 3D structure for the first time in November 2012 . In this manuscript I will describe the 2.5 Å crystal structure of the DiLCrz ribozyme, which we have solved in our laboratory during the last couple of years.

To obtain the 3D structure of DiLCrz, our method of choice has been crystallography. We confirmed the unexpected 3D structure of DiLCrz with small-angle x-ray scattering experiments.

We have also been interested in the role in folding of RNA structural motifs through the study of a motif we have identified in the bacterial RNase P crystal structure (Reiter et al, 2010), the pk-turn (Meyer et al, 2012). RNA 3D structures are mainly organized around A-form helices interspersed with single-stranded junctions. These single-stranded junctions orient the Watson-Crick helices by forming complex motifs or modules defined by a consensus sequence and a set of WC and non-WC base pairs (Leontis & Westhof, 2002) and/or architectural features. Therefore, RNA motifs can be predicted from the secondary structure (Jossinet et al, 2010). However, the RNase P and the LCrz crystal structures show that depending on intramolecular tertiary or intermolecular interactions, RNA modules can adopt unexpected motifs.

A. RIBOZYMES IN THE RNA WORLD

The RNA world as suggested by Gilbert in 1986 (Gilbert, 1986) has been mostly supported by the first observation of catalytic RNAs (Guerrier-Takada et al, 1983; Kruger et al, 1982) and assumed that RNA could have been the first macromolecule of life as it is at the same time a support for the genetic information and a catalyst. *In vitro* experiments even demonstrated that it is possible to transcribe an RNA template with an RNA polymerase ribozyme (Wochner et al, 2011).The RNA world hypothesis can be seen as a solution to the chicken and egg problem of what between DNA or protein was first (Cech, 2012; Lehman, 2010).

But how did DNA and protein arise from the RNA world? This question has not been answered yet (Dworkin et al, 2003) and if the hypothesis of the RNA world in the sense of RNA being the first macromolecule is based on an intellectual hypothesis, it contains some limitations. Primitive conditions on earth might have been favorable for nucleotides to appear (Powner et al, 2009) but not ideal for RNA macromolecules to thrive (Kawamura, 2012). It seems that macromolecules (DNA, RNA and protein) more likely arise from a prebiotic soup in which RNA and peptides would have coexisted. This leads to the alternative hypothesis that RNA catalysts, protein enzymes and modern RNPs are offspring of ancestral RNPs (Cech, 2009). One good example supporting the “RNP world” is RNase P that performs 5’ maturation of pre-tRNA. RNase P are present in the three kingdoms of life and have

structurally related RNA subunits, but different protein entities indicating that at least the RNAs should have been present in LUCA (Evans et al, 2006). Bacterial RNase P are ribozymes as their RNA entity is capable of cleaving the 5' leader sequence of pre-tRNA accurately without any protein help *in vitro* (Guerrier-Takada et al, 1983). However *in vivo*, bacterial RNA P is associated to proteins, which stabilize the RNA entity into an active conformation. With time, some new proteins appear in RNase P of eukaryotes and Archaea sharing more of the catalytic properties with its RNA component (Evans et al, 2006) (Walker & Engelke, 2006). Interestingly protein only RNase P have been found in organelles of some eukaryotes (Gutmann et al, 2012) (Holzmann et al, 2008) but their protein components seem not to be evolutionary related to the protein counterparts of RNP RNase P. From an "RNP world" and depending on selective pressure and macromolecules specificity, evolution would have then chosen different roads to catalyse the same reaction.

Nevertheless, if RNA might not be "THE" first macromolecule of life it has a major and indispensible role in modern biology. Indeed RNA is the obligatory intermediate between DNA and proteins but it is not, from far, its only mission in cells (Kapranov et al, 2007). With the example of RNase P we already understand that RNAs combined with proteins in RNP fulfill diverse processes. These processes range from pre-RNA maturation (Green, 1986), protein synthesis (Ban et al, 2000) (Nissen et al, 2000) to telomerase activity (Miller & Collins, 2002) (Qiao & Cech, 2008) and recognition of the signal sequence of secreted proteins (Walter et al, 1984). ncRNAs can by various ways finely regulate gene expression (Barrett et al, 2012) (Moreira et al, 2012) for example in the case of siRNA and microRNA that by base pairing to mRNAs regulate their life span and availability for translation (Joshua-Tor & Hannon, 2011). Introns interrupting pre-tRNA, pre-rRNA and pre-mRNA are of different kinds and regulate gene expression via splicing. A more detailed description of introns is made in I.A.

NcRNA also encompass ribozymes. Ribozymes are ribonucleic acid enzymes that, for naturally occurring ones, can be indexed in two sub-groups. The small ribozymes group includes the hammerhead, hairpin, Varkud satellite, hepatitis delta

virus and *glmS* ribozymes. The second group of ribozymes encompass group I, II and III introns, ribonuclease P, the ribosome and the lariat-capping ribozyme.

1. Lariat-Capping ribozymes

Lariat-capping ribozymes are a new family of ribozymes on which I focused during my thesis. LCrz, initially named GIR1 for Group-I-like ribozyme 1, have been renamed as neither their 3D structure nor their branching reaction supports their original name.

a) The lariat-capping ribozymes in twin-ribozyme introns

LCrz has been observed for the first time in the myxomycete *Dydimium iridis* (Johansen & Vogt, 1994) where it is inserted in the apical loop of the P2 stem of DiGIR2 a canonical group IE intron. DiLCrz is followed by a homing endonuclease gene (HEG) which itself contains a spliceosomal intron, I51 **(Figure 1)**. The 1.4 kb twin-ribozyme intron is inserted at position 956 (*E. coli* SSU rDNA numbering system) in the SSU of the rDNA. Therefore the twin-ribozyme intron is named Dir.S956-1 (Johansen & Haugen, 2001).

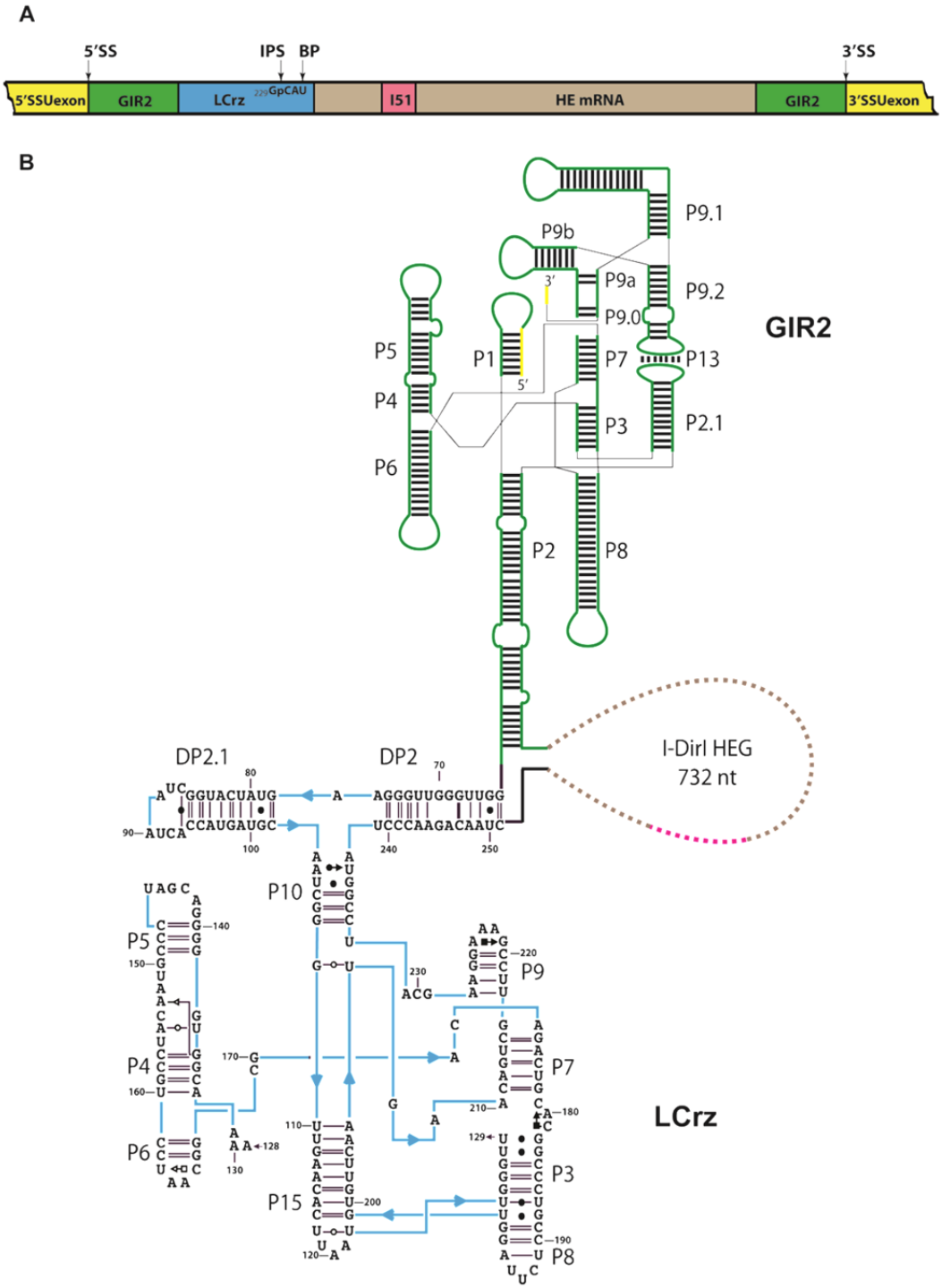


Figure 1: (A) Schematic view of the twin-ribozyme intron with at its extremities the exons (yellow) encoding for the SSU of the ribosome. The splicing group I intron, GIR2, is shown in green, the branching LCrz in blue, the ORF of the HEG in brown with its spliceosomal intron I-51 in pink. 5'splice site (SS), 3'splice site, branching point (BP) and internal processing site (IPS) are pointed out by arrows. Nucleotides forming the lariat (CAU) and G_{229} equivalent of ωG in group I introns are materialized. (B) The twin-ribozyme intron with the secondary structure of GIR2 shown in green interrupted in P2 by the sequence of LCrz and HEG. The 5' and 3' exons are shown in yellow. LCrz secondary structure (blue) is represented as the one of group I introns. HEG (brown) and I-51 (pink) are shown in dashed lines.

Later, LCrz has been found in 29 species of the amoeboflagellate genus *Naegleria* (Einvik et al, 1997) where the 1.2 kb twin-ribozyme intron (or twintron) is inserted at position 516. In Nae.S516 (Johansen & Haugen, 2001), NaGIR2 belongs to the group IC1 introns, LCrz is inserted in the apical loop of the P6 stem of GIR2 and HEG do not contain a spliceosomal intron **(Table 1)**.

	Dir.S956-1	**Nae.S516**
Microorganism	*Dydimium iridis* (Panama 2 isolate)	29 out of 70 *Naegleria* strains sequenced
Insertion site of the twintron	Position 956 in SSU rRNA	Position 516 in SSU rRNA
Insertion site of LCrz-HEG	P2 stem of DiGIR2	P6 stem of NaGIR2
GIR2	IE type group I intron	IC1 type group I intron
Intron in HEG	Spliceosomal intron of 51 nts	

Table 1: Comparisons of the principal features of the twin-ribozyme from *D. iridis* and *Naegleria*.

In the *Naegleria* genus 29 out of 70 sequenced species genomes possess Nae.S516 twin-ribozyme intron (Wikmark et al, 2006). This widespread but sporadic distribution indicates that the twin-ribozyme intron appeared early in the *Naegleria* genus, spread *via* vertical inheritance pattern and has been subsequently lost in the course of evolution in 60% of the strains. In the contrary, only one strain of the *Dydimium* genus, the panama isolate 2, contains the twin-ribozyme intron indicating that Dir.S956-1 has been recently horizontally inherited. Despite their apparent unrelated origin, DiLCrz and NaLCrz share a high sequence similarity **(Figure 2)**. More recently LCrz has been observed in the heteroloblosea *Allovahlkampfia sp.* (unpublished data- Yunjia Tang PhD work). Asp.S516 is organized as Nae.S516 and shares the same principal features.

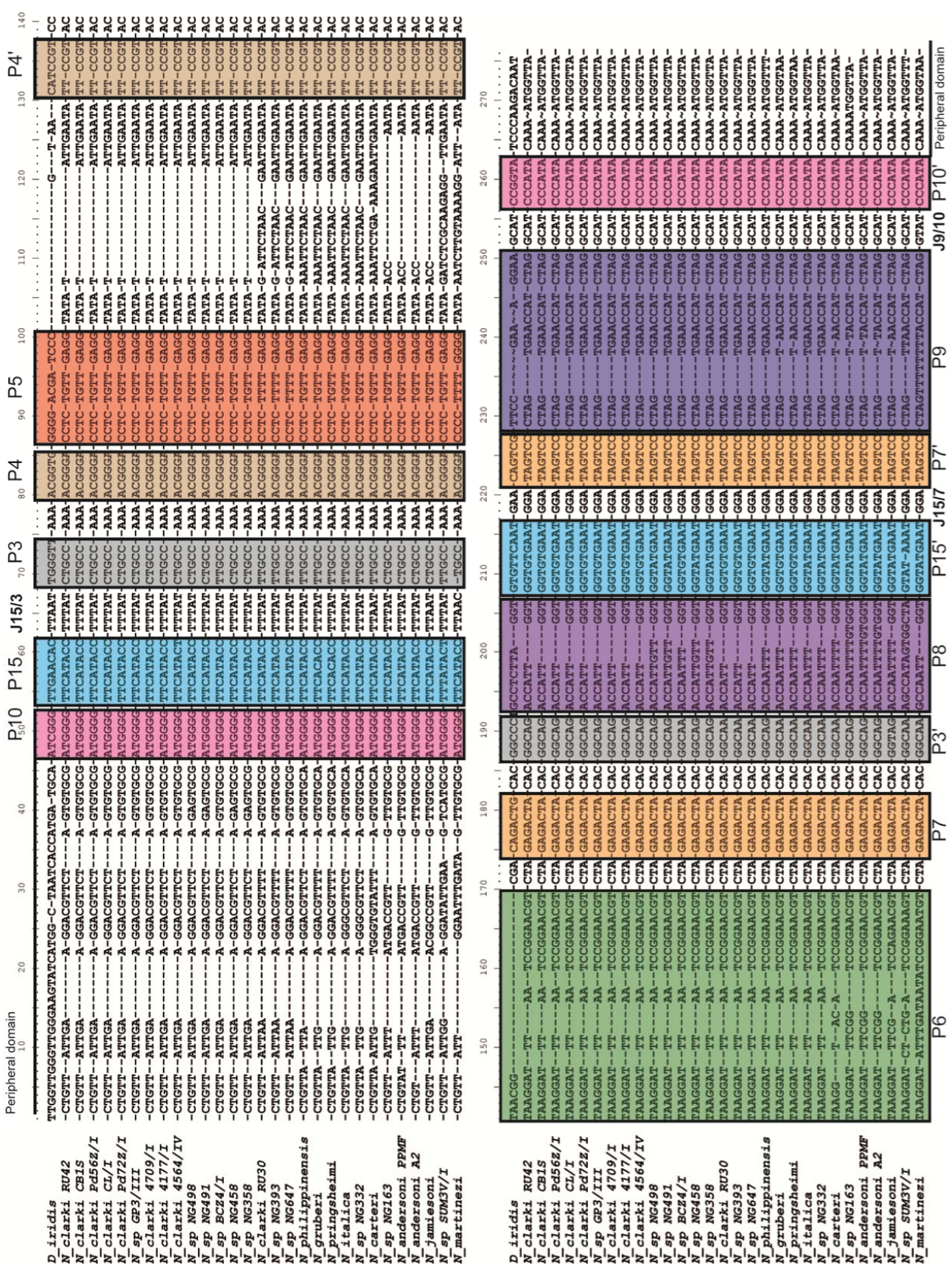


Figure 2: Sequence alignment of DiLCrz and NaLCrz from 29 different species.

b) Processing pathways of twin-ribozyme introns

Processing pathways of the twin-ribozyme introns are tightly connected to the host life cycle **(Figure 3)**. The first and predominant pathway occurs when conditions of growth are optimal for the amoeba or flagellate at the haploid stages **(Figure 4)**. This pathway is the self-splicing pathway during which the GIR2 intron is responsible for the twin-ribozyme excision and exon ligation, thus initiating the 18S pre-rRNA maturation. Once the intron is spliced out, LCrz can catalyse its branching reaction releasing the downstream HE pre-mRNA capped by a lariat of three nucleotides (Nielsen et al, 2005; Pyle, 2005). The 5' three nucleotides lariat substitutes for the conventional m^7G cap of mRNA. The 5' capped HE pre-mRNA is then cleaved at IPS before being polyadenylated at the 3' end. Last step is the removal of I51 spliceosomal intron of the mRNA **(Figure 3)** (Vader et al, 1999). No IPS3 is detected for Nae.S516 (Decatur et al, 2000). This pathway benefits to the host as well as to the intron in the sense that mature 18S rRNA and mature HE mRNA are produced.

The second processing pathway leads to the formation of a circular twin-ribozyme intron in which LCrz is inactive. This circularization pathway is solely mediated by GIR2 and is a general feature of group I introns (Nielsen et al, 2003). Reopening of the circle seems to be required for LCrz branching reaction to occur. This pathway is clearly detrimental to the host because exons are not ligated, but favourable to the intron as the FLC has been suggested to promote group I intron mobility (Birgisdottir & Johansen, 2005). FLC are especially abundant after heat shock (unpublished data) indicating that FLC is biologically significant **(Figure 3)**.

The last pathway occurs under starvation conditions that lead to *D. iridis* encystement. In this pathway, LCrz only is active, leading to the release of a 7.5 kb RNA product cap with the three nucleotides lariat and un-ligated exons (Vader et al, 2002) **(Figure 3)**. The biological relevance of this pathway has not been clarified yet, but it could be a way for the host to down-regulate the level of rRNA expression during starvation while having a stock of HE mRNA ready to be translated upon improvement of growing conditions.

Looking at these three pathways, it is clear that within the twin-intron ribozyme, GIR2 and LCrz actions have to be coordinated. Indeed, LCrz is fully transcribed before GIR2 but should not be active before the latter. Therefore LCrz should adopt an alternative inactive conformation during splicing and FLC pathways to prevent early cleavage of the intron that corresponds to the starvation-induced pathway. How do LCrz and GIR2 coordinate their action? We have no clear evidence yet for RNA-RNA or protein-RNA interactions that could be responsible of this regulation.

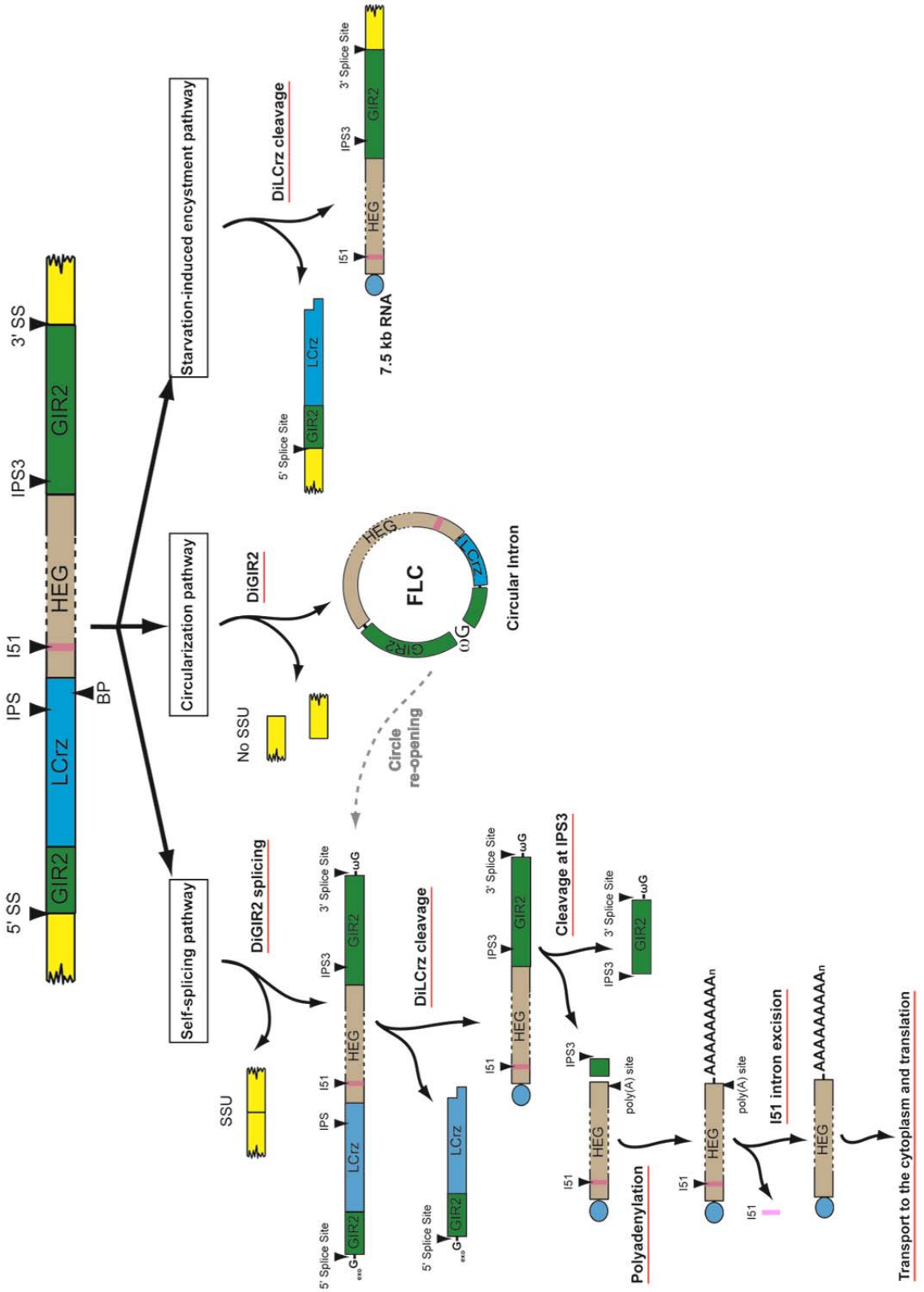


Figure 3: Diagram of the three different pathways followed by the twin-ribozyme intron. 5' and 3' exons of the SSU of the ribosome are shown in yellow, the splicing group I intron GIR2 in green, the branching LCrz in blue, HEG in brown and the spliceosomal intron I-51 in pink.

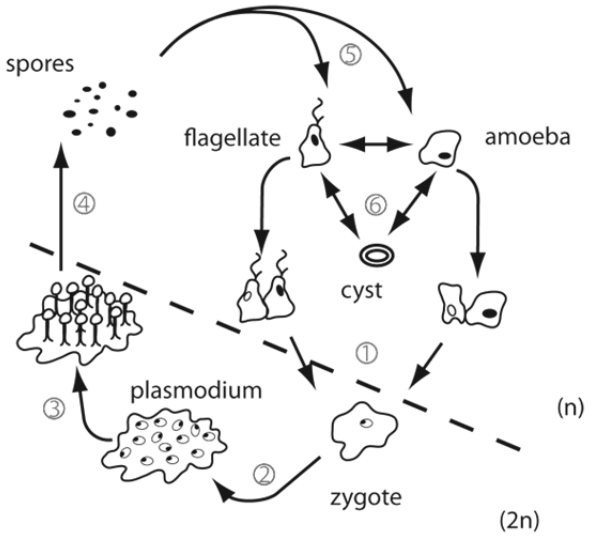


Figure 4: Life cycle of the myxomycete *Didymium iridis* that alternates between a haploid microscopic stage (top part) and a diploid macroscopic stage (bottom part). (1) Sexual fusion of amoeba (drawn on right side) or flagellates (drawn on the left side) carrying different mating alleles (black or white nucleus). (2) Development of the zygote into a multinucleate plasmodium. (3) Development into sporangia. (4) Production of haploid spores by meiosis. Spores will be disseminated by the wind and will germinate into new haploid cells (5). Depending on environmental conditions these new haploid cells will either grow as flagellates, amoeba or cysts (6).

c) Lariat-capping ribozymes' secondary structures

LCrz secondary structure has been determined by comparative sequence analysis, chemical and enzymatic probing and site directed mutagenesis (Einvik et al, 1998b). The secondary structure of LCrz is closely related to the one of group I introns **(Figure 5).** Thus, helical stems of LCrz are designated as for group I introns. The secondary structure of the core of LCrz is organized in three domains: (1) the stabilization domain P4-P5-P6, (2) the substrate domain P10-P15 and (3) the catalytic domain P3-P7-P8-P9. There are however some differences between LCrz and group I introns. First, LCrz lacks the P1 stem and the J8/7 junction. In LCrz P15 is equivalent to P1 and P2 of group I introns. Secondly, LCrz pocesses one additional domain, the peripheral domain DP2-DP2.1. Thirdly, the pseudoknot formed by base pairing of J8/7 from LCrz with the 5' strand of P2 create P15 and allows the formation of the family C (Lescoute & Westhof, 2006) three way junction between P3, P8 and P15. In addition, the J15/7 and the J9/10 junctions are unique features of LCrz.

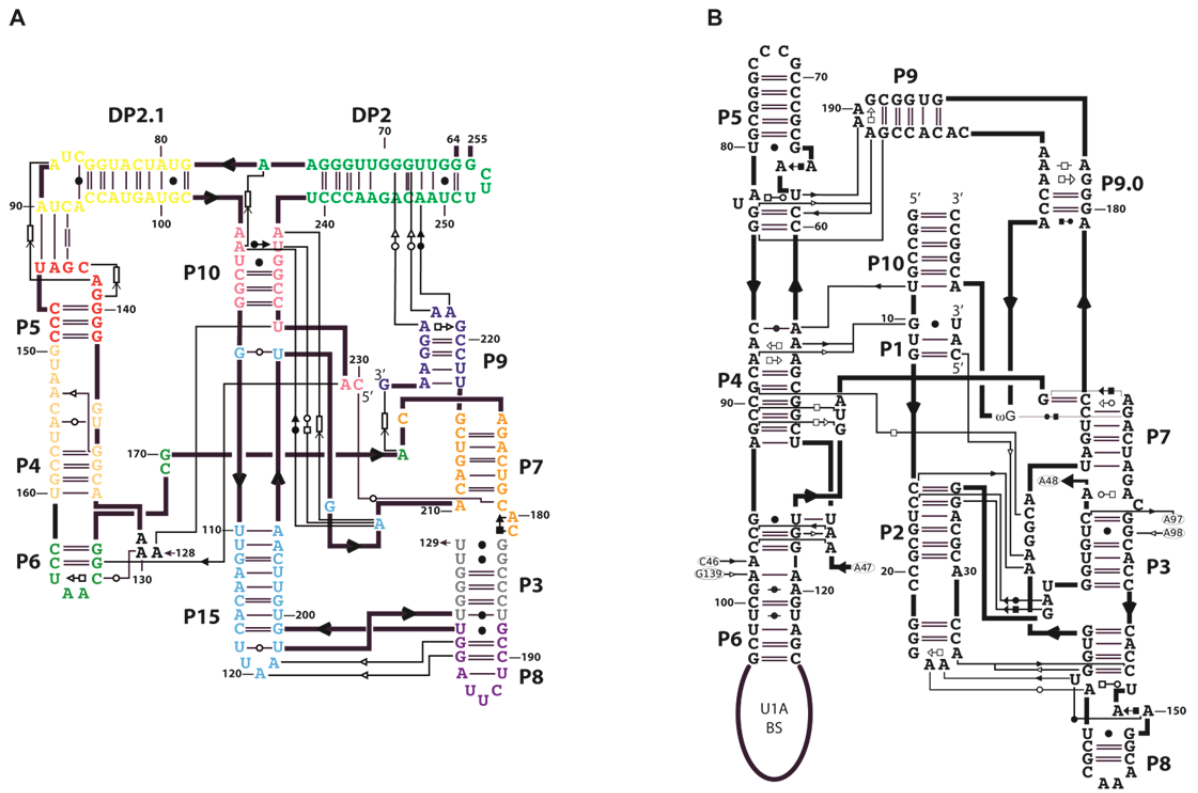


Figure 5: (A) Secondary structure diagrams of DiLCrz compared to *Azoarcus* group IC3 intron (B). Tertiary interactions are displayed using the LW nomenclature (Leontis & Westhof, 2001). Stacking interactions in LCrz are indicated by an arrow pointing into a rectangle.

The critical role of flanking regions of the LCrz for its branching activity has been demonstrated by deletion analysis of the 5' and 3' ends (Einvik et al, 2000; Jabri et al, 1997). One important property of DP2 is that this stem can dissociate to allow the 3' strand to adopt an alternative hairpin fold composed of nucleotides from the 5' end of the downstream HE pre-mRNA. This alternative fold is named HEG P1 and has been observed for the first time in the 5' UTR of the HE pre-mRNA (Vader et al, 1999). HEG P1 is also observed in the pre-catalytic form of DiLCrz (Nielsen et al, 2009) or in the post-clivage state of DiLCrz (Birgisdottir et al, 2011) were it is involved in the release of the HE pre-mRNA. DP2 and HEG P1 are mutually exclusive structures and present only in the active and inactive states respectively **(Figure 6)**. The mechanism by which the on and off switch of LCrz is mediated remains to be elucidated.

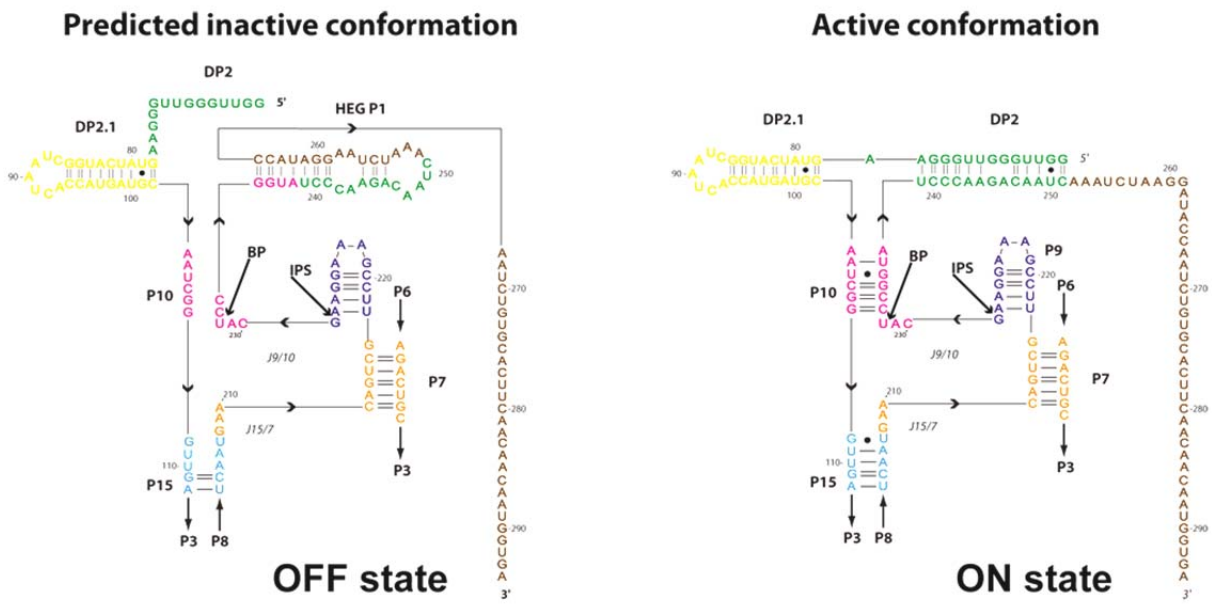


Figure 6: Secondary structure diagram of DiLCrz predicted off state (left panel) compared to the active state (right panel). When HEG P1 is formed, P10 and DP2 are disrupted. HEG P1 encompass nucleotides from mHEG 5'UTR and exclude the 5' strand of DP2. Arrows point out IPS and BP.

LCrz variants from *Naegleria* (Jabri et al, 1997) and *Allovahlkampfia sp.* genus share the same core organization as DiLCrz but harbor up to a hundred additional nucleotides **(Figure 7)**. Main differences are: (1) the NaLCrz and AlLCrz sequence starts in P2.1, (2) a variable J5/4 junction, (3) a longer P6 stem and (4) a variable L9 loop. An additional stem appears in the peripheral domain of AlLCrz. In NaLCrz the structural organization of nucleotides joining P10 to NP2.1 (NP2 plus an additional stem loop) is not clear yet.

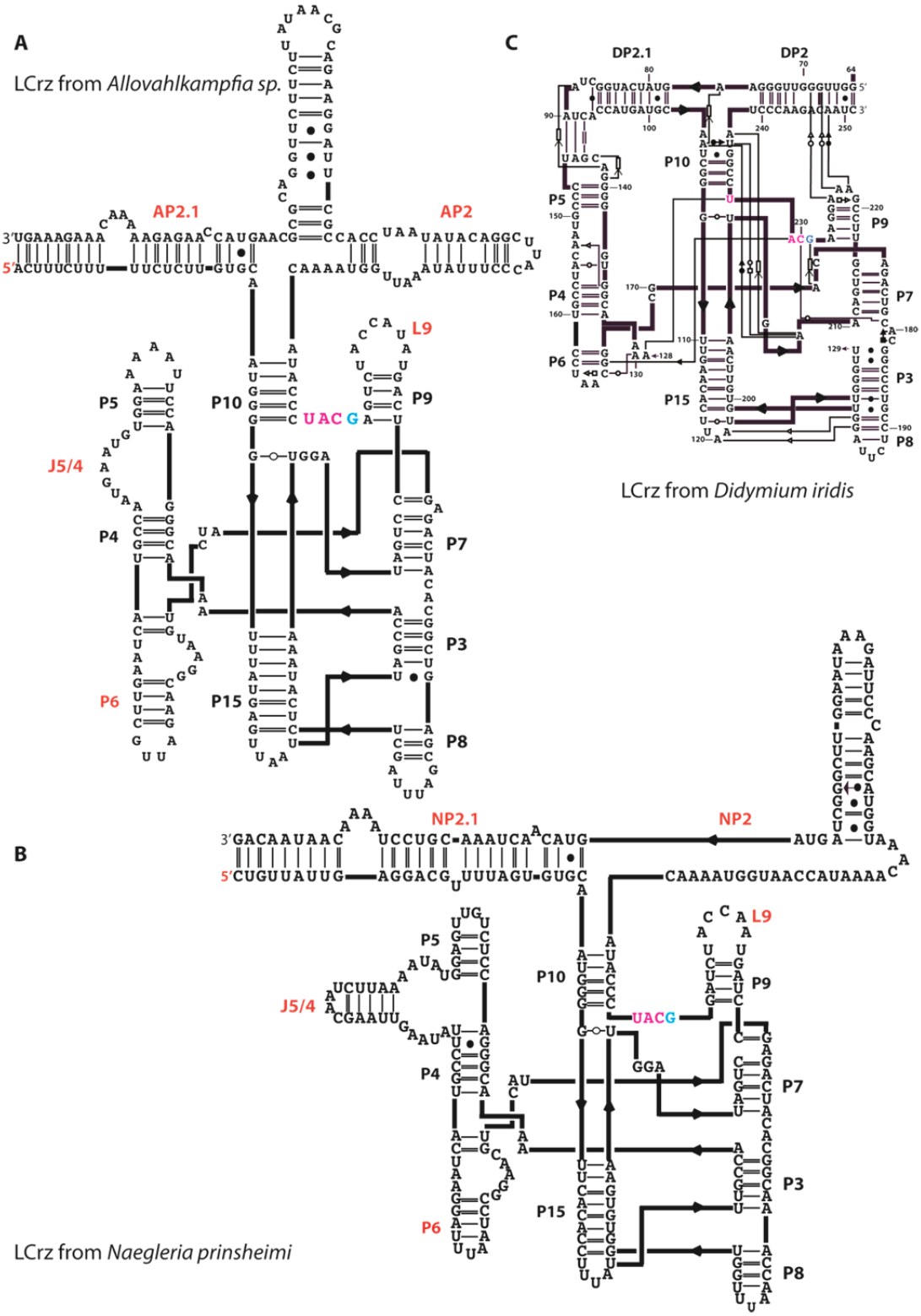


Figure 7: Secondary structure diagram of AlLCrz (A), NaLCrz (B) and DiLCrz (C). Main differences between AlLCrz and NaLCrz in comparison to DiLCrz are highlighted in red: (1) the sequence starts in P2.1, J5/4 is variable, P6 stem in longer, L9 comprise more nts and the peripheral regulatory domain is larger and potentially habor one additional stem. Nucleotides forming the post-cleavage lariat are shown in pink and G_{229} in blue.

2. Group I introns

Group I introns are ribozymes mainly found in the nucleolar rDNA and in mRNA, tRNA and rRNA from organelles of lower eukaryotes. Less than 4% of the identified group I introns are located in prokaryotes and none have been seen so far in Archaea (Haugen et al, 2005). The size of group I ribozyme ranges from ~250 to ~500 nucleotides. Group I introns catalyse a two steps self-splicing event allowing the ligation of the 5' and the 3' exons and the release of the intron (Herschlag & Cech, 1990). The spliced-out intron can further form a long full length circle (Vicens & Cech, 2009) or truncated circles (Tanner & Cech, 1996; Zaug et al, 1983). However group I introns can also undertake the so-called "circularization pathway" that leads to the formation of a full length circle intron but unligated exons (Nielsen et al, 2003; Zaug et al, 1983) **(Figure 8)**.

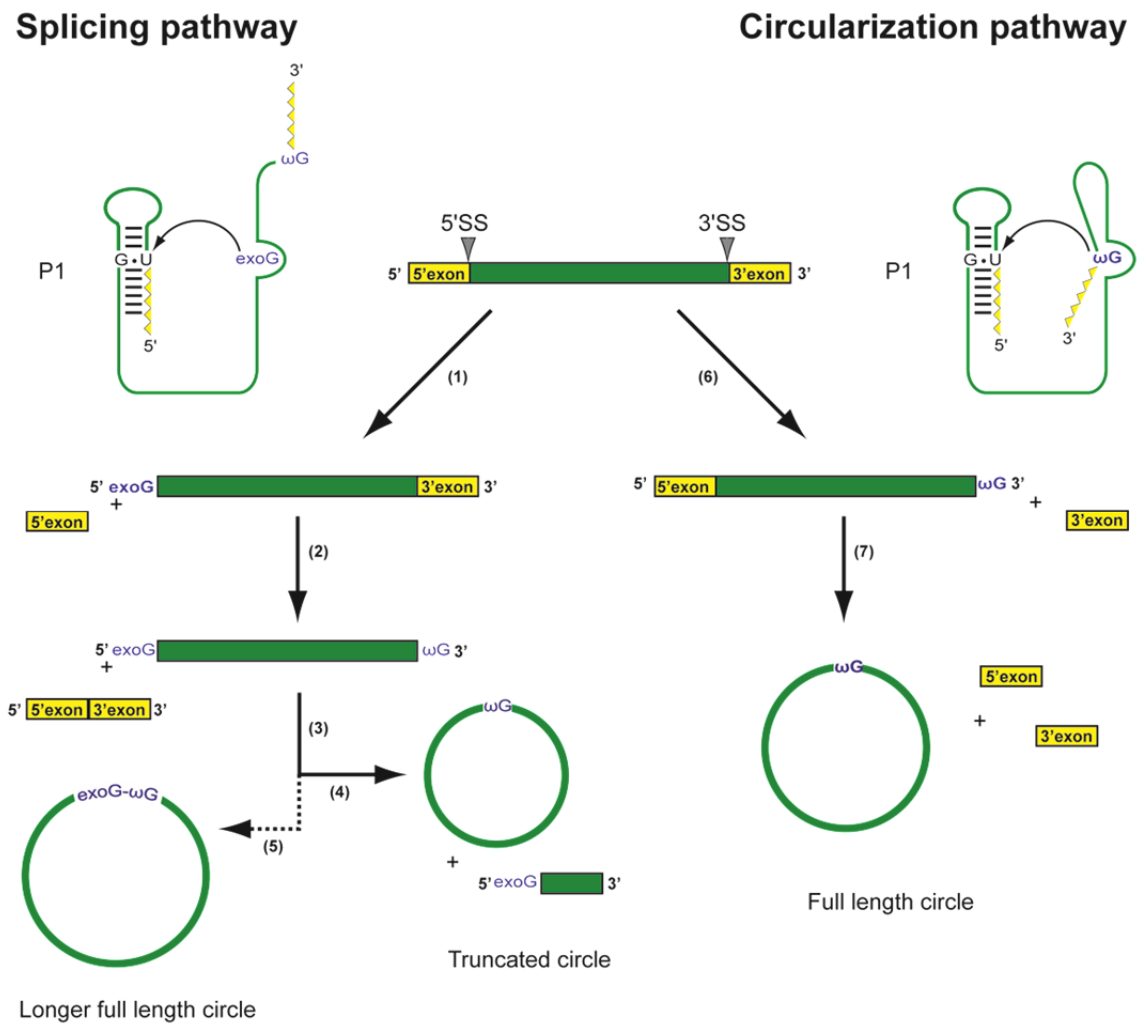


Figure 8: Schematic representation of the group I introns' processing pathways. Figure adapted from (Nielsen et al, 2003). Group I intron is shown in green and exons in yellow. exoG and ωG are colored in blue. On the left panel the splicing pathway is displayed. It takes places in two consecutive transesterification reactions (1 initiated by exoG) and (2 initiated by U) and further processing (3) of the group I intron lead either to a longer than the full length circle circle (5) or a truncated circle (4) (Vicens & Cech, 2009). The circularization pathway leading to unligated exons and formation of a full lenth circularized intron result from a hydrolytic cleavage at the 3' SS (6) followed by a transesterification reaction initiated by ωG (7).

Some group I introns contain ORF encoding homing endonuclease (HE). One role of the HE is cognate intron insertion at DNA level by homing in an intron-less allele at a homologous site via the double strand break repair pathway (Belfort & Perlman, 1995; Dujon, 1989). Evolution of group I introns and their associated HE are correlated as suggested by Goddart and Burt (Goddard & Burt, 1999) **(Figure 9)**. Yet, HEs have two ways to escape the Goddart-Burt cycle. The first one is HEG mobility alone (Haugen et al, 2005) and the second one is to acquire new function

such as maturase activity (Chatterjee et al, 2003). Indeed, HEs having maturase activity are helpful for the folding of its cognate group I intron into its active conformation, therefore improving self-splicing (Belfort, 2003). A way for group I introns to escape the Goddart-Burt cycle is to promote their own mobility at RNA level by reverse splicing in homologous and heterologous sites (Roman et al, 1999; Roman & Woodson, 1995; Roman & Woodson, 1998; Woodson & Cech, 1989) **(Figure 10)**. Interestingly, FLC could also allow intron reverse splicing (Nielsen et al, 2003). Due to group I intron mobility, twintrons are observed when a group I intron is inserted into a pre-existing group I intron (Einvik et al, 1998a; Hafez et al, 2013). Likewise, HEs can acquire maturase activity not to be lost in the course of evolution, the second group I intron in a twintron could evolve to gain new functions e.g. branching.

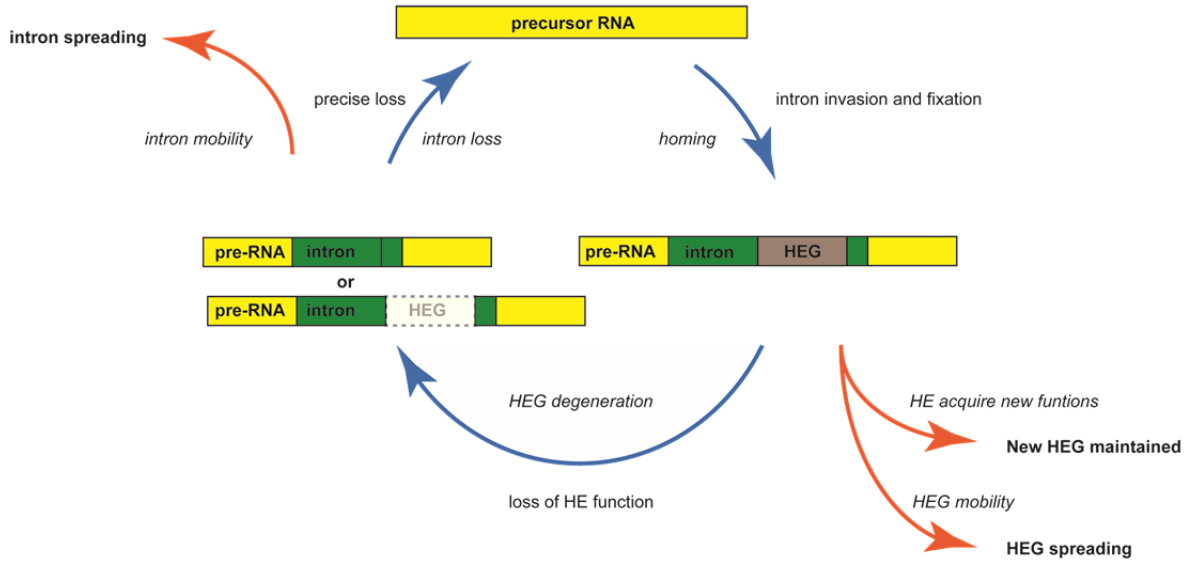


Figure 9: Goddard and Burt cycle (blue arrows) for group I introns (green) and their cognate HEG (brown) showing a correlated evolution of both genetic elements. When all intron-less alleles have been invaded by group I introns thanks to homing endonucleases action, both elements tends to be lost. One way to escape intron loss is for the intron to promote its own spreading at RNA level (red arrow on the left hand side). To persist HE can either acquire new functions or promote its own mobility (red arrows on the right hand side).

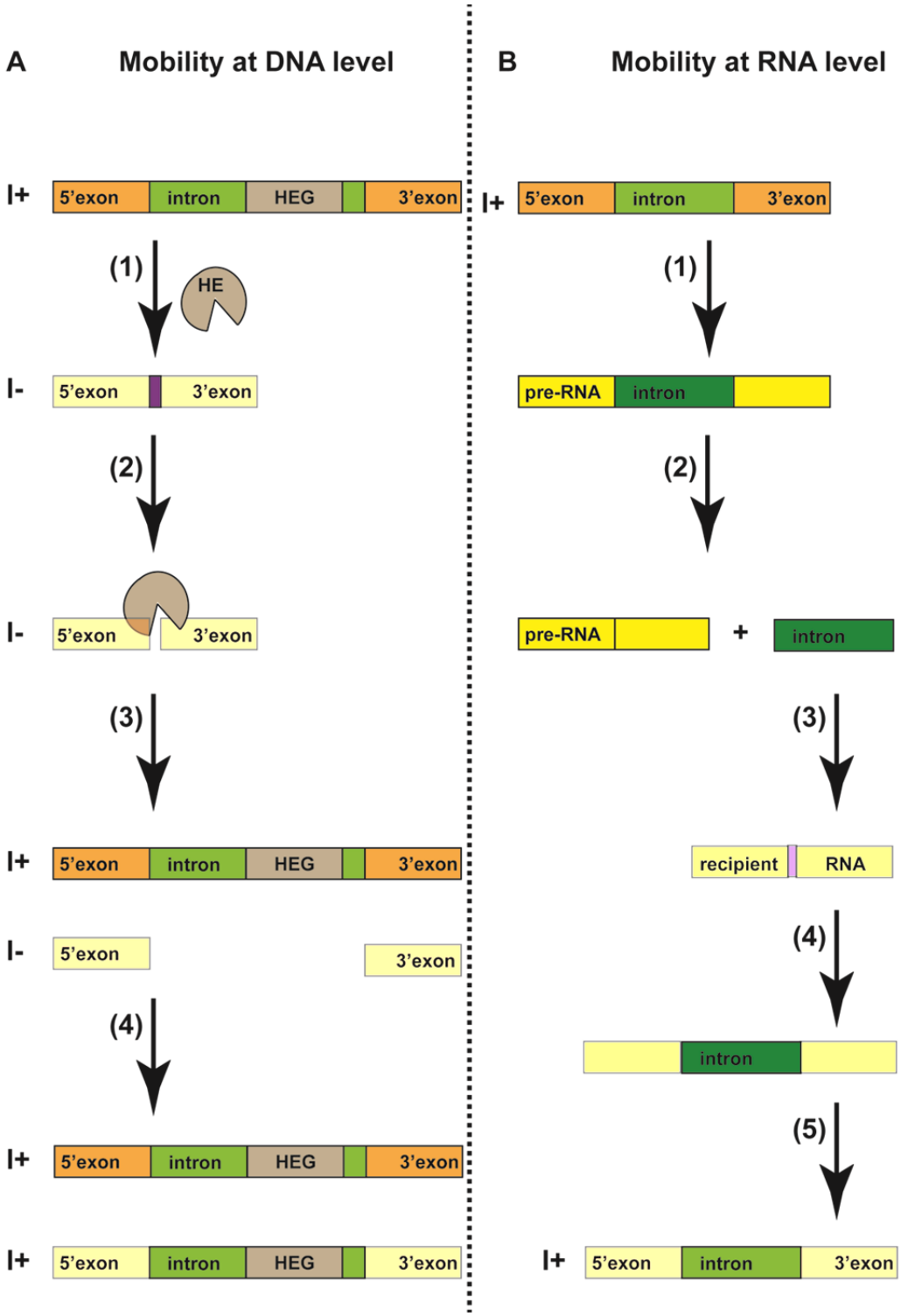


Figure 10: Group I intron mobility (Figure adapted from (Hedberg & Johansen, 2013)). Mobility at the DNA (orange) level promoted by HE (brown) shown on the left panel. It implicates a dsDNA break mediated by the HE binding (1) and cleavage (2) at its recognition site (shown in purple) followed by recombination (3) and repair (4). Mobility at the RNA (yellow) level promoted by the intron (green). Steps for intron self-mobility are: (1) transcricription of an I+ gene, (2) group I intron splicing, (3) spliced-out intron recognize an intron insertion site in an I- RNA, (4) intron insertion by reverse-splicing and (5) reverse transcription. I- is an intron-less allele and I+ is an intron-containing allele.

As mentioned above, some group I introns require the action of a maturase to fold in their active conformation (Russell, 2008). This dependence of the ribozyme toward proteins for folding can be ordered in three categories: (1) self-splicing all-ribozyme introns e.g. *Tetrahymena*, (2) host-factor-dependent optional introns e.g. S1389 in *Diderma* rDNA (Wikmark et al, 2007), (3) host-factor-dependent obligatory introns e.g. ND5-717 intron in hexacoral mitochondrial genomes (Emblem et al, 2011). Host-dependent introns often show extended peripheral loop regions (Wikmark et al, 2007). Additionally, the level of dependence of the group I intron toward proteins seems to be correlated with its GC content (Vicens et al, 2008). If the group I intron in need of a protein co-factor does not encode a maturase it will hijack proteins from the host organism (Lambowitz & Caprara, 1999).

Noteworthy, group I intron sequence can be split into stretches of nucleotides located at different position in the genome with a preserved splicing activity. Examples of composite group I introns capable of *trans*-splicing are all found in the *cox1* gene of mitochondria (Burger et al, 2009; Nadimi et al, 2012). Interestingly this ability of group I intron can be used as a molecular biology tool to perform splicing of genes in bacteria (work presented at the RNA meeting 2013 by Ulrich Müller).

a) Group I introns' secondary structure

Group I introns vary considerably in length, sequence and catalytic efficiency nevertheless they all retain a consensus secondary structure composed of helical domains, numbered P1 through P10, and single stranded junction and/or non-WC base-pairs connecting these helical domains (Cech et al, 1994; Davies et al, 1982; Michel et al, 1982). Domains P1 and P2 compose the substrate domain in which P1 is formed by the 5' end of the ribozyme containing the IGS and the 3' end of the 5' exon. At the top of P1 there is a highly conserved G.U wobble base pair (Michel & Westhof, 1990). The P3-P9 domain is the catalytic domain. It encloses P3, P7, P8 and P9 and is interrupted by the stabilization domain P4-P6 (Kim & Cech, 1987; Murphy & Cech, 1993). The G-binding pocket accommodating exoG and ωG afterwards (detailed catalytic mechanism is described in I.B.1.b) sits at the top of P7

(Michel & Westhof, 1990) **(Figure 11)**. P10 is a helix formed after the first step of splicing by the 5' end of P1 and the 5' end of the 3' exon.

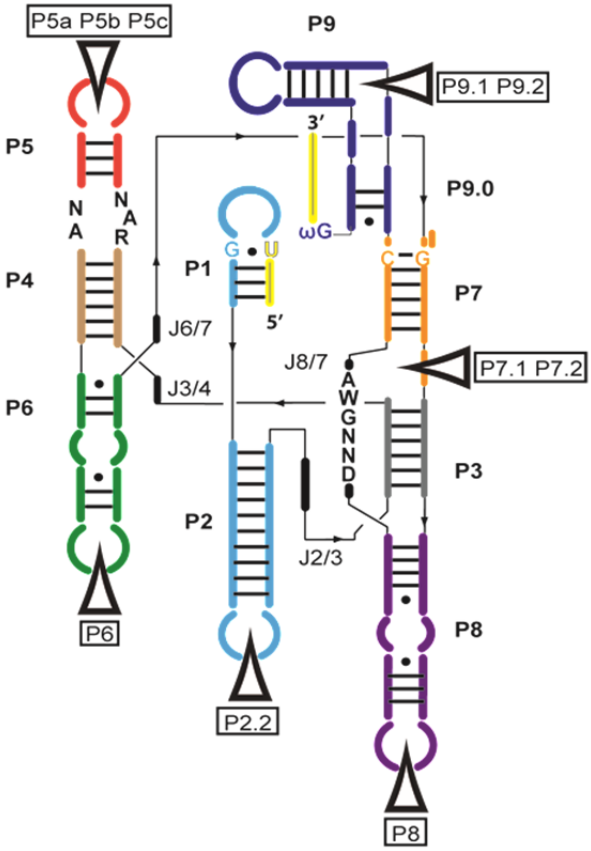


Figure 11: Secondary structure diagram of group I introns as proposed by Cech (Cech et al, 1994). Only conserved nucleotides (Michel & Westhof, 1990) are shown in letters. Stems are colored in the same color as the one of their similar stem in LCrz (see **Figure 5**). Arrows indicated insertion sites for additional peripheral domains.

Group I introns can contain additional domains budding from terminal or internal loops of helical domains as shown in **Figure 11**. Additional domain and co-variation analysis led to the classification of group I introns into 13 subgroups (Michel & Westhof, 1990). Peripheral domains enhance the stability of the folded ribozyme and/or help the folding of the ribozyme in an active conformation (Doudna & Szostak, 1989; van der Horst et al, 1991).

b) Group I introns' three-dimensional structure

Crystal structures of three different group I introns from three diverse origins in three distinct chemical states are available to appreciate group I introns 3D structure. The first one is the group IC1 intron located in rRNA of *Tetrahymena* trapped in a pre-catalytic state in which P1 is not present due to the construct design (Guo et al, 2004) **(Figure 12A)**. To be able to crystallize this construct the Cech lab performed *in vitro* selection for more thermostable ribozymes than the wild-type one (Guo & Cech, 2002). The second crystal structure is the one of the bacteriophage Twort group IA2 intron (Landthaler & Shub, 1999). In this crystal structure, P1 is formed by addition in *trans* of a four nucleotides product analogue, the G.U wobble base pair at the top of P1 is formed and ωG is in the G-binding pocket of P7 (Golden et al, 2005) **(Figure 12B)**. The last group I intron that has been crystallized is the one of the $tRNA^{Ile}$ from *Azoarcus* (Adams et al, 2004). It belongs to the subgroup IC3 and it has been engineered to mimic a transition state just after the first step of self-splicing but right before the second step occurs. Therefore this structure is of particular interest for our work as P10 and P1 are formed and ωG lies into its binding pocket **(Figure 12C)**. These three features allow a fairly straightforward comparison with LCrz crystal structure (see results section, manuscript submitted). All features of tertiary structures of the three group I introns have been described and compared in a book chapter (Golden, 2008) and I will only describe some details used for further comparison with LCrz **(Figure 13)**.

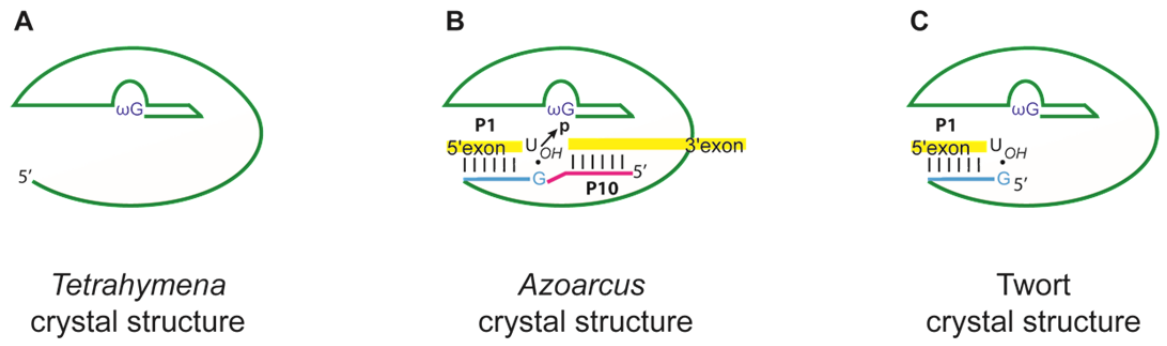


Figure 12: Schematic diagram of the three different catalytic states in which three different group I introns have been trapped for crystallization. The ribozyme core is represented in green, exons in yellow, P1 in cyan, P10 in pink and ωG in blue. Figure adapted from (Golden, 2008)

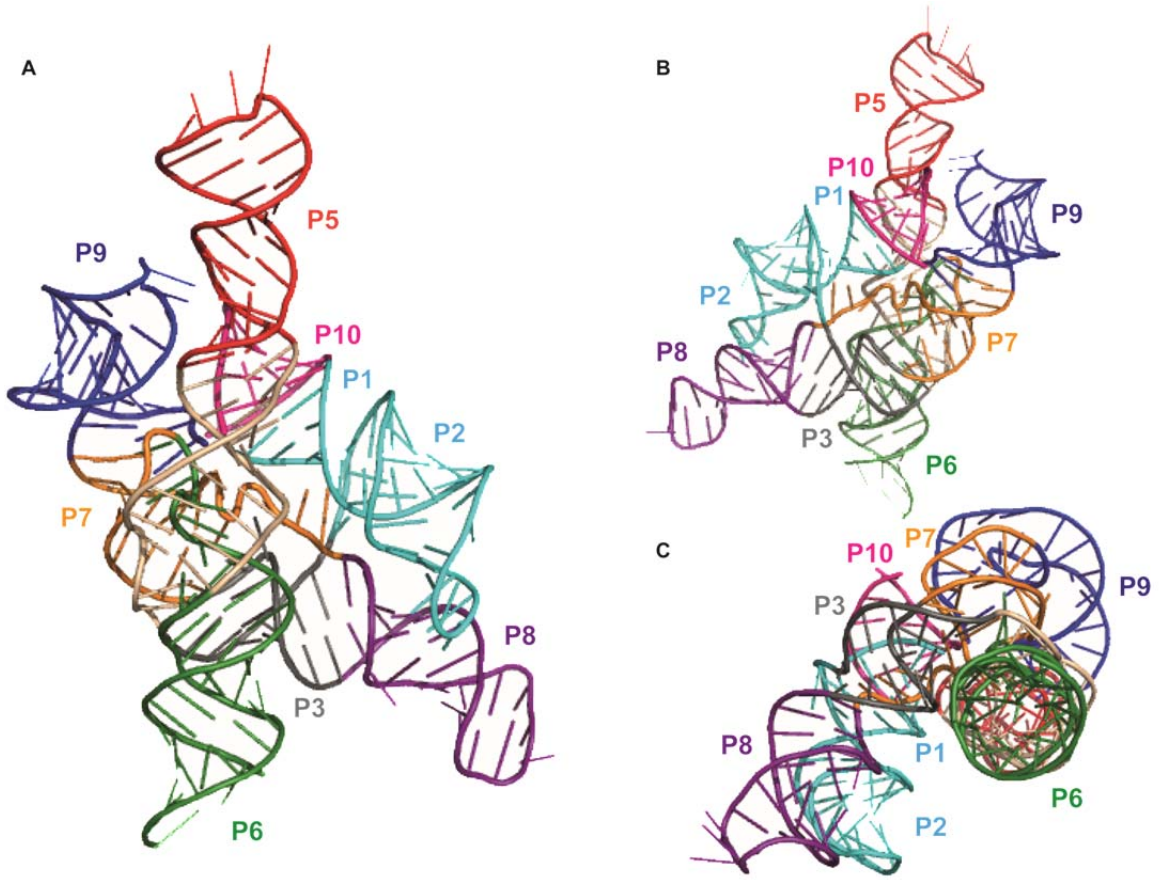


Figure 13: Three dimensional structure of the *Azoarcus* group IC3 intron (Adams et al, 2004) stems are colored as for DiLCrz (**Figure 5**). (A) View of the group I intron with the P5/P4/P6 stack oriented vertically and the catalytic and substrate domains behind. (B) Same view as (A) but after a 180° rotation along the y axis allowing us to observe the catalytic and substrate domains in the front. (C) View of the group I intron with the catalytic and substrate domains wrapped around P5/P4/P6 stack oriented with a 90° rotation along the x axis in comparison to view (A).

The G-binding pocket has been identified biochemically using mutant of the group I intron from *Tetrahymena* and analogues of guanosine (Michel et al, 1989a). In the three crystal structures and as predicted ωG is forming a base triple interaction with the second and highly conserved G.C base pair of P7. On the top of ωG stacks a single bulged adenosine forming an additional base triple with the first base pair of P7. The third (A-U) base pair of P7 forms as well a base triple interaction with the adenosine of the J6/7 junction. Thus the adenosine of J6/7 tethers the stabilization domain to the catalytic domain but it also shares an essential hydrogen bond between its 2'-hydroxyl group and the 2'hydroxyl and N3 of ωG (Bass & Cech, 1986;

Ortoleva-Donnelly et al, 1998). The G-binding pocket is thus composed of three consecutives base triples **(Figure 14)**.

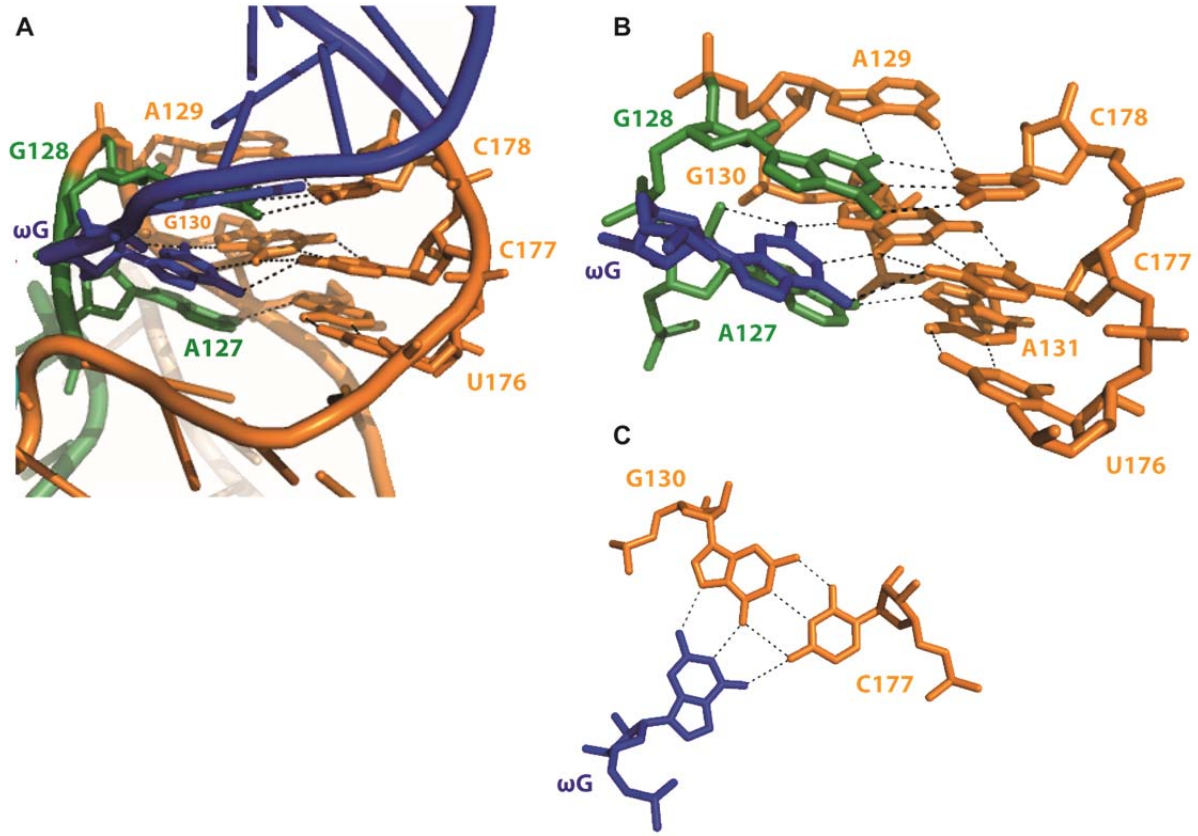


Figure 14: Three-dimensional structure of the *Azoarcus* group IC3 intron G-binding pocket (A). A129.G128-C178 and A127.A131-U176 base triples sandwich the ωG.G130-C177 base triple therefore forming the G-binding pocket. (B). Detailed view of the base triple holding ωG in its binding pocket (C). Nucleotides from P7 are in orange, from J6/7 in green and ωG in blue.

The stabilization domain displays three coaxially stacked helices, P4-P5-P6, around which wraps the catalytic domain. J4/5 of the stabilization domain is made of non-WC base pairs that interact with the 5' splice site. The tandem of A.A sheared base pairs of J4/5 uses two adenosines to make A-minor interactions with the G.U wobble base pair of P1 (Strobel et al, 1998) **(Figure 15A)**. J3/4, J6/6a and J6/7 are tethering P4-P6 to P3-P9 domain **(Figure 15B)**. J2/3 and J8/7 are single stranded junctions important for the recognition of the 5'SS as their nucleotides interact with the minor groove of P1 hence facilitating packing of P1-P2 against P3-P9 domain (Pyle et al, 1992; Szewczak et al, 1998) **(Figure 16)**.

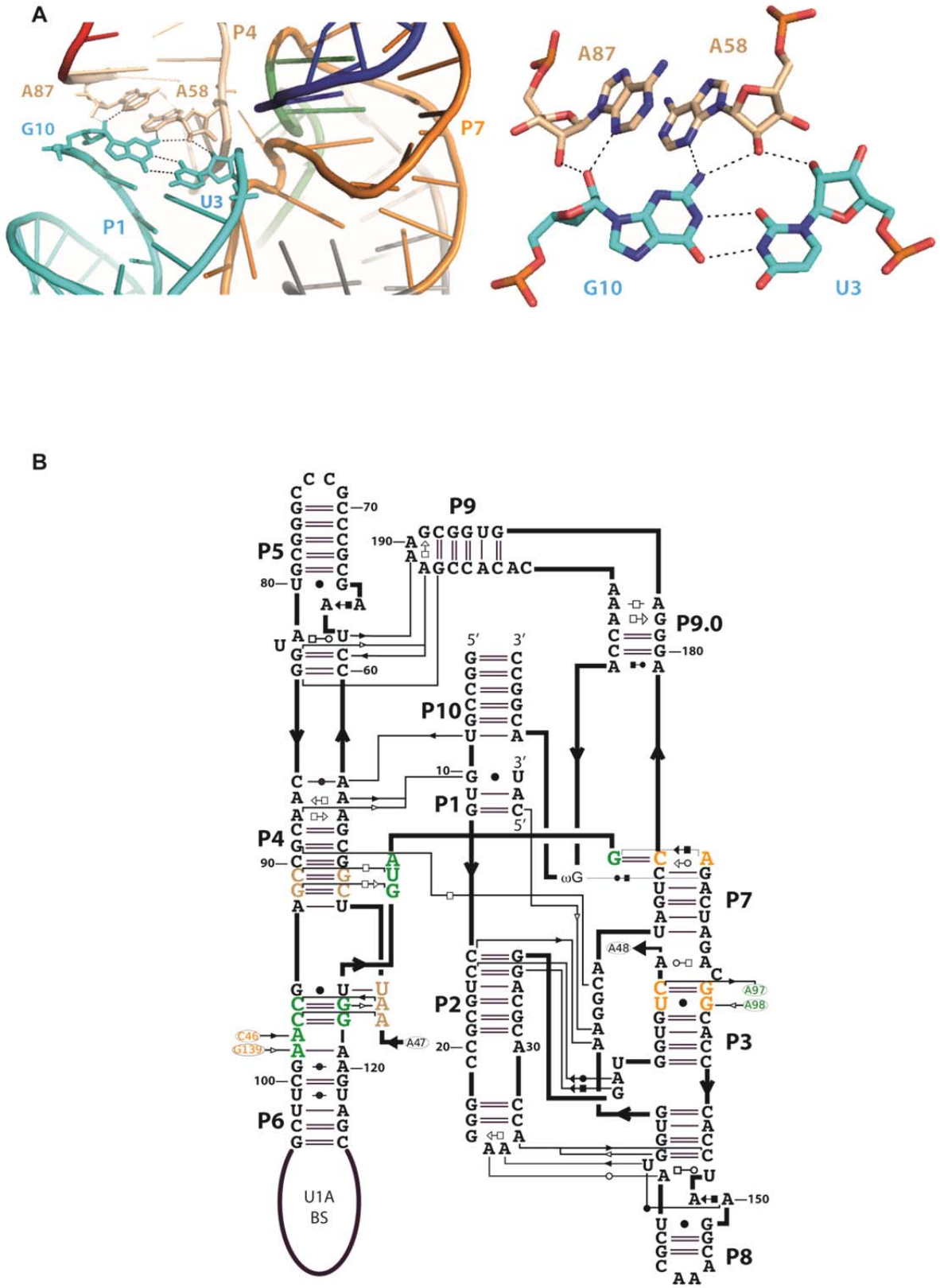


Figure 15: (A) Three-dimensional structure of the *Azoarcus* group IC3 intron showing the implication of J4/5 (beige) at the 5' splice site. Detail of A-minor interactions of A58 and A87 from J4/5 with the G.U wobble base pair of P1 (cyan). (B) Secondary structure diagram of the Azoarcus group I intron with nucleotides tethering P4-P6 domain to P3-P9 domain shown in colors.

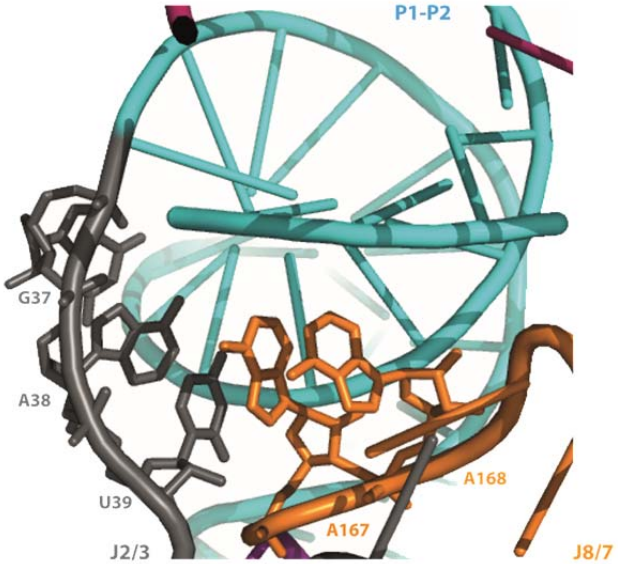


Figure 16: Three-dimensional structure of the *Azoarcus* group IC3 intron showing interaction of J2/3 (grey) and J8/7 (orange) with the minor groove of P1 (cyan).

c) Group I introns and the lariat-capping ribozymes

Group I introns and LCrz are tightly related since their sequence and secondary structure are very similar **(Figure 5)**. Hence, the first name given to LCrz was Group-I-like-intron ribozyme 1. As mentioned in I.A.1.a, LCrz are parts of twintrons located in the SSU rDNA of three species of eukaryotes (Einvik et al, 1998a). Very interestingly, one group I twintron has been described recently in the mtDNA at position mS917 in gene encoding the mitochondrial small ribosomal subunit of several species of fungi (Hafez et al, 2013). However LCrz are not catalyzing splicing, but branching in order to cap the downstream HE pre-mRNA with a lariat of three nucleotides (Nielsen et al, 2005). Moreover, comparison of the three-dimensional structures of group I introns and DiLCrz shows some important differences (see Results section). Consequently, it is reasonable to suggest that LCrz come from a group I intron that has evolved to acquire new functions (Beckert et al, 2008).

3. Group II introns

Group II introns are large ribozymes, ~300 to ~1000 nucleotides, observed in about 25% of bacteria (Dai & Zimmerly, 2002) as well as in mitochondrial and chloroplast genomes of lower eukaryotes, higher plants (Dai et al, 2003) and archaea (Rest & Mindell, 2003) and more recently in an annelid worm (Valles et al, 2008). As for group I introns they have to splice out to allow 5' and 3' exons of a given gene to be ligated. Most group II introns enclose an intron encoded protein essential for *in vivo* self-splicing. Indeed, if *in vitro*, group II introns are capable of self-splicing it is in salt conditions far from cellular concentration (Daniels et al, 1996; Jarrell et al, 1988). The cognate IEP acts as a maturase, which helps the RNA to fold in its active conformation either by stabilizing its active fold or by resolving non-native structures that constitute a kinetic trap (Lambowitz & Belfort, 1993; Wank et al, 1999). Group II introns that do not contain ORFs or encode degenerated IEPs rely, like group I introns, on some of the host proteins to act as chaperones (Lambowitz et al, 1999).

Group II introns are not only self-splicing element. They are mobile genetic elements that are capable of insertion in DNA. Once again, for this retrohoming by reverse-splicing event, group II introns rely on their IEPs (Meunier et al, 1990). On top of their maturase activity IEPs are reverse transcriptases. After the RNA intron self-splicing the IEP stays bound to it and this RNP complex recognises the DNA to invade. The RNA intron uses its ribozyme activity for reverse splicing and the RT IEP is in charge of transcription of the inserted RNA intron back to DNA (Lambowitz & Zimmerly, 2004).

As a result of retrohoming, one group II intron can be inserted into an already existing one to form the so-called twintron (Copertino & Hallick, 1993; Dai & Zimmerly, 2003). Remarkably, depending on the insertion site of the second intron into the first one and to preserve self-splicing capability for both of them they have to adopt a concerted order of reaction (Drager & Hallick, 1993; Nakamura et al, 2002). Interestingly some introns and twintrons do not contain IEP and some IEP have been shown to act in trans on group II introns from unrelated genes (Dai & Zimmerly, 2003; Meng et al, 2005).

Group II introns are not always encoded as one stretch of nucleotides. A functional group II intron, capable of self-splicing, can be reconstituted in *trans* with two or more separate transcripts (Bonen, 1993; Glanz & Kuck, 2009). This last feature combined with evidence for a similar splicing mechanism designate group II introns as ancestors of the eukaryotic spliceosome (Cech, 1986; Collins & Guthrie, 2000; Sharp, 1985).

a) Group II introns' secondary structure

Despite low sequence conservation, group II introns adopt a highly conserved secondary structure in which six helical domain (DI to DVI) are distributed around a large central junction (Michel & Ferat, 1995). Each helical domain has its own function. DI is the largest domain; it contains sequences for recognition of both exons and most recognition sites for long range tertiary interactions with the other domains (Hamill & Pyle, 2006; Su et al, 2005; van der Veen et al, 1987a). DI, DII, DIII, DV, DVI are implicated in the formation of the catalytic site and the two steps of self-splicing (Costa & Michel, 1999; Fedorova et al, 2003; Qin & Pyle, 1998). Nucleotides of the central junction and of DV plus the bulged adenosine of DVI are the most conserved residues of group II introns. DIV is the domain that contains the IEP ORF (Lambowitz & Zimmerly, 2004). Secondary structure of group II introns can be further subdivided into three subgroups namely IIA, IIB and IIC (Michel et al, 1989b; Toor et al, 2001) **(Figure 17A)**.

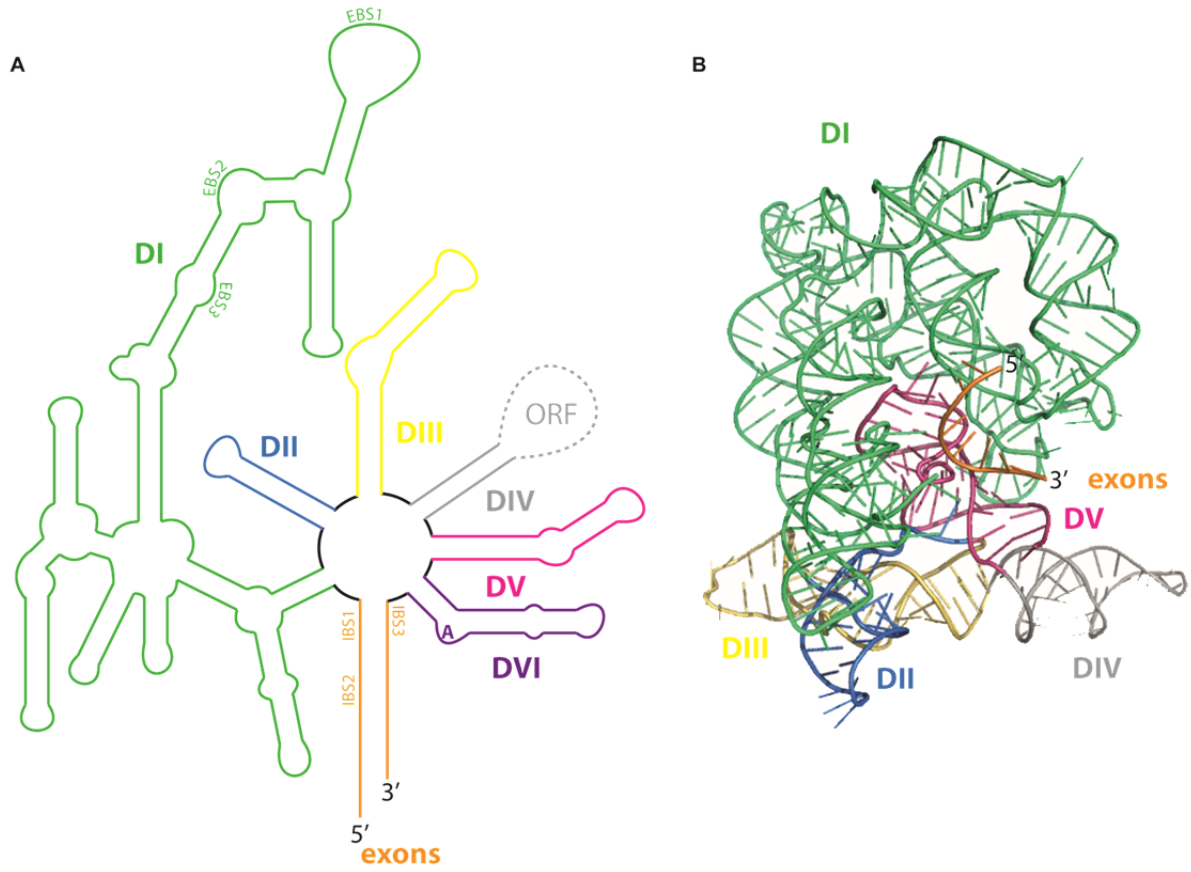


Figure 17: Group II introns secondary structure diagram (A) and 3D structure (B). In green is shown the largest domain DI harboring exons binding sites (EBS 1 to 3). DIV (in grey) can contain an ORF (grey dashes) in its apical loop. In the crystal structure DIV I present as a simple stem loop. DVI (purple) containing the reactive and highly conserved adenosine is missing in the crystal structure. Exons are shown in orange.

b) Group II introns' three-dimensional structure

In 2008 the Pyle lab has published the first crystal structures of a group IIC intron from the deep-sea bacterium *Oceanobacillus iheyensis (Toor et al, 2008)*. This Group IIC intron has been chosen for crystallization because it is smaller than most studied group IIA and IIB introns. At present, seventeen crystal structures of the same group IIC spliced intron are available each of them displays different features of group II introns, which have been reviewed in detail this year (Marcia et al, 2013). Succinctly, in the crystal structures, the DI domain wraps around the other domains to scaffold the ribozyme core. The DIV domain, herein replaced by a hairpin, points outward allowing for DIV to accommodate an ORF without impairing the intron folding. DV, which forms the catalytic site, is buried in the ribozyme core.

Unfortunately, in all available structures, the DVI domain containing the bulged adenosine is always missing due to hydrolysis (Toor et al, 2010) **(Figure 17B)**.

c) Group II introns and the lariat-capping ribozymes

The common feature between group II introns and LCrz is the nature of the branching reaction which leads to the formation of a 2',5' phosphate bond closing an RNA lariat (Pyle, 2005) **(Figure 23 and 25)**. However, if the reaction catalyzed by both ribozymes is similar from a chemical point of view, all other features such as, nucleotides implicated in branching, size of the lariat, global 2D and 3D structures, biological role and context etc. are specific to each ribozyme. This implies that in the course of evolution different RNAs have been selected to catalyze the same reaction. This situation is similar to what is observed for small ribozymes (see I.A.5).

4. Ribonuclease P

RNase P is a large ribonucleoprotein complex observed in the three domains of life (Woese et al, 1990) and is composed of one catalytic RNA (Guerrier-Takada et al, 1983) and one or several proteins. The catalytic RNA is homologous in Bacteria, Archaea and Eukarya but the number of protein subunits varies (Frank & Pace, 1998). The bacterial RNase P has only one ~14kDa (12 to 15kDa) protein while archaeal RNase P counts four protein subunits and eukaryotic RNase P has from six to twelve protein subunits (Hartmann & Hartmann, 2003). Interestingly and as mentioned in the introduction, proteinaceous RNase Ps (PRORP) exist. These PRORP are observed in mitochondria, chloroplasts and nucleus of *Arabidopsis thaliana* (Gutmann et al, 2012) as well as in human mitochondria (Holzmann et al, 2008). However in this manuscript I will only discuss the bacterial RNA-based RNase P from type A (Ellis & Brown, 2009).

The major RNase P role is to process the 5' leader sequence of pre-tRNAs (Robertson et al, 1972). As a ribozyme, the RNA subunit of RNase P bears the catalytic activity (Guerrier-Takada et al, 1983; Kikovska et al, 2007; Pannucci et al,

1999). However the catalytic RNA dependence to protein subunits for efficient processing of the tRNA is increasing with the number of protein subunits in ribonucleoprotein complex. Noteworthy, RNase P is also responsible for, e.g., 4.5S processing (Peck-Miller & Altman, 1991), tmRNAs maturation (Komine et al, 1994) and cleavage of transient structures of *E. coli* and *B. subtilis* riboswitches (Altman et al, 2005).

a) RNase P's secondary structure

The secondary structure of the type A RNase P can be subdivided into two sets of nucleotides: (1) the nucleotides forming the highly conserved core of the RNase P RNA common to all five structural classes of RNase P and (2) the nucleotides forming additional domains specific of each structural class. Type A RNase P is present in Bacteria as well as in Archaea pointing to a probable common ancestor.

The core domain has been defined from comparative sequence analysis of RNase P RNA from phylogenetically diverse bacteria (Haas & Brown, 1998). This minimal core domain has been proven to be active *in vitro* (Siegel et al, 1996). It contains the five conserved regions where most of the highly conserved residues are observed and the A type helical domain P1, P2, P3, P4, P9, P10 and P11 **(Figure 18)**. The secondary structure of this minimal core is shared by RNase P RNAs of the three domains of life and the conserved regions show high covariation rates. This domain is also designated as C-domain (catalytic domain) as it contains the residues responsible for catalysis.

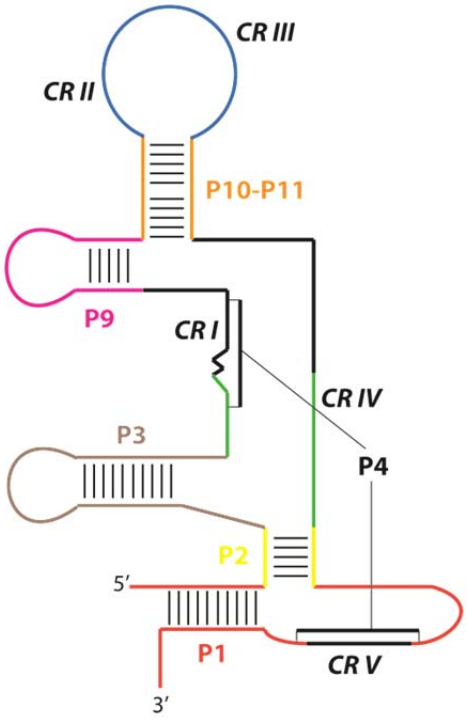


Figure 18: Secondary structure diagram of the consensus structure among RNase P from Bacteria, Archaea and Eukaryotes. The five universally conserved sequence regions are labeled CRI-V. Regions that are colored have homologous structural elements in all three domains of life and regions in black are specific unique element for each phylogenetic group.

About half of the nucleotides of the RNase P RNA form additional domains that stabilise the core. Some of which are shared between A and B type bacterial RNase P (Waugh et al, 1989). This is the case for the loop of P15 implicated in the recognition of the 3'CCA tail of the pre-tRNA and the helices P7, P8, P9, P10 interacting with the T-loop of the substrate pre-tRNA (Harris et al, 1994; Knap et al, 1990). One specific feature of the type A RNase P RNA is the P6 pseudoknot formed by base pairing between residues of the apical loop of P17 with the bulged nucleotides between P5 and P7 **(Figure 19)**. Nucleotides downstream and upstream from the P7 stem define the S-domain (specificity domain) that is responsible for substrate recognition.

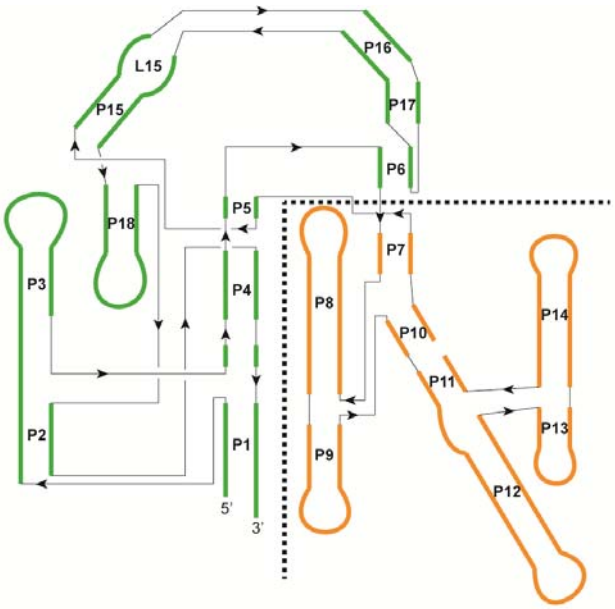


Figure 19: Secondary structure diagram of the bacterial RNase P (Figure adapted from (Reiter et al, 2010)) where all stem-loops belonging to the catalytic domain are shown in green and all stem-loop belonging to the substrate domain in orange.

b) RNase P's three-dimentional structure

The 3.8 Å resolution crystal structure of the RNase P holoenenzyme in complex with a tRNA has been solved recently (Reiter et al, 2010). This first low-resolution structure confirms and details the interaction between the holoenzyme and the pre-tRNA. Indeed, nucleotides G19 and C56 from the pre-tRNA D and T loop interact with nucleotides A112 and G147 from the RNase P RNA, respectively. The 3' $ACCA_{76}$ tail of the pre-tRNA is forming three WC base pairs with nucleotides in L15 of the RNase P RNA and one weak interaction is observed between A76 (pre-tRNA) and G253 (RNA from RNase P). The crystal structure reinforces the hypothesis that the protein subunit of RNase P has a role in recognition and positioning of the 5' leader sequence (Koutmou et al, 2010; Niranjanakumari et al, 1998). Coaxial stacks of helices are: (1) P1/P4/P5, (2) P2/P3, (3) P6/P15/P16, (4) P7/P10/P11/P12, (5) P8/P9 and P13/P14 **(Figure 19 et 20)**. More interestingly, this structure reveals the RNA motifs adopted by the RNA junctions bridging helices. The J16/17 junction has been of particular interest for us and will be detailed in the result section of this manuscript.

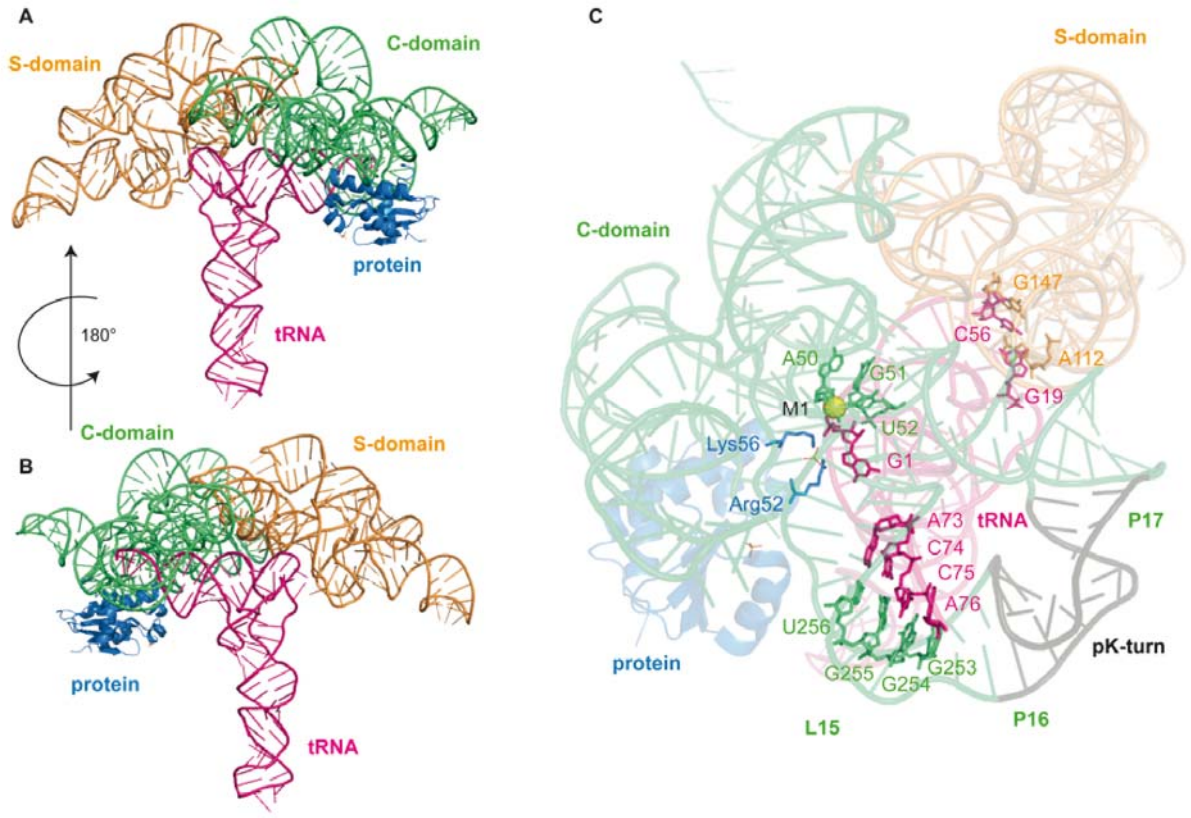


Figure 20: Three-dimensional structure of the type A bacterial RNase P holoenzyme bound to a substrate tRNA (A). (B) is the same as (A) after a 180° rotation along the y axis. (C) is a zoom in the structure after a rotation of 90° along the y axis of view (A). Details shown are, contacts between the tRNA (T and D loop, 5'CCA3' tail and at the 5' maturation site) and the RNase P. The pk-turn motif between P16 and P17 is shown in grey. As in Figure 19, the C-domain is colored in green and the S-domain in orange. The tRNA is shown in pink and the protein component of RNase P in blue.

5. Small ribozymes

Six small ribozyme families are constituted by the hammerhead (Prody et al, 1986), hairpin (Buzayan et al, 1986), hepatitis delta virus (Sharmeen et al, 1988), Varkud satellite (Saville & Collins, 1990), *glmS* (Winkler et al, 2004) and twister (work presented by Breaker lab at the RNA meeting 2013). Their sizes vary from 50 to 150 nucleotides and despite their very diverse 3D structures **(Figures 21),** they all catalyse the same 2'-O transesterification cleavage reaction described in I.B.1.c.

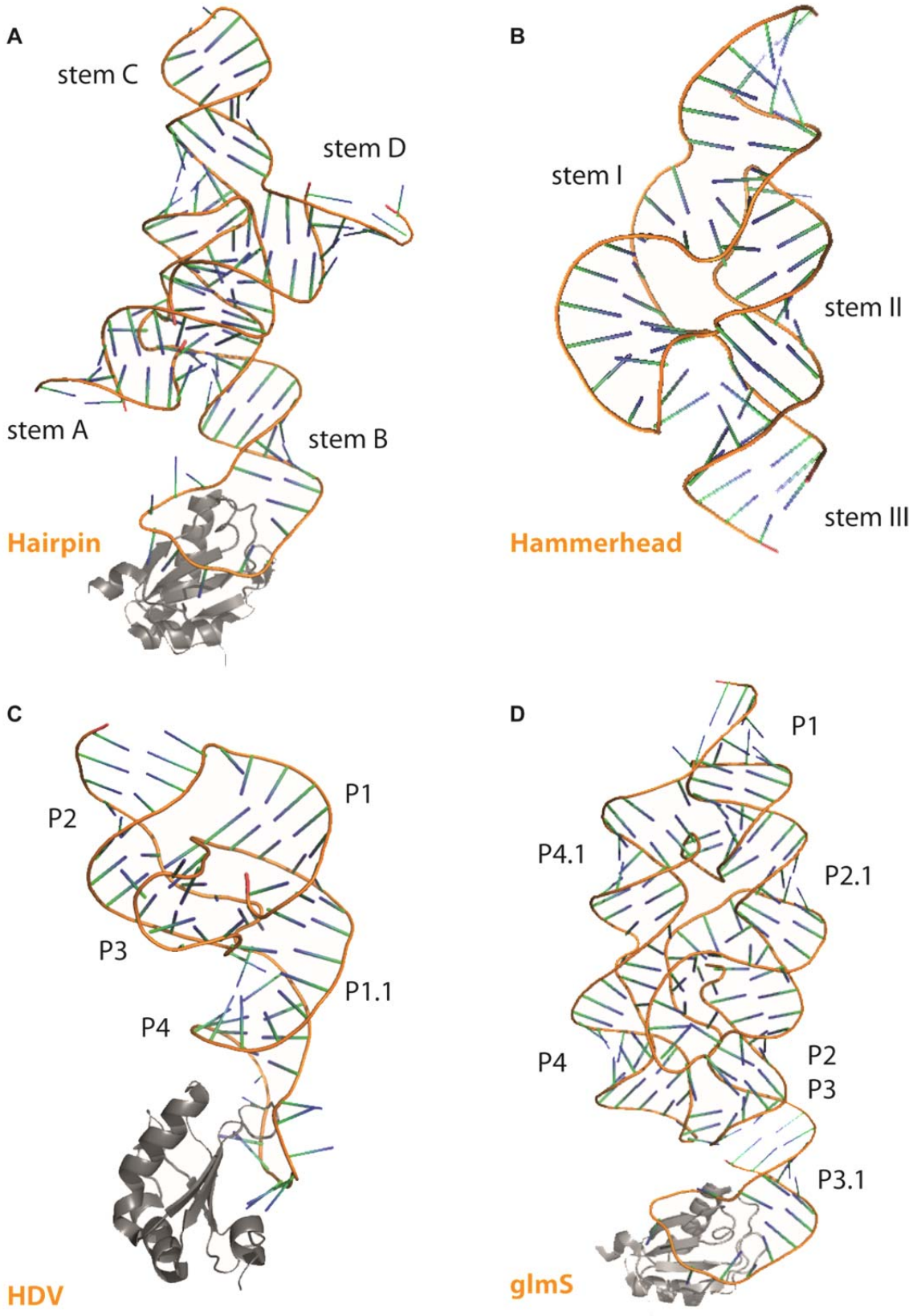


Figure 21: Crystal structures of the hairpin (Rupert & Ferre-D'Amare, 2001), hammerhead (Martick & Scott, 2006), HDV (Ke et al, 2004) and glmS (Cochrane et al, 2009) ribozymes represented in cartoon mode. U1A spliceosomal protein used for crystallization in shown in grey.

a) Biological roles of small ribozymes

HMH and hairpin ribozymes have been observed for the first time in the sense and antisense strands of the satellite RNA of the tobacco ring spot virus and in RNA transcripts of avocado sunblotch viroid. HMH and hairpin are responsible for the separation into monomeric segments of the concatamer product resulting from the rolling-circle replication undergone by the satellite RNA (Buzayan et al, 1986; Hutchins et al, 1986; Prody et al, 1986). Likewise HDV and VS ribozymes have been found in the satellite RNA virus of human pathogen *hepatitis* B (Sharmeen et al, 1988) and mitochondrial genome of *Neurospora crassa,* respectively (Saville & Collins, 1990). Both, HDV and VS ribozymes, cleave the concatameric RNA transcript in which they are inserted (Kennell et al, 1995; Perrotta & Been, 1991). But small-ribozymes are not restricted to satellite RNA. Indeed, the *glmS* ribozyme is located in the 5'UTR of the mRNA encoding the essential enzyme glucosamine-6-phosphate (GlcN6P) synthase in some Gram-positive bacteria. *glmS* is not only a ribozyme but also a riboswitch that needs its co-factor, GlcN6P, to accomplish efficient self-cleavage (Winkler et al, 2004). Moreover, the increasing amount of genomic data and well-suited bioinformatic tools, show that small-ribozymes are ubiquitous in bacteria and eukaryotes (Bourdeau et al, 1999; Ferbeyre et al, 1998; Martick et al, 2008; Webb et al, 2009).

b) Small ribozymes and the lariat-capping ribozymes

One common feature between hammerhead ribozyme and LCrz is the crucial importance of distal tertiary contacts having major effect on the chemical reaction catalysed by each ribozyme. In the case of the hammerhead ribozyme, a minimal version has been firstly proven active *in vitro* and crystallized (Pley et al, 1994; Uhlenbeck, 1987). However, twelve years later the crystal structure of a full-length hammerhead ribozyme revealed a tertiary interaction between the bulge of stem I and the apical loop of stem II (Martick & Scott, 2006) that improves cleavage by a 1000-fold in comparison to the minimal version of the hammerhead ribozyme (Canny et al, 2004). In the case of LCrz the peripheral domain DP2-DP2.1 is important as its conformation determines its on or off conformation (see I.B.1.a and **Figure 24**).

Indeed alternative conformations of HEG P1 and DP2 are observed exclusively in the inactive and active forms of LCrz respectively (see I.A.1.c and **Figure 6**). Moreover, the crystal structure of LCrz revealed that the peripheral domain acts as a clamp tightening the ribozyme core (see Results section).

Each small-ribozyme has been extensively studied ((Lilley & Eckstein, 2008) and (Ferre-D'Amare & Scott, 2010) for review). Therefore it is now possible to use them as powerful tools in molecular biology (Price et al, 1995). During my PhD, I used hammerhead and HDV ribozymes to obtain LCrz to be used for crystallography with well-defined 5' and 3' ends, respectively. This process is detailed in II.A.1.b.

B. CHEMICAL REACTIONS CATALYSED BY RIBOZYMES

The range of natural chemical reactions catalyzed by ribozymes is limited to transesterification reactions and hydrolysis, but depending on the structural environment the resulting products are diverse **(Figure 22)**. One exception is the ribosome in which rRNA catalyses peptidyl transfer.

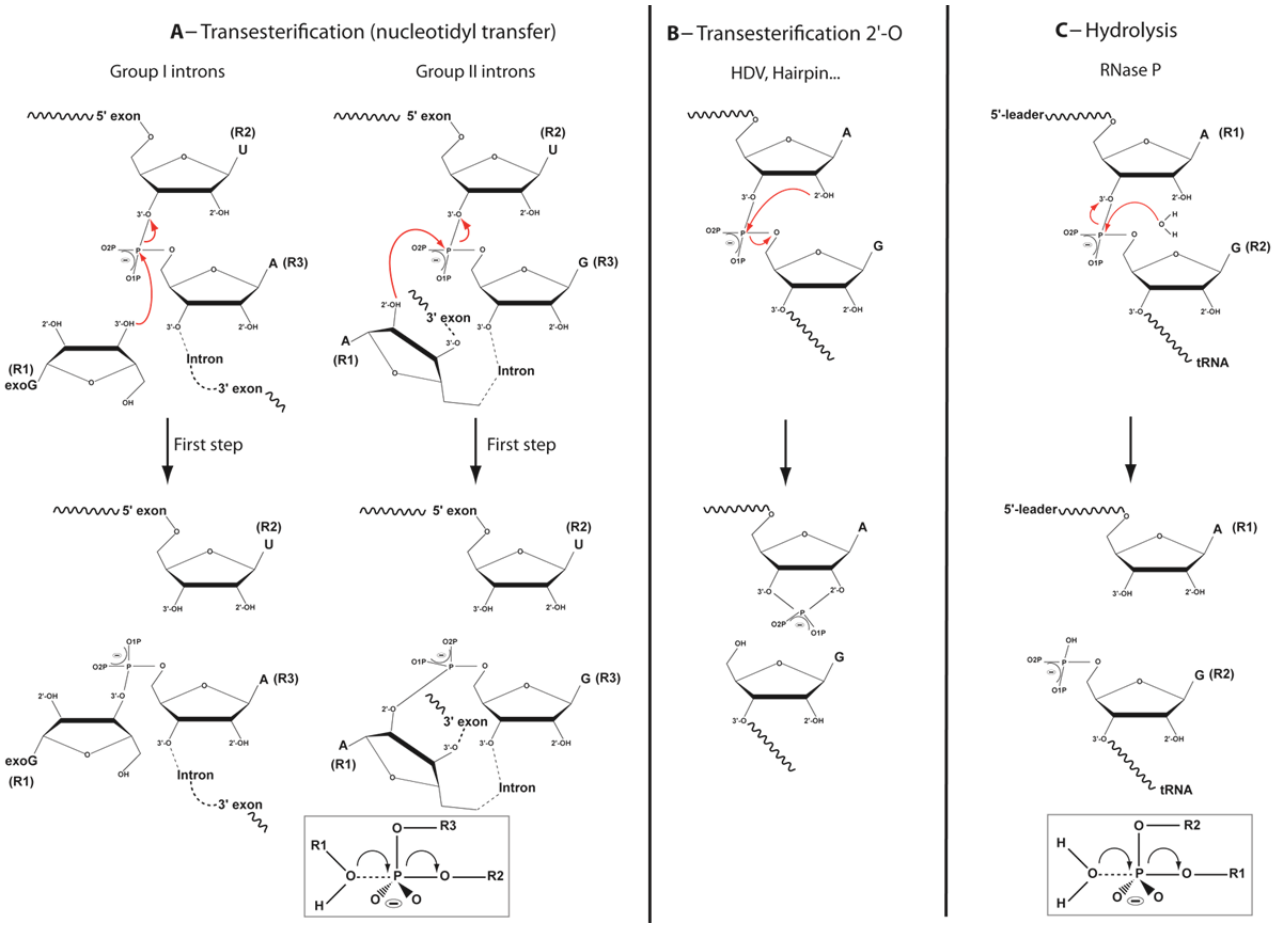


Figure 22: Reactions catalyzed by known naturally occurring ribozymes. (A) Transesterification reaction that occurs during the first step of splicing by group I and II introns is chemically identical. Differences are: (1) the nucleophile in group I intron is provided by an exogenous guanosine (exoG) and is induced by the 3'O when the nucleophile in group II intron is embedded in the intron's sequence and is induced by the 2'O and (2) the 3' product of the first step of splicing is a linear group I intron with exoG bound at the 5' end when group II form a lariat closed by a 2'5' phosphate bond. (B) Transesterification reaction catalyzed by all small naturally occurring ribozymes. It involves attack of the 2'O on the adjacent 3'P leading to the release of a 5' product with a 2'3' cyclic phosphate 3'end and a 3' product with a 5'OH 5'end. (C) Hydrolysis catalyzed by RNase P is triggered by the nucleophilic attack of a water molecule onto the scissile phosphate between the 5' leader sequence of the tRNA and the tRNA. The resulting products are 5'P 5'end for the 3' product and à 3'OH 3'end for the 5' product.

1. Transesterification

a) Branching reaction

LCrz characteristic *in vivo* reaction is the branching reaction leading to the formation of a tiny lariat of three nucleotides ($C_{230}AU$) at the 5' end of the 3' product, the HE pre-mRNA, and a 3'OH end (G_{229}) at the 3' end of the ribozyme catalytic core. The reaction is initiated by a nucleophilic attack of the 2'OH group of U232 at the BP onto the 5'P of C230 at the IPS. A 2',5' phosphodiester bond closes the three nucleotides lariat (Nielsen et al, 2005) **(Figure 23)**.

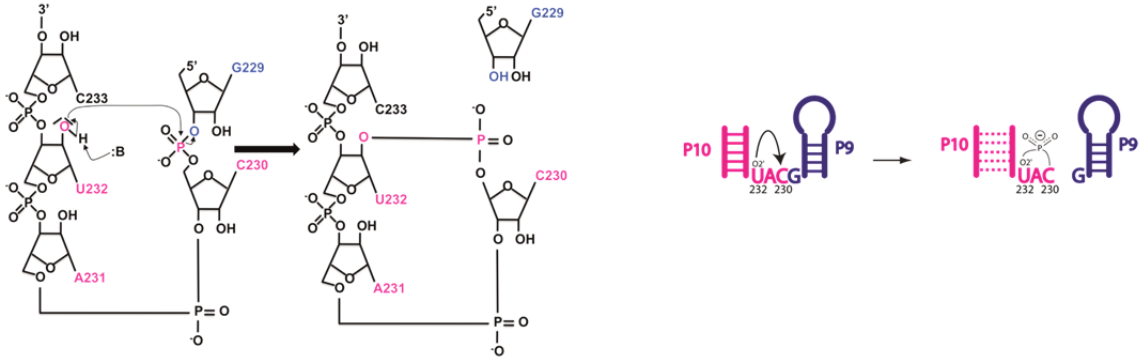


Figure 23: The branching reaction catalyzed by LCrz is triggered by the nucleophilic attack of the 2'O from U232 onto the close in space 5'P of C230. The resulting 5' product is a 3'OH 3'end (blue) and the resulting 3' product is a lariat of three nucleotides closed by a 2'5' phosphate bond (pink).

In vitro, the reverse reaction of branching, ligation is observed and can completely mask the branching reaction (Nielsen et al, 2009; Nielsen et al, 2005). LCrz can also catalyse irreversible hydrolysis at IPS (Decatur, 1995; Johansen & Vogt, 1994). Interestingly the length of the 5' and 3' flanking regions influence the ratio between the three LCrz-catalysed reactions (Einvik et al, 2000; Nielsen et al, 2009) **(Figure 24)**.

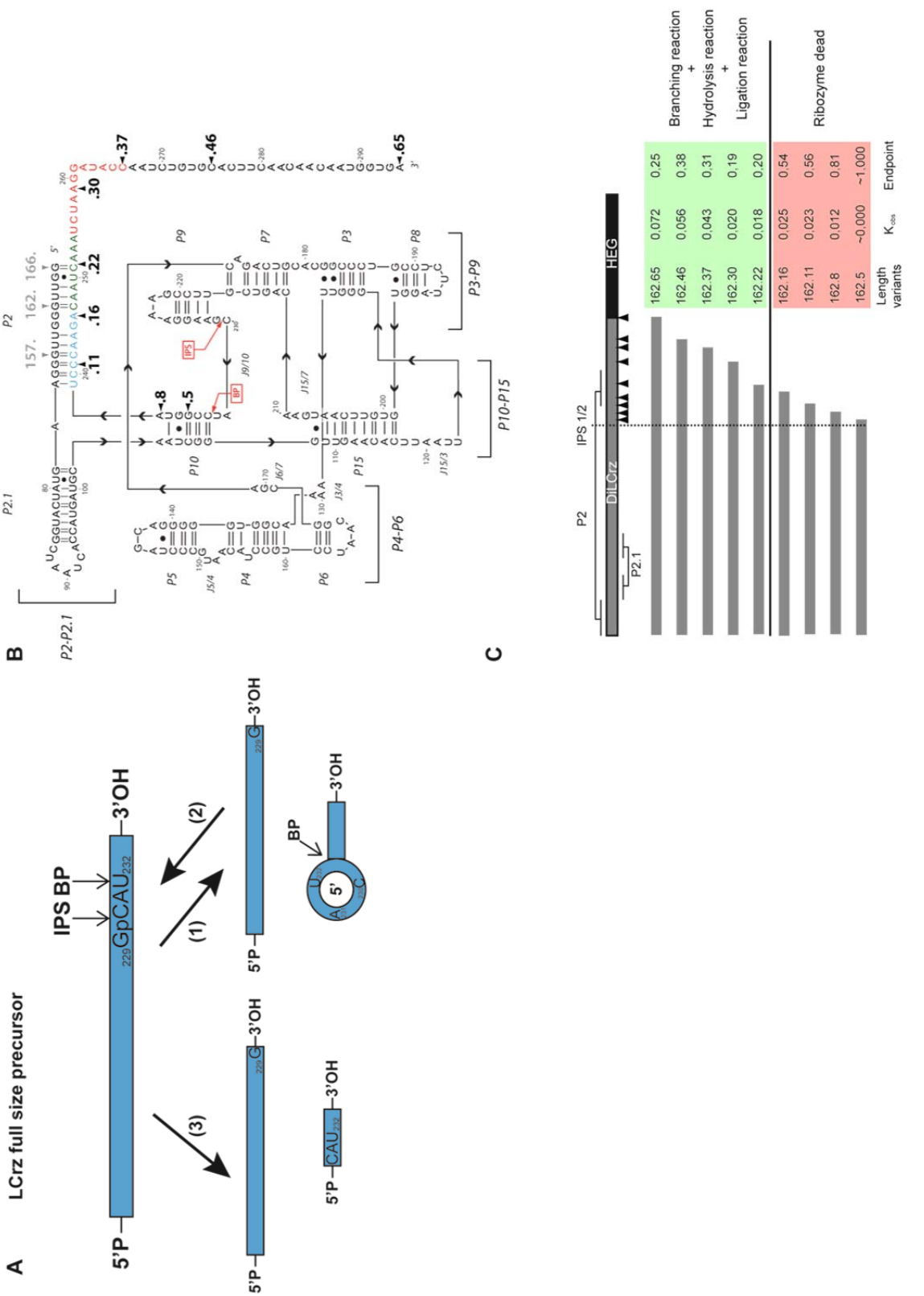


Figure 24: (A) The three types of reaction that LCrz can undertake (1) is branching, (2) ligation and (3) irreversible hydrolysis. (B) Secondary structure diagram of LCrz harboring the DP2/DP2.1 peripheral domain on which systematic deletion has been performed to test the number of nucleotides required of branching or hydrolysis. Numbers pointing at arrows indicate the distance from the IPS in the number of nts. (C) Schematic diagram showing effect on the three reactions that LCrz can undertake depending on the length of its 3'end.

LCrz branching reaction resembles the first step of splicing mediated by group II introns. Group II introns branching reaction is initiated by the nucleophilic attack of the 2'OH group of the highly conserved adenosine of DVI onto the 5' splice site. This first step allows for the release of the 5' exon and leads to the formation of a circularised intron. In the second step of splicing, the 3' end of the 5' exon attacks the 3' splice site consequently joining the exons and releasing the lariat intron (Daniels et al, 1996; Jarrell et al, 1988) **(Figure 22 and 25)**.

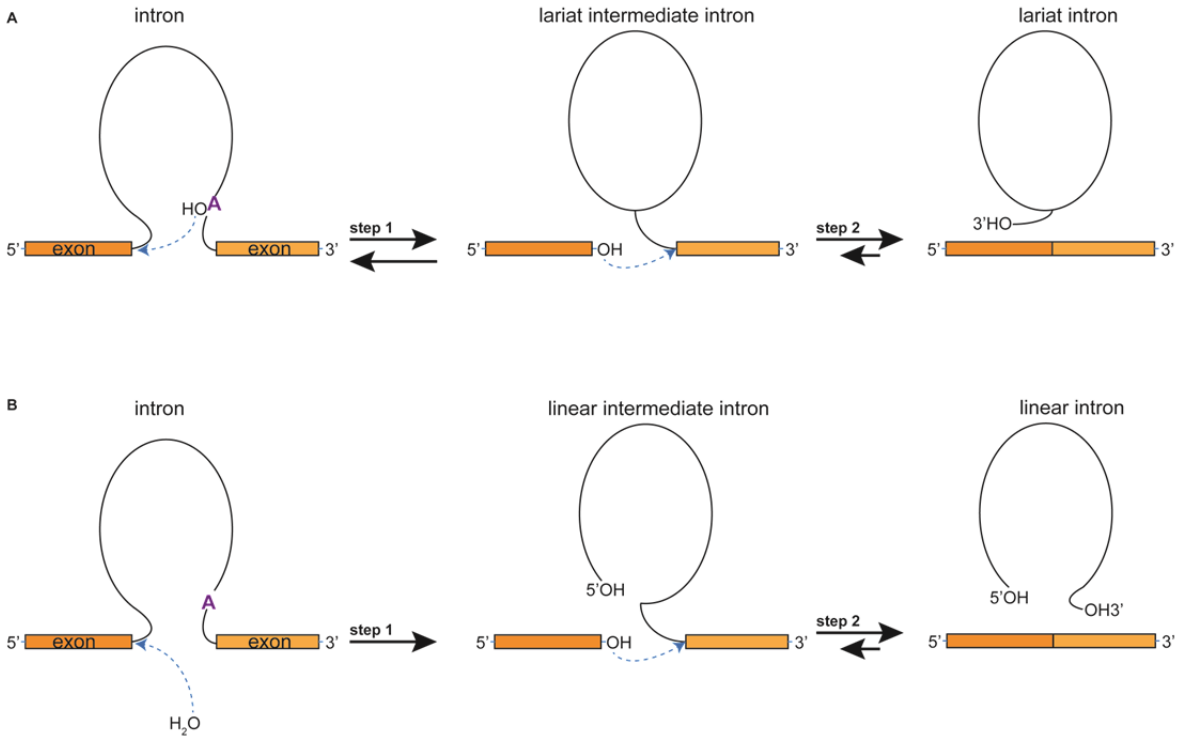


Figure 25: The two pathways that group II introns can undertake for splicing. (A) Step 1: The 2'O of the highly conserved adenosine from DVI domain attacks the 5' splice site. The 5' exon is released with 3'OH 3'end and the intron is circularized. Step (2) the 3'O of the 5' exon attacks the 3' splice site. Exons are ligated and the lariat intron is freed. Both steps are highly reversible. This pathway if reversed allows for group II introns reverse splicing. (B) Hydrolytic pathway in which a water molecule plays the role of the highly conserved adenosine. Resulting intron is linear and no reverse splicing is possible. (Figure adapted from (Pyle, 2008)).

b) Nucleotidyl transfer

Self-splicing is the result of two consecutive phosphotransesterification reactions allowing ligation of the flanking exons and excision of the intron. In the first step of self-splicing, the first stage consists in the recognition of the 5' SS by the IGS therefore forming P1. The last nucleotide of the 5' exon is the uridine that forms the conserved wobble G.U base pair at the end of P1. Then, the 3' hydroxyl group of an exogenous guanosine, exoG, bound into the G-binding pocket of P7 attacks the scissile phosphate right downstream from the uridine at the 5'SS. As a result, exoG becomes the first nucleotide at the 5' end of the intron and the 3' end of the 5' exon is freed (Cech et al, 1981; Zaug & Cech, 1982). After a conformational rearrangement of the catalytic site to release exoG and accommodate ωG in the G-binding pocket the second step of self-splicing takes place (Been & Cech, 1986). ωG is a conserved nucleotide of group I intron at its 3' end. The free 3' hydroxyl terminus of the 5' exon then attacks the scissile phosphate downstream from ωG leading to the ligation of the flanking exons and to the release of the intron with ωG at its 3' end **(Figure 26)**.

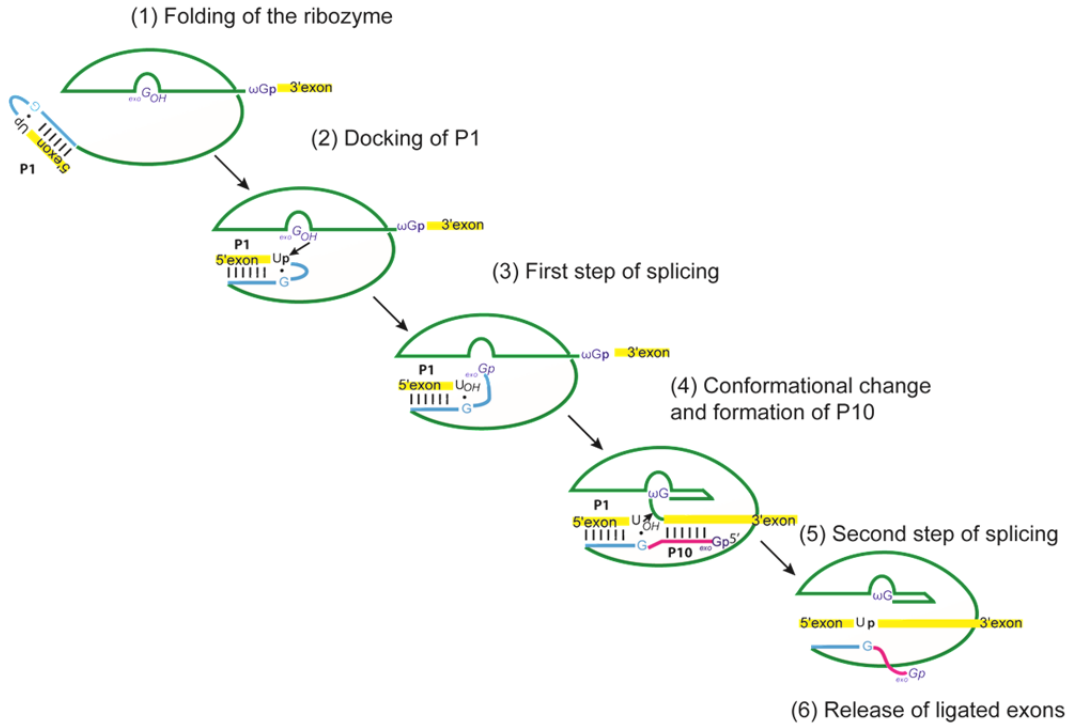


Figure 26: Diagram of the two steps of splicing performed by group I introns. The first step of splicing (3) is triggered by an exogenous guanosine (exoG) when the second step (5) of splicing triggered by the uridine involved in the G.U wobble base pair of P1and belonging to the 3'end of the 5'exon. (Figure adapted from (Golden, 2008)).

Metal ions surely play a role in RNA folding but also in catalysis by either triggering a nucleophilic attack, stabilizing a leaving group or organizing critical nucleotides in the catalytic site (Johnson-Buck et al, 2011; Steitz & Steitz, 1993). In the case of group I introns, different biomolecular approaches allow for the identification of three magnesium ions in the active site of the *Tetrahymena* intron (Hougland et al, 2005; Shan et al, 2001; Szewczak et al, 2002) **(Figure panel 27 A)**. Crystal structures discussed in I.A.2.c confirmed the presence of only two out of three metal ions at the active site (Golden, 2008; Lipchock & Strobel, 2008; Stahley & Strobel, 2005) **(Figure panel 27 B and C)**.

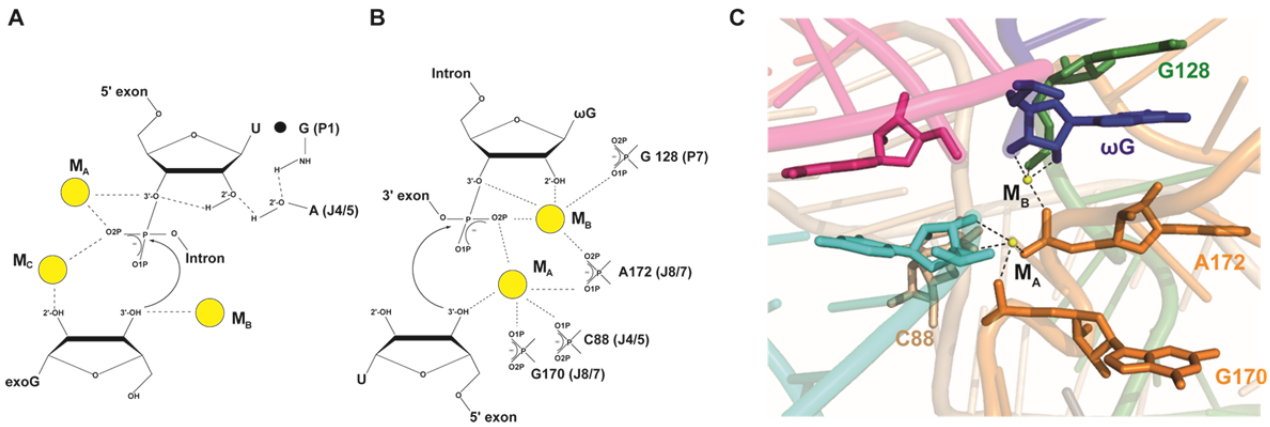


Figure 27: (A) Biochemical model for metal ion interactions in the active site of the *Tetrahymena* intron. Metal ions at the active site of *Azoarcus* group I intron and their interactions shown in a schematic view (B) and in the crystal structure (C). Hydrogen bonds are represented with dashed lines.

Does it mean that there are only two metal ions in the catalytic site of group I introns? It is not possible to rule out either model as, crystal structures give a snapshot of one specific state of the ribozyme when biochemical analysis in solution reflects more the malleability necessary to self-splicing of the ribozyme.

Self-splicing catalyzed by group II introns is, as for group I intron, a two steps process but the resulting products are different. The first step, also known as branching reaction, is described in I.B.1.a. The second step of splicing is the nucleophile attack of the 3' splice by the 3'OH terminus of the free 5' exon. This last step allows for the release of the lariat intron and ligated exons (Daniels et al, 1996; Jarrell et al, 1988) **(Figure 25 A)**. An alternative to this mechanism is the

replacement in the first step of splicing of the adenosine of DVI by a water molecule. Resulting products are ligated exons and a linear intron (Jarrell et al, 1988; van der Veen et al, 1987b) **(Figure 25 B)**.

Recent crystal structures (Chan et al, 2012; Marcia & Pyle, 2012) and previous biochemical studies confirm that two magnesium and two potassium ions are implicated in the self-splicing reaction (Gordon & Piccirilli, 2001; Marcia & Pyle, 2012; Sigel et al, 2000) **(Figure 28)**. Moreover as group II introns have only one catalytic site it should undergo structural rearrangement to allow the two consecutive S_N2 reactions to happen (Chanfreau & Jacquier, 1996; Marcia et al, 2013).

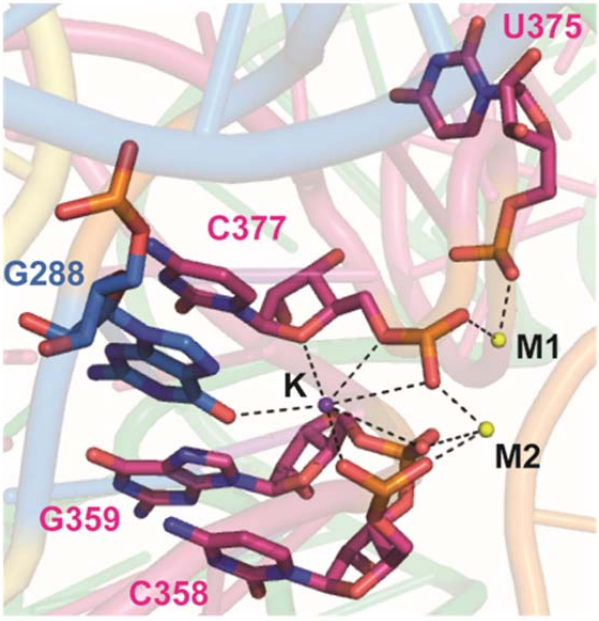


Figure 28: Metal ion cluster at the active site of group II intron from *Oceanobacillus iheyensis*. The three metal ions are interconnected via residues of the intron. Potassium ion is displayed in violet and magnesium in yellow. Hydrogen bonds are shown as dashed lines.

c) Transesterification 2'-O

The transesterification 2'-O reaction follows a S_N2 mechanism during which the oxygen atom at position 2' of the nucleotide N attacks the 5' phosphate of the 3' nucleotide N+1 leading to the formation a 2',3'-cyclic-phosphate 5' product and a 5' hydroxyl group on the 3' product **(Figure 29)**. For this reaction to occur the 2'-oxygen, the 5'-phosphate and the 5'-oxygen have to be aligned. Moreover, to trigger

the reaction, a general base should remove the proton from the 2' hydroxyl group of nucleotide N and a general acid (Kath-Schorr et al, 2012; Lilley, 2011) should protonate the 5' oxygen of nucleotide N+1 to facilitate the release of the leaving group. Roles of general acid and general base are played by nucleotides.

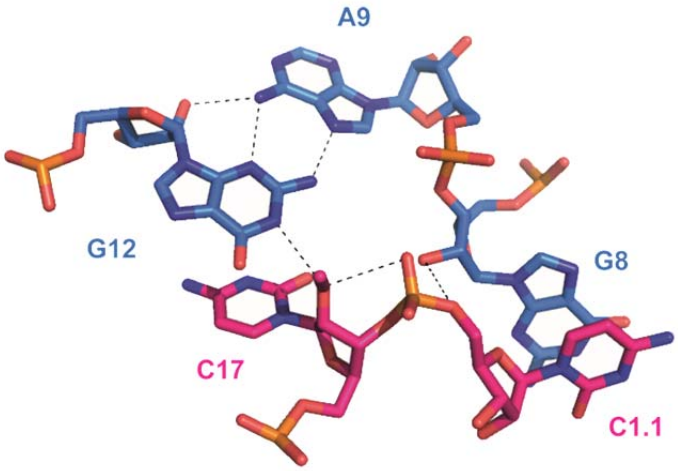


Figure 29: Detailed three-dimensional structure of the hammerhead active site with nucleotide N and N+1 being C17 and C1.1 respectively (pink). A methyl group replaces the 2'O from C17 to prevent cleavage during crystallization. The general base is G12 while the general acid is G8 (blue).

The transesterification 2'-O is common to all small-ribozymes however, the nucleotide that acts as a general base or a general acid is specific to each ribozyme. Likewise the N and N+1 nucleotides are specific to one ribozyme. In every case, the reverse reaction, ligation, can be observed.

2. Hydrolysis

Hydrolysis is the reaction catalysed by RNase P that allows the release of a 5' phosphate mature tRNA and a 3' hydroxyl 5' leader sequence (Altman & Smith, 1971). Remarkably RNase P cleaves its substrates in *trans* and is capable of multiple turnover. Hydrolysis by RNase P RNA is magnesium dependant and suggests an S_N2 -like mechanism (Smith et al, 1992). Nucleotides from the RNase P RNA involved in catalysis are A50, G51 and U52 all located in the P4 stem. The recent crystal structure of the type A RNase P revealed two metal ions at the active site when the pre-cleavage state of the pre-tRNA is mimicked (Reiter et al, 2010) **(Figure 30)**.

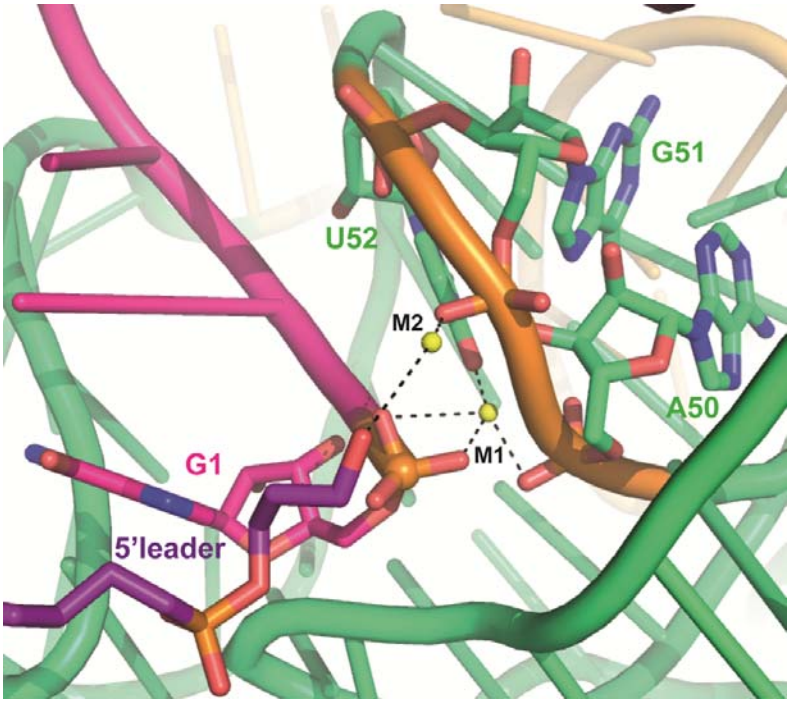


Figure 30: Two metal ions (yellow) at the active site of the type A bacterial RNase P holoenzyme bound to a mature tRNA substrate (pink) and a 5' tRNA leader sequence mimic (violet). RNase P catalytic domain is shown in green and hydrogen bonds in black dashed lines.

MATERIAL & METHODS

II. MATERIAL AND METHODS

A. CRYSTALLIZATION OF THE LARIAT-CAPPING RIBOZYME

1. Construct design

a) DiLCrz with substrate *in trans*

First constructs of the LCrz have been designed with sequences of DiLCrz, which is the most extensively studied LCrz out of the three types known. The DiLCrz 166.22 sequence that includes 166 nucleotides upstream the IPS and 22 nucleotides downstream the IPS has been chosen because it almost exclusively performs the branching reaction **(Figure 24)** (Nielsen et al. Science, 2005). It harbors the peripheral domain DP2-DP2.1 but the alternative hairpin HEG P1 cannot be formed (Nielsen, RNA, 2009). Moreover this sequence is the shortest containing all nucleotides necessary for the branching reaction to take place. The earliest forms of the DiLCrz designed for crystallization were composite with a core ending in the L9 loop and a substrate, added in *trans*, encompassing the 3' strands of P9, P10 and DP2. In order to present an active-like catalytic site with impeded activity on both hydrolysis and branching the substrate strand of 29 nucleotides harbors a phosphorothioate modified C_{230} nucleotide **(Figure 31)**.

To obtain a well-defined 3’ end to overcome the random addition of 3' nucleotides by T7 run-off transcription, an HDV ribozyme is added downstream DiLCrz sequence (Price et al, 1995b; Schurer et al, 2002). The 85 nucleotides HDV ribozyme is placed to cleave right after the last nucleotide, A_{222}, of DiLCrz core **(Figure 31)**.

Lastly, to promote crystallization, the P8 loop of DiLCrz, predicted not to be involved in intramolecular contacts (Beckert et al, 2008), has been engineered in two ways: (1) by adding a binding site for the spliceosomal protein U1A (Ferre-D'Amare &

Doudna, 2000; Oubridge et al, 1994) and (2) by adding a tetraloop receptor motifs (TR) capped by means of a tetraloop (TL) (Ferre-D'Amare et al, 1998) **(Figure 31)**.

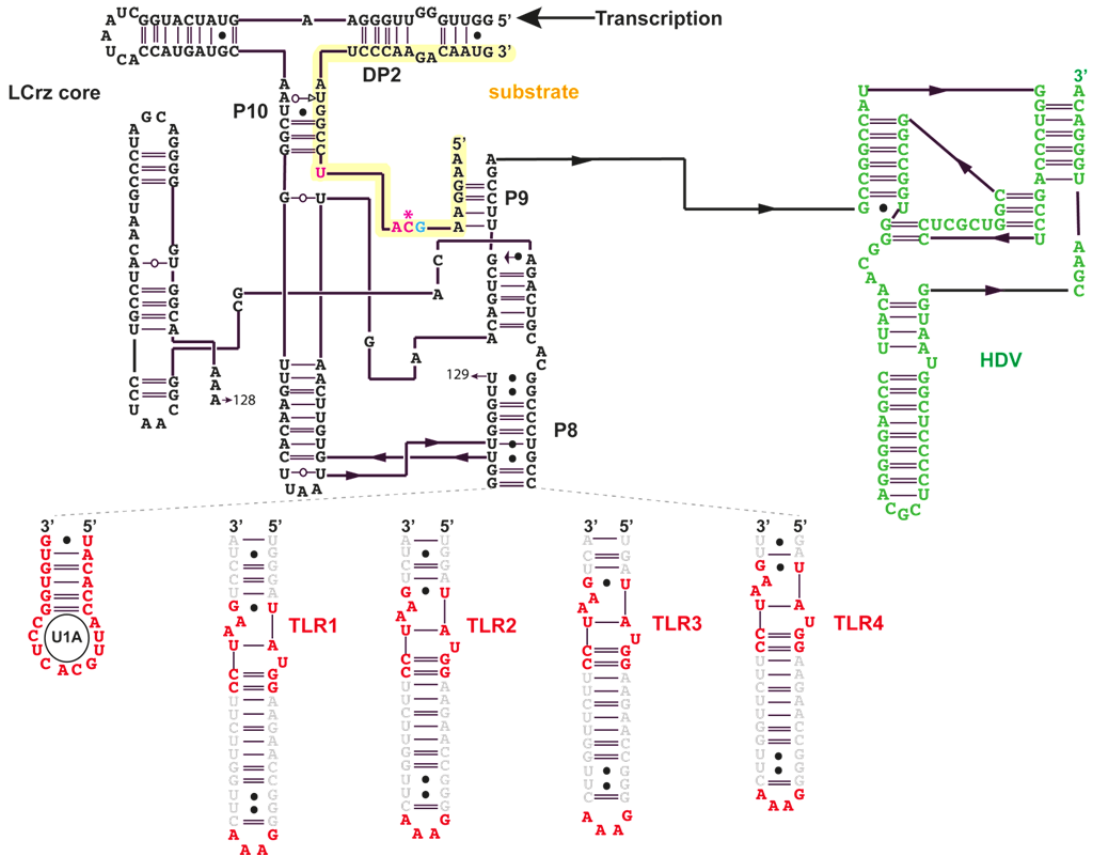


Figure 31: Diagram of the five DiLCrz constructs with the substrate *in trans*. *In vitro* transcription starts with the two guanosines at the 5' end of DP2 of the DiLCrz core shown in black. The 29 nucleotides substrate strand is highlighted in orange and the nucleotide C_{230} harboring a phosphorothioate group is designated by an asterisk. The three nucleotides forming the lariat are shown in pink and the guanosine G_{229} is shown in light blue. Five alternatives engineered P8 loop, U1A binding sequence or TL-TR motif, are shown in red. Spacer sequences between the ribozyme core and TR and between TR and TL are shown in grey. The HDV ribozyme added to avoid heterogeneous 3' end of DiLCrz due to *in vitro* run off transcription with the T7 RNA polymerase is shown in green.

b) Circularly permutated DiLCrz

The second type of LCrz constructs are circularly permutated (cp). In LCrz_cp the 5'- and 3' terminus correspond to residues (5'-OH)-C_{230} and (2',3'-cyclic monophosphate)-G_{229}, respectively. The DP2 stem is closed by a 5'UUCG3' tetraloop **(Figure 32)**. The scissile bond of the ribozyme is thus opened, preventing any kind of spontaneous cleavage. Since the circular permutation of the ribozyme is well folded, as observed by electrophoresis under native conditions, this design is

also expected to prevent RNA degradation. Noteworthy LCrz_cp constructs also allow preparing crystallization drops with a single homogenous macromolecule hence avoiding crystals containing only the U1A protein or crystal packing defects due to the substrate strand added in excess.

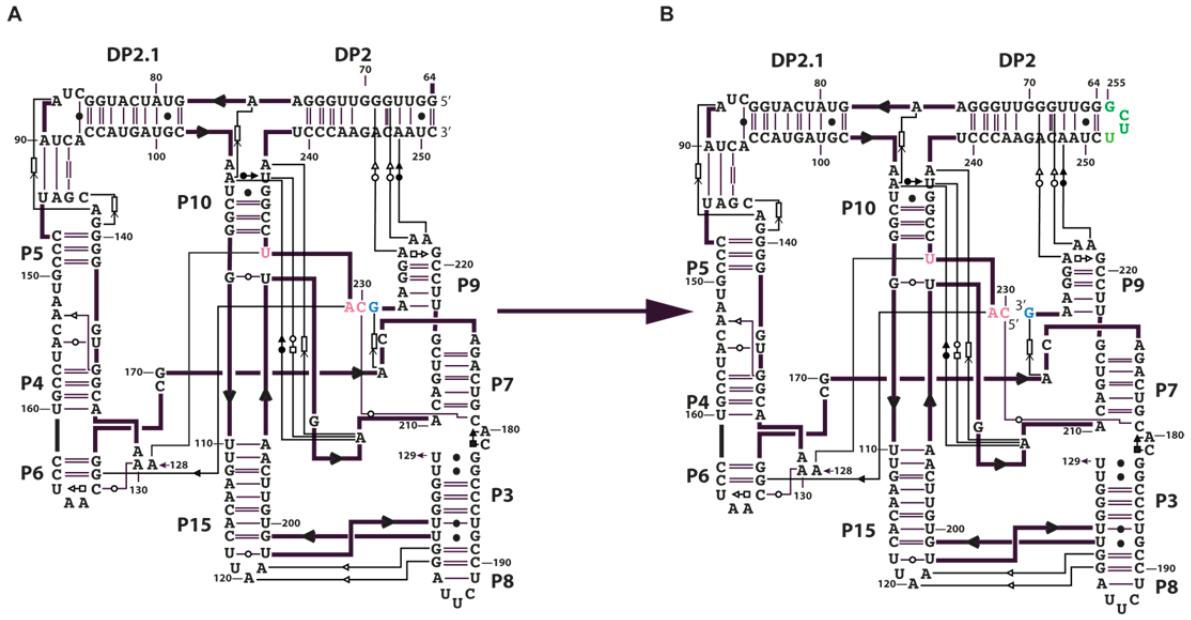


Figure 32: Design of the DiLCrz circular permutation. (A) Secondary structure of the 166.22 wild type sequence of the DiLCrz with the three nucleotides C_{230}, A_{231} and U_{232} forming the post-cleavage lariat in pink and G_{229} equivalent to group I ribozyme ωG in blue. The *in vitro* transcription start site is at G_{64} in DP2. The sequence is annotated as in the twintron. (B) Circular permutation construct of the DiLCrz (DiLCrz_cp) with the *in vitro* transcription start site at C_{230}. The DP2 stem is closed by a 5'UUCG3' tetraloop shown in green. Non-WC base pairs are shown in respect to the Leontis-Westhof nomenclature (Leontis & Westhof, 2001). Empty rectangles represent base stacking interactions.

One major drawback of the LCrz_cp constructs is that their sequences start with a cytosine and this feature is not compatible with efficient *in vitro* transcription using the T7 RNA polymerase. Instead, the sequence should start with at least two guanosines (Helm et al, 1999). To overcome this matter and obtain a well-defined 5' end required for crystallization (Price et al, 1995b) we inserted uptstream from the DiLCrz_cp sequence a minimal HMH (Scott et al, 1995b). However we noticed that minimal HMH co-transcriptional cleavage was very weak probably because of the downstream LCrz_cp RNA that folds rapidly and is highly structured therefore competing with HMH folding. We then designed some extended versions of the HMH ribozymes **(Figure 33)** retaining the important bulge-loop interaction made between stem I and II (Martick & Scott, 2006). Sequence for the HMH extended versions are

derived from naturally occurring HMH ribozymes (Khvorova et al, 2003) and take into account that the 3' strand of stem I has to be the sequence of DiLCrz. The preservation of the essential tertiary interaction is verified *in silico* using the S2S software (Jossinet & Westhof, 2005) **(Figure 34)**.

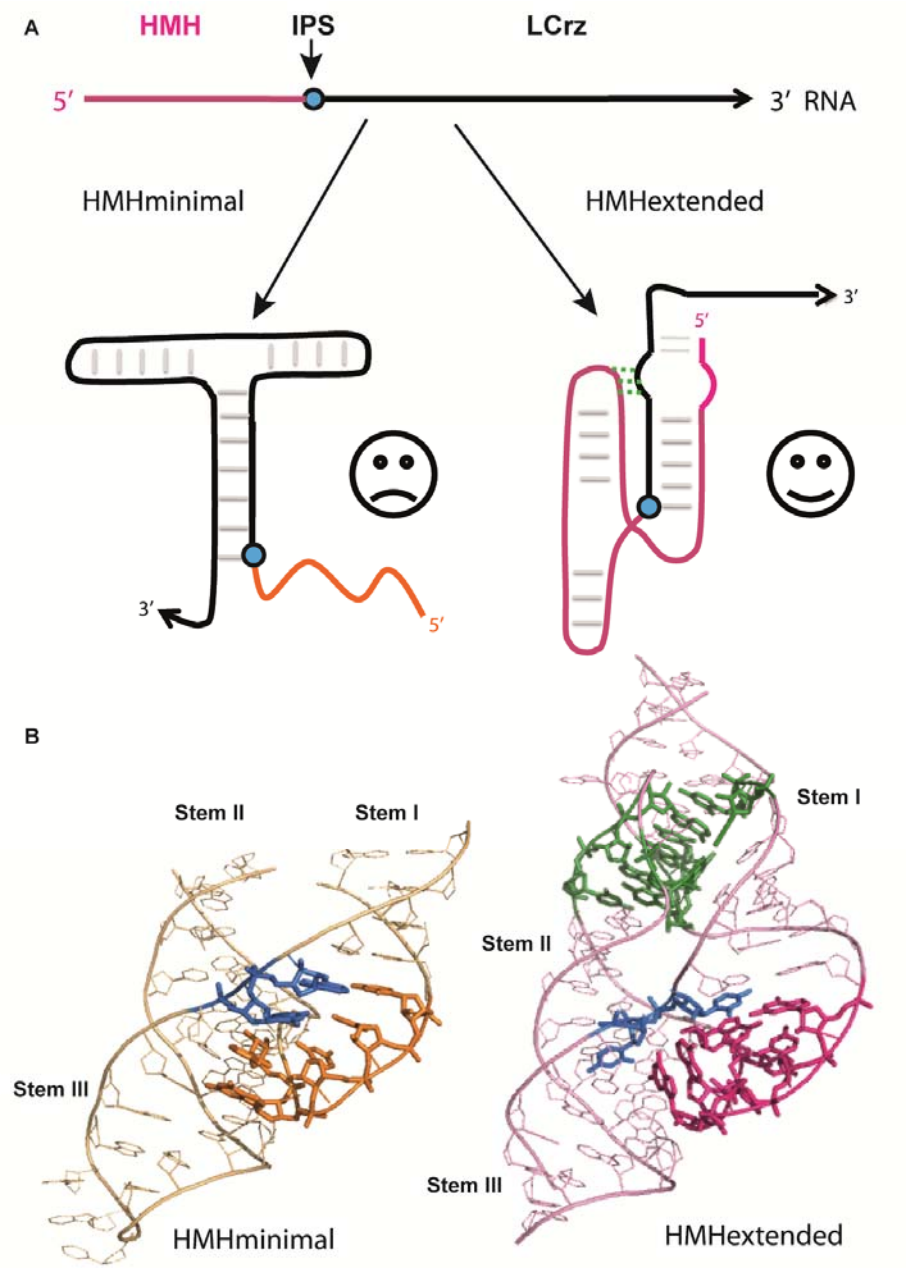


Figure 33: (A) Highly structured LCrz RNA (black) folding in the course of the transcription impairs folding and efficient cleavage of the minimal HMH ribozyme (orange) but not of the extended HMH (pink). The cleavage site for the HMH ribozyme is shown in blue. (B) Crystal structure of the full length version of the HMH ribozyme is shown in pink (Martick & Scott, 2006) with nucleotides implicated in the stem I – stem II tertiary interaction in green. Crystal structure of the minimal HMH is shown in orange (Scott et al, 1995b). For both HMH structures, nucleotides either side of the scissile phosphate bond are shown in blue and nucleotides positioned differently depending on the HMH sequences lengths are shown in pink sticks (full length HMH) or orange sticks (minimal HMH).

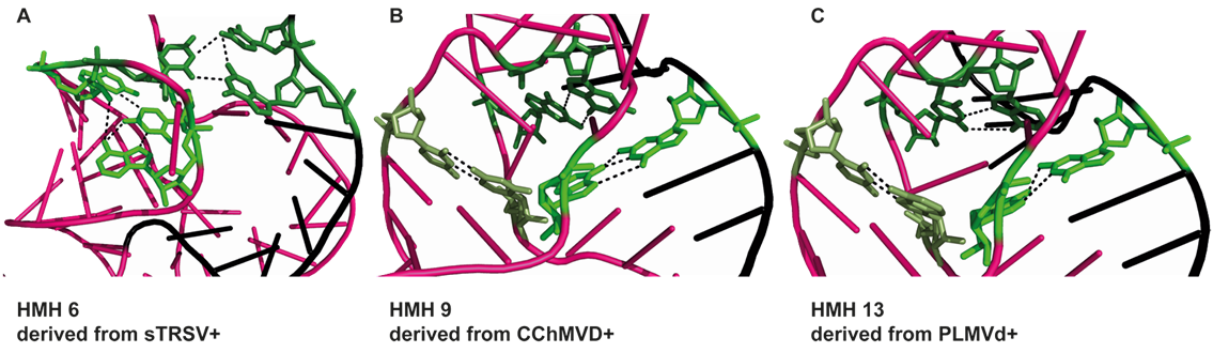


Figure 34: *In silico* predicted tertiary interactions (base pairs implicated are shown in green) between the internal bulge of stem I and the apical loop of stem II in extended versions of the HMH ribozyme (pink) from diverse origins bearing the 5' sequence of DiLCrz_cp (black) as the 3'strand of stem I.

Optimization of these extended versions of the HMH ribozyme lead to the publication of a book chapter "Cis-acting 5' hammerhead optimization for in vitro transcription of highly structured RNAs" for an edition of Methods in Molecular Biology in which we detail the procedure to adapt extended versions of the HMH ribozyme to any highly structured RNA **(Article 1)**.

Cis-acting 5' hammerhead optimization for in vitro transcription of highly structured RNAs

Mélanie Meyer & Benoît Masquida*

* To whom correspondence should be addressed

Benoît Masquida

b.masquida@unistra.fr

tel: +33 3 68 85 14 81

fax: +33 3 68 85 13 65

GMGM UMR 7156, IPCB, CNRS, Université de Strasbourg, 21 rue René Descartes 67084 Strasbourg

Running Title

Use of optimized hammerhead ribozymes for in vitro transcription applications

Keywords

Ribozymes, RNA folding, Lariat capping GIR1 ribozyme, in vitro transcription, RNA tertiary interactions

Abstract

RNA-mediated biological processes usually require precise definition of 5' and 3' ends. RNA ends obtained by *in vitro* transcription using T7-RNA-polymerase are often heterogeneous in length and sequence. An efficient strategy to overcome these drawbacks consists in inserting an RNA with known boundaries in between two ribozymes, usually a 5'-hammerhead and a 3' hepatitis delta virus ribozymes, that cleave off the desired RNA. In practice, folding of the three RNAs challenges each other, potentially preventing thorough processing. Folding and cleavage of the 5' hammerhead ribozyme relies on a sequence of nucleotides belonging to the central RNA making it more sensitive than the usual 3' hepatitis delta virus ribozyme. The intrinsic stability of the central RNA may thus prevent correct processing of the full transcript. Here, we present a method in which incorporation of a full length hammerhead ribozyme with a specific tertiary interaction prevents alternative folding with the lariat capping GIR1 ribozyme and enables complete cleavage in the course of the transcription. This strategy may be transposable for *in vitro* transcription of any highly structured RNA.

1. Introduction

To study the massive amount of RNAs identified by transcriptomics analysis, it is important to identify their boundaries in order to grab the "right" molecules. The history of ribozymes well illustrates this idea. At the beginning, the hammerhead ribozyme (HMH) has been characterized in short versions that prevented determining the full extent of the catalytic properties of the natural versions (Fedor & Uhlenbeck, 1990). Thus, biochemical studies showing that hammerhead ribozymes were orders of magnitude more efficient to fold and cleave under low magnesium concentrations if a tertiary interaction between stem I and the loop from stem II was occurring (Khvorova et al, 2003), came years later as a surprise (Figure 1). This finding was further demonstrated by the crystal structure of the *Schistosoma* hammerhead ribozyme in which setting the correct tertiary interaction leads to a shape of the catalytic site in which the chemical groups responsible for catalysis are ready for in-line attack (Martick & Scott, 2006). This was not the case in crystal structures of the shorter versions (Pley et al, 1994; Scott et al, 1995a) (Figure 1). Other examples like studies on the hairpin (Rupert & Ferre-D'Amare, 2001) and Varkud satellite (Lacroix-Labonte et al, 2012) ribozymes also show that the originally defined molecules were not including all the determinants required for their full efficiencies, and further, the vision of a modular architecture of RNA molecules (Masquida et al, 2010). The loss of structural integrity of important modules thus affects the accuracy of RNA architecture, which in the end perturbs the function.

The potential occurrence of functional modules within the ends of RNA molecules justifies that they should be well defined for most studies. Unfortunately, the most widely used enzyme for *in vitro* production of RNA, the T7 RNA polymerase (T7RNAP),

is prone to incorporate residues not encoded by the template at both the 5' (Helm et al, 1999) and 3' ends (Milligan & Uhlenbeck, 1989). These uncoded residues eventually jeopardize experimental strategies such as crystallographic studies, dye labelling for FRET studies, or in general any technic based on RNA ligation (Solomatin & Herschlag, 2009). Moreover, any sequence immediately downstream from the promoter cannot be synthesized since T7RNAP requires a specific G-rich starting sequence (Beckert & Masquida, 2011).

The errors made by T7RNAP can be concentrated in the tails of ribozymes located 5' and 3' of the RNA of interest that will be cleaved off (Price et al, 1995a). This strategy can also be adapted to various systems including the production of RNAs with unfavorable 5' sequences (Fechter et al, 1998; Mörl et al, 2005), and/or to the use of ribozymes in *trans* (Ferré-D'Amaré & Doudna, 1996). On the 5' side, a hammerhead ribozyme is usually chosen (Figure 2). Only few nucleotides with any sequence downstream from the cleavage site are required for hybridization to fold the active ribozyme. On the 3' side, the hepatitis delta virus (HδV) ribozyme is preferred because it is entirely located 3' from the cleavage site and cleaves after any nucleotide, which leaves total freedom to the user. After cleavage, the central RNA is left with 5' OH and 2',3' cyclic phosphate groups.

As long as the central RNA adopts a simple structure, this strategy does not need specific adaptation. However, if the structure of the central RNA is complex, it may interfere with proper folding of the flanking ribozymes and prevent efficient processing. While a short incubation at a temperature favoring structural rearrangements (~60°C) usually restores cleavage by the 3' HδV ribozyme, it is not the case for the 5' hammerhead. This situation happened when we tried to express a lariat capping GIR1 ribozyme (Nielsen et

al, 2005) with well defined ends (Figure 2). Nearly no cleavage was observed with a hammerhead version bearing a rather short stem I (5 bp). We assumed that the used truncated hammerhead had sub-optimal activity because it could not form the critical tertiary interactions taking place between stem I and the loop of stem II. Our purpose was to design 5' hammerhead constructs so that their substrate (stem I) resulted from hybridization to a longer sequence of the central lariat capping GIR1 ribozyme (LCrz) while preserving unpaired nucleotides, which interact with the facing loop from stem II. This tertiary interaction, thereby preserved in the fusion between the hammerhead and the lariat capping GIR1 ribozymes, accelerates folding and catalysis and allows decreasing the required magnesium concentration (Khvorova et al, 2003).

Design of the fusion hammerhead constructs using the program RNAfold (Hofacker et al, 1994) and their comparison to wild-type sequences of full length hammerhead ribozymes was carried out. Selected hammerhead sequences with a stem I integrating the 5' end of the lariat capping GIR1 ribozyme (Nielsen et al, 2005) were assayed for efficient cleavage in the course of *in vitro* transcription reactions (Figure 4). The direct involvement of the bulge in stem I was investigated by determining the observed kinetic constants of the successful candidates (Figure 5(Canny et al, 2004)). Cloning of efficient fusion hammerhead ribozymes upstream from the lariat capping GIR1 ribozyme, followed by the HδV ribozyme, was performed in plasmid pUC19, further transcribed *in vitro* as linearized templates. Northern blot analysis showed that the optimized hammerhead ribozymes were very efficient in the full context since they cleave to completion in the course of transcription, thus validating our design (Figure 7). This method validated on two different sequences of the lariat capping GIR1 ribozyme is a

practical addition to the molecular biologist's toolbox that addresses defining accurate ends to highly structured RNAs.

2. Materials

Prepare all solutions using Millipore or DEPC water. All buffers are filtered using a 0.22 μm sterile filter and kept at room temperature unless otherwise specified.

2.1 In silico design of fusion hammerhead ribozyme sequences

1. RNAfold (Hofacker et al, 1994) web server can be used http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi.

   Sequence of the full-length wild type hammerhead ribozyme from *Schistosoma*:
   5'GGAUGUACUACCAGCUGAUGAGUCCCAAAUAGGACGAAACGCCGCAAGGCGUCCU
   GGUAUCCAAUCC3'

   Sequence of the full-length fusion hammerhead 6 ribozyme derived from sTRSV+:
   5'GGGGUGCUUCGGAUGCUGAUGAGUCCGUGAGGACGAAACAGGGCAACCUGUCCAU
   CCGGUAUCCC3'.

   Sequence of the full-length fusion hammerhead 9 ribozyme derived from CChMVD+:
   5'GGGGACGUCCGGAUGCUGAUGAAGAUCCAUGACAGGAUCGAAACCUCUUGCAAAA
   GAGGUCCAUCCGGUAUCCC3'

   Sequence of the full-length ΔC15 fusion hammerhead 10 ribozyme derived from sLTSV-:
   5'GGGAGGUGCGGAUGAUGAGUCCGAAAGGACGAAACAGUAGGAAUACUGUCCAUCC
   GGUAUCCC3'

Sequence of the full-length fusion hammerhead 13 ribozyme derived from PLMVd+:

5'GGGUAACGGAUGCUGACGAGUCUCUGAGAUGAGACGAAACUCUUCGCAAGAAGAG UCCAUCCGGUAUCCC3'

2.2 In vitro determination of fusion hammerhead ribozymes cleavage efficiencies

2.2.1Co-transcriptional cleavage of fusion hammerhead ribozymes

1. DNA oligos are used as template for amplification by PCR creating the dsDNA template required for the *in vitro* transcriptions of the full-length hammerhead ribozymes from *Schistosoma* and full-length fusion hammerhead ribozymes HMH6, HMH9, HMH10 and HMH13.

   Forward (+) and reverse (-) strands for each hammerhead ribozyme:

   HMHSchistosoma+

   5'TAATACGACTCACTATAGGATGTACTACCAGCTGATGAGTCCCAAATAGGACGAAA CGCCGCAAGGCGTCCTGGTATCCAATCC3'

   HMHSchistosoma-

   5'GGATTGGATACCAGGACGCCTTGCGGCGTTTCGTCCTATTTGGGACTCATCAGCTGG TAGTACATCCTATAGTGAGTCGTATTA3'

   HMH6+

   5'TAATACGACTCACTATAGGGGTGCTTCGGATGCTGATGAGTCCGTGAGGACGAAAC AGGGCAACCTGTCCATCCGGTATCCC3'

   HMH6-

   5'GGGATACCGGATGGACAGGTTGCCCTGTTTCGTCCTCACGGACTCATCAGCATCCGA AGCACCCCTATAGTGAGTCGTATTA3'

HMH9+

5’TAATACGACTCACTATAGGGGACGTCCGGATGCTGATGAAGATCCATGACAGGATC

GAAACCTCTTGCAAAAGAGGTCCATCCGGTATCCC3’

HMH9-

5’GGGATACCGGATGGACCTCTTTTGCAAGAGGTTTCGATCCTGTCATGGATCTTCATC

AGCATCCGGACGTCCCCTATAGTGAGTCGTATTA3’

HMH10+

5’TAATACGACTCACTATAGGGAGGTGCGGATGATGAGTCCGAAAGGACGAAACAGTA

GGAATACTGTCCATCCGGTATCCC3’

HMH10-

5’GGGATACCGGATGGACAGTATTCCTACTGTTTCGTCCTTTCGGACTCATCATCCGCA

CCTCCCTATAGTGAGTCGTATTA3’

HMH13+

5’TAATACGACTCACTATAGGGTaaCGGATGCTGACGAGTCTCTGAGATGAGACGAAA

CTCTTCGCAAGAAGAGTCCATCCGGTATCCC3’

HMH13-

5’GGGATACCGGATGGACTCTTCTTGCGAAGAGTTTCGTCTCATCTCAGAGACTCGTCA

GCATCCGTTACCCTATAGTGAGTCGTATTA3’

2. DNA primers used for PCR amplification of the dsDNA template:

Primer 1: 5’ TAATACGACTCACTATA3’

Primer 2 for fusion hammerhead ribozymes: 5’GGGATACCGGATGGAC3’

Primer 2 for hammerhead from Schistosoma: 5’GGATTGGATACCAG3’

3. Phusion® High-Fidelity DNA Polymerase at a concentration of 2000 U/mL (New England BioLabs *inc*) stored at -20 °C.

4. Phusion® High-Fidelity DNA Polymerase HF reaction buffer 5x (New England BioLabs *inc*) stored at -20 °C.
5. 25 mM dNTPs mix stock solution is stored at -20 °C.
6. Glycerol blue loading buffer 6x: 50% glycerol, 0.1% xylene cyanol, 0.1% bromophenol blue.
7. 2% agarose gel: 2 g of agarose for routine use (Sigma Aldrich) dissolved in 100 mL of TBE 1x.
8. Steady red ^{32}P-αUTP at a concentration of 10 mCi/mL and a specific activity of 3000 Ci/mmol (Hartmann Analytic) is stored at 4 °C.
9. Micro Bio-Spin ® 6 Chromatography Columns (Bio-Rad).
10. His-tag-T7-RNA polymerase is produced in the laboratory from plasmid pT7-911Q (Ichetovkin et al, 1997).
11. TMSDT buffer 5x: 200 mM Tris-HCl pH 8.1, 110 mM $MgCl_2$, 5 mM spermidine, 25 mM DTT, 0.05% Triton stored at -20 °C.
12. Alkaline phosphatase (Sigma-Aldrich) is stored at a concentration of 1 U/µL in TMSDT 1x at 4 °C.
13. 100 mM ATP, 100 mM UTP, 100 mM CTP and 100 mM GTP stock solution are stored at -20 °C.
14. Urea loading buffer 2x: 8 M urea, 0.025% xylene cyanol, 0.025% bromophenol blue.
15. Tris Borate EDTA buffer (TBE 10x) to run PAGE: 216 g of Tris base, 55 g of Boric acid, 40 mL of 0.5 M EDTA pH 8.0, adjust with water up to 1 L.
16. Denaturing PAGE: 10% acrylamide/bis-acrylamide (19:1), 8 M urea in TBE 1x.
17. Gel size: 35 cm width * 42 cm length * 0.08 cm thickness.

18. Apparatus to vacuum dry gels: Slab Gel Dryer (HAEFER Scientific Instruments).
19. 35 x 43 cm Imaging Plate (FujiFilm).
20. Bio-imager (FujiFilm).

2.2.2 Kinetic analysis of fusion hammerhead ribozymes with trans substrates

1. Substrates RNA were purchased at Dharmacon (Thermo Scientific).
2. Shaker MixMate® (Eppendorf).
3. 3 M KCl stock solution.
4. T4 RNA ligase at a concentration of 5 U/µL (Ambion ®) stored at -20 °C.
5. T4 RNA ligase reaction buffer 10x: 500 mM Tris-HCl pH 7.8, 100 mM $MgCl_2$, 100 mM DTT, 10 mM ATP (Ambion ®) stored at -20 °C.
6. Steady red ^{32}pCp at a concentration of 10 mCi/mL and a specific activity of 3000 Ci/mmol (Hartmann Analytic) is stored at 4 °C.
7. X-ray autoradiograms Super RX (Fuji Medical).
8. Kinetic buffer 10x: 500 mM Tris-HCl pH 8.0, 1 M NaCl, 1 mM EDTA.
9. 1 M $MgCl_2$ stock solution.
10. Formamide-EDTA blue 2x: 80% formamide, 50 mM EDTA pH 8.0, 0.02% bromophenol blue.
11. Quantifications were done using Science Lab 2003 Image Gauge 4.2 (FujiFilm).
12. The curve fit was done using KaleidaGraph 4.02.

2.3 Cloning of ternary ribozyme construct

1. DNA oligos used to create the insert corresponding to the ternary ribozyme constructs to be cloned:

Forward primers:

EcoRI_T7p_HMH2

5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGGTAACGGATGCTGACGAGTCT

CTGAGA3’

HMH1_LCrz

5’CTGACGAGTCTCTGAGATGAGACGAAACTCTTCGCAAGAAGAGTCCATCCGGTATC

CCAAGA3’

Reverse primers:

BamHI_HδV3

5’GGATCCTCTAGAGTCGACCTTGTCCCATTCGCCATTACCGAGGGGACGGTCCCCTCG

GAATGTT3’

HδV2

5’GTCCCCTCGGAATGTTGCCCACCGGCCGCCAGCGAGGAGGCTGGGACCATGGCCGGC

3’

HδV1_LCrz

5’GGACCATGGCCGGCGATTGTCTTGGGATACCGGATGCTTCCTTTCGGAACGACT3’

2. QIAquick® Gel Extraction Kit (Qiagen).
3. pUC19 vector (New England Biolabs) stored at a concentration of 1 mg/mL at -20 °C.
4. In-Fusion® HD Cloning Kit (Clontech).
5. QIAprep® Spin Miniprep Kit (Qiagen).
6. DNA plasmids were sequenced at GATC Biotech.

2.4 In vitro transcription of the full length RNA using linearized plasmids as templates

Material necessary for this section is the same as the one used in section 2.1.1.

2.5 Northern blot analysis of the cleavage pattern of the ternary ribozyme transcripts

1. DNA probes for Northern blot:

   HMH 5'GACTCTTCTTGCGAAGAGTTTC3'

   LCrz 5'CATGATACTTCCCAACCCAACC3'

   HδV 5'CTCGGAATGTTGCCCACCGGCC3'

2. Steady red ^{32}P-γATP at a concentration of 10 mCi/mL and a specific activity of 3000 Ci/mmol (Hartmann Analytic) is stored at 4 °C.
3. T4 polynucleotide kinase is at a concentration of 10 U/µL (Thermo scientific) is stored at – 20 °C.
4. T4 polynucleotide kinase buffer A (forward reaction) 10x: 500 mM Tris-HCl pH 7.6, 100 mM $MgCl_2$, 50 mM DTT, 1 mM spermidine (Thermo scientific) is stored at – 20 °C.
5. Nylon membrane for Northern blot analysis: Amersham Hybond-H+ (GE Healthcare Life Sciences).
6. Stratalinker® UV Crosslinker (Stratagene).
7. Hybridization bottle: 35 x 300 mm (Schott).
8. SSPE 6x: 3 M NaCl, 0.2 M NaH_2PO_4 and 0.02 M EDTA. pH is adjusted to 7.4 with NaOH.
9. Pre-hybridization buffer: SSPE 1x, 2.5% SDS, 2.5% Denhart.

   Pre-Hybridization buffer is stored at 30°C to avoid SDS precipitation. Denhart solution should be freshly made.

10. Hybridization incubator "Isotemp" (Fisher Scientific)

3. Methods

3.1 In silico design of fusion hammerhead ribozyme sequences

The aim of this step is to design fusion hammerhead ribozymes presenting a hammerhead catalytic core fused to a substrate stem corresponding to the 5' sequence of the lariat capping GIR1 ribozyme. The ribozyme strand presents both ends rooted in stem I, while loops close stems II and III. Stems II and III and the single-stranded regions form the catalytic core and are directly copied from genuine hammerhead sequences presenting tertiary contacts between stem I and the loop of stem II (sTRSV+, CChMVD+, PLMVD+, (Khvorova et al, 2003)). Stem I should be approximately a dozen nucleotides long with a 3' strand being the 5' end of the central RNA, the lariat-capping GIR1 ribozyme in this case (Nielsen et al, 2005). Consequently the 5' strand of stem I is designed to be partialy complementary to the latter and starts with three G residues for transcriptional efficiency (Figure 2). Moreover, a bulge intervening 6 nucleotides after the cleavage site has to be preserved in stem I. The purpose of this bulge is to avoid formation of a perfect duplex which would destroy the tertiary contacts with the loop of stem II (Figure 1) that are crucial for efficient folding and cleavage of the ribozyme (Figure 3 (Canny et al, 2004)). In the sequences resulting from the fusion between hammerhead ribozymes and the lariat-capping GIR1 ribozyme, the presence of (i) a correct three-way junction forming the catalytic core and of (ii) the critical unpaired nucleotides in stem I was systematically checked using RNAfold (Hofacker et al, 1994) (*see* **Note 1**).

3.2 In vitro determination of fusion hammerhead ribozymes cleavage efficiencies

Successful sequences were transcribed from commercial template DNA oligos. The extent of cleavage was monitored in the course of the transcription reactions (Figure 4). More detailed kinetic analysis of the most efficient transcripts, HMH13, were performed under single-turnover conditions using fusion hammerhead ribozymes consisting of distinct catalytic and substrate strands (Figure 5).

3.2.1 Co-transcriptional cleavage of fusion hammerhead ribozymes

To produce RNAs corresponding to candidate sequences selected *in silico* by *in vitro* transcription (Beckert & Masquida, 2011; Milligan et al, 1987), DNA templates preceded by a Class III T7 promoter (5' TAA TAC GAC TCA CTA TA 3') should be produced (Figure 4).

1. Design full-length template DNA oligos, forward and reverse strand that will be used as dsDNA templates for *in vitro* transcription of the candidates sequences selected *in silico* (*see* **note 2**).
2. Design primers for PCR amplification of the template DNA oligos. (*see* **note 3**).
3. Prepare 50 µL of PCR reaction containing: 300 ng of each DNA template, 0.5 µL of the Phusion® High-Fidelity DNA Polymerase (2000 U/mL), 10 µL of the Phusion® HF Buffer 5x, 1.2 µM of forward primer, 1.2 µM of reverse primer, 400 µM dNTPs mix, water up to 50 µL.
4. Run the PCR reaction in a thermal cycler following the set up displayed in Table 1.
5. Check the efficiency of amplification by mixing 5 µL of the PCR reaction to 1 µL glycerol blue loading buffer 6x and load on to a 2% (w/v) agarose gel.
6. After ethidium bromide staining, observe the DNA migration pattern by UV shadowing.
7. Clean the PCR reaction using the Micro Bio-Spin Columns.

8. Set up 50 µL transcription reactions containing: 1 µg of the cleaned PCR reaction, 10 µL of TMSDT 5x Buffer, 5 mM ATP, 5 mM GTP, 5 mM CTP, 2.5 mM UTP, 25 µCi ^{32}P-αUTP (10 mCi/mL, 3000 Ci/mmol), 0.2 µL of alkaline phosphatase (1 U/µL) (*see* **note 4**), T7 RNA polymerase as indicated by your provider (*see* **note 5**), water up to 50 µL.
9. Incubate at 37 °C for 2 h.
10. Remove a 12 µL aliquot of each transcription reaction after 30 min, 1 h and 2 h.
11. Stop the transcription by mixing the sample to an equal volume of urea-containing loading buffer 2x.
12. Clean the transcription reaction using the Micro Bio-Spin Columns (*see* **note 6**).
13. Fractionate each sample on a 10% denaturing PAGE. Run the gel for two hours at 60 W (*see* **note 7**).
14. Vacuum dry the gel onto Whatman paper for four hours at 80°C or alternatively seal it in a plastic film and store it at -80°C.
15. Expose an imaging plate for 10 min in a cassette (*see* **note 8**).
16. Read the imaging plate using a Bio-imager.

3.2.2 Kinetic analysis of fusion hammerhead ribozymes with trans substrates

This step can be performed as in (Canny et al, 2004) to check whether the bulge in stem I is functional or to determine an optimal magnesium concentration for cleavage (*see* **Note 9**). The hammerhead cores are obtained by *in vitro* transcription of dsDNA templates while substrates are obtained by RNA solid phase chemical synthesis (Figure 5).

1. Design two sets of fusion hammerhead ribozymes and their substrates as separate RNA molecules. From this step the catalytic cores of the fusion hammerhead ribozymes will be handled separately from their substrate domains.
2. Purify by 10% denaturing PAGE the complementary DNA oligos that will form the dsDNA templates used for *in vitro* transcription of the different ribozymes' cores.
3. Mix 20 µg of the forward and the reverse pure template DNA oligos and add KCl up to a final concentration of 500 mM. To hybridize the template DNA oligos heat the sample for 1 min at 95°C and allow it to cool down to room temperature for 10 min.
4. Set up 200 µL transcription reactions containing: 40 µg of dsDNA template, 40 µL of TMSDT 5x Buffer, 5 mM ATP, 5 mM GTP, 5 mM CTP, 5 mM UTP, 0.8 µL of alkaline phosphatase (1 U/µL) (*see* **note 4**), T7 RNA polymerase as indicated by your provider (*see* **note 5**), water up to 200 µL.
5. Purify transcription reactions by 10% denaturing PAGE.
6. Identify RNA bands corresponding to the fusion hammerhead core by UV shadowing.
7. Cut the band of interest out of the gel and place it into a 1.5 mL tube.
8. Immerse the gel band in about 800 µL of water and place the tube in a shaker over-night at 4°C.
9. Decant the elution water and transfer it to a fresh tube.
10. Recover the fusion hammerhead cores by ethanol precipitation of the elution water.

11. Set a 3' radio-labeling reaction for the RNA substrates: 500 ng of RNA substrate, 40 µCi ^{32}pCp (10 mCi/mL, 3000 Ci/mmol), 10% DMSO, 1 µL T4 RNA ligase (5 U/µL), 3 µL T4 RNA ligase buffer 10x, water up to 30 µL.
12. Incubate at 16 °C over-night.
13. Clean the labeling reactions using the Micro Bio-Spin Columns (*see* **note 6**).
14. Purify the labeled RNA substrates by 20% denaturing PAGE at 50 W for 2 hours.
15. Localize the positions of the labeled RNAs using an X-ray sensitive film.
16. Perform steps 7-10 of section 3.2.2.
17. Set up a reaction for analyzing the kinetics of the hammerhead cleavage under single turnover conditions as follows: 0.5 µM of purified fusion hammerhead core RNA, 10 nM of the 3' purified radio-labeled substrate RNA, 1.5 µL of kinetic buffer 10x, water up to 13 µL.
18. Incubate the reaction at 70 °C for 2 min and then at 25 °C for 10 min.
19. Take a 2 µL aliquot of the reaction.
20. Add $MgCl_2$ to a final concentration of 0.1 mM.
21. Incubate the reaction at 25 °C for 2 h.
22. Take a 2 µL aliquot of the reaction after, 15 s, 1 min, 3 min, 15 min, 2 h.
23. Quench the reaction by mixing the aliquot with 2 µL EDTA-formamide blue 2x.
24. Place samples in ethanol cooled down by dry ice.
25. Separate the cleaved from uncleaved substrates by 12% denaturing PAGE at 50 W for 2 hours (*see* **note 7**).
26. Vacuum dry the gel onto Whatman paper for 4 hours at 80 °C or alternatively seal it in a plastic film and store it at -80 °C.
27. Expose an imaging plate over-night in a cassette (*see* **note 10**).

28. Reveal the radioactivity pattern on the imaging plate using a Bio-imager.
29. Quantify the amount of cleaved substrate using appropriate software.
30. Plot the fraction of substrate cleaved *versus* time and determine k_{obs} by fitting the data to the equation, $F_{Cl} = F_{max}(1-e^{-kt})+F_0$ using the KaleidaGraph software (*see* **note 11**).

3.3 Cloning of ternary ribozyme constructs

The most efficient fusion hammerhead ribozyme is cloned upstream from the lariat capping GIR1 ribozyme followed by an HδV ribozyme ending with a restriction endonuclease site used for linearization to permit run-off transcription. Since cloning is the best choice for storing matrices in the laboratory's library, it is advised to anticipate the cloning step at the very beginning of the sequence design by adding a 5' sequence either corresponding to the desired restriction enzyme recognition sequence or to the recombination sequence specified in the cloning kit.

All lariat capping GIR1 RNA constructs from *Didymium iridis* (DiGIR1) were transcribed from plasmid DNA templates obtained by cloning of an insert generated by 3 successive PCR amplifications. The 5’ PCR primers were designed to contain an EcoRI restriction site, a T7 promoter, the sequence of the co-transcriptional self-cleaving hammerhead ribozyme and the 5’ sequence of the DiGIR1 ribozyme. The 3’ PCR primers included a BamHI restriction site, a post-transcriptional self-cleaving HδV ribozyme and the 3’ sequence of the lariat capping DiGIR1 ribozyme (Figure 6).

1. Design forward and reverse primers for the successive PCRs (*see* **note 12**).
2. Prepare 50 µL of PCR reaction containing: 50 pg of template plasmid DNA, 0.5 µL of the Phusion® High-Fidelity DNA Polymerase (2000 U/mL), 10 µL of the

Phusion® HF Buffer 5x, 1.2 µM of forward primer, 1.2 µM of reverse primer, 400 µM dNTP mixture, water up to 50 µL (*see* **note 13**).

3. Run the PCR reaction in a thermal cycler following the set up displayed in Table 1 (*see* **note 14**).
4. Mix the PCR reaction to 10 µL glycerol blue loading buffer 6x and load the entire PCR reaction onto a 2% (w/v) agarose gel.
5. After ethidium bromide staining, observe the DNA migration pattern under UV shadowing.
6. Cut the DNA band of interest out of the gel.
7. Recover the DNA insert in 30 µL of water with a gel extraction kit (*see* **note 15**).
8. Insert the purified PCR product into a pUC19 plasmid using a cloning kit.
9. Directly transform XL-1 Blue *Escherichia coli* competent cells with the recombinant plasmid product (*see* **note 16**).
10. Grow the transformed XL-1 Blue *Escherichia coli* competent cells on LB/Ampicillin (100 µg/mL) plates overnight at 37 °C.
11. Grow 3 mL of LB/Ampicillin (100 µg/mL) cultures from individual colonies.
12. Extract DNA using a miniprep spin kit.
13. Sequence the extracted DNA using the M13 sequencing primers.
14. Prepare glycerol stocks of the clones containing the desired sequences by mixing 1 mL of saturated bacterial culture to 600 µL of sterilized 50% glycerol solution in 2 mL screw cap tubes. Store the glycerol stocks at -80 °C.

3.4 In vitro transcription of the ternary ribozyme RNA using linearized plasmids as templates

The transcription reaction of the plasmid template is performed as for the PCR template except that the plasmid is linearized prior to the reaction and that the mass added is greater since the molarity of the template is important.

1. For *in vitro* transcription, linearize the plasmid with BamHI restriction endonuclease following the enzyme provider instructions.
2. Set up 50 µL transcription reactions containing: 2.5 µg of linearized plasmid DNA, 10 µL of TMSDT 5x Buffer, 5 mM ATP, 5 mM GTP, 5 mM CTP, 5 mM UTP, 0.2 µL of alkaline phosphatase (*see* **note 4**), T7 RNA polymerase as indicated by your provider (*see* **note 5**), water up to 50 µL.
3. In parallel set up 50 µL transcription reactions containing: 2.5 µg of linearized plasmid DNA, 10 µL of TMSDT 5x Buffer, 5 mM ATP, 5 mM GTP, 5 mM CTP, 2.5 mM UTP, 25 µmCi ^{32}P-αUTP (10 mCi/mL, 3000 Ci/mmol), 0.2 µL of alkaline phosphatase (1 U/µL) (*see* **note 4**), T7 RNA polymerase as indicated by your provider (*see* **note 5**), water up to 50 µL.
4. Incubate at 37 °C for 2 h.
5. Incubate half of each transcription at 60 °C for 1 h.
6. Stop transcription reaction by mixing it to an equal volume of urea blue loading buffer 2x.

3.5 Northern blot analysis of the cleavage pattern of the ternary ribozyme RNA

1. Design three DNA probes that will be used to hybridize specifically either to the hammerhead ribozyme or to the central RNA or to the HδV ribozyme (*see* **note 17**).
2. Label each DNA probe by mixing: 1 µM of DNA oligo, 2.5 µL of ^{32}P-γATP (10 mCi/mL, 3000 Ci/mmol), 1 µL of T4 polynucleotide kinase (10 U/µL), 5 µL buffer A 10x, water up to 50 µL.
3. Incubate 30 min at 37 °C then 10 min at 65 °C.
4. Clean the labeling reactions using the Micro Bio-Spin Columns (*see* **note 6**).
5. Fractionate samples from *in vitro* transcription reactions of the ternary ribozyme RNA on an 8% denaturing PAGE. Load in four consecutive lanes your 60 °C pre-incubated or not samples in order to be able to compare the cold to the corresponding body-labeled transcription reactions. Repeat the same operation two times. Each set of four lanes will be used for a different DNA probe. Run for 3 hours at 60 W. (*see* **note 7**).
6. Transfer the RNA bands from the gel onto positively charged nylon membrane using a vacuum gel drier for four hours at 80 °C.
7. Wash the membrane in water to remove the dried gel particles (*see* **note 18**).
8. Crosslink the RNA onto the membrane using a UV crosslinker.
9. Cut the membrane to separate each set of four lanes. From this step you will be dealing with three membranes in parallel.
10. Insert delicately each membrane into a hybridization bottle (*see* **note19**).
11. Add 20 mL of pre-hybridization buffer.
12. Incubate at 60 °C for 30 min in a rolling hybridization oven.

13. Add the total amount of one labeled DNA primer per hybridization bottle.
14. Incubate at 37 °C over night in the rolling hybridization oven.
15. Wash each membrane with 20 mL of SSPE 6x 5 min at 37°C in the rolling hybridization oven. Repeat this step three times.
16. Wash each membrane with 20 mL of SSPE 6x 5 min at 42°C in the rolling hybridization oven.
17. Seal each membrane in a plastic film.
18. Expose an imaging plate to the plastic sealed membrane for three hours in a cassette (*see* **note 8, 10**).
19. Reveal the radioactivity pattern on the imaging plate using a Bio-imager.

4 Notes

1. The sequence of the adapted substrate starts after the cleavage site and preserves several unpaired residues in stem I of the ribozyme core to favor tertiary contacts with the loop of stem II. The best situation is when the UAC sequence forms a bulge. When this not possible, avoid building an extended helix by inserting at least non-complementary residues.
2. Do not forget to insert the T7 promoter sequence upstream of the DNA templates.
3. These primers can be the same for all fusion hammerhead ribozymes candidates if you use as primer 1 the sequence of the T7 promoter and as primer 2 the sequence of the central RNA
4. It is not compulsory to add alkaline phosphatase but it is known to increase in vitro transcription yields.
5. We produce our own T7 RNA polymerase, which reduces cost considerably.

6. It is not mandatory to clean transcription reaction although it is preferable to remove not incorporated radio-labeled nucleotides to avoid contamination of the electrophoresis buffer.
7. To have better defined bands, gels are pre-run for one hour at 60 W.
8. The radioactivity pattern of the gel can as well be obtained on X- ray autoradiogram.
9. This step is not mandatory and was carried out as a proof of principle. It can be performed if the ribozyme construct is inactive to further check for efficiency of a set of hammerhead constructs carrying different kinds of bulges in stem I.
10. Exposure time of the imaging plate depends on the amount of ^{32}P-labeled material on the membrane.
11. F_{cl} is the fraction of cleaved substrate, F_0 is the initial amount of substrate cleaved before addition of MgCL2, F_{max} is the maximal amount of substrate cleaved, t is the time in minute and k_{obs} is the observed rate of cleavage.
12. For optimal yield, forward and reverse primers should have close melting temperature.
13. Keep your samples on ice while preparing the PCR reaction.
14. Insert your samples in the thermal cycler when the temperature of the sample holder is at 90°C minimum.
15. To avoid any trouble with next steps of the cloning process, 10% of ammonium acetate 3 M pH 5.2 is added to the recovered DNA and precipitated with 3 volumes of pure ethanol. DNA is then re-suspended in 30 µL of water.
16. Very efficient chemically competent *Escherichia coli* cells can be prepared if all the material is maintained ice-cold throughout the whole preparation process.

17. The three primers should have close melting temperature and the melting temperature must be above 42 °C
18. Be delicate not to scratch the membrane.
19. Be careful that side of the membrane on which the RNA is crosslinked is facing the inside of the hybridization bottle. There should not be any membrane overlap in the hybridization bottle.

Acknowledgement

This work was supported by the Centre National de la Recherche Scientifique and the University of Strasbourg through a Ph.D. grant to MM. We are grateful to Laure Schaeffer for technical assistance.

References

1. Beckert B, Masquida B (2011) Synthesis of RNA by in vitro transcription. Methods Mol Biol 703:29-41
2. Canny MD, Jucker FM, Kellogg E et al. (2004) Fast cleavage kinetics of a natural hammerhead ribozyme. J Am Chem Soc 126:10848-10849
3. Fechter P, Rudinger J, Giege R et al. (1998) Ribozyme processed tRNA transcripts with unfriendly internal promoter for T7 RNA polymerase: production and activity. FEBS Lett 436:99-103
4. Fedor MJ, Uhlenbeck OC (1990) Substrate sequence effects on "hammerhead" RNA catalytic efficiency. Proc Natl Acad Sci U S A 87:1668-1672
5. Ferré-D'amaré AR, Doudna JA (1996) Use of *cis*- and *trans*-ribozymes to remove 5' and 3' heterogeneities from milligrams of *in vitro* transcribed RNA. Nucleic Acids Res. 24:977-978
6. Helm M, Brule H, Giege R et al. (1999) More mistakes by T7 RNA polymerase at the 5' ends of in vitro-transcribed RNAs. RNA 5:618-621
7. Hofacker IL, Fontana W, Stadler PF et al. (1994) Fast folding and comparison of RNA secondary structures. Monatshefte für Chemie / Chemical Monthly 125:167-188
8. Ichetovkin IE, Abramochkin G, Shrader TE (1997) Substrate recognition by the leucyl/phenylalanyl-tRNA-protein transferase. Conservation within the enzyme family and localization to the trypsin-resistant domain. J Biol Chem 272:33009-33014

9. Jossinet F, Ludwig TE, Westhof E (2010) Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. Bioinformatics 26:2057-2059
10. Khvorova A, Lescoute A, Westhof E et al. (2003) Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. Nat Struct Biol 10:708-712
11. Lacroix-Labonte J, Girard N, Lemieux S et al. (2012) Helix-length compensation studies reveal the adaptability of the VS ribozyme architecture. Nucleic Acids Res 40:2284-2293
12. Martick M, Scott WG (2006) Tertiary contacts distant from the active site prime a ribozyme for catalysis. Cell 126:309-320
13. Masquida B, Beckert B, Jossinet F (2010) Exploring RNA structure by integrative molecular modelling. N Biotechnol 27:170-183
14. Milligan JF, Groebe DR, Witherell GW et al. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. Nuleic. Acids Res. 15:8783-8798
15. Milligan JF, Uhlenbeck OC (1989) Synthesis of small RNAs using T7 RNA polymerase. Methods Enzymol. 180:51-62
16. Mörl M, Lizano E, Willkomm DK et al. (2005) Production of RNAs with homogenous 5' and 3' ends. In: Hartmann RK, Bindereif A, Westhof E (eds) Handbook of RNA biochemistry. p 22-35
17. Nielsen H, Westhof E, Johansen S (2005) An mRNA is capped by a 2', 5' lariat catalyzed by a group I-like ribozyme. Science 309:1584-1587
18. Pley HW, Flaherty KM, Mckay DB (1994) Three-dimensional structure of a hammerhead ribozyme. Nature. 372:68-74
19. Price SR, Ito N, Oubridge C et al. (1995) Crystallization of RNA-protein complexes I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. J. Mol .Biol. 249:398-408
20. Rupert PB, Ferre-D'amare AR (2001) Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. Nature 410:780-786.
21. Scott WG, Finch JT, Klug A (1995) The crystal structure of an all-RNA hammerhead ribozyme : A proposed mechanism for RNA catalytic cleavage. Cell. 81:991-1002
22. Solomatin S, Herschlag D (2009) Methods of site-specific labeling of RNA with fluorescent dyes. Methods Enzymol 469:47-68

Figures

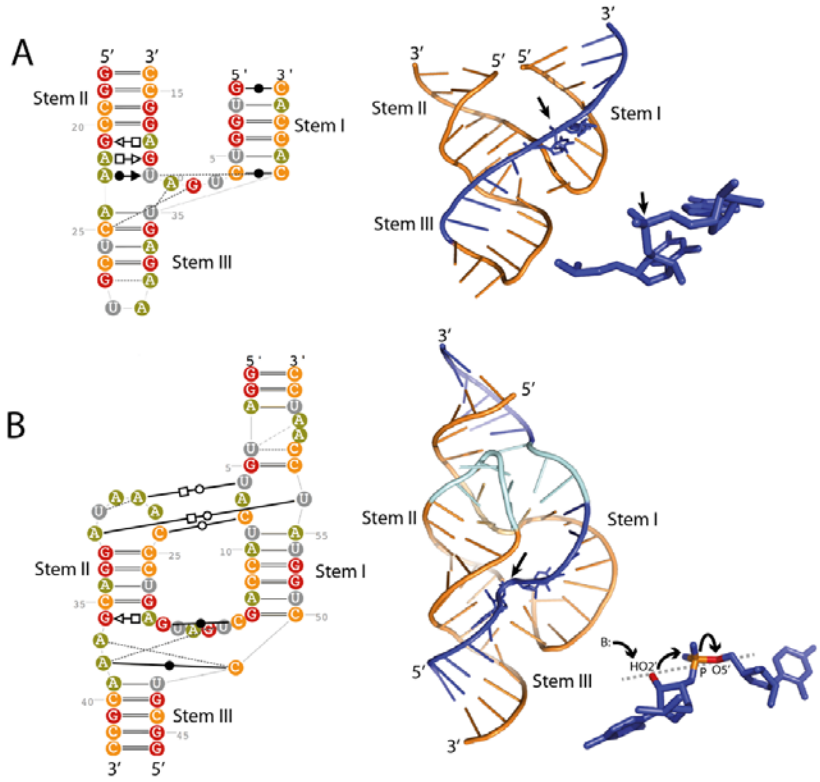


Figure 1: Comparative secondary and three-dimensional structure of the hammerhead ribozyme cores from the minimal (A) versus the full-length (B) versions. Secondary structures (represented using the program Assemble (Jossinet et al, 2010)) indicate the sequence with nucleotides colored by type. For clarity, stem I is depicted with the 5' strand on the left. Schemes in following figures depicts the 5' strand of stem I on the right. The crystal structures of the minimal (Scott et al, 1995a) and full length (Martick & Scott, 2006) hammerhead ribozymes present an orange core and a blue substrate. Nucleotides critical for tertiary interactions taking place in the full-length hammerhead are colored in cyan (B). The cleavage site is indicated with an arrow. The dinucleotide around the scissile bond are enlarged and displayed as sticks. The structures show the different conformations taking place in the two-ribozyme versions. A dashed grey line shows the alignment of the 3 colored atoms (O2', P, O5') involved in the in-line attack mechanism that takes place in the efficient version of the hammerhead ribozyme.

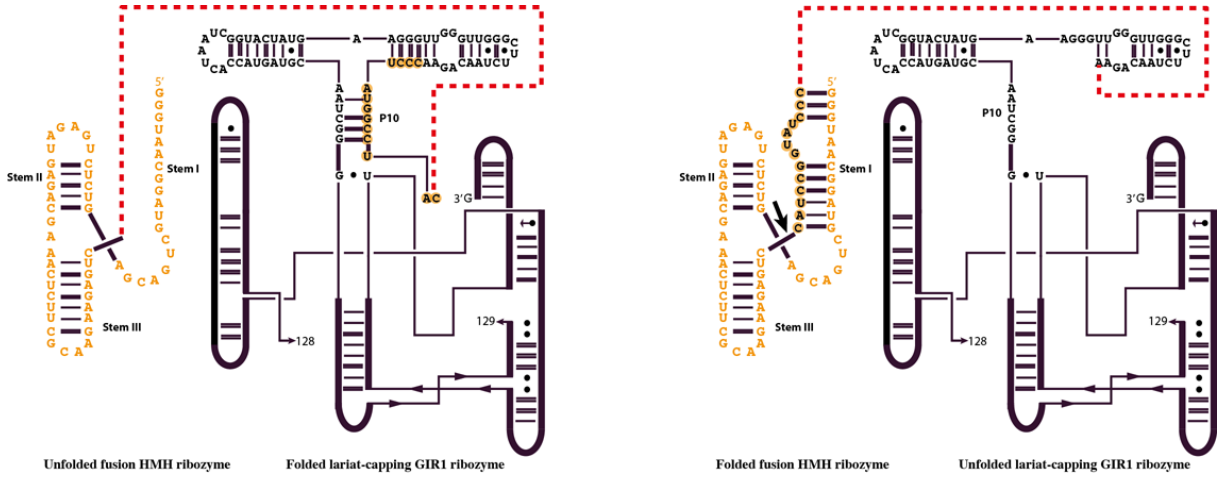


Figure 2: Mutually exclusive theoretical secondary structures of the pre-RNA ternary ribozyme transcript. If the pre-transcript presents an unfolded inactive hammerhead and a folded lariat-capping GIR1 ribozymes, the situation is unfavorable (left panel). However, the opposite situation (right panel) favors maximal cleavage during the *in vitro* transcription reaction. The bulge in stem I potentially forms a more stable structure with the loop closing stem II than if stem I is closed by a couple of additional A-U pairs. An arrow indicates the cleavage site.

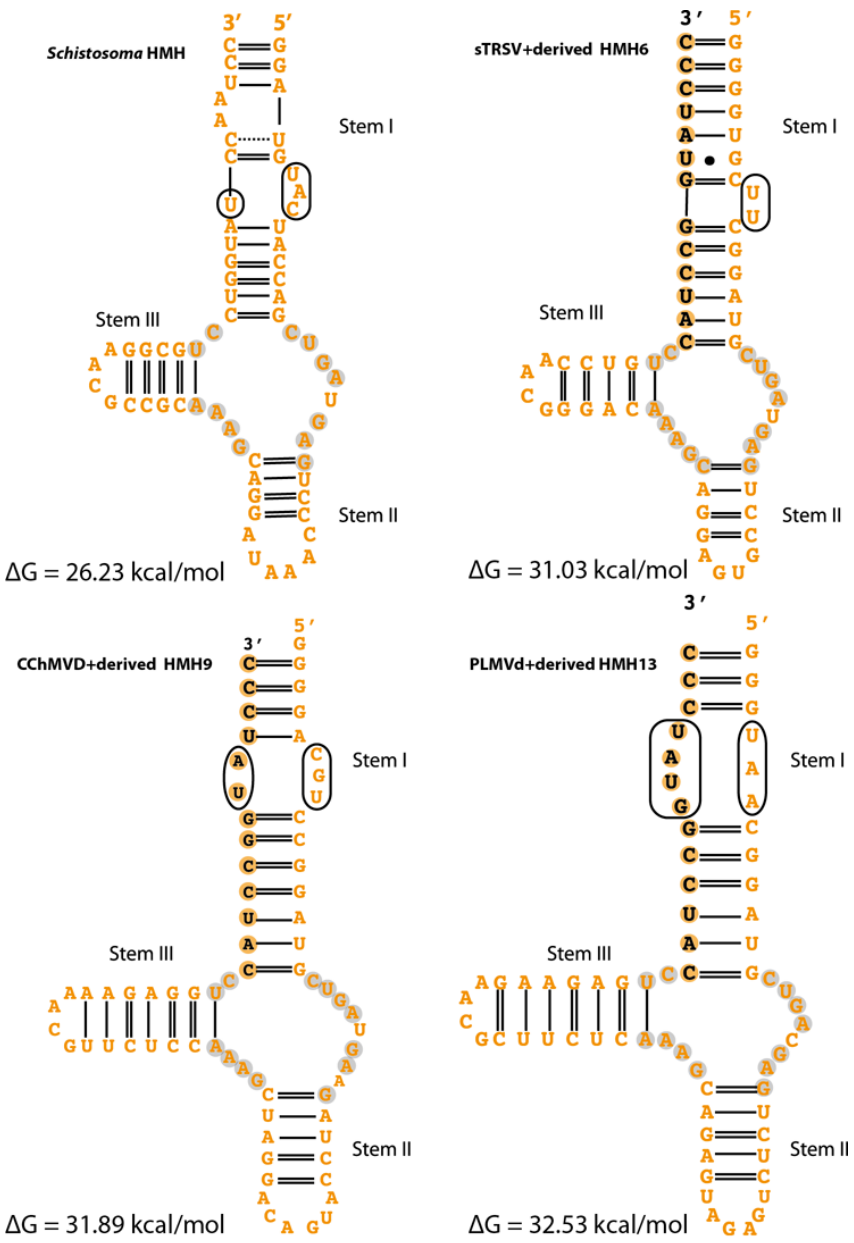


Figure 3: Secondary structure diagrams of the fusion hammerhead ribozyme candidates (outputs from RNAfold) show a correct catalytic core structure and unpaired nucleotides in stem I. The secondary structure of the hammerhead ribozyme from *Schistosoma* is shown as a reference model with nucleotides from the bulge in stem I interacting with the loop from stem II circled. Some of the circled unpaired nucleotides in candidates HMH6, 9 and 13 potentially interact with the loop from stem II.

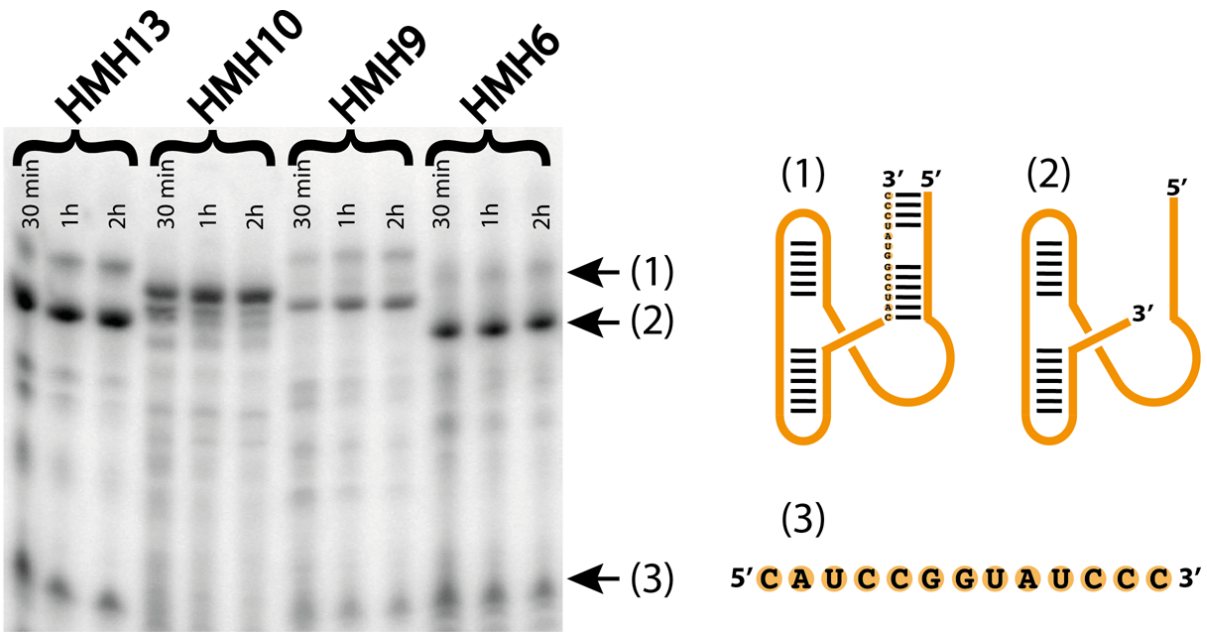


Figure 4: Body-labeled RNA pattern from a standard transcription reaction of the full-length hammerhead ribozymes fused to the 13 nucleotides at the 5' end of the lariat-capping GIR1 ribozyme. This step is meant to check whether ribozymes with the modified sequence, but lacking the main part of the lariat-capping GIR1 ribozyme, cleave in the course of the transcription reaction. The main RNA species numbered 1 to 3 are visible on the gel, (1) uncleaved hammerhead, (2) cleaved hammerhead core, (3) 13-nt cleaved substrate. HMH10, used as a negative control, carries a mutation ΔC15 with respect to the *Schistosoma* hammerhead ribozyme numbering (*see* **Figure 1B**).

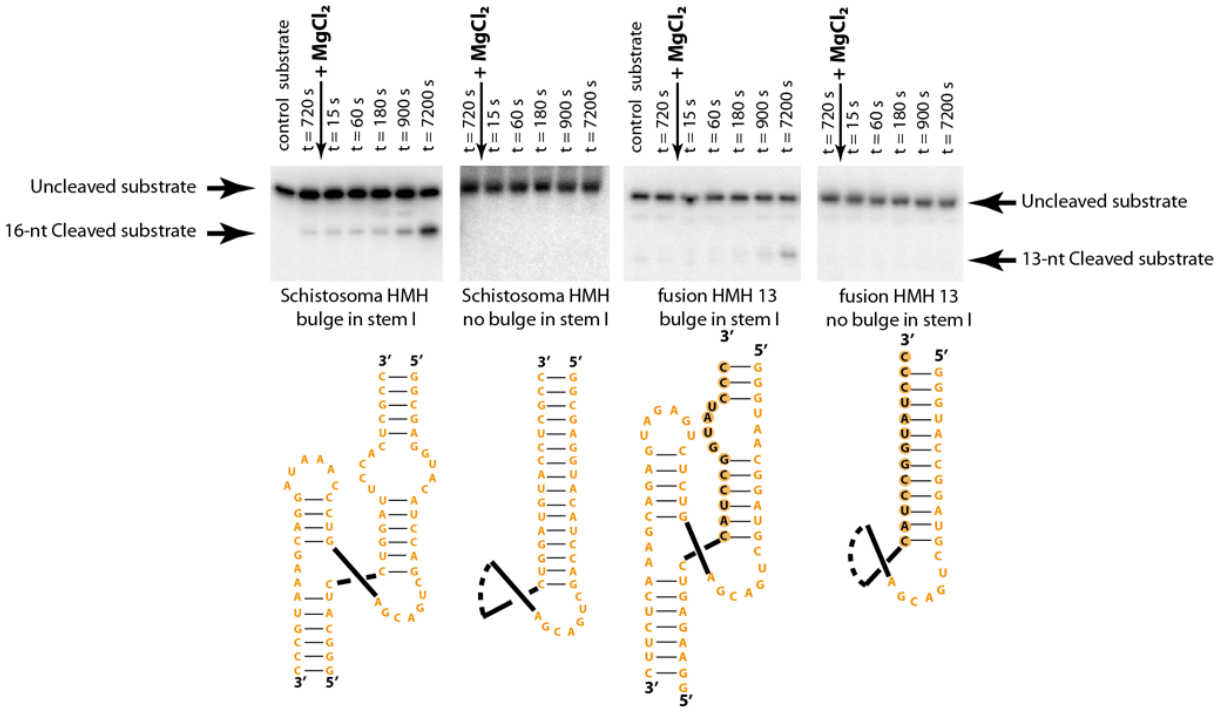


Figure 5: Example of kinetic analysis of *trans* fusion hammerhead lariat capping GIR1 ribozymes. The model hammerhead from *Schistosoma* (left panel) is compared to HMH13 (right panel). A 50-fold excess of the *in vitro* transcribed hammerhead ribozyme core was added to the 3' ^{32}P-labeled RNA substrate substrate to achieve single turnover conditions. Two kinds of cores (and substrates in the case of the *Schistosoma* ribozyme) were used. The first kind preserved the bulge in stem I whereas the second kind forms a duplex with perfect complementarity. Cleavage is observed only when the bulge is present. The fraction of cleaved substrate was plotted versus time. Data were fitted to the equation $F_{cl}=F_{max}(1-e^{-kt})+F_0$ to determine k_{obs} values. The *Schistosoma* ribozyme presents a k_{obs} similar to HMH13, $k_{obsSch}=0.013\pm4.7\cdot10^{-3}$ min^{-1} and $k_{obsHMH13}=0.016\pm8.3\cdot10^{-3}$ min^{-1} showing that our designed hammerheads are as efficient as the *Schistosoma* model ribozyme at 0.1 mM $MgCl_2$.

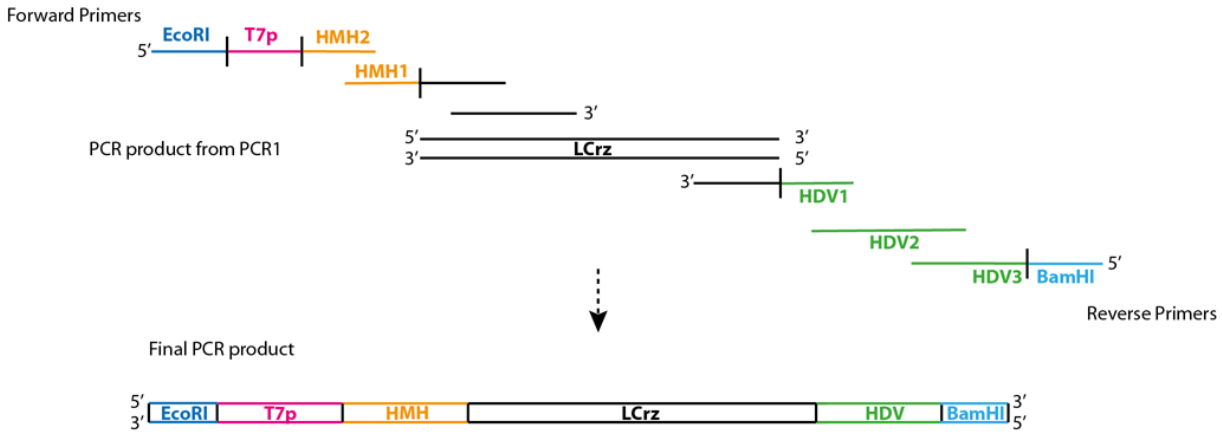


Figure 6: Description of the PCR strategy of the DNA template encoding the ternary ribozyme system. A plasmid containing a 5' truncated lariat-capping GIR1 ribozyme (LCrz) was used as a template for the first PCR amplification. Two further PCR steps created the flanking hammerhead and hepatitis δ virus (HδV) ribozymes, a T7-promotor and restriction sites for cloning (EcoRI, BamHI).

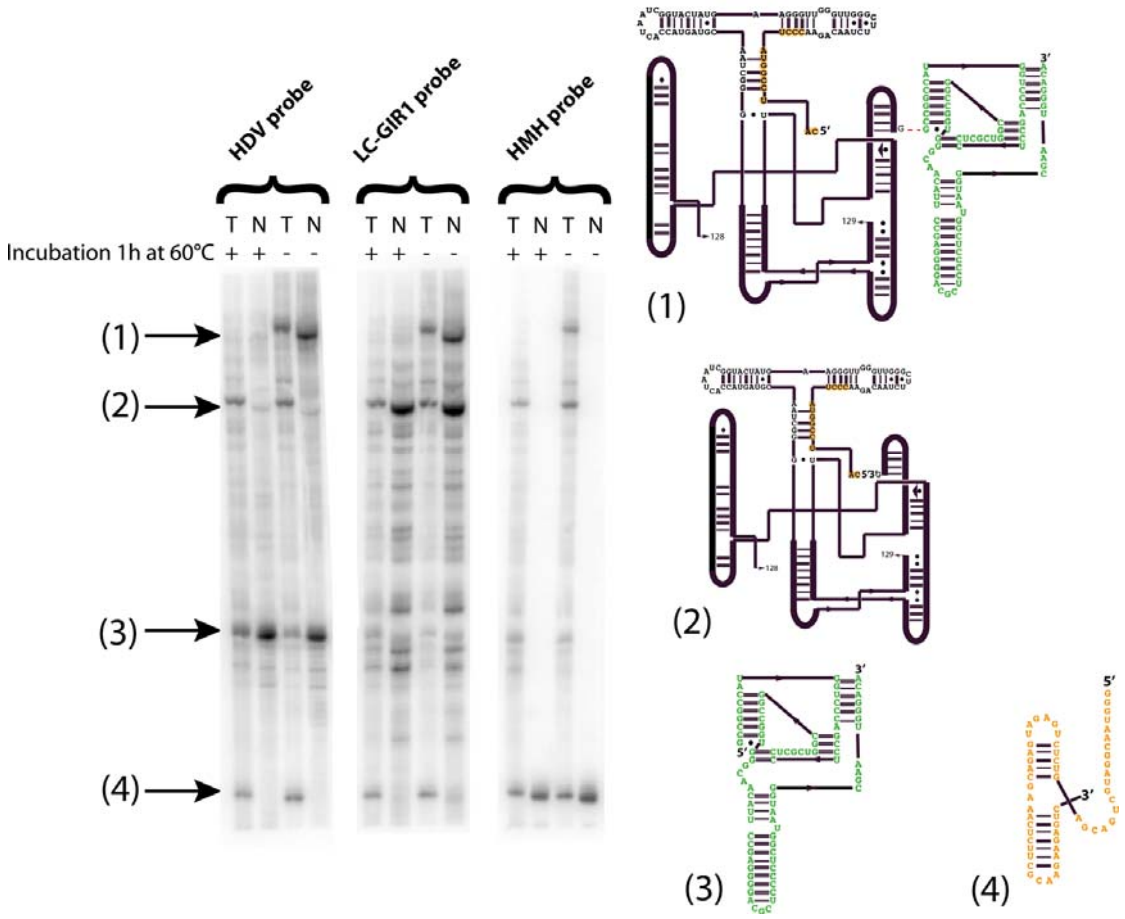


Figure 7: Northern blot analysis of a transcription reaction of the ternary ribozyme construct. Two hours transcription Body-labeled (T) and cold (N) reactions were performed and fractionated side by side by PAGE. Northern blot analysis was performed using a set of probes specific for (i) HδV, (ii) LCrz, and (iii) the hammerhead ribozymes. To complete HδV cleavage reactions, samples were incubated 1 hour at 60°C. Four main RNA products are visible on the membranes. (1) corresponds to the lariat-capping GIR1 ribozyme tethered to the HδV ribozyme. (2) is the correctly processed lariat-capping GIR1 ribozyme. (3) and (4) corresponds to the two cleaved flanking ribozymes. It is important to note that cleavage by the hammerhead ribozyme goes to completion during the transcription reaction implying that the ternary ribozyme transcript cannot be detected.

Table 1: PCR program

Step	Temperature (°C)	Time	Nb of cycles
Initial denaturing	95	3 min	1
Denaturing	95	30 s	30
Annealing	45*/50**	1 min	
Elongation	72	1 min	
End	4	10 min	1

*Annealing temperature for PCR of section 3.2.1 step 4.

**Annealing temperature for PCR of section 3.3 step 3.

As for constructs described in II.A.1.a, a HDV ribozyme is added downstream of LCrz_cp constructs to obtain a well-defined 3' end. 109 circularly permutated constructs have been designed out of which 17 have been tested for crystallization. The simplest construct is DiLCrz_cp "wild type" that is the 166.22 sequence plus the 5'UUCG3' tetralopp of DP2 **(Figure 35)**. One variant of this DiLCrz_cp construct is DiLCrz_cp_U1A that harbors a U1A binding site in the loop of P8 **(Figure 35)**.

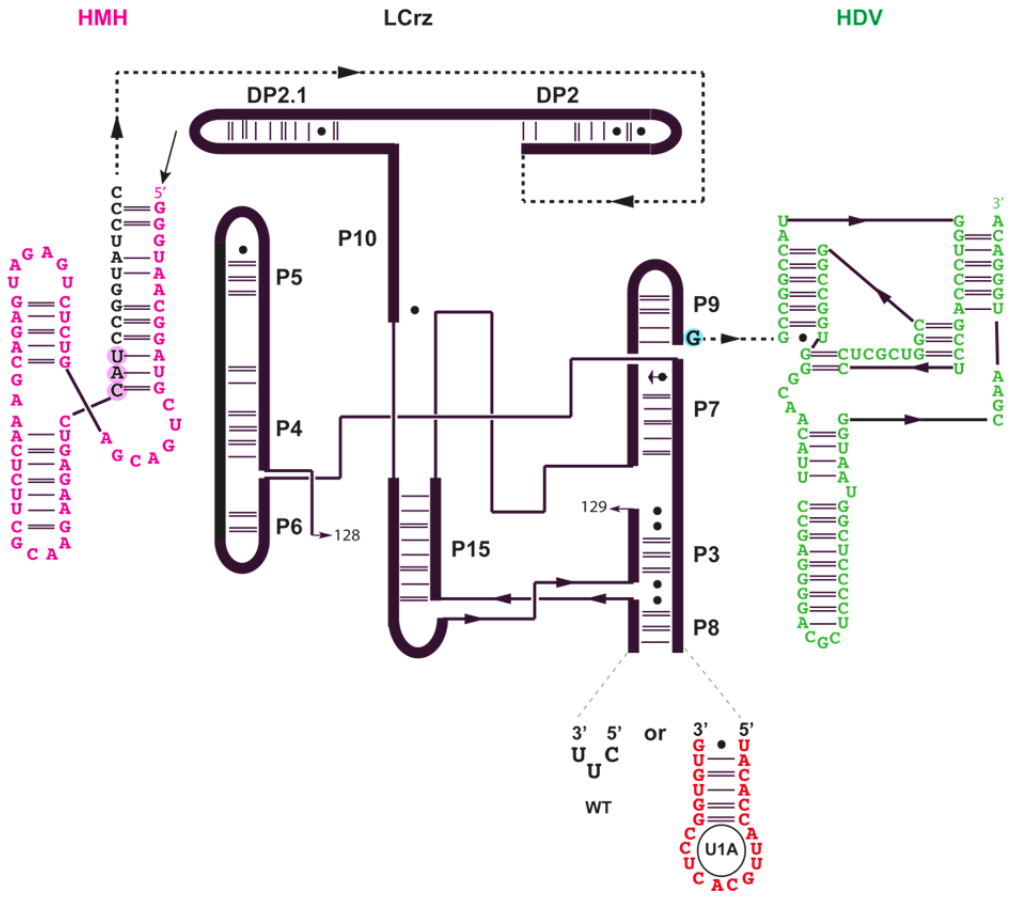


Figure 35: Diagram of two DiLCrz circularly permutated constructs: DiLCrz wild type in black with the HMH13 ribozyme upstream (pink) and the HDV ribozyme downstream (green). Unmodified P8 loop in shown in black and the alternative U1A binding sequence in red. The three nucleotides forming the post-cleavage lariat are circled in pink and G_{229} in blue. *In vitro* transcription starts at HMH ribozyme 5' end (black arrow).

Although the diffraction limit of our first crystals was poor (15 Å), DiLCrz_TLR3 construct **(Figure 31)** conferred a first success at crystallizing the LC ribozyme. We thus decided to take advantage of the circularly permutated constructs and of the tetraloop/tetraloop-receptor motifs to avoid crystallizing a mix of different RNA in order to simplify our protocols. To do so, we engineered stem-loops containing either a tetraloop or a tetraloop receptor, or both **(Figure 36)**. U-rich sequences between the ribozyme core and the inserted motifs and between motifs themselves have been chosen to influence the flexibility of the engineered stems (Jaeger et al, 2001). Engineered stems can be inserted in the loop of DP2 and / or in the loop of P8 **(Figure 37)**. 96 different constructs have been designed, out of which we used four for crystallization trials **(Figure 38)**.

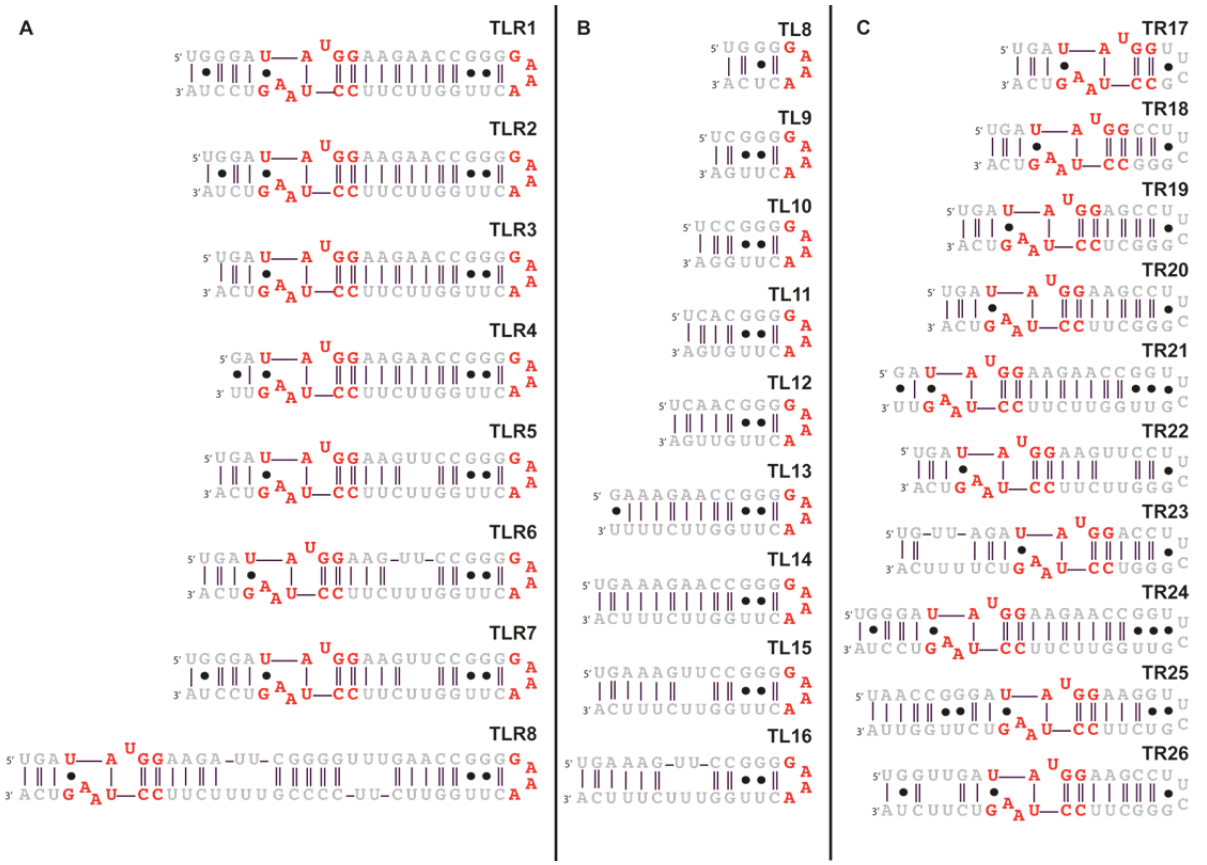


Figure 36: Engineered stems to be inserted in place of L8 and/or LP2. TL and TR motifs are shown in red and spacer sequences in grey. (A) Stems containing TL and TR motifs. (B) Stems containing TL motifs. These stems will be inserted in combination with a stem containing a TR motif in the same RNA construct. (C) Stems containing TR motifs.

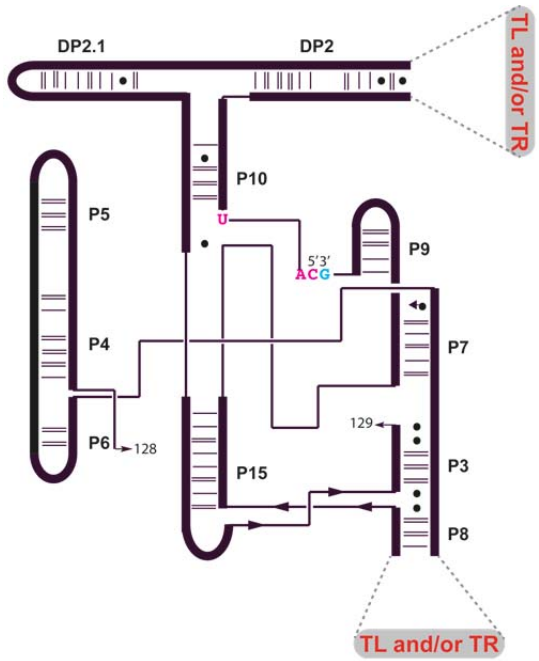


Figure 37: Diagram of DiLCrz_cp construct with open L8 and DP2 where the engineered stems presented in Figure 37 are to be inserted. Nucleotides forming the lariat are shown in pink and G_{229} in blue.

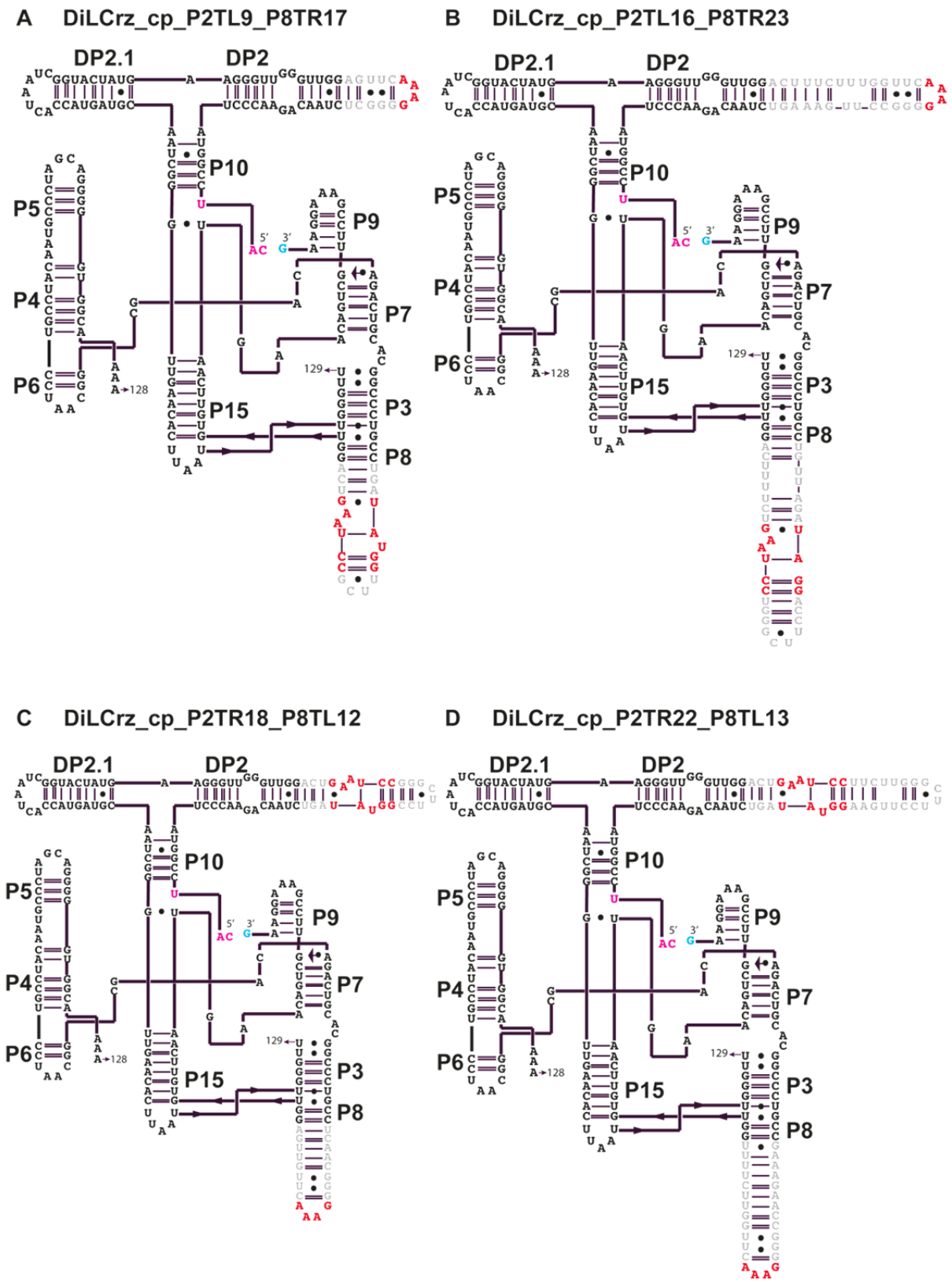


Figure 38: Four DiLCrz_cp harboring engineered stems shown in Figure 37 that have been used for crystallization trials. (A) and (B) harbor a TL motif in place of LP2 and a TR motif in place of L8. (C) and (D) harbor a TR motif in place of LP2 and a TL motif in place of L8. Nucleotides forming the lariat are shown in pink and G_{229} in blue.

After solving the structure of the DiLCrz_cp construct (see Results section) we noticed that because of crystal packing clashes the first uridine of the 5'UUCG3' DP2 loop and the two uridines from L8 are missing in the density map. To check whether better crystal packing would increase the diffraction limit, we designed a DiLCrz_cp version that is shorter by one AU base pair in DP2 **(Figure 39)**.

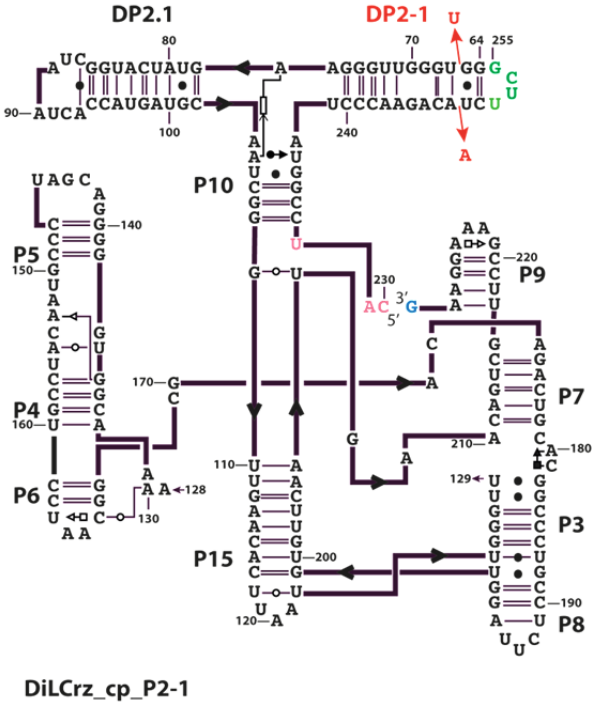


Figure 39: Diagram of DiLCrz_cp_P2-1 construct that has one A-U base pair less (in red) in the DP2 stem in comparison the DiLCrz_cp construct. The 5'UUCG3' tetraloop closing DP2 is shown in green, the nucleotides forming the post-cleavage lariat are shown in pink and G_{229} in blue.

c) Circularly permutated NaLCrz and AlLCrz

As for DiLCrz, circularly permutated constructs of NaLCrz and AlLCrz have been designed with an optimized extended HMH upstream and a HDV ribozyme downstream of the LCrz sequences. One difference between DiLCrz and the two other LCrz is that naturally occurring NaLCrz and AlLCrz sequences start in P2.1 therefore the 5'UUCG3' tetraloop is added to this stem. Because NaLCrz_cp and AlLCrz_cp are larger than DiLCrz_cp and because the secondary structure of their peripheral domains have not yet been extensively characterized we designed two versions of each ribozyme with more or less base pairs in their P2.1 stems **(Figure 40 and 41)**.

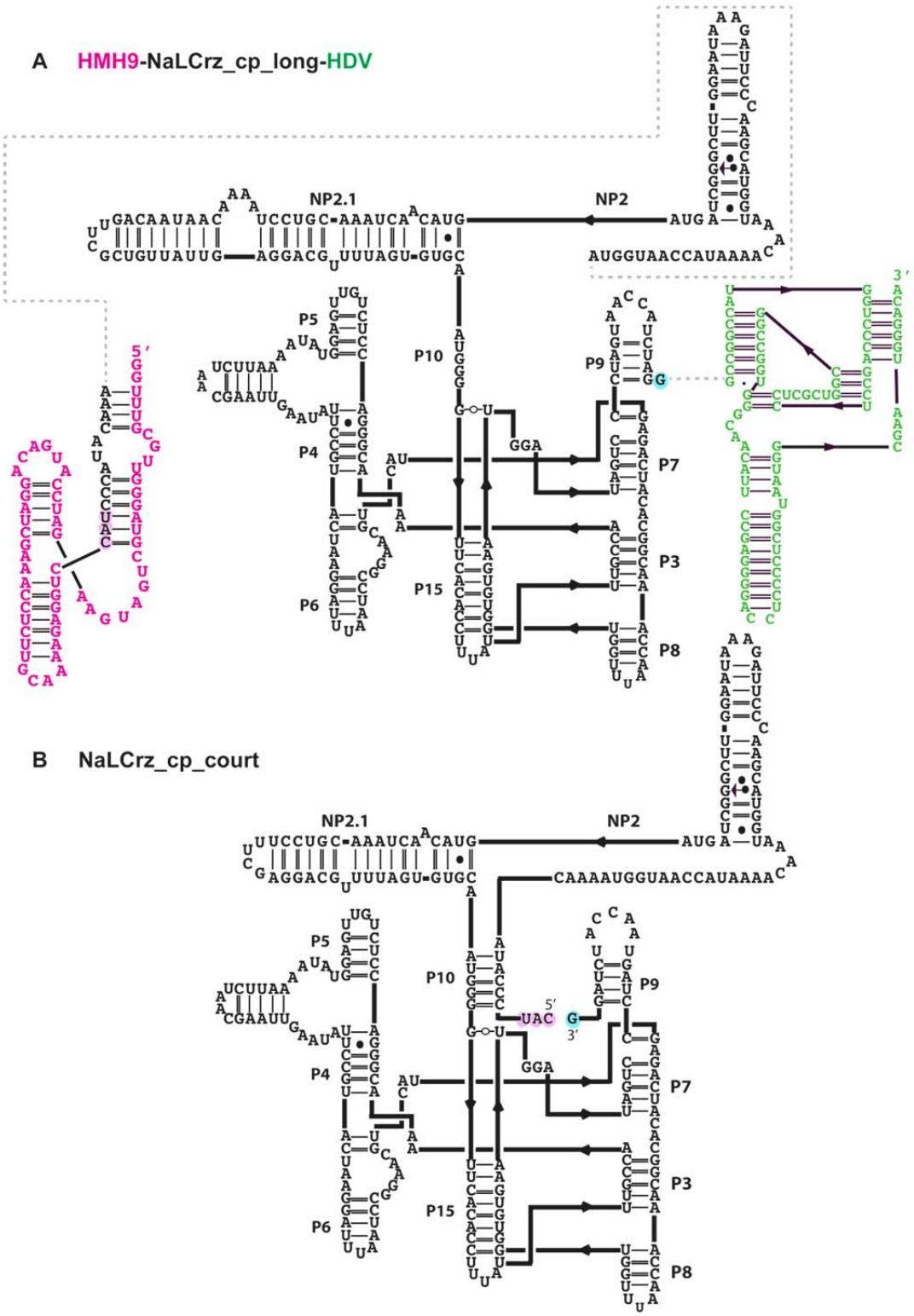


Figure 40: Diagram of two NaLCrz circularly permutated constructs. (A) NaLCrz_cp_long in black with the HMH9 ribozyme upstream (pink) and the HDV ribozyme downstream (green). The NP2.1 stem in closed by a 5'UUCG3' tetraloop. The three nucleotides forming the post-cleavage lariat are circled in pink and G_{229} in blue. *In vitro* transcription starts at HMH ribozyme 5' end. (B) NaLCrz_cp_court after cleavage of both, upstream and downstream, ribozymes in the course of the transcription. NaLCrz_cp_court is shorter of 22 nucleotides in NP2.1 in comparison to NaLCrz_cp_long.

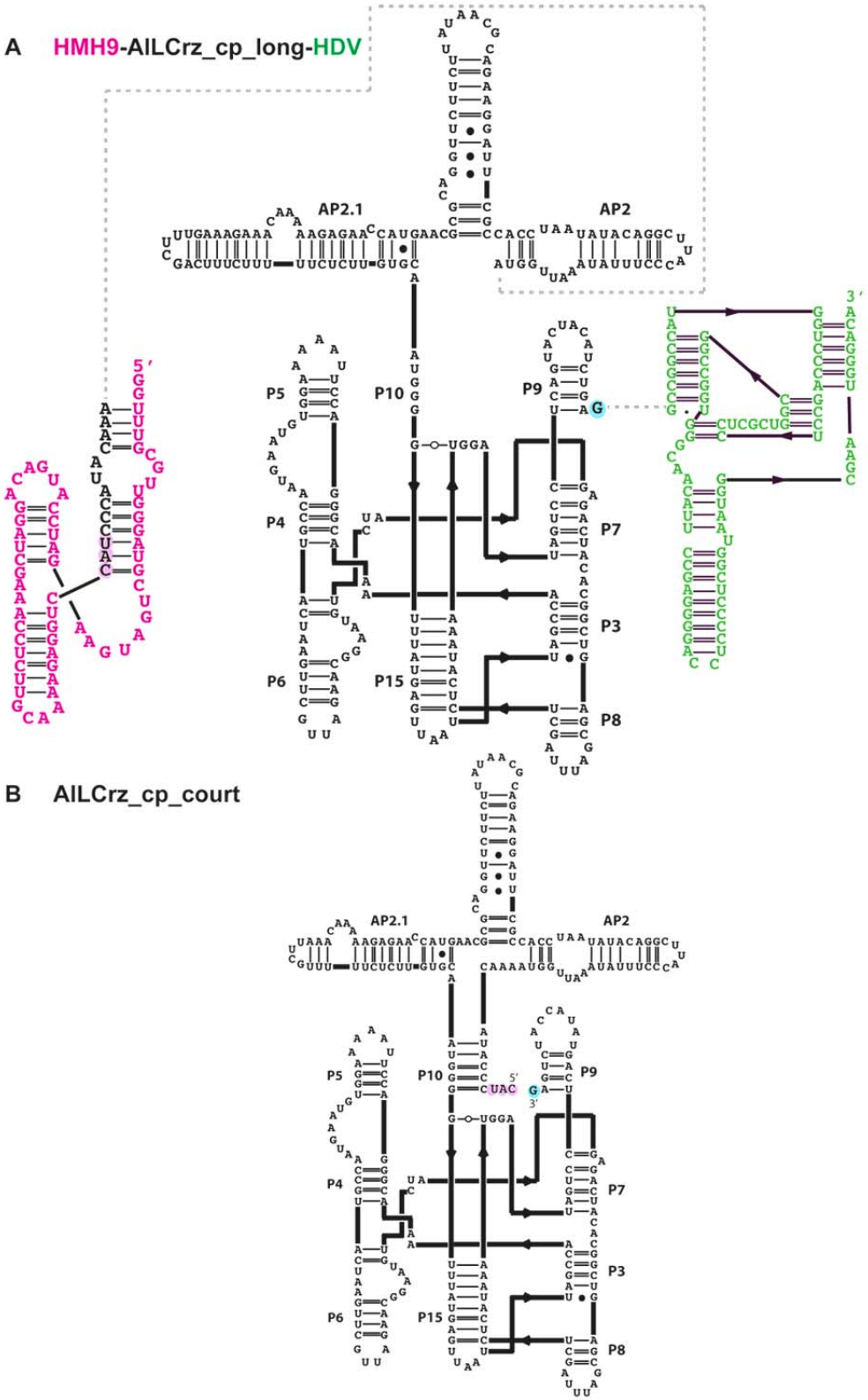


Figure 41: Diagram of two AlLCrz circularly permutated constructs. (A) AlLCrz_cp_long in black with the HMH9 ribozyme upstream (pink) and the HDV ribozyme downstream (green). The AP2.1 stem in closed by a 5'UUCG3' tetraloop. The three nucleotides forming the post-cleavage lariat are circled in pink and G_{229} in blue. *In vitro* transcription starts at HMH ribozyme 5' end. (B) AlLCrz_cp_court after cleavage of both, upstream and downstream, ribozymes in the course of the transcription. AlLCrz_cp_court is shorter of 12 nucleotides in AP2.1 in comparison to AlLCrz_cp_long.

2. LCrz RNA production and purification

All molecular biology methods have been adapted from (Sambrook, 1989). All solutions have been filtered on 22 µm filter and sterile glass or plastic material has been used wherever possible.

a) Competent *Escherichia coli* XL-1 Blue cells preparation

Competent *Escherichia coli* XL-1 Blue cells are prepared in the laboratory according to the below protocol below (Inoue et al, 1990).

1. Inoculate a starter culture of 3 mL of LB/Tetracycline (30 µL/mL) with 50 µL of competent cells from the initial strain.
2. Incubate at 18 °C for 6 hours.
3. Inoculate 250 mL of LB/Tetracycline (30 µL/mL) with the 3 mL of starter culture.
4. Incubate at 18 °C until the OD of the bacterial culture reaches 0.6 at a λ= 600 nm.
5. Pre-cool at 4 °C all materials and solutions that will be used from this step and work in the cold room (4 °C). These are very important precautions to take in order to obtain efficient competent cells.
6. Centrifuge the bacterial culture at 2500 g fro 10 min at 4 °C. Discard the supernatant.
7. Resuspend the pelleted cells in 80 mL of TBjap (10 mM Pipes, 15 mM $CaCl_2$, 250 mM KCl pH adjust pH to 6.7, add $MnCl_2$ to a concentration of 55 mM and DMSO to 1.96%).
8. Keep the suspension on ice for 10 min.
9. Centrifuge the bacterial culture at 2500 g for 10 min at 4 °C. Discard the supernatant.
10. Resuspend the pelleted cells in 18.6 mL of TBjap and DMSO to a final concentration of 7%.
11. Keep the suspension on ice for 10 min.

12. Very delicately aliquot the competent cells in Eppendorf tubes and immediately freeze them in liquid nitrogen.
13. Quickly store aliquots of competent cells at -80 °C.

b) Cloning of the full length constructs

All LCrz RNA constructs have been transcribed from plasmid DNA templates obtained by cloning of an insert generated by 2 or 4 successive PCRs. The 5' PCR primers have been designed to contain an EcoRI restriction site, a T7 promoter, the sequence of the co-transcriptional self-cleaving HMH ribozyme (for the circularly permutated constructs only) and the 5' sequence of LCrz. 3' PCR primers included a BamHI restriction site, a post-transcriptional self-cleaving HDV ribozyme and the 3' sequence of the LCrz **(Figure 42 and Appendix I)**.

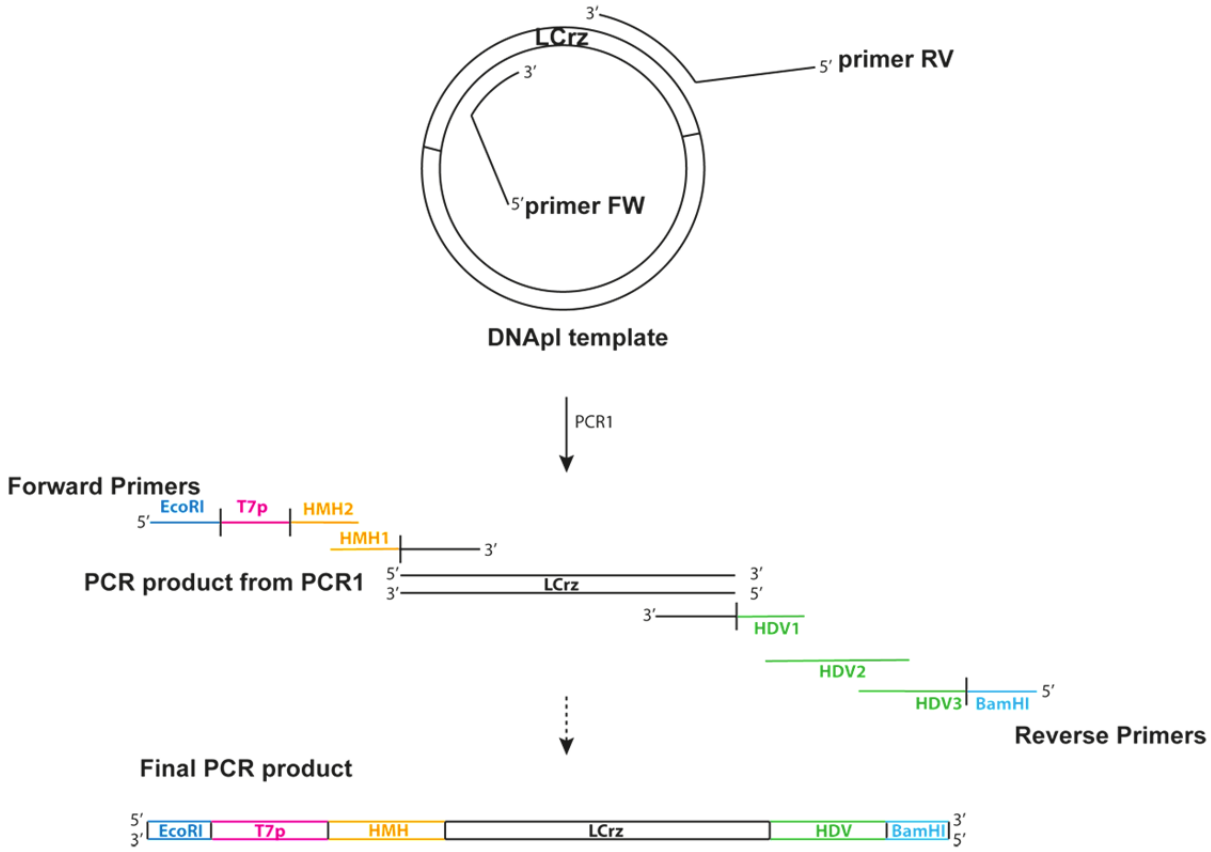


Figure 42: Scheme of the successive PCRs necessary to clone full length constructs including EcoRI restriction site (blue), the T7 promoter (pink), extended HMH ribozyme (orange), LCrz (black), HDV ribozyme (green) and BamHI restriction site (cyan).

1. Design forward and reverse primers for the successive PCRs **(Appendix II)**. For optimal yield, forward and reverse primers have been designed to have identical melting temperature.
2. Prepare 50 µL of PCR1 reaction containing: 50 pg of template plasmid DNA, 0.5 µL of Phusion® High-Fidelity DNA Polymerase (2000 U/mL - New England BioLabs *inc*), 10 µL of Phusion® HF Buffer 5x (New England BioLabs *inc)*, 1.2 µM of forward primer, 1.2 µM of reverse primer, 400 µM dNTP mixture, water up to 50 µL. Samples are kept on ice while preparing the PCR reaction.
3. Run the PCR reaction in a thermal cycler following the set up displayed in **Table 2**. Insert samples in the thermal cycler when the temperature of the sample holder is at 90 °C minimum.

Step	Temperature (°C)	Time	Nb of cycles
Initial denaturing	95	3 min	1
Denaturing	95	30 s	30
Annealing	50	1 min	
Elongation	72	1 min	
End	4	10 min	1

Table 2: PCR program to produce the full-length constructs inserts

4. Mix the PCR reaction to 10 µL glycerol blue loading buffer 6x (50% glycerol, 0.1% xylene cyanol, 0.1% bromophenol blue) and load the entire PCR reaction onto a 2% (w/v) agarose gel (2 g of agarose for routine use - Sigma Aldrich - dissolved in 100 mL of TBE 1x).
5. After ethidium bromide staining, observe the DNA migration pattern with UV light.
6. Cut the DNA band of interest out of the gel.
7. Recover the DNA insert in 30 µL of water with the QIAquick® Gel Extraction Kit (Qiagen). To avoid any trouble with next steps of the cloning process, 10% of ammonium acetate 3 M pH 5.2 is added to the recovered DNA and precipitated with 3 volumes of pure ethanol. DNA is then re-suspended in 30 µL of water.

8. Prepare 50 μL of PCR2 reaction containing: 50 pg of PCR1 product as template DNA and perform step 2 to 7 until the insert contains the full length sequence. Alternatively if the product from PCR1 is clean and show no additional product from various sizes, 1 μL of crude PCR1 reaction can be used as template for PCR2.
9. Insert the purified PCR product into a pUC19 plasmid using In-Fusion® HD Cloning Kit (Clontech). An alternative protocol has also been used: mix a ½ molar ratio of pUC19 plasmid that has been double digested by EcoRI and BamHI and purified on agarose gel and the insert that has also been double digested by EcoRI and BamHI and purified on agarose gel with 0.5 μL of PEG 6000 24% w/v (Hampton Research) and water up to 17.5 μL. Incubate the sample at 65 °C for 5 min. Add 0.5 μL of T4 DNA ligase (30 U/μL- Thermo Scientific) and 2 μL of T4 DNA ligase 10x (400 mM Tris-HCl, 100 mM MgCl2, 100 mM DTT, 5 mM ATP pH 7.8 at 25 °C-Thermo Scientifc). Incubate the sample at 20 °C for 1 hour.
10. Directly transform competent *Escherichia coli* XL-1 Blue cells (prepared as described in II.B.2.a) with either the In-Fusion® HD Cloning Kit product or with 10 μL of the ligation reaction.
11. Spread the transformed competent *Escherichia coli* XL-1 Blue cells samples on Petri dish containing LB/Ampicillin (100 μg/mL) and incubate overnight at 37 °C.
12. Grow 3 mL of LB/Ampicillin (100 μg/mL) cultures from individual colonies.
13. Extract DNA using a Miniprep Spin Kit (QIAprep® Spin Miniprep Kit – Qiagen).
14. The extracted DNA is sequenced using the M13 sequencing primers by GATC Biotech (Jakob-Stadler-Platz 7 - 78467 Konstanz – Germany).
15. Prepare glycerol stocks of the clones containing the desired sequences by mixing 1 mL of saturated bacterial culture to 600 μL of sterilized 50% glycerol solution in 2 mL screw cap tubes. Store the glycerol stocks at -80 °C.

c) Site directed mutagenesis

Because the U203C mutation shown to stabilize NaLCrz (Jabri & Cech, 1998) resulting from a PCR artefact in DiLCrz_cp, has been established indispensable for

crystallization of DiLCrz **(Figure 43)** and does not impair branching **(Figure 44)** we decided to mutate the DiLCrz_active, DiLCrz_P2-1, AlLCrz_cp_long, AlLCrz_cp_court and NaLCrz_cp_long and NaLCrz_cp_court at this equivalent position in P15 **(Figure 45 and Appendix I)**.

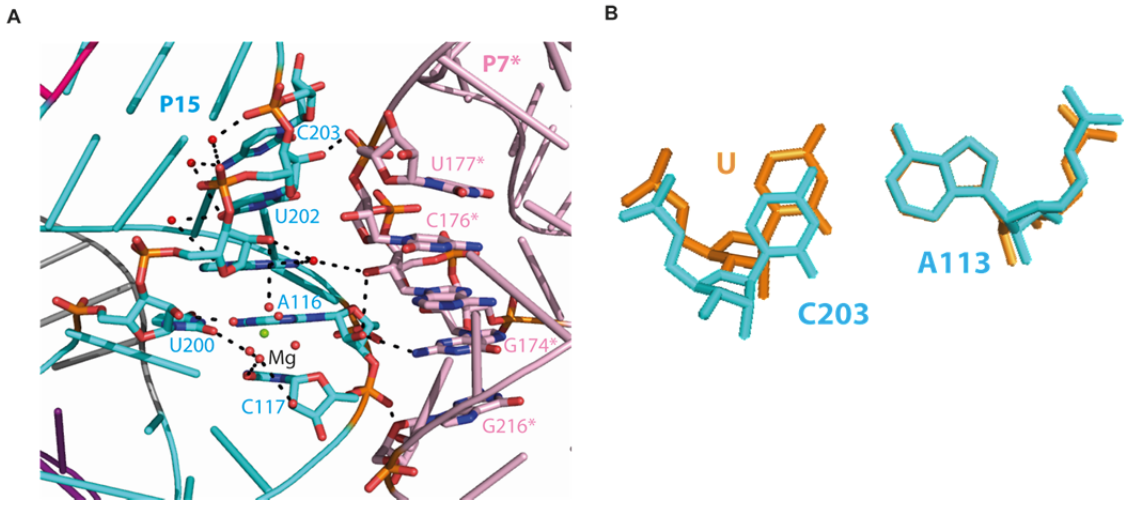


Figure 43: (A) Crystal packing contacts between nucleotides neighboring the U203>C mutation in P15 of one DiLCrz_cp (cyan) and nucleotides of P7 of a symmetry-related DiLCrz_cp molecule (light pink). (B) Superposition of the A113-C203 WC *trans* base paire (cyan) of P15 to a canonical A-U WC *trans* base pair (orange).

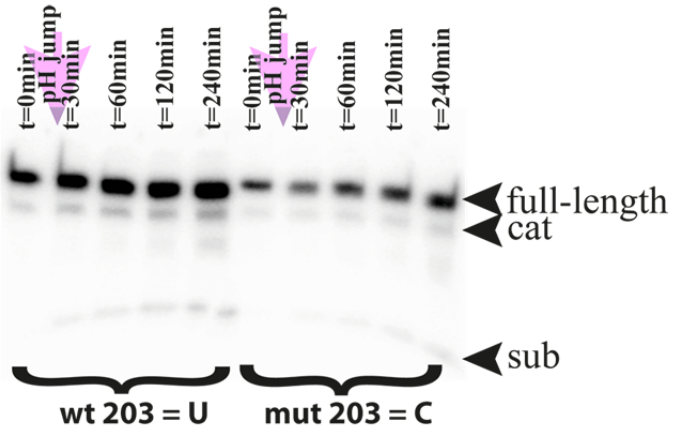


Figure 44: Cleavage assay of the U203C mutated DiLCrz_active. [α-32P] UTP body labeled wild type active DiLCrz and mutated U203C active DiLCrz have been assayed for cleavage efficiency. RNA is denatured at 95°C for 1 min and refolded in 1 M KCl, 25 mM MgCl2, 10 mM Acetate pH 5.0 at 45 °C for 5 min. An aliquot is taken right after refolding, t=0 min. To trigger the cleavage reaction, 4 volumes of 1 M KCl, 25 mM MgCl2, 47,5 mM HEPES pH 7.5 is added to the folded RNA. Cleavage reaction takes place at 45 °C for 4 h (Decatur, 1995). Aliquots are taken at t=30 min, t=60 min, t=120 min and t=240 min. The reaction is stopped by addition of EDTA-urea blue loading buffer. Aliquots are loaded on a 8% denaturing PAGE. Three distinct products are observed: full-length DiLCrz is the uncleaved ribozyme (188 nt), cat is the catalytic core (166 nt) and sub is the cleaved substrate (22 nt).

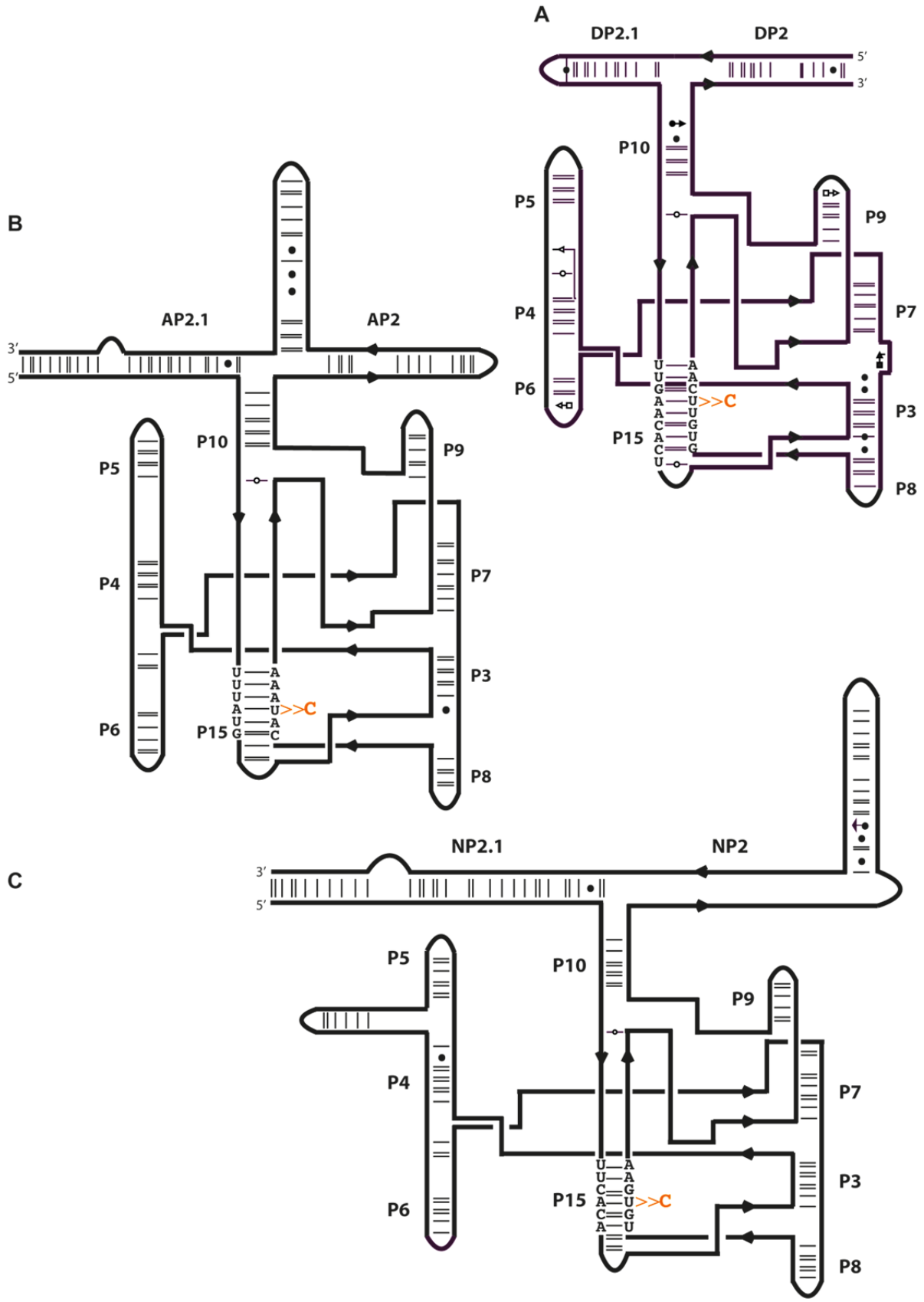


Figure 45: Diagram showing the position of the U to C mutation (orange) in P15 stem of (A) DiLCrz_active, (B) AlLCrz and (C) NaLCrz.

1. Design forward and reverse primers so that they are perfectly complementary and respect the following rules: (1) primers should be between 25 and 45 bases in length, with a melting temperature superior to 75 °C, (2) the desired point mutation should be close to the middle of the primer with ~15 bases of template-complementary sequence on both sides and (3) the sequences of the primers should have at least a G or C at its ends and a minimum GC content of 40% **(Appendix III)**.
2. Prepare 50 μL of PCR reaction containing: 50 ng of template plasmid DNA, 0.5 μL of the Phusion® High-Fidelity DNA Polymerase (2000 U/mL - New England BioLabs *inc*), 10 μL of the Phusion® HF Buffer 5x (New England BioLabs *inc)*, 1.2 μM of forward primer, 1.2 μM of reverse primer, 400 μM dNTP mixture, 10% DMSO, water up to 50 μL. Samples are kept on ice while preparing the PCR reaction.
3. Run the PCR reaction in a thermal cycler following the set up displayed in **Table 3**. Insert samples in the thermal cycler when the temperature of the sample holder is at 90 °C minimum.

Step	Temperature (°C)	Time	Nb of cycles
Initial denaturing	99	3 min	1
Denaturing	95	30 s	15
Annealing	50	1 min	
Elongation	72	1 min 30 s	
End	4	10 min	1

Table 3: PCR program for site directed mutagenesis

4. Mix the PCR reaction to 10 μL glycerol blue loading buffer 6x (50% glycerol, 0.1% xylene cyanol, 0.1% bromophenol blue) and load the entire PCR reaction onto a 2% (w/v) agarose gel (2 g of agarose for routine use - Sigma Aldrich - dissolved in 100 mL of TBE 1x).
5. After ethidium bromide staining, observe the DNA migration pattern with UV light.
6. Cut the DNA band of interest out of the gel.

7. Recover the DNA plasmid in 30 μL of water with the QIAquick® Gel Extraction Kit (Qiagen). To avoid any trouble with next steps of the cloning process, 10% of ammonium acetate 3 M pH 5.2 is added to the recovered DNA and precipitated with 3 volumes of pure ethanol. DNA is then resuspended in 30 μL of water.
8. Remove the PCR DNA template by mixing: 30 µL of the recovered DNA, 4 µL of the Tango buffer 10x (33 mM Tris-acetate pH 7.9 at 37 °C, 10 mM Mg-acetate, 66 mM K-acetate, 0.1 mg/mL BSA – Thermo Scientific), 1 µL DpnI (10 U/µL – Thermo Scientific), water up to 40 µL.
9. Directly transform competent *Escherichia coli* XL-1 Blue cells with 20 µL of the DpnI reaction.
10. Spread the transformed competent *Escherichia coli* XL-1 Blue cells on LB/Ampicillin (100 µg/mL) Petri dish and grow overnight at 37 °C.
11. Grow 3 mL of LB/Ampicillin (100 µg/mL) cultures from individual colonies.
12. Extract DNA using a Miniprep Spin Kit (QIAprep® Spin Miniprep Kit – Qiagen).
13. The extracted DNA is sequenced using the M13 sequencing primers by GATC Biotech (Jakob-Stadler-Platz 7 - 78467 Konstanz – Germany).
14. Prepare glycerol stocks of the clones containing the desired sequences by mixing 1 mL of saturated bacterial culture to 600 µL of sterilized 50% glycerol solution in 2 mL screw cap tubes. Store the glycerol stocks at -80 °C.

d) Plasmid preparation

The amount of plasmid needed, hence the amount of bacterial culture, depends on the size of the *in vitro* transcription you are planning to perform. Typically, I used 300 µg of plasmid DNA to perform a 6 mL *in vitro* transcription to obtain micrograms of pure RNA for crystallization.

1. Prepare a starter bacterial culture by inoculating 3 mL of LB/Ampicillin (100 µg/mL) with a tip dipped in the glycerol stock of your plasmid of interest. Grow the 3 mL cultures at 37 °C for 6 hours. Starter bacterial culture is not mandatory however it generally leads to a better bacterial growth hence a better DNA plasmid yield.

2. Inoculate 250 mL of LB/Ampicillin (100µg/mL) with the 3 mL of starter culture. Grow over-night at 37°C.
3. Centrifuge the bacterial culture at 4000 g for 20 min at 4 °C. Discard the supernatant.
4. Extract plasmid from the bacterial pellet with the PureLink® HiPure Plasmid Filter Maxiprep Kit (Invitrogen).
5. For *in vitro* transcription, linearize the extracted plasmid with PstI restriction enzyme following the enzyme provider instructions.
6. Check on an agarose gel that the DNA plasmid is fully digested unless this will decrease the yield of the upcoming transcription.

e) Large scale *in vitro* transcription of the full length RNA using linearized plasmids as templates

Large scale *in vitro* transcription (up to 6 mL) is indispensable to obtain micrograms of pure RNA for crystallization

1. Set up 6 mL transcription reaction containing: 300 µg of linearized plasmid DNA, 1.2 mL of TMSDT 5x Buffer (200 mM Tris-HCl pH 8.1, 110 mM $MgCl_2$, 5 mM spermidine, 25 mM DTT, 0.05% Triton), 5 mM ATP, 5 mM GTP, 5 mM CTP, 5 mM UTP, 6 µL of alkaline phosphatase (1 U/µL in TMSDT 1x - Sigma-Aldrich), T7 RNA polymerase as indicated by your provider (we produce our T7 RNA polymerase in the laboratory), water up to 6 mL.
2. Incubate at 37 °C for 4 h.
3. Instantly incubate at 60 °C for 1 hour to allow HDV ribozyme cleavage.

f) LCrz RNA purification

Four different methods have been tried to purify the LCrz RNA: native PAGE, denaturing PAGE, denaturing PAGE combined to HPLC and BioTrap (Schleicher & Schuell). Only the following protocol using denaturing PAGE purification allows us to obtain RNA suitable for crystallization.

1. Add 10% of ammonium acetate 3 M pH 5.2 to the 6 mL transcription and precipitated with 3 volumes of pure ethanol.
2. Centrifuge at 20 000 g for 20 min in 50 mL BD Falcon ™ conical tubes.
3. Discard the supernatant and resuspend the pellet in 1 mL of water.
4. Add to the resuspended pellet 1 mL of urea blue loading buffer 2x (8 M urea, 0.025% xylene cyanol, 0.025% bromophenol blue).
5. To check that transcription of LCrz and cleavage of the 5' HMH and 3' HDV ribozymes were efficient load 10 µL of the transcription mixed to urea blue on an 8% denaturing PAGE (8% acrylamide/bis-acrylamide (19:1), 8 M urea in TBE 1x). Run the gel at 50 W until the bromophenol blue reaches the bottom of the gel (approximately 6 hours). Gel size: 35 cm width * 42 cm length * 0.08 cm thickness.
6. To purify LCrz transcription, load the totality of the transcription mixed to urea blue loading buffer on an 8% denaturing PAGE. Run the gel at 50 W until the bromophenol blue reaches the bottom of the gel (approximately 6 hours). Gel size: 35 cm width * 42 cm length * 0.2 cm thickness.
7. Place the gel on a silica plate and observe the RNA migration pattern with UV light.
8. Cut the RNA band of interest out of the gel.
9. Crush the gel band through a syringe.
10. For passive elution of the RNA, add 11 mL of water to the crushed gel in a 50 mL tube placed at 4 °C on a roller mixer (Stuart ™ Scientific) over-night.
11. Filter the sample through a 150 mL Analytical Filter Unit (Nalgene ®)
12. Add 10% of ammonium acetate 3 M pH 5.2 to the filtrate and precipitated with 3 volumes of pure ethanol.
13. Centrifuge at 27 500 g for 20 min in 50 mL BD Falcon ™ conical tubes.
14. Discard the supernatant and resuspend the RNA pellet in 50 µL of water.
15. Determine the RNA concentration from OD measurements at λ = 260 nm (Nanodrop – LabTech).
16. Store the pure RNA at -20 °C.

3. Purification of the substrate strand RNA

The modified RNA substrate (5' A A G G A A G * C A U C C G G U A U C C C A A G A C A A U G 3') added *in trans* to the catalytic core (constructs described in II.A.1.a **Figure 31**) has been chemically synthesized (Dharmacon – Thermo Scientific) and purified by HPLC in the laboratory.

1. Deprotect the 2'-ACE® protected RNA by dissolving the dry pellet of RNA in 400 µL of the 2'-deprotection buffer (100 mM acetic acid adjusted to pH 3.8 with TEMED – Thermo Scientific) and incubating the sample at 60 °C for 30 min.
2. Dry the sample using a SpeedVac® (Savant).
3. Resuspend the dry pellet of deprotected RNA in 1 mL of buffer A (4 M urea, 15 mM $NaClO_4$, 1% acetonitrile, 20 mM Mes-KOH ph 6.5).
4. Equilibrate the ion exchange chromatography column (Nucleopac PA-100 – Dionex) with buffer A and heat the column to 60 °C (CROCOCIL).
5. Inject the RNA sample in the chromatography column.
6. Inject elution buffer B (4 M urea, 400 mM $NaClO_4$, 1% acetonitrile, 20 mM Mes-KOH pH 6.5) to allow progressive elution of the RNA.
7. Collect fractions of pure RNA sample.
8. Mix an aliquot of each fraction to an equal volume of urea blue loading buffer 2x (8 M urea, 0.025% xylene cyanol, 0.025% bromophenol blue) and analyse their purity by running a 20% denaturing PAGE (20% acrylamide/bis-acrylamide (19:1), 8 M urea in TBE 1x).
9. Reveal the pattern of migration in the gel of the pure RNA fractions by staining the gel with toluidine blue staining solution (1 g of tuluidine blue, 400 mL ethanol, 10 mL acetic acid, water up to 1 L).
10. Desalt the pure RNA fractions using inverse phase columns (Sep-Pak® C18 Classic Cartridges short body – Waters) following the provider instructions.

4. Over-expression and purification of the U1A protein

The U1A protein mutated at positions 31 (Y31H) and 36 (Q36R) (MAVPETRPNH TIYINNLNEK IKKDELKKSL HAIFSRFGQI LDILVSRSLK MRGQAFVIFK EVSSATNALR SMQGFPFYDK PMRIQYAKTD SDIIAKMK) that is added to the RNA harboring a U1A binding site in the loop of P8 (constructs described in II.A.1.a **Figure 31** and II.A.1.b **Figure 35**) has been produced in the laboratory. We thank Adrian Ferré d'Amaré for giving us the plasmid pU1A_31H36R containing the U1A gene.

1. Transform BL21 Star™ (DE3)pLysS One Shot® (Invitrogen™) competent cells with 6 ng of plasmid containing the sequence for U1A over-expression.
2. Spread the transformed competent cells on a LB/Ampicillin (100 µg/mL) /Chloramphenicol (34 µg/mL) agar plate. Incubate the plate at 37 °C over-night.
3. Prepare a starter bacterial culture by inoculating 3 mL of LB/Ampicillin (100 µg/mL) /Chloramphenicol (34 µg/mL) medium with a tip dipped in selected colony from the agar plate. Grow the 3 mL cultures at 37 °C over-night.
4. Inoculate 2 L of bacterial culture with 1 mL of the starter bacterial culture.
5. Allow bacteria to grow until the DO at λ = 600 nm reaches 0.4.
6. Induce U1A gene over expression by adding IPTG at a final concentration of 0.5 mM to the 2 L of medium. Grow at 37 °C for 4 hours.
7. Centrifuge at 4000 g for 30 min at 4 °C.
8. Resuspend the bacterial pellet corresponding to 1 L of bacterial culture in 25 mL of lysis buffer (20 mM Tris-HCl pH 7.4, 500 mM NaCl, 10% glycerol, 1 mM EDTA, 1 mM PMSF, 5 µg/mL Leupeptine, 1% Aprotinine (v/v), 0.1% NP-40).
9. Sonicate the sample 10 times for 30 s.
10. Centrifuge at 10 000 g for 10 min at 4 °C.
11. Collect the supernatant and keep it at 4 °C.
12. Check for the presence of the U1A protein in the supernatant by running a SDS 12% PAGE gel (12% acrylamide/bis-acrylamide (37.5:1), 1.5% Tris, 9.4% Glycine, 0.5% SDS).
13. Add PEI dropwise to the supernatant to reach 0.5% finale concentration of PEI. During this step shake gently and continuously the sample at 4 °C.

14. Centrifuge at 20 000 g for 5 min at 4 °C.
15. Dialyze the lysate twice in buffer A (100 mM KCl, 25 mM HEPES-KOH pH 7.5, 10 mM DTT, 0.5 mM EDTA, 0.5 mM PMSF)
16. Purify by FPLC (ÄKTA™ – GE Healthcare Life Sciences) using an ion exchange column (Sepharose Fast Flow - GE Healthcare Life Sciences).
17. Inject elution buffer B (500 mM KCl, 25 mM HEPES-KOH pH 7.5, 10 mM DTT, 0.5 mM EDTA, 0.5 mM PMSF) to allow progressive elution of the U1A protein.
18. Perform a second purification step using a gel-filtration column (HiLoad™ Superdex™ 75 pg - GE Healthcare Life Sciences) and buffer A.
19. Dialyze the U1A protein in buffer C (10 mM Phosphate buffer pH 7.5, 50 mM KCl, 1 mM DTT).
20. Perform a cation exchange chromatography (Bio-Scale Ceramic Hydroxyapatite CHT5-I – BIORAD).
21. Perform an elution gradient with buffer D (10 mM Phosphate buffer pH 7.5, 50 mM KCl, 1 mM DTT, 500 mM $(NH_4)_2SO_4$) to improve the U1A protein purity.
22. Dialyze the U1A protein in storage buffer (10 mM HEPES-KOH pH 7.5, 1 mM DTT, 0.1 mM EDTA).
23. Concentrate the U1A protein by tangential centrifugation at 5000 g for 10 min (Amicon MWCO 3kDa – Millipore). Store the U1A protein at -80 °C.

5. Crystallization

a) Sample preparation

In the case of purified LCrz_cp, the RNA has been prepared in the desired volume at the necessary concentration (100 μM in most cases) in a solution containing 100 mM KCL and 10 mM MES-KOH pH 6.5. The sample has been heated during 1 min at 95 °C followed by 4 min on ice. $MgCl_2$ up to 4 mM was then added, and the solution was placed at 55 °C for 10 min then at 4 °C overnight. The homogenous folding of the ribozyme has been verified by running a native 8% PAGE (8% acrylamide/bis-acrylamide (19:1) in TB 1x).

In the case of LCrz core, the RNA was prepared as the LCrz_cp, except that if necessary, the U1A protein was added at the same final concentration as the ribozyme core and the sample was further incubated at 25 °C for 20 min. RNA substrate strand is always added in excess. The sample is then placed at 4 °C overnight. The formation of the binary or ternary complex is verified by gel shift assay.

b) Sitting drops

Initial crystallization conditions were identified using commercial screens (Natrix HT, Index HT, PEGRx HT, Crystal screen HT, Crystal screen cryo HT from Hampton Research© and Nuc-Pro HTS from Jena Bioscience). Crystallization trays were prepared with the Mosquito® robot (TTP Labtech) in 96 wells Corning© plates by mixing a volume of LCrz RNA with an equal volume of reservoir solution. Crystallization plates were kept at different temperatures. All crystallization trials are enlisted in **Table 4 and 5**.

LCrz RNA construct	RNA concentration in µM	Incubation temperature in °C	Drop size in nL	Crystallization screen Index HT - Hampton Research	Crystallization screen PEGRx HT - Hampton Research	Crystallization screen Natrix HT - Hampton Research	Crystallization screen Crystal screen HT - Hampton Research	Crystallization screen Crystal screen Cryo HT - Hampton Research	Crystallization screen JBScreen Nuc-Pro HTS - JenaBioscience
DiLCrz_cp	100	20	400	x	x	x			x
DiLCrz_cp_U203C	88	20	400			x	x	x	
DiLCrz_cp_U203C	100	20	400	x	x	x			x
DiLCrz_cp_U203C	100	20	400			x			
DiLCrz_cp_U203C	200	20	400			x			
DiLCrz_cp_U203C	300	20	400			x			
DiLCrz_cp_U203C	500	20	400			x			
DiLCrz_cp_U203C	100	30	400			x			
DiLCrz_cp_U203C	200	30	400			x			
DiLCrz_cp_U203C	300	30	400			x			
DiLCrz_cp_U203C	500	30	400			x			
DiLCrz_cp_P2-1	100	20	400	x	x	x			x
DiLCrz_cp_U203C_P2-1	100	20	400	x	x	x			x
DiLCrz_cp_P2TL9_P8TR17	100	20	400	x	x	x			x
DiLCrz_cp_P2TL16P8_TR23	100	20	400	x	x	x			x
DiLCrz_cp_P2TR18P8_TL12	100	20	400	x	x	x			x
DiLCrz_cp_P2TR22_P8TL13	100	20	400	x	x	x			x
DiLCrz_active	100	20	400	x	x	x			x
DiLCrz_U203C_active	100	20	400	x	x	x	x	x	x
AlLCrz_cp_U252C_long	100	20	400	x	x	x			x
AlLCrz_cp_court	100	20	400	x	x	x			x
AlLCrz_cp_U252C_court	100	20	400	x	x	x			x
NaLCrz_cp_U261C_long	100	20	400	x	x	x			x
NaLCrz_cp_U261C_court	100	20	400	x	x	x			x

Table 4: Summary table of the crystallization trials done with the Mosquito® robot in 96 wells sitting drop set-ups. Crosses indicate that this crystallization screen has been tested. Drop size as indicated corresponds to the volume of RNA plus the volume of crystallization solution.

LCrz ARN	LCrz core RNA concentration in µM	LCrz substrate RNA concentration in µM	U1A protein concentration in µM	Incubation temperature in °C	Drop size in nL	Crystallization screen Index HT - Hampton Research	Crystallization screen PEGRx HT - Hampton Research	Crystallization screen Natrix HT - Hampton Research	Crystallization screen JBScreen Nuc-Pro HTS - JenaBioscience
DiLCrz_TLR1	180	240		20	400	x	x	x	x
DiLCrz_TLR1	133	200		20	400			x	
DiLCrz_TLR1	180	240		37	400			x	
DiLCrz_TLR1	180	240		20	200			x	
DiLCrz_TLR2	100	150		20	400	x	x	x	x
DiLCrz_TLR3	72	96		20	400	x	x	x	x
DiLCrz_TLR3	133	200		20	400			x	
DiLCrz_TLR3	72	96		37	400			x	
DiLCrz_TLR3	72	96		20	200			x	
DiLCrz_TLR3	72	96		20	2000		x	x	
DiLCrz_TLR3	50	75		20	400		x	x	
DiLCrz_TLR4	120	160		20	400	x	x	x	x
DiLCrz_TLR4	133	200		20	400			x	
DiLCrz_TLR4	120	160		37	400			x	
DiLCrz_TLR4	120	160		20	200			x	
DiLCrz_cp_U1A	100		100	20	400	x	x	x	x
DiLCrz_cp_U1A	100		100	20	1000		x		
DiLCrz_cp_U1A	50		50	20	1000		x		

Table 5: Summary table of the crystallization trials done with the Mosquito® robot in 96 wells sitting drop set-ups for LCrz constructs in complex with an RNA substrate or the U1A protein. Crosses indicate that this crystallization screen has been tested. Drop size as indicated corresponds to the volume of RNA plus the volume of crystallization solution.

To optimize crystals size and diffraction sparse matrix have been designed to span the phase diagram by varying the concentration of salt, precipitant and RNA but also the pH or the temperature **(Table 6 to 13)**. Sitting drops were made as for the initial screening. Initial screening with commercial kits plus screening using sparse matrix screens represent 15 760 sitting drops.

		1	2	3	4	5	6	7	8	9	10	11	12
A	PEG 400	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%
	$MgCl_2$ (mM)	60	60	60	60	60	60	60	0	15	30	90	120
	NaCl (mM)	0	30	60	90	190	260	130	130	130	130	130	130
B	PEG 400	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%
	$MgCl_2$ (mM)	60	60	60	60	60	60	60	0	15	30	90	120
	NaCl (mM)	0	30	60	90	190	260	130	130	130	130	130	130
C	PEG 400	15%	15%	15%	15%	15%	15%	15%	15%	15%	15%	15%	15%
	$MgCl_2$ (mM)	60	60	60	60	60	60	60	0	15	30	90	120
	NaCl (mM)	0	30	60	90	190	260	130	130	130	130	130	130
D	PEG 400	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%
	$MgCl_2$ (mM)	60	60	60	60	60	60	60	0	15	30	90	120
	NaCl (mM)	0	30	60	90	190	260	130	130	130	130	130	130
E	PEG 400	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%
	$MgCl_2$ (mM)	60	60	60	60	60	60	60	0	15	30	90	120
	NaCl (mM)	0	30	60	90	190	260	130	130	130	130	130	130
F	PEG 400	35%	35%	35%	35%	35%	35%	35%	35%	35%	35%	35%	35%
	$MgCl_2$ (mM)	60	60	60	60	60	60	60	0	15	30	90	120
	NaCl (mM)	0	30	60	90	190	260	130	130	130	130	130	130
G	PEG 400	45%	45%	45%	45%	45%	45%	45%	45%	45%	45%	45%	45%
	$MgCl_2$ (mM)	60	60	60	60	60	60	60	0	15	30	90	120
	NaCl (mM)	0	30	60	90	190	260	130	130	130	130	130	130
H	PEG 400	55%	55%	55%	55%	55%	55%	55%	55%	55%	55%	55%	55%
	$MgCl_2$ (mM)	60	60	60	60	60	60	60	0	15	30	90	120
	NaCl (mM)	0	30	60	90	190	260	130	130	130	130	130	130

Table 6: 96 crystallization conditions sparse matrix the condition A9 (highlighted in purple) from the JBScreen Nuc-Pro HTS (JenaBioscience) used as a starting point. Sodium citrate pH 5.6 is fixed at 100 mM. This screen has been used to crystallize DiLCrz_TLR3 at 20 °C, 30 °C and 37°C in 400 nL drops at concentration of:
- [DiLCrz_TLR3] = 40 µM, [RNA substrate strand] = 53 µM
- [DiLCrz_TLR3] = 72 µM, [RNA substrate strand] = 96 µM
- [DiLCrz_TLR3] = 140 µM, [RNA substrate strand] = 187 µM

		1	2	3	4	5	6	7	8	9	10	11	12
A	Sodium Malonate (mM)	50	100	150	50	100	150	50	100	150	50	100	150
	(w/v) PEG 400 (%)	10	10	10	20	20	20	30	30	30	40	40	40
B	Sodium Formate (mM)	50	100	150	50	100	150	50	100	150	50	100	150
	(w/v) PEG 1000 (%)	10	10	10	20	20	20	30	30	30	40	40	40
C	Sodium Formate (mM)	50	100	150	50	100	150	50	100	150	50	100	150
	(w/v) PEG 3350 (%)	10	10	10	20	20	20	30	30	30	40	40	40
D	Sodium Formate (mM)	50	100	150	50	100	150	50	100	150	50	100	150
	(w/v) PEG 8000 (%)	10	10	10	20	20	20	30	30	30	40	40	40

Table 7: 48 crystallization conditions sparse matrix with condition F3 (highlighted in purple) from the PEGRx HT (Hampton research ©) used as a starting point. Tris pH 8.0 is fixed at 100 mM. This screen has been used to crystallize DiLCrz_cp_U203C at 20 °C in 400 nL drops.

		1	2	3	4	5	6	7	8	9	10	11	12
A	NaCl (mM) (w/v) PEG 400 (%)	100 5	200 5	100 15	200 15	100 20	200 20	100 25	200 25	100 35	200 35	100 45	200 45
B	NaCl (mM) (w/v) PEG 1500 (%)	100 5	200 5	100 15	200 15	100 20	200 20	100 25	200 25	100 35	200 35	100 45	200 45
C	NaCl (mM) (w/v) PEG 3350 (%)	100 5	200 5	100 15	200 15	100 20	200 20	100 25	200 25	100 35	200 35	100 45	200 45
D	NaCl (mM) (w/v) PEG 8000 (%)	100 5	200 5	100 15	200 15	100 20	200 20	100 25	200 25	100 35	200 35	100 45	200 45

Table 8: 48 crystallization conditions sparse matrix with condition F11 (highlighted in purple) from the Index HT (Hampton research ©) used as a starting point. BIS-TRIS pH 6.5 is fixed at 100 mM. This screen has been used to crystallize DiLCrz_cp_U203C at 20 °C in 400 nL drops.

		1	2	3	4	5	6	7	8	9	10	11	12
A	NaCl (mM) (w/v) PEG 400 (%)	100 5	200 5	100 15	200 15	100 20	200 20	100 25	200 25	100 35	200 35	100 45	200 45
B	NaCl (mM) (w/v) PEG 1500 (%)	100 5	200 5	100 15	200 15	100 20	200 20	100 25	200 25	100 35	200 35	100 45	200 45
C	NaCl (mM) (w/v) PEG 3350 (%)	100 5	200 5	100 15	200 15	100 20	200 20	100 25	200 25	100 35	200 35	100 45	200 45
D	NaCl (mM) (w/v) PEG 8000 (%)	100 5	200 5	100 15	200 15	100 20	200 20	100 25	200 25	100 35	200 35	100 45	200 45

Table 9: 48 crystallization conditions sparse matrix with condition F12 (highlighted in purple) from the Index HT (Hampton research ©) used as a starting point. HEPES pH 7.5 is fixed at 100 mM. This screen has been used to crystallize DiLCrz_cp_U203C at 20 °C in 400 nL drops.

		1	2	3	4	5	6	7	8	9	10	11	12
A	Sodium Formate (mM) (w/v) PEG 400 (%)	100 10	200 10	300 10	100 20	200 20	300 20	100 30	200 30	300 30	100 40	200 40	300 40
B	Sodium Formate (mM) (w/v) PEG 1000 (%)	100 10	200 10	0.3 10	100 20	200 20	300 20	100 30	200 30	300 30	100 40	200 40	300 40
C	Sodium Formate (mM) (w/v) PEG 3350 (%)	100 10	200 10	0.3 10	100 20	200 20	300 20	100 30	200 30	300 30	100 40	200 40	300 40
D	Sodium Formate (mM) (w/v) PEG 8000 (%)	100 10	200 10	0.3 10	100 20	200 20	300 20	100 30	200 30	300 30	100 40	200 40	300 40

Table 10: 48 crystallization conditions sparse matrix with condition H6 (highlighted in purple) from the Index HT (Hampton research ©) used as a starting point. This screen has been used to crystallize DiLCrz_cp_U203C at 20 °C in 400 nL drops.

		1	2	3	4	5
A (F1)	100 mM BICINE	pH 7.0	pH 7.5	pH 8.0	pH 8.5	pH 9.0
B (F3)	100 mM Tris	pH 7.0	pH 7.5	pH 8.0	pH 8.5	pH 9.0
C (D11)	100 mM BIS-TRIS	pH 5.5	pH 6.0	pH 6.5	pH 7.0	pH 7.5
D (F11)	100 mM BIS-TRIS	pH 5.5	pH 6.0	pH 6.5	pH 7.0	pH 7.5
E (F12)	100 mM HEPES	pH 6.5	pH 7.0	pH 7.5	pH 8.0	pH 8.5
F	NaCl mM PEG 3350 w/v 100 mM MES	200 25% pH 5.5	200 25% pH 6.0	200 25% pH 6.5	200 25% pH 7.0	

Table 11: 29 crystallization conditions sparse matrix with conditions D11, F11, F12 (highlighted in purple) from the Index HT (Hampton research ©) and F1, F3 from the PEGRx HT (Hampton research ©) used as a starting point. pH of the buffer is the only parameter that has been modified in the sparse matrix. This screen has been used to crystallize DiLCrz_cp, DiLCrz_cp_U203C, DiLCrz_active and DiLCrz_active_U203C at 20 °C in 400 nL drops.

		1	2	3	4
A	100 mM Sodium citrate tribasic dihydrate PEGMME 550 (v/v)	pH 3 30%	pH 4 30%	pH 5 30%	pH 6 30%

Table 12: 4 crystallization conditions sparse matrix with condition A10 (highlighted in purple) from the PEGRx (Hampton research ©) used as a starting point. pH of the buffer is the only parameter that has been modified in the sparse matrix. This screen has been used to crystallize DiLCrz_cp_U203C at 20 °C in 400 nL , 600 nL and 800 nL drops.

		1	2	3	4
A	100 mM Sodium citrate tribasic dihydrate Jeffamine ® ED-2001 pH 7.0 (v/v)	pH 3 30%	pH 4 30%	pH 5 30%	pH 6 30%

Table 13: 4 crystallization conditions sparse matrix with condition C1 (highlighted in purple) from the PEGRx (Hampton research ©) used as a starting point. pH of the buffer is the only parameter that has been modified in the sparse matrix. This screen has been used to crystallize DiLCrz_cp_U203C at 20 °C in 400 nL , 600 nL and 800 nL drops.

c) Hanging drops

Again, to optimize crystal size and diffraction, larger drops in Linbro 24 wells plates (Hampton Research) have been prepared by mixing 1 µL of RNA to 1 µL of the reservoir solution **(Table 14)**. Reservoir volume is 150 µL.

RNA	Reservoir solution	[RNA] µM	T (°C)
DiLCrz_TLR3	75 mM Sodium chloride, 2 mM Calcium chloride dihydrate, 50 mM Sodium cacodylate trihydrate pH 6.0, 30% w/v 1,6 Hexanediol, 0.5 mM Spermine	[LCrz core] = 24 [RNA substrate strand] = 36 and 48	4, 20,37
DiLCrz_TLR3	75 mM Sodium chloride, 2 mM Calcium chloride dihydrate, 50 mM Sodium cacodylate trihydrate pH 6.0, 30% w/v 1,6 Hexanediol, 0.5 mM Spermine	[LCrz core] = 48 [RNA substrate strand] = 64 and 96	4, 20, 37
DiLCrz_TLR3	75 mM Sodium chloride, 2 mM Calcium chloride dihydrate, 50 mM Sodium cacodylate trihydrate pH 6.0, 30% w/v 1,6 Hexanediol, 0.5 mM Spermine	[LCrz core] = 72 [RNA substrate strand] = 96 and 144	4, 20,37
DiLCrz_TLR3	75 mM Sodium chloride, 2 mM Calcium chloride dihydrate, 50 mM Sodium cacodylate trihydrate pH 6.0, 30% w/v 1,6 Hexanediol, 0.5 mM Spermine	[LCrz core] = 96 [RNA substrate strand] = 128 and 192	4, 20,37
RNA	**Reservoir solution**	**[RNA] µM**	**T**

			(°C)
DiLCrz_TLR3	75 mM Sodium chloride, 2 mM Calcium chloride dihydrate, 50 mM Sodium cacodylate trihydrate pH 6.0, 30% w/v 1,6 Hexanediol, 0.5 mM Spermine	[LCrz core] = 120 [RNA substrate strand] = 160 and 240	4, 20,37
DiLCrz_TLR3	75 mM Sodium chloride, 2 mM Calcium chloride dihydrate, 50 mM Sodium cacodylate trihydrate pH 6.0, 30% w/v 1,6 Hexanediol, 0.5 mM Spermine	[LCrz core] = 144 [RNA substrate strand] = 192 and 288	4, 20,37
DiLCrz_cp_U203C	75 mM Sodium chloride, 2 mM Calcium chloride dihydrate, 50 mM Sodium cacodylate trihydrate pH 6.0, 30% w/v 1,6 Hexanediol, 0.5 mM Spermine	100	20
DiLCrz_cp_U203C	100 mM BIS-TRIS pH 6.5, 28% w/v PEGMME 2000	100	20
DiLCrz_cp_U203C	200 mM NaCl, 100 mM BIS-TRIS pH 6.5, 25% w/v PEG 3350	100	20
DiLCrz_cp_U203C	200 mM NaCl, 100 mM HEPES pH 7.5, 25% w/v PEG 3350	100	20
DiLCrz_cp_P2-1_U203C	100 mM BIS-TRIS pH 6.5, 28% w/v PEGMME 2000	100	20
DiLCrz_cp_P2-1_U203C	2% v/v PEG 400, 100 mM Imidazol pH 7.0, 24% w/v PEGMME5000	100	20
DiLCrz_active_U203C	100 mM Sodium Citrate tribasic dihydrate pH 5.0, 30% v/v PEGMME 550	100	20
DiLCrz_active_U203C	100 mM Sodium Citrate tribasic dihydrate pH 5.0, 30% v/v Jeffamine ® ED-2001 pH 7.0	100	20
DiLCrz_active_U203C	100 mM Sodium Malonate pH 8.0, 100 mM Tris pH 8.0, 30% w/v PEG 1000	100	20

Table 14: Summary table of the crystallization trials in 2 µL hanging drops with a reservoir of 150 µL.

d) Microbatches

In addition to hanging drops, microbatch experiments have been set up for most of the successful condition, since different setups can influence crystal quality (Chayen, 1998). Droplets were prepared in minitray Nunc™ plates (60 wells) by mixing 2.5 µl of the LCrz RNA at a concentration of 100 µM with 2.5 µl of crystallization solution at various concentrations. Drops were sealed using either 10 µl paraffin oil or crystal adhesive tape or 0.2% agarose gel. We also set up gel-microbatchs by mixing 2.5 µl of the LCrz RNA with 2.5 µl of crystallisation solution at

various concentrations and 0.2% low-melting agarose gel sealed with paraffin oil (Lorber et al, 2009). Some crystallization conditions have been supplemented with iridium hexamine or various polyamines. All microbatchs were kept at 20°C.

RNA	Crystallization condition	Microbatch type	Additive
DiLCrz_cp_U203C	10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% - D11 – Index HT- Hpt Research- 100 mM BIS-TRIS pH 6.5, 28% w/v PEGMME 2000	-Sealed with paraffin oil -Sealed with crystal tape -Sealed with agarose gel -0.2% low-melting agarose sealed with paraffin oil	
DiLCrz_cp_U203C	10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% - E11 – Natrix HT- Hpt Research- 75 mM Sodium chloride, 2 mM Calcium chloride dihydrate, 50 mM Sodium cacodylate trihydrate pH 6.0, 30% w/v 1,6 Hexanediol, 0.5 mM Spermine	-Sealed with paraffin oil -Sealed with crystal tape -Sealed with agarose gel -0.2% low-melting agarose sealed with paraffin oil	
DiLCrz_cp_U203C	10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% - F11 – Index HT- Hpt Research- 200 mM NaCl, 100 mM BIS-TRIS pH 6.5, 25% w/v PEG 3350	-Sealed with paraffin oil -Sealed with crystal tape -Sealed with agarose gel -0.2% low-melting agarose sealed with paraffin oil	
DiLCrz_cp_U203C	10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% - F12 – Index HT- Hpt Research- 200 mM NaCl, 100 mM HEPES pH 7.5, 25% w/v PEG 3350	-Sealed with paraffin oil -Sealed with crystal tape -Sealed with agarose gel -0.2% low-melting agarose sealed with paraffin oil	
DiLCrz_cp_U203C	50%, 60%, 70%, 80% - F1 – PEGRx HT- Hpt Research- 100 mM Succinic acid pH 7.0, 100 mM BICINE pH 8.5, 30% v/v PEGMME 550	-Sealed with paraffin oil -Sealed with crystal tape -0.2% low-melting agarose sealed with paraffin oil	
DiLCrz_cp_U203C	50%, 60%, 70%, 80% - F3 – PEGRx HT- Hpt Research- 100 mM Sodium malonate pH 8.0, 100 mM Tris pH 8.0, 30% w/v PEG 1000	-Sealed with paraffin oil -Sealed with crystal tape -0.2% low-melting agarose sealed with paraffin oil	
DiLCrz_cp_U203C	50%, 60%, 70%, 80% - H6 – Index HT- Hpt Research- 200 mM Sodium formate, 20% w/v PEG 3350	-Sealed with paraffin oil -Sealed with crystal tape -0.2% low-melting agarose sealed with paraffin oil	
DiLCrz_cp_U203C	50%, 60%, 70%, 80% - D11 – Index HT- Hpt Research- 100 mM BIS-TRIS pH 6.5, 28% w/v PEGMME 2000	-Sealed with paraffin oil -Sealed with crystal tape	1 mM Iridium hexamine
DiLCrz_cp_U203C	50%, 60%, 70%, 80% - E11 – Natrix HT- Hpt Research- 75 mM Sodium chloride, 2 mM Calcium chloride dihydrate, 50 mM Sodium cacodylate trihydrate pH 6.0, 30% w/v 1,6 Hexanediol, 0.5 mM Spermine	-Sealed with paraffin oil -Sealed with crystal tape	1 mM Iridium hexamine
DiLCrz_cp_U203C	50%, 60%, 70%, 80% - F11 – Index HT- Hpt Research- 200 mM NaCl, 100 mM BIS-TRIS pH 6.5, 25% w/v PEG 3350	-Sealed with paraffin oil -Sealed with crystal tape	1 mM Iridium hexamine

RNA	Crystallization condition	Microbatch type	Additive
DiLCrz_cp_U203C	50%, 60%, 70%, 80% - F12 – Index HT- Hpt Research- 200 mM NaCl, 100 mM HEPES pH 7.5, 25% w/v PEG 3350	-Sealed with paraffin oil -Sealed with crystal tape	-1 mM Iridium hexamine -12 mM Spermidine -12 mM Spermine -12 mM Cadaverine -12 mM Putrecine
DiLCrz_cp_U203C	50%, 60%, 70%, 80% - F1 – PEGRx HT- Hpt Research- 100 mM Succinic acid pH 7.0, 100 mM BICINE pH 8.5, 30% v/v PEGMME 550	-Sealed with paraffin oil -Sealed with crystal tape	1 mM Iridium hexamine
DiLCrz_cp_U203C	50%, 60%, 70%, 80% - F3 – PEGRx HT- Hpt Research- 100 mM Sodium malonate pH 8.0, 100 mM Tris pH 8.0, 30% w/v PEG 1000	-Sealed with paraffin oil -Sealed with crystal tape	1 mM Iridium hexamine
DiLCrz_cp_U203C	50%, 60%, 70%, 80% - H6 – Index HT- Hpt Research- 200 mM Sodium formate, 20% w/v PEG 3350	-Sealed with paraffin oil -Sealed with crystal tape	1 mM Iridium hexamine
DiLCrz_cp_P2-1_U203C	50%, 60%, 70%, 80% - F12 – Index HT- Hpt Research- 200 mM NaCl, 100 mM HEPES pH 7.5, 25% w/v PEG 3350	-Sealed with paraffin oil -Sealed with crystal tape	
DiLCrz_cp_P2-1_U203C	60%, 70%, 80% - D11 – Index HT- Hpt Research- 100 mM BIS-TRIS pH 6.5, 28% w/v PEGMME 2000	-Sealed with paraffin oil -Sealed with crystal tape	-1 mM Iridium hexamine -12 mM Spermidine -12 mM Spermine -12 mM Cadaverine -12 mM Putrecine
DiLCrz_cp_P2-1	40%, 50%, 60%, 70%, 80% - F12 – Index HT- Hpt Research- 2% v/v PEG 400, 100 mM Imidazole pH 7.0, 24% w/v PEGMME 5000	-Sealed with paraffin oil -Sealed with crystal tape	
AlLCrz_cp_court	50%, 60%, 70%, 80% - G7 – PEGRx HT- Hpt Research- 200 mM NaCl, 100 mM HEPES pH 7.5, 25% w/v PEG 3350	-Sealed with paraffin oil -Sealed with crystal tape	
DiLCrz_active	50%, 60%, 70%, 80% - F12 – Index HT- Hpt Research- 2% v/v PEG 400, 100 mM Imidazole pH 7.0, 24% w/v PEGMME 5000	-Sealed with paraffin oil -Sealed with crystal tape	
DiLCrz_active_U203C	50%, 60%, 70%, 80% - F12 – Index HT- Hpt Research- 2% v/v PEG 400, 100 mM Imidazole pH 7.0, 24% w/v PEGMME 5000	-Sealed with paraffin oil -Sealed with crystal tape	-1 mM Iridium hexamine
DiLCrz_active_U203C	50%, 60%, 70%, 80% - A10 – PEGRx HT- Hpt Research- 100 mM Sodium citrate tribasic dihydrate pH 5.0, 30% v/v PEGMME 550	-Sealed with paraffin oil -Sealed with crystal tape	-1 mM Iridium hexamine
DiLCrz_active_U203C	50%, 60%, 70%, 80% - C1 – PEGRx HT- Hpt Research- 100 mM Sodium citrate tribasic dihydrate pH 5.0, 30% v/v Jeffamine® ED-2001 pH 7.0	-Sealed with paraffin oil -Sealed with crystal tape	-1 mM Iridium hexamine

RNA	Crystallization condition	Microbatch type	Additive
DiLCrz_active_U203C	50%, 60%, 70%, 80% - F3 – PEGRx HT- Hpt Research- 100 mM Sodium malonate pH 8.0, 100 mM Tris pH 8.0, 30% w/v PEG 1000	-Sealed with paraffin oil -Sealed with crystal tape	-1 mM Iridium hexamine

Table 15: Summary table of the crystallization trials in 5 µL microbatch drops. The crystallization conditions are given at concentration corresponding to 100%. Percentages indicated are final concentration of a given crystallization solution in a microbatch drop.

At a final concentration of 60% or 70% of the crystallization condition, crystals were larger than the ones grown in sitting or hanging drop setups and allow us to collect high resolution diffraction data. One general feature regardless of the crystallization condition of the LCrz RNA is that additives, iridium hexamine or polyamines, allow crystals to grow faster and larger, although their diffraction properties were significantly decreased. Taken together optimization in hanging drops and microbatch drops represent 777 handmade drops. All crystallization conditions that allow RNA crystals to grow, in sitting drops, hanging drops and microbatch drops are summarized in **Appendix IV**. For each successful crystallization condition the following information are given: the LCrz RNA construct, the setup, the crystallization condition and the size of the drop, the time lapse between the day the drops has been made and the picture of the crystals has been taken. When available additional information is given such as the way crystals have been frozen and the best resolution observed.

e) Post-crystallization treatments

Cryoprotective and dehydration treatments have been experienced, as they have been shown to reduce ice-rings and improve diffraction (Abergel, 2004; Alcorn & Juers, 2010). All treatments have been tested on DiLCrz_cp_U203C RNA crystals grown in microbatchs (F12 – Index HT- Hampton Research- 2% v/v PEG 400, 100 mM Imidazole pH 7.0, 24% w/v PEGMME 5000) sealed with crystal adhesive tape. Crystals have been soaked as follows:

- 5 or 10 min in 5 µl of crystallization solution and 5 or 10 % (v/v) PEG 400
- 5 or 10 min in 5 µl of crystallization solution and 5 or 10 % (v/v) PEG 600
- 5 or 10 min in 5 µl of crystallization solution and 5 or 10 % (v/v) glycerol

- 5 or 10 min in 5 μl of crystallization solution and 5 or 10 % (v/v) MPD
- 5 or 10 min in 5 μl of crystallization solution and 5 or 10 % (v/v) ethylene glycol

In order to solve the structure of the DiLCrz_cp using heavy atoms derivatives, crystals have been soaked as follows (Heras & Martin, 2005):

1. 5 min in 5 μl of crystallization solution and 0 or 5 or 10 % (v/v) PEG 400 (optional)
2. 5 min in 5 μl of crystallization solution and 0 or 5 or 10 % (v/v) PEG 400 and 1 mM osmium hexamine or iridium hexamine or cobalt hexamine
3. Few seconds in 5 μl of crystallization solution and 0 or 5 or 10 % (v/v) PEG 400 and 2 mM $MgCl_2$

Crystals were stored in CrystalCap™ HT filled with ethane and frozen in liquid nitrogen.

6. Data collection and processing

DiLCrz_cp_U203C yielded data quality crystals in condition 0.2 Sodium chloride, 0.1 M HEPES pH 7.5, 25% w/v PEG 3,350. Diffraction data were collected at SLS on beam-line X06DA (PXIII) equipped with a PILATUS 2M detector. Oscillation frames (0.1°/frame) were collected on crystals undertaking a 20% attenuated beam flux to decrease radiation damage. Crystals were maintained at 100K under a cryo-flux. To solve the structure of the DiLCrz ribozyme *de novo* by MAD, 4 data sets were collected **(Figure 46)** on the same crystal (see Results section). Native datasets were also collected up to 2.45 Å resolution. Data were processed using XDS (Kabsch, 2010) and heavy atoms sites were determined using SHELX package (Sheldrick, 2008).

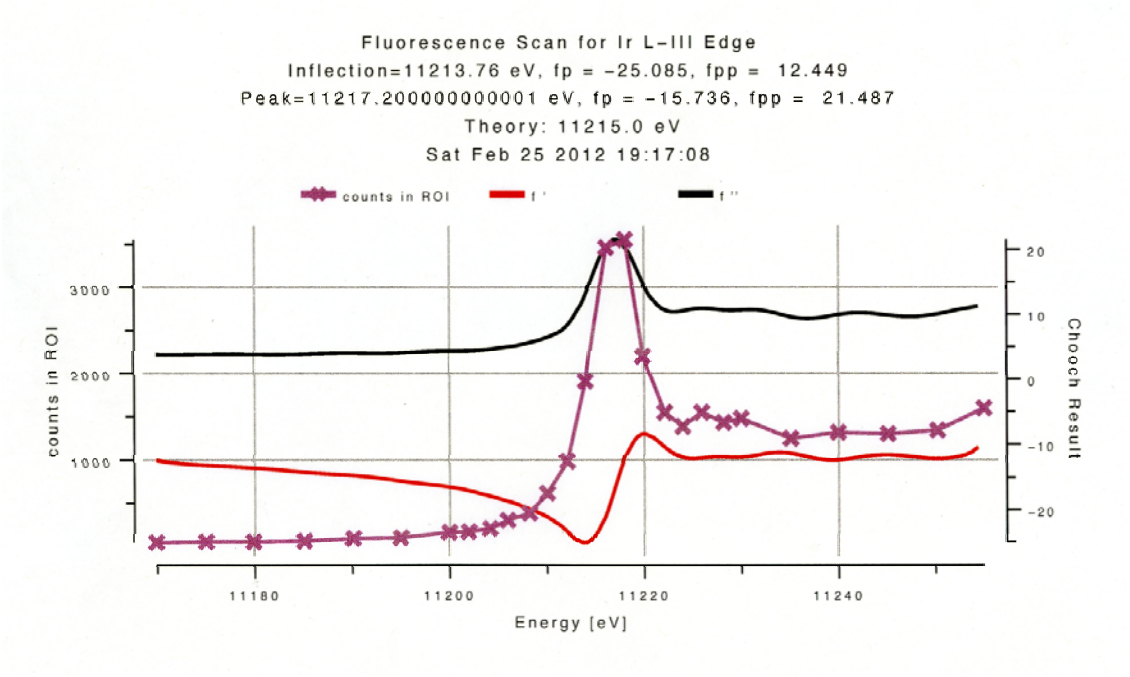


Figure 46: Fluorescence scan performed on iridium hexamine soaked crystals to determine wavelengths at which data should be collected to solve the structure of DiLCrz_cp_U203C *de novo* by MAD.

DiLCrz_active_U203C yielded one data quality crystal in condition 100 mM Sodium citrate tribasic dihydrate pH 5.0, 30% v/v Jeffamine® ED-2001 pH 7.0. Diffraction data were collected at SOLEIL on beam-line Proxima I equipped with a PILATUS 4M detector. Oscillation frames (0.2°/frame) were collected on crystal undertaking an attenuated beam flux to decrease radiation damage. Crystals were maintained at 100K under a cryo-flux. To solve the structure of DiLCrz_active_U203C we performed a molecular replacement with MolRep (Vagin & Teplyakov, 2010) using the model of the DiLCrz_cp_U203C construct.

7. Structure determination and model refinement

The resulting experimental electron density map of DiLCrz_cp_U203C allowed to build an initial model using Coot (Emsley & Cowtan, 2004) which was later refined with Phenix (Adams et al, 2010) against higher resolution data obtained from crystal 2 to final R_{work}/R_{free} values of 18.80%/23.57% (see Results section). Density around J6/7 region as well as the first nucleotides of P7 were poorly defined but were perfectly visible when the structure was refined against the data from crystal 1 (final R_{work}/R_{free} values of 19.81%/26.21%).

Data up to 4 Å were collected in spring 2013 at SOLEIL for DiLCrz_active_U203C. Initial electron density maps were obtained by molecular replacement using MolRep (Kruger et al, 1982) .However the model proved to be difficult to refine because at this resolution the molecular replacement solution largely dominates the data. We now turn towards heavy atom derivatives to solve this problem.

Molecular visualizations and structures illustrations were performed using PyMOL (Schrödinger).

B. SMALL-ANGLE X-RAY SCATTERING EXPERIMENTS

LCrz RNA have been produced and purified as for crystallization. All LCrz RNA have been prepared in 25 µL at the concentration of 6 mg/mL in a solution containing 100 mM KCL and 10 mM MES-KOH pH 6.5. The sample has been heated during 1 min at 95 °C followed by 4 min on ice. $MgCl_2$ up to 4 mM was then added, the solution was placed at 55 °C for 10 min then at 4 °C overnight. The homogenous folding of the ribozyme has been verified by PAGE under native conditions.

All details for SAXS data collection and analysis are given in the supplementary material of the article describing the DiLCrz_cp_U203C crystal structure (see Results section).

C. BIOINFORMATIC ANALYSIS OF THE BACTERIAL TYPE A RNASE P

All details for bioinformatic analysis of the motif kinking RNA helices are given in the material and method section of the article and in the supplementary material (see Results section).

RESULTS &

DISCUSSION

III. RESULTS AND DISCUSSION

A. CRYSTAL STRUCTURE OF THE LARIAT-CAPPING RIBOZYME

1. Introduction

Out of 2208 families of ncRNA identified in Rfam (Gardner et al, 2011) only 60 families have at least one of its member for which the 3D structure has been solved. And a glance at the PBD, www.pdb.org, (Berman et al, 2000) reveals that even if the number of RNA 3D structures solved increases at a rate of approximately 75 new structures per year for the last 10 years it is far less than for proteins **(Figure 47)**. X-ray crystallography is the method of choice to solve 3D structure as for example, 2040 RNA structures, including ribosomes', listed in the NDB are solved by X-ray crystallography in comparison to only 554 RNA structures of rather small size solved by NMR (Berman et al, 1992).

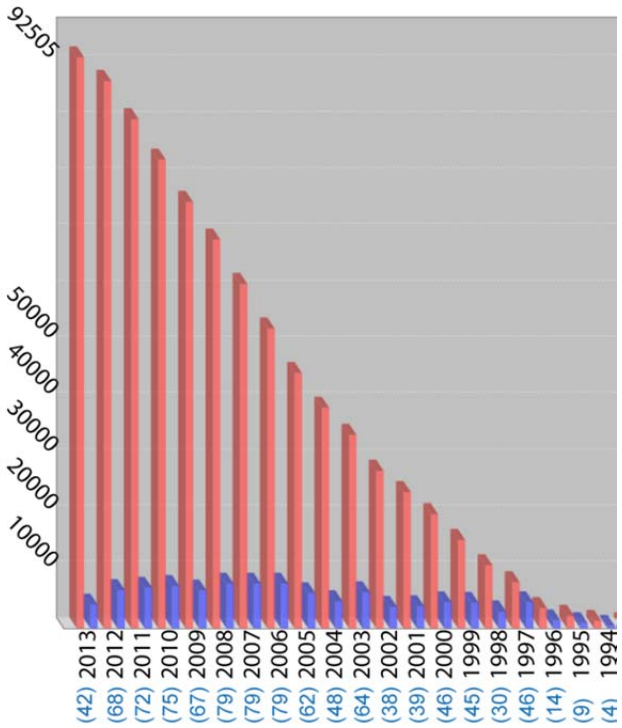


Figure 47: Histogram showing the yearly growth of structures solved by X-ray (Data taken from PDB website). RNA structures are shown in blue with the exact number in brackets. In red is the total number of X-ray structures registered in the PBD up to the 29th of July 2013.

In the below article, we report the crystal structure of the 192 nucleotides DiLCrz at a resolution of 2.5 Å. Size and resolution are comparable with the three known group I introns 3D structures (Cate et al, 1996a; Golden et al, 2005; Lipchock & Strobel, 2008). With this new crystal structure we supplement the collection of 3D structures of the ribozyme family. Moreover, previous work on modeling of the DiLCrz ribozyme (Beckert, 2008) and comparison between DiLCrz and group I intron crystal structure allow us to hypothesize that LCrz ribozyme could result from the evolutionary metamorphosis of a group I intron in the context of a twintron (Decatur et al, 1995).

2. Article

Metamorphosis of a group I intron into a lariat capping ribozyme

Mélanie Meyer[1,2], Henrik Nielsen[3], Vincent Oliéric[4], Pierre Roblin[5,6], Steinar D. Johansen[7], Eric Westhof[1], Benoît Masquida[1,2]*

[1] Architecture et Réactivité de l'ARN, IBMC, Université de Strasbourg, CNRS, Strasbourg, France

[2] *Present address :* Génétique Moléculaire, Génomique, Microbiologie, IPCB, Université de Strasbourg, CNRS, Strasbourg, France

[3] Department of Cellular and Molecular Medicine, The Panum Institute, University of Copenhagen, Copenhagen, Denmark

[4] Swiss Light Source at Paul Scherrer Institute, Villigen, Switzerland

[5] Synchrotron Soleil, Saint-Aubin, France

[6] INRA, Unité Biopolymères, Interactions, Assemblages, Nantes, France

[7] Department of Medical Biology, University of Tromsø, Tromsø, Norway

* Correspondence to b.masquida@unistra.fr

Keywords : RNA structure, LC ribozyme, GIR1, branching reaction, crystallography, SAXS

Summary

The lariat capping (LC) ribozyme is a natural ribozyme isolated from eukaryotic microorganisms. Despite significant structural similarity to group I introns, the LC ribozyme catalyzes cleavage by branching leaving the 3' product with a 3 nt lariat cap that functionally substitutes for a conventional mRNA cap in the downstream pre-mRNA encoding an homing endonuclease (HE). We describe the 2.5 Å LC ribozyme crystal structure, which suggests that structural features inherited from group I ribozymes have undergone metamorphosis due to profound changes in molecular selection pressure on a rather short time-scale to give rise to a new ribozyme family that catalyzes the 2',5' branching reaction. The structure elucidates the role of key elements that regulate the activity of the LC ribozyme by conformational switching, and suggests a mechanism by which the signal for branching is transmitted to the catalytic core. Based on this, a model is derived that coordinates ribosome synthesis and the intron-encoded homing endonuclease expression during a starvation-re-feeding cycle of the host organism.

One sentence summary

The lariat capping (LC) ribozyme crystal structure reveals a new class of ribozyme, which derives from group I introns.

Introduction

Various mechanisms critically regulate splicing, thus controlling the fate of host gene products. Most splicing that relies solely on auto-catalytic RNA introns appears to be unregulated, but control can be achieved at the RNA level since the modular organization of RNA allows for the juxtaposition of functional units working in concert (Masquida et al., 2010). Examples of riboswitches adjoined to spliceosomal (Cheah et al., 2007) or group I (Lee et al., 2010) introns well illustrate this concept and suggest that additional original splicing regulation mechanisms may exist.

A particularly elaborate example of RNA regulated splicing is the group I twin-ribozyme introns found in the small subunit (SSU) part of the ribosomal precursor in several protists (**Fig. 1A**). These are composed of a conventional group I splicing ribozyme (GIR2) into which is inserted a cassette composed of a branching ribozyme and a homing endonuclease (HE) gene (Nielsen et al., 2008). The branching activity results in cleavage and concomitant formation of a tiny (3 nt) lariat capping the 5' end of the cleavage product (**Fig. 1B**) corresponding to the HE pre-mRNA (Nielsen et al., 2005), hence the name lariat capping (LC) ribozyme. The lariat cap appears to substitute for a conventional mRNA m^7G cap in a situation where a mRNA is expressed from within a RNA polymerase I gene (Johansen et al., 2007). The cross-talk between the group I ribozyme (GIR2) and the LC ribozyme is demonstrated by the existence of three distinct processing pathways of the rRNA precursor, depending on environmental conditions. Under favorable conditions, splicing takes place first, followed by branching and further processing and eventually translation of the HE mRNA (Vader et al., 2002; Vader et al., 1999). The transcriptional order of the ribozymes implies that the activity of the branching ribozyme is specifically suppressed until splicing has taken place. Cellular stress conditions induce formation of full-length intron circles by a circularization

pathway (Nielsen et al., 2003) that leave the ribosomal exons unligated. Finally starvation conditions induce branching by the LC ribozyme without splicing taking place (Vader et al., 2002). Since splicing only enables ligation of ribosomal exons, the interplay between GIR2 and LC ribozymes influences the fate of both ribosome and HE mRNA production. Thus, the key question is how such critical regulation is achieved and in particular what structural features of the underlying RNA mediate this.

Originally, the LC ribozyme was described as a group I – like ribozyme (GIR1) based on striking similarities at both sequence and secondary structure levels (Johansen et al., 2002; Johansen and Vogt, 1994)(**Fig. 1C-D**). At first sight, the LC ribozyme adopts a overall architecture similar to that of group I ribozymes with three stacks of helices condensing to form a compact core (Beckert et al., 2008). One major difference is the presence of a group I intron–unrelated domain named DP2-DP2.1 in the LC ribozyme found in *Didymium iridis* tethered to P10 and involved in conformational switching (Birgisdottir et al., 2011; Nielsen et al., 2009). The other main difference is the P15 pseudoknot which is the result of recruitment of the group I – intron characteristic J8/7 junction as a pairing partner for P2'. This, together with the P3-P7 pseudoknot, creates a double-pseudoknotted structure with a strand topology that is distinctly different from a group I ribozyme. Based on this, we have previously hypothesized a model for the emergence of the LC ribozyme from within a group I scaffold (Beckert et al., 2008). The evolutionary context for this was suggested to be the insertion of a splicing group I intron into another group I intron harboring a HE gene (Haugen et al., 2005). In this scenario, selection pressure for the splicing activity of the invading intron would be lost and replaced by selection pressure for the promotion of HE gene expression. The model predicts drastic changes in the architecture of the ribozyme because it replaces a two-

step reaction depending on an exogenous co-factor (the exo-G) with a one-step reaction regulated by conformational switching.

Here, we report the 2.5 Å crystal structure of a 192 nt LC ribozyme from *Didymium iridis* (DiLC). The formation of the double pseudoknot leads to profound rearrangement of the ribozyme core, resulting in the absence of almost all structural hallmarks of group I splicing ribozymes. The structure shows how the regulatory domain twists the core of the LC ribozyme by mediating two sets of tertiary interactions, which organize the 3WJ specific of the LC ribozyme so that it acts as a receptor for a critical residue from the core, itself intimately connected to residues forming the lariat fold. The present fold departs so drastically from group I intron structures that it justifies to define a new ribozyme family for LC ribozymes. Moreover, the structure illustrates how domains inherited from an ancestral group I intron have evolved to adopt new shapes and functions that cooperated to develop and maintain a branching reaction. The biological consequence of this is the adaptation of expression of an mRNA from a polI context in a regulated fashion that is coordinated with the cells production of ribosomes.

Results

The overall structure of the LC ribozyme is stretched by the regulatory domain

The 2.5 Å crystal structure of a 192 nt LC ribozyme from *Didymium iridis* (DiLC) crystallizes in space group P212121 and presents one molecule per asymmetric unit (Table S1). To facilitate crystallization, the ribozyme was circularly permutated such that the scissile bond was broken and the segment connecting DiLC to the cognate splicing ribozyme closed by a UUCG loop (**Fig. S1**). The construct is thus locked into the active fold (Nielsen et al., 2009), although the branching reaction cannot take place since the scissile bond is already open. Similarly to group I introns, DiLC is organized around

stacks of helices that form three distinct domains (**Fig. 1C and 1D**). DiLC presents the group I typical P3-P7 pseudoknot that, together with an additional pseudoknot (P15-P3) and the associated three-way junction (3WJ, P15-P3-P8), leads to a highly constrained, double pseudoknotted (P15-P3-P7) core (Einvik et al., 1998). In group I introns, the core is invariably clamped by a direct interaction of the tips of P2 and P5 with P8 and P9, respectively (Stahley and Strobel, 2006). In DiLC, the P2/P8 interaction is replaced by a 3WJ encompassing P15-P3-P8, in which Watson-Crick UoU pairs mediate stacking continuity. In this context, P15 is equivalent to P2 and P1 stems in group I introns. Distal from P15, a regulatory domain tethered to P10 - built by the specific elements DP2 and DP2.1 involved in switching on the activity of DiLC during processing of the twin-ribozyme intron (Nielsen et al., 2009) - bridges P5 and P9 (**Fig. 1C**). DP2 and DP2.1 interact respectively with P9 and P5. The curvature of this domain provokes stretching of the P4-P6 domain, and results in an overall signet ring shape in which DP2.1 and the P4-P6 domain become perpendicular to the core (**Fig. 2A-C, Fig. S2A-D**). Instead, in group I ribozymes, the P4-P6 domain adopts a smaller angle with respect to the core and forms a tertiary interaction with P9, due to the systematic presence of a structural module that bends P9 towards P5 (**Fig. 1D**).

Two specific tertiary interactions are responsible for this altered shape (**Fig. 3A-B)**. The L9 GAAA tetraloop interacts in the shallow groove of the DP2 stem where two Watson-Crick purine-purine (R-R) pairs are observed (**Fig. 3A**). The tertiary interaction propagates through a ribose zipper involving G221 and G70 (paired to G245). The interaction is supported by mutation of the R-R pairs and lengthening or shortening of P9 (**Fig. S3**) as well as by analysis of an analogous interaction between L9 and the regulatory domain in the *Naegleria* LC ribozyme (Tang et al., 2011). The second tertiary interaction involves the DL2.1 and the L5 loops, which form three *cis* WC base pairs

between residues 5'-A90UC-3' from DL2.1 and 5'-G144AU-3' from L5. This kissing complex (**Fig. 3B**) is potentially conserved among the known instances of LC ribozymes at the sequence level.

The internal loop splitting P4 from P5 folds to let L5 kiss L2.1

In group I introns, the J5/4 junction makes tertiary interactions with the splice site GoU pair in P1 (**Fig. 1D**). In DiLC, this interaction is absent and instead J5/4 adopts a zigzag motif remote from the catalytic site and serves bending the P4-P6 domains to place L5 favorably to form a kissing complex with the loop of DP2.1 (**Fig. 3B, Fig. S4A**). As a consequence of the P4-P6 reorganization, the characteristic set of base triple interactions taking place in group I introns between P4 and the core are absent. Instead, the domain is anchored only at the tip of P6. The J5/4 zigzag forms two loop-like conformations, each one capping the flanking helices and allowing them to unstack. The facing strand (J4/5) adopts a typical A-form conformation. U136 locks the zigzag by making a Watson-Crick/Watson-Crick *trans* base pair with A155 that is stacked on the unpaired G137 (**Fig. S4B**). Mutational analysis shows that A153 from the zig-zag motif is critical for catalysis despite its distant location from the active site (Beckert et al., 2008). A153 resides at the hinge of the motif where it participates in a tertiary contact through an A-minor interaction (Doherty et al., 2001; Nissen et al., 2001) to the apical G=C pair of P4, with A152 providing the second A-minor contact with the second base pair of P4 (**Fig. S4C**). Two Mg^{2+} ions occupy the deep groove of P4, 9.5 Å apart from each other (**Fig. S4A**).

The original architecture of the LC ribozyme is supported by small angle X-ray scattering on the active and inactive ribozymes

The unexpected position of the P4-P6 domain resulting from the DL2.1/L5 kissing complex and from the conformation of the internal loop between P4 and P5 prompted

us to further investigate the conformation of the LC ribozyme in solution to address a potential crystallization artifact. Small angle X-ray scattering (SAXS) experiments were thus carried out both on the circularly permutated (CP) LC ribozyme solved in the present study and on the active wild-type 166.22 construct (**Fig. 2**) (Nielsen et al., 2005). Both forms show the same kind of SAXS shapes (**Fig. S2I**). The active ribozyme is slightly more compact than the CP form with a *Rg Guinier* of 29.8 Å and *Dmax* of 100 Å versus 30.9 Å and 105 Å. Strikingly, the fit between the crystal structure of the CP form and the SAXS curves was better in the case of the active form ($\mathbf{X}^2$=1.2) than in the case of the CP form ($\mathbf{X}^2$=2.5). This observation indicates that the crystal environment promotes the CP form to adopt a conformation closer to the active form than in solution.

The DP2-DP2.1 bridge participates to the folding of the ribozyme core

The tertiary interactions mediated between the regulatory domain and the core elements P5 and P9 decrease the flexibility of the 3WJ involving P10. As being connected to the regulatory domain, to the core and also directly to the 3 residues forming the lariat fold, P10 is a central element of the LC ribozyme architecture. Interestingly, the P10 3WJ is mostly stabilized by stacking mediated by five unpaired adenosines located at the interface of the stems (residues 76, 77, 103, 104, 238) (**Fig. 3C**). One side of A76 and A103 stack on the last base pair of DP2 and DP2.1, respectively. A77 stacks with A104, which stabilizes the sharp turn between DP2.1 and P10, and A238 stacks upon the unpaired U239 at the inlet of the 3' strand from DP2. This architecture forms a snug fit for residue A209, which belongs to the junction between P15 and P7, and is responsible for the most specific interactions within the junction (**Fig. 3C, Fig. S2F**). A209 is sandwiched between the GoU pair of P10 and A238. The Hoogsteen edge of A209 contacts the Watson-Crick edge of U237. Moreover, the A209 *syn* orientation permits an additional H-bond to A104(O2'). Thus, the 3WJ acts as a receptor recognizing A209.

The interaction from A209 with the P10 3WJ also results from an unusual conformation of the GoU pair of P15. As a mimic of P1/P2, P15 harbors a terminal GoU pair supposed to correspond to the Watson-Crick GoU pair involved in the second step of splicing by group I ribozymes (**Fig. S2E-H**). Nonetheless, in DiLC, this GoU pair adopts a *trans* instead of *cis* geometry. Consequently, U207 presents its O2' group in the deep groove of P15 where it interacts with the N7 group of A206, and is thus located opposite to the position of the scissile bond, i.e. from the lariat fold. This conformation creates a S-turn (Correll et al., 1998) that projects A209 into the P10 3WJ. The *trans* conformation of the GoU pair in P15 is thus necessary for the interaction of A209 within the P10 3WJ. This interaction indicates to the core the conformational state of the regulatory domain. A210 follows A209 and is directed towards the lariat fold. In the vicinity of the latter, an iridium hexammine ion that could indicate a metal ion site involved in catalysis is observed (**Fig. S2F**). The crystal structure conformation of J15/7 sheds light on the strategy undertaken by DiLC to create an interaction between the lariat fold and the active conformation of the regulatory domain.

Stabilization of the three-way junction also settles the position of residue U232, because it is directly connected to P10. U232, which bears the O2’ nucleophilic group in LC ribozyme, builds up the lariat fold together with C230 and A231. The positions of these residues are stabilized by specific interactions, bringing the nucleophile U232(O2’) at 5.1 Å from the C230(O5’) that bears the scissile phosphate involved in branching in the natural form of the ribozyme (**Fig. 3D**). This distance leaves the exact space needed to accommodate the phosphate group missing in the crystal structure due to the construct design (**Fig. S1**). The high level of internal constraints necessary to pre-form a 3-nt lariat fold generates a structure in which the sugar moieties are very close to each other while

the base moieties pop out to interact with partners in fairly distant regions of the DiLC ribozyme core.

In group I introns, the scissile phosphate is placed in proximity of the nucleophile at the second step of splicing by binding of the ωG as a triple in the G-binding site in P7 (**Fig. S2E**). Mutational analysis has shown the homologous LC G229 to be a critical nucleotide (Einvik et al., 1998), and expected to be bound in the G-binding site prior to branching. Yet, in the crystal structure, G229 is not bound to P7 and is at a considerable distance (10Å) from the nucleophile. G216 does not form either the first base triple with C171 or A172 that is characteristic of P7 in splicing (**Fig. 3D, Fig. S2E-F**). Together, these observations indicate that the crystal structure reflects the post-catalytic state in LC ribozyme with a relaxed P7 as is also observed in group I ribozymes in the post-catalytic state (Lipchock and Strobel, 2008).

Discussion

We have solved the 2.5 Å crystal structure of the lariat capping ribozyme from *Didymium iridis* in a version comprising both the core and the appended regulatory domain. The solved circular permutated form of the ribozyme corresponds to a post-cleavage state since the regulatory domain DP2-DP2.1 is locked into its active conformation (Nielsen et al., 2009) and the scissile bond is disconnected. Our structure reveals how two tertiary interactions mediated by the active conformation of the regulatory domain are critical for assembly of the active site of the ribozyme through folding of a three-way junction, which serves as a receptor for an internal residue triggering folding nucleation of the catalytic core. SAXS analysis of the active ribozyme suggests that its conformation deviates only locally from the present structure. Thus, the observed better fit between

the crystal structure and the SAXS shape corresponding to the active ribozyme indicates that the crystal environment restricts relaxation of the circularly permutated construct.

The LC ribozyme can also be viewed as a riboswitch, since its activity is turned on by the flanking splicing ribozyme GIR2 in a manner that is responsive to different cellular conditions (Decatur et al., 2000; Nielsen et al., 2003; Vader et al., 2002). The relaying of the switch from GIR2 is currently unknown but likely involves the formation of the DP2.1/L5 kissing loop that is highly conserved. Growing *Didymium* cells produce large amounts of HE mRNA, which end up on polysomes, although no homing function is attached to the protein at this stage (Vader et al., 2002). This observation makes from the HE a putative candidate to operate the switch since some HE families have been shown to contain members acting as splicing factors (Ho et al., 1997; Ho and Waring, 1999; Lambowitz et al., 1999). Ligation of the SSU ribosomal RNA exons used in translation occur during the splicing pathway and is associated to translation of the HE mRNA (**Fig. 4**). We assume that the increasing HE pool may favour the inactive conformation of the LC ribozyme by forming an intron-HE complex, thus preventing pre-splicing cleavage by branching. Splicing would then trigger branching, processing and further translation of the HE mRNA. In the opposite situation encountered during starvation, down-regulation of translation would reduce the HE pool below the threshold necessary to prevent formation of the LC ribozyme active form. Consequently, starvation would promote LC ribozyme cleavage of the unspliced pre-rRNA by branching, which is indeed observed *in vivo,* further splicing inhibition and translation arrest. Getting off starvation could partly rely on processing of the accumulated HE-encoding lariat-capped pre-mRNAs, which could be readily translated to enhance splicing and further translation (Vader et al., 2002). This pathway constitutes an

example where the intron intervenes on the host's biology, although introns are usually thought as neutral, selfish elements (Burt and Koufopanou, 2004).

The present structure illustrates how co-evolution between the LC core and the regulatory domain apparently permitted the accommodation of the lariat-forming residues within the ribozyme by means of a specific set of tertiary interactions with residues from the core not directly involved in group I splicing. Beyond architectural features, our structure reveals the surprising metamorphosis of elements known from two-step splicing group I intron ribozymes into a structure that supports a one-step branching reaction. Group I derived ribozymes can perform hydrolytic cleavage and splicing both in *cis* and in *trans* (Inoue et al., 1985), act as a ligase (Vicens and Cech, 2009) and work as an allosteric enzyme regulated by a second messenger (Lee et al., 2010). It can assemble itself from pieces (Draper et al., 2008) and can even form cooperative networks that support self-replication (Vaidya et al., 2012). Thus, the present structure of the LC ribozyme adds to the versatility of group I ribozyme scaffolds and completes the collection of crystal structures of large ribozyme classes that comprise group I (Adams et al., 2004; Golden et al., 2005; Guo et al., 2004) and group II (Toor et al., 2008) splicing ribozymes, and RNase P (Reiter et al., 2010).

The astonishing sequence and secondary structure similarities between the *Naegleria* LC and the *Azoarcus* group I ribozymes underline the fairly recent appearance of the LC ribozyme family, which could have been driven by (i) the insertion of a group I ribozyme within another (Hafez et al., 2013), (ii) sequence drift due to the absence of selection pressure towards splicing (Beckert et al., 2008) and (iii) the fact that intra-chain 2' OH groups can be activated by elements from the environment (Zaher et al., 2011) while 3' OH groups are activatable only at 3' terminal residues. The readiness with which 2',5'

bond formation emerges in RNA scaffolds, also seen with *in vitro* selected ribozymes (Carlomagno et al., 2013; Lee et al., 2011), could indicate that splicing may have developed more efficiently from group II introns or from the spliceosome than from group I introns due to a more flexible lariat-based mechanism with catalytic RNAs (Anokhina et al., 2013).

Experimental Procedures

Crystallization, X-ray data collection, structure determination and refinement

Parallelepipedal crystals grew in 6 weeks from a 100 µM solution of purified LC RNA transcripts in microbatch set-up containing 0.28 M Sodium chloride, 0.14 M HEPES pH 7.5, 35% w/v PEG 3,350.

Diffraction data were collected at 100 K at the macromolecular crystallography beamline X06DA at the Swiss Light Source (SLS) and processed using XDS (Kabsch, 2010). Phase information was obtained with crystal 1 (Table S1) by a 4-wavelength MAD experiment around the Iridium edge using the SHELX package (Sheldrick, 2008).

The initial model was refined with Phenix (Adams et al., 2010) and iteratively built using Coot (Emsley and Cowtan, 2004). The model obtained with data from crystal 2 was refined to final R_{work}/R_{free} values of 18.80%/23.57% (See Table S1). Density around J6/7 region was perfectly visible when the structure was refined against the data from crystal 1 (final R_{work}/R_{free} values of 19.81%/26.21%). The two structures are similar with an overall root mean square deviation of all atoms (*rmsd*) calculated with lsqman (Kleywegt and Jones, 1994) of 0.67 (*nrmsd* = 0.25).

SAXS data collection and analysis

SAXS measurements were undertaken to compare the crystal structure to the solution structure of LC ribozyme. SAXS data from pure LC ribozyme samples were collected directly at the exit of a size exclusion chromatography HPLC system (David and Perez, 2009). The Guinier curve was obtained from averaging the frames of equal R_g within the main peak of elution. The corresponding gyration radii (R_g) drawn from the Guinier domain (Guinier, 1939) are close to 29 Å and the autocorrelation functions p(r), determined from the total set of scattering points, give the same value of R_g and a maximal intramolecular distance D_{max} equal to 115 Å for the circularly permutated LC ribozyme.

The theoretical scattering pattern was calculated from the crystal structure, using CRYSOL (Svergun et al., 1995), and compared with the experimental data. This pattern fits well the experimental data and displays a systematic deviation value of 1.2 or 2.5 for the catalytic or the circularly permutated LC ribozymes, respectively. Consistent with these findings, the crystallographic model does match the low-resolution shape model constructed *ab initio* with DAMMIF (Franke and Svergun, 2009; Inoue et al., 1985).

Accession Numbers

Coordinates and structure factors have been deposited with the PDB ID 4HHK and 4HHN.

Acknowledgments

We thank scientists at ESRF, Max Nanao; at SOLEIL, Andrew Thompson, Pierre Legrand as well as Jean Cavarelli for support in the synchrotron Blocked Allocation Group (BAG) from Strasbourg. We wish to deeply thank Marat Yusupov for the gift of osmium

hexammine, Stéphane Bellemin-Laponaz for the gift of iridium hexammine. The crystallographic experiments were performed on the X06DA beamline at the Swiss Light Source, Paul Scherrer Institut, Villigen, Switzerland. SAXS experiments were performed on the SWING beamline at synchrotron SOLEIL. We thank the Centre National de la Recherche Scientifique for support and the University of Strasbourg for a Ph. D. grant to MM. This work was partly supported by ANR-10-BLAN-1502-02 "GRP2CONF" (to EW) and by the Danish Council for Independent Research in Natural Sciences (to HN).

References

The DP2 and DP2.1 elements constitute a regulatory domain involved in switching on the activity of LCrz during processing of the twin-ribozyme intron. The switching mechanism and the architecture of the domain differ among the known instances of LCrz.

Materials and methods are available as supplementary material on Science Online.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Adams, P.L., Stahley, M.R., Gill, M.L., Kosek, A.B., Wang, J., and Strobel, S.A. (2004). Crystal structure of a group I intron splicing intermediate. RNA *10*, 1867-1887.

Anokhina, M., Bessonov, S., Miao, Z., Westhof, E., Hartmuth, K., and Luhrmann, R. (2013). RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. EMBO J *advance online publication*.

Beckert, B., Nielsen, H., Einvik, C., Johansen, S.D., Westhof, E., and Masquida, B. (2008). Molecular modelling of the GIR1 branching ribozyme gives new insight into evolution of structurally related ribozymes. EMBO J *27*, 667-678.

Birgisdottir, A.B., Nielsen, H., Beckert, B., Masquida, B., and Johansen, S.D. (2011). Intermolecular interaction between a branching ribozyme and associated homing endonuclease mRNA. Biol Chem *392*, 491-499.

Burt, A., and Koufopanou, V. (2004). Homing endonuclease genes: the rise and fall and rise again of a selfish element. Curr Opin Genet Dev *14*, 609-615.

Carlomagno, T., Amata, I., Codutti, L., Falb, M., Fohrer, J., Masiewicz, P., and Simon, B. (2013). Structural principles of RNA catalysis in a 2'-5' lariat-forming ribozyme. Journal of the American Chemical Society *135*, 4403-4411.

Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Szewczak, A.A., Kundrot, C.E., Cech, T.R., and Doudna, J.A. (1996). RNA tertiary structure mediation by adenosine platforms. Science. *273*, 1696-1699.

Cheah, M.T., Wachter, A., Sudarsan, N., and Breaker, R.R. (2007). Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. Nature *447*, 497-500.

Correll, C.C., Munishkin, A., Chan, Y.-L., Ren, Z., Wool, I.G., and Steitz, T.A. (1998). Crystal structure of the ribosomal RNA domain essential for binding of elongation factors. Proc. Natl. Acad. Sci. USA *95*, 13436-13441.

David, G., and Perez, J. (2009). Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. J Appl Cryst *42*, 892-900.

Decatur, W.A., Johansen, S., and Vogt, V.M. (2000). Expression of the Naegleria intron endonuclease is dependent on a functional group I self-cleaving ribozyme. RNA *6*, 616-627.

Doherty, E.A., Batey, R.T., Masquida, B., and Doudna, J.A. (2001). A universal mode of helix packing in RNA. Nat Struct Biol *8*, 339-343.

Draper, W.E., Hayden, E.J., and Lehman, N. (2008). Mechanisms of covalent self-assembly of the Azoarcus ribozyme from four fragment oligonucleotides. Nucleic Acids Res *36*, 520-531.

Einvik, C., Nielsen, H., Westhof, E., Michel, F., and Johansen, S. (1998). Group I-like ribozymes with a novel core organization perform obligate sequential hydrolytic cleavages at two processing sites. RNA *4*, 530-541.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.
Franke, D., and Svergun, D.I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. Journal of Applied Crystallography *42*, 342-346.
Golden, B.L., Kim, H., and Chase, E. (2005). Crystal structure of a phage Twort group I ribozyme-product complex. Nat Struct Mol Biol *12*, 82-89.
Guinier, A. (1939). La diffraction des rayons X aux très petits angles: application à l'étude de phénomènes ultra microscopiques. Ann Phys *12*, 161-237.
Guo, F., Gooding, A.R., and Cech, T.R. (2004). Structure of the tetrahymena ribozyme; base triple sandwich and metal ion at the active site. Mol Cell *16*, 351-362.
Hafez, M., Majer, A., Sethuraman, J., Rudski, S.M., Michel, F., and Hausner, G. (2013). The mtDNA rns gene landscape in the Ophiostomatales and other fungal taxa: twintrons, introns, and intron-encoded proteins. Fungal genetics and biology : FG & B *53*, 71-83.
Haugen, P., Simon, D.M., and Bhattacharya, D. (2005). The natural history of group I introns. Trends Genet *21*, 111-119.
Ho, Y., Kim, S.J., and Waring, R.B. (1997). A protein encoded by a group I intron in Aspergillus nidulans directly assists RNA splicing and is a DNA endonuclease. Proceedings of the National Academy of Sciences of the United States of America *94*, 8994-8999.
Ho, Y., and Waring, R.B. (1999). The maturase encoded by a group I intron from Aspergillus nidulans stabilizes RNA tertiary structure and promotes rapid splicing. Journal of molecular biology *292*, 987-1001.
Inoue, T., Sullivan, F.X., and Cech, T.R. (1985). Intermolecular exon ligation of the rRNA precursor of Tetrahymena: oligonucleotides can function as 5' exons. Cell *43*, 431-437.
Johansen, S., Einvik, C., and Nielsen, H. (2002). DiGIR1 and NaGIR1: naturally occurring group I-like ribozymes with unique core organization and evolved biological role. Biochimie *84*, 905-912.
Johansen, S., and Vogt, V.M. (1994). An intron in the nuclear ribosomal DNA of Didymium iridis codes for a group I ribozyme and a novel ribozyme that cooperate in self-splicing. Cell *76*, 725-734.
Johansen, S.D., Haugen, P., and Nielsen, H. (2007). Expression of protein-coding genes embedded in ribosomal DNA. Biol Chem *388*, 679-686.
Kabsch, W. (2010). Xds. Acta Crystallogr D Biol Crystallogr *66*, 125-132.
Kleywegt, G.J., and Jones, T.A. (1994). A super position. CCP4/ESF-EACBM Newsletter on protein crystallography, 9-14.
Lambowitz, A.M., Zimmerly, S., and Perlman, P.S. (1999). Group I and group II ribozymes as RNPs: Clues to the past and guides to the future. In The RNA world second edition, R.F. Gesteland, T.R. Cech, and J.F. Atkins, eds. (Cold Spring Harbor, New York: CSHL Press), pp. 451-486.
Lee, C.S., Mui, T.P., and Silverman, S.K. (2011). Improved deoxyribozymes for synthesis of covalently branched DNA and RNA. Nucleic Acids Research *39*, 269-279.
Lee, E.R., Baker, J.L., Weinberg, Z., Sudarsan, N., and Breaker, R.R. (2010). An allosteric self-splicing ribozyme triggered by a bacterial second messenger. Science *329*, 845-848.
Leontis, N.B., and Westhof, E. (2001). Geometric nomenclature and classification of RNA base pairs. RNA *7*, 499-512.
Lipchock, S.V., and Strobel, S.A. (2008). A relaxed active site after exon ligation by the group I intron. Proceedings of the National Academy of Sciences of the United States of America *105*, 5699-5704.

Masquida, B., Beckert, B., and Jossinet, F. (2010). Exploring RNA structure by integrative molecular modelling. N Biotechnol *27*, 170-183.
Nielsen, H., Beckert, B., Masquida, B., and Johansen, S.D. (2008). The GIR1 branching ribozyme. In Ribozymes and RNA catalysis, D.M.J. Lilley, and F. Eckstein, eds. (London: Royal Society of Chemistry), pp. 229-252.
Nielsen, H., Einvik, C., Lentz, T.E., Hedegaard, M.M., and Johansen, S.D. (2009). A conformational switch in the DiGIR1 ribozyme involved in release and folding of the downstream I-DirI mRNA. RNA *15*, 958-967.
Nielsen, H., Fiskaa, T., Birgisdottir, A.B., Haugen, P., Einvik, C., and Johansen, S. (2003). The ability to form full-length intron RNA circles is a general property of nuclear group I introns. RNA *9*, 1464-1475.
Nielsen, H., Westhof, E., and Johansen, S. (2005). An mRNA is capped by a 2', 5' lariat catalyzed by a group I-like ribozyme. Science *309*, 1584-1587.
Nissen, P., Ippolito, J.A., Ban, N., Moore, P.B., and Steitz, T.A. (2001). RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. Proceedings of the National Academy of Sciences of the United States of America *98*, 4899-4903.
Reiter, N.J., Osterman, A., Torres-Larios, A., Swinger, K.K., Pan, T., and Mondragon, A. (2010). Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. Nature *468*, 784-789.
Schrodinger, LLC (2010). The PyMOL Molecular Graphics System, Version 1.3r1.
Sheldrick, G.M. (2008). A short history of SHELX. Acta Crystallogr A *64*, 112-122.
Stahley, M.R., and Strobel, S.A. (2006). RNA splicing: group I intron crystal structures reveal the basis of splice site selection and metal ion catalysis. Current Opinion in Structural Biology *16*, 319-326.
Svergun, D., Barberato, C., and Koch, M.H.J. (1995). CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. Journal of Applied Crystallography *28*, 768-773.
Tang, Y., Nielsen, H., Birgisdottir, A.B., and Johansen, S. (2011). A natural fast-cleaving branching ribozyme from the amoeboflagellate Naegleria pringsheimi. RNA Biol *8*, 997-1004.
Toor, N., Keating, K.S., Taylor, S.D., and Pyle, A.M. (2008). Crystal structure of a self-spliced group II intron. Science *320*, 77-82.
Vader, A., Johansen, S., and Nielsen, H. (2002). The group I-like ribozyme DiGIR1 mediates alternative processing of pre-rRNA transcripts in Didymium iridis. Eur J Biochem *269*, 5804-5812.
Vader, A., Nielsen, H., and Johansen, S. (1999). In vivo expression of the nucleolar group I intron-encoded I-dirI homing endonuclease involves the removal of a spliceosomal intron. EMBO J *18*, 1003-1013.
Vaidya, N., Manapat, M.L., Chen, I.A., Xulvi-Brunet, R., Hayden, E.J., and Lehman, N. (2012). Spontaneous network formation among cooperative RNA replicators. Nature *491*, 72-77.
Vicens, Q., and Cech, T.R. (2009). A natural ribozyme with 3',5' RNA ligase activity. Nat Chem Biol *5*, 97-99.
Zaher, H.S., Shaw, J.J., Strobel, S.A., and Green, R. (2011). The 2[prime]-OH group of the peptidyl-tRNA stabilizes an active conformation of the ribosomal PTC. EMBO J *30*, 2445-2453.

Figure captions

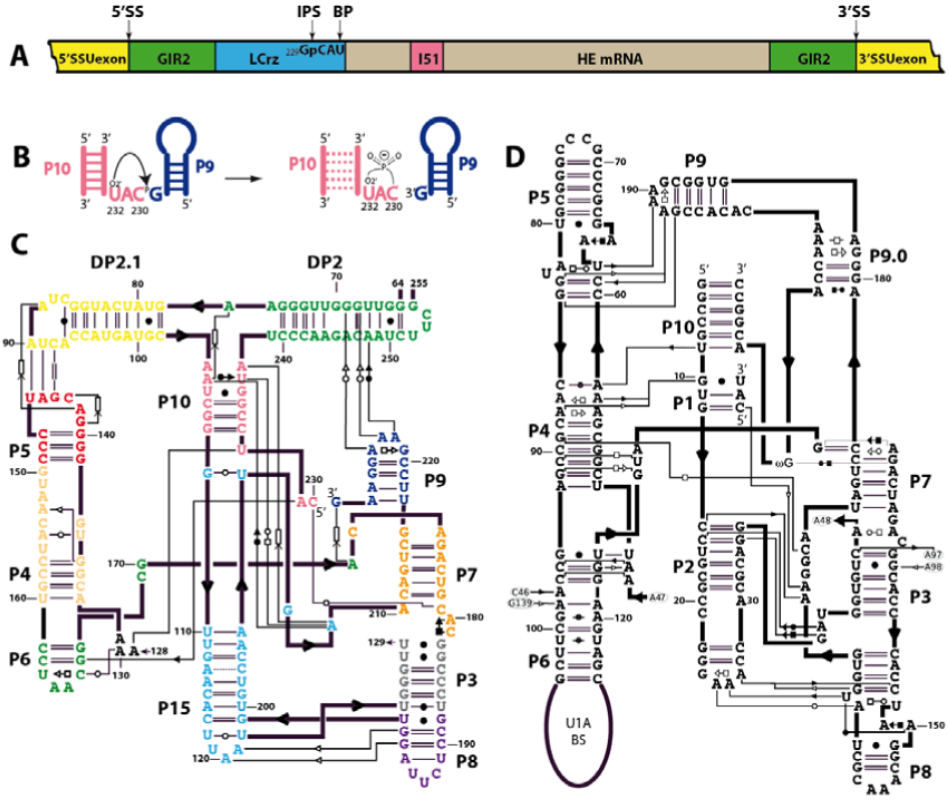


Figure 1, Meyer *et al.*

Figure 1 : (A) The lariat capping ribozyme (LC ribozyme) RNA is followed by a homing endonuclease mRNA (HE mRNA), which contains a spliceosomal intron (I51) in the case of *Didymium iridis*. This composite RNA is embedded in a regular group I intron (GIR2) interrupting the SSU rRNA precursor. The internal processing site (IPS) of LC ribozyme is only 3 nucleotides away from the branching point (BP). (1) and (2) point at the 5' and 3' splice sites, respectively. (B) Schematic representation of the branching reaction. Secondary structure diagrams of the DiLC ribozyme (C) and group IC3 ribozyme from *Azoarcus* (D) Note that a C residue was introduced as a PCR artifact in place of U203. Stem P1 is formed by the interaction of the intron with the 5' exon, which has the same CAU sequence as the lariat fold in LC ribozyme. Tertiary interactions are displayed using the Leontis-Westhof nomenclature (Leontis and Westhof, 2001). Stacking interactions in LC ribozyme are indicated by an arrow pointing into a rectangle.

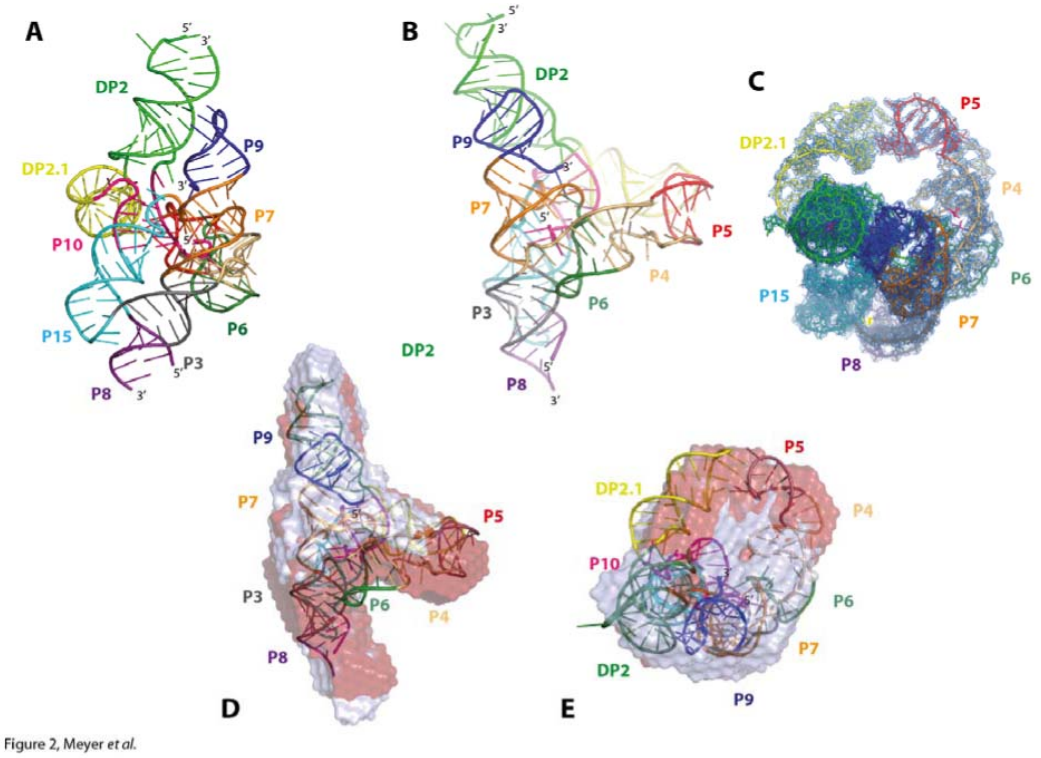


Figure 2 : (A) 3D structure of DiLC ribozyme is represented according to the orientation usually used for group I introns with P9 in the upper right corner and P15 (equivalent to the P1 and P2 stems in group I ribozymes) in the lower left corner. (B) Rotation by 90° around a vertical axis shows P9 spaced out from P5 by the specific orientation of P4-P6 and DP2-DP2.1, as resulting from the L2.1/L5 kissing complex. (C) Further 90° rotation around a horizontal axis shows the signet ring shape resulting from the orientation of P4-P6 and DP2-DP2.1 domains. The model is inscribed in a weighted 2Fo-Fc electron density map contoured to 1.0 σ. Statistics on data collection and refinement are summarized in Supplementary Table 1. (D) Fit of the low resolution shape obtained by SAXS calculated with DAMMIF (grey or red shapes correspond to the CP or wild-type LC ribozymes) and the crystal structure represented in ribbon and sticks model. The same fit rotated by 90° along *y* and *x* axis (E). Colors are as in **Fig. 1**. Drawing are prepared with Pymol (Schrodinger, 2010).

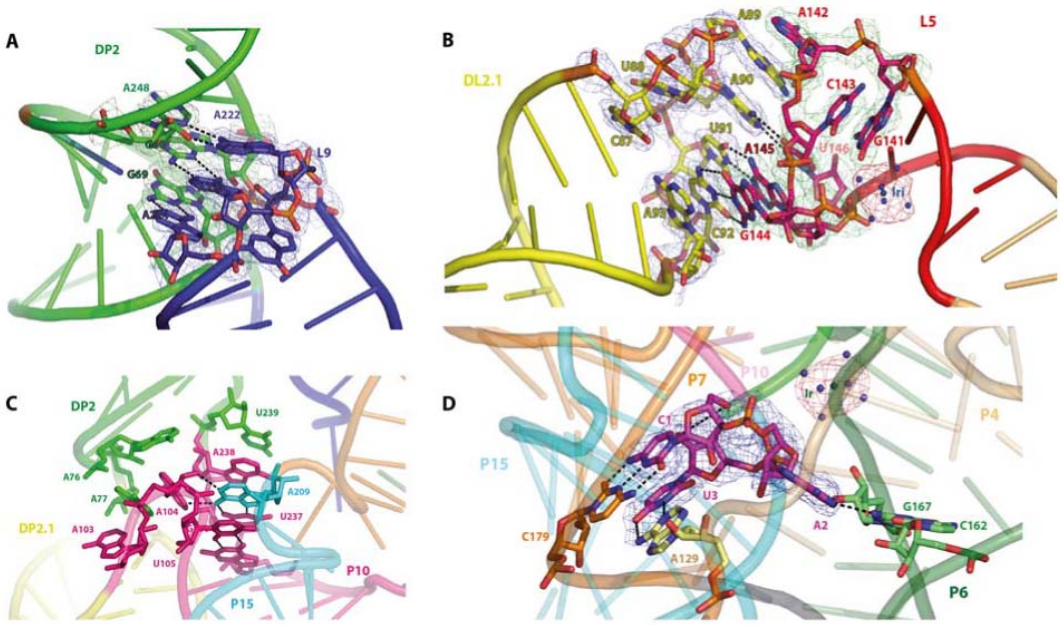


Figure 3, Meyer *et al*

Figure 3: Tertiary interactions stabilize the three-way junction (3WJ) involving P10, a critical element during catalysis. (A) The DP2/L9 tertiary interaction involves the Watson-Crick edge of residues A222, A223 and A224 from the L9 tetraloop (deep blue), which interact in the shallow groove of A248, G68 and G69 in DP2 (green), respectively. (B) The DL2.1 loop (yellow) and the L5 loop (red) form a kissing complex consisting of residues 5'-A_{90}UC-3' from DL2.1 and residues 5'-G_{144}AU-3' from L5. The iridium hexammine ion stabilizing the L5 fold can be seen in the deep groove (NH2: blue; Ir: pale green). (C) The 3WJ involving the regulatory domain DP2/DP2.1 is tethered to the core at the level of P10. It is mainly stabilized by means of stacking interactions between A residues belonging to the single strands linking the three paired elements. However A209 from the J15/7 junction mediates a critical contact necessary for the correct structuring of the 3WJ (**Fig. 3C**). Dark dashed lines indicate H bonds. Residues directly involved in tertiary interactions are represented as sticks. Helices are drawn in cartoon mode. (D) The lariat residues are depicted together with their interacting residues,

namely A129, the C162=G167 pair and C179 that belong to J3/4, P6 and J7/3, respectively. The iridium hexammine ion close to the lariat fold is also shown. The *cis* Watson-Crick/Watson-Crick U232-A129 base pair is sandwiched between the C230·C179 pair and an AC-platform formed by A180 and C181 (similar to the AA-platform observed in the *Tetrahymena* P4-P6 crystal structure (Cate et al., 1996).

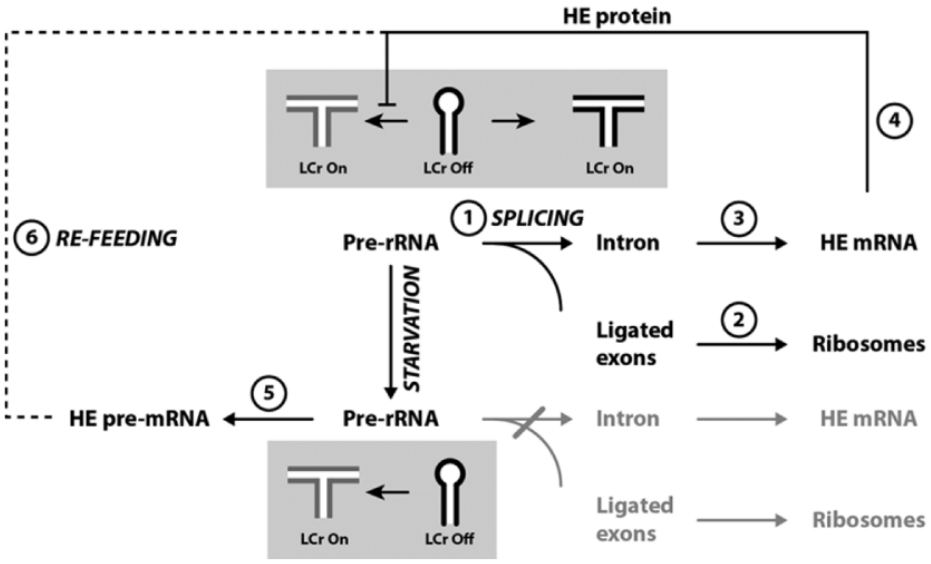


Figure 4: Switching of the Lariat capping ribozyme during a starvation-refeeding cycle. The LCrz can switch between an on- and an off-state. During normal growth, the ribozyme is in the off-state in pre-rRNA due to binding of the intron-encoded HE protein. Splicing (1) produces spliced out introns as well as ligated exons that are further processed (2) and incorporated into ribosomes. As a result of splicing the context of the LC ribozyme is changed leading to switching to the on-state. This leads to further processing of the intron (3) into a HE mRNA. The HE mRNA is translated (4) into HE protein that binds to the pre-rRNA stabilizing the off-state of LCrz. This feed-back loop is broken during starvation-induced encystment. Reduction in trancription elongation slows down splicing and reduces translation of the HE mRNA, which results in pre-rRNA transcripts outnumbering the HE protein. This in turn activates the LC ribozyme and results in alternative processing of the pre-rRNA by branching (5) producing HE pre-rRNA without rRNA production. Upon re-feeding and excystment, the HE pre-mRNA may be processed into mature mRNA that upon translation into HE protein re-establish the feed-back loop characteristic of growing cells.

This pdf file includes

- **Supplemental figures and legends**

  Figure S1 is related to Figure 1
  Figure S2 is related to Figure 2
  Figure S3 is related to Figure 3
  Figure S4 is related to Figure 3

- **Supplemental tables**

  Table S1 contains data collection and refinement statistics

- **Supplemental Experimental Procedures**

- **Supplemental References**

2

Supplemental figures and legends

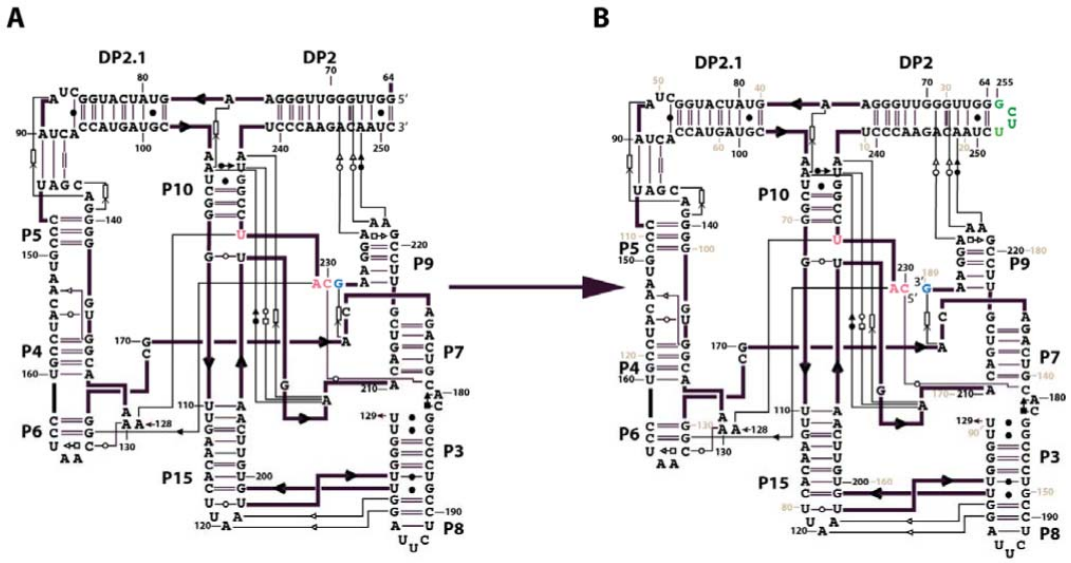


Figure S1, Meyer *et al*

Figure S1: Design of the lariat capping ribozyme (LCrz) circular permutation. (A) Secondary structure of the wild type sequence of the LCrz with the three nucleotides C230, A231 and U232 forming the post-cleavage lariat in pink and G229 equivalent to group I ribozyme ωG in blue. The sequence is annotated as in the intron. (B) Circular permutation construct of the LC ribozyme with T7 *in vitro* transcription start site at C230. The DP2 stem is closed by a 5'UUCG3' tetraloop shown in green. The sequence is annotated as in the intron in black and as in the PDB file (4HHK) in grey. Note that residues are numbered continuously in the PDB files despite discontinuities in the structure at two places. The first U of the UUCG loop of DP2 is missing as well as the two U residues from L8.

2

3

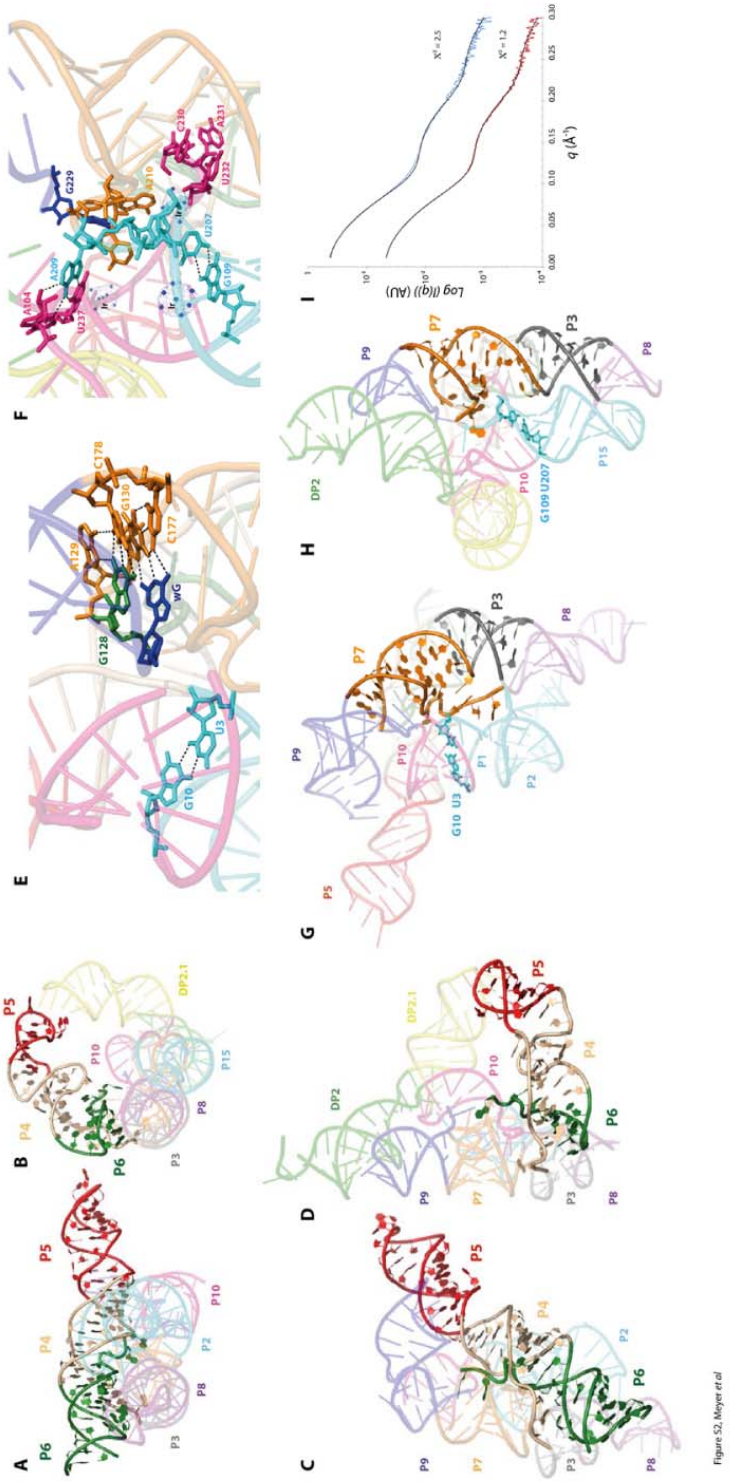


Figure S2, Meyer et al

3

4

Figure S2: Global comparison of group I intron from *Azoarcus* (Azo) and DiLCrz. To compare 3D structures, the Azo (A) ribozyme was oriented similarly to DiLCrz (B) by superimposing their P3 and P7 stems using all backbone atoms in program *lsqman* (Kleywegt and Jones, 1994). The resulting *rmsd* for the considered 153 atoms is 3.14 Å. Views of the two ribozymes can thus be analyzed pairwise. P15 of DiLCrz is equivalent to the P1/P2 stack in Azo. The GoU base pair of each ribozyme is shown as sticks (cyan). Colors of each domain are the same in the two ribozymes. The P4-P6 domain is laid perpendicularly on the Azo core (C), contrarily to the same domain in DiLCrz, in which L5 forms a kissing complex with DL2.1, conferring to DiLCrz its signet-ring shape (D). A 90° rotation around a horizontal axis shows that J5/4 is ideally positioned to play its role in stabilization of the crucial GoU base pair of P1 in Azo (E). Rather, in DiLCrz, J5/4 adopts a zigzag motif remote from the catalytic site (F). ωG is in the G-binding pocket of P7 and close to the GoU base pair of P1. It forms a base triple with G130 and C177 of Azo. G128 from J6/7 forms a base triple with A129 and C178 from P7 (G). The GoU *trans* Watson-Crick base pair at the top of P15 allows A209 to interact with A104 and U237 in the 3WJ involving DP2, DP2.1, and P10. C230, A231 and U232 together form a properly folded lariat. In the post-catalytic state G229, homologous to ωG in Azo, does not occupy the G-binding pocket of P7 as in group I intron. Indeed, the G-binding pocket is not formed in P7 of DiLCrz. Noteworthy, A210 can adopt two conformations one of which pointing at the lariat indicating that it may play a role in the catalysis (H). (I) displays the fits between the I(q) SAXS curves obtained in solution with the circularly permutated (CP, red curve) and the wild-type (blue curve) forms of the LC ribozyme and the the crystal structure model of the CP LC ribozyme. Noteworthy a better fit is observed in the case of the wild-type active ribozyme ($\mathbf{X}^2$ = 1.2), which indicates that the crystal packing may restrain the dynamics of the CP LC ribozyme.

4

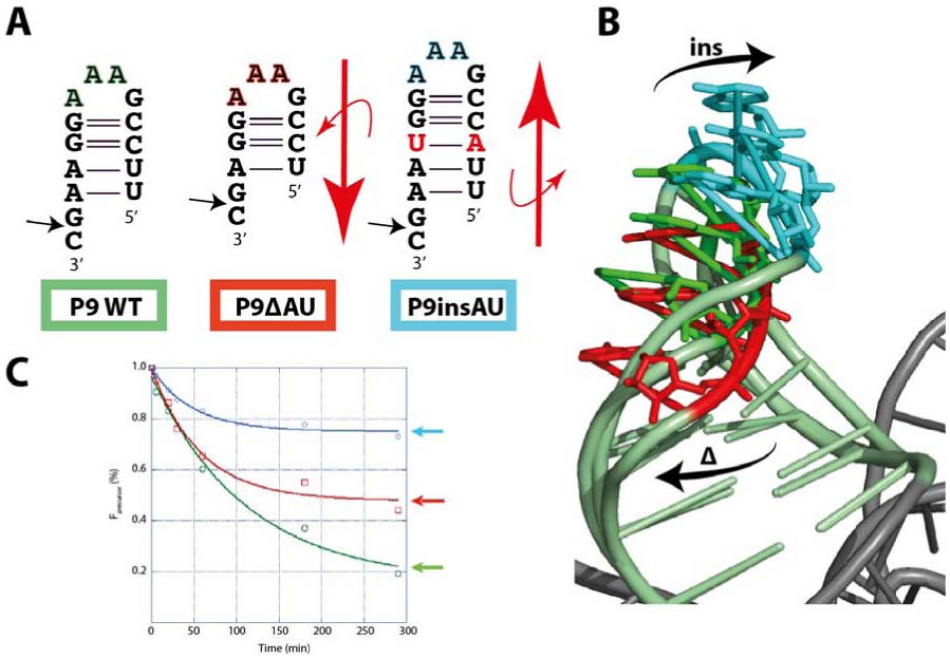


Figure S3: P9 mutants demonstrate the role of the P2/L9 interaction. (A, B) Deleting or adding one base pair to P9 translates and rotates L9 and therefore disturbs its interaction with DP2. Black arrows point at the cleavage site. Insertion of one base pair orients away from the receptor the adenine residues from the loop, while deletion of one base-pair maintains the interaction potential of two adenine residues (B). These effects are suggested by the catalytic tests determined under single turnover conditions (C).

6

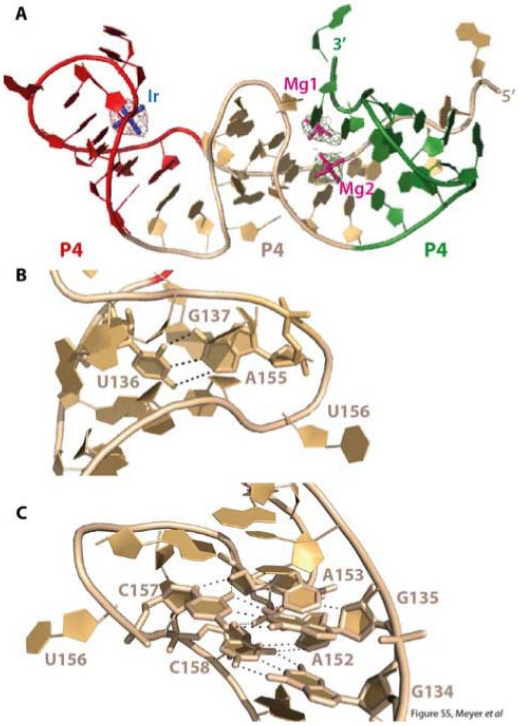


Figure S4: (A) Overall view of the P4/P6 zig-zag motif. The three metal ions stabilizing the motif are represented within the corresponding weighted 2Fo-Fc densities countoured at 1.0 sigma. The first Mg^{2+} ion is penta-hydrated and bound to the O2P atom from G170 while the coordinated water molecules form H-bonds to O6 positions of bases G134 and G135. The second Mg^{2+} ion is hexa-hydrated and bridges the bases of C159 and G169 from P4 and J6/7, respectively. The iridium hexammine ion in the deep groove of P4 stabilizes the loop by binding to the unpaired G141 and the preceding G=C base pair. (B) Detail of the motif showing the inflexion point where a trans Watson-Crick U-A base pair intervenes promoted by stacking with G137 and protruding of U156. (C) From another point of view, the motif presents the A-minor interactions mediated by A152 and A153 to the G=C pairs involving G134 and 135 within P4.

6

7

Supplemental tables

Table S1: Statistics on Crystallographic Data Collection and Refinement

	Crystal 1 (4 wavelength Ir MAD)				*Crystal 2*
	peak	***inflection***	***high remote***	***low remote***	
Data Collection					
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell a (Å)	59.44	59.44	59.39	59.47	57.63
b (Å)	86.26	86.32	86.10	86.38	85.71
c (Å)	109.78	109.82	109.65	109.88	108.53
Beamline	X06DA/SLS	X06DA/SLS	X06DA/SLS	X06DA/SLS	X06DA/SLS
Wavelength (Å)	1.10530	1.10564	1.09553	1.11525	1.6
Resolution limits (Å)	48.9 - 2.9	48.9 - 2.9	48.9 - 2.8	48.9 - 2.9	47.8 - 2.5
high resolution shell	3.0 - 2.9	3.0 - 2.9	2.9 - 2.8	3.0 - 2.9	2.6 - 2.5
Reflections: measured	82329	82184	188727	85513	494356
unique	23432	23367	26765	24203	34693
Completeness (%)	96.8 (69.3)*	96.4 (65.9)*	99.9 (99.7)*	99.6 (99.7)*	96.4 (94.5)*
R_{meas} (%)	15.1 (92.1)*	16.7 (114.6)*	17.1 (118.2)*	16.9 (124.9)*	10.4 (110.9)*
I / σ	9.81 (1.49)*	9.27 (1.19)*	13.71 (1.89)*	9.33 (1.19)*	24.88 (2.49)*
CC(1/2)	99.3 (57.4)*	99.2 (42.5)*	99.6 (60.3)*	99.3 (40.9)*	99.9 (76.5)*
Phasing					
Resolution range used (Å)	48.9 - 3.4				
Heavy atom sites§	4 Ir				
Correlation coefficient (all/weak)	47.0 / 32.2				
FOM	0.7				
Pseudo-free CC	72.23				
Refinement					
Reflections R_{cryst} / R_{free}	15298/766				34663/1968
R_{cryst} (%)	19.81				18.80
R_{free} (%)†	26.21				23.57
RNA atoms	4076				4076
Heterogen atoms	111				121
Water molecules	21				88
Average B factor (Å^2)	16.40				18.60
R.m.s.d. bond length (Å)	0.08				0.002
R.m.s.d. angles length (°)	1.571				0.629

7

8

* Values in parentheses correspond to high resolution shell in data collections
† 5 % of the reflections were set aside for an R_{free} test before initiating any refinement
§ Shelxd output

8

9

Supplemental Experimental Procedures

Preparation of the circularly permutated form of the LCrz ribozyme

RNA transcripts were obtained by *in vitro* transcription using a His-tagged version of the T7-RNA polymerase (Ichetovkin et al., 1997). An insert obtained by PCR corresponding to the circularly permutated form of the LCrz inserted between a 5' hammerhead ribozyme and a 3' HδV ribozyme (Price et al., 1995) was inserted in an EcoRI/BamHI digested pUC19 (NEB) using the InFusion cloning kit (Clontech). The PCR reaction introduced an U203C mutation resulting in the presence of an A-C base pair in P15. This mutation had been previously reported as stabilizing the Naegleria lariat-capping ribozyme (Jabri and Cech, 1998). Fast cleavage in the course of the transcription reaction by the 5' hammerhead ribozyme was possible because the sequence of stem I was changed to the LCrz while preserving the tertiary interactions between stems I and II (Martick and Scott, 2006). One hour incubation at 60°C was necessary to complete HδV cleavage. Thus, the RNA of interest presents 5' hydroxyl and 2',3' cyclic phosphodiester ends which do not correspond to the branching reaction products, in the sense that the missing phosphate group of C230 prevents tethering. The lariat is expected to be folded but not 2',5'-tethered in the structure. This is actually what is observed in the crystal structure. The transcript was purified by 8% polyacrylamide gel electrophoresis, identified by UV-shadowing, recovered by passive elution overnight, and filtrated on dispensable sterile 0.2µm filtration units (Nalgene). The RNA sample was concentrated by ethanol precipitation.

9

10

Crystallization of the LCrz

A 100 µM RNA solution was heated during 1 min at 95°C followed by 4 min on ice. $MgCl_2$ up to 4 mM was then added, and the solution was placed at 55°C for 10 min, then at 4°C overnight. Initial conditions were obtained from commercial high-throughput screening kits from *Hampton research* (PEGRx, Index, Natrix) and *Jena Biosciences* (Nuc-Pro). Parallelepiped shaped crystals grew in ~20 days in microbatch settings in 0.2 M sodium chloride, 0.1 M HEPES pH 7.5, 25% w/v PEG 3,350.

Data collection, structure determination and model refinement

Crystals were flash-frozen by immersion in liquid ethane. Diffraction data were collected at 100 K at the macromolecular crystallography beamline X06DA at the Swiss Light Source (SLS) and processed using XDS (Kabsch, 2010). Phase information was determined with data obtained from crystal 1 (Table S1) by a 4-wavelength MAD experiment around the iridium edge using the SHELX package (Sheldrick, 2008).

The resulting experimental electron density map allowed to build an initial model that was later refined with Phenix (Adams et al., 2010) and iteratively built using Coot (Emsley and Cowtan, 2004) against higher resolution data obtained from crystal 2 to final R_{work}/R_{free} values of 18.80%/23.57% (see Table S1). Density around J6/7 region as well as the first nucleotides of P7 was poorly defined but was perfectly visible when the structure was refined against the data from crystal 1 (final R_{work}/R_{free} values of 19.81%/26.21%).

Model statistics were obtained with MolProbity (Chen et al., 2010). Molecular visualizations

10

and structures illustrations were performed using Pymol (Schrodinger, 2010). Data collection and refinement statistics are given in Table S1. Coordinates and structure factors have been deposited with the PDB ID 4HHK and 4HHN.

SAXS data collection and analysis

X-ray scattering data were collected on beamline SWING at Synchrotron SOLEIL (Gif-sur-Yvette, France). The beam wavelength was set to $\lambda = 1.033$ Å. The 17 x 17 cm^2 low-noise Aviex CCD detector was positioned at a distance of 1800 mm from the sample, with the direct beam off-centred. The resulting exploitable Q-range was 0.014 – 0.4 $Å^{-1}$, where $Q = 4\pi \sin\theta / \lambda$, and 2θ is the scattering angle.

To remove the aggregation contribution, concentrated solutions of circularly permutated or active wild-type LC ribozyme (25 µl, 8mg/ml) were injected into a Size Exclusion Column (Agilent Bio-sec3) using an Agilent © High Performance Liquid Chromatography (HPLC) system and eluted directly into the SAXS flow-through capillary cell at a flow rate of 200 µl / min (David and Perez, 2009). SAXS data were collected online throughout the whole elution time, with frame duration of 0.5 s to prevent X-ray damage and a dead time between frames of 0.5 s. The frames corresponding to the main elution peak were inspected and averaged by using FOXTROT, a homemade program dedicated to data processing. In total, 20 frames were averaged to produce a curve with the same Q-range as previously, with no attractive interactions at low angle.

12

All subsequent data processing, analysis and modeling steps were carried out with PRIMUS (Konarev et al., 2003), and other programs of the ATSAS suite (Konarev et al., 2006). The radius of gyration *Rg* was evaluated using the Guinier approximation (Guinier, 1939), assuming that at very small angles ($Q < 1/3\ R_g$), intensity may be represented as $I(Q) = I(0) \exp(-(Q.R_g)^2/3)$. The same parameter was also calculated from the entire scattering pattern using the indirect transform package GNOM (Svergun, 1992), which also provides the distance distribution function *p(r)* of the particle.

Low-resolution models of circularly permutated or active wild-type LC ribozymes were generated *ab initio*, using the dummy atom model method, as implemented in the program DAMMIF (Franke and Svergun, 2009), which uses simulated annealing to search a compact model that fits the experimental data to minimize the discrepancy

$$\chi^2 = \frac{1}{N-1}\sum_j \left[\frac{I_{\exp}(s_j) - cI_{\mathrm{calc}}(s_j)}{\sigma(s_j)}\right]^2$$

where N is the number of experimental points, $I_{exp}(Q)$, $I_{calc}(Q)$ and $\sigma(Qj)$ are the experimental and calculated intensity and the experimental error at the momentum transfert Qj, respectively, and *c* is a scaling factor. The experimental errors $\sigma(Q)$ were evaluated for each frame from the scattered number of photons N(Q) as $\sigma(Q) = \sqrt{N(Q)}$ and adequately propagated.

A set of low resolution models were calculated and compared with DAMAVER, a procedure to analyze and align *ab initio* models in order to select the most typical one (Volkov and Svergun, 2003). Scattering from the crystallographic model of the LC ribozyme was calculated using CRYSOL (Svergun et al., 1995) and compared to the circularly permutated LC ribozyme or to

12

13

the active wild-type experimental curves. The program fits the experimental intensity I(Q) with the atomic coordinates by adjusting the excluded volume, averaged atomic radius and the contrast of the hydration layer surrounding the particle in solution to minimize discrepancy as described in Eq. (1).

The superimposition of low-resolution molecular shapes and crystallographic structures were performed with SUPCOMB, a method to optimize the alignment of three-dimensional objects. A normalized spatial discrepancy (NSD) is introduced as a proximity measure between three-dimensional objects, and gives a quantitative estimate of similarity between the objects after minimization (Kozin and Svergun, 2001).

13

Supplemental References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.
Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr *66*, 12-21.
David, G., and Perez, J. (2009). Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. Journal of Applied Crystallography *42*, 892-900.
Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.
Franke, D., and Svergun, D.I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. Journal of Applied Crystallography *42*, 342-346.
Guinier, A. (1939). La diffraction des rayons X aux très petits angles: application à l'étude de phénomènes ultra microscopiques. Ann Phys *12*, 161-237.
Ichetovkin, I.E., Abramochkin, G., and Shrader, T.E. (1997). Substrate recognition by the leucyl/phenylalanyl-tRNA-protein transferase. Conservation within the enzyme family and localization to the trypsin-resistant domain. J Biol Chem *272*, 33009-33014.
Jabri, E., and Cech, T.R. (1998). In vitro selection of the Naegleria GIR1 ribozyme identifies three base changes that dramatically improve activity. RNA *4*, 1481-1492.
Kabsch, W. (2010). Xds. Acta Crystallogr D Biol Crystallogr *66*, 125-132.
Kleywegt, G.J., and Jones, T.A. (1994). A super position. CCP4/ESF-EACBM Newsletter on protein crystallography, 9-14.
Konarev, P.V., Petoukhov, M.V., Volkov, V.V., and Svergun, D.I. (2006). ATSAS 2.1, a program package for small-angle scattering data analysis. Journal of Applied Crystallography *39*, 277-286.
Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. Journal of Applied Crystallography *36*, 1277-1282.
Kozin, M.B., and Svergun, D.I. (2001). Automated matching of high- and low-resolution structural models. Journal of Applied Crystallography *34*, 33-41.
Martick, M., and Scott, W.G. (2006). Tertiary contacts distant from the active site prime a ribozyme for catalysis. Cell *126*, 309-320.
Price, S.R., Ito, N., Oubridge, C., Avis, J.M., and Nagai, K. (1995). Crystallization of RNA-protein complexes I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. J. Mol .Biol. *249*, 398-408.
Schrodinger, LLC (2010). The PyMOL Molecular Graphics System, Version 1.3r1.
Sheldrick, G.M. (2008). A short history of SHELX. Acta Crystallogr A *64*, 112-122.
Svergun, D. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. Journal of Applied Crystallography *25*, 495-503.
Svergun, D., Barberato, C., and Koch, M.H.J. (1995). CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. Journal of Applied Crystallography *28*, 768-773.

15

Volkov, V.V., and Svergun, D.I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. Journal of Applied Crystallography *36*, 860-864.

15

3. Additional results

We also collected diffraction data for the DiLCrz_active_U203C construct up to 4 Å and solved the structure by the molecular replacement method using the structure of DiLCrz_cp_U203C. Unit cell parameters change significantly between the cp and the active DiLCrz_U203C constructs crystals **(Table 16)**. Some regions of the DiLCrz_active_U203C fit unambiguously in the density map e.g. the P3 stem or the very important DP2/DP2.1/P10 3WJ **(Figure 48)**. However because some nucleotides are clearly positioned differently in the active LCrz in comparison to the post-cleavage state one **(Figure 49)** it is impossible to confidently build the model at a resolution of 4 Å and hence to proceed to refinement **(Table 16)**. It is for the moment out of question to determine in which state is the LCrz in the crystal as hydrolysis or branching could have occurred during crystallization. We now try to reproduce this crystal to solve the structure *de novo* using heavy atoms derivatives.

	DiLCrz_cp_U203C	DiLCrz_active_U203C
Resolution (Å)	2.5	4
Space group	P $2_12_12_1$	P $2_12_12_1$
Angles (°)	α=90; β-90; γ=90	α=90; β-90; γ=90
Length (Å)	a=57.63; b=85.71; c=108.53	a=52.54; b=88.82; c=105.65
Rwork (%)	21.5	35.7
Rfree (%)	27.1	45.5
Methods	MAD	MR

Table 16: Table summarizing cell parameters, resolution, refinement statistics and method for 3D structure resolution for DiLCrz_U203C cp and active constructs.

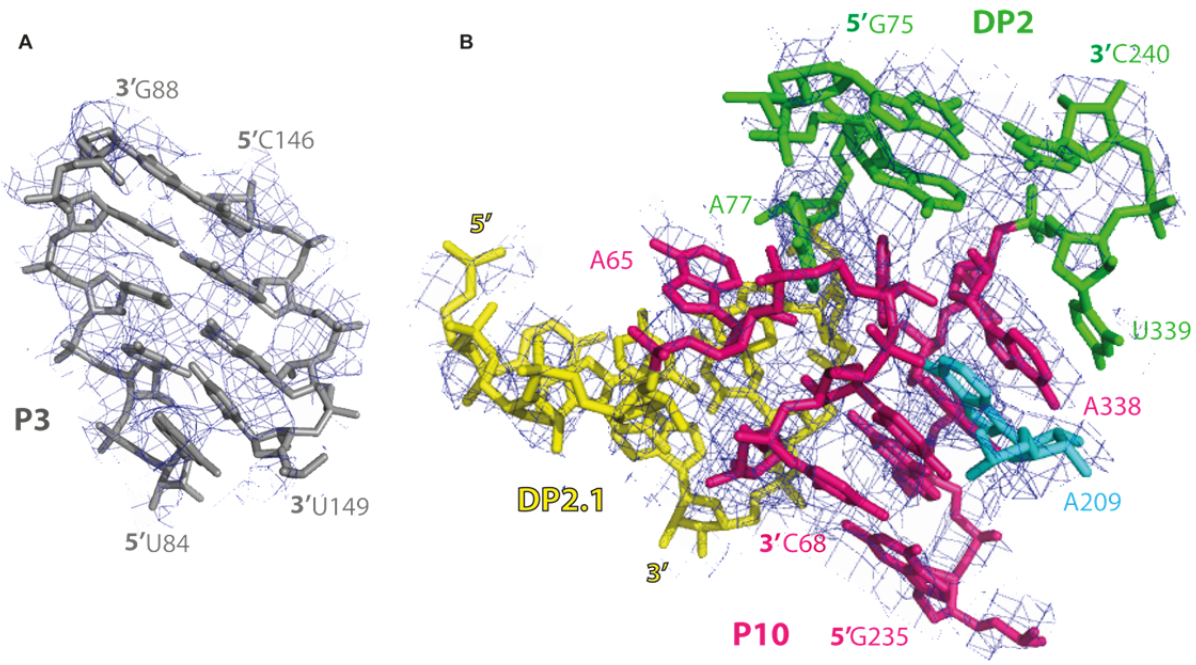


Figure 48: 3D structures of P3 stem (A) and DP2/DP2.1/P10 3WJ (B) in the 2Fo-Fc composite omit maps of DiLCrz_active_U203C. Stems are colored as in the article of the crystal structure of DiLCrz.

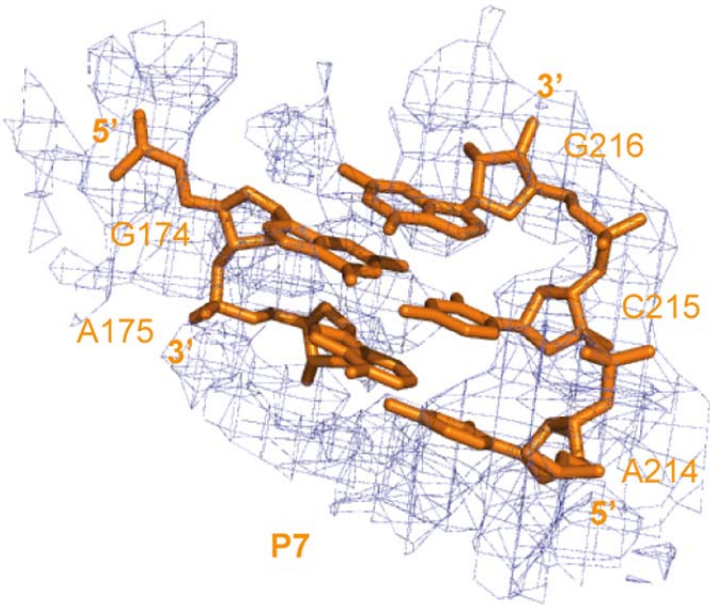
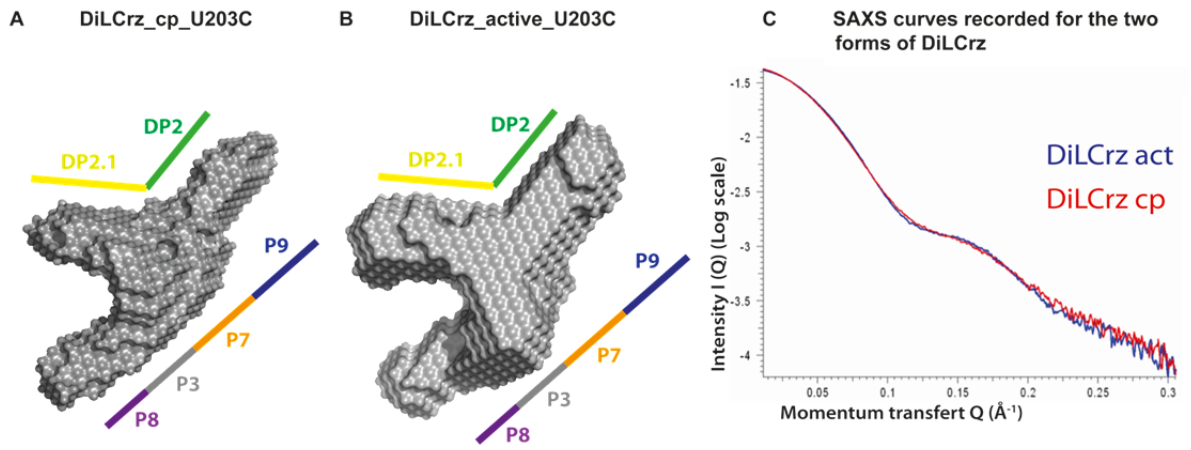


Figure 49: 3D structures of the two first base pairs of P7 stem clearly shifted from the 2Fo-Fc composite omit maps of DiLCrz_active_U203C.

SAXS data have been collected as well for the DiLCrz_active_U203C construct and we can conclude that the pre- and post-cleavage ribozymes fold in the same global shape **(Figure 50)**. Indeed, the crystal structure of DiLCrz_cp_U203C fits without ambiguity into SAXS shapes of both forms of DiLCrz. Interestingly the fit

is slightly better in the case of DiLCrz_active_U203C SAXS shape. This indicates that crystal packing could favor DiLCrz_cp_U203C to adopt a conformation closer than in solution to the active one.



D

	Rg (Guinier) (Å)	Rg (P(r)) (Å)	Dmax (Å)	Volume (Å^3)
DiLCrz cp	30.9	30.8	105	87850
DiLCrz act	29.8	29.7	100	87850

Figure 50: SAXS shapes of DiLCrz_cp_U203C (A) and DiLCrz_active_U203C (B). Stem loops of DiLCrz are schematically drawn to orient the SAXS shapes and colored as in the article of the crystal structure of DiLCrz. (C) Superposition of SAXS curves recorded for the two forms of DiLCrz. (D) Table that summarizes SAXS parameters for each DiLCrz.

4. Discussion

As described in the introduction, LCrz is part of a twintron as it is embedded into a canonical group I intron, GIR2 (Decatur et al, 1995). There are other examples of a heterogenous twintrons, in the sense that both introns within the twintron are not identical, found in the gene coding for the mitochondrial small subunit ribosomal RNA of fungi where a group IIA intron is inserted within the P9.1 stem of a group IC2 intron (Hafez et al, 2013). More interesting to us, a homogenous group I twintron is as well observed in the SSU rRNA of two species of the *Ophiostomatales* taxa. The twintron is composed of an external ORF-less group ID intron and an internal group ID intron followed by the ORF of a HEG inserted in the P1 stem of the external intron (Hafez et al, 2013). The latter observation in addition to the previous sequence alignments,

secondary structure comparison with *Azoarcus* group IC3 intron, mutagenesis and modeling data (Beckert et al, 2008) lead us to the assumption that LCrz results from the evolutionary drift of a group I intron that under new selective pressure acquired new functions to escape from a probable loss (Goddard & Burt, 1999) **(Figure 51)**.

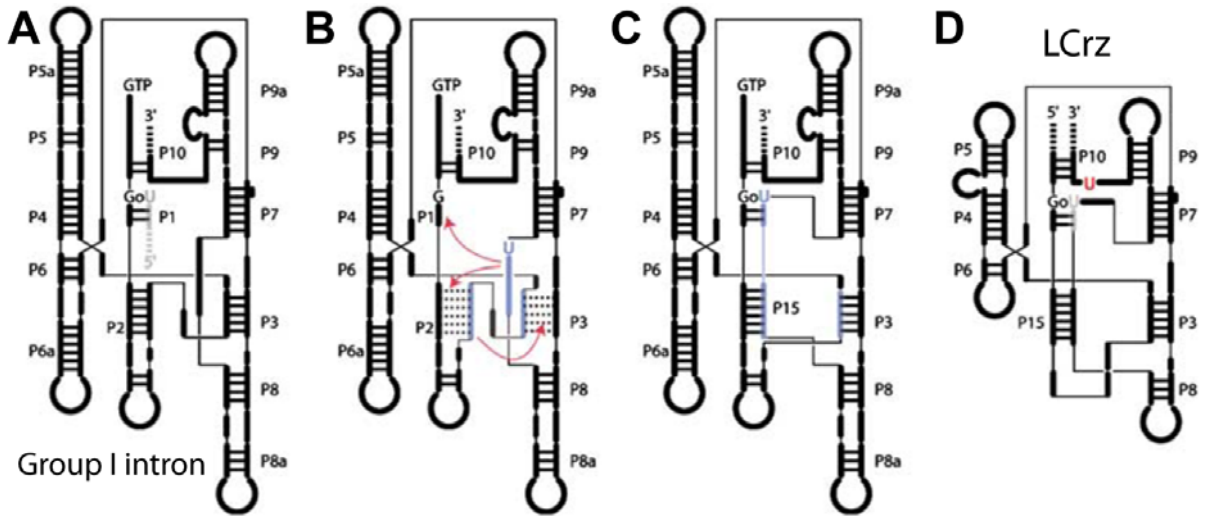


Figure 51: A model for the metamorphosis of a group I intron into LCrz. Figure adapted from (Beckert et al, 2008). (A) Azoarcus group I intron a step just before the second step of splicing undertakes some mutation in J8/7, P2 and P3 that allows for partial alternatives base pairing between the substrate and the catalytic domains. (B) Loss of the 5' exon favors secondary structure's remodeling especially of the long stretch of nucleotides left unpaired in P1. (C) All structural elements are in place to perform the branching reaction. (D) Loss of structural element not necessary to the energetic stabilization of the active conformation of the new LCrz.

Some preliminary experiments have been performed to test *in vitro* "retro-evolution" of LCrz. A proof of concept has been established (B. Beckert thesis 2010) using *Azoarcus* pre-tRNA splicing ribozyme to demonstrate that biotin-GTP can be used as a cofactor to promote the splicing pathway in a SELEX experiment (Wilson & Szostak, 1999). It would also be interesting to solve the 3D structure of LCrz deprived of its DP2/DP2.1 peripheral domain to check whether it would fold into a group I intron fold.

Because LCrz followed by the ORF of a HEG is inserted either in P2 of DiGIR2 or in the P6 stem of NaGIR2 and AlGIR2, we suggest that LCrz together with the gene of its cognate HE invaded GIR2. This event could have happened at least twice in evolution as the GIR2 recipient group I introns are of different class

depending on the eukaryote in which the twintron is observed. Still in agreement with the Goddard and Burt cycle, the HE should acquire new function not to be lost (Chatterjee et al, 2003; Goddard & Burt, 1999). Though HE in *Didymium iridis* is still capable of intron homing *in vivo* (Johansen et al, 1997). With respect to this hypothesis, the question whether GIR2 was an ORF-less intron or has lost its HEG is still open. A possibility is that GIR2 was an ORF-containing group I intron and the sequence between the 5' strand of GIR2 and the HE ORF is encoding for an intron insertion site therefore allowing intron reverse splicing at this specific location. LCrz could have emerged from the sequence drift of a group I intron that has promoted its own mobility at the RNA level (Hedberg & Johansen, 2013).

One prospective function of the HE would be to regulate the on- and off-conformational switch of LCrz. Indeed, since LCrz is fully transcribed before GIR2, it should not react before self-splicing of GIR2 has been completed since this would lead to unligated exons of the SSU rRNA (Vader et al, 1999). Additionaly, the mutually exclusive DP2 and HEG P1 stems are observed in the on and off states, respectively **(Figure 6)** (Nielsen et al, 2009). The HE of the twintron could recognize one or the other conformation and modulate the conformational switch. However we cannot exclude that a completely different enzyme could be recruited to regulated LCrz peripheral conformational switch as e.g. group I introns are capable of hijacking proteins to accomplish a maturase function to help their own splicing (Garriga & Lambowitz, 1986; Paukstelis et al, 2008). Nonetheless, we cannot exclude that the HE has acquired new functions unrelated to its catalytic activity or to the regulation of the twintron processing.

Nature is likely not altruistic, therefore co-evolution of LCrz and HE for mutual interest is possible especially as LCrz is always found followed by HEG. NaHEG or NaLCrz insertions are not observed solely in Nae.S516 twintron (Haugen et al, 2004; Wikmark et al, 2006). Additionally, having a look back at the three processing pathways that the twintron GIR2/LCrz can undertake **(Figure 3),** we observe that during the starvation-induced encystement pathway the production of mature SSU rRNA is down regulated while there is an accumulation of the capped 7.5 kb mRNA of the HE (Nielsen et al, 2008; Vader et al, 2002). The capped mRNA of the HE is

protected from degradation at its 5’ end by the lariat of three nucleotides and at the 3’ end by additional nucleotides belonging to the 3’ exon and GIR2. The accumulated mRNA would therefore be ready to be translated when the myxomycete comes out of encystement and a large amount of HE could be produced (Nielsen et al, 2008). If the HE regulates the on and off switch of LCrz this phenomenon would allow for rapid up-regulation of the mature SSU rRNA production allowing the cell to quickly restore the optimal level of ribosome and therefore of the resulting cellular functions.

B. THE PK-TURN, A NEW RNA KINKING MOTIF

1. Introduction

RNA motifs are of two types: (1) RNA motifs that only implicate primary sequences as for example the Shine-Dalgarno motif or (2) those that adopt a 3D structure characterized by a set of specific structural features. We focused on the second ones (reviewed in (Masquida et al, 2010)) and when RNA motifs are later discussed in the manuscript they always imply structural characteristics in the sense of the three-dimensional structure.

WC base pairs, mostly distributed in A-type helices, represent 60 to 70% of the total base pairs of a structured RNA (Westhof et al, 2011) and RNA motifs arose from the match of single stranded nucleotides linking one helix to the others. RNA motifs are defined by a set of weak interactions including non-WC base pairs folding the phosphodiester backbone of the considered RNA strands in a constrained and characteristic manner (Leontis et al, 2006; Leontis & Westhof, 2003). RNA motifs consensus sequence encompass all isosteric non-WC base pairs ordered and directed identically to fold in the same three-dimensional structure (Leontis & Westhof, 2003). RNA motifs allow additional nucleotides to be inserted if they do not disturb the overall geometry of their folding (Abramovitz & Pyle, 1997; Klein et al, 2001). RNA motifs not only join and orient RNA edging helices; they also mediate RNA-RNA or RNA-protein interactions and can be drug targets (Ban et al, 2000; Wimberly et al, 2000). Among the well-known RNA motifs are: the loop E (Correll et al, 1997; Leontis et al, 1986), the sarcin-ricin motif (Leontis et al, 2002), the GNRA tetraloop (Abramovitz & Pyle, 1997), and its receptor (Cate et al, 1996b), the kink-turn or k-turn (Vidovic et al, 2000), the G-ribo (Steinberg & Boutorine, 2007), the RA motif (Grabow et al, 2012).

K-turns are widespread motifs present e.g. in the spliceosomal RNA U4 (Vidovic et al, 2000), in riboswitches (Blouin & Lafontaine, 2007; Montange & Batey, 2006), in the human RNAse P (Davila Lopez et al, 2009) and multiple times in rRNA (Klein et al, 2001). A database dedicated to k-turns is available online: www.dundee.ac.uk/biocentre/nasg/kturn/ (Schroeder et al, 2010). The k-turn motif is

defined by a 15 nucleotides sequence organized in two helices, the canonical one (C-stem) bridged by a bulge of 3 nucleotides to the non-canonical one (NC-tem) capped by two G.A base pairs **(Figure 52)** (Lilley, 2012). The k-turn motif can be inserted in both directions but the angle between C- and NC-stem is always around 60° and the minor grooves of each helix face each other. L1 and L2 nucleotides are stacked onto C- and NC-stem respectively when L3 is bulged out (Lilley, 2012). The two G.A pairs are *trans* sugar edge (G) to Hoogsteen edge (A) pairs and sequence variation of these base pairs is mostly observed at position 2b.2n (Cannone et al, 2002; Schroeder et al, 2012; Wimberly et al, 2000). Stacking and hydrogen bonds hold tightly the RNA motif into its fold. Three hydrogen bonds are crucial: (1) H bond between 2'O of L1 and N1 of A1n, (2) H bond between 2'O of L3 and the *pro*S non-bridging oxygen of the phosphate group between L1 and L2 and (3) an A-minor interaction between the nucleotide at position -1n and either N1 or N3 of the highly conserved adenosine 2b **(Figure 52B)**. K-turn folding can be induced by metal ions (Liu & Lilley, 2007), protein interaction (Turner et al, 2005) and/or long range tertiary interactions (Antonioli et al, 2010; Schroeder et al, 2011). K-turns are at least related to two motifs: reverse k-turn (Strobel et al, 2004) and pk-turn.

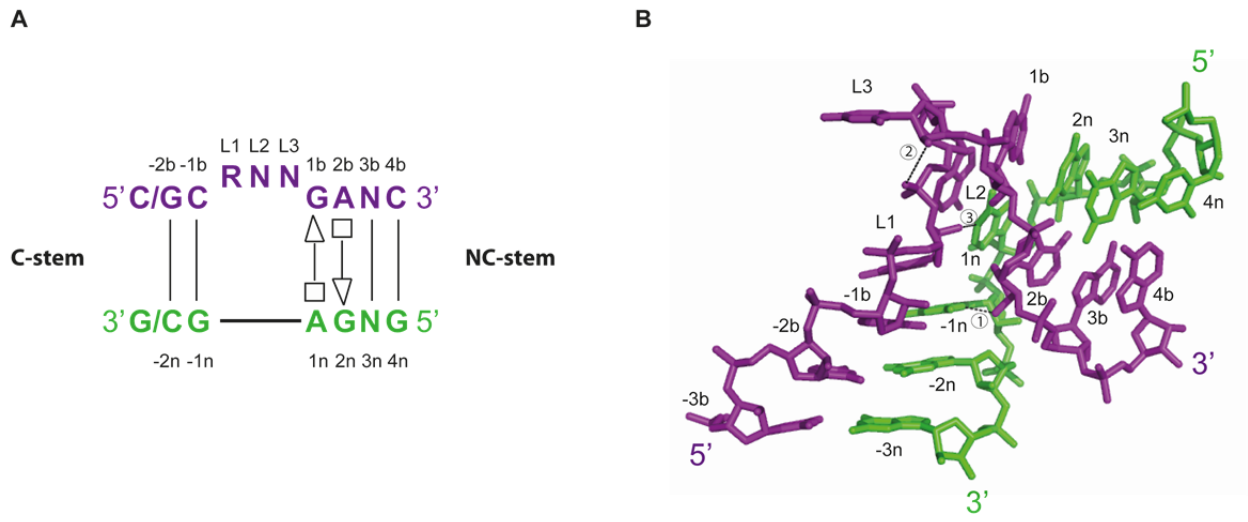


Figure 52: (A) Consensus sequence of the k-turn motif numbered as defined by David Lilley (Lilley, 2012). Characteristic G.A base pairs at the beginning of the NC-stem are represented according to LW nomenclature (Leontis & Westhof, 2001). (B) 3D structure of the k-turn motif present in the SAM-I riboswitch (Stoddard et al, 2010) labeled as the consensus sequence. Circled number show: (1) H bond between 2'O of L1 and N1 of A1n, (2) H bond between 2'O of L3 and the *pro*S non-bridging oxygen of the phosphate group between L1 and L2 and (3) an A-minor interaction between the nucleotide at position -1n and N1 of the highly conserved adenosine 2b. In both panels, the longest strand that contains the three bulged nucleotides is shown in purple and the shortest in green.

The pk-turn that we describe in the following article is not just another k-turn, but a new motif that expands the variety of motif able to sharply kink RNA helices. The pk-turn motif observed in the P16-P17 domain of the 3D structure of the bacterial RNAse P bound to its tRNA substrate (Reiter et al, 2010) does not respect any sequence or base pairs geometry requirements of the k-turn motif (Lilley, 2012). Nevertheless the pk-turn motif allows the same kinking and twisting of the edging helices than the k-turn.

Because RNA motifs are important from a structural and functional point of view, it is of great interest to be able to spot them in available 3D structures (Apostolico et al, 2009; Cruz & Westhof, 2011b; Petrov et al, 2011; Sarver et al, 2008; Wadley & Pyle, 2004) but also to be able to model them from sequencing data to help determining the 3D structure of large RNA (Cruz & Westhof, 2011b; Masquida et al, 2010; Westhof et al, 2011). New RNA 3D structure is always the time for new motifs to be observed (Holbrook, 2005).

2. Article

RNA Biology 9:3, 254–260; March 2012; © 2012 Landes Bioscience

A structural module in RNase P expands the variety of RNA kinks

Mélanie Meyer, Eric Westhof and Benoît Masquida*

Architecture et Réactivité de l'ARN, Université de Strasbourg, IBMC, CNRS; Strasbourg, France

Keywords: kink-turn, RNase P, RNA structure, RNA motif, RNA module, structural alignment

RNA structures are built from recurrent modules that can be identified by structural and comparative sequence analysis. In order to assemble sets of helices in compact architectures, modules that introduce bends and kinks are necessary. Among such modules, kink-turns form an important family that presents sequence and structural characteristics. Here, we describe an internal loop in the bacterial type A RNase P RNA that sets helices bound at the junctions exactly in the same relative positions as in kink-turns but without the structural signatures typical of kink-turns. Our work suggests that identifying a structural module in a subset of RNA sequences constitutes a strategy to identify distinct sequential motifs sharing common structural characteristics.

Introduction

It is now well accepted that RNA structures are formed by the assembly of specific building blocks and that, among those, the A-form helix is the most represented. Standard Watson-Crick helices are separated by structural modules that organize the different domains in the three-dimensional space and confer to the molecule interaction capacities within the same RNA or with other RNAs, proteins and ligands (see ref. 1 for review). Comparative sequence analysis and crystallographic structures of RNA modules embedded in large RNAs have permitted the development of several ab initio and homology modeling strategies.[2-5] The success of ab initio modeling strategies relies on the correct attribution of a sequence to a given structural module. To achieve this goal, knowing all the sequences subjected to the formation of specific module is important.

RNA modules display a limited number of covariation rules. They are generally more difficult to characterize in terms of sequence because they are often complicated (1) by sequence variations (insertions or deletions of bulges); (2) by the complex network of interactions maintaining their integrity; and (3) by their propensity to make long distance contacts with RNAs or proteins. Based on precise structural alignments of sequences, progress toward the identification of known modules has been made.[6,7] This applies to the kink-turn (k-turn) module which generally connects two helices by means of a G-A tandem preceded on one side by an average 3 nt bulge (**Fig. 1**) with almost no sequence restrictions except those that arise from the anchoring of proteins which is systematically observed for k-turns found in rRNAs.[8,9] The G-A tandem accepts sequence variations in some cases, but those are restricted to the G-A pair distal from the bulge.[9] The two framing A-form helices adopt a specific relative orientation consisting of a ~90° kink and a roughly ~180° twist between the Watson-Crick base pairs contiguous to the non-Watson-Crick part of the module.

Here we report an RNA module fully distinct from all other modules kinking RNA that have been described to date.[9] It consists of an asymmetrical internal loop that orients the framing helices exactly as regular k-turns do. This module occurs in the crystal structure of the bacterial RNase P holoenzyme bound to its tRNA product[10] and is located in the P16-P17 domain. This motif is distinct from the putative k-turn that has been described in the P12 domain of the human RNase P RNA.[11] For these reasons, it was named the pk-turn (p stems from RNase P). With respect to kinking and twisting of the framing helices, the pk-turn presents the same architectural properties as k-turns, albeit with completely different base pairs and sequence requirements. If structural modules are defined as sets of ordered non-Watson-Crick pairs embedded within Watson-Crick pairs,[7] then the pk-turn could be viewed as a new module. However comparative sequence analysis shows that some RNase P sequences present a standard k-turn module at this specific place. Following this definition, the pk-turn increases the sequence diversity of the known kinking modules. This observation helps to understand how RNA modules are selected and evolve from a pool of sequences and comforts the view that module conservation throughout the phylogeny is not pre-required to prove the occurrence of a motif. On the opposite, the finding of a module in a narrow subset of the phylogenetic tree seems to point to regions where distinct modules sharing common structural characteristics intervene. This idea could be applied to interpret sequence alignment results, to refine computational search of RNA modules in sequence databases, to infer some of their functional aspects, and to improve molecular modeling strategies.

*Correspondence to: Benoît Masquida; Email: b.masquida@ibmc-cnrs.unistra.fr
Submitted: 12/14/11; Revised: 01/19/12; Accepted: 01/20/12
http://dx.doi.org/10.4161/rna.19434

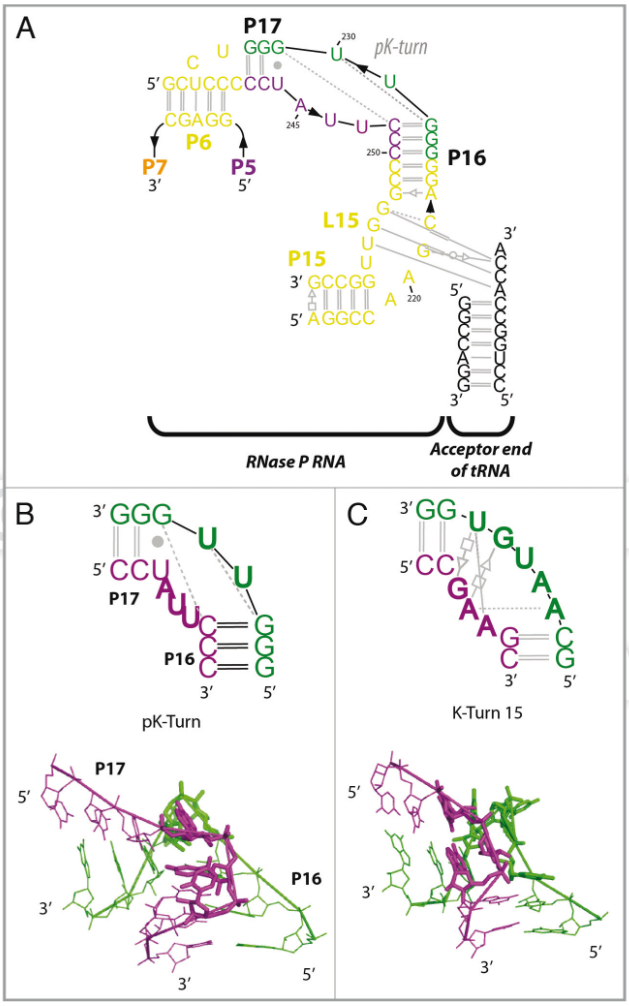


Figure 1. Secondary structure of the region comprising the pk-turn (A). Residue numbering is according to.[10] The scheme shows how the pk-turn is critical for the simultaneous correct presentation of the P6 pseudoknot and of the L15 loop that interacts with the acceptor end of the pre-tRNA substrate (black lettering). Comparative secondary and 3D structures of the pk-turn (B) and of the classical k-turn number 15[6] (C) from the *Haloarcula marismortui* 23S ribosomal crystal structure.[33] Residues unpaired or involved in non-canonical interactions are in bold face. The sequence motifs are completely distinct although they produce the same kink-turn between the bound helices. On the 3D panel (bottom), residues forming helices are drawn in thin lines and residues constituting the non-canonical motif are represented by sticks.

Material and Methods

The pk-turn module was identified in the RNase P holoenzyme crystal structure[10] and superimposed with the k-turn-15[6] using all backbone atoms from residues listed in **Table S1**. The *rmsd* and the translation-rotation matrices were calculated using lsqman.[12] 3D pictures were prepared using PyMOL[13] and Assemble.[3] All secondary structures from bacterial RNase P type A RNAs from the RNase P database[14] were checked and sorted according to their specific features (see **Supplemental Material**). A sequence alignment was derived using S2S.[15] The pk-turn candidate sequences were aligned to define a consensus using the program RNAalifold.[16] Sequences with extended bulges with respect to the reference *T. maritima* structure were considered as outliers and not considered in the consensus definition.

Results

Early modeling studies[17,18] and the crystal structure of the naked bacterial type A RNase P RNA[19] have shown that a kink has to occur in the region of the RNase P RNA encompassing stems P15/P16/P17 responsible for the binding of the acceptor end of the pre-tRNA substrate. However, its precise location as well as its architecture have only been revealed by the recent crystal structure from the *Thermotoga maritima* RNase P ribozyme,[10] in which the addition of the tRNA to the holoenzyme contributes to the stabilization of this region. Despite its average resolution (3.8 Å), the crystal structure shows without ambiguity that the kink takes place at the internal loop separating P16 from P17. Close inspection of the kink shows that it is distinct from the known k-turns[8,9] in terms of sequence and base pair geometry of the internal loop. However, the relative positions and orientations of the framing helices are common to the standard k-turn module.

The formation of the RNase P P16/P17 kink results from the flexibility of the two strands composing the internal loop. One U residue of each strand bulges out (U229 and U247) while U246 is oriented within the helical stack and could form a *trans* Watson-Crick base pair with G228 if the latter were not already involved in pairing with C248 (**Fig. 1**). The only purine of the motif (A245) provides stacking continuity by capping P17. The absence of G residues obviously departs from the characteristic k-turn G-A pair directly following the bulge. Most importantly, P16 and P17, the stems bound to the RNase P kink-turn, superimpose very well with the stems bound to a regular k-turn (*rmsd* of 1.5 Å; **Fig. 2**), as is the case for the superimpositions performed between different k-turns (reported in ref. 8).

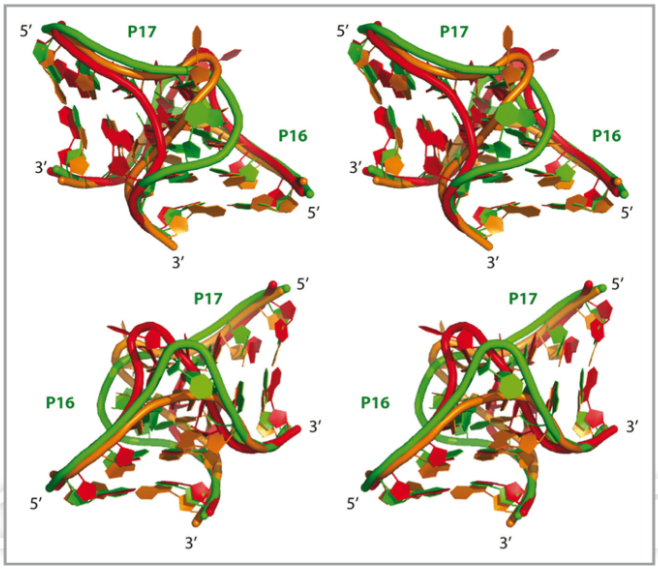


Figure 2. Wall-eye stereo view of the superimposition between the pk-turn (green ribbon, residues 235–242 and 251–259; PDB ID code 3OK7) and k-turn 15 (kt-15, residues 245–249 and 259–265; PDB ID code 1S72) in two opposite orientations (orange and red ribbons). Superimposition of the backbone atoms of the two Watson-Crick base-pairs of the helices bound to each motif were performed using *Isqman*.[12] The view on the lower panel is rotated 180° toward the view on the upper panel. *Rmsd* values and residues used in their determination are listed in **Table S1**.

Analysis of 295 secondary structures representative of the bacterial phylogeny (secondary structures from the RNase P database[14]) reinforces this observation (**Table 1**). A significant portion of bacterial type A RNase P RNA seemingly carries at this position a motif that fully respects the structural constraints characteristic of a regular k-turn module in several organisms (Deinococcus, Thermus, high G + C Gram$^+$, *Mycobacterium tuberculosis*). Interestingly, both positions of the 3 nt bulge with respect to the G residues from the G-A tandem are observed in the sequences bearing a k-turn (**Fig. 3**, upper panel, and **Fig. 4**). In RNase P, the most abundant k-turns have their G-A tandem rooted in P17 (**Table 1**). Inspection of the secondary structures also shows that three-way junctions (3WJ, in Chlamydia, *Clostridium sporogenes*, α purple bacteria) and four-way junctions (4WJ, in Bacteroides and Planctomycetes), occur between P16 and P17 (**Fig. 3**, lower panel).

Among internal loops, k-turns are represented in ~20% and pk-turns in 5% of the considered structural alignments. Hence, those modules do not represent the majority of the internal loops observed in this region. Roughly 40% of the aligned sequences exhibit a 2 to 5 nt bulge in J16/17 not evenly scattered within phyla but mostly occurring in γ purple bacteria and cyanobacteria. A couple of sequences with a bulge in J17/16 is also observed (*Propioniferax innocua, Aeromicrobium erythreum* and *fastidiosum*). It is difficult to infer the structure of these bulges directly from the alignments. Nonetheless, it seems very likely that a kink should also occur at this place since the secondary structure from the *Thermotoga maritima* RNase P enzyme is very similar to the one from γ purple bacteria and cyanobacteria. Moreover, P16 and P17 are systematically longer in the latter phyla (6 and 5 bp or 6 and 3 bp, respectively) than in Thermotogales (5 and 3 bp). This observation could indicate that fraying of some base pairs may take place to facilitate the proper conformation.

A structural alignment restricted to internal loops (**Fig. 4**) illustrates how the four main families of modules could interchange and play equivalent roles. The k-turn modules are split between two families, having the bulge either in J16/17 (KT1) or in J17/16 (KT2). KT1 largely dominates (**Table 1**). In these two situations, the G-A tandem extends P17 or P16, respectively (**Fig. 3**). A similar situation is observed for the pk-turn module. The majority of the observed pk-turns (11 in 15) is composed of an asymetrical internal loop with a J17/16 strand longer than the J16/17 one while fewer sequences have the opposite configuration (4 in 15, **Table 1**). The *Borrelia burgorferi* motif (**Fig. 3**) certainly represents an alternative sequence to the *T. maritima* pk-turn since the sequence of the closely related *Borrelia hermsii* is indeed very close to it (**Fig. 5A**). Some Prochlorococcus representatives seem to represent an intermediate stage between the Thermotoga and Borrelia pk-turn motifs. It is important to note that the kink between the helices allows easy accomodation of slightly longer bulges in J16/17 and J17/16 explaining why some pk-turns present one or two nucleotides insertions. The sequences from these three genus were used to produce a consensus of the pk-turn motif (**Fig. 5B**). On the opposite,

Table 1. Classification and distribution by phylum of motifs occurring in the secondary structures of the RNase P RNA P16-P17 region

	Purple Bacteria						Gram+G+C		Cyano	Bacte	Spiro	Chlam	Plancto	Green S	Green NS	Deino	Therm	UA	Total
	α	β	γ	δ	ε	U	High	Low											
3WJ1	15	3	10	1	-	1	6	-	-	1	-	-	6	2	-	-	-	12	57
3WJ2	-	-	-	-	-	-	-	1	-	2	-	24	-	-	-	-	-	5	32
K-T1	-	9	1	2	-	1	38	-	-	-	-	-	-	-	3	-	-	2	56
K-T2	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	3	-	-	4
pKT1	-	-	-	-	-	-	-	-	5	-	2	-	-	-	-	-	2	2	11
pKT2	-	-	-	-	-	-	-	-	4	-	-	-	-	-	-	-	-	-	4
npKT1	7	-	65	-	-	5	-	5	22	1	2	1	-	-	-	-	-	8	116
npKT2	-	-	-	1	-	-	3	-	-	-	2	-	-	-	-	-	-	-	6
4WJ	-	-	-	-	-	-	-	-	-	1	-	-	1	-	-	-	-	2	4
DP16	-	-	-	-	3	-	-	-	-	-	-	-	-	-	-	-	-	2	5
total	22	12	76	4	3	7	47	6	31	5	6	25	8	2	3	3	2	33	295

3WJ1, Three-way junction with additional hairpin inserted between in J16/17; 3WJ2, Three-way junction with additional hairpin inserted between in J17/16; KT1, k-turn with GA tandem anchored in P17; KT2, k-turn with GA tandem anchored in P16; pKT1, pk-turn found in the *Thermotoga maritima* RNase P RNA crystal structure with J16/17 shorter than J17/16 or inserted in opposite orientation (pKT2); npKT1, Motifs expected to form a k-turn based on a 3 to 4 nt bulge in J16/17 or J17/16 (npKT2); 4WJ, Four-way junction with either two hairpins inserted in J17/16 or one hairpin in each junction between P16 and P17; ΔP16, Case specific to the ε subdivision of purple bacteria in which P16 is deleted resulting in a large L15 loop directly connected to P17. U, purple bacteria of uncertain affiliation; Cyano, cyanobacteria; Bacte, Bacteroides; Spiro, Spirochaetes; Plancto, Planctomycetes; Green S, green sulfur bacteria; Green NS, green non-sulfur bacteria; Deino, Deinococci; Therm, Thermotogales; UA, bacteria of uncertain affiliation.

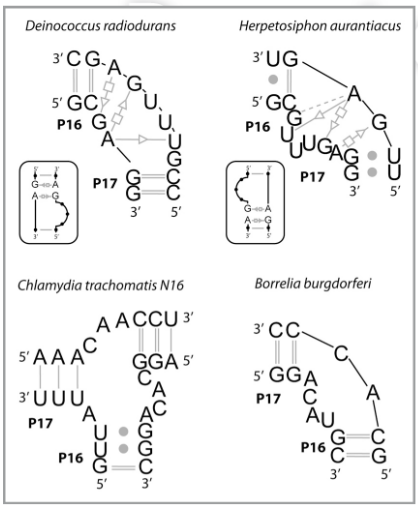


Figure 3. Examples of secondary structures of different motifs intervening at the P16-P17 junction in the bacterial type A RNase P RNA. *D. radiodurans* and *H. aurantiacus* RNase P RNAs show the presence of a k-turn in place of the pk-turn motif. Interestingly these k-turns are inserted in opposite orientations (see insets corresponding to each secondary structure). The relative orientations of P16 and P17 seems to be conserved also by means of a 3-way junction (*C. trachomatis*) or by other motifs of adopting a structure presumably close to the pk-turn (*B. burgdorferi*).

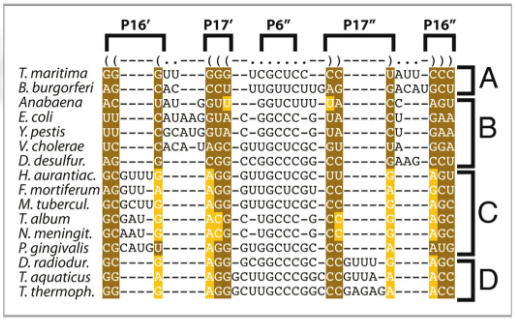


Figure 4. Alignement of a set of sequences of type A RNase P RNAs representative of the bacterial phylogeny. Sequences were gathered from the RNase P database.[14] Brown boxes indicate Watson-Crick conservation of base pairs forming P16 and P17 stems. Orange residues within these boxes indicate non-isosteric *cis* WW base pairs (Anabaena, *T. album N. meningiditis*) or non-canonical *trans* SH G-A pairs (sequences in groups C and D) according to the LW base pair classification.[34] Sequences are sorted according to the type of motif contained in the loop connecting P16 to P17. (A) pk-turn, (B) kink-turns with unknown structures, (C) k-turns with the bulge upstream of the G-A tandem in the junction between P16 and P17 (J16/17), (D) k-turns in orientation opposite to (C), with the bulge upstream of the G-A tandem in J17/16. Purple bacteria: *Neisseria meningitidis, Escherichia coli, Vibrio cholerae, Yersinia pestis, Desulfovibrio desulfuricans* ; Gram$^+$ bacteria: *Mycobacterium tuberculosis*; cyanobacteria: *Anabaena*; Bacteroides and relatives: *Porphyromonas gingivalis*; Spirochaetes: *Borrelia burgdorferi;* green non-sulfur bacteria: *Herpetosiphon aurantiacus;* Deinococcus and relatives: *Deinococcus radiodurans, Thermus thermophilus, Thermus aquaticus;* Thermotogales: *Thermotoga maritima*; bacteria of uncertain affiliation: *Fusobacterium mortiferum.*

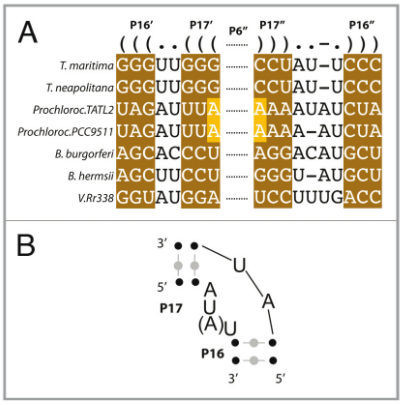


Figure 5. Alignement of a subset of pk-turn candidate sequences in the RNase P P16/P17 region. (A) The 7 sequences shown are representative of the three genus displaying the pk-turn. Sequences with extended bulges were removed and the same number of sequences for each genus was taken in order to avoid increasing the impact of outliers as described in.[16] Sequences not used in **Figure 4** include *Thermotoga neapolitana, Prochlorococcus TATL2, Prochlorococcus PCC9511, Borrelia hermsii, Volunteer Rr338* (Seq ID: AF296062). (B) Consensus secondary structure derived from the alignment using the program RNAalifold.[16] The A residue in parenthesis corresponds to an insertion observed in the Prochlorococcus and Borrelia genus.

hypothetical non-*Thermotoga maritima* pk-turns (npk-turn) are characterized by a 2–5 nt bulge in J16/17 structurally closer to k-turns with the bulge in J16/17. Thus, all these sequence motifs are expected to form genuine kink-turns accommodating the formation of the P6 pseudoknot together with the interaction with the pre-tRNA substrate. However, unlike for the pk-turn, the absence of a npk-turn crystal structure prevents to assess whether npk-turns promote folding of the P16/P17 region by means of positioning the stems as the pk-turn does. Along the same line of idea, it is reasonable to hypothesize that the 3WJ and 4WJ should also preserve the relative orientation of P16 and P17. Rare exceptions to this general trend are seen in organisms (ε purple bacteria, *Campylobacter jejuni* and *Helicobacter pylori*) with an RNase P RNA in which P16 is seemingly absent resulting in a large L15 loop directly connected to P17. In these cases, it is likely that the extended L15 may form a kink-turn as in RNase P RNAs from other phylogenetic groups, although, no evidence can be presented.

Discussion

The impressive molecular resemblance between the pk-turn and k-turns constitutes a strong piece of evidence that RNase P RNAs with a P16/P17 sequence containing the k-turn constraints indeed fold as such. This study thus shows how the pk-turn replace k-turns to play an equivalent structural role in an homologous region of the RNase P RNAs. This provides a convincing illustration of how evolution selects different RNA sequences to promote the formation of equivalent modules,[1] in this case, a k-turn or k-turn-like sequential motifs to favor the presentation of the pre-tRNA substrate to the RNase P enzyme. Moreover, the k-turn strands can be exchanged without affecting the relative orientation of P16 and P17 (KT1 or KT2, **Table 1**). Interestingly, the strands composing a loop E motif in the GOLLD RNA[20] are also found exchanged in a sequence subset. The exchange between the k-turn strands suggests that the recognition of the pre-tRNA substrate together with the formation of the P6 pseudoknot establish the main selection pressures that drive the evolution of the sequence in the junction between P16 and P17. This conclusion could lead to a strategy to elucidate whether given RNA modules interact or not with proteins. Most k-turns indeed interact with proteins to form asymmetric complexes as between the 15.5 kD protein and the snRNA U4 k-turn[21] or between k-turns found in rRNAs and their cognate proteins.[8] In these cases, inverting the k-turn is equivalent to swapping the bulge and the G-A tandem from one stem to the other and thus to change the orientation of the bound protein. We can then conclude from the observation that the P16/P17 junction harbors k-turns in both orientations that it does not require any cognate protein during catalysis since it would result in two different binding modes for a candidate protein across the phylogeny. Developing analysis methods to search for inverted motifs within homologs would thus help systematically finding domains that may favor protein binding.

It is worth noting that the *Thermotoga maritima* lysine riboswitch also presents a kink-like module between stem P2a and loop L2 that is replaced by a regular k-turn in other lysine riboswitches.[22] However, superimposition of the considered P2a-L2 motif with kt-7 from 16S rRNA or with the kink from the P4P6 domain from the Tetrahymena group I intron[23] fails to show a straightforward structural relationship between those three motifs.[24] The lysine riboswitch P2a-L2 motif takes place in a context very different from the k-turn since the L2 loop forms a pseudoknot with L3 and thus cannot be considered as an internal loop per se. The resulting different topology in which this motif ascribes itself raises the question of what precise structure would adopt the L2 k-turn sequence. Along the same line of idea, k-turn forming sequences can also fold in a reverse orientation[25] as seen in the P9 domain of the Azoarcus group I ribozyme crystal structure.[26] Yet, the structural context forces it to adopt a different base-pairing scheme to support the formation of the critical tertiary interaction between the L9 tetraloop and its receptor located in P5. Grafting of a consensus k-turn sequence in the P9 domain does not modify the overall fold as proven by the crystal structure of an Azoarcus group I ribozyme mutant.[27] If k-turns and most probably other RNA modules can be forced to adopt other conformations, it is thus likely that the opposite situation may also be true. Sequences not respecting the consensus for a k-turn may be forced by the structural context to adopt a k-turn conformation. The energetic constraints imposed by the whole RNA architecture and contacts stabilizing the crystal packing may thus force proper folding of a given module. This idea is supported by a recent experimental study on the SAM-I riboswitch in which mutants of the k-turn motif which prevent

correct folding in isolation have no effect in the context of the full length riboswitch.[28]

As mentioned above, crystal structures of the P16/P17 region based on 2–5 nt bulges (npk-turn) are missing. Nevertheless, modeling of these bulges shows that fraying of at least one base pair from either P16 or P17 is necessary to force the stems to adopt a conformation favoring the formation of contacts between the shallow grooves of the stems characteristic of k-turns (data not shown). Without fraying, the kink serves to move away the two bound stems so that the deep grooves face each other as observed in two recent crystal structures.[29,30] Fraying actually occurs in the *Thermotoga maritima* RNase P RNA[10] in which nucleotides involved in the first base pair of P17 do not interact in the crystal structure in contradiction with the initial secondary structure. Rather these residues help mediate smooth connections between P16 and P17 in order to facilitate the pk-turn fold. These conclusions bring the idea that the overall architecture of the RNase P RNA may energetically force the folding of the P16/P17 domain to form a kink-turn explaining why this region accepts a high level of sequence variations. The consensus deduced in this study (**Fig. 5B**) may now be used to identify new occurrences of the pk-turn in other RNAs using a variety of strategies.[7,31]

Other motifs (2–5 nt bulges, 3WJ and 4WJ) found within the secondary structure data set should also be instrumental in precisely orienting P16 and P17 so that they roughly form a 90° angle and their shallow grooves interact. The 2–5 nt bulge in J16/17 may represent a family of motifs aiming at inserting a local energetic instability that will appropriately respond to P6 folding by acquiring a kinked conformation. 3WJ are known for mediating fairly precise angles between their stems.[32] One of the three families of 3WJ (family A) specifically fulfills both the angle between the helical stack and the third helix and the contacts between the shallow grooves.

It has been noted that k-turns are not always conserved in all phylogenetic groups for rRNAs.[6] The example found in the RNase P holoenzyme crystal structure describes how k-turns are replaced in certain branches of the phylogenetic tree by other structurally equivalent modules. The definition of k-turns and RNA modules in general is based on the network of specific interactions taking place between nucleotides and results in a consensus sequence. Their structural features like the relative positions and orientations of connected helices are usually considered as facilitated by the interaction network. Here, we illustrate how defining a motif based on its structural features offers an original strategy to identify new interaction networks that offset the consensus and result in the same overall relative positions and orientations of helices. Thorough conservation of a module throughout the phylogeny is thus not required to conclude for its existence. On the opposite, the presence of a given motif even in a small subset of sequences may contribute to identify other motifs that are not homologous in sequence but may share common structural features.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

This work was supported by the University of Strasbourg (PhD grant CDOC 2010_117 to M.M.), the Centre National de la Recherche Scientifique and the Université de Strasbourg (to B.M. and E.W.). We thank Pascal Auffinger for helpful discussions and critical reading of the manuscript.

Supplemental Materials

Supplemental materials may be downloaded here:
www.landesbioscience.com/journals/rnabiology/article/19434

References

1. Masquida B, Beckert B, Jossinet F. Exploring RNA structure by integrative molecular modelling. N Biotechnol 2010; 27:170-83; PMID:20206310; http://dx.doi.org/10.1016/j.nbt.2010.02.022
2. Das R, Karanicolas J, Baker D. Atomic accuracy in predicting and designing noncanonical RNA structure. Nat Methods 2010; 7:291-4; PMID:20190761; http://dx.doi.org/10.1038/nmeth.1433
3. Jossinet F, Ludwig TE, Westhof E. Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. Bioinformatics 2010; 26:2057-9; PMID:20562414; http://dx.doi.org/10.1093/bioinformatics/btq321
4. Rother M, Rother K, Puton T, Bujnicki JM. ModeRNA: a tool for comparative modeling of RNA 3D structure. Nucleic Acids Res 2011; 39:4007-22; PMID:21300639; http://dx.doi.org/10.1093/nar/gkq1320
5. Sijenyi F, Saro P, Ouyang Z, Damm-Ganamet K, Wood M, Jiang J, et al. The RNA Folding Problems: Different levels of RNA Structure Prediction. In: Leontis N, Westhof E, eds. RNA 3D Structure Analysis and Prediction: Springer, 2011.
6. Lescoute A, Leontis NB, Massire C, Westhof E. Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. Nucleic Acids Res 2005; 33:2395-409; PMID:15860776; http://dx.doi.org/10.1093/nar/gki535
7. Cruz JA, Westhof E. Sequence-based identification of 3D structural modules in RNA with RMDetect. Nat Methods 2011; 8:513-21; PMID:21552257; http://dx.doi.org/10.1038/nmeth.1603
8. Klein DJ, Schmeing TM, Moore PB, Steitz TA. The kink-turn: a new RNA secondary structure motif. EMBO J 2001; 20:4214-21; PMID:11483524; http://dx.doi.org/10.1093/emboj/20.15.4214
9. Schroeder KT, McPhee SA, Ouellet J, Lilley DM. A structural database for k-turn motifs in RNA. RNA 2010; 16:1463-8; PMID:20562215; http://dx.doi.org/10.1261/rna.2207910
10. Reiter NJ, Osterman A, Torres-Larios A, Swinger KK, Pan T, Mondragón A. Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. Nature 2010; 468:784-9; PMID:21076397; http://dx.doi.org/10.1038/nature09516
11. Rosenblad MA, López MD, Piccinelli P, Samuelsson T. Inventory and analysis of the protein subunits of the ribonucleases P and MRP provides further evidence of homology between the yeast and human enzymes. Nucleic Acids Res 2006; 34:5145-56; PMID:16998185; http://dx.doi.org/10.1093/nar/gkl626
12. Kleywegt GJ, Jones TA. A super position. CCP4/ESF-EACBM Newsletter on protein crystallography 1994; 9-14.
13. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.3r1. 2010.
14. Brown JW. The Ribonuclease P Database. Nucleic Acids Res 1999; 27:314; PMID:9847214; http://dx.doi.org/10.1093/nar/27.1.314
15. Jossinet F, Westhof E. Sequence to Structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. Bioinformatics 2005; 21:3320-1; PMID:15905274; http://dx.doi.org/10.1093/bioinformatics/bti504
16. Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. RNAalifold: improved consensus structure prediction for RNA alignments. BMC Bioinformatics 2008; 9:474; PMID:19014431; http://dx.doi.org/10.1186/1471-2105-9-474
17. Massire C, Jaeger L, Westhof E. Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. J Mol Biol 1998; 279:773-93; PMID:9642060; http://dx.doi.org/10.1006/jmbi.1998.1797
18. Harris ME, Kazantsev AV, Chen JL, Pace NR. Analysis of the tertiary structure of the ribonuclease P ribozyme-substrate complex by site-specific photoaffinity cross-linking. RNA 1997; 3:561-76; PMID:9174092
19. Torres-Larios A, Swinger KK, Krasilnikov AS, Pan T, Mondragón A. Crystal structure of the RNA component of bacterial ribonuclease P. Nature 2005; 437:584-7; PMID:16113684; http://dx.doi.org/10.1038/nature04074

20. Weinberg Z, Perreault J, Meyer MM, Breaker RR. Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis. Nature 2009; 462: 656-9; PMID:19956260; http://dx.doi.org/10.1038/nature08586
21. Vidovic I, Nottrott S, Hartmuth K, Lührmann R, Ficner R. Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. Mol Cell 2000; 6:1331-42; PMID:11163207; http://dx.doi.org/10.1016/S1097-2765(00)00131-3
22. Garst AD, Héroux A, Rambo RP, Batey RT. Crystal structure of the lysine riboswitch regulatory mRNA element. J Biol Chem 2008; 283:22347-51; PMID:18593706; http://dx.doi.org/10.1074/jbc.C800120200
23. Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, et al. Crystal structure of a group I ribozyme domain: principles of RNA packing. Science 1996; 273:1678-85; PMID:8781224; http://dx.doi.org/10.1126/science.273.5282.1678
24. Serganov A, Huang L, Patel DJ. Structural insights into amino acid binding and gene control by a lysine riboswitch. Nature 2008; 455:1263-7; PMID:18784651; http://dx.doi.org/10.1038/nature07326
25. Strobel SA, Adams PL, Stahley MR, Wang J. RNA kink turns to the left and to the right. RNA 2004; 10:1852-4; PMID:15547133; http://dx.doi.org/10.1261/rna.7141504
26. Adams PL, Stahley MR, Kosek AB, Wang J, Strobel SA. Crystal structure of a self-splicing group I intron with both exons. Nature 2004; 430:45-50; PMID:15175762; http://dx.doi.org/10.1038/nature02642
27. Antonioli AH, Cochrane JC, Lipchock SV, Strobel SA. Plasticity of the RNA kink turn structural motif. RNA 2010; 16:762-8; PMID:20145044; http://dx.doi.org/10.1261/rna.1883810
28. Schroeder KT, Daldrop P, Lilley DMJ. RNA tertiary interactions in a riboswitch stabilize the structure of a kink turn. Structure 2011; 19:1233-40; PMID:21893284; http://dx.doi.org/10.1016/j.str.2011.07.003
29. Dibrov SM, Johnston-Cox H, Weng YH, Hermann T. Functional architecture of HCV IRES domain II stabilized by divalent metal ions in the crystal and in solution. Angew Chem Int Ed Engl 2007; 46:226-9; PMID:17131443; http://dx.doi.org/10.1002/anie.200603807
30. Dibrov SM, McLean J, Parsons J, Hermann T. Self-assembling RNA square. Proc Natl Acad Sci U S A 2011; 108:6405-8; PMID:21464284; http://dx.doi.org/10.1073/pnas.1017999108
31. Bernhart SH, Hofacker IL. From consensus structure prediction to RNA gene finding. Brief Funct Genomic Proteomic 2009; 8:461-71; PMID:19833701; http://dx.doi.org/10.1093/bfgp/elp043
32. Lescoute A, Westhof E. Topology of three-way junctions in folded RNAs. RNA 2006; 12:83-93; PMID:16373494; http://dx.doi.org/10.1261/rna.2208106
33. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science 2000; 289:905-20; PMID:10937989; http://dx.doi.org/10.1126/science.289.5481.905
34. Leontis NB, Westhof E. Geometric nomenclature and classification of RNA base pairs. RNA 2001; 7:499-512; PMID:11345429; http://dx.doi.org/10.1017/S1355838201002515

Supplemental Material to:
Mélanie Meyer, Eric Westhof and Benoît Masquida. A structural module in RNase P expands the variety of RNA kinks. RNA Biology 2012; 9(3);
http://dx.doi.org/10.4161/rna.9.3.19434
http://www.landesbioscience.com/journals/rnabiology/article/19434

Supplementary Online Material

Mélanie Meyer, Eric Westhof & Benoît Masquida

Typology of P16-P17 bulge in the secondary structures of the type A RNase P RNA present in the RNase P database (14)

Abbreviations are defined as follows:

- Pk-turn: k-turn-like identified in the Thermotoga maritima P16/P17 junction
- npk-turn: Non *Thermotoga maritima* pk-turn
- k-turn P17: k-turn whith a G-A tandem rooted in P17
- k-turn P16: k-turn whith a G-A tandem rooted in P16
- 3WJ: 3-way junction (the region designated between the parentheses indicates where the additional stem is inserted)
- 4WJ: 4-way junction (the region(s) designated between the parentheses indicate(s) where the additional stem(s) is(are) inserted)
- ΔP16: Indicates secondary structures in which P16 is apparently absent

The number of secondary structures corresponding to each motif in each bacterial branch is indicated between italic bold-face brackets ***[n]***.

Bacteria which names are underlined have been used to build up Figure 4.

For internal loops, the number of base pairs in P16 and P17, and the number of unpaired bases in J16/17 and J17/16 were counted. The numbers are indicated between parentheses for each strain respecting the format below:

$(n_{bpP16},n_{b1},n_{b2},n_{bpP17})$

n_{bpP16}: *number of base pairs in P16*

n_{b1}: *number of bases in J16/17*

n_{b2}: *number of bases in J17/16*

n_{bpP17}: *number of base pairs in P17*

Purple Bacteria & relatives

- *Alpha subdivision*

3WJ (J16/17) *Agrobacterium tumefaciens* , *Caulibacter crescentus*, *Rhodobacter capsulatus*, *Rhodospirillum rubrum, Rickettsia prowazekii, Pond scum #1, #6, #24, #26, #27, #31, Volunteer ESH212C, A18(31), A46(74), Rhodopseudomonas palustris* ***[15]***

npk-turn *Pond scum #2 (7,5,0,4), Salt Marsh A12(14) (6,4,0,4), A48(76) (6,5,1,4), Wolbachia (7,3,0,4), Lake Griffy B #23 (7,4,1,6), #41 (7,4,0,4), W #113 (7,4,0,4)* ***[7]***

- *Beta subdivision*

k-turn P17 *Ralstonia eutrophus (6,5,2,3), Neisseria gonnorhoeae (6,6,3,2), Neisseria meningitidis (6,6,3,2), Comamonas testosteroni (6,5,2,3), Thiobacillus thioparus (6,5,2,3), Volunteer ESH167F (6,6,3,2), ESH183D (6,5,2,3), vB11 (6,5,2,3), Nitrosomonas europeae (6,6,3,2)* ***[9]***

3WJ (J16/17) *Bordetella pertussis, Volunteer MXA1, ESH26-4* ***[3]***

- *Gamma subdivision*

npk-turn, bulge in J16/17 *Acidothiobacillus ferrooxidans (6,3,0,5), Azotobacter vinlandii (6,3,0,5), paspali (6,3,0,5), salinestris (6,3,0,5), chroococcum (6,3,0,5), Buchnera sp. ((6,7,4,5) large asymetric bulge), Azomonas macrocytogenes (6,3,0,5), Haemophilus influenza (6,2,0,5), Klebsiella pneumoniae (6,4,0,5), Escherichia coli (6,4,0,5), Leclercia adecarboxylata (6,4,0,5), Erwinia agglomerulans (6,4,0,5), Pasteurella aerogenes (6,2,0,5), Providencia alcalifaciens (6,4,0,5), Pseudomonas shigelloides (6,3,0,5), Proteus vulgaris (6,4,0,5), Pseudomonas aeruginosa (6,4,0,5), fluorescens (6,2,0,5), putida (6,4,0,5), azotoformans (6,2,0,5), stutzeri (6,3,0,5), Salmonella tiphy (6,4,0,5), Salmonella tiphymurium (same as E. coli (6,4,0,5)), Serratia liquefaciens (6,4,0,5), marescens (6,4,0,5), Xyelella fastidiosa (6,3,0,5), Shewanela putrefasciens (6,4,0,5), Vibrio cholerae (6,3,0,6), Yersinia pestis (6,4,0,5), Volunteer ESH7-9 (6,4,0,5), ESH7-16 (6,4,0,5), ESH17b-7 (6,2,0,5), ESH20b-4 (6,2,0,5), ESH21b-4 (6,2,0,5), Yellowstone Black Hole #145 (6,2,0,5), Yellowstone Pit #126 (6,2,0,5), Pit #131 (6,4,0,5), Volunteer PurpleX (6,4,0,5), Hge2 (6,4,0,5), Salt Marsh A03 (6,4,0,5), A07 (6,2,0,5), A09(11) (6,3,0,5), A14(17) (6,4,0,5), A15(19) (6,3,0,5), A20(33) (6,3,0,5), A32(54) (6,4,0,5), A38(62) (6,3,0,5), A40(65) (6,4,0,5), A47(75) (6,3,0,5), A53(82) (6,3,0,5), Compost Pile A54 (7,3,0,4), B139 (6,4,0,5), B141 (6,4,0,5), A52 (6,3,0,5), B147 (6,4,0,5), B71 (6,4,0,5), B73 (6,4,0,5), B74 (6,4,0,5), B75 (6,4,0,5), B80 (6,4,0,5), B86 (6,4,0,5), Salt Marsh B-B1 (6,4,0,5), Vet School Drainage Water (6,4,0,5), Drainage Fluid (6,2,0,5), Salt Marsh A05 (= E. coli) (6,4,0,5)* ***[65]***

3WJ (J16/17) *Allochromatium vinosum, Lake Griffy W #23, Salt Marsh A02, A19(32), A22(37), A29(46), A33(55), A34(56), A51(79) (Long hairpin inserted), Compost Pile A63* ***[10]***

k-turn P17 *Salt Marsh A36(59) (6,5,2,4)* ***[1]***

- *Delta subdivision*

npk-turn *Desulfovibrio desulfuricans (6,0,3,5)* ***[1]***

k-turn P17 *Desulfovibrio vulgaris (7,4,2,4), Geobacter sulfurreducens (6,5,2,4)* ***[2]***

3WJ (J16/17) *Myxococcus xanthus,* ***[1]***

- *Epsilon subdivision*

ΔP16 *Campylobacter jejuni, Helicobacter pylori 26695 and J99* (P16 is fused on L15. Atypical case corresponding to a different architecture), *Salt Marsh A30(48)* ***[3]***

- *Uncertain affiliation*

npk-turn *Volunteer ESH7-4 (6,3,0,5), Pond Scum #22 (6,3,0,5), #33 (6,3,0,5), Lake Griffy A#2 (6,3,0,5), W#17 (6,4,0,5)* ***[5]***

3WJ (J16/17) *Salt Marsh A17(26)* ***[1]***

k-turn P17 *Volunteer ESH46a-1 (6,6,3,2)* ***[1]***

Gram⁺ Bacteria

- *High G+C subdivision*

[k-turn P17 *(L15 = 3WJ15/16) [Corynebacterium diphteriae (6,6,2,4), Micrococcus luteus (6,7,2,4), Mycobacterium avium (6,6,2,4), bovis (6,6,2,4), leprae (6,6,2,4), tuberculosis (6,6,2,4), Nocardoides sp. (6,6,2,4), Saccharomonospora sp. StrainK180 (7,5,2,4), azurea (7,5,2,4), caesia (7,5,2,4), glauca (7,5,2,4), cyanea K168T (7,5,2,4), viridis (7,5,2,4), Nocardoides jensenii (6,6,2,4), fulvus (6,6,2,4), Luteococcus japonicus (6,6,2,4), Microlunatus phosphovorus (6,6,3,4), Friedmaniella antartica (6,6,2,4), Saccharomonospora sp.K180 (7,5,2,4), Saccharomosnospora azurea (7,5,2,4), caesia (7,5,2,4), glauca (7,5,2,4), cyanea (7,5,2,4), viridis (7,5,2,4), Streptomyces bikiniensis (6,6,2,4), Streptomyces lividans (6,6,2,4), Volunteer ESH210B (6,6,2,4), Salt Marsh A04 (6,6,3,4), A10(12) (6,6,2,4), A27(43) (6,6,3,4), A49(77) (6,6,3,4), A62(96) (6,6,2,4)]* ***[33]***

(L15 = 3WJ16/15) [Nocardoides pyridinolyticus (6,6,2,4)] ***[1]***

(No L15 insertion) [Nocardoides luteus (6,6,2,4), flavus (6,6,2,4), Intrasporagium calvum (6,6,2,4)] ***[3]***

k-turn P16 *(L15 = 3WJ15/16) [Volunteer Rr327 (5,2,5,5)]* ***[1]*****]** ***[38]***

npk-turn *(L15 = 3WJ15/16, bulge J17/16) [Aeromicrobium erythreum (5,0,3,5), fastidiosum (5,0,3,5), Propioniferax innocua (6,0,2,4)]* ***[3]***

3WJ (J16/17) *Nocardoides simplex, plantarum, albus, Sporichthya polymorpha, Terracoccus luteus, Terrabacter tumescens (derived from k-turnP17)* ***[6]***

- *Low G+C subdivision*

npk-turn *Carboxydothermus hydrogenoformans (6,3,0,5), Clostridium acetobutylicum (6,5,0,5), difficult to judge because of the highly non-canonical P16 secondary structure, Eubacterium thermomarinus (5,7,0,6), Salt Marsh A01 (6,4,0,5)* ***[5]***

3WJ (J17/16) *[Clostridium sporogenes]* ***[1]***

Cyanobacteria

npk-turn (Large L15 or 3WJ J16/15 or J15/16, 7 bp P6) *Anabaena (5,2,0,3), Anabaena variabilis (5,2,0,3), Tolypothrix PCC 7601 (5,2,0,3), Dermocarpa PCC 7437 (5,2,0,3), Fisherella UTEX 1829 (5,2,0,3), Nostoc PCC7107 (5,2,0,3), PCC6705 (5,2,0,3), PCC7413 (5,2,0,3), Oscillatoria PCC 7515 (5,2,0,3), Pseudoanabaena PCC 6903 (9,1,0,3), PCC6903 (9,1,0,3), Synechococcus PCC 6301 (5,2,0,3), PCC 7001 (6,3,0,3), PCC 6717 (3+6,2,1,3 See 2D), PCC 7003 (6,4,1,4), PCC 7942 (5,2,0,3), PCC 6803 (5,2,0,2), PCC 6308 (5,2,0,2), MINOS1, Prochlorothrix hollandica (5,2,0,3), Pond Scum #4 (8,4,1,4), Volunteer ESH183A (3+3,5,2,4)* ***[22]***

pk-turn *Prochlorococcus marinus, Prochlorococcus TATL2, TAK9803-2, TATL2-2, PCC9511, NATL2, PAC1B, PAC1A, NATL1* ***[9]***

Bacteroides & relatives

npk-turn *Bacteroides thetaiotaomicron 5482 (7,5,1,4)* ***[1]***

[3WJ (J17/16) *Flavobacterium yabuuchiae, Lake Griffy B#27* ***[2]***

(J16/17) *Salt Marsh A16(21)* ***[1]***] ***[3]***

k-turn P17 *Porphyromonas gingivalis (6,6,2,4), Volunteer ESH30-3 (5,6,3,4), 167^{E} (5,7,2,4), Pond Scum #8 (6,6,2,4), Lake Griffy A#1 (7,5,2,4)* ***[5]***

Undetermined 3WJ or 4WJ *Salt Marsh-A21(35)* ***[1]***

Spirochaetes

pk-turn *Borrelia burgdorferi, Borrelia hermsii,* ***[2]***

npk-turn *Leptospira borgpetersenii (6,3,0,4), Leptospira weilii (6,3,0,4), Treponema pallidum (5,0,3,3), Salt Marsh A25(39) (7,0,3,4)* ***[4]***

Chlamydiae

3WJ (J17/16) *Chlamydia trachomatis N16 (5 strains), pecorum H3 (4 strains), muridarum MoPn (2 strains), suis S45, Chlamydophila pneumoniae (5 strains), psittaci (4 strains), caviae, felis, abortus B577* ***[24]***

npk-turn *Parachlamydia acanthamoebae (5,3,0,6)* ***[1]***

Planctomycetes

3WJ (J16/17) *Pirellula stayeli, Planctomyces maris, Lake Griffy A#6, A#8, W#116, Salt Marsh A58(88)* ***[6]***

k-turn P16 *Lake Griffy W#46 (5,2,6,6)* ***[1]***

4WJ (J17/16) *Salt Marsh A37(61)* ***[1]***

Green sulfur Bacteria

3WJ (J16/17) *Chlorobium limicola, Chlorobium tepidum* ***[2]***

Green non-sulfur Bacteria

k-turn P17 *Chloroflexus aurantiacus (6,6,2,4), Herpetosiphon aurantiacus (6,6,2,4), Salt Marsh Bb-3 (5,7,2,3)* ***[3]***

Unclassified *Thermoleophilum album, Thermomicrobium roseum* ***[2]***

Deinococcus & relatives

k-turn P16 *Deinococcus radiodurans (5,2,6,4), Thermus aquaticus (5,2,7,5), thermophilus (5,2,7,5)* ***[3]***

Thermotogales

pk-turn *Thermotoga maritima (5,2,3,3), Thermotoga neapolitana* ***[2]***

Bacteria of uncertain affiliation

3WJ (J16/17) *Leptospirillum ferrooxidans MK, CF12, Octopus Spring Electric Monk 14b-9, 14b-11, Lake Griffy B#21, Salt Marsh A55(84), B-b2, B-a3, Compost Heap A58, A62, B70, B72* ***[12]***

3WJ (J17/16) *Verrucomicrobium spinosum, Lake Griffy W#18, Salt Marsh A31(53), Compost Heap B69, A49* ***[5]***

k-turn P17 *Lake Griffy B#5 (6,6,3,2), Compost Heap A59 (6,6,2,4)* ***[2]***

pk-turn *Compost Heap A61, Volunteer Rr368* ***[2]***

Non Tm pk-turn *Volunteer ESH20b-1 (6,3,0,4), Volunteer Hge3-5 (5,4,0,4), Salt Marsh A11(13) (6,5,1,4), A13(15) (6,3,0,5), A26(42) (6,3,0,5), A60(90) (6,3,0,5), Compost Heap B144 (6,4,0,5), salt Marsh B-A5 (6,4,0,5)* ***[8]***

4WJ (J17/16-J17/16) *Compost Heap A56, B7776* ***[2]***

Unclassified *Thermodesulfotobacterium commune, Fusobacterium mortiferum (11233979), Octopus Spring Pink Filaments #101, Lake Griffy A#10, B#32, Compost Heap B7677* ***[6]***

Supplementary Table 1 : Residue numbers used in the calculation of the *rmsd* and for superposition between pk-turn and k-turn 15 in (6).

	5' strand 1	3' strand 1	5' strand 2	3' strand 2
pk-turn	235-236	241-242	251-252	258-259
k-turn 15	244-245	249-250	259-260	266-267
k-turn 15 (opp)	259-260	266-267	244-245	249-250

Rmsd pk-turn- k-turn 15 = 1.32 Å

Rmsd k-turn 15- k-turn 15 (opp) = 1.53 Å

Supplementary references

6. *Lescoute, A., Leontis, N.B., Massire, C. and Westhof, E. (2005) Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. Nucleic Acids Res,* ***33****, 2395-2409.*

14. *Brown, J.W. (1999) The Ribonuclease P Database. Nucleic Acids Res,* ***27****, 314.*

3. Discussion

As suggested in our paper and in collaboration with David Lilley's laboratory the commutability of k-turn and pk-turn has been tested (Daldrop et al, 2013). As a result, the pk-turn can functionally substitute for the k-turn in the SAM-I riboswitch from *Thermoanaerobacter tengcongensis* (Schroeder et al, 2011). Vice-versa a k-turn inserted in either orientation between P16 and P17 of the RNAse P RNA from *Thermotoga maritima* can functionally substitute for the pk-turn (Reiter et al, 2010). However in both experiments the activity of recipient catalytic RNA was decreased when the innate RNA motif was replaced (Daldrop et al, 2013). If the outcome of the pk-turn as a RNA 3D motif inserted between two canonical helices is the same as a k-turn on the overall 3D structure, the intrinsic structural features of the pk-turn differ from the k-turn. The pk-turn motif is suggested to be more flexible than the k-turn as FRET analysis show that it does not fold in the context of a simple duplex in presence of magnesium ions while k-turns do so (Liu & Lilley, 2007). This apparent flexibility is compensated by RNAse P folding when it adopts its active conformation as well as by the SAM-I riboswitch. K-turn and pk-turn are not the only motifs found to link and orient P16 to P17 in RNase P. Indeed, 3WJ and 4WJ are as well observed. Taken together these motifs are examples that there are several alternative RNA motifs that can be naturally selected to answer the same overall three-dimensional constraints depending on contextual factors.

Because the pk-turn motif has never been observed before the one revealed by the three-dimensional structure of bacterial RNAse P, any bio-informatic tool could not have detected it. On the contrary sequences expected to fold in a given motif can fold differently under new contextual factors (Antonioli et al, 2010). A well-described example is the reverse k-turn that has only a G to A mutation in a critical G.A pair (Strobel et al, 2004). The unexpected folding of the J5/4 junction in DiLCrz provides another example. Indeed, when this junction as been modeled with respect to sequence alignments, secondary structure comparison with group I introns and mutagenesis data, the junction has been built very close to the group I intron junction which serves in recognition of the GoU base pair at the active site (Einvik et al, 1998). Though because the L5-DL2.1 interaction had not been anticipated, the

peripheral domain DP2-DP2.1 has been left out in the model. The crystal structure of DiLCrz however revealed that L5-DL2.1 forms a kissing complex that constrains the J5/4 junction to fold differently and to stand away from the ribozyme core **(Figure 53)**.

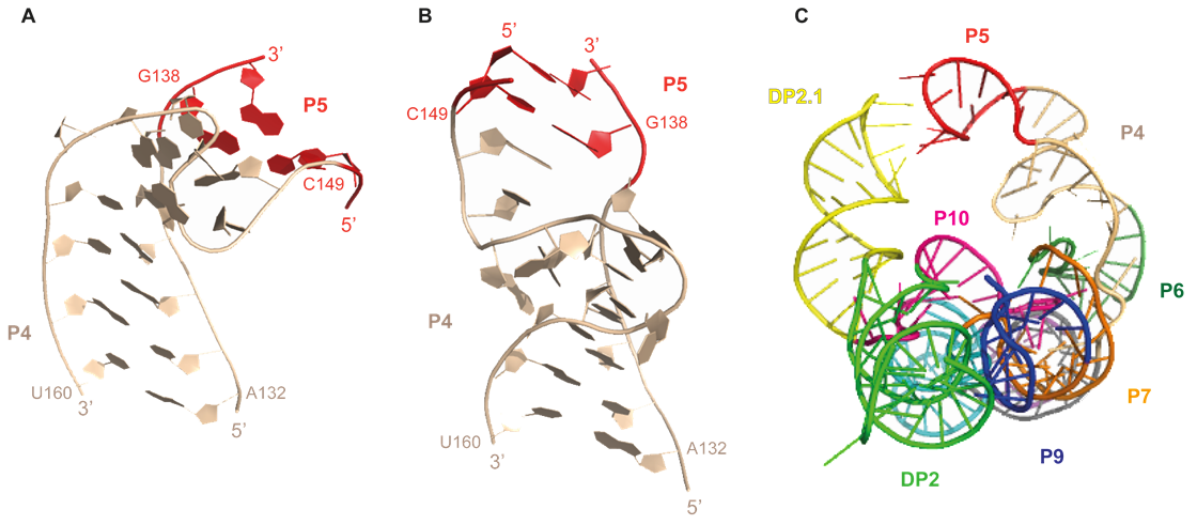


Figure 53: Comparison between the modeled J5/4 junction of DiLCrz (A) (Beckert, 2008) and the J5/4 junction observed in the crystal structure (B). J5/4 in the 3D structure comprising the peripheral domain DP2-DP2.1 is shaped to tailor the L5/DL2.1 kissing complex (C) and is made of two A-minor and base stacking interactions.

In conclusion, it is possible to predict *in silico* in which known 3D motif a sequence could fold and to identify known 3D motifs in new RNA 3D structures. But because nature offers too many subtle variations, there should be many motifs supposed to adopt known architectures that remain hidden in databases or on the opposite, many motifs may be incorrectly predicted (Leontis et al, 2002). This demonstrates that experimental evidences remain critical to avoid being misled for 3D motif analysis and their prediction. Despite the high recurrence of known RNA 3D motifs, many more may be observed with the increasing number of studied functional ncRNA (Cruz & Westhof, 2011a) and of large RNA three-dimensional structures solved (Holbrook, 2005).

CONCLUSION & PERSPECTIVES

IV. CONCLUSION AND PERSPECTIVES

We have solved the 2.5 Å crystal structure of the Lariat-Capping ribozyme that was one of the last ribozyme for which no X-ray structure was available. The structure revealed the crucial role of the DP2-DP2.1 peripheral domain that clamps the ribozyme core together by means of two tertiary interactions, the L5-L2.1 kissing complex and the L9-DP2 GNRA-TR interaction, and the family B DP2-DP2.1-P10 3WJ. The other 3WJ P3-P8-P15 belonging to the C family junction (Lescoute & Westhof, 2006) is part of the double pseudoknot P7-P15-P3 crucial to the structure of the catalytic core. The half S-turn mediated by the GoU *trans* WC base pair atop P15 constitutes a motif related to the sarcin-ricin motif (Correll et al, 1999). Even though the 3D structure of a post-clevage state mimic of LCrz has been solved, the three nucleotides lariat $C_{230}A_{231}U_{232}$ is folded and tightly constrained by specific interactions of each nucleotides of the lariat with nucleotides from the ribozyme core. Importantly most crucial interactions are supported by mutagenesis data (Jabri & Cech, 1998). The structural context thus well influences the 3D motif present in RNAs. One striking example is the pk-turn that we describe in (Meyer et al, 2012), which is structurally related to the k-turn (Lilley, 2012) and reverse k-turn (Antonioli et al, 2010), altough it does not fold in isolation but rather needs its harbouring RNA to force it to adopt the right structure conferring it highly dynamic properties that may be important for the function of the ribozyme where it has been spotted. RNA 3D folding and structure result from the interplay between sequences and contextual constraints.

Lariat-Capping ribozymes catalyze a reaction equivalent to the first step of splicing catalyzed by group II introns however in a structural context related to group I introns (Nielsen et al, 2005). Structural similarities and differences between LCrz and group I introns in addition to its localization in a twintron leads us to assume that LCrz results from the evolutionary drift of a group I intron. The branching reaction that arose from the metamorphosis of a group I intron into LCrz is of great interest as very little is known about the evolutionary routes that gave rise to ribozymes in general and more specifically to RNA lariat formation (Carlomagno et al, 2013) thought to be involved in essential processes such as splicing by group II introns (Lambowitz &

Zimmerly, 2011) or by the spliceosome in eukaryotes (van der Feltz et al, 2012). LCrz 3D structure exemplifies that chemically identical reactions can arise from very different structural context. This work also shows that the structure of RNAs is maintained *in vivo* by constant exposure to qualitative selection pressure and that variation of those may result in the appearance of specific RNAs with new functions. Considering ecological niches populated by microorganisms helps thus embracing the unknown wealth that should exist in nature when they travel horizontally from a niche to another in various media including oceans which contain billions of viruses and microbes per liter.

In a near future, practical objectives will be to solve the crystal structures of the active form of the LCrz using heavy atoms derivatives and also the crystal structures and SAXS shapes of the two others LCrz from *Naegleria sp.* and *Allovahlkampfia sp*. Indeed, if LCrz appeared independently several times in the course of evolution, these structures will help better understand how different structural solutions have been selected to respond to the same problem which is to trigger LCrz catalysis mainly after splicing of the whole intron. The structures will unravel how domains that differ from DiLCrz are organized but also if the L5-L2.1 kissing complex conferring to DiLCrz its unique signet ring shape is preserved or whether elements inherited from the ancestral group I ribozyme inserted within the ribosomal intron have evolved to form different motifs of interaction. Crystallization of NaLCrz and AlLCrz might require more engineering than DiLCrz implying high-throughput probing experiments to refine the secondary structure of their large peripheral domain. In a more distant future experiments working with the full-length twintron should help understanding how splicing by GIR2 and branching by LCrz are coordinated to regulated SSU rRNA production.

In collaboration with partners' laboratory upcoming objectives are to develop a strategy exploiting LCrz as a biotechnological tool. With the 3D crystal structure of LCrz now in hands we can, as we did for HMH ribozyme, optimize new versions of LCrz capable of cleaving unrelated substrate. Stability of the thereby mRNA capped by the three nucleotides lariat can be exploited to increase production of protein in prokaryotes and eukaryotes. LCrz can even be imagined as a therapeutic tool if

engineered to be an efficient *trans* acting ribozyme capable of cleaving undesirable viral RNA or defective innate RNA in cells in a multiple turnover manner.

Moreover mRNA capped by a three nucleotides lariat (native HE mRNA or reporter gene mRNA) can be used to explore all steps of translation from translation initiation to mRNA degradation. Indeed complex systems could be investigated exploiting the fact that pre-mRNA of the HE in GIR2/LCrz twintron is a RNA polymerase I transcript although it is matured and translated as if it wore a RNA polymerase II transcript. One question to be addressed in the native context of this twintron is how transcription, translation and mRNA decay are regulated in eukaryotes when RNA polymerase II is not involved? It would then be possible to imagine a system in which regulation of a transcript level depends on the RNA polymerase I rather than on RNA polymerase II. LCrz could be as well used for gene therapy if inserted upstream the ORF of protein that is not sufficiently expressed because e.g. its corresponding mRNA is degraded before translation since it lacks a cap. Stable capped mRNA can as well be used as RNA sponges e.g. to modulate siRNA level. Encouraging preliminary results show that lariat-capped mRNA can be expressed and tracked in *E. coli*, yeast and mammalian cells.

REFERENCES

REFERENCES

Abergel C (2004) Spectacular improvement of X-ray diffraction through fast desiccation of protein crystals. *Acta crystallographica Section D, Biological crystallography* **60:** 1413-1416

Abramovitz DL, Pyle AM (1997) Remarkable morphological variability of a common RNA folding motif: the GNRA tetraloop-receptor interaction. *Journal of molecular biology* **266:** 493-506

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66:** 213-221

Adams PL, Stahley MR, Kosek AB, Wang J, Strobel SA (2004) Crystal structure of a self-splicing group I intron with both exons. *Nature* **430:** 45-50

Alcorn T, Juers DH (2010) Progress in rational methods of cryoprotection in macromolecular crystallography. *Acta crystallographica Section D, Biological crystallography* **66:** 366-373

Altman S, Smith JD (1971) Tyrosine tRNA precursor molecule polynucleotide sequence. *Nature: New biology* **233:** 35-39

Altman S, Wesolowski D, Guerrier-Takada C, Li Y (2005) RNase P cleaves transient structures in some riboswitches. *Proceedings of the National Academy of Sciences of the United States of America* **102:** 11284-11289

Antonioli AH, Cochrane JC, Lipchock SV, Strobel SA (2010) Plasticity of the RNA kink turn structural motif. *RNA* **16:** 762-768

Apostolico A, Ciriello G, Guerra C, Heitsch CE, Hsiao C, Williams LD (2009) Finding 3D motifs in ribosomal RNA structures. *Nucleic acids research* **37:** e29

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **289:** 905-920

Barrett LW, Fletcher S, Wilton SD (2012) Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cellular and molecular life sciences : CMLS* **69:** 3613-3634

Bass BL, Cech TR (1986) Ribozyme inhibitors: deoxyguanosine and dideoxyguanosine are competitive inhibitors of self-splicing of the Tetrahymena ribosomal ribonucleic acid precursor. *Biochemistry* **25:** 4473-4477

Beckert B, Nielsen H, Einvik C, Johansen SD, Westhof E, Masquida B (2008) Molecular modelling of the GIR1 branching ribozyme gives new insight into evolution of structurally related ribozymes. *The EMBO journal* **27:** 667-678

Been MD, Cech TR (1986) One binding site determines sequence specificity of Tetrahymena pre-rRNA self-splicing, trans-splicing, and RNA enzyme activity. *Cell* **47:** 207-216

Belfort M (2003) Two for the price of one: a bifunctional intron-encoded DNA endonuclease-RNA maturase. *Genes & development* **17:** 2860-2863

Belfort M, Perlman PS (1995) Mechanisms of intron mobility. *The Journal of biological chemistry* **270:** 30237-30240

Berman HM, Olson WK, Beveridge DL, Westbrook J, Gelbin A, Demeny T, Hsieh SH, Srinivasan AR, Schneider B (1992) The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophysical journal* **63:** 751-759

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic acids research* **28:** 235-242

Birgisdottir AB, Johansen S (2005) Site-specific reverse splicing of a HEG-containing group I intron in ribosomal RNA. *Nucleic acids research* **33:** 2042-2051

Birgisdottir AB, Nielsen H, Beckert B, Masquida B, Johansen SD (2011) Intermolecular interaction between a branching ribozyme and associated homing endonuclease mRNA. *Biological chemistry* **392:** 491-499

Blouin S, Lafontaine DA (2007) A loop loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control. *RNA* **13:** 1256-1267

Bonen L (1993) Trans-splicing of pre-mRNA in plants, animals, and protists. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **7:** 40-46

Bourdeau V, Ferbeyre G, Pageau M, Paquin B, Cedergren R (1999) The distribution of RNA motifs in natural sequences. *Nucleic acids research* **27:** 4457-4467

Burger G, Yan Y, Javadi P, Lang BF (2009) Group I-intron trans-splicing and mRNA editing in the mitochondria of placozoan animals. *Trends in genetics : TIG* **25:** 381-386

Buzayan JM, Gerlach WL, Bruening G (1986) Non-enzymatic cleavage and ligation of RNAs complementary to a plant virus satellite RNA. *Nature* **323:** 349-353

Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Muller KM, Pande N, Shang Z, Yu N, Gutell RR (2002) The comparative

RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC bioinformatics* **3:** 2

Canny MD, Jucker FM, Kellogg E, Khvorova A, Jayasena SD, Pardi A (2004) Fast cleavage kinetics of a natural hammerhead ribozyme. *Journal of the American Chemical Society* **126:** 10848-10849

Carlomagno T, Amata I, Codutti L, Falb M, Fohrer J, Masiewicz P, Simon B (2013) Structural principles of RNA catalysis in a 2'-5' lariat-forming ribozyme. *Journal of the American Chemical Society* **135:** 4403-4411

Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA (1996a) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273:** 1678-1685

Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Szewczak AA, Kundrot CE, Cech TR, Doudna JA (1996b) RNA tertiary structure mediation by adenosine platforms. *Science* **273:** 1696-1699

Cech TR (1986) The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell* **44:** 207-210

Cech TR (2009) Crawling out of the RNA world. *Cell* **136:** 599-602

Cech TR (2012) The RNA worlds in context. *Cold Spring Harbor perspectives in biology* **4:** a006742

Cech TR, Damberger SH, Gutell RR (1994) Representation of the secondary and tertiary structure of group I introns. *Nature structural biology* **1:** 273-280

Cech TR, Zaug AJ, Grabowski PJ (1981) In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **27:** 487-496

Chan RT, Robart AR, Rajashankar KR, Pyle AM, Toor N (2012) Crystal structure of a group II intron in the pre-catalytic state. *Nature structural & molecular biology* **19:** 555-557

Chanfreau G, Jacquier A (1996) An RNA conformational change between the two chemical steps of group II self-splicing. *The EMBO journal* **15:** 3466-3476

Chatterjee P, Brady KL, Solem A, Ho Y, Caprara MG (2003) Functionally distinct nucleic acid binding sites for a group I intron encoded RNA maturase/DNA homing endonuclease. *Journal of molecular biology* **329:** 239-251

Chayen NE (1998) Comparative studies of protein crystallization by vapour-diffusion and microbatch techniques. *Acta crystallographica Section D, Biological crystallography* **54:** 8-15

Cochrane JC, Lipchock SV, Smith KD, Strobel SA (2009) Structural and chemical basis for glucosamine 6-phosphate binding and activation of the glmS ribozyme. *Biochemistry* **48:** 3239-3246

Collins CA, Guthrie C (2000) The question remains: is the spliceosome a ribozyme? *Nature structural biology* **7:** 850-854

Copertino DW, Hallick RB (1993) Group II and group III introns of twintrons: potential relationships with nuclear pre-mRNA introns. *Trends in biochemical sciences* **18:** 467-471

Correll CC, Freeborn B, Moore PB, Steitz TA (1997) Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell* **91:** 705-712

Correll CC, Wool IG, Munishkin A (1999) The two faces of the Escherichia coli 23 S rRNA sarcin/ricin domain: the structure at 1.11 A resolution. *Journal of molecular biology* **292:** 275-287

Costa M, Michel F (1999) Tight binding of the 5' exon to domain I of a group II self-splicing intron requires completion of the intron active site. *The EMBO journal* **18:** 1025-1037

Cruz JA, Westhof E (2011a) Identification and annotation of noncoding RNAs in Saccharomycotina. *Comptes rendus biologies* **334:** 671-678

Cruz JA, Westhof E (2011b) Sequence-based identification of 3D structural modules in RNA with RMDetect. *Nature methods* **8:** 513-521

Dai L, Toor N, Olson R, Keeping A, Zimmerly S (2003) Database for mobile group II introns. *Nucleic acids research* **31:** 424-426

Dai L, Zimmerly S (2002) Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. *Nucleic acids research* **30:** 1091-1102

Dai L, Zimmerly S (2003) ORF-less and reverse-transcriptase-encoding group II introns in archaebacteria, with a pattern of homing into related group II intron ORFs. *RNA* **9:** 14-19

Daldrop P, Masquida B, Lilley DM (2013) The functional exchangeability of pk- and k-turns in RNA structure. *RNA biology* **10**

Daniels DL, Michels WJ, Jr., Pyle AM (1996) Two competing pathways for self-splicing by group II introns: a quantitative analysis of in vitro reaction rates and products. *Journal of molecular biology* **256:** 31-49

Davies RW, Waring RB, Ray JA, Brown TA, Scazzocchio C (1982) Making ends meet: a model for RNA splicing in fungal mitochondria. *Nature* **300:** 719-724

Davila Lopez M, Rosenblad MA, Samuelsson T (2009) Conserved and variable domains of RNase MRP RNA. *RNA biology* **6:** 208-220

Decatur WA, Einvik C, Johansen S, Vogt VM (1995) Two group I ribozymes with different functions in a nuclear rDNA intron. *The EMBO journal* **14:** 4558-4568

Decatur WA, Johansen S, Vogt VM (2000) Expression of the Naegleria intron endonuclease is dependent on a functional group I self-cleaving ribozyme. *RNA* **6:** 616-627

Decatur WAE, C.; Johansen, S.; Vogt, M.V (1995) Two group Iribozymeswithdifferentfunctionsina nuclear rDNA intron. *EMBO* **14:** 4558-4568

Doudna JA, Szostak JW (1989) Miniribozymes, small derivatives of the sunY intron, are catalytically active. *Molecular and cellular biology* **9:** 5480-5483

Drager RG, Hallick RB (1993) A complex twintron is excised as four individual introns. *Nucleic acids research* **21:** 2389-2394

Dujon B (1989) Group I introns as mobile genetic elements: facts and mechanistic speculations--a review. *Gene* **82:** 91-114

Dworkin JP, Lazcano A, Miller SL (2003) The roads to and from the RNA world. *Journal of theoretical biology* **222:** 127-134

Einvik C, Decatur WA, Embley TM, Vogt VM, Johansen S (1997) Naegleria nucleolar introns contain two group I ribozymes with different functions in RNA splicing and processing. *RNA* **3:** 710-720

Einvik C, Elde M, Johansen S (1998a) Group I twintrons: genetic elements in myxomycete and schizopyrenid amoeboflagellate ribosomal DNAs. *Journal of biotechnology* **64:** 63-74

Einvik C, Nielsen H, Nour R, Johansen S (2000) Flanking sequences with an essential role in hydrolysis of a self-cleaving group I-like ribozyme. *Nucleic acids research* **28:** 2194-2200

Einvik C, Nielsen H, Westhof E, Michel F, Johansen S (1998b) Group I-like ribozymes with a novel core organization perform obligate sequential hydrolytic cleavages at two processing sites. *RNA* **4:** 530-541

Ellis JC, Brown JW (2009) The RNase P family. *RNA biology* **6:** 362-369

Emblem A, Karlsen BO, Evertsen J, Johansen SD (2011) Mitogenome rearrangement in the cold-water scleractinian coral Lophelia pertusa (Cnidaria, Anthozoa) involves a long-term evolving group I intron. *Molecular phylogenetics and evolution* **61:** 495-503

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60:** 2126-2132

Evans D, Marquez SM, Pace NR (2006) RNase P: interface of the RNA and protein worlds. *Trends in biochemical sciences* **31:** 333-341

Fedorova O, Mitros T, Pyle AM (2003) Domains 2 and 3 interact to form critical elements of the group II intron active site. *Journal of molecular biology* **330:** 197-209

Ferbeyre G, Smith JM, Cedergren R (1998) Schistosome satellite DNA encodes active hammerhead ribozymes. *Molecular and cellular biology* **18:** 3880-3888

Ferre-D'Amare AR, Doudna JA (2000) Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. *Journal of molecular biology* **295:** 541-556

Ferre-D'Amare AR, Scott WG (2010) Small self-cleaving ribozymes. *Cold Spring Harbor perspectives in biology* **2:** a003574

Ferre-D'Amare AR, Zhou K, Doudna JA (1998) A general module for RNA crystallization. *Journal of molecular biology* **279:** 621-631

Frank DN, Pace NR (1998) Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annual review of biochemistry* **67:** 153-180

Gardner PP, Daub J, Tate J, Moore BL, Osuch IH, Griffiths-Jones S, Finn RD, Nawrocki EP, Kolbe DL, Eddy SR, Bateman A (2011) Rfam: Wikipedia, clans and the "decimal" release. *Nucleic acids research* **39:** D141-145

Garriga G, Lambowitz AM (1986) Protein-dependent splicing of a group I intron in ribonucleoprotein particles and soluble fractions. *Cell* **46:** 669-680

Gilbert W (1986) Origin of life: The RNA world. *Nature* **319:** 618

Glanz S, Kuck U (2009) Trans-splicing of organelle introns--a detour to continuous RNAs. *BioEssays : news and reviews in molecular, cellular and developmental biology* **31:** 921-934

Goddard MR, Burt A (1999) Recurrent invasion and extinction of a selfish gene. *Proceedings of the National Academy of Sciences of the United States of America* **96:** 13880-13885

Golden B (2008) Group I Introns : Biochemical and Christallographic Characterization of the Active Site Structure. *Ribozymes and RNA Catalysis***:** 178-200

Golden BL, Kim H, Chase E (2005) Crystal structure of a phage Twort group I ribozyme-product complex. *Nature structural & molecular biology* **12:** 82-89

Gordon PM, Piccirilli JA (2001) Metal ion coordination by the AGC triad in domain 5 contributes to group II intron catalysis. *Nature structural biology* **8:** 893-898

Grabow WW, Zhuang Z, Swank ZN, Shea JE, Jaeger L (2012) The right angle (RA) motif: a prevalent ribosomal RNA structural pattern found in group I introns. *Journal of molecular biology* **424:** 54-67

Green MR (1986) Pre-mRNA splicing. *Annual review of genetics* **20:** 671-708

Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35:** 849-857

Guo F, Cech TR (2002) Evolution of Tetrahymena ribozyme mutants with increased structural stability. *Nature structural biology* **9:** 855-861

Guo F, Gooding AR, Cech TR (2004) Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. *Molecular cell* **16:** 351-362

Gutmann B, Gobert A, Giege P (2012) PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis. *Genes & development* **26:** 1022-1027

Haas ES, Brown JW (1998) Evolutionary variation in bacterial RNase P RNAs. *Nucleic acids research* **26:** 4093-4099

Hafez M, Majer A, Sethuraman J, Rudski SM, Michel F, Hausner G (2013) The mtDNA rns gene landscape in the Ophiostomatales and other fungal taxa: twintrons, introns, and intron-encoded proteins. *Fungal genetics and biology : FG & B* **53:** 71-83

Hamill S, Pyle AM (2006) The receptor for branch-site docking within a group II intron active site. *Molecular cell* **23:** 831-840

Harris ME, Nolan JM, Malhotra A, Brown JW, Harvey SC, Pace NR (1994) Use of photoaffinity crosslinking and molecular modeling to analyze the global architecture of ribonuclease P RNA. *The EMBO journal* **13:** 3953-3963

Hartmann E, Hartmann RK (2003) The enigma of ribonuclease P evolution. *Trends in genetics : TIG* **19:** 561-569

Haugen P, Reeb V, Lutzoni F, Bhattacharya D (2004) The evolution of homing endonuclease genes and group I introns in nuclear rDNA. *Molecular biology and evolution* **21:** 129-140

Haugen P, Simon DM, Bhattacharya D (2005) The natural history of group I introns. *Trends in genetics : TIG* **21:** 111-119

Hedberg A, Johansen SD (2013) Nuclear group I introns in self-splicing and beyond. *Mobile DNA* **4:** 17

Helm M, Brule H, Giege R, Florentz C (1999) More mistakes by T7 RNA polymerase at the 5' ends of in vitro-transcribed RNAs. *RNA* **5:** 618-621

Heras B, Martin JL (2005) Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta crystallographica Section D, Biological crystallography* **61:** 1173-1180

Herschlag D, Cech TR (1990) Catalysis of RNA cleavage by the Tetrahymena thermophila ribozyme. 1. Kinetic description of the reaction of an RNA substrate complementary to the active site. *Biochemistry* **29:** 10159-10171

Holbrook SR (2005) RNA structure: the long and the short of it. *Current opinion in structural biology* **15:** 302-308

Holzmann J, Frank P, Loffler E, Bennett KL, Gerner C, Rossmanith W (2008) RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell* **135:** 462-474

Hougland JL, Kravchuk AV, Herschlag D, Piccirilli JA (2005) Functional identification of catalytic metal ion binding sites within RNA. *PLoS biology* **3:** e277

Hutchins CJ, Rathjen PD, Forster AC, Symons RH (1986) Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic acids research* **14:** 3627-3640

Inoue H, Nojima H, Okayama H (1990) High efficiency transformation of Escherichia coli with plasmids. *Gene* **96:** 23-28

Jabri E, Aigner S, Cech TR (1997) Kinetic and secondary structure analysis of Naegleria andersoni GIR1, a group I ribozyme whose putative biological function is site-specific hydrolysis. *Biochemistry* **36:** 16345-16354

Jabri E, Cech TR (1998) In vitro selection of the Naegleria GIR1 ribozyme identifies three base changes that dramatically improve activity. *RNA* **4:** 1481-1492

Jackson DA, Pombo A, Iborra F (2000) The balance sheet for transcription: an analysis of nuclear RNA metabolism in mammalian cells. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **14:** 242-254

Jaeger L, Westhof E, Leontis NB (2001) TectoRNA: modular assembly units for the construction of RNA nano-objects. *Nucleic acids research* **29:** 455-463

Jarrell KA, Peebles CL, Dietrich RC, Romiti SL, Perlman PS (1988) Group II intron self-splicing. Alternative reaction conditions yield novel products. *The Journal of biological chemistry* **263:** 3432-3439

Johansen S, Elde M, Vader A, Haugen P, Haugli K, Haugli F (1997) In vivo mobility of a group I twintron in nuclear ribosomal DNA of the myxomycete Didymium iridis. *Molecular microbiology* **24:** 737-745

Johansen S, Haugen P (2001) A new nomenclature of group I introns in ribosomal DNA. *RNA* **7:** 935-936

Johansen S, Vogt VM (1994) An intron in the nuclear ribosomal DNA of Didymium iridis codes for a group I ribozyme and a novel ribozyme that cooperate in self-splicing. *Cell* **76:** 725-734

Johnson-Buck AE, McDowell SE, Walter NG (2011) Metal ions: supporting actors in the playbook of small ribozymes. *Metal ions in life sciences* **9:** 175-196

Joshua-Tor L, Hannon GJ (2011) Ancestral roles of small RNAs: an Ago-centric perspective. *Cold Spring Harbor perspectives in biology* **3:** a003772

Jossinet F, Ludwig TE, Westhof E (2010) Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics* **26:** 2057-2059

Jossinet F, Westhof E (2005) Sequence to Structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. *Bioinformatics* **21:** 3320-3321

Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* **66:** 125-132

Kapranov P, Willingham AT, Gingeras TR (2007) Genome-wide transcription and the implications for genomic organization. *Nature reviews Genetics* **8:** 413-423

Kath-Schorr S, Wilson TJ, Li NS, Lu J, Piccirilli JA, Lilley DM (2012) General acid-base catalysis mediated by nucleobases in the hairpin ribozyme. *Journal of the American Chemical Society* **134:** 16717-16724

Kawamura K (2012) Drawbacks of the ancient RNA-based life-like system under primitive earth conditions. *Biochimie* **94:** 1441-1450

Ke A, Zhou K, Ding F, Cate JH, Doudna JA (2004) A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature* **429:** 201-205

Kennell JC, Saville BJ, Mohr S, Kuiper MT, Sabourin JR, Collins RA, Lambowitz AM (1995) The VS catalytic RNA replicates by reverse transcription as a satellite of a retroplasmid. *Genes & development* **9:** 294-303

Khvorova A, Lescoute A, Westhof E, Jayasena SD (2003) Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nature structural biology* **10:** 708-712

Kikovska E, Svard SG, Kirsebom LA (2007) Eukaryotic RNase P RNA mediates cleavage in the absence of protein. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 2062-2067

Kim SH, Cech TR (1987) Three-dimensional model of the active site of the self-splicing rRNA precursor of Tetrahymena. *Proceedings of the National Academy of Sciences of the United States of America* **84:** 8788-8792

Klein DJ, Schmeing TM, Moore PB, Steitz TA (2001) The kink-turn: a new RNA secondary structure motif. *The EMBO journal* **20:** 4214-4221

Knap AK, Wesolowski D, Altman S (1990) Protection from chemical modification of nucleotides in complexes of M1 RNA, the catalytic subunit of RNase P from E coli, and tRNA precursors. *Biochimie* **72:** 779-790

Komine Y, Kitabatake M, Yokogawa T, Nishikawa K, Inokuchi H (1994) A tRNA-like structure is present in 10Sa RNA, a small stable RNA from Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **91:** 9223-9227

Koutmou KS, Zahler NH, Kurz JC, Campbell FE, Harris ME, Fierke CA (2010) Protein-precursor tRNA contact leads to sequence-specific recognition of 5' leaders by bacterial ribonuclease P. *Journal of molecular biology* **396:** 195-208

Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31:** 147-157

Lambowitz AM, Belfort M (1993) Introns as mobile genetic elements. *Annual review of biochemistry* **62:** 587-622

Lambowitz AM, Caprara MG (1999) Group I and II ribozymes as RNPs: Cluesto the Past and Guides and to the Future. *The RNA World***:** 451-485

Lambowitz AM, Caprara MG, Zimmerly S, Perlman PS (1999) Group I and Group II ribozymes as RNPs: clues to the past and guides to the future. *The RNA World*

Lambowitz AM, Zimmerly S (2004) Mobile group II introns. *Annual review of genetics* **38:** 1-35

Lambowitz AM, Zimmerly S (2011) Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harbor perspectives in biology* **3:** a003616

Landthaler M, Shub DA (1999) Unexpected abundance of self-splicing introns in the genome of bacteriophage Twort: introns in multiple genes, a single gene with three introns, and exon skipping by group I ribozymes. *Proceedings of the National Academy of Sciences of the United States of America* **96:** 7005-7010

Lehman N (2010) RNA in evolution. *Wiley interdisciplinary reviews RNA* **1:** 202-213

Leontis NB, Ghosh P, Moore PB (1986) Effect of magnesium ion on the structure of the 5S RNA from Escherichia coli. An imino proton magnetic resonance study of the helix I, IV, and V regions of the molecule. *Biochemistry* **25:** 7386-7392

Leontis NB, Lescoute A, Westhof E (2006) The building blocks and motifs of RNA architecture. *Current opinion in structural biology* **16:** 279-287

Leontis NB, Stombaugh J, Westhof E (2002) Motif prediction in ribosomal RNAs Lessons and prospects for automated motif prediction in homologous RNA molecules. *Biochimie* **84:** 961-973

Leontis NB, Westhof E (2001) Geometric nomenclature and classification of RNA base pairs. *RNA* **7:** 499-512

Leontis NB, Westhof E (2002) The annotation of RNA motifs. *Comparative and functional genomics* **3:** 518-524

Leontis NB, Westhof E (2003) Analysis of RNA motifs. *Current opinion in structural biology* **13:** 300-308

Lescoute A, Westhof E (2006) Topology of three-way junctions in folded RNAs. *RNA* **12:** 83-93

Lilley DM (2011) Catalysis by the nucleolytic ribozymes. *Biochemical Society transactions* **39:** 641-646

Lilley DM (2012) The structure and folding of kink turns in RNA. *Wiley interdisciplinary reviews RNA* **3:** 797-805

Lilley DM, Eckstein F (2008) *Ribozymes and RNA Catalysis*: RSCPublishing.

Lipchock SV, Strobel SA (2008) A relaxed active site after exon ligation by the group I intron. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 5699-5704

Liu J, Lilley DM (2007) The role of specific 2'-hydroxyl groups in the stabilization of the folded conformation of kink-turn RNA. *RNA* **13:** 200-210

Lorber B, Sauter C, Theobald-Dietrich A, Moreno A, Schellenberger P, Robert MC, Capelle B, Sanglier S, Potier N, Giege R (2009) Crystal growth of proteins, nucleic acids, and viruses in gels. *Progress in biophysics and molecular biology* **101:** 13-25

Marcia M, Pyle AM (2012) Visualizing group II intron catalysis through the stages of splicing. *Cell* **151:** 497-507

Marcia M, Somarowthu S, Pyle AM (2013) Now on display: a gallery of group II intron structures at different stages of catalysis. *Mobile DNA* **4:** 14

Martick M, Horan LH, Noller HF, Scott WG (2008) A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA. *Nature* **454:** 899-902

Martick M, Scott WG (2006) Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell* **126:** 309-320

Masquida B, Beckert B, Jossinet F (2010) Exploring RNA structure by integrative molecular modelling. *New biotechnology* **27:** 170-183

Meng Q, Wang Y, Liu XQ (2005) An intron-encoded protein assists RNA splicing of multiple similar introns of different bacterial genes. *The Journal of biological chemistry* **280:** 35085-35088

Meunier B, Tian GL, Macadre C, Slonimski P, Lazowska J (1990) Group II intron stranspose in yeast mitochondria. *Structure, function and biogenesis of energy transfer systems***:** 169-174

Meyer M, Westhof E, Masquida B (2012) A structural module in RNase P expands the variety of RNA kinks. *RNA biology* **9:** 254-260

Michel F, Ferat JL (1995) Structure and activities of group II introns. *Annual review of biochemistry* **64:** 435-461

Michel F, Hanna M, Green R, Bartel DP, Szostak JW (1989a) The guanosine binding site of the Tetrahymena ribozyme. *Nature* **342:** 391-395

Michel F, Jacquier A, Dujon B (1982) Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie* **64:** 867-881

Michel F, Umesono K, Ozeki H (1989b) Comparative and functional anatomy of group II catalytic introns--a review. *Gene* **82:** 5-30

Michel F, Westhof E (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *Journal of molecular biology* **216:** 585-610

Miller MC, Collins K (2002) Telomerase recognizes its template by using an adjacent RNA motif. *Proceedings of the National Academy of Sciences of the United States of America* **99:** 6585-6590

Montange RK, Batey RT (2006) Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature* **441:** 1172-1175

Moreira S, Breton S, Burger G (2012) Unscrambling genetic information at the RNA level. *Wiley interdisciplinary reviews RNA* **3:** 213-228

Murphy FL, Cech TR (1993) An independently folding domain of RNA tertiary structure within the Tetrahymena ribozyme. *Biochemistry* **32:** 5291-5300

Nadimi M, Beaudet D, Forget L, Hijri M, Lang BF (2012) Group I intron-mediated trans-splicing in mitochondria of Gigaspora rosea and a robust phylogenetic affiliation of arbuscular mycorrhizal fungi with Mortierellales. *Molecular biology and evolution* **29:** 2199-2210

Nakamura Y, Kaneko T, Sato S, Ikeuchi M, Katoh H, Sasamoto S, Watanabe A, Iriguchi M, Kawashima K, Kimura T, Kishida Y, Kiyokawa C, Kohara M, Matsumoto M, Matsuno A, Nakazaki N, Shimpo S, Sugimoto M, Takeuchi C, Yamada M, Tabata S (2002) Complete

genome structure of the thermophilic cyanobacterium Thermosynechococcus elongatus BP-1. *DNA research : an international journal for rapid publication of reports on genes and genomes* **9:** 123-130

Nielsen H, Beckert B, Masquida B, Johansen SD (2008) The GIR1 banching ribozyme. In *Ribozymes and RNA Catalysis*, Lilley DM, Eckstein F (eds), pp 229-252. London, Uk: The Royal Society of Chemistry

Nielsen H, Einvik C, Lentz TE, Hedegaard MM, Johansen SD (2009) A conformational switch in the DiGIR1 ribozyme involved in release and folding of the downstream I-DirI mRNA. *RNA* **15:** 958-967

Nielsen H, Fiskaa T, Birgisdottir AB, Haugen P, Einvik C, Johansen S (2003) The ability to form full-length intron RNA circles is a general property of nuclear group I introns. *RNA* **9:** 1464-1475

Nielsen H, Westhof E, Johansen S (2005) An mRNA is capped by a 2', 5' lariat catalyzed by a group I-like ribozyme. *Science* **309:** 1584-1587

Niranjanakumari S, Stams T, Crary SM, Christianson DW, Fierke CA (1998) Protein component of the ribozyme ribonuclease P alters substrate recognition by directly contacting precursor tRNA. *Proceedings of the National Academy of Sciences of the United States of America* **95:** 15212-15217

Nissen P, Hansen J, Ban N, Moore PB, Steitz TA (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science* **289:** 920-930

Ortoleva-Donnelly L, Szewczak AA, Gutell RR, Strobel SA (1998) The chemical basis of adenosine conservation throughout the Tetrahymena ribozyme. *RNA* **4:** 498-519

Oubridge C, Ito N, Evans PR, Teo CH, Nagai K (1994) Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372:** 432-438

Pannucci JA, Haas ES, Hall TA, Harris JK, Brown JW (1999) RNase P RNAs from some Archaea are catalytically active. *Proceedings of the National Academy of Sciences of the United States of America* **96:** 7803-7808

Paukstelis PJ, Chen JH, Chase E, Lambowitz AM, Golden BL (2008) Structure of a tyrosyl-tRNA synthetase splicing factor bound to a group I intron RNA. *Nature* **451:** 94-97

Peck-Miller KA, Altman S (1991) Kinetics of the processing of the precursor to 4.5 S RNA, a naturally occurring substrate for RNase P from Escherichia coli. *Journal of molecular biology* **221:** 1-5

Perrotta AT, Been MD (1991) A pseudoknot-like structure required for efficient self-cleavage of hepatitis delta virus RNA. *Nature* **350:** 434-436

Petrov AI, Zirbel CL, Leontis NB (2011) WebFR3D--a server for finding, aligning and analyzing recurrent RNA 3D motifs. *Nucleic acids research* **39:** W50-55

Pley HW, Flaherty KM, McKay DB (1994) Three-dimensional structure of a hammerhead ribozyme. *Nature* **372:** 68-74

Powner MW, Gerland B, Sutherland JD (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459:** 239-242

Price SR, Ito N, Oubridge C, Avis JM, Nagai K (1995) Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *Journal of molecular biology* **249:** 398-408

Prody GA, Bakos JT, Buzayan JM, Schneider IR, Bruening G (1986) Autolytic processing of dimeric plant virus satellite RNA. *Science* **231:** 1577-1580

Pyle AM (2005) Capping by branching: a new ribozyme makes tiny lariats. *Science* **309:** 1530-1531

Pyle AM (2008) *Group II introns: Catalysts for Splicing, Genomic Change and Evolution. Ribozymes and RNA catalysis.*: RSCPublishing.

Pyle AM, Murphy FL, Cech TR (1992) RNA substrate binding site in the catalytic core of the Tetrahymena ribozyme. *Nature* **358:** 123-128

Qiao F, Cech TR (2008) Triple-helix structure in telomerase RNA contributes to catalysis. *Nature structural & molecular biology* **15:** 634-640

Qin PZ, Pyle AM (1998) The architectural organization and mechanistic function of group II intron structural elements. *Current opinion in structural biology* **8:** 301-308

Reiter NJ, Osterman A, Torres-Larios A, Swinger KK, Pan T, Mondragon A (2010) Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* **468:** 784-789

Rest JS, Mindell DP (2003) Retroids in archaea: phylogeny and lateral origins. *Molecular biology and evolution* **20:** 1134-1142

Robertson HD, Altman S, Smith JD (1972) Purification and properties of a specific Escherichia coli ribonuclease which cleaves a tyrosine transfer ribonucleic acid presursor. *The Journal of biological chemistry* **247:** 5243-5251

Roman J, Rubin MN, Woodson SA (1999) Sequence specificity of in vivo reverse splicing of the Tetrahymena group I intron. *RNA* **5:** 1-13

Roman J, Woodson SA (1995) Reverse splicing of the Tetrahymena IVS: evidence for multiple reaction sites in the 23S rRNA. *RNA* **1:** 478-490

Roman J, Woodson SA (1998) Integration of the Tetrahymena group I intron into bacterial rRNA by reverse splicing in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **95:** 2134-2139

Rupert PB, Ferre-D'Amare AR (2001) Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature* **410:** 780-786

Russell R (2008) RNA misfolding and the action of chaperones. *Frontiers in bioscience : a journal and virtual library* **13:** 1-20

Sambrook J, Fritsch, E.F. (1989) *Molecular cloning: a laboratory manual.* : Cold Spring Harbor Laboratory Press, New York.

Sarver M, Zirbel CL, Stombaugh J, Mokdad A, Leontis NB (2008) FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *Journal of mathematical biology* **56:** 215-252

Saville BJ, Collins RA (1990) A site-specific self-cleavage reaction performed by a novel RNA in Neurospora mitochondria. *Cell* **61:** 685-696

Schroeder KT, Daldrop P, Lilley DM (2011) RNA tertiary interactions in a riboswitch stabilize the structure of a kink turn. *Structure* **19:** 1233-1240

Schroeder KT, Daldrop P, McPhee SA, Lilley DM (2012) Structure and folding of a rare, natural kink turn in RNA with an A*A pair at the 2b*2n position. *RNA* **18:** 1257-1266

Schroeder KT, McPhee SA, Ouellet J, Lilley DM (2010) A structural database for k-turn motifs in RNA. *RNA* **16:** 1463-1468

Schurer H, Lang K, Schuster J, Morl M (2002) A universal method to produce in vitro transcripts with homogeneous 3' ends. *Nucleic acids research* **30:** e56

Scott WG, Finch JT, Klug A (1995) The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell* **81:** 991-1002

Shan S, Kravchuk AV, Piccirilli JA, Herschlag D (2001) Defining the catalytic metal ion interactions in the Tetrahymena ribozyme reaction. *Biochemistry* **40:** 5161-5171

Sharmeen L, Kuo MY, Dinter-Gottlieb G, Taylor J (1988) Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *Journal of virology* **62:** 2674-2679

Sharp PA (1985) On the origin of RNA splicing and introns. *Cell* **42:** 397-400

Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* **64:** 112-122

Siegel RW, Banta AB, Haas ES, Brown JW, Pace NR (1996) Mycoplasma fermentans simplifies our view of the catalytic core of ribonuclease P RNA. *RNA* **2:** 452-462

Sigel RK, Vaidya A, Pyle AM (2000) Metal ion binding sites in a group II intron core. *Nature structural biology* **7:** 1111-1116

Smith D, Burgin AB, Haas ES, Pace NR (1992) Influence of metal ions on the ribonuclease P reaction. Distinguishing substrate binding from catalysis. *The Journal of biological chemistry* **267:** 2429-2436

Stahley MR, Strobel SA (2005) Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science* **309:** 1587-1590

Steinberg SV, Boutorine YI (2007) G-ribo: a new structural motif in ribosomal RNA. *RNA* **13:** 549-554

Steitz TA, Steitz JA (1993) A general two-metal-ion mechanism for catalytic RNA. *Proceedings of the National Academy of Sciences of the United States of America* **90:** 6498-6502

Stoddard CD, Montange RK, Hennelly SP, Rambo RP, Sanbonmatsu KY, Batey RT (2010) Free state conformational sampling of the SAM-I riboswitch aptamer domain. *Structure* **18:** 787-797

Strobel SA, Adams PL, Stahley MR, Wang J (2004) RNA kink turns to the left and to the right. *RNA* **10:** 1852-1854

Strobel SA, Ortoleva-Donnelly L, Ryder SP, Cate JH, Moncoeur E (1998) Complementary sets of noncanonical base pairs mediate RNA helix packing in the group I intron active site. *Nature structural biology* **5:** 60-66

Su LJ, Waldsich C, Pyle AM (2005) An obligate intermediate along the slow folding pathway of a group II intron ribozyme. *Nucleic acids research* **33:** 6674-6687

Szewczak AA, Kosek AB, Piccirilli JA, Strobel SA (2002) Identification of an active site ligand for a group I ribozyme catalytic metal ion. *Biochemistry* **41:** 2516-2525

Szewczak AA, Ortoleva-Donnelly L, Ryder SP, Moncoeur E, Strobel SA (1998) A minor groove RNA triple helix within the catalytic core of a group I intron. *Nature structural biology* **5:** 1037-1042

Tanner M, Cech T (1996) Activity and thermostability of the small self-splicing group I intron in the pre-tRNA(lle) of the purple bacterium Azoarcus. *RNA* **2:** 74-83

Toor N, Hausner G, Zimmerly S (2001) Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* **7:** 1142-1152

Toor N, Keating KS, Fedorova O, Rajashankar K, Wang J, Pyle AM (2010) Tertiary architecture of the Oceanobacillus iheyensis group II intron. *RNA* **16:** 57-69

Toor N, Keating KS, Taylor SD, Pyle AM (2008) Crystal structure of a self-spliced group II intron. *Science* **320:** 77-82

Turner B, Melcher SE, Wilson TJ, Norman DG, Lilley DM (2005) Induced fit of RNA on binding the L7Ae protein to the kink-turn motif. *RNA* **11:** 1192-1200

Uhlenbeck OC (1987) A small catalytic oligoribonucleotide. *Nature* **328:** 596-600

Vader A, Johansen S, Nielsen H (2002) The group I-like ribozyme DiGIR1 mediates alternative processing of pre-rRNA transcripts in Didymium iridis. *European journal of biochemistry / FEBS* **269:** 5804-5812

Vader A, Nielsen H, Johansen S (1999) In vivo expression of the nucleolar group I intron-encoded I-dirI homing endonuclease involves the removal of a spliceosomal intron. *The EMBO journal* **18:** 1003-1013

Vagin A, Teplyakov A (2010) Molecular replacement with MOLREP. *Acta crystallographica Section D, Biological crystallography* **66:** 22-25

Valles Y, Halanych KM, Boore JL (2008) Group II introns break new boundaries: presence in a bilaterian's genome. *PloS one* **3:** e1488

van der Feltz C, Anthony K, Brilot A, Pomeranz Krummel DA (2012) Architecture of the spliceosome. *Biochemistry* **51:** 3321-3333

van der Horst G, Christian A, Inoue T (1991) Reconstitution of a group I intron self-splicing reaction with an activator RNA. *Proceedings of the National Academy of Sciences of the United States of America* **88:** 184-188

van der Veen R, Arnberg AC, Grivell LA (1987a) Self-splicing of a group II intron in yeast mitochondria: dependence on 5' exon sequences. *The EMBO journal* **6:** 1079-1084

van der Veen R, Kwakman JH, Grivell LA (1987b) Mutations at the lariat acceptor site allow self-splicing of a group II intron without lariat formation. *The EMBO journal* **6:** 3827-3831

Vicens Q, Cech TR (2009) A natural ribozyme with 3',5' RNA ligase activity. *Nature chemical biology* **5:** 97-99

Vicens Q, Paukstelis PJ, Westhof E, Lambowitz AM, Cech TR (2008) Toward predicting self-splicing and protein-facilitated splicing of group I introns. *RNA* **14:** 2013-2029

Vidovic I, Nottrott S, Hartmuth K, Luhrmann R, Ficner R (2000) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Molecular cell* **6:** 1331-1342

Wadley LM, Pyle AM (2004) The identification of novel RNA structural motifs using COMPADRES: an automated approach to structural discovery. *Nucleic acids research* **32:** 6650-6659

Walker SC, Engelke DR (2006) Ribonuclease P: the evolution of an ancient RNA enzyme. *Critical reviews in biochemistry and molecular biology* **41:** 77-102

Walter P, Gilmore R, Blobel G (1984) Protein translocation across the endoplasmic reticulum. *Cell* **38:** 5-8

Wank H, SanFilippo J, Singh RN, Matsuura M, Lambowitz AM (1999) A reverse transcriptase/maturase promotes splicing by binding at its own coding segment in a group II intron RNA. *Molecular cell* **4:** 239-250

Waugh DS, Green CJ, Pace NR (1989) The design and catalytic properties of a simplified ribonuclease P RNA. *Science* **244:** 1569-1571

Webb CH, Riccitelli NJ, Ruminski DJ, Luptak A (2009) Widespread occurrence of self-cleaving ribozymes. *Science* **326:** 953

Westhof E, Masquida B, Jossinet F (2011) Predicting and modeling RNA architecture. *Cold Spring Harbor perspectives in biology* **3**

Wikmark OG, Einvik C, De Jonckheere JF, Johansen SD (2006) Short-term sequence evolution and vertical inheritance of the Naegleria twin-ribozyme group I intron. *BMC evolutionary biology* **6:** 39

Wikmark OG, Haugen P, Lundblad EW, Haugli K, Johansen SD (2007) The molecular evolution and structural organization of group I introns at position 1389 in nuclear small subunit rDNA of myxomycetes. *The Journal of eukaryotic microbiology* **54:** 49-56

Wilson DS, Szostak JW (1999) In vitro selection of functional nucleic acids. *Annual review of biochemistry* **68:** 611-647

Wimberly BT, Brodersen DE, Clemons WM, Jr., Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V (2000) Structure of the 30S ribosomal subunit. *Nature* **407:** 327-339

Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428:** 281-286

Wochner A, Attwater J, Coulson A, Holliger P (2011) Ribozyme-catalyzed transcription of an active ribozyme. *Science* **332:** 209-212

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America* **87:** 4576-4579

Woodson SA, Cech TR (1989) Reverse self-splicing of the tetrahymena group I intron: implication for the directionality of splicing and for intron transposition. *Cell* **57:** 335-345

Zaug AJ, Cech TR (1982) The intervening sequence excised from the ribosomal RNA precursor of Tetrahymena contains a 5-terminal guanosine residue not encoded by the DNA. *Nucleic acids research* **10:** 2823-2838

Zaug AJ, Grabowski PJ, Cech TR (1983) Autocatalytic cyclization of an excised intervening sequence RNA is a cleavage-ligation reaction. *Nature* **301:** 578-583

APPENDIX

APPENDIX I

5’EcoRIT7pHMHLCrzHDVBamHI 3’

ECORI_T7P_DiLCrz_TLR1_HDV_BAMHI

**5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTGGGTTGGGAAGTATCATGGCTAA
TCACCATGATGCAATCGGGTTGAACACTTAATTGGGTTAAAACGGTGGGGGACGATCCCGTA
ACATCCGTCCTAACGGCGACAGACTGCACGGCCCTGCCTGGGATATGGAAGAACCGGGGAAA
CTTGGTTCTTCCTAAGTCCTAGGTGTGTTCAATGAACAGTCGTTCCGAAAGGAAGCATCCGG
TATCCCAAGACAATCGCCGGCCATGGTCCCAGCCTCCTCGCTGGCGGCCGGTGGGCAACATT
CCGAGGGGACCGTCCCCTCGGTAATGGCGAATGGGACAAGGTCGACTCTAGAGGATCC 3’**

ECORI_T7P_DiLCrz_TLR2_HDV_BAMHI

**5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTGGGTTGGGAAGTATCATGGCTAA
TCACCATGATGCAATCGGGTTGAACACTTAATTGGGTTAAAACGGTGGGGGACGATCCCGTA
ACATCCGTCCTAACGGCGACAGACTGCACGGCCCTGCCTGGATATGGAAGAACCGGGGAAAC
TTGGTTCTTCCTAAGTCTAGGTGTGTTCAATGAACAGTCGTTCCGAAAGGAAGCATCCGGTA
TCCCAAGACAATCGCCGGCCATGGTCCCAGCCTCCTCGCTGGCGGCCGGTGGGCAACATTCC
GAGGGGACCGTCCCCTCGGTAATGGCGAATGGGACAAGGTCGACTCTAGAGGATCC 3’**

ECORI_T7P_DiLCrz_TLR3_HDV_BAMHI

**5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTGGGTTGGGAAGTATCATGGCTAA
TCACCATGATGCAATCGGGTTGAACACTTAATTGGGTTAAAACGGTGGGGGACGATCCCGTA
ACATCCGTCCTAACGGCGACAGACTGCACGGCCCTGCCTGATATGGAAGAACCGGGGAAACT
TGGTTCTTCCTAAGTCAGGTGTGTTCAATGAACAGTCGTTCCGAAAGGAAGCATCCGGTATC
CCAAGACAATCGCCGGCCATGGTCCCAGCCTCCTCGCTGGCGGCCGGTGGGCAACATTCCGA
GGGGACCGTCCCCTCGGTAATGGCGAATGGGACAAGGTCGACTCTAGAGGATCC 3’**

ECORI_T7P_DiLCrz_TLR4_HDV_BAMHI

**5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTGGGTTGGGAAGTATCATGGCTAA
TCACCATGATGCAATCGGGTTGAACACTTAATTGGGTTAAAACGGTGGGGGACGATCCCGTA
ACATCCGTCCTAACGGCGACAGACTGCACGGCCCTGCCGATATGGAAGAACCGGGGAAACTT
GGTTCTTCCTAAGTTGGTGTGTTCAATGAACAGTCGTTCCGAAAGGAAGCATCCGGTATCCC
AAGACAATCGCCGGCCATGGTCCCAGCCTCCTCGCTGGCGGCCGGTGGGCAACATTCCGAGG
GGACCGTCCCCTCGGTAATGGCGAATGGGACAAGGTCGACTCTAGAGGATCC 3’**

ECORI_T7P_HMH6_DiLCrz_cp_HDV_BAMHI

**5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGGGUGCUUCGGAUGCUGAUGAGUCCG
UGAGGACGAAACAGGAACGCCUGUCCATCCGGTATCCCAAGACAATCTTCGGGTTGGGTTGG
GAAGTATCATGGCTAATCACCATGATGCAATCGGGTTGAACACTTAATTGGGTTAAAACGGT
GGGGGACGATCCCGTAACATCCGTCCTAACGGCGACAGACTGCACGGCCCTGCCTCTTAGGT
GTGTTCAATGAACAGTCGTTCCGAAAGGAAGGCCGGCCATGGTCCCAGCCTCCTCGCTGGCG
GCCGGTGGGCAACATTCCGAGGGGACCGTCCCCTCGGTAATGGCGAATGGGACAAGGTCGAC
TCTAGAGGATCC 3’**

ECORI_T7P_HMH6_DiLCrz_cp_U1A_HDV_BAMHI

**5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGGGUGCUUCGGAUGCUGAUGAGUCCG
UGAGGACGAAACAGGAACGCCUGUCCATCCGGTATCCCAAGACAATCTTCGGGTTGGGTTGG**

GAAGTATCATGGCTAATCACCATGATGCAATCGGGTTGAACACTTAATTGGGTTAAAACGGT
GGGGGACGATCCCGTAACATCCGTCCTAACGGCGACAGACTGCACGGCCCTGCCTACACCAT
TGCACTCCGGTCTGGGTGTGTTCAATGAACAGTCGTTCCGAAAGGAAGGCCGGCCATGGTCC
CAGCCTCCTCGCTGGCGGCCGGTGGGCAACATTCCGAGGGGACCGTCCCCTCGGTAATGGCG
AATGGGACAAGGTCGACTCTAGAGGATCC 3'

ECORI T7P HMH9 DiLCrz_cp HDV BAMHI

5'GACGGCCAGTGAATTCTAATACGACTCACTATAGGGGACGUCCGGAUGCUGAUGAAGAUC
CAUGACAGGAUCGAAACCUCUUGCAAAAGAGGUCCATCCGGTATCCCAAGACAATCTTCGGG
TTGGGTTGGGAAGTATCATGGCTAATCACCATGATGCAATCGGGTTGAACACTTAATTGGGT
TAAAACGGTGGGGGACGATCCCGTAACATCCGTCCTAACGGCGACAGACTGCACGGCCCTGC
CTCTTAGGTGTGTTCAATGAACAGTCGTTCCGAAAGGAAGGCCGGCCATGGTCCCAGCCTCC
TCGCTGGCGGCCGGTGGGCAACATTCCGAGGGGACCGTCCCCTCGGTAATGGCGAATGGGAC
AAGGTCGACTCTAGAGGATCC 3'

ECORI T7P HMH9 DiLCrz_cp_U1A HDV BAMHI

5'GACGGCCAGTGAATTCTAATACGACTCACTATAGGGGACGUCCGGAUGCUGAUGAAGAUC
CAUGACAGGAUCGAAACCUCUUGCAAAAGAGGUCCATCCGGTATCCCAAGACAATCTTCGGG
TTGGGTTGGGAAGTATCATGGCTAATCACCATGATGCAATCGGGTTGAACACTTAATTGGGT
TAAAACGGTGGGGGACGATCCCGTAACATCCGTCCTAACGGCGACAGACTGCACGGCCCTGC
CTACACCATTGCACTCCGGTCTGGGTGTGTTCAATGAACAGTCGTTCCGAAAGGAAGGCCGG
CCATGGTCCCAGCCTCCTCGCTGGCGGCCGGTGGGCAACATTCCGAGGGGACCGTCCCCTCG
GTAATGGCGAATGGGACAAGGTCGACTCTAGAGGATCC 3'

ECORI T7P HMH13 DiLCrz_cp HDV BAMHI

5'GACGGCCAGTGAATTCTAATACGACTCACTATAGGGTAACGGATGCTGACGAGTCTCTGA
GATGAGACGAAACTCTTCGCAAGAAGAGTCCATCCGGTATCCCAAGACAATCTTCGGGTTGG
GTTGGGAAGTATCATGGCTAATCACCATGATGCAATCGGGTTGAACACTTAATTGGGTTAAA
ACGGTGGGGGACGATCCCGTAACATCCGTCCTAACGGCGACAGACTGCACGGCCCTGCCTCT
TAGGTGTGTTCAATGAACAGTCGTTCCGAAAGGAAGGCCGGCCATGGTCCCAGCCTCCTCGC
TGGCGGCCGGTGGGCAACATTCCGAGGGGACCGTCCCCTCGGTAATGGCGAATGGGACAAGG
TCGACTCTAGAGGATCC 3'

ECORI T7P HMH13 DiLCrz_cp_U1A HDV BAMHI

5'GACGGCCAGTGAATTCTAATACGACTCACTATAGGGTAACGGATGCTGACGAGTCTCTGA
GATGAGACGAAACTCTTCGCAAGAAGAGTCCATCCGGTATCCCAAGACAATCTTCGGGTTGG
GTTGGGAAGTATCATGGCTAATCACCATGATGCAATCGGGTTGAACACTTAATTGGGTTAAA
ACGGTGGGGGACGATCCCGTAACATCCGTCCTAACGGCGACAGACTGCACGGCCCTGCCTAC
ACCATTGCACTCCGGTCTGGGTGTGTTCAATGAACAGTCGTTCCGAAAGGAAGGCCGGCCAT
GGTCCCAGCCTCCTCGCTGGCGGCCGGTGGGCAACATTCCGAGGGGACCGTCCCCTCGGTAA
TGGCGAATGGGACAAGGTCGACTCTAGAGGATCC 3'

ECORI T7P HMH13 DiLCrz_cp U203C HDV BAMHI

5'GACGGCCAGTGAATTCTAATACGACTCACTATAGGGTAACGGATGCTGACGAGTCTCTGA
GATGAGACGAAACTCTTCGCAAGAAGAGTCCATCCGGTATCCCAAGACAATCTTCGGGTTGG
GTTGGGAAGTATCATGGCTAATCACCATGATGCAATCGGGTTGAACACTTAATTGGGTTAAA
ACGGTGGGGGACGATCCCGTAACATCCGTCCTAACGGCGACAGACTGCACGGCCCTGCCTCT
TAGGTGTGTCCAATGAACAGTCGTTCCGAAAGGAAGGCCGGCCATGGTCCCAGCCTCCTCGC
TGGCGGCCGGTGGGCAACATTCCGAGGGGACCGTCCCCTCGGTAATGGCGAATGGGACAAGG
TCGACTCTAGAGGATCC 3'

ECORI T7P HMH13 DiLCrz_cp_P2-1 HDV BAMHI

5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGGTAACGGATGCTGACGAGTCTCTGA
GATGAGACGAAACTCTTCGCAAGAAGAGTCCATCCGGTATCCCAAGACATCTTCGGGTGGGT
TGGGAAGTATCATGGCTAATCACCATGATGCAATCGGGTTGAACACTTAATTGGGTTAAAAC
GGTGGGGGGACGATCCCGTAACATCCGTCCTAACGGCGACAGACTGCACGGCCCTGCCTCTTA
GGTGTGTTCAATGAACAGTCGTTCCGAAAGGAAGGCCGGCCATGGTCCCAGCCTCCTCGCTG
GCGGCCGGTGGGCAACATTCCGAGGGGACCGTCCCCTCGGTAATGGCGAATGGGACAAGGTC
GACTCTAGAGGATCC 3’

ECORI T7P HMH13 DiLCrz_cp_P2-1 U203C HDV BAMHI

5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGGTAACGGATGCTGACGAGTCTCTGA
GATGAGACGAAACTCTTCGCAAGAAGAGTCCATCCGGTATCCCAAGACATCTTCGGGTGGGT
TGGGAAGTATCATGGCTAATCACCATGATGCAATCGGGTTGAACACTTAATTGGGTTAAAAC
GGTGGGGGACGATCCCGTAACATCCGTCCTAACGGCGACAGACTGCACGGCCCTGCCTCTTA
GGTGTGTCCAATGAACAGTCGTTCCGAAAGGAAGGCCGGCCATGGTCCCAGCCTCCTCGCTG
GCGGCCGGTGGGCAACATTCCGAGGGGACCGTCCCCTCGGTAATGGCGAATGGGACAAGGTC
GACTCTAGAGGATCC 3’

ECORI T7P DiLCrz actif HDV BAMHI

5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTGGGTTGGGAAGTATCATGGCTAA
TCACCATGATGCAATCGGGTTGAACACTTAATTGGGTTAAAACGGTGGGGGACGATCCCGTA
ACATCCGTCCTAACGGCGACAGACTGCACGGCCCTGCCTCTTAGGTGTGTTCAATGAACAGT
CGTTCCGAAAGGAAGCATCCGGTATCCCAAGACAATCGCCGGCCATGGTCCCAGCCTCCTCG
CTGGCGGCCGGTGGGCAACATTCCGAGGGGACCGTCCCCTCGGTAATGGCGAATGGGACAAG
GTCGACTCTAGAGGATCC 3’

ECORI T7P DiLCrz actif U203C HDV BAMHI

5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTGGGTTGGGAAGTATCATGGCTAA
TCACCATGATGCAATCGGGTTGAACACTTAATTGGGTTAAAACGGTGGGGGACGATCCCGTA
ACATCCGTCCTAACGGCGACAGACTGCACGGCCCTGCCTCTTAGGTGTGTCCAATGAACAGT
CGTTCCGAAAGGAAGCATCCGGTATCCCAAGACAATCGCCGGCCATGGTCCCAGCCTCCTCG
CTGGCGGCCGGTGGGCAACATTCCGAGGGGACCGTCCCCTCGGTAATGGCGAATGGGACAAG
GTCGACTCTAGAGGATCC 3’

ECORI T7P HMH13 DiLCrz_cp_P2TL9_P8TR17 HDV BAMHI

5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGGTAACGGATGCTGACGAGTCTCTGA
GATGAGACGAAACTCTTCGCAAGAAGAGTCCATCCGGTATCCCAAGACAATCTCGGGGAAAC
TTGAGGTTGGGTTGGGAAGTATCATGGCTAATCACCATGATGCAATCGGGTTGAACACTTAA
TTGGGTTAAAACGGTGGGGGACGATCCCGTAACATCCGTCCTAACGGCGACAGACTGCACGG
CCCTGCCTGATATGGTTCGCCTAAGTCAGGTGTGTTCAATGAACAGTCGTTCCGAAAGGAAG
GCCGGCCATGGTCCCAGCCTCCTCGCTGGCGGCCGGTGGGCAACATTCCGAGGGGACCGTCC
CCTCGGTAATGGCGAATGGGACAGGATCCTCTAGAGTCGACCT 3’

ECORI T7P HMH13 DiLCrz_cp_P2TR18_P8TL12 HDV BAMHI

5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGGTAACGGATGCTGACGAGTCTCTGA
GATGAGACGAAACTCTTCGCAAGAAGAGTCCATCCGGTATCCCAAGACAATCTGATATGGCC
TTCGGGCCTAAGTCAGGTTGGGTTGGGAAGTATCATGGCTAATCACCATGATGCAATCGGGT
TGAACACTTAATTGGGTTAAAACGGTGGGGGACGATCCCGTAACATCCGTCCTAACGGCGAC
AGACTGCACGGCCCTGCCTCAACGGGGAAACTTGTTGAGGTGTGTTCAATGAACAGTCGTTC
CGAAAGGAAGGCCGGCCATGGTCCCAGCCTCCTCGCTGGCGGCCGGTGGGCAACATTCCGAG
GGGACCGTCCCCTCGGTAATGGCGAATGGGACAGGATCCTCTAGAGTCGACCT 3’

ECORI T7P HMH13 DiLCrz_cp_P2TR22_P8TL13 HDV BAMHI

5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGGTAACGGATGCTGACGAGTCTCTGA
GATGAGACGAAACTCTTCGCAAGAAGAGTCCATCCGGTATCCCAAGACAATCTGATATGGAA
GTTCCTTCGGGTTCTTCCTAAGTCAGGTTGGGTTGGGAAGTATCATGGCTAATCACCATGAT
GCAATCGGGTTGAACACTTAATTGGGTTAAAACGGTGGGGGACGATCCCGTAACATCCGTCC
TAACGGCGACAGACTGCACGGCCCTGCCGAAAGAACCGGGGAAACTTGGTTCTTTTGGTGTG
TTCAATGAACAGTCGTTCCGAAAGGAAGGCCGGCCATGGTCCCAGCCTCCTCGCTGGCGGCC
GGTGGGCAACATTCCGAGGGGACCGTCCCCTCGGTAATGGCGAATGGGACAGGATCCTCTAG
AGTCGACCT 3’

ECORI T7P HMH13 DiLCrz_cp_P2TL16_P8TR23 HDV BAMHI

5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGGTAACGGATGCTGACGAGTCTCTGA
GATGAGACGAAACTCTTCGCAAGAAGAGTCCATCCGGTATCCCAAGACAATCTGAAAGTTCC
GGGGAAACTTGGTTTCTTTCAGGTTGGGTTGGGAAGTATCATGGCTAATCACCATGATGCAA
TCGGGTTGAACACTTAATTGGGTTAAAACGGTGGGGGACGATCCCGTAACATCCGTCCTAAC
GGCGACAGACTGCACGGCCCTGCCTGTTAGATATGGACCTTCGGGTCCTAAGTCTTTTCAGG
TGTGTTCAATGAACAGTCGTTCCGAAAGGAAGGCCGGCCATGGTCCCAGCCTCCTCGCTGGC
GGCCGGTGGGCAACATTCCGAGGGGACCGTCCCCTCGGTAATGGCGAATGGGACAGGATCCT
CTAGAGTCGACCT 3’

ECORI T7P HMH9 NaLCrz_cp HDV BAMHI

5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTTGCGTTGGGATGCTGATGAAGAT
CCATGACAGGATCGAAACCTCTTGCAAAAGAGGTCCATCCCATACAAAATGGTAACCATAAA
ACAAATGGTACGAACCCTTAGAAATAAGGTTCGGGCTAGTAGTACAACTAAACGTCCTAAAA
CAATAACAGTTGCCTGTTATTGAGGACGTTTTAGTGTGCAATGGGGTTCACACCTTTATTTG
CCAAAACGGGACCTCTGTTGAGGTATAAAATTCTAACGAATTGAATATTCCGTACTAAGGAT
TTAATCCGGAACGTCATGAGACTACACGGCAAACCAATTTGGTGGTGTGAATGGATAGTCCC
TAGTAACCATCTAGGGCCGGCCATGGTCCCAGCCTCCTCGCTGGCGGCCGGTGGGCAACATT
CCGAGGGGACCGTCCCCTCGGTAATGGCGAATGGGACAAGGTCGACTCTAGAGGATCC 3’

ECORI T7P HMH9 NaLCrz_cp U261C HDV BAMHI

5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTTGCGTTGGGATGCTGATGAAGAT
CCATGACAGGATCGAAACCTCTTGCAAAAGAGGTCCATCCCATACAAAATGGTAACCATAAA
ACAAATGGTACGAACCCTTAGAAATAAGGTTCGGGCTAGTAGTACAACTAAACGTCCTAAAA
CAATAACAGTTGCCTGTTATTGAGGACGTTTTAGTGTGCAATGGGGTTCACACCTTTATTTG
CCAAAACGGGACCTCTGTTGAGGTATAAAATTCTAACGAATTGAATATTCCGTACTAAGGAT
TTAATCCGGAACGTCATGAGACTACACGGCAAACCAATTTGGTGGCGTGAATGGATAGTCCC
TAGTAACCATCTAGGGCCGGCCATGGTCCCAGCCTCCTCGCTGGCGGCCGGTGGGCAACATT
CCGAGGGGACCGTCCCCTCGGTAATGGCGAATGGGACAAGGTCGACTCTAGAGGATCC 3’

ECORI T7P HMH9 NaLCrz_cp_court HDV BAMHI

5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTTGCGTTGGGATGCTGATGAAGAT
CCATGACAGGATCGAAACCTCTTGCAAAAGAGGTCCATCCCATACAAAATGGTAACCATAAA
ACAAATGGTACGAACCCTTAGAAATAAGGTTCGGGCTAGTAGTACAACTAAACGTCCTTTGC
AGGACGTTTTAGTGTGCAATGGGGTTCACACCTTTATTTGCCAAAACGGGACCTCTGTTGAG
GTATAAAATTCTAACGAATTGAATATTCCGTACTAAGGATTTAATCCGGAACGTCATGAGAC
TACACGGCAAACCAATTTGGTGGTGTGAATGGATAGTCCCTAGTAACCATCTAGGGCCGGCC
ATGGTCCCAGCCTCCTCGCTGGCGGCCGGTGGGCAACATTCCGAGGGGACCGTCCCCTCGGT
AATGGCGAATGGGACAAGGTCGACTCTAGAGGATCC 3’

ECORI T7P HMH9 NaLCrz_cp_court U261C HDV BAMHI

5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTTGCGTTGGGATGCTGATGAAGAT
CCATGACAGGATCGAAACCTCTTGCAAAAGAGGTCCATCCCATACAAAATGGTAACCATAAA
ACAAATGGTACGAACCCTTAGAAATAAGGTTCGGGCTAGTAGTACAACTAAACGTCCTTTGC
AGGACGTTTTAGTGTGCAATGGGGTTCACACCTTTATTTGCCAAAACGGGACCTCTGTTGAG
GTATAAAATTCTAACGAATTGAATATTCCGTACTAAGGATTTAATCCGGAACGTCATGAGAC
TACACGGCAAACCAATTTGGTGGTGCGAATGGATAGTCCCTAGTAACCATCTAGGGCCGGCC
ATGGTCCCAGCCTCCTCGCTGGCGGCCGGTGGGCAACATTCCGAGGGGACCGTCCCCTCGGT
AATGGCGAATGGGACAAGGTCGACTCTAGAGGATCC 3’

ECORI T7P HMH9 AlLCrz_cp HDV BAMHI

5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTTGCGTTGGGATGCTGATGAAGAT
CCATGACAGGATCGAAACCTCTTGCAAAAGAGGTCCATCCCATACAAAATGGTTAAATATTT
CCCATTCGGACATATAATCCACCGCTTAGGAAGACGCAATATTCTTCTTGGACGCGCAAGTA
CCAAGAGAAAAACAAAGAAAGTTTCGACTTTCTTTTTCTCTTGTGCAATGGGGTTTATGAGT
TAATTAGCCAAAACGGGACCTTAAAAAGGTGTAAGTAACCGTACTAAGTTCGTTAGAACGGA
ATGTCTAGAGACTACACGGCTGAGCGATTTAGCTCTCATAAATGGATAGTCCTCAGTATACC
ATCTGAGGCCGGCCATGGTCCCAGCCTCCTCGCTGGCGGCCGGTGGGCAACATTCCGAGGGG
ACCGTCCCCTCGGTAATGGCGAATGGGACAAGGTCGACTCTAGAGGATCC 3’

ECORI T7P HMH9 AlLCrz_cp U252C HDV BAMHI

5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTTGCGTTGGGATGCTGATGAAGAT
CCATGACAGGATCGAAACCTCTTGCAAAAGAGGTCCATCCCATACAAAATGGTTAAATATTT
CCCATTCGGACATATAATCCACCGCTTAGGAAGACGCAATATTCTTCTTGGACGCGCAAGTA
CCAAGAGAAAAACAAAGAAAGTTTCGACTTTCTTTTTCTCTTGTGCAATGGGGTTTATGAGT
TAATTAGCCAAAACGGGACCTTAAAAAGGTGTAAGTAACCGTACTAAGTTCGTTAGAACGGA
ATGTCTAGAGACTACACGGCTGAGCGATTTAGCTCTCACAAATGGATAGTCCTCAGTATACC
ATCTGAGGCCGGCCATGGTCCCAGCCTCCTCGCTGGCGGCCGGTGGGCAACATTCCGAGGGG
ACCGTCCCCTCGGTAATGGCGAATGGGACAAGGTCGACTCTAGAGGATCC 3’

ECORI T7P HMH9 AlLCrz_cp_court HDV BAMHI

5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTTGCGTTGGGATGCTGATGAAGAT
CCATGACAGGATCGAAACCTCTTGCAAAAGAGGTCCATCCCATACAAAATGGTTAAATATTT
CCCATTCGGACATATAATCCACCGCTTAGGAAGACGCAATATTCTTCTTGGACGCGCAAGTA
CCAAGAGAAAAACAAATTCGTTTTTCTCTTGTGCAATGGGGTTTATGAGTTAATTAGCCAAA
ACGGGACCTTAAAAAGGTGTAAGTAACCGTACTAAGTTCGTTAGAACGGAATGTCTAGAGAC
TACACGGCTGAGCGATTTAGCTCTCATAAATGGATAGTCCTCAGTATACCATCTGAGGCCGG
CCATGGTCCCAGCCTCCTCGCTGGCGGCCGGTGGGCAACATTCCGAGGGGACCGTCCCCTCG
GTAATGGCGAATGGGACAAGGTCGACTCTAGAGGATCC 3’

ECORI T7P HMH9 AlLCrz_cp_court U252C HDV BAMHI

5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTTGCGTTGGGATGCTGATGAAGAT
CCATGACAGGATCGAAACCTCTTGCAAAAGAGGTCCATCCCATACAAAATGGTTAAATATTT
CCCATTCGGACATATAATCCACCGCTTAGGAAGACGCAATATTCTTCTTGGACGCGCAAGTA
CCAAGAGAAAAACAAATTCGTTTTTCTCTTGTGCAATGGGGTTTATGAGTTAATTAGCCAAA
ACGGGACCTTAAAAAGGTGTAAGTAACCGTACTAAGTTCGTTAGAACGGAATGTCTAGAGAC
TACACGGCTGAGCGATTTAGCTCTCACAAATGGATAGTCCTCAGTATACCATCTGAGGCCGG
CCATGGTCCCAGCCTCCTCGCTGGCGGCCGGTGGGCAACATTCCGAGGGGACCGTCCCCTCG
GTAATGGCGAATGGGACAAGGTCGACTCTAGAGGATCC 3’

APPENDIX II

ECORI T7P DiLCrz_TLR1 HDV BAMHI

Matrix PCR1 : DiGR1M103 (Substrate in trans and C2 U1A loop in P8)

PCR1
TLRgT7ECORI
5'GACGGCCAGTGAATTCTAATACGACTCACTATAGG 3'
TLR1-2
5'CTTAGGAAGAACCAAGTTTCCCCGGTTCTTCCATATCCCAGGCAGGGCCGTGCA 3'

PCR2
TLRgT7ECORI
5'GACGGCCAGTGAATTCTAATACGACTCACTATAGG 3'
TLR1-1HdVGIR1TLR
5'ATGGCCGGCTCGGAACGACTGTTCATTGAACACACCTAGGACTTAGGAAGAACCAAGTT 3'

PCR3
TLRgT7ECORI
5'GACGGCCAGTGAATTCTAATACGACTCACTATAGG 3'
TLRgHDVGIR1
5'ATGTTGCCCACCGGCCGCCAGCGAGGAGGCTGGGACCATGGCCGGCTCGGAA 3'

PCR4
TLRgT7ECORI
5'GACGGCCAGTGAATTCTAATACGACTCACTATAGG 3'
TLRgBAMHIHDV
5'AGGTCGACTCTAGAGGATCCTGTCCCATTCGCCATTACCGAGGGGACGGTCCCCTCGGAA TGTTGCCCACCGGC 3'

ECORI T7P DiLCrz_TLR2 HDV BAMHI

Matrix PCR1 : DiGR1M103 (Substrate in trans and C2 U1A loop in P8)

PCR1
TLRgT7ECORI
5'GACGGCCAGTGAATTCTAATACGACTCACTATAGG 3'
TLR2-2
5'ACTTAGGAAGAACCAAGTTTCCCCGGTTCTTCCATATCCAGGCAGGGCCGTGCA 3'

PCR2
TLRgT7ECORI
5'GACGGCCAGTGAATTCTAATACGACTCACTATAGG 3'
TLR2-1HdVGIR1TLR
5'ATGGCCGGCTCGGAACGACTGTTCATTGAACACACCTAGACTTAGGAAGAACCAAGT 3'

PCR3
TLRgT7ECORI
5'GACGGCCAGTGAATTCTAATACGACTCACTATAGG 3'
TLRgHDVGIR1
5'ATGTTGCCCACCGGCCGCCAGCGAGGAGGCTGGGACCATGGCCGGCTCGGAA 3'

PCR4
TLRgT7ECORI
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGG 3’
TLRgBAMHIHDV
5’AGGTCGACTCTAGAGGATCCTGTCCCATTCGCCATTACCGAGGGGACGGTCCCCTCGGAA
TGTTGCCCACCGGC 3’

ECORI T7P DiLCrz_TLR3 HDV BAMHI

Matrix PCR1 : DiGR1M103 (Substrate in trans and C2 U1A loop in P8)

PCR1
TLRgT7ECORI
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGG 3’
TLR3-2
5’CTTAGGAAGAACCAAGTTTCCCCGGTTCTTCCATATCAGGCAGGGCCGTGCA 3’

PCR2
TLRgT7ECORI
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGG 3’
TLR3-1HdVGIR1TLR
5’ATGGCCGGCTCGGAACGACTGTTCATTGAACACACCTGACTTAGGAAGAACCAAGTT 3’

PCR3
TLRgT7ECORI
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGG 3’
TLRgHDVGIR1
5’ATGTTGCCCACCGGCCGCCAGCGAGGAGGCTGGGACCATGGCCGGCTCGGAA 3’

PCR4
TLRgT7ECORI
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGG 3’
TLRgBAMHIHDV
5’AGGTCGACTCTAGAGGATCCTGTCCCATTCGCCATTACCGAGGGGACGGTCCCCTCGGAA
TGTTGCCCACCGGC 3’

ECORI T7P DiLCrz_TLR4 HDV BAMHI

Matrix PCR1 : DiGR1M103 (Substrate in trans and C2 U1A loop in P8)

PCR1
TLRgT7ECORI
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGG 3’
TLR4-2
5’CTTAGGAAGAACCAAGTTTCCCCGGTTCTTCCATATCGGCAGGGCCGTGCA 3’

PCR2
TLRgT7ECORI
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGG 3’
TLR4-1HdVGIR1TLR
5’ATGGCCGGCTCGGAACGACTGTTCATTGAACACACCGACTTAGGAAGAACCAAGTT 3’

PCR3
TLRgT7ECORI

5’GACGGCCAGTGAATTCTAATACGACTCACTATAGG 3’
TLRgHDVGIR1
5’ATGTTGCCCACCGGCCGCCAGCGAGGAGGCTGGGACCATGGCCGGCTCGGAA 3’

PCR4
TLRgT7ECORI
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGG 3’
TLRgBAMHIHDV
5’AGGTCGACTCTAGAGGATCCTGTCCCATTCGCCATTACCGAGGGGACGGTCCCCTCGGAA
TGTTGCCCACCGGC 3’

ECORI T7P HMH6 DiLCrz_pc HDV BAMHI

Matrix PCR1 : DiGR1_WT_5-7 (Cp)

PCR1
HMHgHDVBAMHI
5’AGGTCGACTCTAGAGGATCCTGTCCCATTCGCC 3’
HMH6-1
5’CTGATGAGTCCGTGAGGACGAAACAGGGCAACCTGTCCATCCGGTATCCC 3’

PCR2
HMHgHDVBAMHI
5’AGGTCGACTCTAGAGGATCCTGTCCCATTCGCC 3’
HMH6-2T7ECORI
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGGGTGCTTCGGATGCTGATGAGTCCG
TGA 3’

ECORI T7P HMH6 DiLCrz_pc_U1A HDV BAMHI

Matrix PCR1 : DiGIR1_M_2-5-1 (Cp and C2 U1A loop in P8)

PCR1
HMHgHDVBAMHI
5’AGGTCGACTCTAGAGGATCCTGTCCCATTCGCC 3’
HMH6-1
5’CTGATGAGTCCGTGAGGACGAAACAGGGCAACCTGTCCATCCGGTATCCC 3’

PCR2
HMHgHDVBAMHI
5’AGGTCGACTCTAGAGGATCCTGTCCCATTCGCC 3’
HMH6-2T7ECORI
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGGGTGCTTCGGATGCTGATGAGTCCG
TGA 3’

ECORI T7P HMH9 DiLCrz_pc HDV BAMHI

Matrix PCR1 : DiGR1_WT_5-7 (Cp)

PCR1
HMHgHDVBAMHI
5’AGGTCGACTCTAGAGGATCCTGTCCCATTCGCC 3’
HMH9-1
5’ATCCATGACAGGATCGAAACCTCTTGCAAAAGAGGTCCATCCGGTATCCC 3’

PCR2
HMHgHDVBAMHI

5’AGGTCGACTCTAGAGGATCCTGTCCCATTCGCC 3’
HMH9-2T7ECORI
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGGGACGTCCGGATGCTGATGAAGATC
CATGACAGGATC 3’

ECORI T7P HMH9 DiLCrz_pc_U1A HDV BAMHI

Matrix PCR1 : DiGIR1_M_2-5-1 (Cp and C2 U1A loop in P8)

PCR1
HMHgHDVBAMHI
5’AGGTCGACTCTAGAGGATCCTGTCCCATTCGCC 3’
HMH9-1
5’ATCCATGACAGGATCGAAACCTCTTGCAAAAGAGGTCCATCCGGTATCCC 3’

PCR2
HMHgHDVBAMHI
5’AGGTCGACTCTAGAGGATCCTGTCCCATTCGCC 3’
HMH9-2T7ECORI
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGGGACGTCCGGATGCTGATGAAGATC
CATGACAGGATC 3’

ECORI T7P HMH13 DiLCrz_pc HDV BAMHI

Matrix PCR1 : DiGR1_WT_5-7 (Cp)

PCR1
HMHgHDVBAMHI
5’AGGTCGACTCTAGAGGATCCTGTCCCATTCGCC 3’
HMH13-1
5’TCTCTGAGATGAGACGAAACTCTTCGCAAGAAGAGTCCATCCGGTATCCC 3’

PCR2
HMHgHDVBAMHI
5’AGGTCGACTCTAGAGGATCCTGTCCCATTCGCC 3’
HMH13-2T7ECORI
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGGTAACGGATGCTGACGAGTCTCTGA
GATGAGAC 3’

ECORI T7P HMH13 DiLCrz_pc U203C HDV BAMHI

Matrix PCR1 : DiGR1_WT_5-7 (Cp)

PCR1
HMHgHDVBAMHI
5’AGGTCGACTCTAGAGGATCCTGTCCCATTCGCC 3’
HMH13-1
5’TCTCTGAGATGAGACGAAACTCTTCGCAAGAAGAGTCCATCCGGTATCCC 3’

PCR2
HMHgHDVBAMHI
5’AGGTCGACTCTAGAGGATCCTGTCCCATTCGCC 3’
HMH13-2T7ECORI
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGGTAACGGATGCTGACGAGTCTCTGA
GATGAGAC 3’

Note : the U203C mutation in this construct has been introduced randomly during PCRs.

ECORI T7P HMH13 DiLCrz_pc_U1A HDV BAMHI

Matrix PCR1 : DiGIR1_M_2-5-1 (Cp and C2 U1A loop in P8)

PCR1
HMHgHDVBAMHI
5'AGGTCGACTCTAGAGGATCCTGTCCCATTCGCC 3'
HMH13-1
5'TCTCTGAGATGAGACGAAACTCTTCGCAAGAAGAGTCCATCCGGTATCCC 3'

PCR2
HMHgHDVBAMHI
5'AGGTCGACTCTAGAGGATCCTGTCCCATTCGCC 3'
HMH13-2T7ECORI
5'GACGGCCAGTGAATTCTAATACGACTCACTATAGGGTAACGGATGCTGACGAGTCTCTGA
GATGAGAC 3'

ECORI T7P HMH13 DiLCrz_pc_P2-1 HDV BAMHI

Matrix PCR1 : ECORI T7P HMH13 DiLCrz_pc HDV BAMHI

PCR1
FW1DiGIR1pcP2-1
5'CATCCGGTATCCCAAGACATCTTCGGGTGGGTTGGGAAGTATCATGGCTAATCACC 3'
RVgDiGIR1pcP2-1
5'GGATCCTCTAGAGTCGA 3'

PCR2
FW2DiGIR1pcP2-1
5'CTGACGAGTCTCTGAGATGAGACGAAACTCTTCGCAAGAAGAGTCCATCCGGTATCCCAA
GA 3'
RVgDiGIR1pcP2-1
5'GGATCCTCTAGAGTCGA 3'

PCR3
FW3DiGIR1pcP2-1
5'GACGGCCAGTGAATTCTAATACGACTCACTATAGGGTAACGGATGCTGACGAGTCTCTGA
GA 3'
RVgDiGIR1pcP2-1
5'GGATCCTCTAGAGTCGA 3'

ECORI T7P HMH13 DiLCrz_pc_P2-1 U203C HDV BAMHI

Matrix PCR1 : ECORI T7P HMH13 DiLCrz_pc U203C HDV BAMHI

PCR1
FW1DiGIR1pcP2-1
5'CATCCGGTATCCCAAGACATCTTCGGGTGGGTTGGGAAGTATCATGGCTAATCACC 3'
RVgDiGIR1pcP2-1
5'GGATCCTCTAGAGTCGA 3'

PCR2
FW2DiGIR1pcP2-1

5’CTGACGAGTCTCTGAGATGAGACGAAACTCTTCGCAAGAAGAGTCCATCCGGTATCCCAA
GA 3’
RVgDiGIR1pcP2-1
5’GGATCCTCTAGAGTCGA 3’

PCR3
FW3DiGIR1pcP2-1
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGGTAACGGATGCTGACGAGTCTCTGA
GA 3’
RVgDiGIR1pcP2-1
5’GGATCCTCTAGAGTCGA 3’

ECORI T7P DiLCrz actif HDV BAMHI

Matrix PCR1 : ECORI T7P HMH13 DiLCrz_pc HDV BAMHI

PCR1
FWgDiGIR1
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTGGGTTGGGAAGTA 3’
RV1DiGIR1
5’GGACCATGGCCGGCGATTGTCTTGGGATACCGGATGCTTCCTTTCGGAACGACT 3’

PCR2
FWgDiGIR1
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTGGGTTGGGAAGTA 3’
RV2DiGIR1
5’GTCCCCTCGGAATGTTGCCCACCGGCCGCCAGCGAGGAGGCTGGGACCATGGCCGGC 3’

PCR3
FWgDiGIR1
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTGGGTTGGGAAGTA 3’
RV3DiGIR1
5’GGATCCTCTAGAGTCGACCTTGTCCCATTCGCCATTACCGAGGGGACGGTCCCCTCGGAA
TGTT 3’

ECORI T7P HMH13 DiLCrz_pc_P2TL9_P8TR17 HDV BAMHI

Matrix PCR1 : ECORI T7P HMH13 DiLCrz_pc U203C HDV BAMHI

PCR1
FWP2TL9-1
5’ GACAATCTCGGGGAAACTTGAGGTTGGGTTGGGAAGTATCA 3’
RVP8TR17-1
5’ TGACTTAGGCGAACCATATCAGGCAGGGCCGTGCAGTC 3’

PCR2
FWP2TL9-2
5’ TCCGGTATCCCAAGACAATCTCGGGGAAACTTGA 3’
RVP8TR17-2
5’ CGACTGTTCATTGAACACACCTGACTTAGGCGAACCATATCA 3’

PCR3
FWP2-3
5’GCTGACGAGTCTCTGAGATGAGACGAAACTCTTCGCAAGAAGAGTCCATCCGGTATCCCA
AGACAATC 3’

RVP8-3
5’AATGTTGCCCACCGGCCGCCAGCGAGGAGGCTGGGACCATGGCCGGCCTTCCTTTCGGAA
CGACTGTTCATTGAACACACC 3’

PCR4
FWP2-4
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGGTAACGGATGCTGACGAGTCTCTGA
GATG 3’
RVP8-4
5’AGGTCGACTCTAGAGGATCCTGTCCCATTCGCCATTACCGAGGGGACGGTCCCCTCGGAA
TGTTGCCCACCGGCCG 3’

ECORI T7P HMH13 DiLCrz_pc_P2TR18_P8TL12 HDV BAMHI

Matrix PCR1 : ECORI T7P HMH13 DiLCrz_pc U203C HDV BAMHI

PCR1
FWP2TR18-1
5’ TATGGCCTTCGGGCCTAAGTCAGGTTGGGTTGGGAAGTATCA 3’
RVP8TL12-1
5’ CTCAACAAGTTTCCCCGTTGAGGCAGGGCCGTGCAGTC 3’

PCR2
FWP2TR18-2
5’ TCCGGTATCCCAAGACAATCTGATATGGCCTTCGGGCCTAAGT 3’
RVP8TL12-2
5’ CGACTGTTCATTGAACACACCTCAACAAGTTTCCCCGTTGA 3’

PCR3
FWP2-3
5’GCTGACGAGTCTCTGAGATGAGACGAAACTCTTCGCAAGAAGAGTCCATCCGGTATCCCA
AGACAATC 3’
RVP8-3
5’AATGTTGCCCACCGGCCGCCAGCGAGGAGGCTGGGACCATGGCCGGCCTTCCTTTCGGAA
CGACTGTTCATTGAACACACC 3’

PCR4
FWP2-4
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGGTAACGGATGCTGACGAGTCTCTGA
GATG 3’
RVP8-4
5’AGGTCGACTCTAGAGGATCCTGTCCCATTCGCCATTACCGAGGGGACGGTCCCCTCGGAA
TGTTGCCCACCGGCCG 3’

ECORI T7P HMH13 DiLCrz_pc_P2TR22_P8TL13 HDV BAMHI

Matrix PCR1 : ECORI T7P HMH13 DiLCrz_pc U203C HDV BAMHI

PCR1
FWP2TR22-1
5’ AAGTTCCTTCGGGTTCTTCCTAAGTCAGGTTGGGTTGGGAAGTATCA3’
RVP8TL13-1
5’ AAAAGAACCAAGTTTCCCCGGTTCTTTCGGCAGGGCCGTGCAGTC 3’

PCR2
FWP2TR22-2
5’TCCGGTATCCCAAGACAATCTGATATGGAAGTTCCTTCGGGTTCTTCC3’

RVP8TL13-2
5’ CGACTGTTCATTGAACACACCAAAAGAACCAAGTTTCCCCGG 3’

PCR3
FWP2-3
5’GCTGACGAGTCTCTGAGATGAGACGAAACTCTTCGCAAGAAGAGTCCATCCGGTATCCCA
AGACAATC 3’
RVP8-3
5’AATGTTGCCCACCGGCCGCCAGCGAGGAGGCTGGGACCATGGCCGGCCTTCCTTTCGGAA
CGACTGTTCATTGAACACACC 3’

PCR4
FWP2-4
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGGTAACGGATGCTGACGAGTCTCTGA
GATG 3’
RVP8-4
5’AGGTCGACTCTAGAGGATCCTGTCCCATTCGCCATTACCGAGGGGACGGTCCCCTCGGAA
TGTTGCCCACCGGCCG 3’

ECORI T7P HMH13 DiLCrz_pc_P2TL16_P8TR23 HDV BAMHI

Matrix PCR1 : ECORI T7P HMH13 DiLCrz_pc U203C HDV BAMHI

PCR1
FWP2TL16-1
5’ TTCCGGGGAAACTTGGTTTCTTTCAGGTTGGGTTGGGAAGTATCA 3’
RVP8TR23-1
5’ GACTTAGGACCCGAAGGTCCATATCTAACAGGCAGGGCCGTGCAGTC 3’

PCR2
FWP2TL16-2
5’ TCCGGTATCCCAAGACAATCTGAAAGTTCCGGGGAAACTTGGTTTCT 3’
RVP8TR23-2
5’ CGACTGTTCATTGAACACACCTGAAAAGACTTAGGACCCGAAGGTC 3’

PCR3
FWP2-3
5’GCTGACGAGTCTCTGAGATGAGACGAAACTCTTCGCAAGAAGAGTCCATCCGGTATCCCA
AGACAATC 3’
RVP8-3
5’AATGTTGCCCACCGGCCGCCAGCGAGGAGGCTGGGACCATGGCCGGCCTTCCTTTCGGAA
CGACTGTTCATTGAACACACC 3’

PCR4
FWP2-4
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGGTAACGGATGCTGACGAGTCTCTGA
GATG 3’
RVP8-4
5’AGGTCGACTCTAGAGGATCCTGTCCCATTCGCCATTACCGAGGGGACGGTCCCCTCGGAA
TGTTGCCCACCGGCCG 3’

ECORI T7P HMH9 NaLCrz_pc HDV BAMHI

Matrix PCR1 : pUC18 NaGIR1 pr

PCR1

FW1NaGIR1pc
5’GTACAACTAAACGTCCTAAAACAATAACAGTTGCCTGTTATTGAGGACGTTTTAGTGTG
3’
RV1NaGIR1pc
5’CCCCTCGGAATGTTGCCCACCGGCCGCCAGCGAGGAGGCTGGGACCATGGCCGGCCCTAG
ATGGTTACTAGG 3’

PCR2
FW2NaGIR1pc
5’GGTAACCATAAAACAAATGGTACGAACCCTTAGAAATAAGGTTCGGGCTAGTAGTACAAC
TAAACGTCCTA 3’
RV1NaGIR1pc
5’CCCCTCGGAATGTTGCCCACCGGCCGCCAGCGAGGAGGCTGGGACCATGGCCGGCCCTAG
ATGGTTACTAGG 3’

PCR3
FW3NaGIR1pc
5’GATCCATGACAGGATCGAAACCTCTTGCAAAAGAGGTCCATCCCATACAAAATGGTAACC
ATAAAACAAATG 3’
RV2NaAsGIR1pc
5’GGATCCTCTAGAGTCGACCTTGTCCCATTCGCCATTACCGAGGGGACGGTCCCCTCGGAA
TGTTGC 3’

PCR4
FW4NaAsGIR1pc
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTTGCGTTGGGATGCTGATGAAGAT
CCATGACAGGATCG 3’
RV2NaAsGIR1pc
5’GGATCCTCTAGAGTCGACCTTGTCCCATTCGCCATTACCGAGGGGACGGTCCCCTCGGAA
TGTTGC 3’

ECORI T7P HMH9 NaLCrz_pc_court HDV BAMHI

Matrix PCR1 : pUC18 NaGIR1 pr

PCR1
FW1NaGIR1pc_c
5’AACAAATGGTACGAACCCTTAGAAATAAGGTTCGGGCTAGTAGTACAACTAAACGTCCTT
TGCAGGACGTTTTAGTGTGC3’
RV1NaGIR1pc
5’CCCCTCGGAATGTTGCCCACCGGCCGCCAGCGAGGAGGCTGGGACCATGGCCGGCCCTAG
ATGGTTACTAGG 3’

PCR2
FW2NaGIR1pc_c
5’GATCCATGACAGGATCGAAACCTCTTGCAAAAGAGGTCCATCCCATACAAAATGGTAACC
ATAAAACAAATGGTACGAACCC3’
RV1NaGIR1pc
5’CCCCTCGGAATGTTGCCCACCGGCCGCCAGCGAGGAGGCTGGGACCATGGCCGGCCCTAG
ATGGTTACTAGG 3’

PCR3
FW4NaAsGIR1pc
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTTGCGTTGGGATGCTGATGAAGAT
CCATGACAGGATCG 3’

RV2NaAsGIR1pc
5’GGATCCTCTAGAGTCGACCTTGTCCCATTCGCCATTACCGAGGGGACGGTCCCCTCGGAA
TGTTGC 3’

ECORI T7P HMH9 AlLCrz_pc HDV BAMHI

Matrix PCR1 : no matrix
Matrix PCR2 : product of PCR1

PCR1
FW1AsGIR1pc
5’CGCGCAAGTACCAAGAGAAAAACAAAGAAAGTTTCGACTTTCTTTTTCTCTTGTGCAATG
GGGTTTATGAGTTAATTAGCC 3’
RV1AsGIR1pc
5’CCGTTCTAACGAACTTAGTACGGTTACTTACACCTTTTTAAGGTCCCGTTTTGGCTAATT
AACTCATAAAC 3’

PCR2
FW2AsGIR1pc
5’CCCATTCGGACATATAATCCACCGCTTAGGAAGACGCAATATTCTTCTTGGACGCGCAAG
TACCAAGA 3’
RV2AsGIR1pc
5’GGTATACTGAGGACTATCCATTTATGAGAGCTAAATCGCTCAGCCGTGTAGTCTCTAGAC
ATTCCGTTCTAACGAACTTAG 3’

PCR3
FW3AsGIR1pc
5’GATCCATGACAGGATCGAAACCTCTTGCAAAAGAGGTCCATCCCATACAAAATGGTTAAA
TATTTCCCATTCGGACATATAAT 3’
RV3AsGIR1pc
5’CCCCTCGGAATGTTGCCCACCGGCCGCCAGCGAGGAGGCTGGGACCATGGCCGGCCTCAG
ATGGTATACTGAGGACTATC 3’

PCR4
FW4NaAsGIR1pc
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTTGCGTTGGGATGCTGATGAAGAT
CCATGACAGGATCG 3’
RV2NaAsGIR1pc
5’GGATCCTCTAGAGTCGACCTTGTCCCATTCGCCATTACCGAGGGGACGGTCCCCTCGGAA
TGTTGC 3’

ECORI T7P HMH9 AlLCrz_pc_court HDV BAMHI

Matrix PCR1 : no matrix
Matrix PCR2 : product of PCR1

PCR1
FW1AsGIR1pc_c
5’CGCGCAAGTACCAAGAGAAAAACAAATTCGTTTTTCTCTTGTGCAATGGGGTTTATGAGT
TAATTAGCC 3’
RV1AsGIR1pc
5’CCGTTCTAACGAACTTAGTACGGTTACTTACACCTTTTTAAGGTCCCGTTTTGGCTAATT
AACTCATAAAC 3’

PCR2

FW2AsGIR1pc
5’CCCATTCGGACATATAATCCACCGCTTAGGAAGACGCAATATTCTTCTTGGACGCGCAAG
TACCAAGA 3’
RV2AsGIR1pc
5’GGTATACTGAGGACTATCCATTTATGAGAGCTAAATCGCTCAGCCGTGTAGTCTCTAGAC
ATTCCGTTCTAACGAACTTAG 3’

PCR3
FW3AsGIR1pc
5’GATCCATGACAGGATCGAAACCTCTTGCAAAAGAGGTCCATCCCATACAAAATGGTTAAA
TATTTCCCATTCGGACATATAAT 3’
RV3AsGIR1pc
5’CCCCTCGGAATGTTGCCCACCGGCCGCCAGCGAGGAGGCTGGGACCATGGCCGGCCTCAG
ATGGTATACTGAGGACTATC 3’

PCR4
FW4NaAsGIR1pc
5’GACGGCCAGTGAATTCTAATACGACTCACTATAGGTTTGCGTTGGGATGCTGATGAAGAT
CCATGACAGGATCG 3’
RV2NaAsGIR1pc
5’GGATCCTCTAGAGTCGACCTTGTCCCATTCGCCATTACCGAGGGGACGGTCCCCTCGGAA
TGTTGC 3’

APPENDIX III

ECORI T7P DiLCrz actif U203C HDV BAMHI

Matrix : ECORI T7P DiLCrz actif HDV BAMHI
FW203 5'CCCTGCCTCTTAGGTGTGTCCAATGAACAGTCGTTCCGAAAG 3'
RV203 5'CTTTCGGAACGACTGTTCATTGGACACACCTAAGAGGCAGGG 3'

ECORI T7P HMH9 NaLCrz_pc U261C HDV BAMHI

Matrix : ECORI T7P HMH9 NaLCrz_pc HDV BAMHI
2Fw203Na 5'CAAACCAATTTGGTGGTGCGAATGGATAGTCCC3'
2Rv203Na 5'GGGACTATCCATTCGCACCACCAAATTGGTTTG3'

ECORI T7P HMH9 NaLCrz_pc_court U261C HDV BAMHI

Matrix : ECORI T7P HMH9 NaLCrz_pc_court HDV BAMHI
2Fw203Na 5'CAAACCAATTTGGTGGTGCGAATGGATAGTCCC3'
2Rv203Na 5'GGGACTATCCATTCGCACCACCAAATTGGTTTG3'

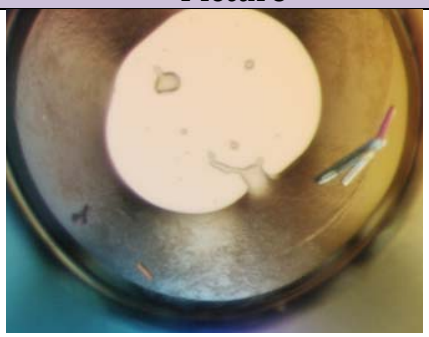
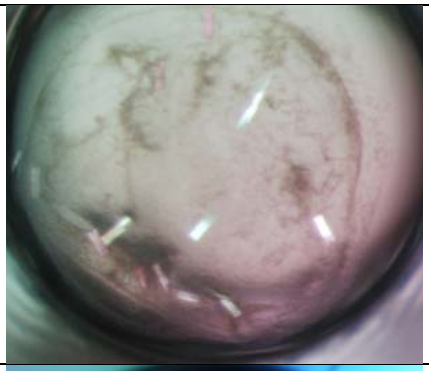
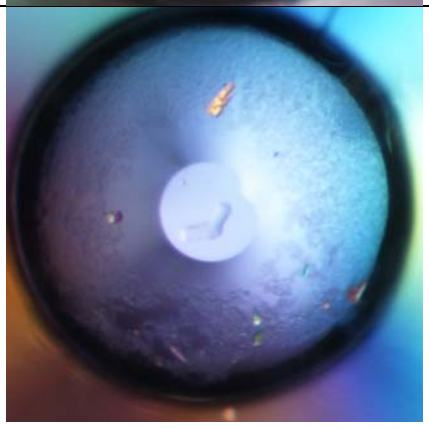
ECORI T7P HMH9 AlLCrz_pc U252C HDV BAMHI

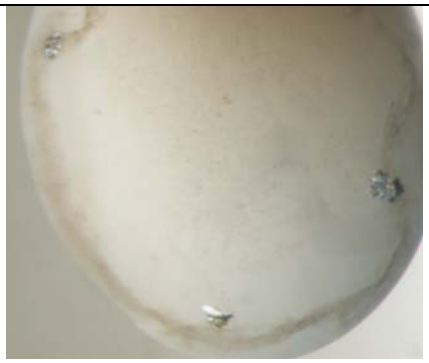
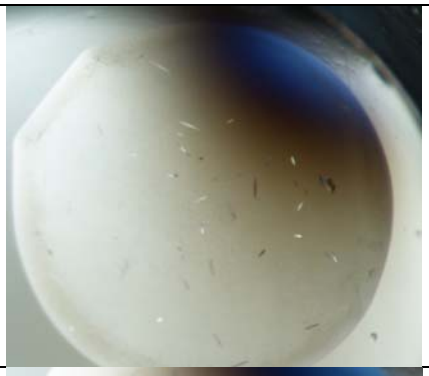
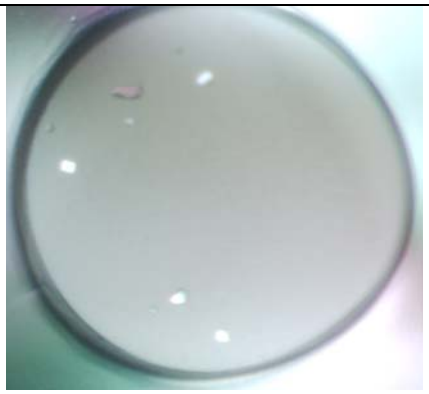
Matrix : ECORI T7P HMH9 AlLCrz_pc HDV BAMHI
2Fw203As 5'GCGATTTAGCTCTCACAAATGGATAGTCCTC3'
2Rv203As 5'GAGGACTATCCATTTGTGAGAGCTAAATCGC3'

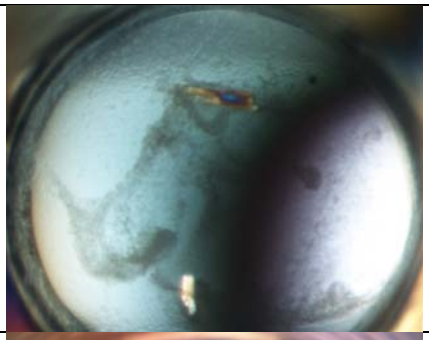
ECORI T7P HMH9 AlLCrz_pc_court U252C HDV BAMHI

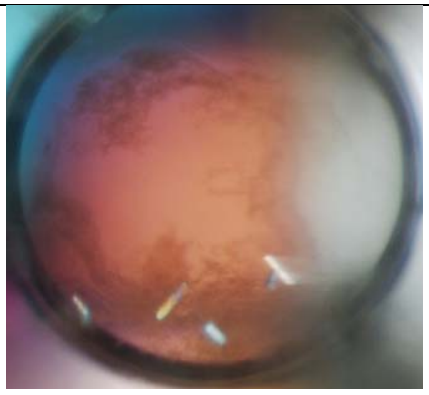
Matrix : ECORI T7P HMH9 AlLCrz_pc_court HDV BAMHI
2Fw203As 5'GCGATTTAGCTCTCACAAATGGATAGTCCTC3'
2Rv203As 5'GAGGACTATCCATTTGTGAGAGCTAAATCGC3'

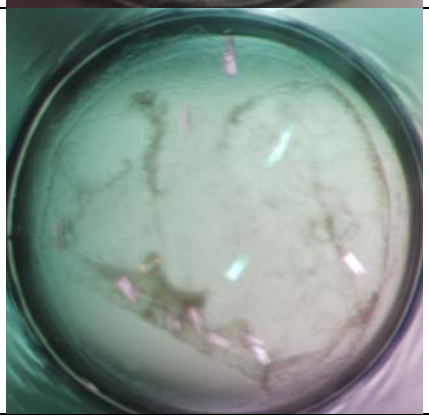
APPENDIX IV

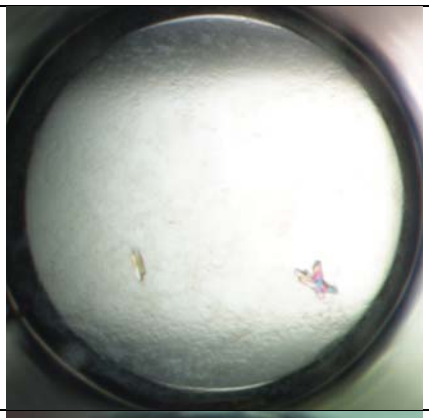
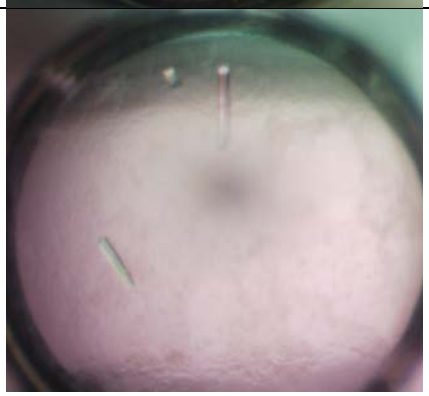
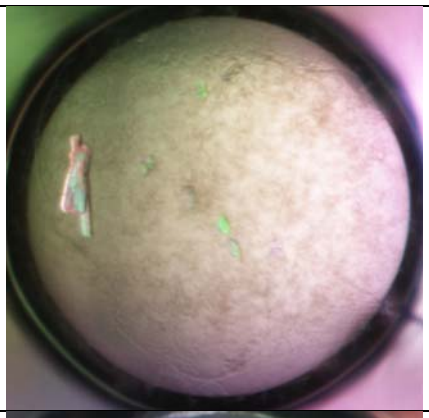
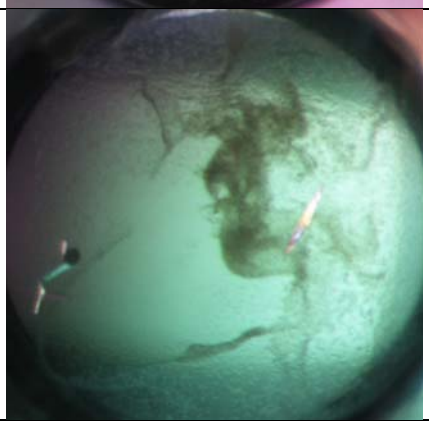
RNA	Picture
• [RNA concentration] • Crystallization drop volume • Crystallization set up • Crystallization temperature • Time lapse between the crystallization trial and the picture • Crystallization condition • Freezing condition • Best resolution	
DiLCrz_cp_U203C	**Picture**
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 months • Natrix HT™ B8: 0.1 M Ammonium acetate, 0.015 M Magnesium acetate tetrahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.5, 10% v/v 2-Propanol	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 months • Natrix HT™ D2: 0.2 M Ammonium acetate, 0.15 M Magnesium acetate tetrahydrate, 0.05 M HEPES sodium pH 7.0, 5% w/v Polyethylene glycol 4,000	
• 100 µM • 400 nL • Sitting drop • 20 °C • 14 weeks • Natrix HT™ E11: 0.075 M NaCl, 0.002 M Calcium chloride dihydrate, 0.05 M Na cacodylate trihydrate pH 6.0, 30% w/v 1,6-Hexanediol, 0.0005 M Spermine • Frozen in liquid ethane • 4 Å	

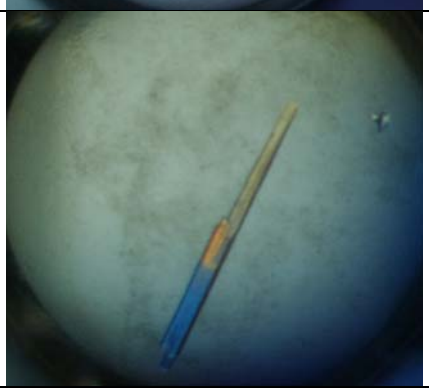
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 months • Natrix HT™ F1: 0.05 M Na cacodylate trihydrate pH 6.0, 30% v/v MPD	
• 100 µM • 400 nL • Sitting drop • 20 °C • 2 weeks • Natrix HT™ G7: 0.08 M Strontium chloride hexahydrate, 0.02 M Magnesium chloride hexahydrate, 0.04 M Na cacodylate trihydrate pH 7.0, 20% v/v MPD, 0.012 M Spermine tetrahydrochloride	
• 100 µM • 400 nL • Sitting drop • 20 °C • 4 weeks • Natrix HT™ G8: 0.08 M NaCl, 0.04 M Na cacodylate trihydrate pH 7.0, 30% v/v MPD, 0.012 M Spermine tetrahydrochloride	
• 100 µM • 400 nL • Sitting drop • 20 °C • 2 weeks • Natrix HT™ G10: 0.04 M LiCl, 0.08 M Strontium chloride hexahydrate, 0.02 M Magnesium chloride hexahydrate, 0.04 M Na cacodylate trihydrate pH 7.0, 30% v/v MPD, 0.012 M Spermine tetrahydrochloride	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 weeks • Natrix HT™ H4: 0.08 M KCl, 0.02 M Magnesium chloride hexahydrate, 0.04 M Na cacodylate trihydrate pH 7.0, 50% v/v MPD, 0.012 M Spermine tetrahydrochloride	

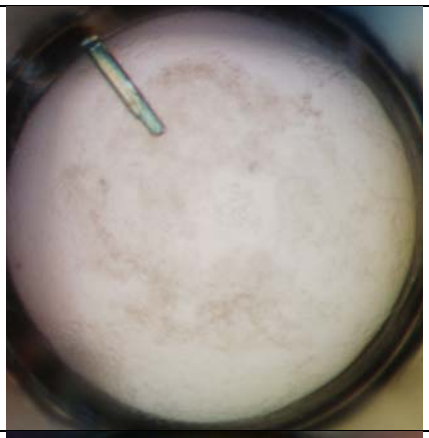
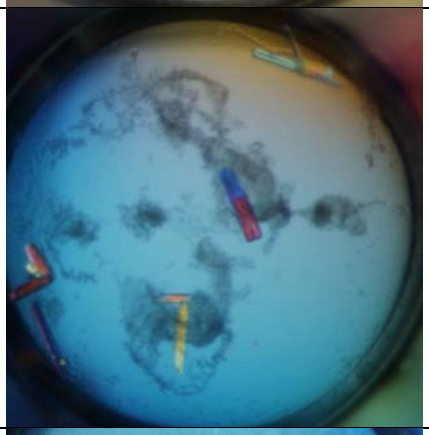
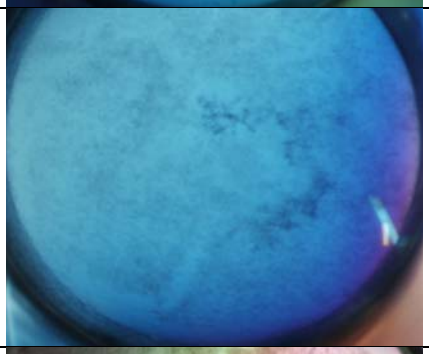
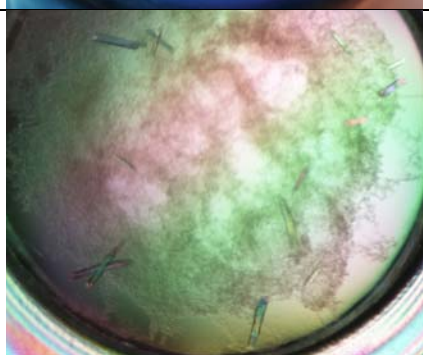
• 100 µM • 400 nL • Sitting drop • 20 °C • 4 weeks • Index HT™ D6: 0.1 M BIS-TRIS pH 5.5, 25% w/v Polyethylene glycol 3,350	
• 100 µM • 400 nL • Sitting drop • 20 °C • 4 weeks • Index HT™ D7: 0.1 M BIS-TRIS pH 6.5, 25% w/v Polyethylene glycol 3,350 • Frozen in liquid ethane • 7 Å	
• 100 µM • 400 nL • Sitting drop • 20 °C • 4 weeks • Index HT™ D8: 0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3,350	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 months • Index HT™ D10: 0.1 M BIS-TRIS pH 6.5, 20% w/v Polyethylene glycol monomethyl ether 5,000	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 days • Index HT™ D11: 0.1 M BIS-TRIS pH 6.5, 28% w/v Polyethylene glycol monomethyl ether 2,000 • Frozen in liquid ethane • 2.8 Å	

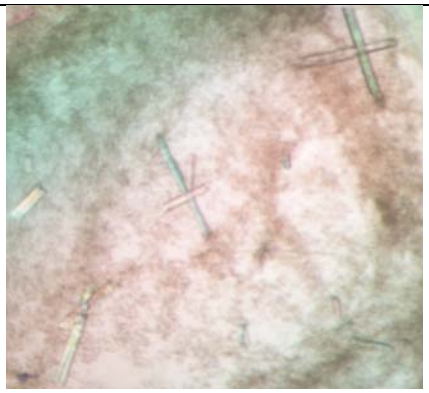
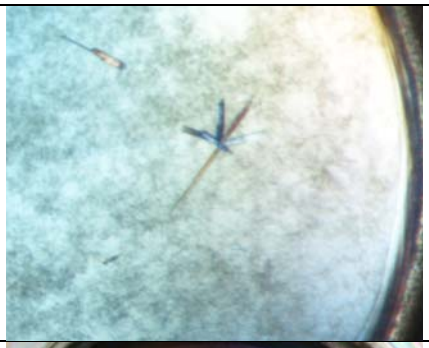
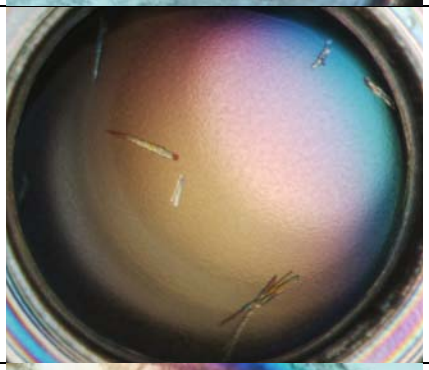
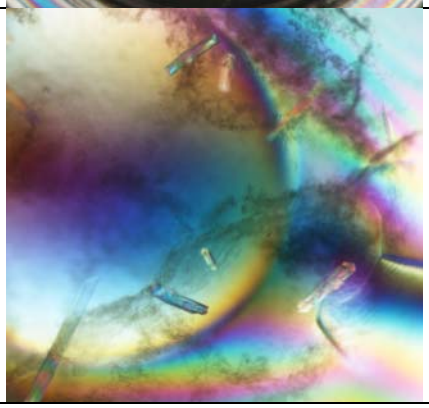
• 100 µM • 400 nL • Sitting drop • 20 °C • 4 weeks • Index HT™ E1: 0.2 M Calcium chloride dihydrate, 0.1 M BIS-TRIS pH 6.5, 45% v/v (+/-)-2-Methyl-2,4-pentanediol	
• 100 µM • 400 nL • Sitting drop • 20 °C • 4 weeks • Index HT™ E4: 0.2 M Ammonium acetate, 0.1 M HEPES pH 7.5, 45% v/v (+/-)-2-Methyl-2,4-pentanediol	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 months • Index HT™ E9: 0.05 M Ammonium sulfate, 0.05 M BIS-TRIS pH 6.5, 30% v/v Pentaerythritol ethoxylate (15/4 EO/OH)	
• 100 µM • 400 nL • Sitting drop • 20 °C • 4 weeks • Index HT™ F10: 0.2 M Sodium chloride, 0.1 M BIS-TRIS pH 5.5, 25% w/v Polyethylene glycol 3,350	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 months • Index HT™ F11: 0.2 M Sodium chloride, 0.1 M BIS-TRIS pH 6.5, 25% w/v Polyethylene glycol 3,350 • Frozen in liquid ethane • 5 Å	

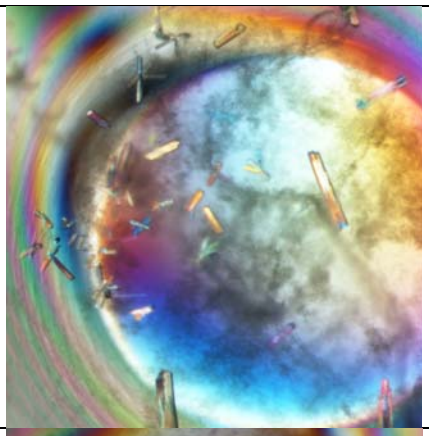
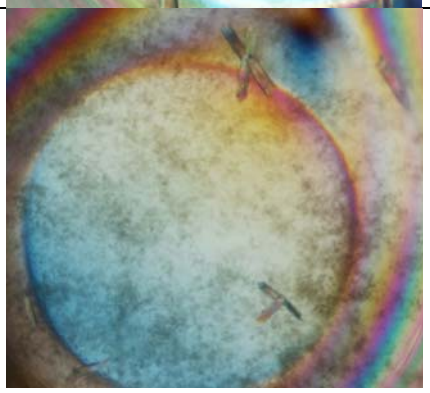
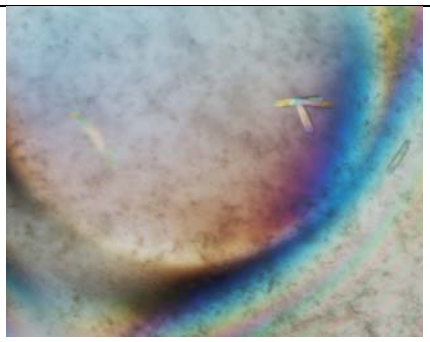
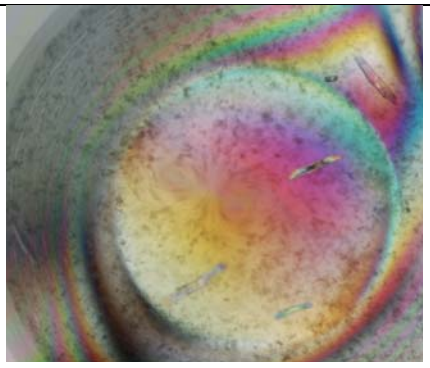
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 months • Index HT™ F12: 0.2 M Sodium chloride, 0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3,350 • Frozen in liquid ethane • 5 Å	
• 100 µM • 400 nL • Sitting drop • 20 °C • 4 weeks • Index HT™ H6: 0.2 M Sodium formate, 20% w/v Polyethylene glycol 3,350 • Frozen in liquid ethane • 2.8 Å	
• 100 µM • 400 nL • Sitting drop • 20 °C • 9 weeks • PEGRx HT™ B8: 0.1 M BIS-TRIS pH 6.5, 20% w/v Polyethylene glycol 1,500	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 weeks • PEGRx HT™ D2: 0.1 M Imidazole pH 7.0, 20% w/v Polyethylene glycol 6,000	

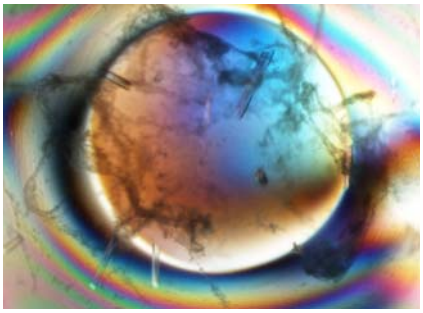
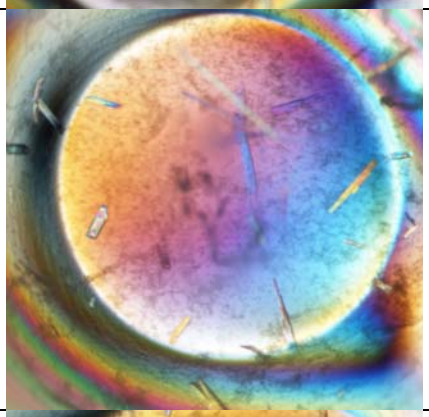
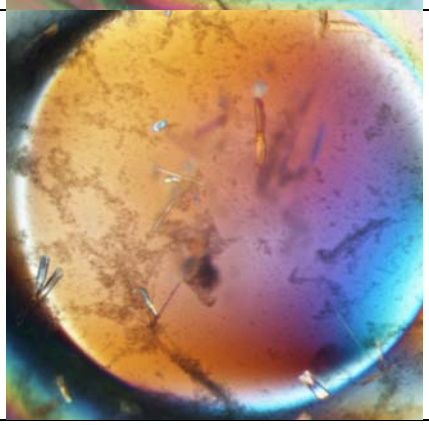
• 100 µM • 400 nL • Sitting drop • 20 °C • 2 weeks • PEGRx HT™ F1: 0.1 M Succinic acid pH 7.0, 0.1 M BICINE pH 8.5, 30% v/v Polyethylene glycol monomethyl ether 550 • Frozen in liquid ethane • 6 Å	
• 100 µM • 400 nL • Sitting drop • 20 °C • 9 weeks • PEGRx HT™ F3: 0.1 M Sodium malonate pH 8.0, 0.1 M Tris pH 8.0, 30% w/v Polyethylene glycol 1,000 • Frozen in liquid ethane • 5 Å	
• 100 µM • 400 nL • Sitting drop • 20 °C • 9 weeks • PEGRx HT™ E11: 1.8 M Ammonium sulfate, 0.1 M BIS-TRIS pH 6.5, 2% v/v Polyethylene glycol monomethyl ether 550	
• 100 µM • 400 nL • Sitting drop • 20 °C • 9 weeks • PEGRx HT™ G7: 2% v/v Polyethylene glycol 400, 0.1 M Imidazole pH 7.0, 24% w/v Polyethylene glycol monomethyl ether 5,000	

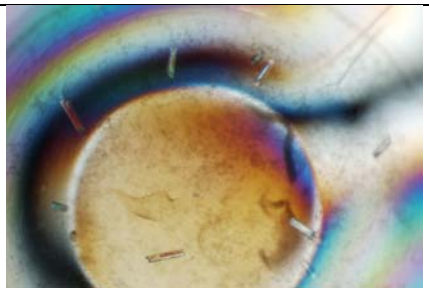
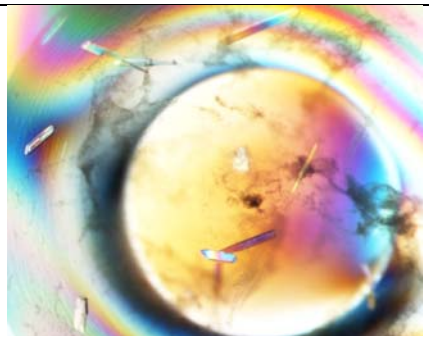
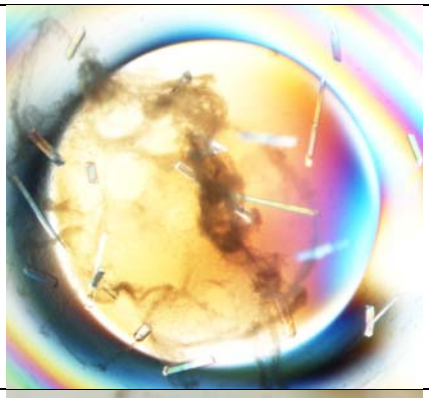
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 weeks • Nuc-Pro HTS C2: 100 mM Sodium chloride, 50 mM MOPS pH 7.0, 20% w/v Polyethylene glycol 4000	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 months • Nuc-Pro HTS G6: 100 mM Sodium acetate, 50mM ADA pH 6.0, 15% v/v MPD	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 months • Nuc-Pro HTS G9: 100 mM Sodium chloride, 50 mM MES Sodium salt pH 6.0, 23% v/v MPD	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 weeks • 0.1 M Sodium malonate pH 8.0, 0.1 M Tris pH 7.0, 30% w/v Polyethylene glycol 1,000 • Frozen in liquid nitrogen • 3 Å	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 weeks • 0.1 M Sodium malonate pH 8.0, 0.1 M Tris pH 7.5, 30% w/v Polyethylene glycol 1,000 • Frozen in liquid nitrogen • 3 Å	

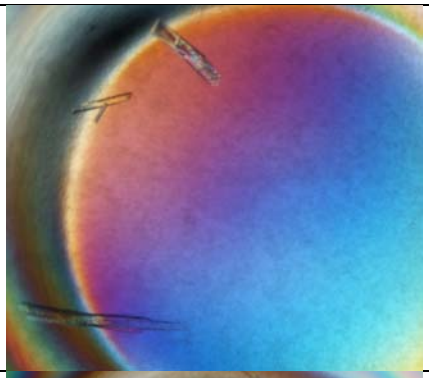
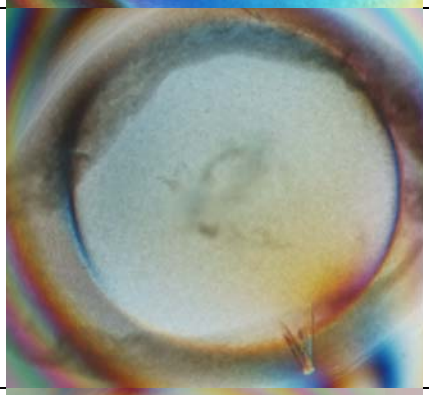
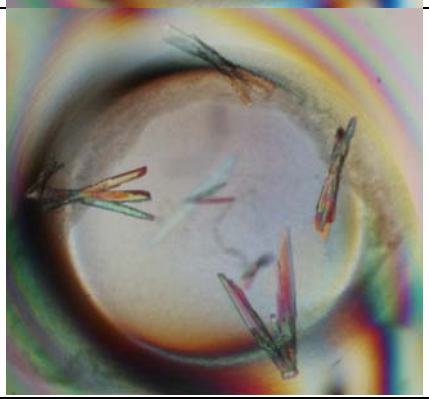
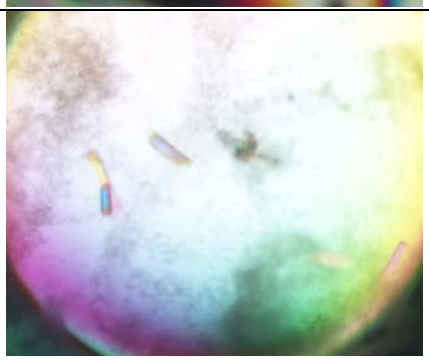
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 weeks • 0.1 M Sodium malonate pH 8.0, 0.1 M Tris pH 8.0, 30% w/v Polyethylene glycol 1,000 • Frozen in liquid nitrogen • 3 Å	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 weeks • 0.2 M Sodium chloride, 0.1 M HEPES pH 8.5, 25% w/v Polyethylene glycol 3,350 • Frozen in liquid nitrogen • 3.5 Å	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 weeks • 0.2 M Sodium chloride, 0.1 M BIS-TRIS pH 7.5, 20% w/v Polyethylene glycol 1,500	
• 100 µM • 2 µL • Hanging drop • 20 °C • 3 weeks • Index HT™ D11: 0.1 M BIS-TRIS pH 6.5, 28% w/v Polyethylene glycol monomethyl ether 2,000 • Frozen in liquid ethane • 3.5 Å	

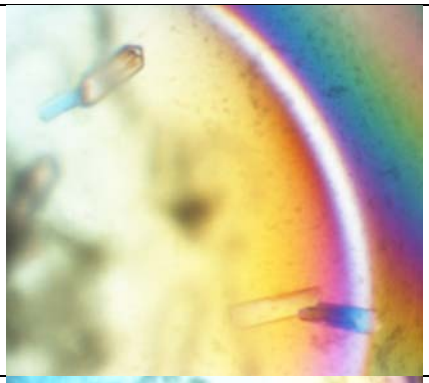
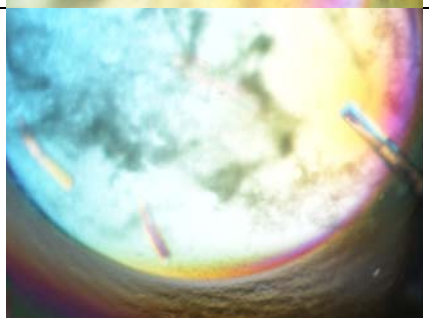
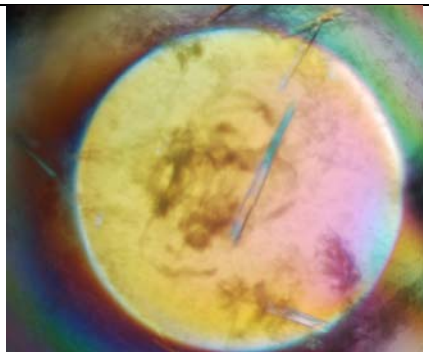
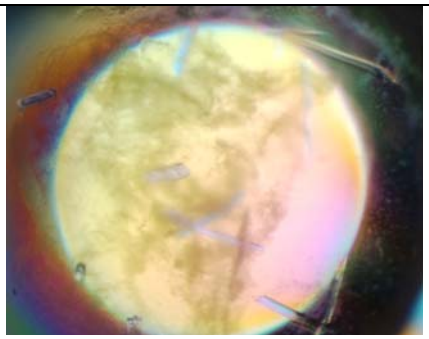
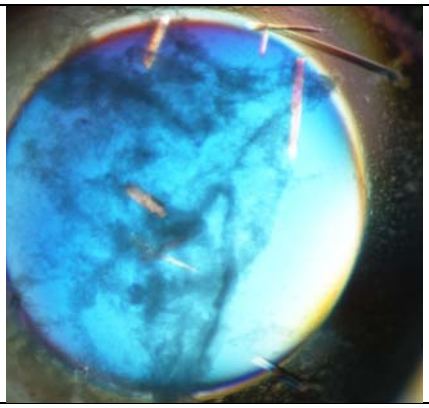
• 100 µM • 2 µL • Hanging drop • 20 °C • 2 weeks • Index HT™ F11: 0.2 M Sodium chloride, 0.1 M BIS-TRIS pH 6.5, 25% w/v Polyethylene glycol 3,350 • Frozen in liquid ethane • 5 Å	
• 100 µM • 2 µL • Hanging drop • 20 °C • 3 weeks • Index HT™ F12: 0.2 M Sodium chloride, 0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3,350	
• 100 µM • 2 µL • Hanging drop • 20 °C • 6 weeks • Natrix HT™ E11: 0.075 M NaCl, 0.002 M Calcium chloride dihydrate, 0.05 M Na cacodylate trihydrate pH 6.0, 30% w/v 1,6-Hexanediol, 0.0005 M Spermine	
• 100 µM • 5 µL • Microbatch sealed with paraffin oil • 20 °C • 3 weeks • 0.12 M BIS-TRIS pH 6.5, 33.6% w/v Polyethylene glycol monomethyl ether 2,000 • Frozen in liquid ethane • 7 Å	

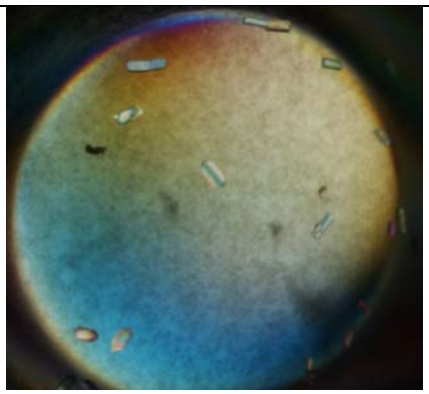
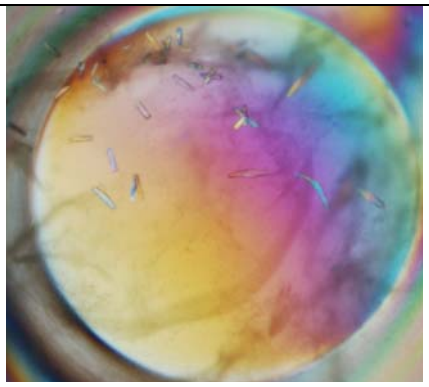
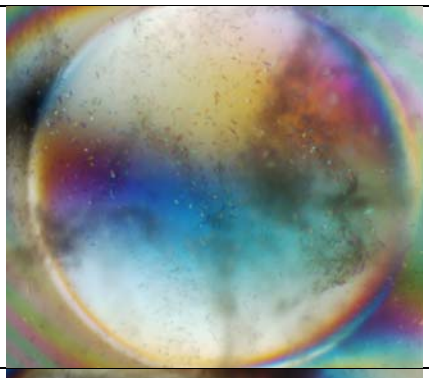
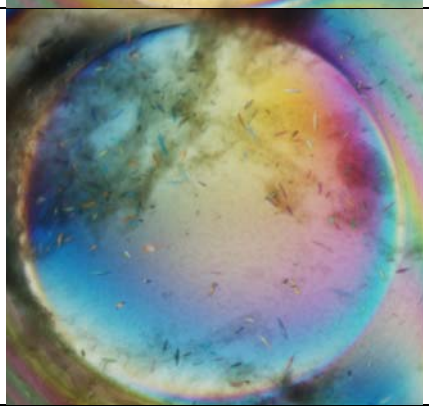
• 100 µM • 5 µL • Microbatch sealed with paraffin oil • 20 °C • 6 days • 0.14 M BIS-TRIS pH 6.5, 39.2% w/v Polyethylene glycol monomethyl ether 2,000	
• 100 µM • 5 µL • Microbatch in gel sealed with paraffin oil • 20 °C • 5 weeks • 0.12 M BIS-TRIS pH 6.5, 33.6% w/v Polyethylene glycol monomethyl ether 2,000, 1%low melting agarose • Frozen in liquid ethane • 10 Å	
• 100 µM • 5 µL • Microbatch in gel sealed with paraffin oil • 20 °C • 5 weeks • 0.14 M BIS-TRIS pH 6.5, 39.2% w/v Polyethylene glycol monomethyl ether 2,000, 1%low melting agarose • Frozen in liquid ethane • 10 Å	
• 100 µM • 5 µL • Microbatch in gel sealed with paraffin oil • 20 °C • 5 weeks • 0.16 M BIS-TRIS pH 6.5, 44.8% w/v Polyethylene glycol monomethyl ether 2,000, 1%low melting agarose • Frozen in liquid ethane • 10 Å	

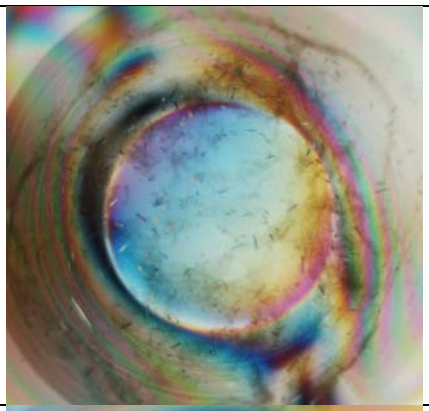
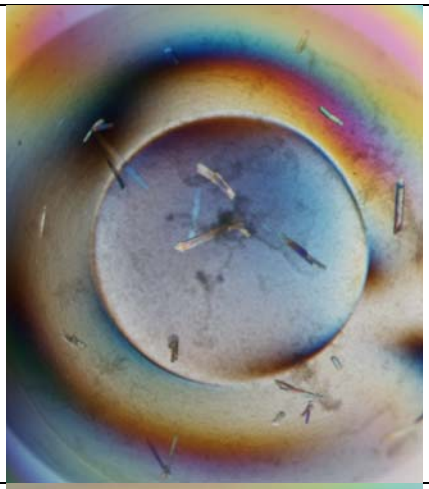
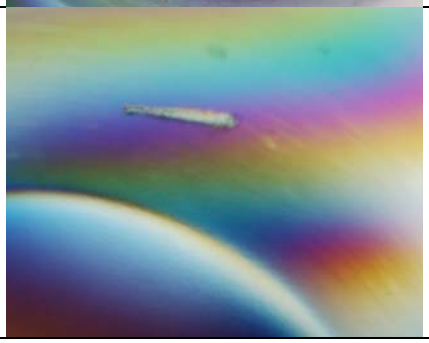
• 100 µM • 5 µL • Microbatch sealed with paraffin oil • 20 °C • 3 weeks • 0.24 M Sodium chloride, 0.12 M BIS-TRIS pH 6.5, 30% w/v Polyethylene glycol 3,350 • Frozen in liquid ethane • 7 Å	
• 100 µM • 5 µL • Microbatch sealed with paraffin oil • 20 °C • 3 weeks • 0.28 M Sodium chloride, 0.14 M BIS-TRIS pH 6.5, 35% w/v Polyethylene glycol 3,350 • Frozen in liquid ethane • 7 Å	
• 100 µM • 5 µL • Microbatch sealed with paraffin oil • 20 °C • 3 weeks • 0.32 M Sodium chloride, 0.16 M BIS-TRIS pH 6.5, 40% w/v Polyethylene glycol 3,350 • Frozen in liquid ethane • 7 Å	
• 100 µM • 5 µL • Microbatch sealed with paraffin oil • 20 °C • 5 days • 0.36 M Sodium chloride, 0.18 M BIS-TRIS pH 6.5, 45% w/v Polyethylene glycol 3,350 • Frozen in liquid ethane • 7 Å	

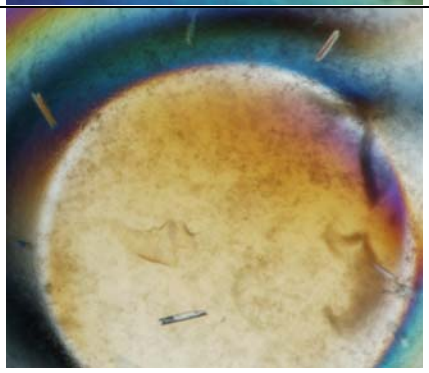
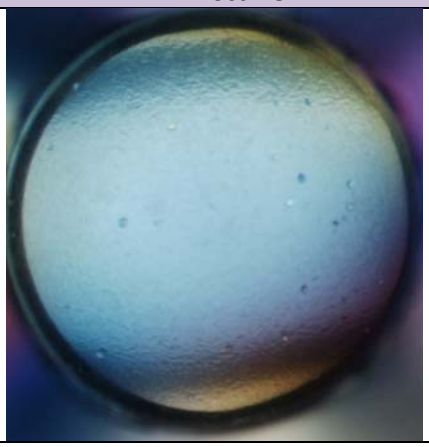
• 100 µM • 5 µL • Microbatch sealed with paraffin oil • 20 °C • 5 days • 0.22 M Sodium chloride, 0.11 M BIS-TRIS pH 6.5, 27.5% w/v Polyethylene glycol 3,350 • Frozen in liquid ethane • 3.6 Å	
• 100 µM • 5 µL • Microbatch sealed with paraffin oil • 20 °C • 3 weeks • 0.28 M Sodium chloride, 0.14 M HEPES pH 7.5, 35% w/v Polyethylene glycol 3,350 • Frozen in liquid ethane • 4 Å	
• 100 µM • 5 µL • Microbatch sealed with paraffin oil • 20 °C • 3 weeks • 0.32 M Sodium chloride, 0.16 M HEPES pH 7.5, 40% w/v Polyethylene glycol 3,350	
• 100 µM • 5 µL • Microbatch sealed with paraffin oil • 20 °C • 3 weeks • 0.36 M Sodium chloride, 0.18 M HEPES pH 7.5, 45% w/v Polyethylene glycol 3,350 • Frozen in liquid ethane • 2.7 Å	

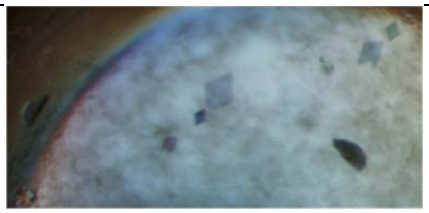
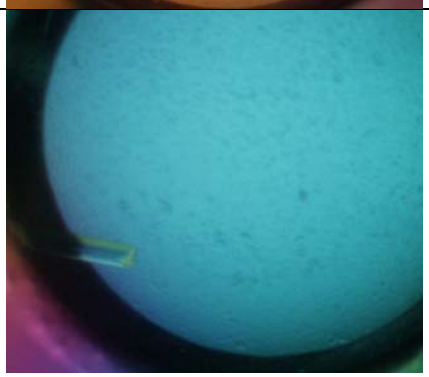
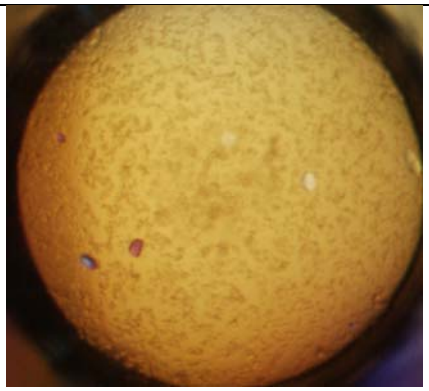
• 100 µM • 5 µL • Microbatch sealed with paraffin oil • 20 °C • 5 weeks • 0.120 M NaCl, 0.0032 M Calcium chloride dihydrate, 0.08 M Na cacodylate trihydrate pH 6.0, 48% w/v 1,6-Hexanediol, 0.008 M Spermine	
• 100 µM • 5 µL • Microbatch sealed with paraffin oil • 20 °C • 5 weeks • 0.135 M NaCl, 0.0036 M Calcium chloride dihydrate, 0.09 M Na cacodylate trihydrate pH 6.0, 54% w/v 1,6-Hexanediol, 0.009 M Spermine	
• 100 µM • 5 µL • Microbatch sealed with paraffin oil • 20 °C • 5 weeks • 0.135 M NaCl, 0.0036 M Calcium chloride dihydrate, 0.09 M Na cacodylate trihydrate pH 6.0, 54% w/v 1,6-Hexanediol, 0.009 M Spermine • Frozen in liquid ethane • To many crystals in the loop	
• 100 µM • 5 µL • Microbatch sealed with crystal tape • 20 °C • 5 weeks • 0.16 M BIS-TRIS pH 6.5, 44.8% w/v Polyethylene glycol monomethyl ether 2,000 • Frozen in liquid ethane • 7 Å	

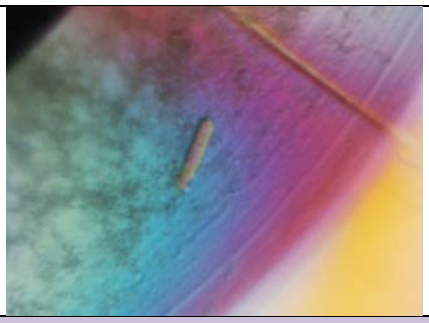
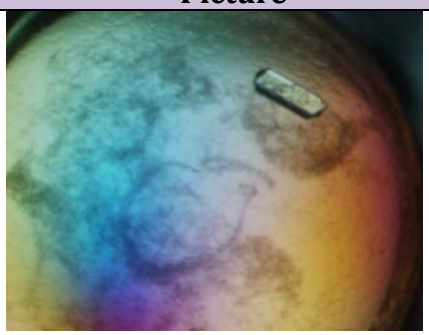
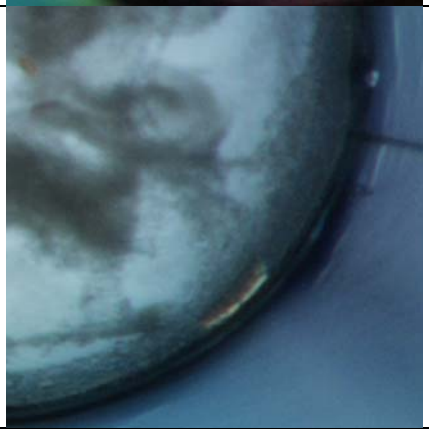
• 100 µM • 5 µL • Microbatch sealed with crystal tape • 20 °C • 3 weeks • 0.28 M Sodium chloride, 0.14 M BIS-TRIS pH 6.5, 35% w/v Polyethylene glycol 3,350 • Frozen in liquid ethane • 7 Å	
• 100 µM • 5 µL • Microbatch sealed with crystal tape • 20 °C • 3 weeks • 0.32 M Sodium chloride, 0.16 M BIS-TRIS pH 6.5, 40% w/v Polyethylene glycol 3,350 • Frozen in liquid ethane • 5 Å	
• 100 µM • 5 µL • Microbatch sealed with crystal tape • 20 °C • 3 weeks • 0.28 M Sodium chloride, 0.14 M HEPES pH 7.5, 35% w/v Polyethylene glycol 3,350 • Frozen in liquid ethane • 2.5 Å	
• 100 µM • 5 µL • Microbatch sealed with crystal tape • 20 °C • 3 weeks • 0.32 M Sodium chloride, 0.16 M HEPES pH 7.5, 40% w/v Polyethylene glycol 3,350 • Frozen in liquid ethane • 3.7 Å	
• 100 µM • 5 µL • Microbatch sealed with crystal tape • 20 °C • 3 weeks • 0.36 M Sodium chloride, 0.18 M HEPES pH 7.5, 45% w/v Polyethylene glycol 3,350 • Frozen in liquid ethane • 4 Å	

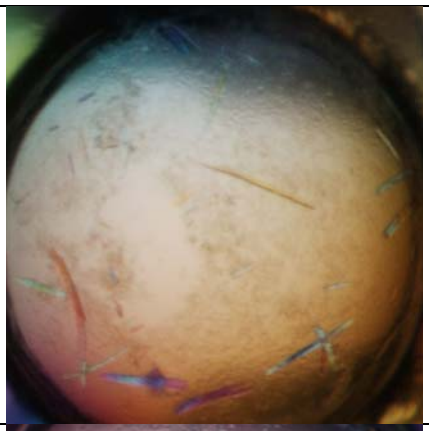
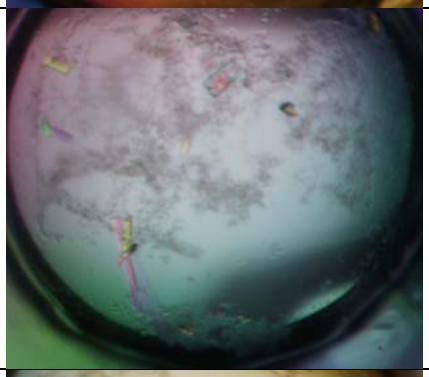
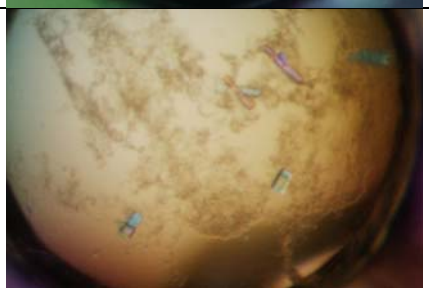
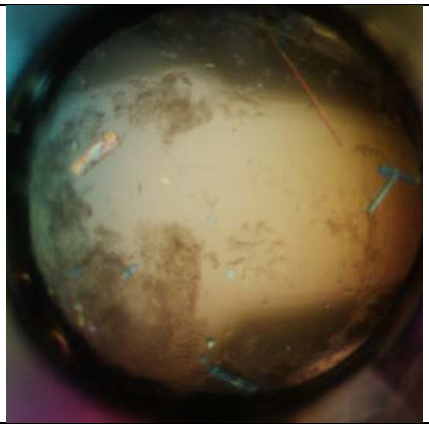
• 100 µM • 5 µL • Microbatch sealed with paraffin oil • 20 °C • 3 weeks • 0.32 M Sodium chloride, 0.16 M BIS-TRIS pH 6.5, 40% w/v Polyethylene glycol 3,350, Spermidine 12 mM • Frozen in liquid nitrogen • 7 Å	
• 100 µM • 5 µL • Microbatch sealed with paraffin oil • 20 °C • 5 days • 0.06 M BIS-TRIS pH 6.5, 16.8% w/v Polyethylene glycol monomethyl ether 2,000, 1 mM Iridium hexamine • Frozen in liquid nitrogen • 7 Å	
• 100 µM • 5 µL • Microbatch sealed with paraffin oil • 20 °C • 5 days • 0.12 M BIS-TRIS pH 6.5, 33.6% w/v Polyethylene glycol monomethyl ether 2,000, 1 mM Iridium hexamine	
• 100 µM • 5 µL • Microbatch sealed with paraffin oil • 20 °C • 5 days • 0.14 M BIS-TRIS pH 6.5, 39.2% w/v Polyethylene glycol monomethyl ether 2,000, 1 mM Iridium hexamine	

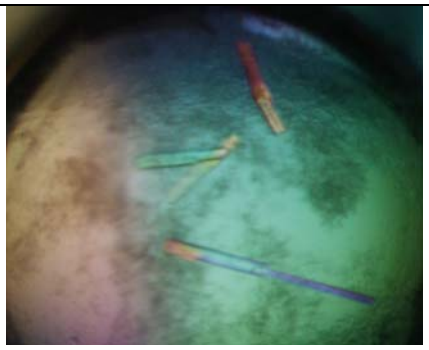
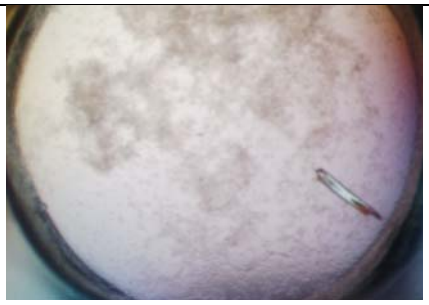
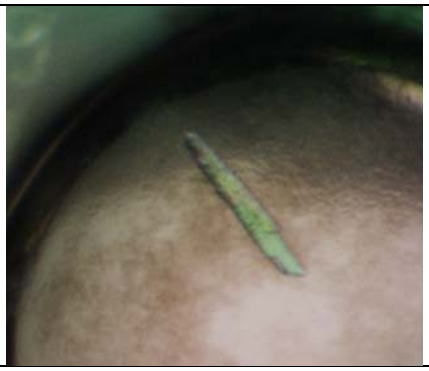
• 100 µM • 5 µL • Microbatch sealed with paraffin oil • 20 °C • 5 days • 0.16 M BIS-TRIS pH 6.5, 44.8% w/v Polyethylene glycol monomethyl ether 2,000, 1 mM Iridium hexamine	
• 100 µM • 5 µL • Microbatch sealed with paraffin oil • 20 °C • 5 days • 0.24 M Sodium chloride, 0.12 M BIS-TRIS pH 6.5, 30% w/v Polyethylene glycol 3,350, 1 mM Iridium hexamine • Frozen in liquid nitrogen • 7 Å	
• 100 µM • 5 µL • Microbatch sealed with paraffin oil • 20 °C • 5 days • 0.32 M Sodium chloride, 0.16 M BIS-TRIS pH 6.5, 40% w/v Polyethylene glycol 3,350, 1 mM Iridium hexamine • Frozen in liquid nitrogen • 7 Å	
• 100 µM • 5 µL • Microbatch sealed with crystal tape • 20 °C • 10 days • 0.12 M Succinic acid pH 7.0, 0.12 M BICINE pH 8.5, 36% v/v Polyethylene glycol monomethyl ether 550, 1 mM Iridium hexamine	

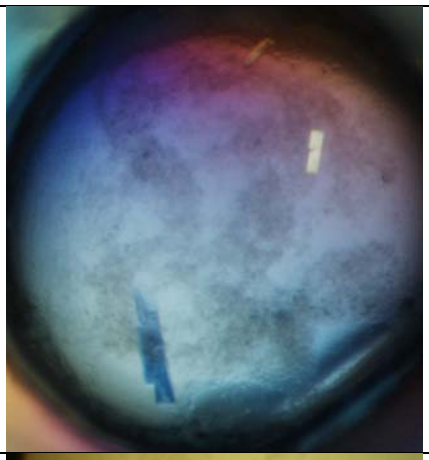
• 100 µM • 5 µL • Gel Microbatch sealed with paraffin oil • 20 °C • 4 weeks • 0.1 M BIS-TRIS pH 6.5, 28% w/v Polyethylene glycol monomethyl ether 2,000, 1% low melting agarose • Frozen in liquid nitrogen • 10 Å	
• 100 µM • 5 µL • Gel Microbatch sealed with paraffin oil • 20 °C • 4 weeks • 0.2 M Sodium chloride, 0.1 M BIS-TRIS pH 6.5, 25% w/v Polyethylene glycol 3,350, 1% low melting agarose • Frozen in liquid nitrogen • 10 Å	
• 100 µM • 5 µL • Gel Microbatch sealed with paraffin oil • 20 °C • 4 weeks • 0.2 M Sodium formate, 20% w/v Polyethylene glycol 3,350, 1%low melting agarose	
DiLCrz_active	**Picture**
• 100 µM • 400 nL • Sitting drop • 20 °C • 12 weeks • Natrix HT™ D5: 0.1 M Potassium chloride, 0.015 M Magnesium chloride hexahydrate, 0.05 M TRIS hydrochloride pH 7.5, 10% v/v Polyethylene glycol monomethyl ether 550	
• 100 µM • 5 µL • Microbatch sealed with crystal tape • 20 °C • 3 weeks	

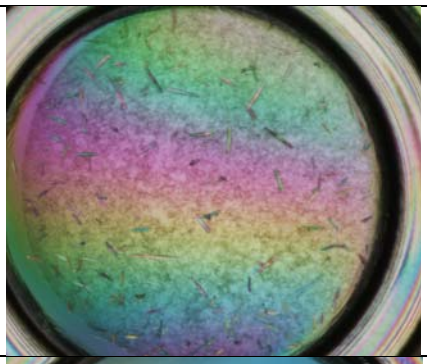
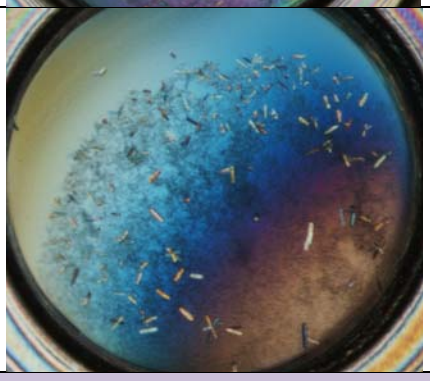
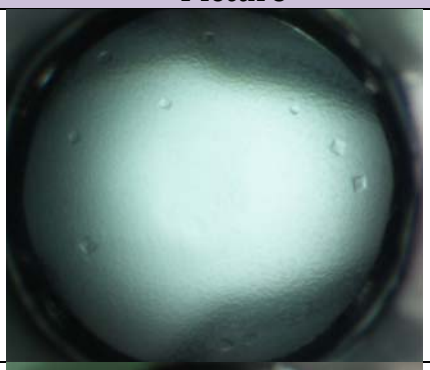
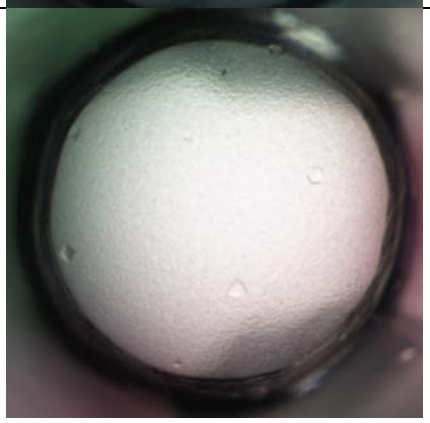
• 0.24 M Sodium chloride, 0.12 M HEPES pH 7.5, 30% w/v Polyethylene glycol 3,350, 12 mM Spermidine • Frozen in liquid nitrogen • 7 Å	
DiLCrz_active_U203C	**Picture**
• 100 µM • 400 nL • Sitting drop • 20 °C • 10 days • PEGRx HT™ A10: 0.1 M Sodium citrate tribasic dihydrate pH 5.0, 30% v/v Polyethylene glycol monomethyl ether 550 • Frozen in liquid nitrogen • 8 Å	
• 100 µM • 400 nL • Sitting drop • 20 °C • 4 weeks • PEGRx HT™ C1: 0.1 M Sodium citrate tribasic dihydrate pH 5.0, 30% v/v Jeffamine ED-2001 pH 7.0 • Frozen in liquid nitrogen • 4 Å	
• 100 µM • 400 nL • Sitting drop • 20 °C • 10 days • PEGRx HT™ F3: 0.1 M Sodium malonate pH 8.0, 0.1 M Tris pH 8.0, 30% w/v Polyethylene glycol 1,000 • Frozen in liquid nitrogen • 9 Å	
DiLCrz_cp_P2-1	**Picture**
• 100 µM • 5 µL • Microbatch sealed with crystal tape • 20 °C • 13 weeks • 0.24 M Sodium chloride, 0.12 M BIS-TRIS pH 6.5, 30% w/v Polyethylene glycol 3,350, 12 mM Putrescine	

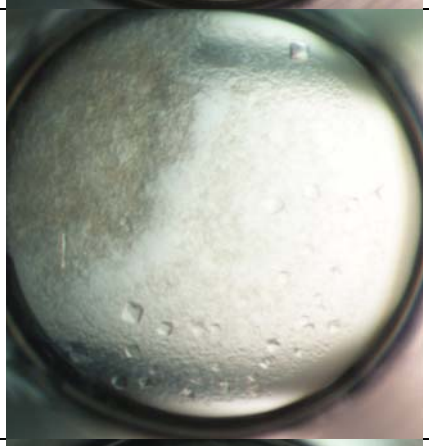
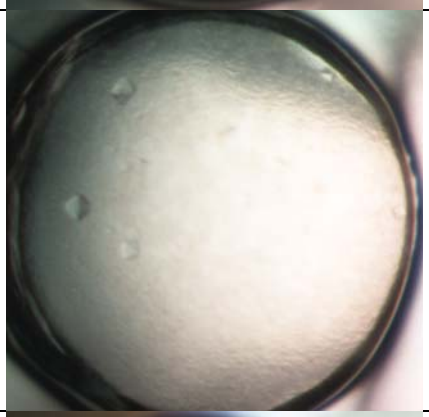
• Frozen in liquid nitrogen • 3.6 Å	
DiLCrz_cp_P2-1_U203C	**Picture**
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 weeks • Index HT™ D6: 0.1 M BIS-TRIS pH 5.5, 25% w/v Polyethylene glycol 3,350 • Frozen in liquid nitrogen • 3.5 Å	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 weeks • Index HT™ D7: 0.1 M BIS-TRIS pH 6.5, 25% w/v Polyethylene glycol 3,350	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 weeks • Index HT™ D8: 0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3,350	
• 100 µM • 400 nL • Sitting drop • 20 °C • 3 weeks • Index HT™ D11: 0.1 M BIS-TRIS pH 6.5, 28% w/v Polyethylene glycol monomethyl ether 2,000 • Frozen in liquid nitrogen • 3 Å	

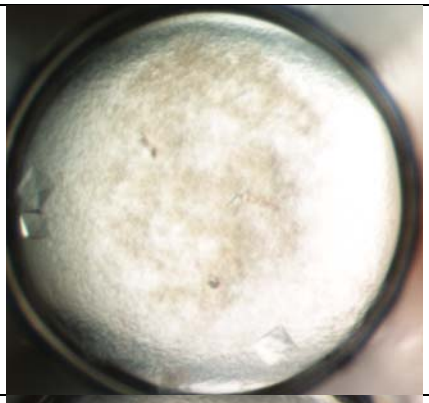
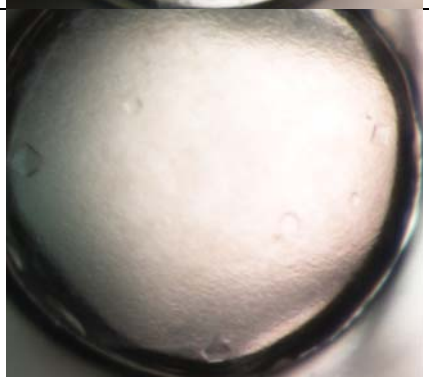
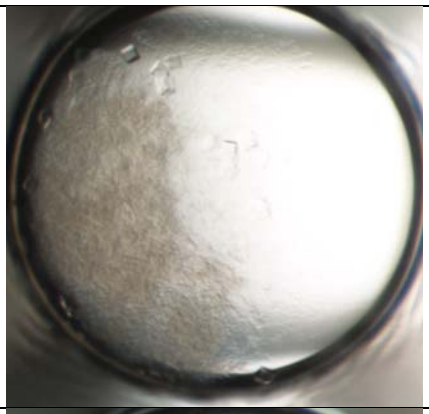
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 weeks • Index HT™ E8: 0.2 M Potassium chloride, 0.05 M HEPES pH 7.5, 35% v/v Pentaerythritol propoxylate (5/4 PO/OH) • Frozen in liquid nitrogen • 3 Å	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 weeks • Index HT™ F10: 0.2 M Sodium chloride, 0.1 M BIS-TRIS pH 5.5, 25% w/v Polyethylene glycol 3,350	
• 100 µM • 400 nL • Sitting drop • 20 °C • 5 days • Index HT™ F11: 0.2 M Sodium chloride, 0.1 M BIS-TRIS pH 6.5, 25% w/v Polyethylene glycol 3,350	
• 100 µM • 400 nL • Sitting drop • 20 °C • 5 days • Index HT™ F12: 0.2 M Sodium chloride, 0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3,350	

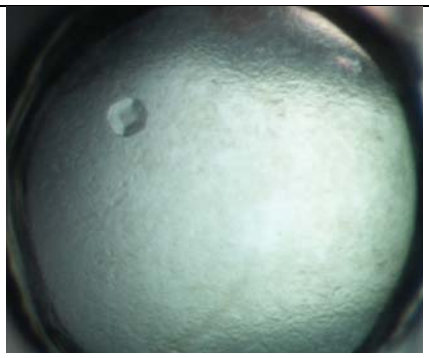
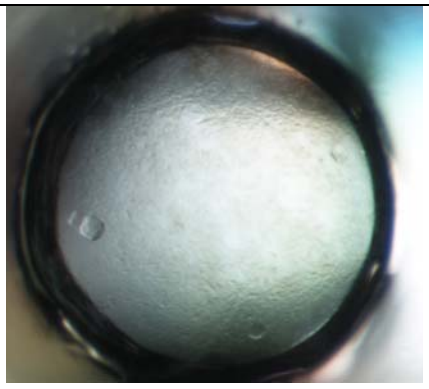
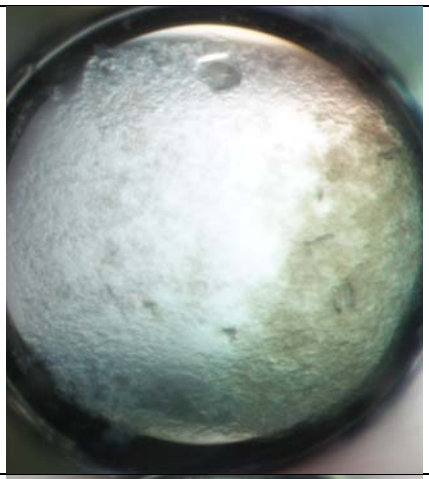
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 weeks • Index HT™ G6: 0.2 M Ammonium acetate, 0.1 M BIS-TRIS pH 5.5, 25% w/v Polyethylene glycol 3,350 • Frozen in liquid nitrogen • 3.9 Å	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 weeks • Index HT™ G7: 0.2 M Ammonium acetate, 0.1 M BIS-TRIS pH 6.5, 25% w/v Polyethylene glycol 3,350 • Frozen in liquid nitrogen • 6 Å	
• 100 µM • 400 nL • Sitting drop • 20 °C • 5 days • Index HT™ G8: 0.2 M Ammonium acetate, 0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3,350	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 weeks • Index HT™ H8: 0.2 M Ammonium acetate, 0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3,350	

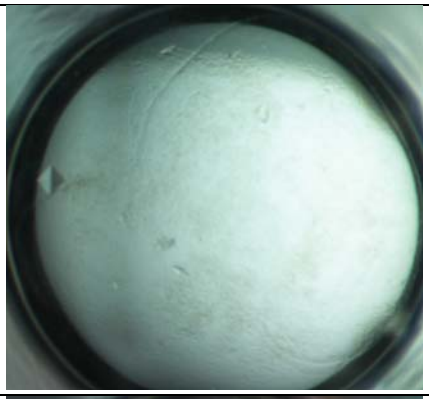
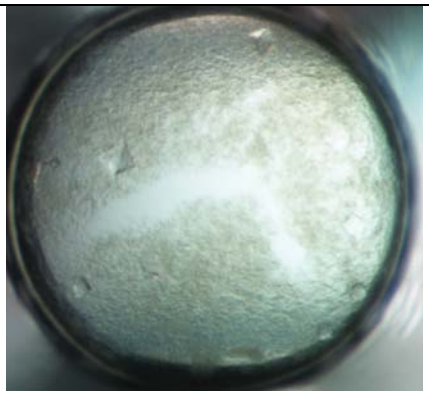
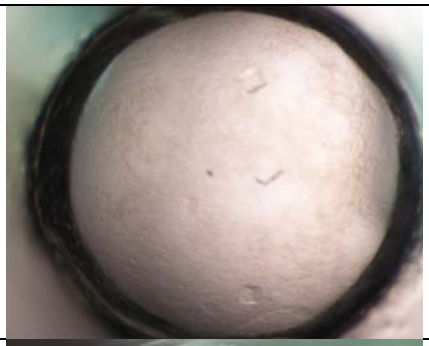
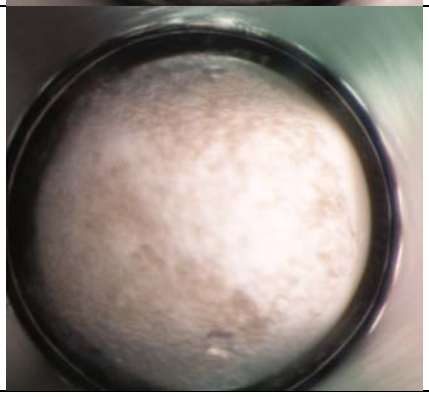
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 weeks • Nuc-Pro HTS D7: 0.2 M Potassium chloride, 0.05 M MES Sodium salt pH 6.5, 10% w/v Polyethylene glycol 8,000	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 weeks • PEGRx HT™ A10: 0.1 M Sodium citrate tribasic dihydrate pH 5.0, 30% v/v Polyethylene glycol monomethyl ether 550	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 weeks • PEGRx HT™ B11: 0.1 M MES monohydrate pH 6.0, 20% w/v Polyethylene glycol monomethyl ether 2,000	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 weeks • PEGRx HT™ G7: 2% v/v Polyethylene glycol 400, 0.1 M Imidazole pH 7.0, 24% w/v Polyethylene glycol monomethyl ether 5,000	

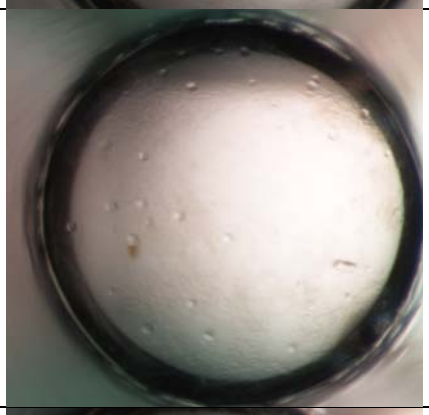
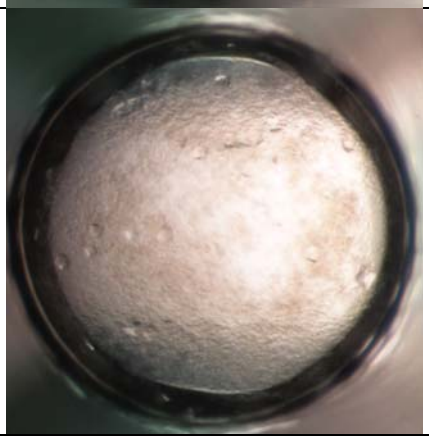
• 100 µM • 2 µL • Hanging drop • 20 °C • 4 weeks • 0.1 M BIS-TRIS pH 6.5, 28% w/v Polyethylene glycol monomethyl ether 2,000	
• 100 µM • 2 µL • Hanging drop • 20 °C • 4 weeks • 0.2 M Ammonium acetate, 0.1 M BIS-TRIS pH 6.5, 25% w/v Polyethylene glycol 3,350	
DiLCrz_TLR3	**Picture**
• 72 µM • 200 nL • Sitting drop • 20 °C • 5 weeks • Natrix HT™ E11: 0.075 M NaCl, 0.002 M Calcium chloride dihydrate, 0.05 M Na cacodylate trihydrate pH 6.0, 30% w/v 1,6-Hexanediol, 0.0005 M Spermine	
• 72 µM • 400 nL • Sitting drop • 20 °C • 3 weeks • Natrix HT™ E11: 0.075 M NaCl, 0.002 M Calcium chloride dihydrate, 0.05 M Na cacodylate trihydrate pH 6.0, 30% w/v 1,6-Hexanediol, 0.0005 M Spermine • Frozen in liquid ethane • 25 Å	

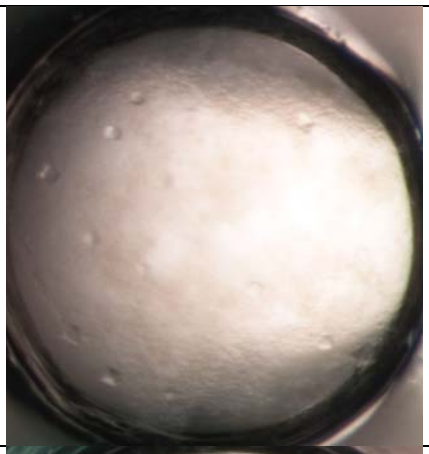
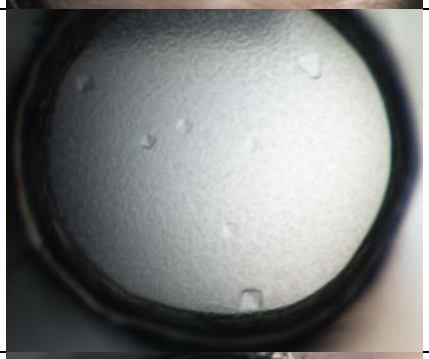
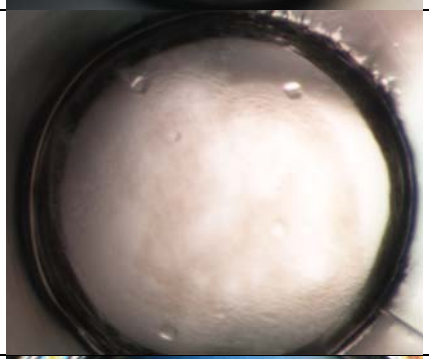
• 140 µM • 400 nL • Sitting drop • 20 °C • 4 months • 60 mM Magnesium choride, 30 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 15% v/v Polyethylene glycol 400	
• 140 µM • 400 nL • Sitting drop • 20 °C • 4 months • 60 mM Magnesium choride, 60 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 15% v/v Polyethylene glycol 400	
• 72 µM • 400 nL • Sitting drop • 20 °C • 4 months • 60 mM Magnesium choride, 90 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 15% v/v Polyethylene glycol 400 • Frozen in liquid ethane • 25 Å	
• 140 µM • 400 nL • Sitting drop • 20 °C • 4 months • 60 mM Magnesium choride, 90 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 15% v/v Polyethylene glycol 400	

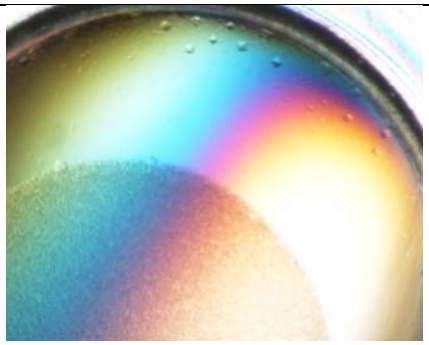
• 140 µM • 400 nL • Sitting drop • 20 °C • 4 months • 60 mM Magnesium choride, 910 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 15% v/v Polyethylene glycol 400	
• 72 µM • 400 nL • Sitting drop • 20 °C • 4 months • 60 mM Magnesium choride, 260 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 15% v/v Polyethylene glycol 400 • Frozen in liquid ethane • 25 Å	
• 140 µM • 400 nL • Sitting drop • 20 °C • 10 weeks • 60 mM Magnesium choride, 30 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 20% v/v Polyethylene glycol 400	
• 140 µM • 400 nL • Sitting drop • 20 °C • 10 weeks • 60 mM Magnesium choride, 60 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 20% v/v Polyethylene glycol 400 • Frozen in liquid ethane • 25 Å	

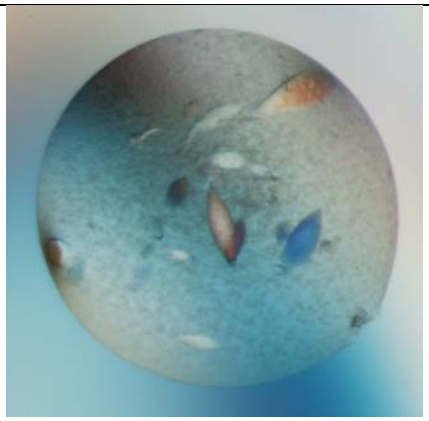
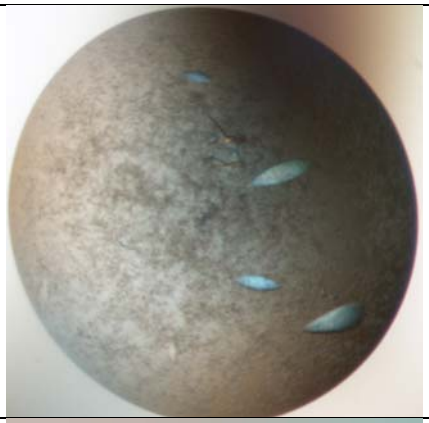
• 72 µM • 400 nL • Sitting drop • 20 °C • 10 weeks • 60 mM Magnesium choride, 90 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 20% v/v Polyethylene glycol 400 • Frozen in liquid ethane • 25 Å	
• 72 µM • 400 nL • Sitting drop • 20 °C • 10 weeks • 60 mM Magnesium choride, 190 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 20% v/v Polyethylene glycol 400 • Frozen in liquid ethane • 20 Å	
• 140 µM • 400 nL • Sitting drop • 20 °C • 10 weeks • 60 mM Magnesium choride, 190 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 20% v/v Polyethylene glycol 400	
• 40 µM • 400 nL • Sitting drop • 20 °C • 10 weeks • 60 mM Magnesium choride, 130 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 20% v/v Polyethylene glycol 400	

• 40 µM • 400 nL • Sitting drop • 20 °C • 10 weeks • 30 mM Magnesium choride, 130 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 20% v/v Polyethylene glycol 400 • Frozen in liquid ethane • 19.7 Å	
• 140 µM • 400 nL • Sitting drop • 20 °C • 10 weeks • 60 mM Magnesium choride, 260 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 25% v/v Polyethylene glycol 400 • Frozen in liquid ethane • 25 Å	
• 40 µM • 400 nL • Sitting drop • 20 °C • 10 weeks • 60 mM Magnesium choride, 90 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 25% v/v Polyethylene glycol 400	
• 72 µM • 400 nL • Sitting drop • 20 °C • 10 weeks • 60 mM Magnesium choride, 90 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 25% v/v Polyethylene glycol 400	
• 140 µM • 400 nL • Sitting drop • 20 °C • 10 weeks • 60 mM Magnesium choride, 90 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 25% v/v Polyethylene glycol 400	

• 40 µM • 400 nL • Sitting drop • 20 °C • 10 weeks • 60 mM Magnesium choride, 190 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 25% v/v Polyethylene glycol 400	
• 40 µM • 400 nL • Sitting drop • 20 °C • 10 weeks • 130 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 25% v/v Polyethylene glycol 400	
• 72 µM • 400 nL • Sitting drop • 20 °C • 10 weeks • 130 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 25% v/v Polyethylene glycol 400	
• 140 µM • 400 nL • Sitting drop • 20 °C • 10 weeks • 130 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 25% v/v Polyethylene glycol 400	

• 72 µM • 400 nL • Sitting drop • 20 °C • 10 weeks • 15 mM Magnesium choride, 130 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 25% v/v Polyethylene glycol 400	
• 140 µM • 400 nL • Sitting drop • 20 °C • 10 weeks • 15 mM Magnesium choride, 130 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 25% v/v Polyethylene glycol 400	
• 72 µM • 400 nL • Sitting drop • 20 °C • 10 weeks • 15 mM Magnesium choride, 130 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 5% v/v Polyethylene glycol 400	
• 72 µM • 400 nL • Sitting drop • 20 °C • 10 weeks • 30 mM Magnesium choride, 130 mM Sodium choride, 100 mM Sodium citrate pH 5.6, 25% v/v Polyethylene glycol 400	
• 48 µM • 2 µL • Hanging drop • 20 °C • 7 weeks • Natrix HT™ E11: 0.075 M NaCl, 0.002 M Calcium chloride dihydrate, 0.05 M Na cacodylate trihydrate pH 6.0, 30% w/v 1,6-Hexanediol, 0.0005 M Spermine	

• 72 µM • 2 µL • Hanging drop • 20 °C • 7 weeks • Natrix HT™ E11: 0.075 M NaCl, 0.002 M Calcium chloride dihydrate, 0.05 M Na cacodylate trihydrate pH 6.0, 30% w/v 1,6-Hexanediol, 0.0005 M Spermine	
• 72 µM • 2 µL • Hanging drop • 20 °C • 7 weeks • Natrix HT™ E11: 0.075 M NaCl, 0.002 M Calcium chloride dihydrate, 0.05 M Na cacodylate trihydrate pH 6.0, 30% w/v 1,6-Hexanediol, 0.0005 M Spermine • Frozen in liquid ethane • 25 Å	
DiLCrz_cp_P2TL9_P8TR17	**Picture**
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 months • Index HT™ E9: 0.05 M Ammonium sulfate, 0.05 M BIS-TRIS pH 6.5, 30% v/v Pentaerythritol ethoxylate (15/4 EO/OH)	
DiLCrz_cp_P2TR18_P8TL12	**Picture**

• 100 µM • 400 nL • Sitting drop • 20 °C • 6 months • Natrix HT™ A11: 0.01 M Magnesium chloride hexahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.0, 1.0 M Lithium sulfate monohydrate • Frozen in liquid ethane • 15 Å	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 months • Natrix HT™ B6: 0.01 M Magnesium acetate tetrahydrate, 0.05 M Sodium cacodylate trihydrate pH 6.5, 1.3 M Lithium sulfate monohydrate • Frozen in liquid ethane • 4.3 Å	
• 100 µM • 400 nL • Sitting drop • 20 °C • 8 weeks • Natrix HT™ C4: 0.05 M Magnesium sulfate hydrate, 0.05 M HEPES sodium pH 7.0, 1.6 M Lithium sulfate monohydrate	
• 100 µM • 400 nL • Sitting drop • 20 °C • 6 months • Index HT™ E9: 0.05 M Ammonium sulfate, 0.05 M BIS-TRIS pH 6.5, 30% v/v Pentaerythritol ethoxylate (15/4 EO/OH)