

Université de Montréal

Études d'ingénierie du ribozyme VS de *Neurospora*

par

Julie Lacroix-Labonté

Département de Biochimie et Médecine Moléculaire

Faculté de Médecine

Thèse présentée à la Faculté de Médecine
en vue de l'obtention du grade de Philosophiae Doctor (Ph.D)
en Biochimie

Décembre 2015

© Julie Lacroix-Labonté, 2015

Résumé

Les ribozymes sont des ARN catalytiques fréquemment exploités pour le développement d'outils biochimiques et d'agents thérapeutiques. Ils sont particulièrement intéressants pour effectuer l'inactivation de gènes, en permettant la dégradation d'ARNm ou d'ARN viraux associés à des maladies. Les ribozymes les plus utilisés en ce moment pour le développement d'agents thérapeutiques sont les ribozymes hammerhead et hairpin, qui permettent la reconnaissance spécifique d'ARN simple brin par la formation de structures secondaires stables. *In vivo*, la majorité des ARN adoptent des structures secondaires et tertiaires complexes et les régions simples brins sont parfois difficiles d'accès. Il serait intéressant de pouvoir cibler des ARN repliés et un motif d'ARN intéressant à cibler est la tige-boucle d'ARN qui peut être importante dans le repliement global des ARN et pour accomplir des fonctions biologiques.

Le ribozyme VS de *Neurospora* fait la reconnaissance de son substrat replié en tige-boucle de façon spécifique par une interaction *kissing-loop*, mais il n'a jamais été exploité pour faire la reconnaissance d'un ARN cible très différent de son substrat naturel. Le but des travaux présentés dans cette thèse est de déterminer si le ribozyme VS possède l'adaptabilité nécessaire pour l'ingénierie de ribozymes qui clivent des ARN cibles différents du substrat naturel. Dans le cadre de cette thèse, le ribozyme VS a été modifié pour l'adapter à différents substrats et des études de cinétiques ont été réalisées pour évaluer l'impact de ces modifications sur l'activité de clivage du ribozyme. Dans un premier temps, le ribozyme a été modifié pour faire la reconnaissance et le clivage de substrats possédant différentes longueurs de tiges Ib. Le ribozyme a été adapté avec succès à ces substrats de différentes longueurs de tige Ib, avec une activité qui est similaire à celle du ribozyme avec un substrat sans modification. Dans un deuxième temps, c'est l'interaction *kissing-loop* I/V du ribozyme qui a été substituée de façon rationnelle, dans le but de savoir si un ribozyme VS mutant peut reconnaître et cliver un substrat ayant une boucle différente de celle de son substrat naturel. L'interaction *kissing-loop* I/V a été substituée pour les interactions *kissing-loop* TAR/TAR* de l'ARN du VIH-1 et L22/L88 de l'ARN 23S de *Deinococcus radiodurans*. La réaction de

clivage des ribozymes comportant ces nouvelles interactions *kissing-loop* est toujours observée, mais avec une activité diminuée. Finalement, la sélection *in vitro* (SELEX) de ribozymes a été effectuée pour permettre un clivage plus efficace d'un substrat mutant avec une nouvelle boucle. Le SELEX a permis la sélection d'un ribozyme qui clive un substrat avec une boucle terminale mutée pour celle de l'ARN TAR du VIH-1 et cela avec une activité de clivage très efficace. L'ensemble de ces études démontre que le ribozyme VS peut être modifié de diverses façons pour la reconnaissance spécifique de différents substrats, tout en conservant une bonne activité de clivage. Ces résultats montrent le grand potentiel d'ingénierie du ribozyme VS et sont prometteurs pour la poursuite d'études d'ingénierie du ribozyme VS, en vue du clivage d'ARN cibles repliés en tige-boucle complètement différents du substrat naturel du ribozyme VS.

Mots-clés : Ribozyme VS de *Neurospora*, ingénierie d'ARN, interaction *kissing-loop*, études de cinétiques, SHAPE, SELEX

Abstract

Ribozymes are catalytic RNAs frequently exploited for the development of biochemistry tools and therapeutic agents. They are particularly interesting in gene inactivation strategies for the degradation of mRNA and viral RNA genome. Currently, the most common ribozymes used in the development of therapeutic agents are the hammerhead and hairpin ribozymes, which can specifically recognize and cleave target single-stranded RNAs through the formation of stable secondary structures. *In vivo*, single-stranded RNAs can be difficult to target because most RNA adopt complex secondary and tertiary structures. It could be worthwhile to target folded RNA motifs, and an interesting target is the stem-loop structure because stem-loops are important for the overall folding of RNA molecules and they can perform important biological functions.

The *Neurospora* VS ribozyme recognizes its stem-loop folded substrate via a specific *kissing-loop* interaction, but it has never been exploited for the recognition of target RNA very different from its natural substrate. The goal of the work presented in this thesis is to determine whether the VS ribozyme possesses the necessary adaptability for engineering ribozymes that target RNAs different from its natural substrate. For this thesis, the VS ribozyme was adapted for the recognition of different substrates and kinetic studies were performed to evaluate the effect of these modifications on the cleavage activity. In a first study, the VS ribozyme was modified to recognize and cleave substrates with different stem I_b lengths. The VS ribozyme was successfully adapted to these substrates of different lengths with a cleavage activity similar to the unmodified ribozyme and substrate. In a second study, the I/V *kissing-loop* interaction was substituted by rational design, in order to establish if the VS ribozyme variants could recognize and cleave a substrate with a different loop than its natural substrate. The I/V *kissing-loop* interaction was substituted for the HIV-1 TAR/TAR* and the *Deinococcus radiodurans* RNA large rRNA subunit L22/L88 *kissing-loop* interactions. The cleavage reaction was observed for the ribozymes with these new interactions, but with reduced activity. Finally, *in vitro* selection (SELEX) was used to enable more efficient cleavage of a mutant substrate with a new loop. SELEX experiments enabled

the selection of ribozyme variants that cleave a substrate with a terminal loop mutated to that of the HIV-1 TAR RNA with a very efficient cleavage activity. All of the studies presented in this thesis show that the VS ribozyme can be modified in various ways for the specific recognition of different substrates, while maintaining efficient cleavage activity. These results demonstrate the great potential of VS ribozyme engineering and are very promising for further engineering studies of VS-derived ribozymes for the cleavage of stem-loop folded target RNA completely different from its natural VS ribozyme substrate.

Keywords : *Neurospora* VS ribozyme, RNA engineering, *kissing-loop interaction*, kinetic studies, SHAPE, SELEX

Table des matières

Résumé.....	ii
Abstract.....	iv
Table des matières.....	vi
Liste des tableaux.....	ix
Liste des figures.....	x
Liste des sigles.....	xiii
Liste des abréviations.....	xv
Remerciements.....	xviii
1.0 Introduction.....	1
1.1 L'importance de l'ARN.....	2
1.1.1 ARN non codants.....	2
1.1.2 Le monde à ARN.....	4
1.1.3 Ribozymes.....	5
1.2 Le ribozyme VS de <i>Neurospora</i>	16
1.2.1 Structure secondaire.....	17
1.2.2 Interaction <i>kissing-loop</i> I/V.....	18
1.2.3 Activation du substrat.....	21
1.2.4 Réaction de clivage.....	24
1.2.5 Études structurales.....	28
1.2.6 Mutagenèse du ribozyme VS.....	35
1.3 Ingénierie des ribozymes.....	37
1.3.1 Ingénierie rationnelle.....	39
1.3.2 Ingénierie par SELEX.....	40
1.3.3 Ingénierie computationnelle.....	41
1.3.4 Ingénierie des ribozymes autoclivants.....	42
1.4 Objectifs de la thèse.....	44

1.5	Techniques utilisées	47
1.5.1	Purification de l'ARN	47
1.5.2	Cinétiques enzymatiques en <i>single-turnover</i>	47
1.5.3	SHAPE	50
1.5.4	SELEX	50
2.0	Études de compensation de la longueur des hélices chez le ribozyme VS	52
	Mise en contexte	53
	ABSTRACT	56
	INTRODUCTION	57
	MATERIALS AND METHODS	59
	RESULTS	64
	DISCUSSION	69
	ACKNOWLEDGMENTS	72
	REFERENCES	73
	TABLES	76
	FIGURES	77
	SUPPLEMENTARY INFORMATION	81
3.0	Substitution de l'interaction <i>kissing-loop</i> chez le ribozyme VS	84
	Mise en contexte	85
	ABSTRACT	88
	INTRODUCTION	89
	MATERIALS AND METHODS	92
	RESULTS	95
	DISCUSSION	101
	ACKNOWLEDGEMENTS	105
	REFERENCES	106
	TABLES	109
	FIGURES	110
	SUPPLEMENTARY INFORMATION	115

4.0	Études de SELEX pour la reconnaissance de l'ARN TAR du VIH-1 par le ribozyme VS	122
	Mise en contexte	123
	ABSTRACT	126
	INTRODUCTION	127
	RESULTS AND DISCUSSION	130
	CONCLUDING REMARKS	135
	MATERIALS AND METHODS	136
	ACKNOWLEDGMENTS	137
	REFERENCES	138
	TABLE	140
	FIGURES	141
5.0	Discussion et perspectives	144
5.1	Ingénierie du ribozyme VS	145
5.1.1	Ingénierie rationnelle	146
5.1.2	Ingénierie par SELEX	148
5.1.3	Règles de l'ingénierie	150
5.2	Comparaison avec l'ingénierie des ribozymes autoclivants	154
5.3	Perspectives	155
5.3.1	Ingénierie du ribozyme VS pour le clivage d'un ARN cible	155
5.3.2	ARN cibles intéressants pour le ribozyme VS	157
5.3.3	Utilisation du ribozyme VS <i>in vivo</i>	159
5.3.4	Utilisation du ribozyme VS comme outil biochimique	162
5.4	Conclusion générale	164
	Bibliographie	i
Annexe 1.0	Étude du rôle des nucléotides de la boucle V dans la reconnaissance du substrat par le ribozyme VS	i
	Mise en contexte	ii

Liste des tableaux

Table II.I Summary of kinetic data for R_{θ} ^a	76
Table II.II Summary of multiple-substrate kinetic data ^a	76
Table III.I. Filtered results of the bioinformatics search to identify KLIs from the PDB.....	109
Supplementary Table III.I	115
Supplementary Table III.II. Values of k_{obs} (min^{-1}) for selected S/R pairs measured with $[S] = 0.250$ nM and $[R] = 0.5$ μM under standard conditions. ^a	115
Table IV.I Kinetic results for the S/R pairs ¹	140

Liste des figures

Figure 1.1 Diverses fonctions des ARNnc dans les domaines de la vie.....	3
Figure 1.2 Mécanisme de clivage des petits ribozymes autoclivants.	7
Figure 1.3 Caractéristiques structurales et modèles du site de clivage proposés pour les ribozymes autoclivants.....	8
Figure 1.4 Structures secondaires des ribozymes découverts récemment.	13
Figure 1.5 Mécanisme de réplication du ribozyme VS.....	17
Figure 1.6 Structure secondaire du ribozyme VS.	18
Figure 1.7 Structure RMN d'un complexe I/V.	20
Figure 1.8 Réarrangement de la structure de SLI lors de la liaison de la tige-boucle V.	22
Figure 1.9 Comparaison des structures des conformations active et inactive de la boucle interne de SLI.....	23
Figure 1.10 Mécanisme proposé de clivage au site actif du ribozyme VS.	25
Figure 1.11 Représentation schématique des structures des sites actifs des ribozymes hairpin et VS.	26
Figure 1.12 Modèles de la structure tertiaire du ribozyme VS.	29
Figure 1.13 Modèles des structures tertiaires du ribozyme VS basés sur le FRET, le SAXS et la RMN.....	30
Figure 1.14 Structure cristalline de la forme dimérique du ribozyme VS.	33
Figure 1.15 Séquences et structures secondaires proposées des ribozymes VS qui clivent rapidement.....	37
Figure 1.16 Séquences des ribozymes hammerhead et hairpin pour l'ingénierie.....	43
Figure 1.17 Exemple d'études de cinétiques enzymatiques du ribozyme VS.	48
Figure 1.18 Études de SELEX du ribozyme VS.....	51
Figure 2.1 Primary and secondary structures of the substrate/ribozyme (S/R) pairs characterized here.	77
Figure 2.2 Single-substrate kinetic analysis of R_0	78
Figure 2.3 Simultaneous kinetic characterization of R_{3bp} with multiple substrates.	79
Figure 2.4 Identification of SLI/SLV models compatible with the kinetic data.....	79

Figure 2.5 Model of an SLI/SLV complex compatible the kinetic data (common-core group 13).	80
Supplementary Figure 2.1 Conformational analysis of the substrates used in this study by native gel electrophoresis.	81
Supplementary Figure 2.2 Identification of a common cleavage site for all functioning S/R pairs.	82
Supplementary Figure 2.3 Flowchart describing the stepwise procedure used to generate three-dimensional models of the SLI/SLV complex and select those compatible with kinetic data.	83
Figure 3.1 Rational engineering of KLIs for substrate recognition by the <i>Neurospora</i> VS ribozyme.	110
Figure 3.2 Schematics and thermodynamic stability of selected KLIs.	111
Figure 3.3 Sequences and kinetic data of the S_{TAR}/R_{TAR^*} variants.	112
Figure 3.4 SHAPE analysis of the R_{TAR^*} ribozyme variants.	113
Figure 3.5 Three-dimensional models of substrate/ribozyme complexes.	114
Supplementary Figure 3.1 Schematic representation of the computational workflow used to identify and characterize KLIs present in the PDB.	116
Supplementary Figure 3.2 ITC studies of the L22/L88 interaction.	117
Supplementary Figure 3.3 Sequences and kinetic data of the S_{L88}/R_{L22} variants.	118
Supplementary Figure 3.4 Sequences and kinetic data of the S_{L22}/R_{L88} variants.	119
Supplementary Figure 3.5 Conformational analysis of the substrate variants used in this study.	120
Supplementary Figure 3.6 SHAPE analysis of the R_0 ribozyme and related variants.	121
Figure 4.1. Sequences and secondary structures of the VS ribozyme variants used in this study.	141
Figure 4.2 Evolution of the ribozyme cleavage activity for the different rounds of selection.	142
Figure 4.3 Sequence alignment of the randomized region of the SLV loop for all the selected ribozymes characterized kinetically.	142
Figure 4.4 Schematics of kissing-loop interactions.	143
Figure 5.1 Résumé des expériences d'ingénierie du ribozyme VS.	147
Figure 5.2 Consensus des modifications du substrat SLI sans changement compensatoire.	152

Figure 5.3 Expériences de SELEX pour la reconnaissance d'un substrat entièrement modifié.
..... 156

Figure 5.4 ARN intéressants pour le clivage par le ribozyme VS 159

Figure 5.5 Modifications au sucre des nucléotides pour augmenter la résistance aux nucléases
chez le ribozyme hammerhead..... 161

Figure 5.6 Exemple d'un ribozyme VS contrôlé par un ligand. 163

Liste des sigles

ADN : acide désoxyribonucléique

ARN : acide ribonucléique

ARNi : ARN interférent

ARNlnc : long ARN non codant (*long non-coding RNA*)

ARNm : ARN messenger

ARNnc : ARN non codant

ARNr : ARN ribosomal

ARNsh : petit ARN en tige-boucle (*small hairpin RNA*)

ARNsn : petit ARN nucléaire (*small nuclear RNA*)

ARNsno : petit ARN nucléolaire (*small nucleolar RNA*)

ARNt : ARN de transfert

CPEB3 : protéine 3 de liaison de l'élément cytoplasmique de polyadénylation (*cytoplasmic polyadenylation element-binding protein*)

CRISPR : courtes répétitions palindromiques groupées et régulièrement espacées (*clustered regularly interspaced short palindromic repeats*)

FRET : transfert d'énergie par résonance Förster (*Förster resonance energy transfer*)

GlcN6P : glucosamine-6-phosphate

glmS : synthase glucosamine-6-phosphate (*glucosamine-6-phosphate synthase*)

HDV : virus de l'hépatite delta

ITC : titrage calorimétrique isotherme (*isothermal titration calorimetry*)

LNA : acide nucléique bloqué (*locked nucleic acid*)

miARN : micro ARN

NMIA : anhydride *N*-méthylisatoïque (*N-methylisatoic anhydride*)

PCR : réaction en chaîne par polymérase (*polymerase chain reaction*)

pb : paire de base

PNA : acide nucléique peptidique (*peptide nucleic acid*)

RNase A : ribonucléase A

RNase P : ribonucléase P

SAXS : diffusion des rayons-X aux petits angles (*small angle X-ray scattering*)

sARN : petit ARN (*small RNA*)

SELEX : évolution systématique de ligands par enrichissement exponentiel (*systematic evolution of ligands by exponential enrichment*)

SHAPE : acylation sélective des 2'-hydroxyle analysée par extension d'amorce (*selective 2'-hydroxyl acylation analyzed by primer extension*)

siARN : petit ARN interférent

TAR : *trans-activation response element*

UTR : région non traduite (*untranslated region*)

VIH : virus de l'immunodéficience humaine

VS : *Varkud Satellite*

W-C : Watson-Crick

Xist : transcrit spécifique du chromosome X inactif (*X inactive specific transcript*)

Liste des abréviations

A : adénine

C : cytosine

G : guanine

H : hydrogène

M : molaire

Mg²⁺ : ion magnésium

Min : minute

N : n'importe quel nucléotide

Nt : nucléotide

O : oxygène

OH : hydroxyle

R : purine

U : uridine

Y : pyrimidine

À ma famille,

qui m'a toujours soutenue dans tous mes projets.

Remerciements

Je tiens à remercier en premier lieu ma directrice de recherche Pascale Legault de m'avoir accepté dans son laboratoire pour un stage d'été au baccalauréat qui s'est prolongé en projet de doctorat. Je la remercie de m'avoir encadrée dans les différents projets de ma thèse de doctorat et de m'avoir inculqué de bonnes pratiques de recherches. J'ai beaucoup appris sur la conception de projets recherches, l'écriture des protocoles, l'analyse et la présentation des résultats, ainsi que la rédaction d'articles.

Un merci spécial à Philippe Lampron qui s'est occupé de ma formation lors de mon entrée au laboratoire et qui m'a initié à la vie de laboratoire en m'apprenant les différentes techniques de base que j'ai utilisées durant mes années de doctorat. Je voudrais aussi remercier tous les membres actuels et passés des laboratoires Legault-Omichinski, que j'ai côtoyés avec beaucoup de plaisirs au laboratoire depuis le début de ma formation et lors de nos nombreuses activités scientifiques et non scientifiques à l'extérieur du laboratoire. Je remercie les membres du laboratoire qui étaient présents lors de mon stage d'été Philippe, Patricia, Geneviève, Alexandre, Caroline et Chantale qui m'ont inspirés à poursuivre mes études au doctorat. Je remercie Geneviève DT, Pierre, Elyse, Dominique, Patricia, Alix, Vanessa, Ao, Jonathan, Nicolas, Alexandre et Éric qui ont contribué à mettre de la vie dans le laboratoire et pour faire en sorte qu'il était agréable de venir y travailler. Un merci spécial à Nicolas, mon collaborateur principal pour cette thèse, mais aussi à Patricia avec qui j'ai collaboré pour mon premier article et à Pierre pour les nombreuses discussions sur l'ingénierie de ribozyme et à qui je lègue la suite de ce projet.

En dernier lieu, je tiens à remercier ma famille et mes amis qui m'ont soutenue durant les nombreuses années de cette thèse et en m'encourageant à poursuivre malgré les difficultés. Je remercie particulièrement mon mari Simon, ma sœur Isabelle et mes parents Marie et Sylvain qui sont toujours présents pour moi et qui s'efforcent à essayer de comprendre de leur mieux mes expériences sur le ribozyme VS.

1.0 Introduction

1.1 L'importance de l'ARN

Le dogme central de la biologie a longtemps été défini par l'ADN qui est le porteur de l'information génétique, qui est transcrit en ARN complémentaire et l'ARN est ensuite traduit en protéines, qui sont responsables de plusieurs fonctions cellulaires. Selon ce dogme, l'ARN joue un rôle d'intermédiaire entre le génome et la synthèse de protéine. Par contre, il est maintenant connu que l'ARN ne se limite pas à n'être qu'un simple intermédiaire et qu'il occupe un rôle central dans de nombreux processus biologiques.

1.1.1 ARN non codants

Auparavant, les protéines étaient considérées comme les seules biomolécules responsables des fonctions catalytiques et structurales. Par la suite, il a été observé que les régions de l'ADN entre les séquences codantes pour les protéines sont nombreuses à être transcrites en ARN. Ces ARN non codants (ARNnc), qui sont passés inaperçus durant de nombreuses années, ont été grandement sous-estimés. Aujourd'hui, il y a des milliers d'ARNnc identifiés dans diverses sphères de la biochimie et ils sont associés à diverses fonctions (figure 1.1). Parmi les ARNnc les plus connus, il y a les ARN ribosomiaux (ARNr) et les ARN de transfert (ARNt), qui sont impliqués dans la traduction des ARN messagers (ARNm) en protéines, ainsi que la RNase P qui est impliquée dans la maturation des ARNt [1-3]. On retrouve aussi les ARNnc impliqués dans l'épissage de l'ARN (ARNsn, ARNsno, introns du group I et II) et la structure des chromosomes (ARN de la télomérase, ARN Xist) [4-9]. Une autre sphère importante des ARNnc est la régulation des gènes qui peut s'effectuer en *trans* (ARNi, miARN et ARN antisens) ou en *cis*, en agissant dans la région non transcrite des ARNm (riborégulateur) [10-14]. Les ARNnc sont même impliqués dans la défense du génome (CRISPR et ARN interagissant avec les protéines piwi) [15, 16]. Le projet ENCODE (*encyclopedia of DNA elements*) qui étudie les fonctions des gènes humains, a identifié 9600 longs ARN non codants (ARNlnc) de plus de 200 nucléotides en 2014, mais l'attribution individuelle de leurs fonctions est une tâche immense qui demandera du temps [17, 18].

L'étendue des fonctions des ARNnc n'est pas encore totalement connue et il reste encore beaucoup à découvrir dans ce domaine [19, 20].

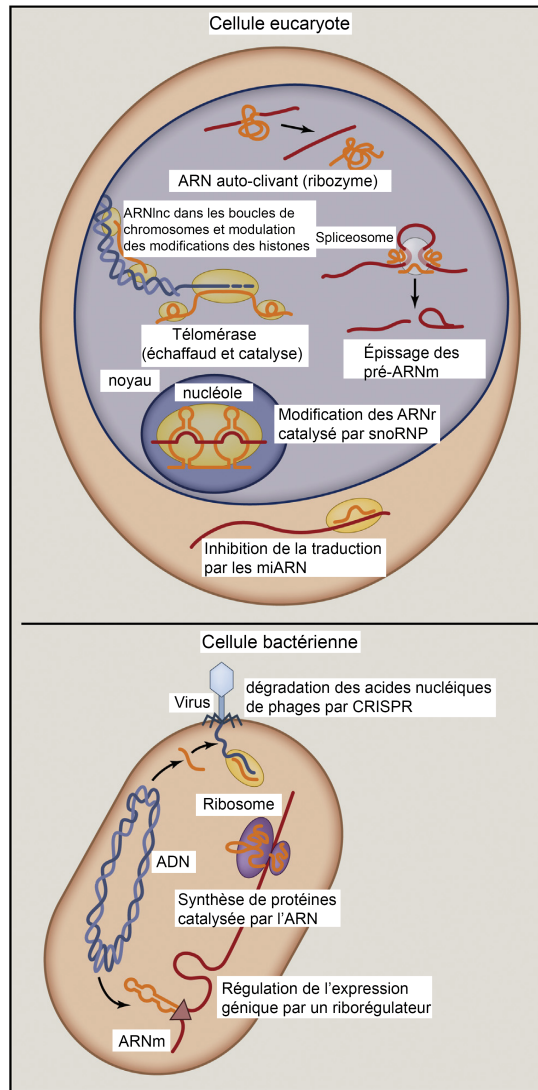


Figure 1.1 Diverses fonctions des ARNnc dans les domaines de la vie.

Les ARNnc ont des fonctions diverses dans la régulation de l'expression génique comme dans la transcription, l'épissage, la traduction et en contribuant à l'organisation et la stabilité du génome. Les brins orange représentent les ARNnc exécutant l'action indiquée, les brins rouges sont les ARN ciblés par les ARNnc et les brins bleus représentent l'ADN. Le triangle représente un métabolite interagissant avec le riborégulateur et les ovales représentent les complexes ARN-protéines comme le spliceosome en blanc et le ribosome en mauve. (Adaptée de [19]).

Plusieurs ARNnc vont agir avec la collaboration de protéines dans des complexes ARN-protéine comme le ribosome et le spliceosome, mais il y a des ARNnc qui peuvent catalyser des réactions chimiques par eux-mêmes, sans l'aide de protéine, ce qui les qualifie d'enzymes à ARN ou plus spécifiquement de ribozymes [21].

1.1.2 Le monde à ARN

Une question à laquelle il n'y a toujours pas de réponse est comment la vie est apparue sur terre. Il semble peu probable que le système complexe de transcription et de traduction, que l'on retrouve aujourd'hui avec de nombreuses molécules impliquées ait pu émerger dans des conditions de vie minimales. Pour décrire comment la vie a pu apparaître dans les conditions d'un monde primitif, l'hypothèse du monde à ARN a été proposée dans les années 1960 par Woese, Orgel et Crick [2, 22, 23]. Le monde à ARN est l'hypothèse selon laquelle au tout début de la vie sur terre, il y avait un monde dans lequel l'ARN ou une molécule chimiquement semblable constitue la substance primaire vivante. Dans ce monde, l'ARN joue à la fois les rôles d'enzyme et de génome. Il est logique de penser à l'ARN pour ces rôles, puisqu'il occupe aujourd'hui une place centrale dans le dogme de la vie. La découverte des ARN catalytiques dans les années 1980 [1, 24] et plus tard, la découverte que la sous-unité responsable de l'activité chez le ribosome est composée uniquement d'ARN [3] apportent des preuves supplémentaires pour soutenir l'hypothèse d'un monde dans lequel l'ARN joue un rôle central. Par la suite, les expériences de sélection *in vitro* ont démontré que l'ARN peut avoir un répertoire de fonctions très variées et compatibles avec le monde à ARN [25-29]. Par contre, il est difficile de déterminer comment la vie est réellement apparue et l'hypothèse du monde à ARN représente l'un des mécanismes possibles pour l'apparition de la vie sur terre [30].

L'ARN est un domaine de recherche important, puisque sa présence est essentielle dans de nombreuses sphères de la biologie, en plus d'avoir pu jouer un rôle dans l'apparition de la vie. L'étude des ARN catalytiques comme les ribozymes est particulièrement

intéressante pour découvrir comment les motifs d'ARN ont développé des activités catalytiques et qu'elles sont les fonctions liées à ses propriétés.

1.1.3 Ribozymes

Les ribozymes sont des molécules avec une activité enzymatique dont le site catalytique est composé uniquement d'ARN, même s'ils sont parfois en complexe avec des protéines. Leur structure leur permet d'accomplir une activité catalytique sans l'assistance d'une protéine. Les ribozymes représentent un bon système pour faire l'étude de structure fonction de l'ARN, puisque des changements dans la structure du ribozyme ont des répercussions sur l'activité de catalyse. L'étude de ces répercussions permet de mieux comprendre les fonctions liées aux motifs de l'ARN, mais aussi de mieux comprendre les mécanismes catalytiques et les fonctions liées à leur activité. Par la suite, ces informations peuvent être mises à profit pour faire l'ingénierie de nouveaux ribozymes possédant de nouvelles propriétés ou encore, pour améliorer l'activité de catalyse de ribozymes existants.

1.1.3.1 Ribozymes naturels

Le premier ribozyme a été découvert 1982 par le Dr Thomas R. Cech en étudiant l'épissage des introns dans les gènes ARNr de certaines espèces de *Tetrahymena thermophila* [24]. Durant l'épissage, l'intron est excisé en une molécule linéaire et ensuite converti sous une forme circulaire. Son équipe de recherche a découvert que pour ces fonctions de clivage et de ligation, l'ARN n'avait pas besoin de l'assistance de protéines : c'est la structure de la molécule d'ARN elle-même qui est responsable de l'activité [24]. Peu de temps après, en 1983, le Dr Sidney Altman fait la découverte d'un deuxième ribozyme, la RNase P, qui est responsable de la maturation des ARNt en hydrolysant l'extrémité 5' des précurseurs des ARNt, et cela même en absence de sa sous-unité protéique [1]. Par la suite, la liste des ribozymes naturels s'est allongée et on les retrouve maintenant chez de nombreuses espèces.

Les ribozymes naturels sont impliqués dans des processus biochimiques importants pour la viabilité cellulaire, puisqu'il a été démontré qu'ils sont responsables de la synthèse protéique (ribosome), de la maturation des ARNt (RNase P) ainsi que de l'épissage des introns (spliceosome et introns des groupes I et II). Il existe aussi une classe de neuf petits ribozymes nucléolytiques qui possèdent des activités d'autoclivage, dont quatre ont été découverts dans les deux dernières années. Les descriptions des membres de ce groupe de petits ribozymes autoclivants se trouvent dans la section ci-dessous (section 1.1.3.2).

1.1.3.2 Ribozymes autoclivants

Les petits ribozymes autoclivants de source naturelle sont les ribozymes hammerhead, hairpin, *glmS*, hepatitis delta virus (HDV), Varkud satellite [31], twister, twister sister, le pistol et le hatchet [32]. La découverte récente des ribozymes twister sister, pistol et hatchet suggère qu'il reste encore d'autres ARN catalytiques à découvrir et que cette liste pourrait s'allonger dans les prochaines années [33].

La réaction de clivage est commune à tous les petits ribozymes autoclivants avec une catalyse de type acide-base générale et il s'agit de la même réaction de clivage que celle effectuée par la première étape de la RNase A [34-36]. Bien que les ribozymes possèdent tous le même mécanisme de clivage, ils ont des séquences différentes pour la composition de leur site actif. La réaction de clivage est amorcée avec la déprotonation du nucléophile 2'-OH par une base, suivi de l'attaque du 2'-oxyanion activé sur le 3'-phosphate adjacent qui brise le lien oxygène-phosphate et de la protonation du groupe oxygène partant par un acide (figure 1.2). Les produits générés par le clivage ont des extrémités 2',3'-cyclique phosphate et 5'-hydroxyle analogues aux produits obtenus suite au clivage par la protéine RNase A. La réaction de clivage s'effectue avec une inversion de configuration stéréochimique, qui implique un mécanisme de type S_N2 , dans lequel les nucléotides au site actif de clivage doivent être alignés dans un état de transition trigonal bipyramide, c'est-à-dire que l'angle de torsion entre le nucléophile 2'-OH, le phosphate au site de clivage et le groupe partant 5'-O doit être près de 180°. En général, les ribozymes qui effectuent la réaction de clivage sont aussi capables

d'effectuer la réaction inverse qui correspond à la ligation [34-36]. Les ribozymes utilisent donc les mêmes stratégies catalytiques que les enzymes comme la RNase A, pour effectuer le clivage. Il existe quatre grandes stratégies catalytiques qui peuvent être utilisées pour accélérer cette réaction de transestérification [37]. Ces stratégies sont la formation de la géométrie alignée (*in line*) au site de clivage, la neutralisation de la charge négative des oxygènes non liés du phosphate, la déprotonation du nucléophile 2'-OH et la neutralisation de la charge du groupe 5'-O partant. Les ribozymes peuvent employer des combinaisons de ces différentes stratégies, pour catalyser leur réaction de clivage [37].

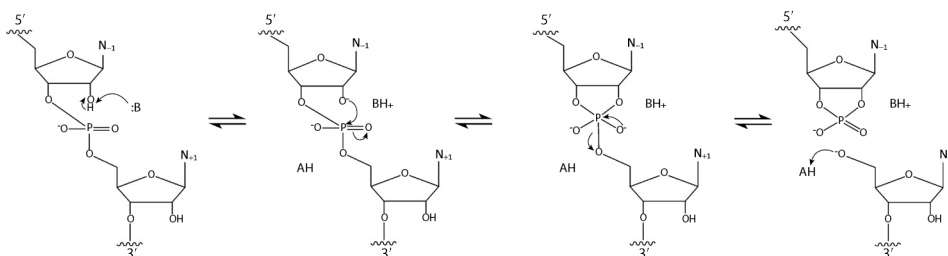


Figure 1.2 Mécanisme de clivage des petits ribozymes autoclivants.

La catalyse acide-base générale implique la déprotonation du nucléophile 2'-OH par une base, l'attaque du 2'-O sur le 3'-phosphate dans la géométrie alignée et la protonation du groupe partant 5'O par l'acide. (Provenant de [33]).

La réaction de clivage des ribozymes dépend d'une variété de cations métalliques ou non métalliques, mais ceux-ci semblent souvent jouer un rôle indirect sur la réaction de clivage. Ils sont plutôt nécessaires pour le bon repliement des structures secondaires et tertiaires des ribozymes, afin de permettre la bonne orientation des groupes réactifs au site actif. Pour stabiliser les structures d'ARN compactes, les cations divalents sont meilleurs que les cations monovalents pour lesquels des concentrations de l'ordre du molaire sont requises pour permettre le clivage. Le magnésium est le métal le plus efficace pour l'activité de la majorité des ribozymes autoclivants, bien que dans certains cas, ce soit par peu de différence par rapport à d'autres métaux [38]. Chaque ribozyme démontre des besoins différents pour les

métaux, les groupes fonctionnels requis ou les intervalles de pH, ce qui signifie que les compositions de leurs sites actifs sont toutes différentes.

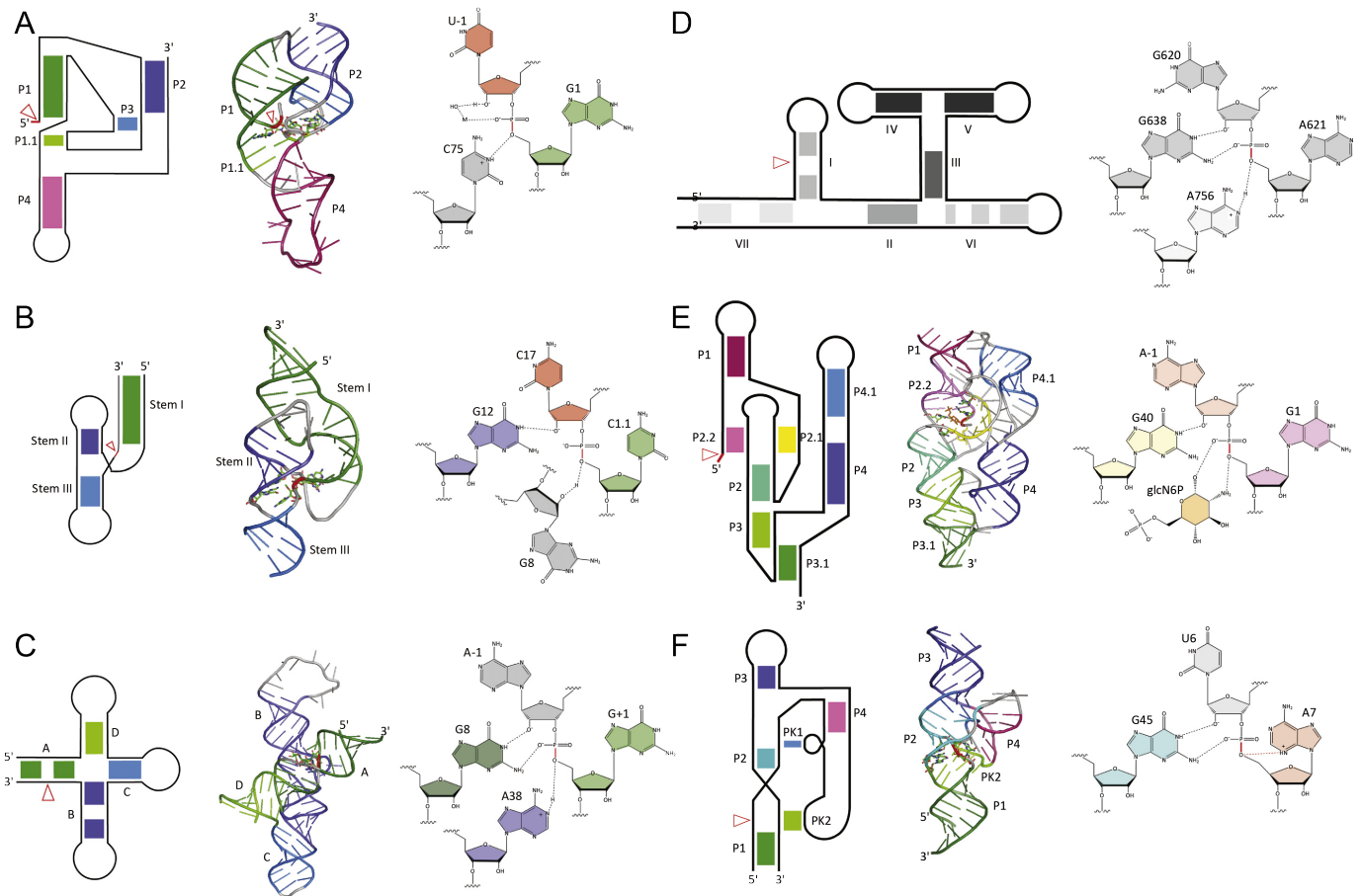


Figure 1.3 Caractéristiques structurales et modèles du site de clivage proposés pour les ribozymes autoclivants.

Les structures secondaires (à gauche), les structures cristallines (au centre) des ribozymes complets et les modèles du site de clivage (à droite) sont présentées pour les ribozymes (A) HDV (B) hammerhead (C) hairpin (D) VS (E) *glmS* et (F) twister. Les sites de clivage sont indiqués par des flèches rouges dans les structures secondaires et des brins rouges dans les structures cristallines. Les pseudonœuds sont indiqués par PK. (Adaptée de [33]).

1.1.3.2.1 Le ribozyme hammerhead

Le ribozyme hammerhead est très répandu dans la nature et il a été le premier petit ribozyme autoclivant à avoir été découvert chez un pathogène de plante. La première fonction qui a été associée au ribozyme hammerhead est dans le traitement des transcrits de la réplication en cercle roulant (*rolling circle*) [39]. On le retrouve maintenant dans divers génomes chez les procaryotes et les eucaryotes et en fonction du contexte dans lequel il se trouve, les fonctions biologiques reliées à ses réactions d'autoclivage et d'autoligation peuvent varier [40-42]. Par exemple, il y a une séquence du ribozyme hammerhead, dans le 3'UTR des gènes C-type lectin type II (*Clec2*) chez les mammifères, avec une séquence qui est séparée en deux sections, car elle est interrompue par une centaine de nucléotides [43]. Le repliement est fonctionnel pour ce ribozyme dans le génome et le clivage a pour conséquence de réduire l'expression protéique liée au gène. La fonction pour le ribozyme hammerhead dans ce contexte pourrait être d'apporter une contribution dans un système de régulation génique complexe [43]. La structure secondaire minimale associée au ribozyme hammerhead, est composée d'un centre catalytique comprenant une dizaine de nucléotides conservés et essentiels à la fonction de clivage, duquel émergent trois hélices (figure 1.3 B). Par contre, la séquence naturelle contient aussi une interaction boucle-boucle (entre la boucle terminale de la tige II et la boucle interne de la tige I) qui est importante pour un repliement efficace du ribozyme et une activité de clivage rapide en conditions physiologiques. Il est possible aussi de faire le clivage avec une version *trans* du ribozyme, mais avec une activité réduite à cause de l'absence de l'interaction boucle-boucle [44]. Dans l'une des structures du ribozyme hammerhead obtenue par cristallographie, il est possible d'observer l'interaction boucle-boucle qui induit l'organisation structurale du centre catalytique et l'orientation alignée (*in line*) au site de clivage [46]. Cette structure est en accord avec le mécanisme enzymatique proposé, car les nucléotides impliqués dans la réaction de clivage, les nucléotides G8 (2'-OH de G8) et G12 (N1 du G12), qui agissent comme acide et base se retrouvent près du site de clivage [46]{Anderson, 2013 #620}.

1.1.3.2.2 Le ribozyme hairpin

Le ribozyme hairpin est présent dans l'un des brins des ARN satellites du virus des taches en anneaux du tabac (*Tobacco ringspot virus*), du virus de la mosaïque jaune de la chicorée (*chicory yellow mottle*) et du virus de la mosaïque de l'arabette (*arabis mosaic*) [39, 47, 48]. Sa présence dans l'ARN viral est accompagnée d'un ribozyme hammerhead dans le brin opposé et ensemble, ces ribozymes sont impliqués dans le traitement des ARN multimériques à la suite de la réplication en cercle roulant. La structure secondaire du ribozyme hairpin se compose de quatre hélices connectées par une jonction à quatre voies (figure 1.3 C). Par contre, la structure secondaire minimale du ribozyme hairpin est composée seulement de deux domaines hélice-boucle-hélice, A et B, qui peuvent être situés sur le même brin ou sur deux brins différents pour une version *trans* du ribozyme [49]. La structure cristalline montre que le site actif se forme par une interaction boucle-boucle entre les boucles internes des domaines A et B, qui contiennent le site de clivage et les nucléotides importants pour la catalyse [50]. Dans le mécanisme enzymatique proposé, le nucléotide G8 (N1 du G8) se trouvant dans la boucle A, est la base et le nucléotide A38 (N1 de A38) de la boucle B, est l'acide permettant la réaction de clivage [51]{Rupert, 2002 #170}.

1.1.3.2.3 Le ribozyme *glmS*

Le ribozyme *glmS* est retrouvé chez plusieurs bactéries à Gram positif et il se situe dans le 5'UTR de l'ARNm *glmS* qui encode de la synthase glucosamine-6-phosphate [52]. Ce ribozyme est aussi un riborégulateur, puisque sa catalyse est activée par la liaison d'un ligand et cette liaison permet la régulation d'un gène. Pour le ribozyme *glmS*, le ligand est la glucosamine-6-phosphate (GlcN6P) et la réaction de clivage est impliquée dans la régulation du gène *glmS*. En réponse à la liaison de la GlcN6P, le clivage du ribozyme va induire la dégradation de la séquence codante en aval et entraîner la diminution de la synthèse de l'enzyme qui produit la GlcN6P. Puisque la protéine *glmS* est instable, la destruction de l'ARNm mène à une régulation négative de l'activité enzymatique. La structure secondaire minimale pour l'activité du ribozyme comprend les tiges P1 et P2, mais l'activité de clivage est meilleure avec les tiges P3 et P4 (figure 1.3 E) [52, 53]. La structure cristalline a démontré

que le ribozyme forme trois empilements d'hélices et deux pseudonoeuds entrelacés [54, 55]. Le centre du ribozyme est composé des tiges P1, P2, P2.1 et P2.2 et il contribue au site de liaison de la GlcN6P, en plus de contenir le site de clivage. Contrairement à un riborégulateur pour lequel la liaison du ligand induit un changement de conformation majeur, la liaison de la GlcN6P joue un rôle biochimique plutôt que structural chez le ribozyme *glmS*. Pour la réaction de clivage, les expériences ne permettent toujours pas de déterminer la base et l'acide, mais le nucléotide G640 (N1 de G640) et la GlcN6P sont suggérés pour remplir ces deux rôles [56, 57].

1.1.3.2.4 Le ribozyme HDV

Le ribozyme HDV provient du génome ARN du virus de l'hépatite delta (HDV) qui est considéré comme un satellite du virus de l'hépatite B humaine [58, 59]. Les brins génomiques et antigénomiques de l'ARN du virus du HDV codent chacun pour des ribozymes HDV distincts, mais qui ont une structure secondaire similaire. Ils ont pour fonction le traitement des intermédiaires de réplication d'ARN viraux, qui ont été générés par un mécanisme de cercle roulant. Une version du ribozyme HDV, le ribozyme CPEB3, a été découverte chez l'humain dans l'intron du gène CPEB3, qui appartient à une famille de gènes qui régulent la polyadénylation des ARNm [60]. Le site catalytique du ribozyme HDV est composé de six nucléotides conservés et la structure secondaire minimale du ribozyme requiert environ 60 nucléotides, pour permettre son repliement caractéristique dans une structure de pseudonoeuds double entrelacé (figure 1.3 A) [61]. Il existe plusieurs versions *trans* du ribozyme qui sont fonctionnelles. Les structures cristallines révèlent que le ribozyme se replie dans deux empilements d'hélices : le premier étant P1, P1.1 et P4 et le deuxième, P2 et P3 [62, 63] {Thaplyal, 2015 #621}. Le mécanisme de clivage du ribozyme HDV est unique à travers les ribozymes autoclivants, car il est le seul petit ribozyme à avoir besoin d'un métal divalent pour la catalyse. Le nucléotide C75 (N3 de C75) joue le rôle d'acide et le magnésium hydraté est la base. L'acide et la base ont pu être repérés dans la structure cristalline à proximité du site de clivage. Par conséquent, ce ribozyme a besoin de métaux non seulement pour permettre son repliement comme les autres ribozymes, mais aussi pour accomplir son activité de clivage.

1.1.3.2.5 Le ribozyme VS

Le ribozyme VS a été découvert dans certaines souches du champignon microscopique *Neurospora* et il est impliqué dans le traitement des transcrits de la réplication en cercle roulant [64]. Il est le plus gros des petits ribozymes autoclivants avec une séquence minimale d'environ 150 nucléotides (figure 1.3 D). Sa structure secondaire se divise en six domaines hélicoïdaux (I à VI) et la structure tertiaire déterminée par cristallographie montre qu'une interaction *kissing-loop* (interaction boucle-boucle entre deux boucles terminales) ainsi que deux jonctions à trois voies facilitent le repliement du ribozyme et la formation du site actif [65, 66]. Le site actif contient les nucléotides importants pour la catalyse et il est formé par l'interaction entre la boucle interne du domaine hélicoïdal I (qui contient le site de clivage) et la boucle interne du domaine hélicoïdal VI. La version *trans* du ribozyme est composée des domaines hélicoïdaux II à VI et le substrat est la tige-boucle I [67]. Les nucléotides impliqués dans le mécanisme de la réaction de clivage sont le nucléotide G638 (N1 du G638) situé dans la boucle interne de SLI et le nucléotide A756 (N1 du A756) dans la boucle interne de SLVI et ils sont proposés comme base et acide, respectivement [68]{Suslov, 2015 #200}.

1.1.3.2.6 Le ribozyme twister

Le ribozyme twister est un motif d'ARN qui a été identifié par une approche bio-informatique dans diverses espèces eucaryotes et bactériennes, en cherchant les bases de données pour des motifs de structures secondaires récurrents [69]. Il existe des milliers de représentants dans les systèmes vivants et cette grande présence laisse croire qu'ils pourraient avoir des fonctions biologiques variées. Plusieurs ribozymes twister et hammerhead sont situés à quelques kilobases de distance près de certains gènes codants pour des protéines, ce qui suggère qu'ils pourraient avoir des fonctions jumelées. La structure secondaire minimale pour l'activité du motif se divise en 3 tiges (P1, P2 et P3) et 3 boucles (L1, L2 et L4). La présence de deux pseudonoeuds (PK1 et PK2) est importante pour la structure tertiaire et l'activité de clivage (figure 1.3 F). Plusieurs structures cristallographiques ont confirmé la présence de ces pseudonoeuds, en plus de révéler une structure globale dans laquelle les hélices P1 et P2 sont coaxialement alignées [70-72]. Il existe une version *trans* du ribozyme qui est active et qui

démontre un clivage très rapide [69]. Une guanine est suggérée pour occuper la place de base. L'identité de l'acide n'a pas été déterminée, mais il a été suggéré qu'une adénine ou un ion divalent pourrait participer directement à la catalyse acide-base générale [70-73].

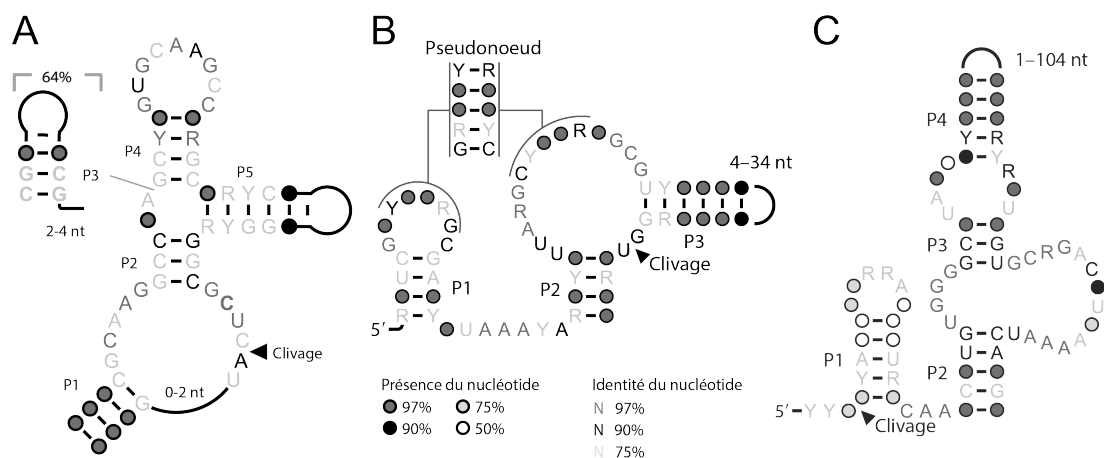


Figure 1.4 Structures secondaires des ribozymes découverts récemment.

Séquences consensus et modèles des structures secondaires des ribozymes (A) twister sister, (B) pistol et (C) hatchet. Les purines sont indiquées par la lettre R et les pyrimidines par un Y. Le site de clivage est indiqué par une flèche noire. (Adaptée de [32]).

1.1.3.2.7 Le ribozyme twister sister

Le ribozyme twister sister a hérité son nom en raison de sa ressemblance au ribozyme twister, qui lui ressemble autant au niveau de la séquence que de la structure secondaire [32]. Il a été découvert en explorant les génomes d'ADN bactérien, à la recherche de structures secondaires conservées qui se trouvent près des séquences de ribozymes hammerhead ou twister (moins de 6 kilobases). Sa structure secondaire ressemble beaucoup au ribozyme twister, car l'arrangement des tiges P1 à P5 est similaire dans les deux ribozymes et ils ont des nucléotides en commun dans la boucle P4 (ou L4) (figure 1.4 A). La différence majeure avec le ribozyme twister est que le ribozyme twister sister ne contient aucun des pseudonoeuds, qui jouent un rôle crucial chez le ribozyme twister. Il existe une version *trans* du ribozyme, avec

un ribozyme d'environ 50 nucléotides et un substrat d'environ 20 nucléotides. Les nucléotides impliqués dans la catalyse n'ont pas encore été déterminés, puisque la découverte du ribozyme est récente. Des expériences supplémentaires seront nécessaires, pour connaître la fonction du ribozyme et la composition de son site actif [32].

1.1.3.2.8 Le ribozyme pistol

Les ribozymes pistol et hatchet ont été découverts de la même façon que le ribozyme twister sister et tout comme pour le ribozyme twister sister, il n'y a pas encore de fonctions qui sont associées à leurs activités de clivage [74]. La structure secondaire du ribozyme pistol comporte trois tiges (P1, P2 et P3), ainsi qu'un pseudonoeud entre la boucle P1 et la jonction P2-P3 (figure 1.4 B). Il est possible de faire une version *trans* du ribozyme qui comporte environ 50 nucléotides et son substrat environ 15 nucléotides. Comme pour le ribozyme twister sister, il faudra faire plus d'expériences pour connaître la composition du site actif [74].

1.1.3.2.9 Le ribozyme hatchet

La séquence consensus et la structure secondaire du ribozyme hatchet ont été déterminées grâce à l'alignement de plus de 159 séquences du ribozyme hatchet [75]. La séquence minimale active pour la catalyse est composée de quatre tiges (P1 à P4). Les nucléotides les plus conservés sont situés dans la jonction entre P1 et P2 et dans le *bulge* entre P2 et P3 (figure 1.4 C). Il existe deux versions *trans* du ribozyme, selon la location de la séparation des brins qui peut être dans la boucle P1 ou P4. Les nucléotides impliqués au site actif n'ont pas encore été déterminés, de même que la fonction de ce ribozyme [75].

Les ribozymes comme le hammerhead, le HDV et le twister se retrouvent répartis dans différents organismes et peuvent posséder des fonctions biologiques différentes selon l'organisme dans lequel ils se trouvent [33]. Par conséquent, les fonctions des ribozymes ne sont pas encore toutes connues. Ces ribozymes sont même localisés à proximité les uns des autres dans certaines régions introniques, ce qui pourrait suggérer de la coopération entre les ribozymes dans l'accomplissement de leurs fonctions. Si les ribozymes hammerhead, HDV et twister sont abondants dans la nature, ce n'est pas le cas des ribozymes hairpin et VS, qui n'ont été retrouvés qu'à peu d'endroits jusqu'à maintenant et qui sont associés à seulement une fonction. C'est ce qui explique que ces deux ribozymes, avec le ribozyme *glmS*, sont les seuls ribozymes qui démontrent une seule fonction biologique constante et claire : le ribozyme hairpin est important pour le traitement des génomes multimériques de l'ARN satellite viral, le *glmS* est un élément de contrôle génétique responsable de la boucle de rétroaction négative et le ribozyme VS est important pour le traitement des transcrits satellite du plasmide Varkud qui sont abondants. La découverte récente des ribozymes twister sister, pistol et hatchet suggère qu'il doit exister dans la nature d'autres motifs d'ARN avec des fonctions catalytiques à identifier et des motifs déjà connus qui restent à découvrir chez de nouveaux organismes, dans lesquels ils pourraient avoir des fonctions différentes [32].

Un intérêt des recherches effectuées sur les ribozymes est de pouvoir faire de l'ingénierie pour modifier leur activité ou pour leur faire accomplir de nouvelles fonctions. La plupart des ribozymes peuvent ainsi être modifiés de leur forme naturelle pour les utiliser pour le clivage en *trans*, pour les rendre plus petits, plus rapides ou pour développer des outils biochimiques en changeant leur substrat pour des ARN cibles (sujet de la section 1.3). Un ribozyme pour lequel il y a eu très peu d'ingénierie pour l'instant est le ribozyme VS de *Neurospora*, qui est l'objet des études de cette thèse.

1.2 Le ribozyme VS de *Neurospora*

Le plasmide VS a été découvert en 1990 dans le laboratoire du Dr Collins, dans les mitochondries de plusieurs souches du champignon *Neurospora* [64]. On le retrouve en présence du plasmide Varkud, dont il dépend puisqu'on ne le retrouve jamais sans sa présence, bien que le plasmide Varkud puisse se retrouver seul. Le plasmide VS est un ADN double brin, qui est présent en peu de copies et qui n'est pas dérivé du chromosome mitochondrial de *Neurospora*. *In vivo*, l'ARN complémentaire au plasmide VS se trouve majoritairement sous la forme d'un ARN simple brin circulaire de 881 nt et *in vitro*, il a été déterminé que la séquence du ribozyme actif occupe une région d'environ 150 nucléotides [76, 77]. Les expériences *in vivo* et *in vitro* montrent que le ribozyme VS effectue une réaction d'autoclivage et une réaction d'autoligation [78]. Ces réactions seraient impliquées dans sa fonction de traitement des transcrits d'ARN multimériques *in vivo*. La réplication du ribozyme VS se fait par un mécanisme qui s'apparente à la réplication en cercle roulant utilisée par d'autres ARN satellites, avec pour différence que la transcription se fait à partir d'un gabarit d'ADN et non à partir d'un gabarit d'ARN comme pour les autres ARN satellites (figure 1.5) [79]. La transcription du plasmide VS se fait à l'aide de l'ARN polymérase mitochondriale de *Neurospora* et elle est amorcée à un site relié à la séquence consensus promoteur de la polymérase, situé en amont du site de clivage du ribozyme à la position 601. La transcription génère un long transcrit d'ARN multimérique. La réaction d'autoclivage permet ensuite d'obtenir plusieurs copies monomériques et la réaction d'autoligation permet la formation de la forme circulaire de l'ARN VS qui est la plus stable *in vivo*. Par la suite, la transcriptase inverse du plasmide Varkud permet de produire l'ADN VS complémentaire correspondant à l'ARN VS, qui va ultimement permettre de générer le plasmide d'ADN double brin. L'ARN VS est répliqué comme un satellite du plasmide Varkud, puisqu'il dépend de celui-ci pour sa réplication [79].

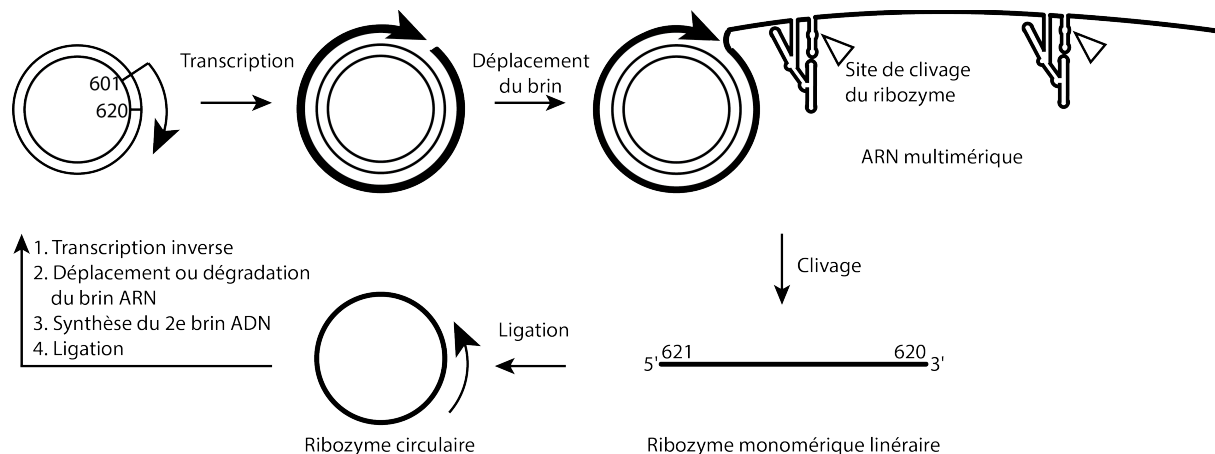


Figure 1.5 Mécanisme de réplication du ribozyme VS.

Le plasmide VS d'ADN double brin est transcrit dans un ARN linéaire contenant plusieurs copies du ribozyme et la réaction d'autoclivage aux nombreux sites de clivage (entre les nucléotides 620 et 621) permet d'obtenir les ARN monomériques. La forme circulaire du ribozyme est obtenue par autoligation et la transcription inverse permet d'amorcer la production du plasmide d'ADN. (Adaptée de [79] et [33]).

1.2.1 Structure secondaire

La séquence minimale de la partie active du ribozyme VS peut être réduite à environ 150 nucléotides [77]. La présence d'un seul nucléotide est requise en amont du site de clivage et le reste de la séquence du ribozyme est situé en aval. Les études de modification chimique et de mutagenèse ont démontré que la structure secondaire est divisée en six domaines hélicoïdaux nommés I à VI (figure 1.6) [80]. Le site de clivage se situe entre les nucléotides G620 et A621 dans la boucle interne de la tige-boucle SLI [76, 80] [76]. Cette séquence minimale composée de six domaines hélicoïdaux est efficace pour la réaction de clivage, mais elle est moins efficace pour la réaction inverse de ligation. En allongeant la séquence du ribozyme en 3' et en 5', un nouveau domaine hélicoïdal VII est formé et l'ajout de cet élément crée un ribozyme qui favorise la réaction de ligation [81].

In vivo et *in vitro*, le ribozyme VS effectue la réaction de clivage et de ligation en *cis*, de façon intramoléculaire, avec la séquence complète du ribozyme composé des domaines hélicoïdaux I à VI. Par contre, il est possible d'effectuer *in vitro* une réaction en *trans* avec la tige-boucle SLI contenant le site de clivage synthétisé séparément du reste du ribozyme [67]. Dans la réaction en *trans*, SLI est le substrat et les domaines hélicoïdaux II à VI forment le domaine catalytique, aussi appelé le ribozyme *trans*.

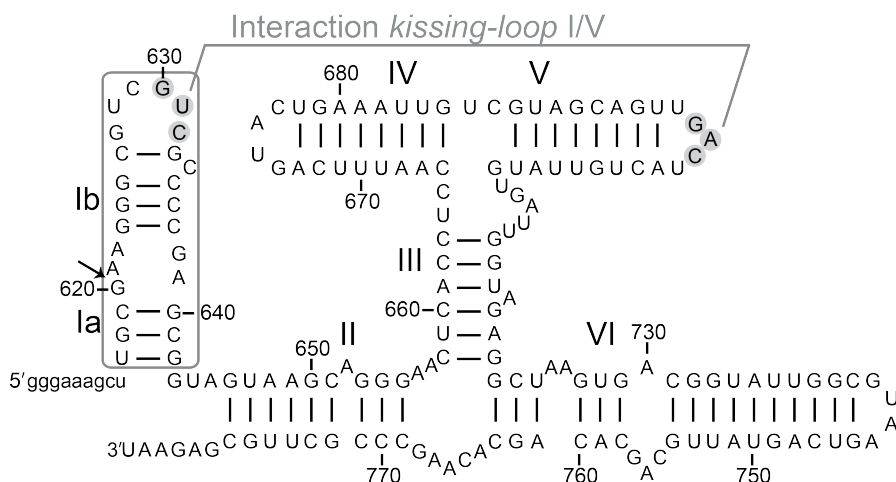


Figure 1.6 Structure secondaire du ribozyme VS.

Le domaine catalytique du ribozyme VS est composé des domaines hélicoïdaux II à VI et le substrat est composé de la tige-boucle I (Ia et Ib) dans l'encadré. Le site spécifique de clivage est localisé entre les nucléotides G620 et A621 dans SLI et il est indiqué par une flèche noire. L'interaction *kissing-loop* I/V implique des interactions W-C entre les nucléotides encadrés en gris dans les boucles I et V. (Adaptée de [80] et [65]).

1.2.2 Interaction *kissing-loop* I/V

La reconnaissance du substrat SLI se fait par une interaction *kissing-loop* entre les boucles de SLI et de SLV (I/V) [65]. Ce mode de reconnaissance du substrat impliquant une interaction *kissing-loop* est unique parmi les ribozymes autoclivants. L'interaction *kissing-loop* se définit par la formation de paires de bases entre les séquences complémentaires des boucles apicales provenant de deux tige-boucles différentes. Ce type d'interaction est très

répandu dans l'ARN et permet d'accomplir des fonctions biologiques diverses, telles que la dimérisation du génome du VIH-1 [82].

L'interaction *kissing-loop* dans le cadre du ribozyme VS a comme fonction principale de permettre la reconnaissance du substrat SLI par le ribozyme, en plus de faciliter la formation du site actif du ribozyme [83]. L'interaction *kissing-loop* I/V est également accompagnée d'un réarrangement du substrat SLI, celui-ci passant d'une conformation inactive vers une conformation active pour la réaction de clivage [84, 85]. Pour toutes ces raisons, la présence de l'interaction *kissing-loop* I/V est requise pour obtenir une activité de clivage optimale. En l'absence de l'interaction *kissing-loop* I/V, la réaction de clivage par le ribozyme est quand même possible, mais la vitesse de réaction est grandement diminuée [85, 86]. La présence de métaux, comme le magnésium, est importante pour permettre la formation de l'interaction *kissing-loop* et le réarrangement de SLI, mais aussi pour permettre la catalyse.

Des études de mutagenèse ont permis d'identifier trois paires de bases W-C impliquées dans l'interaction *kissing-loop* I/V du ribozyme VS. Ces paires de bases impliquent les nucléotides G630, U631 et C632 de la boucle SLI et les nucléotides G697, A698 et C699 de la boucle SLV (figure 1.6) [65]. Ces études de mutagenèse ont aussi proposé que les boucles I et V contiennent des motifs *U-turn*, dont la présence a été confirmée plus tard par des études de RMN [87, 88]. Les motifs *U-turn* facilitent la formation des paires de bases entre les nucléotides de l'interaction *kissing-loop*. Le motif *U-turn* se définit par une séquence UNR dans laquelle le U représente une uridine, N peut être n'importe quel nucléotide et R est une purine. Ce motif va imposer un virage brusque dans le squelette ribose-phosphate entre les deux premiers nucléotides du motif (U et N) et provoquer l'empilement des deux nucléotides suivants (N et R). Ce motif permet ainsi d'exposer des nucléotides au solvant et de faciliter leur interaction avec d'autres nucléotides. La séquence UNR chez le ribozyme VS est composée des nucléotides U696-G697-A698 dans la boucle de SLV et des nucléotides U628-C629-G630 dans la boucle de SLI [65, 87-89]. La structure RMN d'un complexe I/V a permis de confirmer qu'en plus des deux motifs *U-turn* et de la formation des trois paires de bases W-C de l'interaction *kissing-loop*, l'interaction I/V est stabilisée par une paire de base *sugar*

edge/W-C, des ponts hydrogènes, de l'empilement de base, ainsi que par deux triplets de bases impliquant les nucléotides C629-A701-U695 et C626-G633-G634 (correspond aux nucléotides C626-G633-C634 dans le ribozyme VS naturel) (figure 1.7) [87]. Des études de thermodynamiques d'ITC ont permis d'établir que l'interaction *kissing-loop* I/V est très stable avec un K_D de 0.24 μM à 25 °C et 20 mM MgCl_2 , bien qu'elle ne possède que quatre paires de bases [90].

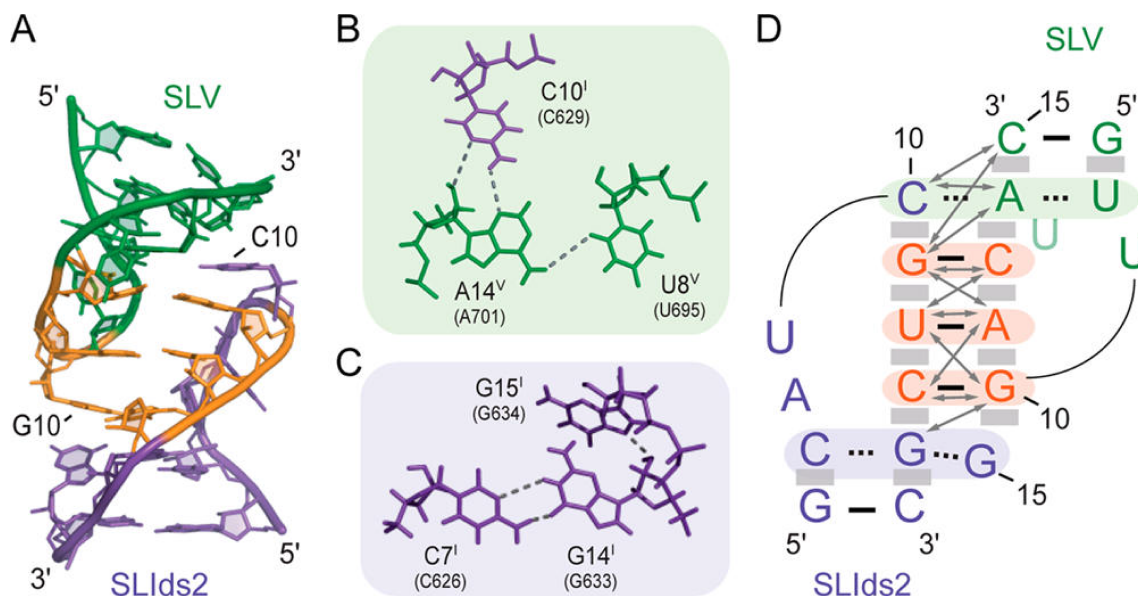


Figure 1.7 Structure RMN d'un complexe I/V.

Structure tridimensionnelle de l'interaction *kissing-loop* entre SLI et SLV déterminée par RMN. Les lignes pointillées indiquent l'emplacement des ponts hydrogènes. Les nucléotides ombragés en orange représentent les interactions W-C et les nucléotides ombragés en vert et mauve représentent les triplets de bases. (A) Structure d'énergie minimale de l'interaction *kissing-loop* I/V. Les deux triplets de bases qui stabilisent la structure de l'interaction *kissing-loop* sont représentés en (B) et (C). (D) Présentation schématique de l'interaction *kissing-loop* I/V. Les lignes pointillées indiquent des ponts hydrogènes définis sur la base de courtes distances dans l'ensemble des structures, les flèches grises indiquent les NOE (*nuclear Overhauser effect*) observés entre les résidus et les boîtes grises représentent l'empilement de bases. (Tirée de [87]).

Les boucles SLI et SLV de l'interaction *kissing-loop* I/V forment une unité modulaire du ribozyme VS, puisqu'il est possible de reconstituer l'interaction entre les deux tige-boucles isolées sans la présence du ribozyme complet [85]. Des études de protection avec des radicaux hydroxyles en présence et en absence de l'interaction *kissing-loop* I/V ont démontré que cette interaction est importante pour l'interaction de SLI avec le reste du ribozyme, mais que les boucles I et V ne semblent pas interagir avec le reste du ribozyme [83]. De plus, les études de titrage calorimétrique isotherme confirment la modularité de l'interaction, car la constante de dissociation pour l'interaction entre SLI et le ribozyme complet est similaire à la constante de dissociation de l'interaction SLI avec une tige-boucle SLV isolée [90].

1.2.3 Activation du substrat

Le premier indice qui a suggéré un changement de conformation dans le substrat SLI, est relié au nucléotide C634, dont la protection par des modifications chimiques des faces W-C change suite à l'ajout de magnésium [80]. En absence de magnésium, il est protégé des modifications, mais il devient réactif suite à l'ajout de magnésium. Ce changement de comportement s'effectue dans SLI en présence du ribozyme et du magnésium (figure 1.8). Il est maintenant clair qu'il existe deux conformations du substrat qui ont été définies par des expériences de SELEX, de mutagenèse et de modifications chimiques et il s'agit d'une conformation active nommée *shifted* et d'une conformation inactive nommée *unshifted* [84].

1.2.3.1 Conformation *unshifted*

Dans la conformation *unshifted*, les nucléotides de la tige Ib G623-G624-G625 sont appariés avec les nucléotides C634-C635-C636 et il y a six nucléotides qui forment la boucle interne de SLI (figure 1.8 A). Les structures RMN du SLI dans la conformation *unshifted* révèlent que la boucle interne contient trois paires de bases non canoniques, qui impliquent les six nucléotides de la boucle interne [91, 92]. Il s'agit d'une paire de base *cis* W-C/W-C A⁺-C avec les nucléotides A622⁺ et C637 et de deux interactions *trans Hoogsteen-sugar edge* G-A impliquant les nucléotides G620-A639 et A621-G638 (figure 1.9 A). Flinders *et al* ont

proposé que la liaison de SLI à la tige-boucle SLV du ribozyme change l'équilibre vers une conformation ouverte de la boucle interne du substrat qui libère C637 pour lui permettre de s'apparier avec G623, ce qui libère à son tour C636 et amorce une séquence de glissement de brin qui se propage dans le reste de l'hélice Ib pour induire le changement de conformation vers la conformation *shifted* [92]. Ce modèle est également compatible avec les études RMN plus récentes effectuées avec plusieurs complexes SLI/SLV [87].

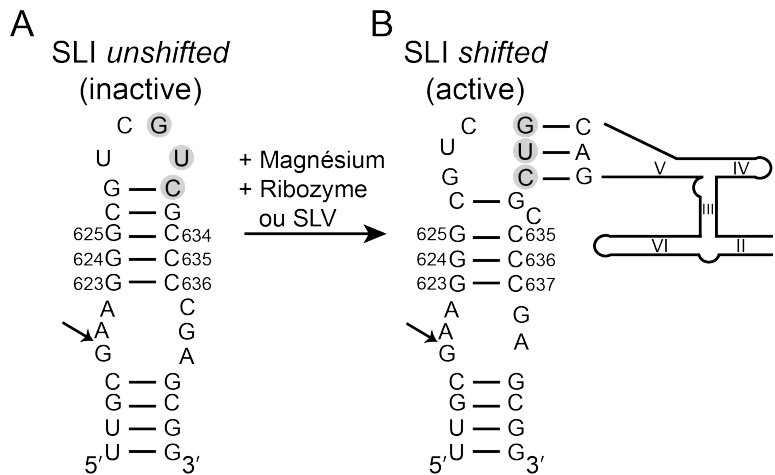


Figure 1.8 Réarrangement de la structure de SLI lors de la liaison de la tige-boucle V.

Changement de conformation du substrat SLI induit par l'ajout du magnésium et du ribozyme ou de la tige-boucle SLV isolé. (A) Dans la conformation *unshifted* du substrat, les nucléotides G623-G624-G625 sont appariés avec les nucléotides C634-C635-C636 et (B) dans la conformation *shifted*, les mêmes nucléotides G623-G624-G625 sont maintenant appariés avec les nucléotides C635-C636-C637. Suite à ce réarrangement du substrat, le nucléotide C634 se retrouve exclu de l'hélice et il y a un nucléotide en moins dans la boucle interne du site de clivage. (Adaptée de [84]).

1.2.3.2 Conformation *shifted*

Lors de l'activation du substrat SLI vers la conformation *shifted*, il se produit un réarrangement dans l'appariement des nucléotides de la tige Ib qui fait en sorte que les nucléotides G623-G624-G625 interagissent maintenant avec les nucléotides C635-C636-C637 (figure 1.8 B). Dans la conformation *shifted*, le nucléotide C634 n'est plus impliqué dans une

paire de base et il y a cinq nucléotides dans la boucle interne du site de clivage. Dans la structure RMN de SLI de la conformation *shifted*, les cinq nucléotides de la boucle interne forment trois paires de bases *trans hoogsteen-sugar edge* G-A (figure 1.9 A et B). Deux de ces paires de bases sont communes avec la conformation *unshifted*, G620-A639 et G638-A621, tandis que pour la troisième paire de base, le nucléotide G638 se partage pour faire une nouvelle paire de base G-A avec A622. Une autre similarité avec la structure *unshifted* est l'empilement interbrin entre les guanines (G620-G638) et les adénines (A621-A639) qui est typique des motifs associés aux paires de bases G-A. La différence entre les structures implique le nucléotide C637 qui fait une interaction avec A622 dans la conformation *unshifted*, mais qui fait une interaction W-C avec G623 de l'hélice Ib dans la conformation *shifted*. La structure RMN de la boucle interne du site de clivage dans la conformation active comporte beaucoup de similarités avec les structures RMN de la conformation inactive du substrat et seulement quelques différences subtiles les différencient (figure 1.9 B). Par contre, il n'y a que la forme active de SLI qui peut être clivée en présence de magnésium [93].

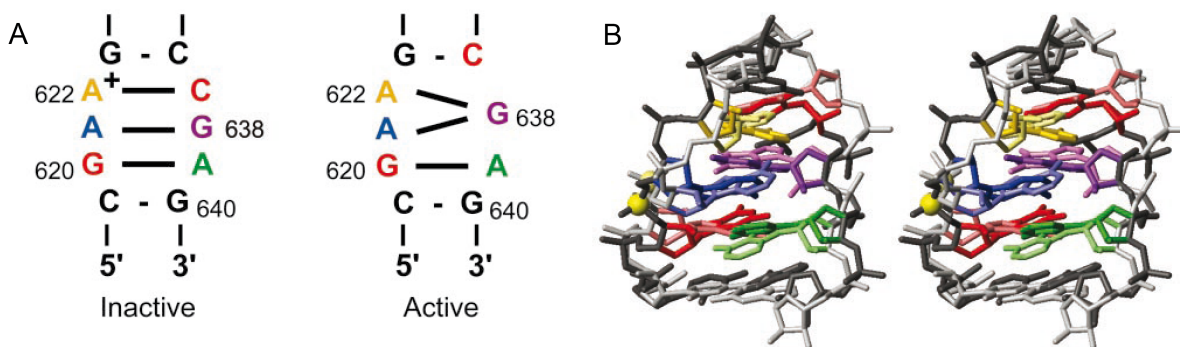


Figure 1.9 Comparaison des structures des conformations active et inactive de la boucle interne de SLI.

(A) Structures secondaires qui montrent l'appariement des nucléotides dans la boucle interne du substrat SLI qui contient le site de clivage dans la conformation inactive à gauche et la conformation active à droite. (B) Superposition des structures RMN (*stereoview*) de la boucle interne de SLI dans la conformation inactive (couleurs pastel) et dans la conformation active (couleurs foncées). (Tirée de [93]).

L'activation du substrat et l'interaction *kissing-loop* I/V peuvent s'effectuer sans la présence du ribozyme complet, avec les tige-boucles SLI et SLV isolées en présence de magnésium [85]. Par contre, la tige-boucle SLV isolée n'est pas suffisante pour permettre le clivage de SLI. Il est possible de faire le design d'un substrat SLI dans une conformation *preshifted* avec la mutation C634G, c'est-à-dire un substrat qui est toujours dans la conformation active. Ce substrat *preshifted* peut être utile pour simplifier des études cinétiques enzymatiques ou thermodynamiques, par exemple, pour étudier l'effet d'une mutation seulement sur l'interaction *kissing-loop* I/V. Par contre, ce substrat *preshifted* ne peut pas compenser pour une absence de l'interaction *kissing-loop* I/V, bien qu'un substrat *shifted* soit plus facilement lié par le ribozyme [94]. Par conséquent, pour obtenir une activité de clivage optimale, la présence du ribozyme, du magnésium, de l'interaction *kissing-loop* I/V intacte et d'un substrat *shifted* ou *preshifted* est requise [85, 94, 95].

1.2.4 Réaction de clivage

Le mécanisme de clivage du ribozyme VS implique une catalyse de type acide-base générale qui est commune aux ribozymes autoclivants et qui a été décrite de façon générale à la section 1.1.3.2.

1.2.4.1 Le nucléotide A756

Le site actif du ribozyme VS a été identifié initialement en cherchant un nucléotide qui pourrait être à proximité du site de clivage, situé entre les nucléotides G620 et A621. Une expérience de *cross-link* a permis de constater que le nucléotide A621 en aval du site de clivage effectue un *cross-link* avec le nucléotide A756 de la boucle A730 de SLVI, autant dans des conditions de clivage que de ligation [96]. Ce *cross-link* est une preuve que la boucle A730 est à proximité du site de clivage et qu'il fait partie du site actif du ribozyme. Des études précédentes de mutagenèse avaient démontré que le nucléotide A756 et la boucle A730 jouent un rôle critique dans la catalyse, mais elles n'avaient pas encore identifié la boucle A730 comme étant une partie intégrante du site actif du ribozyme [97-99]. Dans la

boucle A730, c'est le nucléotide A756 qui semble le plus important, car c'est le nucléotide de la boucle dont la substitution par un autre nucléotide provoque le plus gros effet sur l'activité de clivage et de ligation [99-103]. Des études supplémentaires sur A756 révèlent que c'est la base qui est importante, plus particulièrement la face W-C et que le ribose pourrait avoir seulement une fonction structurale dans le repliement de la boucle A730 [104].

1.2.4.2 Le nucléotide G638

Les expériences de mutagenèse ont également identifié le nucléotide G638 comme étant impliqué dans l'étape limitante de transfert de proton de la réaction [105, 106]. Les deux nucléotides impliqués dans la catalyse acide-base générale sont le nucléotide G638 de la boucle SLI et le nucléotide A756 de la boucle SLVI, par contre ces expériences n'étaient pas suffisantes pour identifier lequel est l'acide et lequel est la base. Des études de cinétique enzymatique en fonction du pH pour différentes mutations des nucléotides A756 et G638 proposent que le nucléotide A756 est l'acide de la réaction de clivage et le nucléotide G638 est la base (figure 1.10) [68].

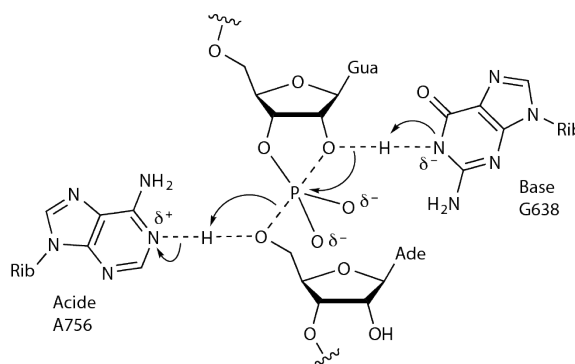


Figure 1.10 Mécanisme proposé de clivage au site actif du ribozyme VS.

Représentation schématisée de la catalyse acide-base générale au site actif du ribozyme, qui implique A756 comme acide et G638 comme base. Les lignes pointillées et les flèches indiquent les liaisons qui sont formées et défaits durant l'état de transition. (Tirée de [66]).

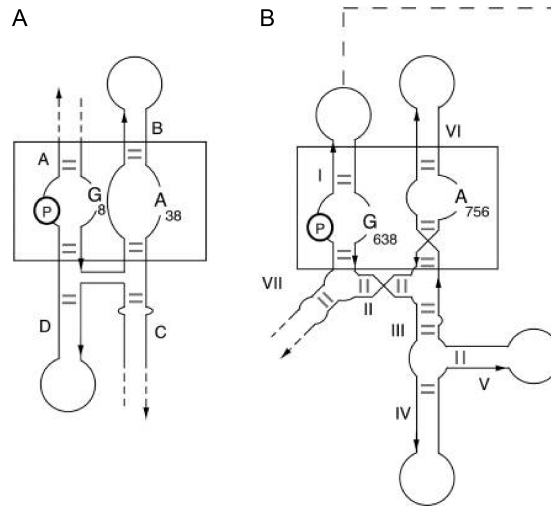


Figure 1.11 Représentation schématique des structures des sites actifs des ribozymes hairpin et VS.

Les sites actifs des ribozymes sont encadrés pour (A) le ribozyme hairpin et pour (B) le ribozyme VS. Le phosphate au site de clivage est encerclé et les nucléotides importants pour la catalyse acide-base générale sont identifiés (adénine et guanine). (Tirée de [107]).

1.2.4.3 Comparaison avec le site actif du ribozyme hairpin

Le site actif du ribozyme VS partage de grandes similarités avec le site actif du ribozyme hairpin, en plus de partager le même mécanisme de clivage [107]. Pour les deux ribozymes, l'acide est une adénine et la base est une guanine. Les deux sites actifs sont générés par l'interaction entre deux boucles internes de deux hélices différentes, qui permettent d'amener à proximité le site de clivage, ainsi que les deux nucléotides impliqués dans la catalyse acide-base générale (voir figure 1.11). Dans le cas du ribozyme VS, ce sont les boucles internes de SLI et SLVI qui forment le site actif et pour le ribozyme hairpin, ce sont les boucles internes A et B. Les sites actifs adoptent une topologie identique avec la guanine, qui joue le rôle de base, dans le brin opposé au site de clivage et l'adénine, qui joue le rôle d'acide, dans la seconde boucle. Dans le ribozyme VS, on retrouve le site de clivage et la base G638, dans la boucle SLI et l'acide A756, dans la boucle SLVI. Dans le ribozyme hairpin, c'est dans la boucle A que l'on retrouve le site de clivage de même que la base G8 et

dans la boucle B se trouve l'acide A38. Les ressemblances vont jusque dans les structures tertiaires des ribozymes dans lesquelles la guanine catalytique est empilée sous le nucléotide N-1 et l'adénine catalytique est empilée sur le nucléotide N+1. De plus, dans les deux ribozymes, un motif *S-turn* est responsable de l'extrusion de l'adénine catalytique dans un sillon mineur élargi et lui permet de s'orienter vers le groupe partant. Finalement, chez ces deux ribozymes, un motif *L-platform* est observé qui permet l'orientation de la guanine catalytique vers le nucléophile 2'-OH et forme un empilement avec la base du *L* formée par la paire de base non canonique entre le nucléotide en 5' du site de clivage et le nucléotide immédiatement en 3' de la base [66].

1.2.4.4 Conditions pour la réaction de clivage

La réaction de clivage nécessite des conditions particulières pour atteindre son activité optimale et ces conditions varient en fonction de la séquence du ribozyme utilisé. Par exemple, la température optimale pour la réaction de clivage se situe entre 40 °C et 50 °C pour un ribozyme mutant en *cis*, mais elle est de 30 °C pour un ribozyme mutant en *trans* [67, 108]. La réaction de clivage optimale nécessite des métaux divalents comme le magnésium avec des concentrations variant de 10 à 25 mM, même si celui-ci n'a pas de rôle direct dans la catalyse [67, 108]. La nécessité des métaux est plutôt associée à la formation de la structure active du ribozyme. Les conditions pour le clivage optimal comprennent la présence de KCl, de spermidine et de Tris-HCl pour l'ajustement du pH entre 7.5 et 8 [108]. Des métaux divalents autres que le magnésium, comme le manganèse et le calcium peuvent aussi soutenir l'activité, mais ils sont moins efficaces [108]. Pour le ribozyme en *trans*, des ions monovalents comme le sodium, le lithium et les ions ammonium peuvent permettre le clivage, mais à des concentrations de l'ordre du molaire [109]. Les dénaturants comme l'urée et le formamide inhibent la réaction de clivage [108]. Le cobalt hexammine qui est un analogue du magnésium hydraté permet un repliement du ribozyme similaire à celui avec le magnésium, mais déstabilise l'interaction *kissing-loop* et ne permet pas le clivage du ribozyme, ce qui signifie qu'il y a des sites de liaisons de métaux importants pour le clivage, qui ne peuvent pas être liés par le cobalt hexammine [110]. La viomycine et la spermine permettent d'obtenir une phase de

clivage très rapide pour le ribozyme en *cis*, qui est suivi d'une phase de clivage moins rapide [111, 112]. Ces deux phases sont expliquées par la formation de multimères qui permettent le clivage en *trans* des ribozymes en *cis*. Cependant, une fraction des ribozymes est inactive et ne peut pas cliver, possiblement attribuable à l'adoption d'une conformation inactive [111, 112].

1.2.5 Études structurales

1.2.5.1 Premier modèle de basse résolution du ribozyme VS

Le premier modèle de la structure tertiaire a été obtenu sur la base de données biochimiques impliquant des expériences de *footprinting* avec des radicaux hydroxyles [83]. Ce modèle a mis en valeur un centre protégé du solvant qui est composé de régions près des jonctions II-III-VI et III-IV-V, de l'interaction *kissing-loop* I/V et des domaines hélicoïdaux II et VI. De plus, ces expériences montrent que le repliement des deux jonctions est indépendant de la formation de l'interaction *kissing-loop* I/V. À partir de l'information obtenue des expériences de *footprinting* avec les radicaux hydroxyles, ainsi que l'information obtenue à partir de d'autres expériences de modification chimique et de mutagenèse, un modèle a pu être généré pour la structure tertiaire du ribozyme (figure 1.12 A et B). Dans ce modèle, les domaines hélicoïdaux II et VI forment une crevasse qui va permettre au substrat SLI de venir se placer pour la formation du site actif du ribozyme. L'interaction *kissing-loop* I/V est requise pour la formation de ce centre catalytique composé des domaines hélicoïdaux II, VI et I.

1.2.5.2 Deuxième modèle de basse résolution du ribozyme VS

Un second modèle de basse résolution a été généré, en se basant sur des études de comparaison de mobilité sur gel d'électrophorèse à l'état natif et sur des études de FRET (*Förster resonance energy transfer*) effectuées sur les jonctions à trois voies isolées II-III-VI et III-IV-V (figure 1.13 A) [113, 114]. Le modèle représente les orientations relatives des hélices par rapport aux autres hélices de leur jonction et il n'inclut pas les nucléotides compris

dans les *bulges* et la boucle A730. Le substrat SLI a ensuite été placé en respectant deux règles : (1) positionner l'extrémité 3' du substrat SLI à proximité de l'extrémité 5' du domaine hélicoïdal II comme dans le ribozyme en *cis* et (2) placer la boucle terminale du substrat SLI dans une position crédible pour effectuer l'interaction *kissing-loop* I/V. La structure globale du ribozyme est organisée par les deux jonctions à trois voies et la formation de cette structure est dépendante du magnésium. Le modèle montre que les hélices III, IV et VI sont entassées coaxialement et que les hélices II et VI sont bien positionnées afin de former la crevasse pour accueillir le substrat SLI et permettre la formation du site actif.

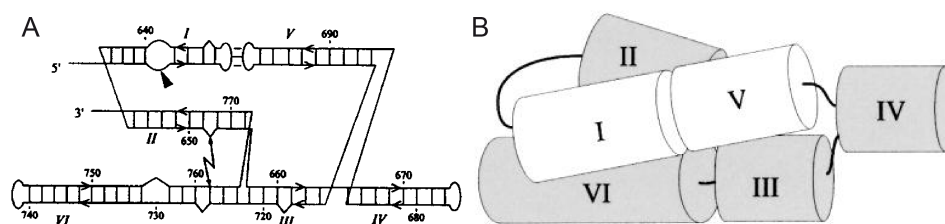


Figure 1.12 Modèles de la structure tertiaire du ribozyme VS.

La structure tridimensionnelle est déterminée en accord avec les données biochimiques, incluant des études de *footprinting* avec des radicaux libres. (A) Le site de clivage est indiqué par la flèche noire dans la boucle de SLI et (B) les hélices sont représentées par des cylindres et les connections entre les hélices sont représentées par des lignes noires. (Tirée de [115] et [83]).

1.2.5.3 Troisième modèle de basse résolution du ribozyme VS

Le troisième modèle de basse résolution a été effectué par la méthode de diffusion des rayons X aux petits angles (SAXS), qui permet d'obtenir de l'information structurale sur des macromolécules en solution, mais qui ne permet pas d'avoir des structures de résolution atomique (figure 1.13 B) [116]. Pour les expériences de SAXS, un ribozyme contenant les domaines hélicoïdaux I à VII est utilisé pour la version *cis* et un ribozyme contenant les domaines hélicoïdaux II à VII est utilisé pour la version *trans*. Les expériences de SAXS ont permis de générer des enveloppes moléculaires dans lesquelles les domaines hélicoïdaux ont

été positionnés d'abord sous forme de cylindres et ceux-ci sont ensuite remplacés par des hélices. Le modèle global montre que les deux jonctions occupent une fonction clé dans l'architecture du ribozyme et le modèle est compatible avec les études de modifications chimiques et les études de FRET. De plus, ce modèle place à proximité la boucle du site de clivage de SLI et la boucle A730 de SLVI, ce qui est cohérent avec la formation du site actif pour la réaction de clivage.

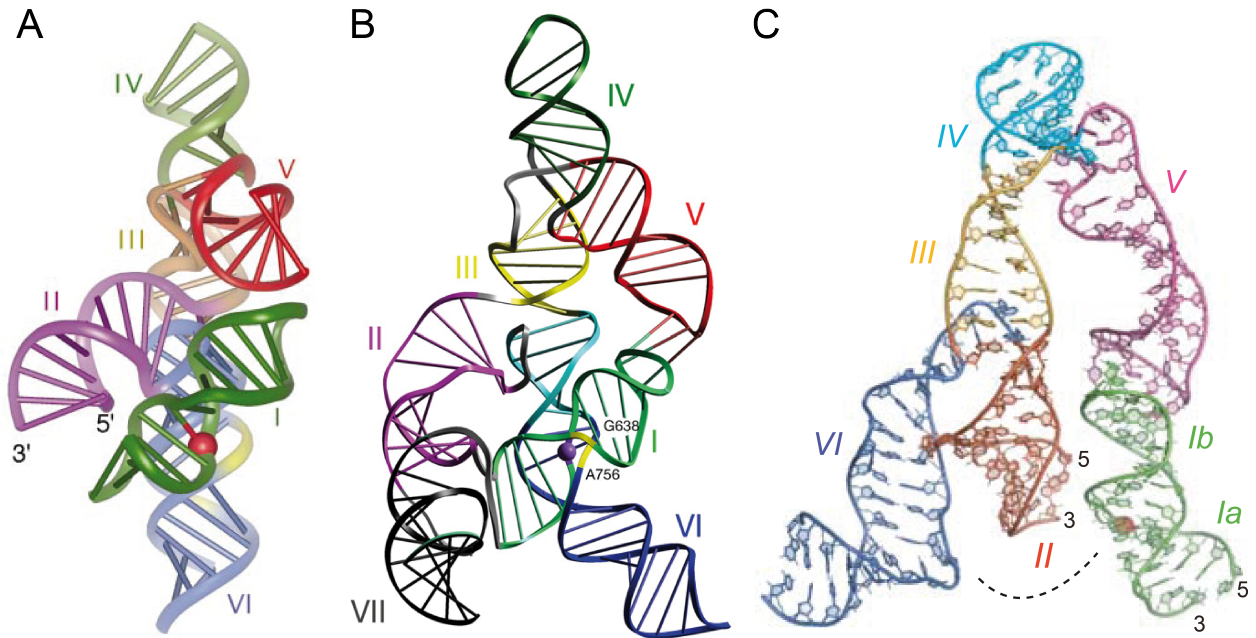


Figure 1.13 Modèles des structures tertiaires du ribozyme VS basés sur le FRET, le SAXS et la RMN.

Les modèles des structures tridimensionnelles sont basés sur des expériences de (A) FRET qui inclut les domaines hélicoïdaux I à VI, de (B) SAXS qui inclut les domaines hélicoïdaux I à VII et de (C) RMN qui inclut les domaines hélicoïdaux I à VI. Le site de clivage est indiqué par une sphère mauve pour le modèle de SAXS et par une sphère rouge pour les modèles de FRET et RMN. (Tirée de [113, 116, 117]).

1.2.5.4 Structures RMN

Face à la difficulté d'obtenir une structure du ribozyme complet à haute résolution, une stratégie de « diviser pour mieux régner » a été employée pour obtenir un modèle global RMN

à partir de structures RMN de haute résolution des sous-domaines isolés. Cette approche fait la supposition que le repliement d'un sous-domaine est comparable entre le contexte du sous-domaine isolé et celui du ribozyme entier.

1.2.5.4.1 Structures RMN des domaines hélicoïdaux I et V

Tout d'abord, les structures RMN de la tige-boucle SLI ont été résolues dans les conformations inactive et active (voir section 1.2.3). Pour la tige-boucle V, deux structures ont été obtenues, en présence et en absence de magnésium [88, 118]. Les similarités entre les structures de la tige-boucle SLV avec et sans magnésium sont la présence d'empilement entre G697, A698 et C699 et la présence d'un virage brusque dans le squelette ribose-phosphate après le U696. Dans les deux structures, il y a la présence d'un motif *U-turn*, qui est relâché en absence de magnésium et il devient compact suite à l'ajout du magnésium qui promeut la formation de ponts H et d'empilements de bases additionnels. La plus grande différence entre les deux structures de SLV est la position du nucléotide U700 qui passe du sillon mineur au sillon majeur, suite à l'ajout de magnésium. En comparant les structures RMN des tige-boucles SLI et SLV isolées et la structure des tige-boucles SLI et SLV dans le complexe I/V (décrite à la section 1.2.2), il y a des changements mineurs au niveau de la boucle V et le substrat SLI passe d'un état libre désordonné à un état plus ordonné dans le complexe. Cet état désordonné peut contribuer à sa reconnaissance par la boucle V pour favoriser l'interaction *kissing-loop* I/V [87].

1.2.5.4.2 Structures RMN du domaine hélicoïdal VI

Il existe deux structures RMN du domaine hélicoïdal VI qui ont été effectuées sur différentes séquences tronquées de SLVI et les structures obtenues représentent des conformations active et inactive de SLVI. La structure RMN de Flinders *et al* représente SLVI dans une conformation inactive, dans laquelle les nucléotides de la boucle A730 sont impliqués dans des paires de bases non canoniques, ce qui est en désaccord avec les données biochimiques. La structure de Desjardins *et al* représente une conformation active et elle est en accord avec les données biochimiques [119, 120]. Dans la structure RMN de la tige-boucle

SLVI active, toutes les paires de bases des hélices prédites dans la structure secondaire sont présentes [119]. La boucle A730 est stabilisée par des ions magnésium et la formation d'une paire de base G-A. Dans la structure de SLVI, il y a la présence d'un motif *S-turn* qui projette les faces W-C des nucléotides C755 et A756 dans le sillon mineur, ce qui est compatible avec le rôle de A756 en tant qu'acide dans la réaction de clivage.

1.2.5.4.3 Structures RMN des jonctions III-IV-V et II-III-VI

Les deux dernières structures RMN obtenues sont celles des jonctions III-IV-V et II-III-VI [117, 121]. Dans les deux structures, les jonctions sont organisées par un réseau complexe d'interactions et les structures sont compatibles avec les données biochimiques. Le rôle principal de la jonction III-IV-V est de positionner la boucle V pour la reconnaissance du substrat SLI. La structure de la jonction III-IV-V inclut un motif *U-turn* canonique composé de U710, G711 et A712. Les tiges III et IV font un empilement coaxial et il y a un angle aigu entre les tiges III et V, ce qui est compatible avec le modèle de SAXS. Dans cette structure, la jonction fonctionne comme une charnière dynamique, qui confère à SLV un certain degré de flexibilité par rapport à la tige III et lui permettrait d'explorer l'espace conformationnel pour la liaison du substrat SLI. La deuxième jonction du ribozyme VS, la jonction II-III-VI, a pour fonction l'orientation des hélices II et VI pour la formation du site actif. La structure de la jonction II-III-VI est stabilisée par la présence d'une interaction tertiaire entre les *bulges* d'adénines des hélices II et VI et par l'empilement coaxial des tiges II et III. L'empilement coaxial entre les tiges II et III n'est pas compatible avec le modèle de FRET pour lequel les *bulges* d'adénines étaient absents [113]. Dans un modèle tridimensionnel du ribozyme complet, basé sur les structures RMN de haute résolution des différents sous-domaines du ribozyme VS, le ribozyme en complexe avec le substrat adopte une conformation ouverte dans laquelle le substrat ne participe pas au site actif du ribozyme (figure 1.13 C). Dans ce modèle, il est possible d'observer un empilement coaxial des domaines hélicoïdaux II, III et IV pour créer un segment d'hélice continue, avec les hélices V et VI qui se retrouvent aux côtés des hélices III et II respectivement. L'interaction *kissing-loop* I/V crée aussi une structure hélicoïdale quasi continue dans ce modèle [117, 121].

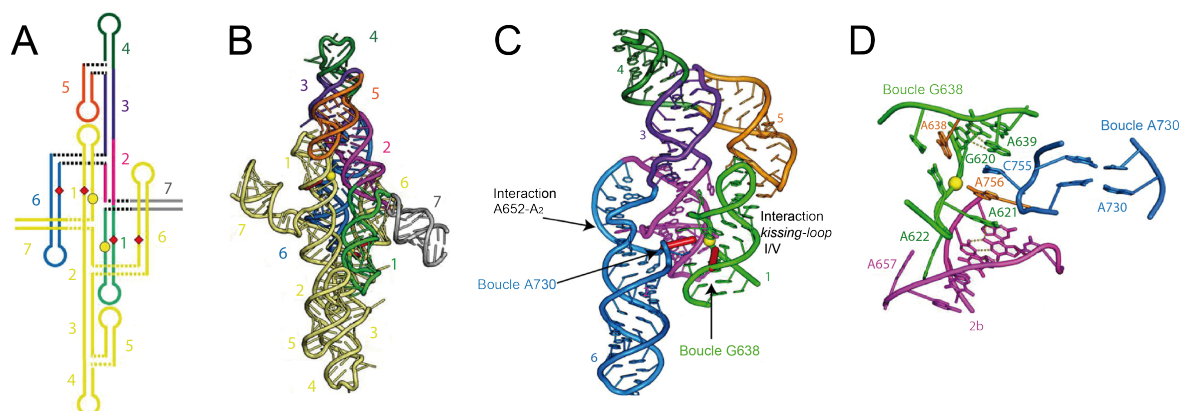


Figure 1.14 Structure cristalline de la forme dimérique du ribozyme VS.

(A) La structure secondaire représentant l'organisation du dimère dans (B) la structure du cristal de la forme dimérique du ribozyme VS. La structure en jaune représente l'un des deux ribozymes du dimère. (C) Structure cristalline du ribozyme en *trans* montrant les domaines hélicoïdaux II à VI de l'un des ribozymes et la tige-boucle SLI du second ribozyme. (D) Structure du site actif du ribozyme VS. Le code de couleur pour les hélices est le même en (A), (B), (C) et (D). Dans les structures cristallines, la sphère jaune représente le site de clivage et les nucléotides en rouge pour (A), (B) et (C) et en orange pour (D) sont impliqués dans la catalyse acide-base générale. (Adaptée de [66]).

1.2.5.5 Structure cristalline du ribozyme VS

Malgré toutes les études structurales effectuées sur le ribozyme VS mentionnées dans les sections précédentes (1.2.5.1 à 1.2.5.4), l'architecture précise du ribozyme VS complet et la configuration du site actif à haute résolution sont demeurées inconnues pendant longtemps, puisqu'il a été difficile d'obtenir une structure haute résolution du ribozyme complet. La structure cristalline du ribozyme VS a été résolue pour la forme dimérique du ribozyme VS complet en *cis* contenant les domaines hélicoïdaux I à VII à une résolution de 3.1 Å (figure 1.14 A et B) [66]. Pour faciliter l'obtention d'une population homogène de ribozymes non clivés, des mutants non fonctionnels (G638A et A756G) ont été cristallisés. Par conséquent, les structures des ribozymes G638A et A756G représentent les structures dans un

état précatalytique. Les structures du cristal révèlent que le ribozyme adopte une forme dimérique symétrique qui contient deux sites actifs hybrides dans lesquels le substrat de l'un des ribozymes est dans le domaine catalytique du second ribozyme. Le repliement global du dimère inclut deux regroupements de plusieurs hélices qui sont similaires et qui consistent à l'empilement des hélices I, II, III et IV du premier ribozyme, qui s'empile avec l'hélice SLV du second ribozyme via l'interaction *kissing-loop* I/V. Les hélices V, VI et VII émergent de chaque empilement à leurs jonctions respectives. Les deux interactions *kissing-loop* I/V présentes dans le dimère permettent une association rapprochée des deux groupements d'hélices. Dans l'interaction *kissing-loop* I/V, il y a la présence de trois paires de bases W-C et d'une paire de base non canonique C629-A701, tel que retrouvé dans la structure RMN du complexe I/V. La jonction I-II-VII permet d'éloigner le substrat SLI de son propre domaine catalytique et de le diriger au domaine catalytique du second ribozyme, pour la formation du dimère. L'orientation du substrat vers l'extérieur du domaine catalytique de son ribozyme est cohérente avec la situation du ribozyme *in vivo*, qui favorise la réaction de ligation pour que le ribozyme soit majoritairement sous une forme circulaire monomérique, qui est la forme la plus stable du ribozyme *in vivo*. Cette orientation du substrat est aussi en accord avec la préférence du ribozyme VS pour le clivage d'un substrat situé en aval de sa séquence, lorsqu'il y a plusieurs copies du substrat dans un même brin [122]. La jonction II-III-VI oriente les hélices II et VI à un angle de 55°, afin de permettre de créer un sillon pour l'amarrage du substrat et la formation du site actif. L'orientation des hélices II et VI est renforcée par une interaction tertiaire entre le *bulge* A652 de l'hélice II et le *bulge* A₂ de l'hélice VI. Le centre catalytique est composé de la boucle interne de l'hélice I, de la boucle A730 de l'hélice VI et du sillon mineur de l'hélice II (figure 1.14 D). L'interaction entre les hélices I et VI permet d'empiler les bases des nucléotides C755 et A756 entre celles des nucléotides A621 et A639. Au site actif, la présence de deux empilements de bases crée un arrangement en forme de V, qui organise G638 et A756 autour du phosphate au site actif et écarte les nucléotides G620 et A621 de chaque côté du site de clivage. Les nucléotides G638 et A756 sont bien placés pour agir dans la catalyse acide-base générale avec G638 au nucléophile 2'-OH et A756 au groupe partant 5'-O, puisqu'ils sont tous les deux non appariés et orientés vers le phosphate du site de clivage situé entre eux. La géométrie alignée (180°) au site de clivage n'est pas respectée pour le mutant G638A avec un angle de 97° entre le 2'-O, le phosphate au site de clivage et le 5'-O,

mais le mutant G756C est plus proche de cette géométrie avec un angle de 129°. Puisque ces mutants représentent un état avant le clivage, il pourrait y avoir des ponts hydrogènes supplémentaires au site actif, qui se forment pour permettre d'atteindre la géométrie alignée et de faciliter le clivage [66].

1.2.5.6 Comparaison entre les structures de haute résolution

Il y a quelques différences importantes entre la structure cristalline et les structures RMN. Les structures obtenues par RMN sont celles de sous-domaines isolés du ribozyme et elles ne sont pas en relation avec les autres éléments du site actif, tandis que la structure cristalline a été effectuée sur le ribozyme complet. Il est donc normal de voir des différences, car les interactions tertiaires dans le ribozyme complet vont influencer la structure, surtout au niveau du site actif dans lequel les tige-boucles I et VI sont impliquées. Dans la boucle A730 de la tige-boucle VI, le modèle du cristal montre un élargissement plus prononcé du sillon mineur, des courbures différentes du squelette phosphate-ribose près du *S-turn*, mais surtout, les nucléotides C755 et A756 sont projetés plus loin à l'extérieur de l'hélice pour interagir avec SLI. Une autre différence attribuable à l'interaction entre les boucles internes I et VI est que dans la structure RMN de la tige-boucle I isolée, les nucléotides A621 et A622 se trouvent à l'intérieur de la boucle interne, tandis que dans la structure cristalline, ils sont sortis de l'hélice pour faire des interactions avec le centre catalytique. Les plus grandes différences entre les structures RMN et cristallines sont observées au niveau des nucléotides impliqués dans la catalyse, ce qui est attribuable majoritairement par le fait que la RMN est effectuée sur des complexes isolés, mais ces différences pourraient aussi refléter une dynamique au niveau de la formation du site actif [66].

1.2.6 Mutagenèse du ribozyme VS

Depuis sa découverte, le ribozyme VS de 881 nucléotides en *cis* découvert chez *Neurospora* a subi de multiples modifications, afin de l'adapter pour l'étude de sa fonction et de sa structure. Le ribozyme VS a tout d'abord été réduit à une séquence minimale active et il a ensuite été modifié pour en faire un ribozyme en *trans* capable de cliver un substrat SLI

isolé. Par la suite, des mutants ont été créés pour le rendre plus rapide ou encore pour favoriser la réaction de clivage ou de ligation. Voici des exemples des principales modifications effectuées sur le ribozyme VS afin de changer son activité.

1.2.6.2 Réduction de la taille du ribozyme VS et mutations pour un clivage plus rapide

Il a été initialement démontré par des expériences de délétion en 5' et en 3', qu'il est possible de réduire la taille du ribozyme VS en *cis* à un seul nucléotide en amont du site de clivage et à environ 150 nucléotides en aval, tout en conservant l'activité de clivage ($k_{\text{obs}} = 0.06-0.08 \text{ min}^{-1}$ à 37 °C) [77]. Par après, des études ont démontré que la taille du ribozyme VS peut être encore plus réduite jusqu'à environ 120 nucléotides et avoir un clivage en *cis* de 30 à 40 fois plus rapide. C'est le cas du ribozyme nommé TR-1 (figure 1.15 A) [86]. Les modifications de TR-1 sont la réduction de la taille des tiges II, IV et VI, l'élimination de la tige Ia du substrat SLI, le déplacement du substrat SLI en 3' du ribozyme et l'ajout de nucléotides entre les hélices I et II ($k_{\text{obs}} = 2.4 \text{ min}^{-1}$ à 37 °C). Il est aussi possible d'améliorer la rapidité de la réaction de clivage en *cis* en effectuant seulement l'insertion d'une séquence pour espacer les hélices I et II ($k_{\text{obs}} = 6-8 \text{ min}^{-1}$ à 37 °C) (figure 1.15 B) [122]. La modification avec le plus grand effet sur la rapidité du clivage est le déplacement du substrat SLI en 3', car un ribozyme RS19 avec des domaines hélicoïdaux de longueurs normales, un substrat en 3' et un espacement entre les hélices I et II donne un clivage en *cis* encore plus rapide que TR-1 (figure 1.15 C) ($k_{\text{obs}} = 3 \text{ s}^{-1}$ à 37 °C) [123]. L'utilisation d'un substrat *preshifted* dans un ribozyme en *cis* (H3) ne change pas l'activité de clivage du ribozyme qui reste similaire à celle d'un substrat activable ($k_{\text{obs}} = 1 \text{ min}^{-1}$ pour les deux à 37 °C) [85].

1.2.6.2 Mutation pour favoriser la réaction de ligation

Il est possible d'apporter des changements au ribozyme VS afin de favoriser la réaction de ligation par rapport à la réaction de clivage. L'ajout de la tige VII à un ribozyme en *cis*, en prolongeant les séquences du ribozyme en 5' et en 3', favorise la réaction de ligation [81]. Une deuxième façon de favoriser la ligation est l'élongation de l'hélice Ia dans le substrat SLI en *trans*. En effet, l'ajout de paires de bases dans l'hélice Ia (jusqu'à 14 pb) favorise la

probabilité de religation en retenant la partie 5' du produit clivé du substrat lors de la réaction de clivage [100].

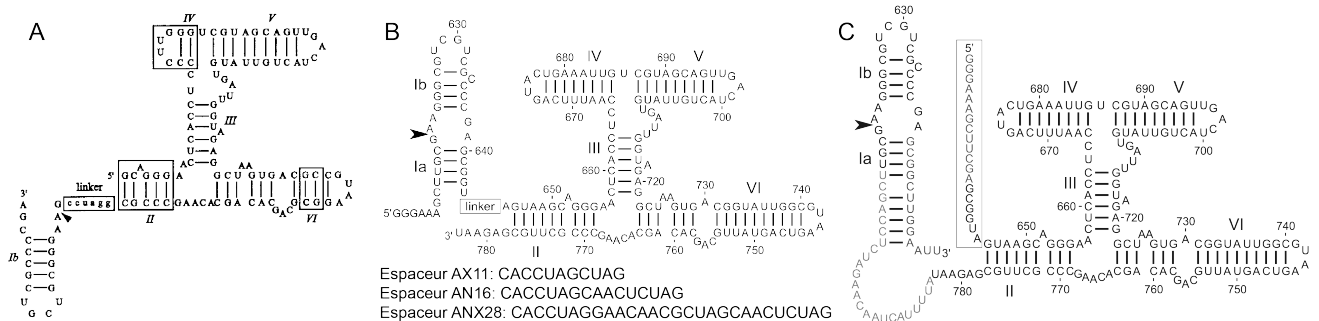


Figure 1.15 Séquences et structures secondaires proposées des ribozymes VS qui clivent rapidement.

Structures secondaires proposées pour les ribozymes qui clivent rapidement (A) TR-1, (B) G11 avec une séquence pour espacer les hélices I et II et (C) RS19. Le site de clivage est indiqué par une flèche noire. Les séquences encadrées en (A), représentent les régions dans lesquelles les domaines hélicoïdaux ont été modifiés. La région encadrée en (B) est la séquence d’espacement dont les séquences sont indiquées sous la figure. (Adaptée de [86] et [123]).

1.3 Ingénierie des ribozymes

Les ribozymes ont tout d’abord été étudiés afin de mieux comprendre leur structure, leur mécanisme de catalyse, ainsi que leurs fonctions biologiques. Il est ensuite possible d’utiliser ces informations dans le but de faire de l’ingénierie de ribozyme, pour le développement d’outils biochimiques et d’agents thérapeutiques. Le but est d’utiliser l’ARN pour développer des outils plus rapides, moins chers et plus fiables que ce qui est utilisé en ce moment [124]. L’ingénierie des ARN peut s’avérer plus facile que l’ingénierie des protéines, car le repliement de ceux-ci est généralement plus simple à comprendre que celui des protéines. De plus, ils sont utilisés plus rapidement pour les stratégies de régulation post-transcriptionnelle et ils drainent moins les ressources de la cellule hôte, puisqu’ils n’ont pas

besoin d'être traduits comme les protéines. Il y a plusieurs molécules d'ARN qui sont en développement comme outils biochimiques ou comme agents thérapeutiques. Ces ARN sont les ribozymes, les ARN interférents (ARNi), les aptamères et les ARN antisens [124].

Les ribozymes sont particulièrement intéressants pour effectuer l'inactivation de l'expression génique, en interagissant au niveau des ARNm ou des ARN viraux, tout comme les ARNi (siARN, miRNA, ARNsh) et les ARN antisens [125-127]. Les ribozymes présentent plusieurs avantages pour faire l'ingénierie puisqu'ils sont très polyvalents. Ils peuvent même servir de plateforme pour ajouter des fonctions de reconnaissance de ligand, afin de servir comme biosenseur ou riborégulateur [128, 129]. Ils peuvent parfois être moins efficaces que les siARN, mais ils ont l'avantage de créer moins d'effets indésirables liés à des erreurs de ciblage (*off-target effect*) dans la cellule [126]. Une façon de réduire les impacts des inconvénients est de faire l'utilisation conjointe de plusieurs stratégies (ribozyme, siARN, protéine, petite molécule, antisens, DNazyme, aptamère...) : premièrement, pour contrer les lacunes de chacune des méthodes et deuxièmement, pour amplifier l'effet thérapeutique, qui sera plus grand qu'avec l'utilisation d'une seule méthode. Un autre avantage des ribozymes est qu'ils peuvent cliver plus d'une fois, donc l'effet du ribozyme est prolongé, comparativement à des outils d'ARN qui agissent avec un ratio de 1:1.

Il y a trois ribozymes qui ont été testés en essais cliniques et ils sont tous dérivés du ribozyme hammerhead, dans lequel plusieurs nucléotides ont été modifiés pour le rendre plus stable. Le premier ribozyme synthétique testé dans des essais cliniques est le ribozyme ANGIOZYME et il s'agit d'un ribozyme contre le cancer. Ce ribozyme régule l'activité du VEGF (*Vascular endothelial growth factor*), en clivant l'ARNm qui encode les protéines du récepteur VEGF primaire, ce qui provoque l'inhibition de l'angiogenèse [125, 130, 131]. Deux autres ribozymes testés en essais cliniques sont le ribozyme OZ1, qui agit contre le virus du VIH [132] et le ribozyme HEPTAZYME, qui cible l'ARNm du génome du virus de l'hépatite C (HCV) [133, 134].

Il existe trois grandes façons de faire l'ingénierie des ribozymes et elles seront décrites dans les prochaines sections : l'ingénierie rationnelle, la sélection *in vitro* et l'ingénierie computationnelle. Il est aussi possible de combiner plusieurs de ces méthodes selon le ribozyme.

1.3.1 Ingénierie rationnelle

L'ingénierie rationnelle est basée sur la connaissance des activités catalytiques de l'ARN [124, 128]. Il s'agit d'utiliser l'activité de catalyse naturelle d'un ARN et d'adapter cette activité à un nouveau contexte pour obtenir une nouvelle activité. Pour cette technique, il est nécessaire de faire preuve de logique afin de modifier des cibles ou de combiner des modules d'ARN. Il faut aussi parfois agir par essais et erreurs, pour arriver au résultat escompté. Voici quelques exemples d'ingénierie rationnelle.

Un bon exemple d'ingénierie rationnelle est de combiner deux modules d'ARN fonctionnels pour créer de nouvelles propriétés. Par exemple, il est possible de combiner un aptamère liant l'ATP à un ribozyme hammerhead, pour que la liaison de l'ATP contrôle la catalyse du ribozyme [128]. La liaison de l'ATP va induire un changement de conformation de l'aptamère, qui va inhiber l'activité du ribozyme hammerhead. L'inverse a aussi été fait, avec une tige-boucle du ribozyme hammerhead instable, qui devient stable et activé, suite au changement de conformation de l'aptamère provoqué par la liaison à l'ATP [128].

Un autre exemple de combinaisons de modules d'ARN est le ribozyme twin, qui est un ribozyme qui contient deux séquences du ribozyme hairpin. Il est capable de faire le clivage d'un ARN à deux sites spécifiques différents [135-137]. Puisque le ribozyme est aussi capable de faire la ligation, il est possible d'insérer une séquence différente en remplacement de la séquence clivée de l'ARN cible. Par conséquent, le ribozyme twin peut être utilisé pour effectuer de petites réparations dans des substrats d'ARN ou pour insérer une séquence d'ARN synthétique marqué pour faire du marquage à des sites spécifiques.

Il est possible de combiner plus de deux modules d'ARN, comme avec les ribozymes multitarget, qui sont composés de plusieurs séquences (jusqu'à neuf) du ribozyme hammerhead et chacune possédant une séquence cible différente [138]. Le ribozyme multitarget possède la même spécificité qu'un seul ribozyme hammerhead, mais il est plus efficace pour cliver une séquence d'ARN cible.

Il est possible d'utiliser la température pour contrôler l'activité d'un gène en utilisant un ARN thermomètre, qui est composé d'une tige-boucle contenant la séquence Shine-Dalgarno [139, 140]. Un changement de température modifie la stabilité de la tige-boucle, ce qui expose le site de liaison du ribosome et permet l'expression du gène en aval. Il est possible de coupler cet ARN thermomètre à un ribozyme hammerhead, pour créer un thermozyyme, un ribozyme contrôlé par la température. Dans ce cas, c'est le clivage du ribozyme qui expose le site de liaison du ribosome pour permettre l'expression du gène et un changement de température inhibe l'expression du gène en empêchant le clivage du ribozyme.

1.3.2 Ingénierie par SELEX

La sélection *in vitro* ou SELEX (*systematic evolution of ligands by exponential enrichment*) est une technique permettant d'identifier des aptamères pour la liaison de ligand avec une grande affinité [141-143]. Les aptamères générés par SELEX peuvent lier une gamme variée de ligands comme des nucléotides, des peptides, des protéines, des petits composés organiques ou des récepteurs. Le SELEX permet aussi de modifier l'activité de catalyse des ribozymes et de créer des ribozymes artificiels qui vont exprimer de nouvelles propriétés catalytiques. Le grand avantage du SELEX, c'est qu'il permet l'exploration de propriétés insoupçonnées de l'ARN. Les expériences de SELEX s'effectuent avec l'utilisation d'une librairie d'ARN ou d'ADN contenant des séquences aléatoires et qui va évoluer dans des conditions de plus en plus strictes, jusqu'à l'obtention des meilleurs candidats pour l'activité recherchée.

Le SELEX a été utilisé dans les expériences sur le monde à ARN, en créant de nouvelles classes de ribozymes. Ces ribozymes artificiels ont des activités compatibles avec un monde dans lequel l'ARN aurait assumé toutes les fonctions nécessaires à l'apparition de la vie. Ces ribozymes démontrent des activités de synthèse de nucléosides, de triphosphorylation et de polymérisation de nucléoside triphosphates [26].

Le SELEX est fréquemment utilisé pour créer des senseurs pour toutes sortes de molécules ou comme agent thérapeutique pour cibler des ARN liés à des maladies. Un exemple de l'utilisation du SELEX pour le développement d'un agent thérapeutique, est l'aptamère pegaptanib sodium (ou Macugen) [144]. Il est le premier aptamère approuvé pour l'usage chez l'humain par la FDA en 2004 et il est utilisé comme médicament contre la dégénérescence maculaire liée à l'âge.

1.3.3 Ingénierie computationnelle

Le développement d'outils computationnels a pour but de faciliter le design d'ARN et de ribozymes pour des tâches précises. Ils sont très pratiques pour faire des prédictions là où l'ingénierie rationnelle ne suffit pas pour la conception de nouveaux ribozymes. Par contre, ces outils sont parfois peu précis et par conséquent, il devient nécessaire de générer expérimentalement plusieurs ARN et les tester, afin de dépister ceux qui vont manifester l'activité désirée.

Un exemple d'ingénierie computationnelle est de créer de nouveaux motifs de ribozymes hammerhead capables de clivage ou encore pour prédire des motifs de ribozymes hammerhead pouvant servir de senseurs pour des oligonucléotides ou des petites molécules [31, 145]. Cette méthode permet d'obtenir des séquences plus rapidement que par le SELEX et ces séquences peuvent ensuite être testées en laboratoire [31].

Un autre exemple d'outils basé sur l'ingénierie computationnelle et qui est très utilisé est l'outil fourni par ThermoFisher scientifique pour le design des miARN et des ARNsh [124]. En donnant la séquence du gène cible, les programmes vont prédire des régions permettant une répression spécifique et efficace du gène par les ARNi. Par la suite, le programme va faire le design d'un miARN ou d'un ARNsh qui cible les séquences d'intérêts et qui contient les éléments structuraux nécessaires pour la biogenèse, la fonction, ainsi que les amorces pour la construction et le clonage.

1.3.4 Ingénierie des ribozymes autoclivants

Le ribozyme hammerhead est souvent utilisé en ingénierie puisqu'il possède une petite taille, une structure bien documentée et la séquence d'ARN cible doit seulement posséder une séquence NUH (N est n'importe quel nucléotide, U est une uridine et H est n'importe quel nucléotide à l'exception d'un G) pour permettre sa reconnaissance par le ribozyme (voir figure 1.16 A) [146]. La reconnaissance de l'ARN cible se fait par l'appariement des nucléotides complémentaires, qui forment les hélices I et III pour ensuite permettre le clivage. La spécificité et le niveau d'activité du ribozyme sont influencés par les longueurs de ces hélices qui varient de 7 nucléotides à plus de 30 nucléotides. Les régions ciblées doivent être accessibles pour permettre la liaison avec le ribozyme. L'accessibilité à ces régions est déterminée par la prédiction computationnelle des structures secondaires, par RNase H et par Mfold. Il y a quatre conditions à remplir qui vont affecter l'efficacité du ribozyme hammerhead : (1) l'identification d'un site de clivage NUH dans l'ARN cible (2) la sélection des séquences autour du site de clivage pour la liaison du ribozyme en respectant l'énergie d'appariement et l'unicité des séquences (3) l'accessibilité du site à lier dans l'ARN cible et (4) la vérification de la séquence du ribozyme pour remplir les exigences de structure fonction de la catalyse [147]. Il existe même des méthodes computationnelles qui permettent de réaliser toutes ces étapes [148].

Le ribozyme hairpin a aussi été utilisé dans le but de cibler des ARN cibles et tout comme pour le ribozyme hammerhead, il se lie à un substrat simple brin par la formation de

paires de bases, pour former deux hélices placées de chaque côté de la boucle A contenant le site de clivage (figure 1.16 B) [149-152]. La séquence requise chez le substrat pour permettre le clivage est RBNGHYB, dans laquelle R est une purine, B est n'importe quel nucléotide sauf A, N est n'importe quel nucléotide, G une guanine, H est n'importe quel nucléotide sauf G et Y est une pyrimidine.

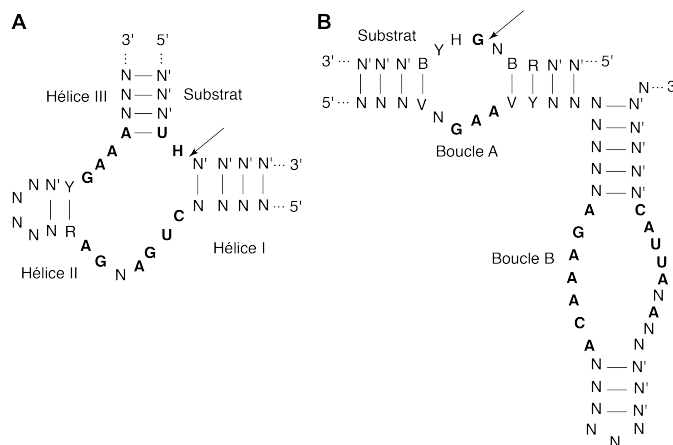


Figure 1.16 Séquences des ribozymes hammerhead et hairpin pour l'ingénierie.

Séquences des structures secondaires représentant les ribozymes (A) hammerhead et (B) hairpin en complexe avec un substrat cible. Le site de clivage est indiqué par une flèche noire, les nucléotides invariables sont indiqués en gras, les Y sont des pyrimidines, R des purines, H n'importe quel nucléotide sauf G, B n'importe quel nucléotide sauf A, V n'importe quels nucléotides sauf U, N et N' indiquent les nucléotides complémentaires. (Adaptée de [146]).

Le ribozyme HDV a été moins utilisé que le hammerhead et le hairpin pour cibler des ARN cibles. Le ribozyme HDV a été utilisé en le couplant avec un biosenseur, qui permet de réguler l'activité du ribozyme par la liaison d'une séquence d'ARN et qui se nomme le ribozyme SOFA-HDV [153-159]. Le ribozyme SOFA-HDV se lie à une première séquence d'ARN cible et suite à l'appariement, le ribozyme libère la séquence du ribozyme responsable de la reconnaissance d'une deuxième séquence d'ARN cible, qui contient le site de clivage. Cette façon de reconnaître l'ARN cible permet un clivage plus spécifique de l'ARN cible.

1.4 Objectifs de la thèse

Les petits ribozymes autoclivants sont considérés comme de bons candidats pour faire l'ingénierie de l'ARN. Les ribozymes les plus utilisés pour l'ingénierie sont les ribozymes hammerhead et hairpin. Ces ribozymes font la reconnaissance d'un substrat simple brin, par la formation de structures secondaires stables, pour ensuite effectuer la réaction de clivage de façon spécifique (voir figure 1.16). Le ribozyme VS ne représente pas un candidat idéal pour la reconnaissance d'un substrat simple brin, puisqu'il fait la reconnaissance de son substrat par une interaction tertiaire *kissing-loop*. Le ciblage d'un ARN simple brin par les ribozymes peut s'avérer difficile *in vivo*, puisque la majorité des ARN adoptent des structures secondaires et tertiaires complexes. L'énergie requise pour déplier ces structures d'ARN peut empêcher la formation d'une interaction entre le ribozyme et le substrat et par conséquent, des régions d'ARN peuvent être inaccessibles aux ribozymes. Pour le ciblage des ARN repliés, le ribozyme VS pourrait représenter un bon candidat, puisqu'il fait la reconnaissance d'un substrat d'ARN replié en tige-boucle, ce que les ribozymes hammerhead et hairpin ne peuvent pas faire. Jusqu'à présent, le ribozyme VS n'a pas été exploité pour faire la reconnaissance d'ARN cibles repliés. Les tige-boucles sont des cibles intéressantes, puisqu'elles sont importantes non seulement pour stabiliser la structure des ARN, mais aussi pour différentes fonctions biologiques. Les tige-boucles peuvent servir de site de reconnaissance pour des protéines comme dans le cas de l'ARN TAR du VIH-1, qui lié par la protéine tat permet le recrutement de la machinerie de transcription. Une autre fonction des tige-boucles est de permettre l'initiation de la dimérisation de l'ARN viral, comme la tige-boucle de l'ARN DIS dans le VIH-1 [160, 161]. L'hypothèse principale de cette thèse est que le ribozyme VS est un bon candidat pour l'ingénierie de ribozymes permettant la reconnaissance et le clivage spécifique d'ARN cibles repliés en tige-boucles. Cette hypothèse a été directement testée dans les trois prochains chapitres de cette thèse. Les chapitres 2.0 et 3.0 ont été effectués en collaboration avec Nicolas Girard, qui s'est occupé du volet bio-informatique des études. Le volet bio-informatique a permis d'enrichir l'étude par la recherche de motifs d'ARN dans les bases de données et pour l'élaboration de modèles des structures d'ARN pour expliquer les résultats des différentes études.

Lors d'études préliminaires auxquelles j'ai participé à titre de co-première auteure et présentées à l'Annexe 1 [89], l'objectif était de déterminer le rôle de certains nucléotides de la boucle V de l'interaction *kissing-loop* en effectuant des mutations de ceux-ci. Les études de cinétiques ont révélé qu'il est possible de modifier l'identité de certains nucléotides de la boucle V de l'interaction *kissing-loop* I/V et même d'ajouter un nucléotide dans la boucle, tout en conservant une activité de clivage similaire. Comme au moment de ces études, il n'y avait pas encore d'informations structurales à haute résolution sur l'interaction *kissing-loop* I/V, un modèle tridimensionnel d'un complexe SLI/SLV a été produit pour interpréter les résultats de mutagenèse [89]. En résumé, ces études préliminaires ont montré que plusieurs modifications au niveau de l'interaction *kissing-loop* sont compatibles avec une activité de clivage optimale et étaient donc prometteuses pour la suite du projet.

Dans le chapitre 2.0, l'objectif était de tester s'il est possible de modifier le ribozyme VS, afin qu'il puisse reconnaître et cliver des substrats avec différentes longueurs de tige Ib. Puisque SLV fait la reconnaissance du substrat SLI et facilite la formation du site actif du ribozyme, il était prévu que le changement de longueur de SLI devrait avoir une incidence sur la formation du site actif et l'activité de clivage. Le modèle du complexe SLI/SLV publié précédemment a permis d'émettre l'hypothèse que les hélices I et V sont coaxiales dans l'interaction *kissing-loop* I/V [89]. Si les hélices sont effectivement coaxiales, il devait être possible de changer la longueur de la tige Ib du substrat, en effectuant une mutation compensatoire dans la tige V, afin de conserver le même nombre de paires de bases total dans l'interaction *kissing-loop* I/V et permettre la formation du site actif. Des cinétiques avec les substrats SLI mutants possédant différentes longueurs de tige Ib et des ribozymes mutants avec différentes longueurs de tiges V ont montré que la compensation des longueurs d'hélices dans l'interaction *kissing-loop* I/V est possible pour plusieurs paires de substrat/ribozyme mutants dans un certain intervalle. Ce projet a donc montré qu'il existe une certaine adaptabilité au niveau de l'interaction *kissing-loop* I/V permettant d'apporter des modifications dans les tiges I et V, sans affecter l'activité de clivage du ribozyme.

Au chapitre 3.0, l'objectif était de continuer à tester l'adaptabilité de l'interaction *kissing-loop* I/V et de pousser l'ingénierie plus loin en substituant l'interaction *kissing-loop* I/V par des interactions *kissing-loop* dont les structures se trouvent dans les bases de données (*Protein data bank*). Puisque l'interaction *kissing-loop* est considérée comme une unité modulaire du ribozyme VS, le remplacement de l'interaction ne devrait pas avoir d'incidence majeure sur le reste du ribozyme. Cet objectif a été testé en effectuant la synthèse de substrats et de ribozymes mutants, dont les boucles I et V sont substituées par les interactions *kissing-loop* TAR/TAR* de l'ARN du VIH-1 et L22/L88 de l'ARN 23S de *Deinococcus radiodurans*, sélectionnées par le biais d'analyses bio-informatiques pour leur compatibilité avec l'interaction I/V. Des cinétiques effectuées avec ces substrats et ribozymes mutants montrent que la réaction de clivage est toujours possible, mais avec une activité réduite. Une modélisation des interactions *kissing-loop* testées suggère que l'interaction a une influence sur l'orientation du substrat et que ces interactions *kissing-loop* contenant plus de paires de base que l'interaction I/V naturelle contraignent partiellement la formation du site actif.

Dans le chapitre 4.0, l'objectif était d'utiliser le SELEX pour modifier la boucle V du ribozyme, afin qu'il fasse la reconnaissance et le clivage d'un substrat S_{TAR} composé des hélices et de la boucle terminale de l'ARN TAR du VIH-1. Le SELEX permet de tester un grand nombre de séquences de la boucle V pour déterminer celles qui sont le mieux adaptées au clivage du substrat S_{TAR} dans le contexte du ribozyme VS. Le SELEX a généré plusieurs séquences de boucles V compatibles avec le clivage du substrat et elles ont permis d'établir une séquence consensus qui permet la liaison et le clivage de l'ARN S_{TAR} dans le contexte du ribozyme VS.

Pour faire suite à ces expériences, nous sommes en mesure de confirmer que le ribozyme VS est un bon candidat pour l'ingénierie de ribozymes permettant la reconnaissance et le clivage spécifique d'ARN cibles repliés en tige-boucles. En effet, ces expériences nous ont permis de connaître les capacités d'adaptation du ribozyme VS pour la reconnaissance et le clivage de nouveaux substrats, ce qui a permis d'établir les bases de l'ingénierie du ribozyme VS. Ces informations sont cruciales afin de poursuivre les expériences d'ingénierie à un niveau supérieur pour réussir à cliver des substrats cibles plus complexes.

1.5 Techniques utilisées

Les expériences décrites dans les chapitres 2.0 et 3.0 ont été effectuées en collaboration avec Nicolas Girard, qui est responsable du volet bio-informatique des études. Je me suis occupée des expériences en laboratoire qui sont décrites dans les prochaines sections (1.5.1 à 1.5.4). Il s'agit de la purification des ARN, des cinétiques enzymatiques en *single-turnover*, des expériences de SHAPE et de SELEX.

1.5.1 Purification de l'ARN

Les substrats du ribozyme VS utilisés dans les différents projets sont transcrits *in vitro* par l'ARN polymérase T7, à partir de gabarits d'ADN partiellement simple brin synthétisé chimiquement. Les ARN sont ensuite purifiés par gels d'électrophorèse en conditions dénaturantes et déphosphorylés avec une phosphatase pour les substrats marqués en 5' avec de la radioactivité. Les ribozymes sont aussi transcrits *in vitro* par l'ARN polymérase T7, mais à partir de plasmides préparés par mutagenèse et coupés avec une enzyme de restriction. Les ribozymes sont ensuite purifiés par gels d'électrophorèse en conditions dénaturantes.

1.5.2 Cinétiques enzymatiques en *single-turnover*

Les réactions de cinétiques enzymatiques sont effectuées dans des conditions dans lesquelles un seul acte catalytique est observé (*single-turnover*), c'est-à-dire, dans des concentrations de ribozyme beaucoup plus élevées que celles du substrat [162]. Dans ces conditions, chaque ribozyme agit sur un seul substrat et elle ignore la dissociation des produits. La cinétique enzymatique est représentative des événements allant de la liaison du substrat, jusqu'à la réaction chimique. C'est un type de cinétique enzymatique représentatif des ribozymes autoclivants, qui performent généralement un seul acte catalytique dans leur fonction biologique. Ces cinétiques enzymatiques permettent de déterminer la constante de vitesse de second ordre k_{cat}/K_M , qui donne une mesure de l'efficacité catalytique de la réaction. Cette constante représente également la spécificité du ribozyme pour le substrat. Le k_{cat}/K_M est

une constante idéale pour comparer différentes combinaisons de ribozymes et de substrats, puisqu'elle inclut autant la réaction chimique que la liaison du substrat.

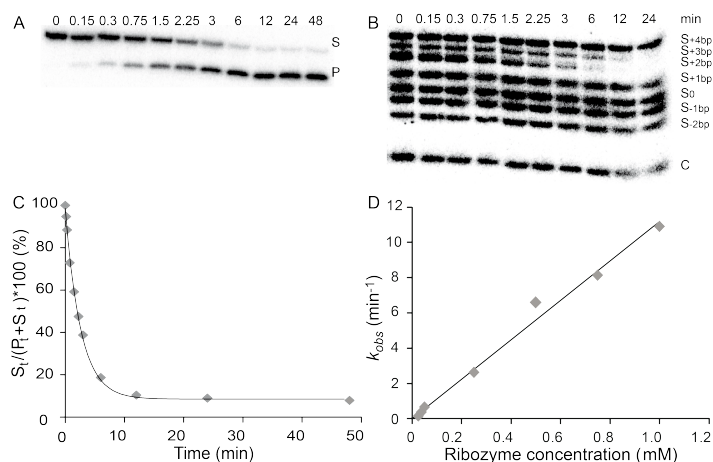


Figure 1.17 Exemple d'études de cinétiques enzymatiques du ribozyme VS.

(A) Le clivage d'un substrat S_0 (250 pM) marqué avec la radioactivité par un ribozyme R_0 (37.5 nM) à 30 °C et 25 mM $MgCl_2$, est suivi par électrophorèse sur gel en conditions dénaturantes. (B) Le clivage d'un mélange de substrats et d'un ADN contrôle par le ribozyme R_{3bp} est suivi par électrophorèse sur gel en conditions dénaturantes. (C) Graphique du pourcentage de substrat restant en fonction du temps. Les données suivent l'équation non-linéaire $F = Ae^{-k_{obs}t} + F_{\infty}$. Le k_{obs} est de $0.332 \pm 0.008 \text{ min}^{-1}$ et le $R^2 = 1.00$ pour le clivage montré en (A). (D) Graphique des k_{obs} en fonction de la concentration de ribozyme. La pente de cette droite donne le k_{cat}/K_M qui est de $11.2 \text{ min}^{-1} \mu\text{M}^{-1}$ et le $R^2 = 0.99$. (Adaptée de [163]).

Les cinétiques enzymatiques pour les différents projets sont effectuées en *trans* avec le substrat SLI synthétisé séparément du ribozyme contenant les domaines hélicoïdaux II à VI (à l'exception des expériences de SELEX qui seront décrites plus tard). La radioactivité est utilisée afin de suivre directement le clivage du substrat et le quantifier. Le marquage perturbe minimalement le substrat, puisqu'il est marqué au ^{32}P à l'extrémité 5'. Les cinétiques enzymatiques sont effectuées en incubant le substrat et le ribozyme séparément, dans le tampon de clivage et au temps zéro, ils sont mélangés et des échantillons sont prélevés à différents temps. Le pourcentage de substrat est calculé pour chaque temps, pour tracer le graphique du pourcentage de substrat en fonction du temps, permettant de déterminer la

constante de vitesse de la réaction k_{obs} à l'aide de l'équation $F = Ae^{-k_{obs}t} + F_{\infty}$ (F est le pourcentage de substrat, A est l'amplitude, t est le temps, k_{obs} est la constante de vitesse et F_{∞} est le pourcentage de substrat qui ne réagit pas au cours de la réaction) (figure 1.17 A et C). Des k_{obs} sont définis pour plusieurs concentrations de ribozyme pour tracer le graphique des k_{obs} en fonction de la concentration de ribozyme et le k_{cat}/K_M est déterminé par la pente de la droite obtenue aux plus faibles concentrations de substrats (figure 1.17 D).

L'avantage d'effectuer les cinétiques enzymatiques en *single-turnover* est qu'elle permet d'effectuer des expériences avec un mélange de plusieurs substrats, comme pour les expériences du chapitre 2.0. Pour ces expériences, afin d'économiser du temps et de limiter le nombre de gels à faire, nous avons mis un ribozyme donné en présence d'un mélange de plusieurs substrats. Puisque les substrats sont de longueurs différentes, il est facile de les séparer sur gel dénaturant et de pouvoir suivre les réactions de clivage de chacun des substrats (figure 1.17 B). Il n'y a pas de problème de compétition entre les différents substrats, puisque dans les conditions de *single-turnover*, la concentration de ribozyme est en excès par rapport à celle du mélange de substrats. Cette méthode a été mise au point pour ce projet et les résultats ont été validés pour s'assurer que les valeurs déterminées avec les substrats multiples correspondent à celles des cinétiques enzymatiques à un seul substrat.

Les expériences de SELEX nécessitent d'avoir un ribozyme en *cis*, composé des domaines hélicoïdaux I à VI, puisque la région avec les nucléotides aléatoires doit être dans le même brin d'ARN que le site de clivage. Dans l'optique d'avoir un meilleur contrôle sur l'activité de clivage, un ribozyme en deux parties a été développé pour ces études (figure 1.18 A). Dans ce système, le substrat contient les domaines hélicoïdaux I à VI, mais avec un seul brin des domaines hélicoïdaux II et VI. Le ribozyme est un ARN complémentaire aux tiges II et VI et il complète les tiges II et VI pour former le ribozyme complet fonctionnel. Pour accomplir les cinétiques enzymatiques, les brins sont préalablement associés ensemble et incubés dans le tampon de clivage sans magnésium. La réaction de clivage est ensuite initiée par l'ajout du magnésium.

1.5.3 SHAPE

Le SHAPE (*selective 2'hydroxyl acylation analyzed by primer extension*) est une technique qui permet d'avoir de l'information sur les structures secondaire et tertiaire d'un ARN. Cette technique consiste à mettre le ribozyme en présence d'un composé qui effectue une modification chimique seulement sur les nucléotides dont le 2'OH est flexible, ce qui signifie que les nucléotides impliqués dans une paire de bases ou une interaction tertiaire stable ne seront pas modifiés [164, 165]. Pour effectuer le SHAPE sur les ribozymes, l'anhydride *N*-méthylisatoïque (NMIA) a été utilisé et les conditions d'expérimentation ont été optimisées pour qu'un seul nucléotide soit modifié par molécule. Par la suite, la transcription inverse est effectuée et la polymérase arrête un nucléotide en amont de la modification par le NMIA. Les résultats sont ensuite visualisés sur gel de polyacrylamide dénaturant. La quantification des bandes avec et sans NMIA permet d'identifier les nucléotides présentant le plus de flexibilité dans la structure de l'ARN. Dans le cas de nos expériences, le SHAPE a été utilisé pour s'assurer que le repliement des différents ribozymes mutants correspond au repliement du ribozyme VS non muté.

1.5.4 SELEX

Le SELEX (voir section 1.3.2) a été utilisé sur le ribozyme VS dans le but de déterminer une interaction *kissing-loop* optimale pour la reconnaissance et le clivage d'un substrat modifié portant la boucle terminale et les hélices Ia et Ib de l'ARN TAR. Pour ces expériences, un système de ribozyme en deux parties a été utilisé (section 1.5.2 et figure 1.18 A). Les nucléotides de la région de la boucle V ont été changés de façon aléatoire, dans le but d'obtenir une librairie contenant toutes les possibilités de nucléotides à chacune des dix positions mutées. Les expériences débutent avec la librairie d'ADN aléatoires amplifiée par PCR et transcrite en ARN, pour obtenir la librairie de ribozymes (figure 1.18 B). La réaction de clivage est initiée par l'ajout de magnésium et le temps de clivage est choisi en fonction du pourcentage de clivage des ribozymes les plus rapides de la librairie [166]. Les ribozymes actifs (clivés) sont ensuite purifiés sur gel de polyacrylamide dénaturant. Une transcription inverse est ensuite effectuée, pour obtenir l'ADN correspondant et recommencer un nouveau

cycle de SELEX. À chaque cycle, les ribozymes qui clivent le plus rapidement sont isolés et le temps de clivage est raccourci d'un cycle à l'autre. Lorsque le groupe de ribozymes actifs a cessé d'évoluer, les meilleurs candidats sont clonés et séquencés.

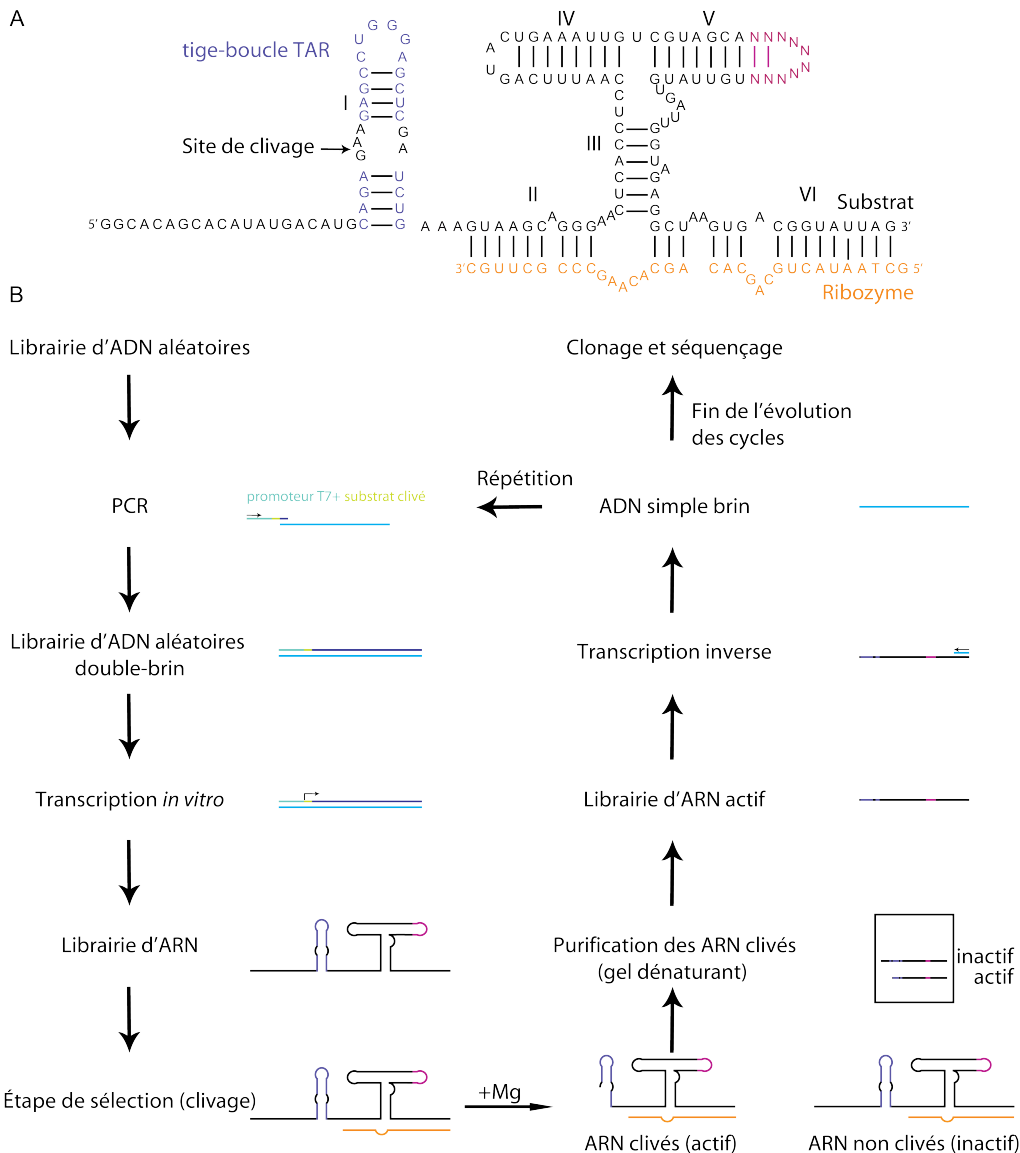


Figure 1.18 Études de SELEX du ribozyme VS.

(A) Séquence du ribozyme VS en deux parties utilisé pour les études de SELEX. Le brin jaune représente la partie ribozyme et le brin noir-mauve-rose représente la partie substrat. Les nucléotides de l'ARN TAR du VIH-1 sont en mauve dans SLI et les nucléotides de SLV qui sont changés de façon aléatoire sont en rose. (B) Schéma représentant les étapes d'un cycle de SELEX.

2.0 Études de compensation de la longueur des hélices chez le ribozyme VS

Mise en contexte

L'objectif de ce premier article est de modifier le ribozyme VS, afin qu'il soit capable de reconnaître et cliver des substrats SLI avec différentes longueurs de tige Ib. Les études préliminaires (en annexe 1) ont démontré que plusieurs mutations étaient tolérées au niveau de l'interaction *kissing-loop*, et elles ont aussi introduit un modèle de l'interaction *kissing-loop* entre les tiges-boucles SLI et SLV. Ce modèle propose que les hélices I et V soient coaxiales, ce qui suggère qu'il est possible de changer les longueurs des tiges Ib et V, pour autant que le nombre de paires de bases total dans les hélices Ib et V reste le même. De cette façon, l'interaction *kissing-loop* pourrait être déplacée, mais le site de clivage du ribozyme demeurerait au même endroit et la formation du site actif serait possible. Des ribozymes mutants avec différentes longueurs de tige V ont été utilisés pour compenser les changements apportés à la longueur des tiges Ib des substrats SLI. Les longueurs de tige Ib testées vont de la délétion de 1 et 2 pb dans la tige Ib de SLI, jusqu'à l'addition de 4 pb. Les ribozymes avec les mutations compensatoires ont des longueurs de tiges V variant de l'addition de 2 pb jusqu'à la délétion de 4 pb.

Les études de cinétiques ont été entreprises pour tester l'activité de clivage de toutes les combinaisons de paires de substrat/ribozyme et ces études ont révélé qu'un même ribozyme peut faire le clivage de trois substrats avec des longueurs de tige Ib différentes et cela avec une activité similaire. De plus, les mutations compensatoires ont bien fonctionné, puisque chaque ribozyme reconnaît de façon préférentielle le substrat avec la mutation compensatoire ou un substrat avec une paire de base de différence. Un nouveau modèle de l'interaction *kissing-loop* I/V a été généré, en accord avec les résultats des cinétiques enzymatiques, et celui-ci permet de constater que contrairement à ce que l'on croyait au départ, l'interaction entre SLI et SLV n'est pas coaxiale, mais plutôt colinéaire. Ce modèle a permis de prédire l'angle entre les hélices et d'illustrer l'adaptabilité de l'architecture du ribozyme VS pour le clivage de substrats mutants. Ce projet a aussi permis d'élaborer une nouvelle méthode de cinétique avec des substrats multiples, qui permet d'effectuer plusieurs cinétiques simultanément, pour économiser du matériel et surtout beaucoup de temps d'expérimentation.

Pour cet article, je suis co-premier auteur avec Nicolas Girard. Les études de modélisation de l'interaction *kissing-loop* I/V ont été effectuées par Nicolas Girard et j'ai été la responsable des études de cinétiques enzymatiques. Dans la partie expérimentale, j'ai effectué la transcription *in vitro* des différents ribozymes mutants à l'aide de plasmides et des substrats mutants à l'aide de gabarits d'ADN. J'ai ensuite mis au point les protocoles pour les cinétiques, analysé les résultats des expériences de cinétiques, vérifié les conformations des substrats par gel natif et la localisation de leur site de clivage avec des digestions à la RNase T1 et par hydrolyse alcaline. J'ai participé à l'écriture des sections du manuscrit pertinentes à ma participation scientifique et Nicolas Girard a fait de même. Pascale Legault a conçu le projet et a supervisé l'ensemble de l'étude, incluant la collecte des données et l'analyse des résultats. Elle a relu et corrigé ce manuscrit.

Helix-Length Compensation Studies Reveal the Adaptability of the VS Ribozyme Architecture

Julie Lacroix-Labonté^{1,3}, Nicolas Girard^{1,3}, Sébastien Lemieux² and Pascale Legault^{1,}*

¹Département de Biochimie, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, QC, Canada, H3C 3J7

²Département d'Informatique et de Recherche Opérationnelle et Institut de Recherche en Immunologie et en Cancérologie, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, QC, Canada, H3C 3J7

³The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

*To whom correspondence should be addressed: Pascale Legault, Phone: 514-343-7326, Fax: 514-343-2210.

This research was originally published in NAR journal:

Nucleic Acids Res. 2012 Mar; 40 (5) : 2284-93. doi: 10.1093/nar/gkr1018. Epub 2011 Nov 15.

ABSTRACT

Compensatory mutations in RNA are generally regarded as those that maintain base pairing, and their identification forms the basis of phylogenetic predictions of RNA secondary structure. However, other types of compensatory mutations can provide higher-order structural and evolutionary information. Here, we present a helix-length compensation study for investigating structure-function relationships in RNA. The approach is demonstrated for stem-loop I and stem-loop V of the *Neurospora* VS ribozyme, which form a kissing-loop interaction important for substrate recognition. To rapidly characterize the substrate specificity (k_{cat}/K_M) of several substrate/ribozyme pairs, a procedure was established for simultaneous kinetic characterization of multiple substrates. Several active substrate/ribozyme pairs were identified, indicating the presence of limited substrate promiscuity for stem Ib variants and helix-length compensation between stems Ib and V. 3D models of the I/V interaction were generated that are compatible with the kinetic data. These models further illustrate the adaptability of the VS ribozyme architecture for substrate cleavage and provide global structural information on the I/V kissing-loop interaction. By exploring higher-order compensatory mutations in RNA our approach brings a deeper understanding of the adaptability of RNA structure, while opening new avenues for RNA research.

Keywords: compensatory mutations, helix-length compensation, *Neurospora* VS ribozyme, multiple-substrate kinetic study, RNA structure adaptability

INTRODUCTION

Compensatory mutations play an important role in the evolution of macromolecular systems. Their identification has proven to be important for the characterization of macromolecular structure and for revealing important aspects of function. Studies of compensatory mutations in functional RNA molecules are particularly informative, because of the simple base-pairing rules that govern the formation of RNA helical structures. Compensatory mutations in RNA are therefore generally seen as those that maintain base pairing interactions, and their identification has been central for comparative sequence analyses in the prediction of RNA secondary structures and for phylogenetic analyses.

Compensatory mutations in RNA are not strictly limited to base-pairing interactions. Other types of structural compensations, such as helix-length compensation, can help provide higher-order structural and evolutionary information. The concept of helix-length compensation was first introduced in the prediction of co-axially stacked helices for rRNAs [1]. It was suggested that since co-axially stacked helices are expected to maintain a constant overall length in a functional RNA architecture, comparative sequence analysis could support co-axially stacked helices when one helix in a group of organisms is shorter, while its partner helix in that group is longer by the same number of base pairs [1]. In support of this theory, comparative sequence analysis previously predicted the existence of co-axially stacked helices in 23S rRNA (helices 2646-2652/2668-2674 and 2675-2680/2727-2732; [2]), which was subsequently confirmed in the X-ray structure of the 50S rRNA subunit [167]. Co-axial stacking of the group I intron P1 and P2 helices were also previously inferred from helix-length variations [4] and later confirmed from crystal structures [5,6]. A more elaborate investigation of tertiary structure compensation has been previously reported as part of a comparative sequence analysis of mitochondrial tRNA sequences that do not fit the standard cloverleaf consensus [7]. 3D modeling helped describe several types of structural compensations, including helix-length compensations, that can help maintain the global L-shape in these atypical tRNAs [7].

In this work, we extend the concept of helix-length compensation for structure-function studies of RNA using the *Neurospora* Varkud satellite [31] ribozyme as a model system (for

reviews, see [8-10]; Figure 2.1A). Like other nucleolytic ribozymes, the VS ribozyme adopts a simple RNA architecture that allows phosphodiester cleavage activity. Notably, the VS ribozyme has the unique ability to specifically recognize its stem-loop I [145] substrate through formation of a kissing-loop interaction with stem-loop V (SLV; [11-13]). This I/V kissing-loop interaction of the VS ribozyme is ideally suited for performing helix-length compensation studies since mutant substrates and ribozymes can be synthesized separately and then mixed in various combinations for kinetic characterization (Figure 2.1B). Furthermore, in the 3D models of the VS ribozyme, helices I and V are extended away from each other in a similar manner as coaxially stacked helices [14-18]. It is expected that adding (or removing) base pairs within stem Ib would disrupt productive docking of the SLI cleavage loop with the A730 loop of stem-loop VI to form the proposed active site [10,19-27] and significantly reduce cleavage activity, as previously observed for stem V length variants [16,28]. We hypothesize that cleavage activity of stem Ib length variants could be restored by adjusting the number of base pairs within stem V of the ribozyme, such that the total number of base pairs within stems Ib and V is preserved (Figure 2.1B). To explore this possibility, we combined kinetic characterization of variants and 3D modeling. Our helix-compensation study with the VS ribozyme provides global structural information and reveals the adaptability of RNA architecture.

MATERIALS AND METHODS

Plasmids

The pAvapl plasmid used for transcription of Avapl (Figure 2.1A) was constructed by PCR from the A-3 clone [18,29]. The plasmids for transcription of Avapl ribozyme variants (Figure 2.1B) were prepared by QuikChangeII site-directed mutagenesis from pAvapl (Stratagene, CA). All plasmids were validated by DNA sequencing.

Preparation of RNA for kinetic studies

The R₀ ribozyme and variant sequences (Figure 2.1) were prepared by *in vitro* T7 RNA polymerase transcription from the pAvapl plasmid (or derivatives) linearized at the *Ava*I site. All ribozymes were purified as previously described [18]. The S₀ substrate and variant sequences (Figure 2.1) were synthesized *in vitro* using T7 RNA polymerase and a single-stranded synthetic oligonucleotide template (Integrated DNA Technologies, IA). They were purified by denaturing gel electrophoresis, dephosphorylated with calf alkaline phosphatase and 5'-end-labeled with γ -(³²P) ATP (PerkinElmer, Ont.) using established procedures [18].

Single-substrate kinetic studies

All reactions were carried out at 30°C in reaction buffer (50 mM Tris pH 7.5, 25 mM KCl, 2 mM spermidine and 25 mM MgCl₂). Prior to starting the cleavage reaction, the 5'-³²P-labeled substrate was heated to 95°C for 1 min in storage buffer (10 mM Tris pH 7.6 and 1 mM EDTA) and snap-cooled to promote hairpin formation. Cleavage reactions were initiated by addition of equal volume (20 μ L) of the 5'-³²P-labeled S and R, either the wild-type R (R₀) or mutants (Figure 2.1B), after a 5-min pre-incubation of S and R at 30°C in reaction buffer. Typically, about 10 aliquots of 3 μ L were removed from the 40 μ L-reaction mixture at specific times, and the reaction was stopped by addition of 12 μ L of stop mix (50 mM EDTA pH 8.0, 0.04% xylene cyanol, 0.04% bromophenol blue and \geq 85% formamide) and transfer to 4°C. Samples were analyzed by denaturing gel electrophoresis [20% acrylamide:bisacrylamide (19:1) / 7 M urea gels]. For each time point, the amounts of substrate and product were

quantified using a Molecular Imager FX and the Quantity One 1D analysis software version 4.6.2 (Bio-Rad Laboratories, Ont.).

Single-turnover kinetic experiments were typically performed with $[S] = 250$ pM and excess R (0.005 to 5 μ M). The percentage of remaining substrate $[F = S_t/(S_t+P_t)*100\%]$; where S_t and P_t are the amount of substrate and product at time t] was plotted against time. The resulting time courses were fitted to the equation $F = Ae^{-(k_{obs} *t)} + F_{\infty}$, where A is the amplitude, k_{obs} is the rate of the reaction and F_{∞} is a variable that represents the amount of substrate at the end point of the cleavage reaction. The value of k_{obs} , the standard error for k_{obs} , and the F_{∞} parameter were estimated by non-linear regression analysis within the OriginPro 8.1 software (OriginLab, MA). The F_{∞} values are below 10% for all S/R combinations.

The dependence of k_{obs} on $[R]$ was analyzed by linear regression to derive the second order rate constant of the reaction, k_{cat}/K_M . The linear dependence of k_{obs} on $[R]$ was observed from 0.005 to 1 μ M for S_0/R_0 and from 0.005 to 3 μ M for both S_{+1bp}/R_0 and S_{-1bp}/R_0 . The quality of the fit for the linear dependence of the k_{obs} on $[R]$ was obtained from the square of the correlation coefficient (R^2), and in all cases R^2 was ≥ 0.99 . Single-substrate kinetic experiments were all performed from at least three independent experiments, and reported k_{cat}/K_M and their errors are average values and standard deviations, respectively, from these multiple measurements.

Multiple-substrate kinetic studies

Multiple-substrate kinetic studies were carried out as single-turnover reactions under the same conditions as the single-substrate kinetic studies. The main difference is that the cleavage reactions were initiated by addition of equal volume (20 μ L) of R (0.005 - 5 μ M of either R_0 or a mutant) and a mixture of 5'-³²P-labeled nucleic acids containing all substrates (S_{-2bp} , S_{-1bp} , S_0 , S_{+1bp} , S_{+2bp} , S_{+3bp} and S_{+4bp} ; 35 pM of each) and, in certain cases, a non-cleavable DNA hairpin control (5'- CGG CGA AGC ATC GTC GGC TTC GCC G -3'; 35 pM). In addition, the percentage of remaining substrate was expressed as $F = (S_t/S_0)*(C_0/C_t)*100\%$ and plotted against time, where S_0 and C_0 are respectively the amount of substrate and non-cleavable S_{-2bp}

control at time $t = 0$ and S_t and C_t represent the corresponding values at time $t \neq 0$. Values for the reaction rate k_{obs} , amount of uncleaved substrate F_∞ and specificity constant k_{cat}/K_M were derived as for the single-substrate kinetic studies. All multiple-substrate kinetic experiments were performed at least three times, and reported k_{cat}/K_M and their errors are average values and standard deviations, respectively, from these multiple measurements.

Localization of the substrate cleavage site

The site of cleavage was determined for all S/R combinations where cleavage was observed. For this purpose, the substrates were subjected to cleavage in single-substrate kinetic conditions chosen to yield approximately ~50% cleavage. Cleavage reactions were analyzed by denaturing gel electrophoresis [20% acrylamide:bisacrylamide (19:1) / 7 M urea gels] and by comparison with partial T1 nuclease digest and alkaline hydrolysis of S_0 .

Native gel studies of RNA substrates

Non-denaturing polyacrylamide gel electrophoresis was carried out on a 7.5% polyacrylamide bisacrylamide (37.5:1) gel in TB buffer (50 mM Tris borate pH 8.0 and 5 mM magnesium acetate). Prior to loading, the RNA substrates were heated at 95°C for 2 min in storage buffer and transferred on ice for 5 min. The gel was pre-run for 30 min at 200 V, run for 2 hrs at 250 V and 4°C and then stained with Stains All (Sigma-Aldrich, Ont.).

Structural modeling of SLI/SLV complexes

3D models of the SLI/SLV complex were generated using a stepwise procedure (Supplementary Figure 2.3). As a first step, RNA fragments were modeled using the constraint-satisfaction software MC-Sym 3.3.2 [30]. Several models of the I/V kissing-loop junction core (residues 626-634 and 695-701) were first obtained separately using the NMR structure of magnesium-bound SLV [31] and constraints derived from biochemical studies, as previously described [18]. A unique structural model was obtained for each type of SLI stem (S_{-1bp} to S_{+4bp} ; residues 619-625 and 635-640 for S_0), in accordance with the secondary structure model (Figure 2.1) and the NMR structure of the active conformation of the cleavage site internal loop [32]. A unique structural model was also obtained for each type of SLV stems (R_{-4bp} to R_{+1bp} ; residues 689-694 and 702-707 for R_0), in accordance with the secondary

structure model (Figure 2.1). Before being grafted together, the core and stem models were energy minimized using the molecular dynamics software Sander [33] based on the Amber force field and including covalent, electrostatic, Van der Waals energies, as well as implicit solvation in the form of a Generalized Born model [34]. Several rounds of minimization (100 000) were first applied including restraints to maintain Watson-Crick base pairing and followed by fewer rounds of minimization (1000) without these restraints. The minimized core and stem models were then filtered using a custom C++ program based on the freely accessible C++ library MC-Core [30]. The filter-selected models that preserved global structural integrity (heavy-atom rmsd between the pre- and post-minimization model less than 2.0 Å) and key structural features (covalent integrity of backbone, Watson-Crick base pairs at the I/V kissing-loop junction, and hydrogen bonds typical of U-turn structure in SLI between U628 2'-O and G630 N7 and between U628 N3 and U631). Each remaining core model was then assembled using MC-Sym with all possible combinations of stem I (S_{-1bp} to S_{+4bp}) and stem V (R_{-4bp} to R_{+1bp}), energy minimized and filtered to create 36 sub-models that form a common-core group.

For all selected SLI/SLV sub-models, helical axes were computed using Curves+ for stem Ib and V [35]. The interhelical angles (θ_{I-V}) and axial displacement (d_{I-V}) between these two axes were computed by vector methods within Pymol using Python scripts. For calculation of d_{I-V} , a reference point O was first set midway between atoms corresponding to A698 N3 and U631 N1 of the kissing-loop interaction. Two segments OA and OB were calculated that start at this reference point to reach the two helical axes at point A and B in a perpendicular orientation. The distance of the AB segment was taken to represent the axial displacement. The reported values of θ_{I-V} and d_{I-V} are average values of all sub-models from a common-core group.

Selection of SLI/SLV models compatible with kinetic data

In order to select common-core groups that are compatible with the kinetic data, the sub-models of each common-core group were first compared to their reference S_0/R_0 sub-

model using an rmsd criteria. The sub-models were first superimposed to the S_0/R_0 sub-model based on alignment of the first base pair of SLV (residues 689 and 707 in S_0/R_0). Heavy atoms pairwise rmsds between the cleavage site region of SLI of the S_0/R_0 sub-model (residues 619-623 and 637-640) and that of each sub-models were then computed. For each common-core group, the rmsd distributions for the high-activity [$(k_{cat}/K_M)/(k_{cat}/K_M) \geq 0.25$] and low-activity [$(k_{cat}/K_M)/(k_{cat}/K_M) \leq 0.01$] classes of S/R pairs were compared using a Welch's t test [36]. The normality assumption was verified using the Shapiro-Wilks test [37] on each distribution, and no significant violation was found ($\alpha = 0.05$). Common-core groups compatible with kinetic data were selected when statistically significant differences (at $\alpha = 0.02$) were observed between the mean rmsd values of high-activity and low-activity classes of S/R pairs.

RESULTS

For helix-length compensation studies, we performed kinetic studies of trans cleavage with the Avapl VS ribozyme and an SLI substrate, herein termed R_0 and S_0 , respectively (Figure 2.1A). The S_0 substrate adopts a pre-shifted active state (Figure 2.1A) that is rapidly cleaved by the VS ribozyme [18]. Helix-length mutations were introduced within stem Ib of S_0 (Figure 2.1B) that remove up to 2 bp (S_{-1bp} and S_{-2bp}) or add up to 4 bp (S_{+1bp} , S_{+2bp} , S_{+3bp} and S_{+4bp}). Helix-length mutations were also introduced within stem V of R_0 (Figure 2.1B) that add up to 2 bp (R_{+1bp} and R_{+2bp}) or remove up to 4 bp (R_{-1bp} , R_{-2bp} , R_{-3bp} and R_{-4bp}). Compensatory substrate/ribozyme (S/R) variant pairs are those for which the total number of base pairs in stems Ib and V is identical to the S_0/R_0 reference pair (Figure 2.1B). We hypothesize that, in the simplest scenario, compensatory S/R variant pairs would retain cleavage activity, whereas other S/R variant pairs would not.

Single-substrate kinetic studies

We first examined the specificity of R_0 for each substrate individually by determining the second-order rate constant (k_{cat}/K_M) under single-turnover conditions [18], as illustrated for S_0 in Figure 2.2. Cleavage reactions were performed with ^{32}P -labeled S_0 and excess R_0 and monitored by denaturing gel electrophoresis (Figure 2.2A). The percentage of remaining substrate was plotted against time to derive the value of the first-order rate constant k_{obs} (Figure 2.2B). For S_{-1bp} , S_0 and S_{+1bp} a linear relationship is observed when k_{obs} is plotted as a function of R_0 concentration, and the values of catalytic efficiency (k_{cat}/K_M) were obtained from the slope of this graph (Figure 2.2C and Table II.I). For all other substrates (S_{-2bp} , S_{+2bp} , S_{+3bp} and S_{+4bp}), the value of k_{obs} is below the detection limit ($\leq 0.001 \text{ min}^{-1}$) at R_0 concentrations up to $5 \mu\text{M}$, and therefore the value of k_{cat}/K_M can not be determined (Table II.I). To insure that the lack of reactivity of certain substrates is not due to folding defects, they were analyzed by native gel electrophoresis (Supplementary Figure 2.1). All substrates, with exception of S_{-2bp} , migrate as a single band on the gel, at a position typical of small RNA hairpins, in agreement with the predicted secondary structure [38]. Most of the S_{-2bp} substrate migrates slower, indicating that it adopts a duplex conformation that is incompatible with cleavage by R_0 .

For R_0 , the highest value of k_{cat}/K_M is obtained with S_0 ($11 \pm 1 \text{ min}^{-1}\mu\text{M}^{-1}$). Surprisingly, the k_{cat}/K_M values for S_{-1} ($5.2 \pm 0.8 \text{ min}^{-1}\mu\text{M}^{-1}$) and S_{+1} ($2.7 \pm 0.3 \text{ min}^{-1}\mu\text{M}^{-1}$) are reduced by only 2- and 4-fold, respectively, with respect to S_0 (Table II.I). The similar substrate specificities of R_0 for S_{-1} , S_0 and S_{+1} point to the adaptability of the VS ribozyme architecture to cleave a range of substrates with 3-5 bp in stem Ib, indicating a certain level of promiscuity in substrate selection.

Multiple-substrate kinetic studies

Single-substrate kinetic studies performed with R_0 are fairly tedious and time-consuming, especially given that several time-courses are needed to derive k_{cat}/K_M values for an individual substrate, each k_{cat}/K_M is determined at least three times, and multiple substrates are being investigated. To accelerate the kinetic characterization of the remaining S/R variant pairs (Figure 2.1B), we set up a multiple-substrate kinetic procedure in which one enzyme is simultaneously incubated with a mixture of substrates [29]. Single-turnover conditions were chosen for our studies because the enzyme is in excess with respect to all substrates, and there should be no competition between the various substrates. Given the different sizes of substrates investigated (Figure 2.1B), the disappearance of any given substrate is easily monitored by denaturing gel electrophoresis (Figure 2.3). The percentage of remaining substrate can be reliably estimated from the ratio of substrate remaining at time t to that at time zero provided that normalization of input RNA in each lane can be achieved. For normalization, we initially added a non-cleavable DNA control in the cleavage reaction, but later found that quantification is more reliable using the non-cleavable S_{-2bp} RNA that gives a better-defined band on the gel (Figure 2.3).

The multiple-substrate kinetic procedure was first performed with R_0 , revealing essentially identical k_{cat}/K_M values to those obtained with the single-substrate kinetic method (Table II.I). These results confirm the reliability of the multiple-substrate procedure and, thus, similar studies were conducted with the variant ribozymes, except for R_{+2bp} because its helix-length complementary substrate S_{-2bp} is not cleavable. Interestingly, substrate cleavage is observed for several S/R pairs, with many having similar k_{cat}/K_M values as the S_0/R_0 pair

(Table II.II). To verify that substrate cleavage for all functional S/R pairs yields the same product, the reaction products were analyzed by comparing their migration with alkaline and T1 digests on denaturing gels. The results clearly demonstrate that cleavage occurs between G620 and A621 for all substrates cleaved by either R_0 or variant ribozymes (Supplementary Figure 2.2). Thus, the helix-length mutations investigated here have no effect on cleavage site selection.

As expected, helix-length compensation is observed for all variant ribozymes, since the highest values of k_{cat}/K_M are observed with the compensatory substrate or one that differs by a single base pair. More precisely, the most active S/R combinations for each ribozyme are: S_{-1bp}/R_{+1bp} , S_0/R_0 , S_{+1bp}/R_{-1bp} , S_{+1bp}/R_{-2bp} , S_{+2bp}/R_{-3bp} and S_{+3bp}/R_{-4bp} . However, k_{cat}/K_M values similar to that of S_0/R_0 (within 4-fold; shaded in Table II.II) are observed only for four of the six ribozymes (R_0 , R_{-1bp} , R_{-2bp} and R_{-3bp}). For the R_{+1bp} and R_{-4bp} ribozymes, the maximal k_{cat}/K_M values are reduced significantly, ~ 80 and ~ 30 fold, respectively, compared to S_0/R_0 .

Structural modeling of SLI/SLV complexes compatible with kinetic data

Catalysis generally requires the precise positioning of chemically active groups within the active site, which in the VS ribozyme results from the precise docking of the cleavage site and A730 internal loops [10,19-27]. Thus, it is likely that the cleavage site loop is presented in a favorable position for productive docking in the most active S/R combinations, but not in the inactive or poorly active S/R combinations. To explore how the relative positioning of the cleavage site loop is affected by helix-length mutations, a structural modeling approach was used. We chose to model an SLI/SLV complex as opposed to models of complete S/R pairs to simplify the modeling procedure and limit the structural consequences of helix-length mutations. Furthermore, the SLI/SLV models allow us to examine the effect of helix-length variations on the positioning of the scissile phosphate in SLI relative to a common base pair in SLV. This approach is consistent with our kinetic studies of S/R pairs in which ribozyme mutants retain a common core sequence and the fact that free substrates interact primarily through the formation of the I/V kissing-loop interaction (Bouchard, P., Delfosse, V. and Legault, P., unpublished data). We assumed that realistic 3D models would be compatible

with the kinetic data, such that the relative positioning of the cleavage site loop in active S/R combinations would be distinct from that found with inactive S/R combinations.

3D models of the SLI/SLV complexes were generated using a stepwise procedure (Supplementary Figure 2.3). First, several models of the I/V kissing-loop junction core were obtained using previously described constraints [18]. Of these core models, 56 were then selected because each could be used to build a common-core group of 36 sub-models that contain all combinations of SLI (S_{-1bp} to S_{+4bp}) and SLV stems (R_{-4bp} to R_{+1bp}) and remain consistent with input structural constraints after energy minimization. For all selected sub-models (56 x 36), the interhelical angles between stems Ib and V were computed and average values were determined for each common-core group. A wide range of axial displacements (d_{I-V} = 12-24 Å), and interhelical angles (θ_{I-V} = 108° to 175°; Figure 2.4) are observed among the common-core groups, pointing to the extensive conformational sampling of the modeling procedure.

We hypothesized that for common-core groups compatible with the kinetic data, the sub-models of the high-activity class [$(k_{cat}/K_M)/(k_{cat}/K_M)_0 \geq 0.25$]; Table II.II] should be significantly different than those of the low-activity class [$(k_{cat}/K_M)/(k_{cat}/K_M)_0 \leq 0.01$]; Table II.II] of S/R pairs. In order to compare the structures of the sub-models, pairwise rmsd's were calculated between the sub-models of each common-core group and their reference S_0/R_0 sub-model. For each common-core group, the rmsd distributions for the high-activity and low-activity classes of S/R pairs were compared using a Welch's *t*-test [36]. Interestingly, we observe a good correlation between *P*-values and interhelical angles (Figure 2.4), where common-core groups with average interhelical angles (θ_{I-V}) closer to 180° show smaller *P*-values and are thus more likely to be consistent with the kinetic data. Group 13 ($\theta_{I-V} = 166.0^\circ \pm 3.6$) was selected for structural analysis, but similar conclusions are obtained for other common-core groups with *P*-values smaller than the α threshold value of 0.02 (group 12: $\theta_{I-V} = 165.4^\circ \pm 3.5^\circ$ and group 20: $\theta_{I-V} = 171.4^\circ \pm 4.5^\circ$).

The 3D structure of the S_0/R_0 sub-model from the selected group reveals that, although the helical axes of stems Ib and V are more or less co-linear, the two stems are not co-axial (Figure 2.5A). Rather, there is a significant perpendicular displacement ($d_{I,V} \sim 20 \text{ \AA}$) between these axes that results from the structure of the kissing-loop junction. The spatial relation between these two axes is reflected in the structure of variant S/R pairs, as illustrated in Figure 2.5B-D. The superposition of all S/R pairs containing R_0 shows that the scissile phosphates of the high-activity substrates (deep orange spheres in Figure 2.5B) cover a $\sim 7.5 \text{ \AA}$ distance on one side of stem Ib, whereas those of low-activity substrates (white spheres in Figure 2.5B) rotate around the helical axis towards a different side of the helix. Thus, this superposition indicates that limited helical displacements of the scissile phosphate are compatible with productive docking of the substrate internal loop with the catalytic domain.

To better understand the helix-length compensation observed between selected S/R pairs, we superposed the S/R pairs that produce maximal cleavage activity for each ribozyme (Figure 2.5C and D). It is clear from both the top-down (Figure 2.5C) and front views (Figure 2.5D) of this superposition that the scissile phosphates are distributed along the arc of a circle. Those of the high-activity S/R pairs (deep orange in Figure 2.5C and D) span a chord length of $\sim 22 \text{ \AA}$. This superposition suggests that large, but restricted displacements of the scissile phosphate along this arc are compatible with productive docking of the substrate internal loop with the catalytic domain.

DISCUSSION

In this study, we explored an unusual type of compensatory mutations in RNA, termed helix-length compensations, using the VS ribozyme as a model system. VS substrate and ribozyme variants were prepared in which the number of base pairs in stems Ib and V were either reduced or increased compared to the reference S_0/R_0 pair. Several S/R combinations were found to be nearly as active as the reference pair, with precise and efficient cleavage obtained for substrate derivatives that contain 3-6 bp in stem Ib. Thus, this work expands the substrate specificity of the VS ribozyme architecture, while revealing its adaptability for efficient substrate cleavage. Two interesting aspects of this adaptability are limited substrate promiscuity with respect to stem Ib length and helix-length compensations between stems Ib and V.

Substrate promiscuity in the VS ribozyme

Substrate cleavage by R_0 occurs most efficiently with a stem Ib containing 3-5 potential base pairs, whereas the catalytic efficiency for substrates with 6-8 bp in stem Ib is reduced at least 50 000-fold compared to S_0 . Thus, there is limited substrate promiscuity with respect to stem Ib length for cleavage by R_0 , and comparable results are obtained with R_{-1} and R_{-2} . A similar phenomenon has previously been observed for the 5'-splicing reaction of the group I intron ribozyme, where the cleavage site coincides with a G-U wobble pair in the first stem-loop (P1) of the intron [39]. Cleavage at the correct site was observed when the G-U wobble pair was inserted within a three-base pair window of the P1 stem [39]. Although the cleavage mechanisms of the group I intron and VS ribozymes differ significantly, both require productive docking of a hairpin substrate with the catalytic domain. Thus, for both ribozymes, helical displacement of the scissile phosphate within a window equivalent to 3 bp is compatible with productive docking and cleavage at the correct site.

Helix-length compensation between stems Ib and V

We initially hypothesized that the reduced cleavage activity of defective stem Ib variants by R_0 could be restored by adjusting the number of base pairs within stem V, such that the total number of base pairs within the two stems would be preserved. Examination of the

kinetic data indicates that there is, as predicted, helix-length compensation between stems Ib and V. Indeed, substrates for which cleavage is not detected with R_0 (S_{+2bp} , S_{+3bp} , S_{+4bp}) are cleaved by a ribozyme variant. However, the cleavage activities of the compensatory S/R pairs are not all restored to the reference level. In agreement with these results, 3D modeling of the I/V kissing-loop complex indicates that the global orientation of stem Ib and V significantly affects the position of the scissile phosphate in the variant S/R pairs.

An unexpected aspect of this helix-length compensation study is that it helps provide global structural information on the I/V kissing-loop interaction. From an ensemble of SLI/SLV models that display a wide range of interhelical angles ($108^\circ \leq \theta_{I,V} \leq 175^\circ$) and axial displacements ($d_{I,V} = 12\text{-}24 \text{ \AA}$), three models were selected that best fit the kinetic data. These models display a much smaller range of interhelical angles and axial displacements ($\theta_{I,V} = 165^\circ\text{-}171^\circ$ and $d_{I,V} = 18\text{-}21 \text{ \AA}$), with near colinearity of their helical axes. By comparison, our previously reported model of the SLI/SLV complex [18] exhibits interhelical parameters ($\theta_{I,V} = 106^\circ$ and $d_{I,V} = 16 \text{ \AA}$) that are somewhat compatible with the ensemble of models, but not with our selected models. In contrast, the approximate colinearity of stem Ib and V is consistent with a 3D model of the VS ribozyme derived from biochemical and mutational data [14,28], however sufficient constraints were not available to precisely define the interhelical angle these two stems nor to rule out alternative models (Collins, R.A., personal communication). Similarly, the global model of the VS ribozyme based on fluorescence energy transfer data does not incorporate sufficient constraints, particularly for the SLI/SLV interaction, to allow a fair comparison with our selected models [15,16]. Interestingly, the model of the complete VS ribozyme derived from small-angle X-ray scattering (SAXS) show an interhelical angle close to that of our selected models [$\theta_{I,V} = 154^\circ$ and $d_{I,V} = 25 \text{ \AA}$; [17]]. This similarity between our selected models and the SAXS model is particularly surprising given that the folding of the complete ribozyme was not accounted for in our modeling studies. In addition, the SAXS model includes helix VII that is not present in the VS used for our kinetic study. The small difference in axial displacement between our selected models and the SAXS model likely reflects structural differences at the kissing-loop junction. Nevertheless, this comparison with the SAXS model suggests that the combination of kinetic

data from helix-length variants with 3D modeling represents a potentially powerful approach to extract interhelical parameters.

Adaptability of the VS ribozyme architecture for substrate cleavage

Given the intrinsic adaptability of the VS ribozymes architecture for substrate cleavage, the modeling results give an appreciation of the level of structural adaptation that is likely at stake for productive docking of the SLI cleavage site internal loop with the A730 loop of SLVI. However, these models do not explain how high-activity S/R pairs move relative to the wild-type S/R pair to allow formation of the active site and efficient cleavage.

An emerging view in RNA structure formation is that tertiary interactions form as a result of conformational sampling of helical domains within a distribution of conformations that is limited, in part, by the inherent flexibility of helical connectivities (for recent reviews see [40,41]). It is likely that in the VS ribozyme, flexible elements such as its bulges, internal loops, three-way junctions, and even its I/V kissing-loop junction [42,43] allow conformational sampling of helical domains, including SLI and SLVI, through which a well-defined active site can emerge. According to this view, the ability to sample the active conformation may be conserved in high-activity S/R pairs but not in those with lower activity. Future investigations are needed to help identify the dynamic features responsible for the conformational adaptability of the VS ribozyme and determine if a similar adaptability exists in other functional RNAs.

Prospects for RNA research

Helix-length compensation represents an alternative approach for investigating the sequence and structural repertoire of functional RNAs that could be applied to other helix pairs. In principle, helix-length compensation could be observed for helices that are more or less co-axially stacked [1-6] or those that are not co-axial but nearly colinear as observed for the VS ribozyme. Although helix-length compensations have been previously identified by comparative sequence analysis [1-6], such analysis may not identify all possible helix-length mutants. Indeed, the genetic mechanism through which such complex mutation pattern evolves likely requires multiple mutational events with an overall lower probability than

simple co-variation of base-pairing residues. Moreover, sequence searches may not always take into account the possibility for helix-length compensation. Systematic helix-length compensation studies could help improve the consensus sequence and structural model for a given functional RNA. Subsequently, revised RNA consensus sequence could be used for bioinformatics database search of additional representatives of known functional RNAs [44-47]. Helix-length compensation studies could also help reveal the inherent adaptability of other complex RNA architectures and be useful for the design of RNAs with novel functions.

ACKNOWLEDGMENTS

The authors thank D.M.J. Lilley for providing the atomic coordinates for the SAXS model of the complete VS ribozyme, J. G. Omichinski for critical reading of the manuscript and R. Richter for computer support.

FUNDING

The Canadian Institutes of Health Research (CIHR MOP-86502 to P.L.); CIHR and the Université de Montréal Ph.D. scholarships (to J.L.-L.); Canada Research Chair in Structural Biology and Engineering of RNA (to P.L.). Funding for open access charge: Canadian Institutes of Health Research.

Conflict of interest statement. None declared.

REFERENCES

1. Woese, C.R., et al., *Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids*. Microbiol. Rev., 1983. **47**(4): p. 621-69.
2. Gutell, R.R., *Comparative sequence analysis and the structure of 16S and 23S rRNA*, in *Ribosomal RNA: Structure, evolution, processing, and function in protein biosynthesis*, R.A. Zimmermann and A.E. Dahlberg, Editors, CRC Press: Boca Raton, Florida, 1995. p. 109-126.
3. Ban, N., et al., *The complete structure of the large ribosomal subunit at 2.4 Å resolution*. Science, 2000. **289**: p. 905-920.
4. Michel, F. and E. Westhof, *Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis*. J. Mol. Biol., 1990. **216**(3): p. 585-610.
5. Adams, P.L., et al., *Crystal structure of a self-splicing group I intron with both exons*. Nature, 2004. **430**: p. 45-50.
6. Golden, B.L., H. Kim, and E. Chase, *Crystal structure of a phage Twort group I ribozyme-product complex*. Nature Struct. & Mol. Biol., 2004. **12**(1): p. 82-88.
7. Steinberg, S., F. Leclerc, and R. Cedergren, *Structural rules and conformational compensations in the tRNA L-form*. J. Mol. Biol., 1997. **266**(2): p. 269-82.
8. Collins, R.A., *The Neurospora Varkud satellite ribozyme*. Bioch. Soc. Trans. Rev., 2002. **30**: p. 1122-1126.
9. Lilley, D.M., *The Varkud satellite ribozyme*. RNA, 2004. **10**: p. 151-158.
10. Wilson, T.J. and D.M. Lilley, *Do the hairpin and VS ribozymes share a common catalytic mechanism based on general acid-base catalysis? A critical assessment of available experimental data*. RNA, 2011. **17**(2): p. 213-21.
11. Dotu, I., et al., *Complete RNA inverse folding: computational design of functional hammerhead ribozymes*. Nucleic Acids Res, 2014. **42**(18): p. 11752-62.
12. Rastogi, T., et al., *A long-range pseudoknot is required for activity of the Neurospora VS ribozyme*. EMBO J., 1996. **15**(11): p. 2820-2825.
13. Andersen, A. and R.A. Collins, *Rearrangement of a stable RNA secondary structure during VS ribozyme catalysis*. Mol. Cell, 2000. **5**: p. 469-478.
14. Andersen, A.A. and R.A. Collins, *Intramolecular secondary structure rearrangement by the kissing interaction of the Neurospora VS ribozyme*. Proc. Natl. Acad. Sci USA, 2001. **98**(14): p. 7730-7735.
15. Hiley, S.L. and R.A. Collins, *Rapid formation of a solvent-inaccessible core in the Neurospora Varkud satellite ribozyme*. EMBO J., 2001. **20**(19): p. 5461-5469.
16. Lafontaine, D.A., D.G. Norman, and D.M. Lilley, *Structure, folding and activity of the VS ribozyme: importance of the 2-3-6 helical junction*. EMBO J., 2001. **20**(6): p. 1415-1424.
17. Lafontaine, D.A., D.G. Norman, and D.M. Lilley, *The global structure of the VS ribozyme*. EMBO J., 2002. **21**(10): p. 2461-2471.
18. Lipfert, J., et al., *The complete VS ribozyme in solution studied by small-angle X-ray scattering*. Structure, 2008. **16**(9): p. 1357-67.
19. Bouchard, P., et al., *Role of SLV in SLI substrate recognition by the Neurospora VS ribozyme*. RNA, 2008. **14**: p. 736-748.

20. Lafontaine, D.A., et al., *The A730 loop is an important component of the active site of the VS ribozyme*. J. Mol. Biol., 2001. **312**: p. 663-674.
21. Lafontaine, D.A., et al., *Functional group requirements in the probable active site of the VS ribozyme*. J. Mol. Biol., 2002. **323**: p. 23-34.
22. Sood, V.D. and R.A. Collins, *Identification of the catalytic subdomain of the VS ribozyme and evidence for remarkable sequence tolerance in the active site loop*. J. Mol. Biol., 2002. **320**: p. 443-454.
23. Jones, F.D. and S.A. Strobel, *Ionization of a critical adenosine residue in the Neurospora Varkud satellite ribozyme active site*. Biochemistry, 2003. **42**: p. 4265-4276.
24. Zhao, Z.Y., et al., *Nucleobase participation in ribozyme catalysis*. J. Am. Chem. Soc., 2005. **127**(14): p. 5026-7.
25. Wilson, T.J., A.C. McLeod, and D.M. Lilley, *A guanine nucleobase important for catalysis by the VS ribozyme*. EMBO J., 2007. **26**(10): p. 2489-500.
26. Smith, M.D. and R.A. Collins, *Evidence for proton transfer in the rate-limiting step of a fast-cleaving Varkud satellite ribozyme*. Proc. Natl. Acad. Sci. USA, 2007. **104**(14): p. 5818-23.
27. Jaikaran, D., et al., *An important role of G638 in the cis-cleavage reaction of the Neurospora VS ribozyme revealed by a novel nucleotide analog incorporation method*. RNA, 2008. **14**(5): p. 938-49.
28. Wilson, T.J., et al., *Nucleobase-mediated general acid-base catalysis in the Varkud satellite ribozyme*. Proc. Natl. Acad. Sci. USA, 2010. **107**(26): p. 11751-6.
29. Rastogi, T. and R.A. Collins, *Smaller, faster ribozymes reveal the catalytic core of Neurospora VS RNA*. J. Mol. Biol., 1998. **277**: p. 215-224.
30. Guo, H.C.T. and R.A. Collins, *Efficient trans-cleavage of a stem-loop RNA substrate by a ribozyme derived from Neurospora VS RNA*. EMBO J., 1995. **14**(12): p. 368-376.
31. Major, F., et al., *The combination of symbolic and numerical computation for three-dimensional modeling of RNA*. Science, 1991. **253**(5025): p. 1255-60.
32. Campbell, D.O., et al., *NMR structure of varkud satellite ribozyme stem-loop V in the presence of magnesium ions and localization of metal-binding sites*. Biochemistry, 2006. **45**(35): p. 10591-605.
33. Hoffmann, B., et al., *NMR structure of the active conformation of the Varkud satellite ribozyme cleavage site*. Proc. Natl. Acad. Sci. USA, 2003. **100**(12): p. 7003-7008.
34. Pearlman, D.A., et al., *Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules*. Comput. Phys. Commun., 1995. **91**(1-3): p. 1-41.
35. Still, W.C., et al., *Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics*. J Am Chem Soc, 1990. **112**(16): p. 6127-6129.
36. Lavery, R., et al., *Conformational analysis of nucleic acids revisited: Curves+*. Nucleic Acids Res., 2009. **37**(17): p. 5917-29.
37. Satterthwaite, F.E., *An approximate distribution of estimates of variance components*. Biometrics, 1946. **2**(6): p. 110-4.
38. Shapiro, S.S. and M.B. Wilk, *An analysis of variance test for normality (complete samples)*. Biometrika, 1965. **52**(3/4): p. 591-611.

39. Beattie, T.L., J.E. Olive, and R.A. Collins, *A secondary-structure model for the self-cleaving region of Neurospora VS RNA*. Proc. Natl. Acad. Sci. USA, 1995. **92**: p. 4686-4690.
40. Doudna, J.A., B.P. Cormack, and J.W. Szostak, *RNA structure, not sequence, determines the 5' splice-site specificity of a group I intron*. Proc. Natl. Acad. Sci. USA, 1989. **86**(19): p. 7402-6.
41. Woodson, S.A., *Compact intermediates in RNA folding*. Annu. Rev. Biophys., 2010. **39**: p. 61-77.
42. Bailor, M.H., et al., *Topological constraints: using RNA secondary structure to model 3D conformation, folding pathways, and dynamic adaptation*. Curr Opin Struct Biol, 2011. **21** (3): p. 296-305.
43. Szewczak, A.A. and T.R. Cech, *An RNA internal loop acts as a hinge to facilitate ribozyme folding and catalysis*. RNA, 1997. **3**(8): p. 838-49.
44. Bailor, M.H., X. Sun, and H.M. Al-Hashimi, *Topology links RNA secondary structure with global conformation, dynamics, and adaptation*. Science, 2010. **327**(5962): p. 202-6.
45. Ferbeyre, G., J.M. Smith, and R. Cedergren, *Schistosome satellite DNA encodes active hammerhead ribozymes*. Mol. Cell. Biol., 1998. **18**(7): p. 3880-8.
46. Gräf, S., et al., *A database search for hammerhead ribozyme motifs*. Biochem. Soc. Trans., 2005. **33**: p. 477-478.
47. Hammann, C. and E. Westhof, *Searching genomes for ribozymes and riboswitches*. Genome Biol., 2007. **8**(4): p. 210.
48. McCown, P.J., A. Roth, and R.R. Breaker, *An expanded collection and refined consensus model of glmS ribozymes*. RNA, 2011. **17**(4): p. 728-36.

TABLES

Table II.I Summary of kinetic data for R_0 ^a

	Values of k_{cat}/K_M ($\text{min}^{-1}\mu\text{M}^{-1}$)	
	Single-substrate kinetics	Multiple-substrate kinetics
S_{-2bp}	n.d. ^a	n.d.
S_{-1bp}	5.2 ± 0.8	5.9 ± 0.7
S_0	11 ± 1	11 ± 2
S_{+1bp}	2.7 ± 0.3	2.7 ± 0.4
S_{+2bp}	n.d.	n.d.
S_{+3bp}	n.d.	n.d.
S_{+4bp}	n.d.	n.d.

^aThe k_{cat}/K_M and their errors (in $\text{min}^{-1}\mu\text{M}^{-1}$) are average values and standard deviations, respectively, from at least three independent experiments. For several S/R pairs cleavage activity was not detected (n.d.; $k_{obs} \leq 0.001 \text{ min}^{-1}$) at $[R] \leq 5 \mu\text{M}$.

Table II.II Summary of multiple-substrate kinetic data^a

	R-4bp		R-3bp		R-2bp		R-1bp		R ₀		R+1bp	
	k_{cat}/K_M	ratio	k_{cat}/K_M	ratio	k_{cat}/K_M	ratio	k_{cat}/K_M	ratio	k_{cat}/K_M	ratio	k_{cat}/K_M	ratio
S_{-1bp}	n.d.		n.d.		n.d.		0.11 ± 0.02	0.0095	5.9 ± 0.7	0.54	0.13 ± 0.02	0.012
S_0	n.d.		n.d.		0.11 ± 0.02	0.01	12.0 ± 0.9	1.1	11 ± 2	1	0.019 ± 0.002	0.0017
S_{+1bp}	n.d.		0.051 ± 0.006	0.0046	10 ± 1	0.87	14 ± 2	1.3	2.7 ± 0.4	0.25	0.013 ± 0.002	0.0012
S_{+2bp}	0.07 ± 0.01	0.0064	3.1 ± 0.2	0.28	7.4 ± 0.8	0.67	1.9 ± 0.3	0.17	n.d.		n.d.	
S_{+3bp}	0.35 ± 0.04	0.032	0.66 ± 0.09	0.064	0.20 ± 0.05	0.018	n.d.		n.d.		n.d.	
S_{+4bp}	0.040 ± 0.004	0.0036	0.020 ± 0.003	0.0018	n.d.		n.d.		n.d.		n.d.	

^aThe k_{cat}/K_M and their errors (in $\text{min}^{-1}\mu\text{M}^{-1}$) are average values and standard deviations, respectively, from at least three independent experiments. The ratios correspond to $(k_{cat}/K_M)/(k_{cat}/K_M)_0$, where $(k_{cat}/K_M)_0$ is the value obtained using R_0 and S_0 . Data for high-activity S/R pairs with ratios ≥ 0.25 are highlighted in grey. For several S/R pairs cleavage activity was not detected (n.d.; $k_{obs} \leq 0.001 \text{ min}^{-1}$) at $[R] \leq 5 \mu\text{M}$.

FIGURES

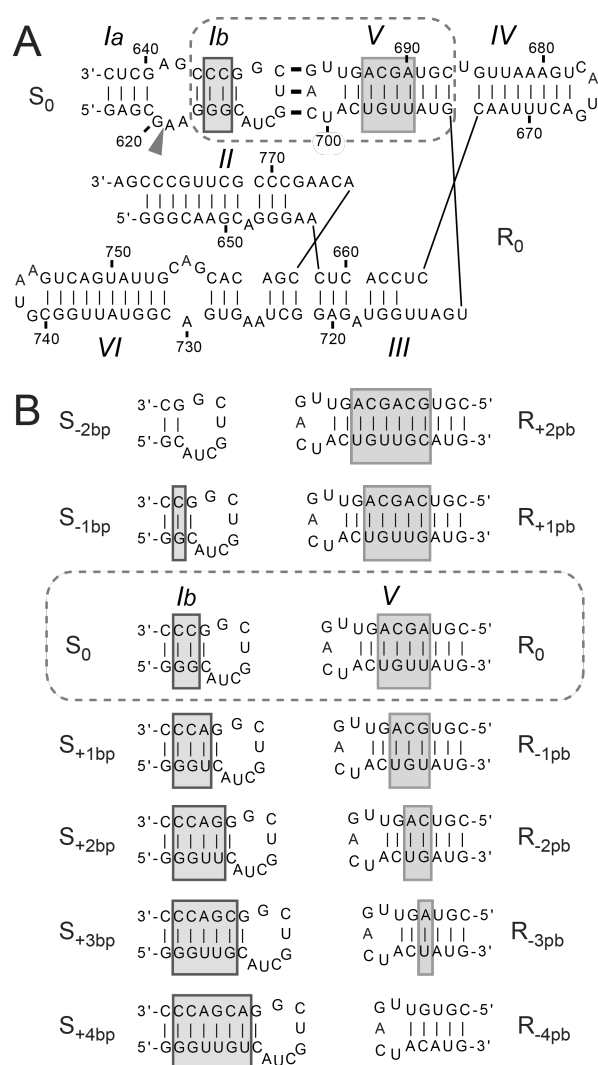


Figure 2.1 Primary and secondary structures of the substrate/ribozyme (S/R) pairs characterized here.

(A) Parental VS ribozyme (R_0) and substrate (S_0). Substrate recognition involves a kissing-loop interaction between residues 630-632 of S_0 and 697-699 of R_0 . The arrowhead points to the cleavage site and the dashed box encloses the region under investigation. (B) Mutant S/R pairs. Only the mutated regions are shown along with the nomenclature of the mutated substrates and ribozymes.

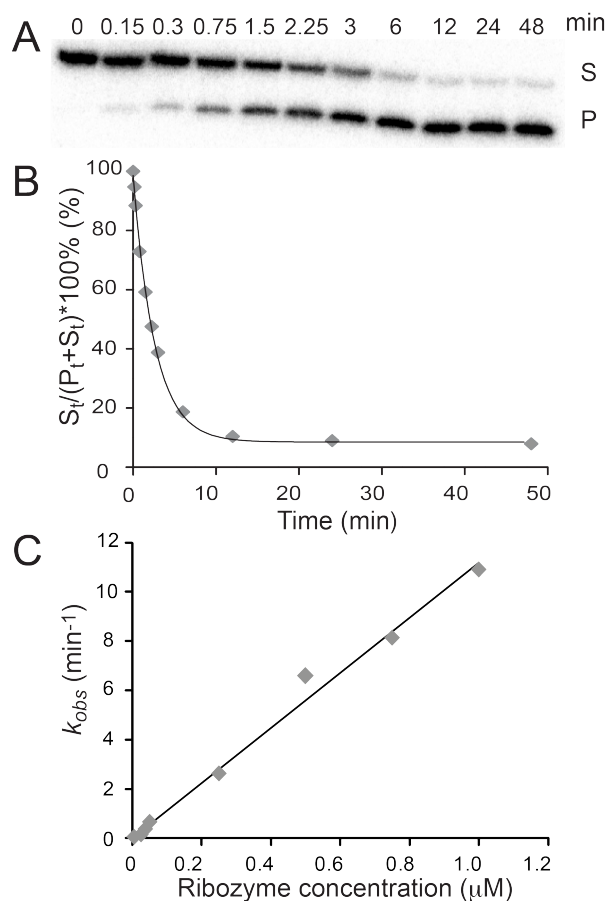


Figure 2.2 Single-substrate kinetic analysis of R_0 .

(A) Cleavage of 5'-³²P-labeled S_0 by R_0 (at 37.5 nM) monitored by denaturing gel electrophoresis. (B) The percentage of remaining substrate was plotted against time. The data were fitted to the non-linear equation $F = Ae^{-(k_{obs} * t)} + F_{\infty}$, to extract the value of k_{obs} ($k_{obs} = 0.332 \pm 0.008 \text{ min}^{-1}$; $R^2 = 1.00$). (C) Cleavage reactions were performed at various concentrations of R_0 , and the value of k_{obs} was plotted against ribozyme concentration. For all k_{obs} values, error bars are smaller than the data point on the graph. The data were fitted to a single exponential equation, and the k_{cat}/K_M is extracted by linear regression ($k_{cat}/K_M = \text{value of the slope} = 11.2 \text{ min}^{-1}\mu\text{M}^{-1}$ and $R^2 = 0.99$).

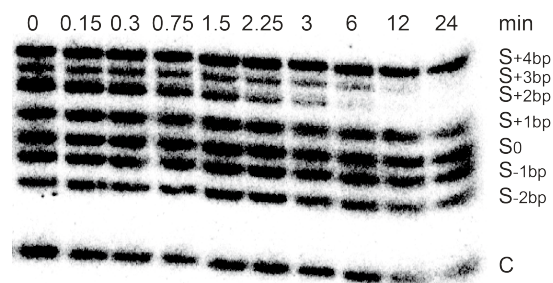


Figure 2.3 Simultaneous kinetic characterization of R_{-3bp} with multiple substrates.

The cleavage reactions of 5'- 32 P-labeled substrates (S) by R_{-3bp} (at 100 nM) were monitored by denaturing gel electrophoresis. A DNA hairpin was used as a normalization control (C).

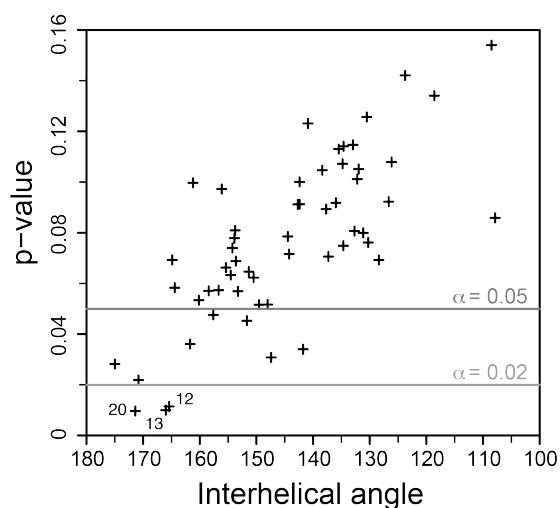


Figure 2.4 Identification of SLI/SLV models compatible with the kinetic data.

Plot of the P -value (Welch's t -test) against the mean interhelical angle. The P -value was obtained from a Welch's t -test that compared the rmsd distributions of the high-activity and low-activity classes of S/R pairs for a given common core group. The mean interhelical angle was obtained by computing the average interhelical angles of the 36 S/R pairs for each common-core group. All common-core groups showing a P -value < 0.02 are numbered. Standard deviations on the interhelical angles vary between 2° and 7° , but were omitted from the graph for clarity.

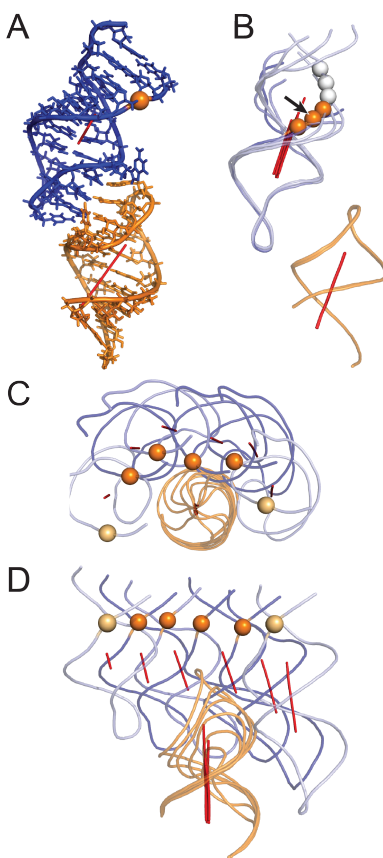
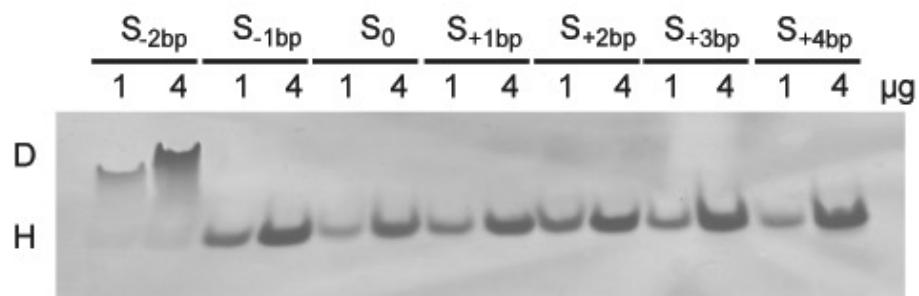


Figure 2.5 Model of an SLI/SLV complex compatible the kinetic data (common-core group 13).

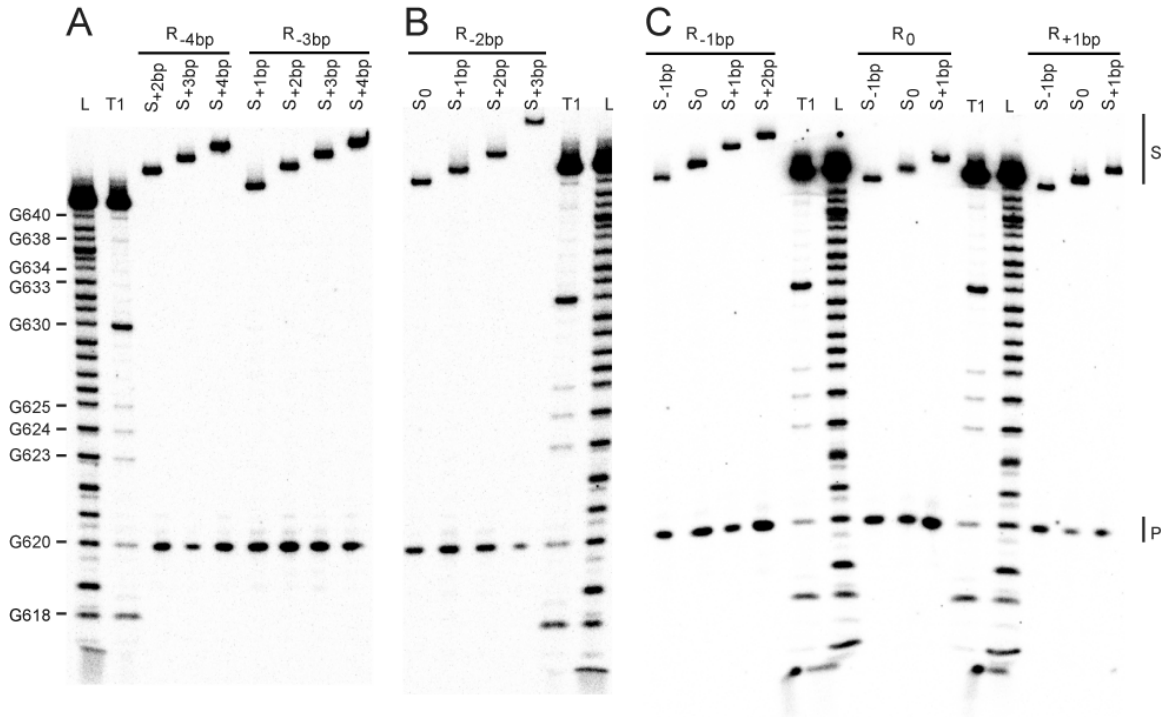
(A) Stick representation of the S_0/R_0 sub-model. (B) Cartoon representation of a superposition of the five mutant S/R_0 sub-models on the S_0/R_0 sub-model. The black arrow points at the scissile phosphate of S_0/R_0 . The scissile phosphates of the following substrates are shown from bottom to top: S_{-1} , S_0 , S_{+1} , S_{+2} , S_{+3} , S_{+4} . (C) Top-down and (D) front views of the superposition of the most catalytically efficient sub-models of each ribozyme. The following S/R pairs are shown from left to right: S_{-1}/R_{+1} , S_0/R_0 , S_{+1}/R_{-1} , S_{+1}/R_{-2} , S_{+2}/R_{-1} , S_{+3}/R_{-4} . In (B), (C) and (D), the sub-models were superposed to the S_0/R_0 sub-model based on heavy atom alignment of the first base pair of SLV (residues 689 and 707 in S_0/R_0). SLI and SLV are represented in deep blue and orange, respectively. Helical axes are represented as red rods and the phosphorus atom at the cleavage site is shown as a sphere. The phosphorus atoms are color-coded according to the catalytic efficiency of the sub-models $\{[(k_{cat}/K_M)/(k_{cat}/K_M) \geq 0.25]$: deep orange; $[0.25 > (k_{cat}/K_M)/(k_{cat}/K_M) > 0.01]$: light orange and $[(k_{cat}/K_M)/(k_{cat}/K_M) \leq 0.01]$: white}.

SUPPLEMENTARY INFORMATION



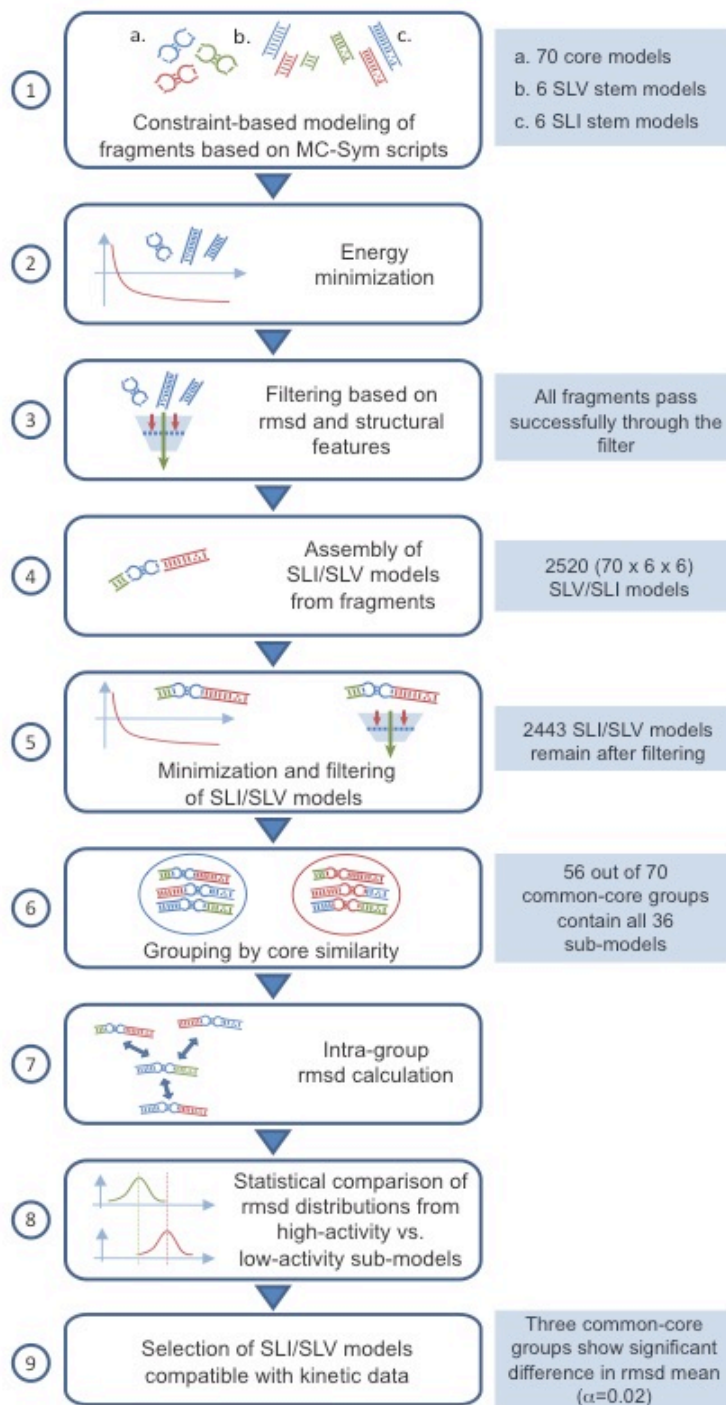
Supplementary Figure 2.1 Conformational analysis of the substrates used in this study by native gel electrophoresis.

For each substrate, two quantities of RNA (1 and 4 μg) were loaded on the gel. All substrates migrate at a similar position on the gel, except for S_{-2bp} which migrates slower, indicating that all but one of the substrates are in a single hairpin (H) conformation, whereas S_{-2bp} preferentially forms a duplex (D).



Supplementary Figure 2.2 Identification of a common cleavage site for all functioning S/R pairs.

Single substrate cleavage reactions of 5'-³²P-labeled substrates by: (A) R_{-4bp} and R_{-3bp}; (B) R_{-2bp}; and (C) R_{-1bp}, R₀ and R_{+1bp}. These reactions were analyzed by denaturing gels, which show the substrates (S) and the 5'-product (P). For comparison, S₀ was subjected to partial alkaline hydrolysis (L) and nuclease T1 digestion (T1).



Supplementary Figure 2.3 Flowchart describing the stepwise procedure used to generate three-dimensional models of the SLI/SLV complex and select those compatible with kinetic data.

Complete details are provided in Materials and Methods.

3.0 Substitution de l'interaction *kissing-loop* chez le ribozyme VS

Mise en contexte

Les études précédentes (au chapitre 2.0) ont mis en évidence l'adaptabilité de l'interaction *kissing-loop* I/V, qui a permis de modifier le ribozyme VS pour le clivage de substrats mutants possédant différentes longueurs de tige Ib. Dans le but de poursuivre les études d'ingénierie, l'objectif de ce deuxième article était de substituer l'interaction *kissing-loop* I/V par des interactions *kissing-loop* dont la structure se trouve dans les bases de données (PDB), pour savoir si le ribozyme VS peut être adapté pour reconnaître des substrats mutants avec des boucles différentes de celle de la tige-boucle SLI naturelle. L'interaction *kissing-loop* I/V est considérée comme une unité modulaire du ribozyme et par conséquent, sa substitution ne devrait pas avoir de conséquence importante sur le reste du ribozyme. Les mutants pour la substitution de l'interaction *kissing-loop* I/V ont été identifiés par une méthode bio-informatique, afin d'obtenir une interaction qui serait compatible avec l'activité de clivage du ribozyme VS. Les deux interactions *kissing-loop* sélectionnées sont : l'interaction TAR/TAR* du VIH-1 et l'interaction L88/L22 de l'ARNr 23S de l'eubactérium *Deinococcus radiodurans*. Ces interactions ont été transposées dans les boucles I et V du ribozyme VS et plusieurs longueurs de tiges Ib et V ont été testées comme au chapitre 2.0 pour déterminer les longueurs les plus appropriées pour accommoder les nouvelles interactions *kissing-loop*.

Les études de cinétiques ont démontré que chacune des substitutions de l'interaction *kissing-loop* permet d'observer une activité de clivage substantielle. Par contre, l'activité de clivage est réduite comparativement à l'activité observée pour le ribozyme et le substrat de référence, respectivement R_0 et S_0 (chapitre 2.0). Cette baisse d'activité n'est pas due à une l'absence de la formation de l'interaction *kissing-loop*, puisque lorsque l'interaction ne peut pas se former, l'activité de clivage est diminuée par un facteur beaucoup plus grand que celui que nous observons. Une compensation des longueurs d'hélices a été observée dans cette étude, tout comme dans le cadre de l'étude précédente (chapitre 2.0). Des modèles des complexes substrat/ribozymes mutants ont été produits pour tenter d'expliquer cette activité plus basse chez les mutants. Ces modèles montrent que les mutants pour les interactions *kissing-loop* limitent vraisemblablement la formation du site actif du ribozyme, comparativement à l'interaction *kissing-loop* I/V.

Pour cet article, je suis première auteure en raison de l'importance de ma contribution à cet article. Nicolas Girard a été responsable des études bio-informatiques pour la recherche des interactions *kissing-loop* potentielles et pour les modèles des interactions *kissing-loop*. Les expériences d'ITC de l'interaction L22/L88 ont été effectuées par Pierre Dagenais. Ma contribution réside dans les études de cinétiques enzymatiques, de gel natif et de SHAPE. J'ai effectué la transcription *in vitro* des ribozymes mutants à l'aide de plasmides et des substrats mutants à l'aide de gabarits d'ADN. J'ai effectué les études de cinétiques enzymatiques et la vérification des conformations des substrats avec des gels natifs et celles des ribozymes avec la technique de SHAPE. J'ai participé à l'écriture des sections du manuscrit pertinentes à ma participation scientifique et Nicolas Girard a fait de même. Pascale Legault a conçu le projet et a supervisé l'ensemble de l'étude, incluant la collecte des données et l'analyse des résultats. Elle a relu et corrigé ce manuscrit.

**Rational engineering of the *Neurospora* VS ribozyme to
allow substrate recognition via different kissing-loop
interactions**

*Julie Lacroix-Labonté, Nicolas Girard, Pierre Dagenais and Pascale Legault**

Département de Biochimie et médecine moléculaire, Université de Montréal, C.P. 6128,
Succursale Centre-Ville, Montréal, QC, Canada, H3C 3J7

*To whom correspondence should be addressed: Pascale Legault, Phone: 514-343-7326,
Fax: 514-343-2210.

This research was originally published in NAR journal:

Nucleic Acids Res. 2016 May; doi: 10.1093/nar/gkw401.

ABSTRACT

The *Neurospora* VS ribozyme is a catalytic RNA that has the unique ability to specifically recognize and cleave a stem-loop substrate through formation of a highly stable kissing-loop interaction. In order to explore the engineering potential of the VS ribozyme to cleave alternate substrates, we substituted the wild-type kissing-loop interaction by other known kissing-loop interactions using an innovative engineering method that combines rational and combinatorial approaches. A bioinformatic search of the protein data bank was initially performed to identify kissing-loop interactions that are structurally similar to the one found in the VS ribozyme. Next, substrate/ribozyme pairs that incorporate these alternative kissing-loop interactions were kinetically and structurally characterized. Interestingly, several of the resulting substrate/ribozyme pairs allowed substrate cleavage with substantial catalytic efficiency, although with reduced activity compared to the reference substrate/ribozyme pair. Overall, this study describes an innovative approach for RNA engineering and establishes that the kissing-loop interaction of the *trans* VS ribozyme can be adapted to cleave other folded RNA substrates.

Keywords: rational engineering of RNA, kissing-loop interactions, helix-length compensation, *Neurospora* VS ribozyme, RNA structure adaptability

INTRODUCTION

The *Neurospora* VS ribozyme belongs to the family of natural nucleolytic ribozymes, which are less than 200 nucleotides in size and include the hairpin, HDV, hammerhead, *glmS*, twister, twister sister, pistol and hatchet ribozymes [1-5]. Like many other ribozymes, the VS ribozyme provides a good model system for fundamental studies of RNA structure and function, which are of value for understanding more complex RNA-mediated processes [6]. So far, a large body of *in vitro* work has led to a comprehensive functional and structural understanding of the VS ribozyme both in its *cis* and *trans* forms, which vary according to the covalent attachment of the stem-loop I (SLI) substrate domain (for reviews see [1,7-9]). These studies have highlighted a unique mode of substrate recognition for the VS ribozyme, which involves specific recognition of the SLI domain in a folded hairpin state [10]. Only a few studies have addressed the possibility of using the VS ribozyme to target and cleave RNA stem-loops that differ from the natural SLI substrate. In contrast, several other small nucleolytic ribozymes, particularly the hairpin, hammerhead and HDV ribozymes, have been modified to cleave numerous alternative single-stranded RNA substrates. For the VS ribozyme, the engineering principles that govern recognition of alternate substrates are likely more complex, but remain to be elucidated.

The catalytic domain or *trans* VS ribozyme (helical domains II-VI; Fig. 3.1A) recognizes its hairpin substrate (stem-loop I or SLI) via the formation of a kissing-loop interaction (KLI) required for efficient cleavage activity [10,11]. In earlier work, mutagenesis and chemical modification studies provided evidence that this KLI contains three WC/WC base pairs (bp) between loop I (G₆₃₀, U₆₃₁ and C₆₃₂) and loop V (G₆₉₇, A₆₉₈ and C₆₉₉), and that both loops likely form U-turn motifs [11]. More recently, the NMR structure of an SLI/SLV complex [12] and the crystal structure of the complete VS ribozyme [13] confirmed formation of the U-turn motifs and three *cis* WC/WC bp, but they also revealed the presence of a *cis* WC/sugar edge bp between C₆₂₉ and A₇₀₁ (between C₁₀ of SLI and A₁₄ of SLV in the SLI/SLV complex; Fig. 3.1B). *In vitro* selection and biochemical studies of the VS ribozyme [14,15] as well as NMR studies of several SLI/SLV complexes [12] have provided evidence that formation of the I/V KLI is

associated with a helix shift in the SLI substrate that is essential for catalysis. SLI variants that adopt a pre-shifted conformation bind with higher affinity to both the *trans* ribozyme [16] and an isolated SLV [17]. Moreover, isothermal titration calorimetry studies of SLI/SLV complexes conducted with a pre-shifted SLI RNA demonstrated that the I/V KLI is remarkably stable ($K_d = 0.24 \pm 0.02 \mu\text{M}$) and constitutes the predominant energetic contribution for substrate binding by the *trans* VS ribozyme [17]. The I/V KLI is also important for substrate docking with the catalytic domain, which allows for the formation of the active site [18]. In the proposed general acid-base mechanism of the VS ribozyme cleavage reaction, G₆₃₈ from the internal loop of SLI (G₆₃₈ loop) acts as the general base and A₇₅₆ from the internal loop of SLVI (A₇₅₆ loop) acts as the general acid [19-26]. The recent crystal structure of the VS ribozyme confirms that the active site is formed through the association of these two internal loops, in which G₆₃₈ and A₇₅₆ are properly positioned with respect to the scissile phosphate to fulfill their proposed roles in catalysis [13].

Only a limited number of studies have investigated the possibility of using the VS ribozyme to selectively cleave RNA stem-loops other than its natural SLI substrate. Although minor changes to the sequence of the natural SLI substrate are compatible with VS ribozyme cleavage [11,14,27-30], larger sequence changes in the substrate may only be possible by incorporating compensating changes in the catalytic domain. To test this possibility, we recently investigated the cleavage of variant SLI substrates, in which the number of base pairs in stem Ib was either increased or decreased, using variant *trans* ribozymes containing a variable number of base pairs in stem V [31]. We observed helix-length compensation between stem Ib and stem V, since the highest activity for each ribozyme variant was observed with a substrate that contains the compensatory number of base pairs in stem Ib (or one less). Moreover, this study establishes that VS ribozyme derivatives efficiently cleave SLI substrates that contain an alternate number of base pairs in stem Ib.

Given these encouraging results with the VS ribozyme, we were interested to gain further insights into the adaptability of this unique ribozyme for specific cleavage of

folded RNA stem-loops. Previous studies have led to the idea that the I/V KLI of the VS ribozyme is an independent RNA module [12,15,17,30], and we hypothesized that this KLI can be substituted by other KLIs to create variant ribozymes that efficiently cleave stem-loop substrates with a modified terminal loop. In the present study, we developed a rational engineering approach to replace the I/V KLI of the *Neurospora* VS ribozyme by other KLIs and achieve efficient substrate cleavage (Fig. 3.1C). This rational approach is supported by the recently available NMR structure of an SLI/SLV complex [12] and the crystal structure of the complete VS ribozyme [13]. We first performed a bioinformatic search for KLIs in the protein data bank (PDB), which allowed us to identify two KLIs that share structural characteristics with the natural KLI of the VS ribozyme. Next, VS substrate/ribozyme (*S/R*) pairs, in which the I/V KLI was substituted by each of these two KLI candidates, were prepared using several combinations of helix length for stems Ib and V [31] in order to fine-tune the structural context of the surrogate KLIs necessary to obtain efficient cleavage of the modified substrates. The engineered *S/R* pairs were then kinetically characterized to identify the most efficient ones and structurally evaluated to validate the procedure. Interestingly, several novel *S/R* pairs were identified that display substantial catalytic activity, further illustrating the adaptability of the VS ribozyme architecture for cleavage of alternate stem-loop substrates.

MATERIALS AND METHODS

Search for KLIs with WebFR3D. A bioinformatic search was performed with WebFR3D [32] to find KLIs in the PDB. The search query was defined as an interaction matrix between a set of six nucleotides with no specific nucleotide identity (NT1 to NT6; Supplementary Fig. 3.1B). Nucleotides NT1 and NT3 form a *cis* WC/WC base pair that closes the first loop, and this loop contains a non self-pairing strand of 4-10 nucleotides. Nucleotides NT4 and NT6 form a *cis* WC/WC base pair that closes the second loop, and this loop also contains a non self-pairing strand of 4-10 nucleotides. Nucleotides NT2 and NT5 are part of the first and second loops, respectively, and must form a *cis* WC/WC base pair. Finally, residues NT1-NT3 must be either sequentially located 5' respectively to residues NT4-NT6 or part of a different chain. For this search, we used the WebFR3D list of non-redundant RNA-containing structures determined by X-ray crystallography, Cryo-EM or NMR spectroscopy.

Clustering of WebFR3D results. A custom Python script was used to parse the HTML output file of WebFR3D. The script extracts two key features about each entry: the PDB ID code of the associated structure and the type and numbering of the nucleotides (NT1-NT6) matching the query. It then uses this information to cluster all fragments associated with a single KLI and quantifies the number of *cis* WC/WC base pairs associated with this interaction.

Structural data retrieval and result listing. A custom Python script was used to automatically gather information about each KLI. The size of each loop and the total number of *cis* WC/WC base pairs in the KLI were retrieved directly from the cluster of WebFR3D results and used to define the secondary structure of each KLI. Additional information was retrieved from the RCSB repository via the PDB ID code of the KLI. The information gathered about each KLI structure was listed as an HTML page enriched with JavaScript (Supplementary Fig. 3.1D and supplementary Table III.I).

Plasmids. The Avapl ribozyme (R_0 , Fig. 3.1A) is a derivative of the Aval ribozyme and its parental plasmid (pAvapl) was constructed by PCR from the A-3 clone [10]. The

plasmids for transcription of the R_{TAR^*} , R_{L22} and R_{L88} ribozyme variants (Fig. 3.3A and Supplementary Figs. 3.3A and 3.4A) were prepared by QuikChange II site-directed mutagenesis (Stratagene, CA). All plasmids were validated by DNA sequencing.

ITC studies. Preparation of RNA and isothermal titration calorimetry (ITC) studies were performed as previously described [17]. Relevant experimental details are provided in the legend of Supplementary Fig. 3.2.

Preparation of RNA for kinetic studies. The R_{TAR^*} , R_{L22} and R_{L88} ribozymes were prepared by *in vitro* T7 RNA polymerase transcription from plasmid templates linearized at the *Ava*I site. All ribozymes were purified by gel electrophoresis and ion-exchange HPLC, as previously described [31]. The S_{TAR} , S_{L88} and S_{L22} substrates were synthesized *in vitro* using the T7 RNA polymerase and single-stranded synthetic oligonucleotide templates (Integrated DNA Technologies, IA). These substrates were purified by gel electrophoresis, dephosphorylated with calf alkaline phosphatase and 5'-end-labeled with γ -(32 P) ATP (PerkinElmer, Ont.) using established procedures [31].

Kinetic studies. Multiple-substrate and single-substrate kinetic studies were carried out as single-turnover reactions at 37°C in 1X Kinetic Reaction buffer (50 mM Tris at pH 7.6, 25 mM KCl, 2 mM spermidine and 25 mM MgCl₂), as previously described [31]. Briefly, multiple-substrate kinetic studies were typically performed with a mixture of 5'- 32 P-labeled nucleic acids containing all purified substrates for a given *S/R* pair (35 pM of each) and excess *R* (0.005 to 25 μ M). These multiple-substrate kinetic studies were used to identify quickly the *S/R* pairs that lead to detectable substrate cleavage. All multiple-substrate kinetic experiments were performed at least three times.

Reactive *S/R* combinations were further investigated by single-substrate kinetic studies to derive the values of k_{obs} and k_{cat}/K_M . These studies were typically performed with $[S] = 250$ pM and excess *R* (0.005 to 15 μ M). The percentage of remaining substrate $[F = S_t/(S_t+P_t)*100\%$; where S_t and P_t are the amount of substrate and product at time t] was plotted against time. The resulting time courses were fit to the equation $F = Ae^{-(k_{obs} * t)} + F_\infty$, where A is the amplitude, k_{obs} is the rate of the reaction and F_∞ is a variable that

represents the unreactive fraction of the substrate. The value of k_{obs} , the standard error for k_{obs} , and the F_{∞} parameter were estimated by non-linear regression analysis within the OriginPro 8.1 software (OriginLab, MA). The F_{∞} values are below 10% for all S/R combinations. A linear dependence of k_{obs} on $[R]$ was observed for each S/R combination at the lowest values of $[R]$ and analyzed by linear regression to derive the second order rate constant of the reaction, k_{cat}/K_M . The quality of the fit for the linear dependence of the k_{obs} on $[R]$ was obtained from the square of the correlation coefficient (R^2), and in all cases R^2 was ≥ 0.99 . For the single-substrate kinetic experiments, three independent experiments were performed, and reported k_{cat}/K_M and their errors are average values and standard deviations, respectively, from these multiple measurements.

Native gel studies of RNA substrates. Non-denaturing polyacrylamide gel electrophoresis was carried out on a 7.5% polyacrylamide bisacrylamide (37.5:1) gel in Tris-glycine buffer (25 mM Tris pH 8.0 and 200 mM glycine). Prior to loading, the RNA substrates (1 and 4 μg) were heated at 95°C for 2 min in storage buffer (10 mM Tris pH 7.6 and 1 mM EDTA) and transferred on ice for 5 min. Then, 4 μL of 2X Kinetic Reaction buffer was added and the samples were incubated at 37°C for 5 min. The gel was pre-run for 30 min at 100 V, run for 2 hrs at 250 V and 4°C and then stained with Stains All (Sigma-Aldrich, Ont.).

SHAPE studies of ribozymes. For SHAPE (Selective 2'-hydroxyl Acylation analyzed by Primer Extension) experiments, variant ribozymes (20 pmol) in 4 μL of H_2O were first incubated at 37°C for 5 min with 5 μL of 2X Kinetic Reaction buffer and then treated with 1 μL of 130 mM *N*-methylisatoic anhydride (NMIA; dissolved in anhydrous DMSO). The resulting mix was incubated for 54 min at 37°C and then transferred on ice. Reverse transcription reactions were performed on mutant ribozymes with 5'-³²P-labeled primers and SuperScript III reverse transcriptase (ThermoFisher Scientific Inc.), as previously described [33]. The primer extension products were separated on an 8% denaturing polyacrylamide gel (7 M urea, 90 mM Tris/Borate, 2 mM EDTA). The gel was visualized by phosphorimager (Personal Molecular ImagerTM System, Bio-Rad) and quantified using Image Lab Software 4.1.

Three-dimensional modeling of S/R complexes with a substituted KLI. Models of the S_0/R_0 , S_{L88-0}/R_{L22-0} and S_{TAR-0}/R_{TAR*-0} complexes were constructed by fragment assembly in PyMOL (The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.) based on heavy-atom superpositions with the S_X/R_X complex obtained from the X-ray structure of the VS ribozyme (PDB ID code 4R4P; [13]) by taking the SLI substrate of one protomer (S_X) and the *trans* VS ribozyme (helical domains II-VI) of the other protomer (R_X). Starting with representative high-resolution structures of the studied KLIs (PDB ID codes 2MI0 -model 2 [12]), (1KIS - model 2 [34] and 4IOA [35]), the hairpin stems within these complexes were modified by stems generated with MC-Sym [36] that match those of SLI (residues 623-626 and 634-637) and SLV (residues 688-694 and 702-708) from the S_0/R_0 complex (Fig. 3.1A), and the resulting hybrid structures retained the hairpin loop residues of the original high-resolution structures (2MI0: residues 4-17 and 104-118; 1KIS: residues 5-12 and 21-28; and 4IOA: residues 423-430 and 2383-2392). The hairpin stem corresponding to stem V within these hybrid structures was superposed to stem V of R_X . Subsequently, S_X was copied one time for each hybrid structure and its stem Ib was superposed to stem Ib of each hybrid structure. The final S/R complexes retain all residues from the hybrid structures, all residues of R_X except for residues 688-708 of SLV and only residues 616-622 and 638-643 of S_X (according to residue numbering in Fig. 3.1).

RESULTS

Search for KLIs in the PDB database. We searched for three-dimensional structures of KLIs in the PDB using a 3-step computational workflow (Supplementary Fig. 3.1A) that is based on the motif-search tool WebFR3D [32]. This online version of the FR3D program was chosen because it has the capacity to perform searches of RNA motifs composed of non-contiguous nucleotides in a curated ensembles of non-redundant structures. In the first step, a search query was launched with WebFR3D to identify single *cis* Watson-Crick (WC) base pairings between any two residues belonging to two different terminal loops of 4-10 residues in length (Supplementary Fig. 3.1B). This search used the WebFR3D list of non-redundant RNA-containing structures determined by X-ray crystallography, Cryo-EM or NMR spectroscopy. In the second step, the individual

entries of the WebFR3D output were clustered to identify all existing WC pairings for a specific pair of terminal loops, and thereby provide a list of KLIs. In the third and final step, additional structural data and related information were retrieved for each KLI, and these data were listed in an interactive HTML/JavaScript table (Supplementary Fig. 3.1D and Supplementary Table III.I). From this motif search of the PDB, 113 KLIs were identified, and these form between one and ten *cis* WC/WC base pairs at the kissing-loop interface.

Selection of KLIs that could substitute for the VS ribozyme I/V interaction. To identify KLIs that could substitute for the VS ribozyme I/V KLI, we automatically filtered the list of 113 structures to eliminate all entries originating from X-ray and Cryo-EM structures with resolution ≥ 3.6 Å and only selected those that form heterodimeric KLI interactions with 3 to 10 *cis* WC/WC base pairs. Heterodimeric KLIs were selected because substitutions of SLI and SLV in the VS ribozyme *S/R* system (Fig. 3.1A) by stem-loops that form an homodimeric KLI might lead to self-association of substrates and ribozymes, which could interfere with *trans* cleavage. In addition, KLIs with a minimum of three *cis* WC/WC base pairs were selected to prevent formation of an unstable KLI that could lead to inefficient *trans* cleavage [11,30]. This filtering step produced a list of 17 structures that represent 6 types of KLIs (Table III.I): the TAR/TAR* interaction derived from HIV-1; the L6/L7 and L22/L88 interactions from the large subunit of rRNAs; the L5/L13 interaction from the cobalamin riboswitch; the L3/L4 interaction from the human SRP Alu domain; and the Cole1 KLI. To identify suitable KLIs for our engineering studies, we selected those that share structural similarities with the VS ribozyme I/V KLI. More specifically, a suitable KLI must fit certain structural characteristics: *i*) form consecutive *cis* WC/WC base pair with non-modified nucleotides, *ii*) the adjoining stems must contain at least three consecutive *cis* WC/WC base pairs and *iii*) the interhelical angle must fall between 125° and 175° [12]. Such structural criteria limited our choice to two classes of KLIs: the HIV-1 TAR/TAR* interaction and the L22/L88 (loop 22/loop 88) interaction from the large subunit of rRNAs, and those with PDB ID codes 1KIS [34] and 4IOA [35] were selected for our engineering studies.

Thermodynamic stability of the selected KLIs. To ascertain that the two selected KLIs are good candidates for our engineering studies, we verified their thermodynamic stability. Although the TAR/TAR* and L22/L88 KLIs incorporate a larger number of WC base pairs than the VS ribozyme I/V KLI (Figs. 3.1B and 3.2AB), the latter is known to be exceptionally stable [17]. Based on isothermal titration calorimetry studies, the K_d of the TAR/TAR* and L22/L88 complexes are 40-fold lower and 16-fold higher, respectively, than that of the SLI/SLV complex (Fig. 3.2C and Supplementary Fig. 3.2). These results confirm that the thermodynamic stability of the two selected KLIs is not considerably different from that of the I/V KLI. Thus, these two KLIs were considered suitable for our engineering studies.

Kinetic characterization of VS ribozyme S/R pairs with a modified I/V KLI. Next, kinetic studies were performed to test the hypothesis that the two selected KLIs can substitute for the natural I/V KLI interaction and yield active VS ribozyme variants. For these studies, we used the previously-characterized S_0/R_0 pair formed between a pre-shifted SLI substrate and a *trans* VS ribozyme [31]. Since the selected KLIs may functionally replace the I/V KLI only when substituted in the proper orientation (e.g. either the L22 or the L88 loop replacing loop I) and with supporting helices of proper lengths, we prepared three combinatorial pools of S/R pairs that consist of 25 S_{TAR}/R_{TAR^*} pairs, 30 S_{L88}/R_{L22} pairs and 36 S_{L22}/R_{L88} pairs (Fig. 3.3A and Supplementary Figs. 3.3A and 3.4A). Within each pool, helix-length variations were introduced in both stem Ib of the variant substrates (S_{TAR} , S_{L88} and S_{L22}) and stem V of the variant ribozymes (R_{TAR^*} , R_{L22} and R_{L88}). In this context, compensatory S/R pairs are those for which the total number of base pairs in stems Ib and V is identical to that found for the S_0/R_0 reference pair (Fig. 3.1). For all S/R combinations, the second-order rate constant k_{cat}/K_M was determined under single turnover conditions. It is important to note that the catalytic parameter k_{cat}/K_M is very useful when comparing the relative activity of a given enzyme for different competing substrates (e.g. R_{TAR^*-0} acting on the different S_{TAR} substrates), but may be misleading when comparing the relative activity of different enzymes [37], particularly those of different S/R families. Thus, to compare the catalytic activity of

different S/R families, we also relied on the values of the cleavage rate constant (k_{obs}) measured under identical experimental conditions.

Kinetic studies of the VS ribozyme S_{TAR}/R_{TAR^*} pairs. Single turnover kinetic studies were performed to derive the second-order rate constant k_{cat}/K_M for all 25 S_{TAR}/R_{TAR^*} combinations. ^{32}P -labeled S_{TAR} substrates were cleaved with an excess of R_{TAR^*} ribozymes, and the cleavage reaction monitored by denaturing gel electrophoresis. Initially, we used a multiple-substrate kinetic approach, where one enzyme is simultaneously incubated with a mixture of substrates [31]. Given that the substrates are of different sizes (Fig. 3.3A), the fate of any given substrate can be monitored by gel electrophoresis [31]. For several of the S_{TAR}/R_{TAR^*} combinations, no cleavage activity was detected after 24 min at $[R] \leq 15 \mu\text{M}$ ($k_{obs} \leq 0.002 \text{ min}^{-1}$; light gray shading in Fig. 3.3B). For the remaining S_{TAR}/R_{TAR^*} combinations that displayed detectable cleavage activity, a single-substrate kinetic approach was used to precisely derive k_{cat}/K_M values. The highest k_{cat}/K_M value ($0.53 \pm 0.03 \text{ min}^{-1}\mu\text{M}^{-1}$) was obtained for the S_{TAR-1}/R_{TAR^*-0} combination (Fig. 3.3B), which is 46-fold lower than that of S_0/R_0 ($k_{cat}/K_M = 24.5 \pm 0.9 \text{ min}^{-1}\mu\text{M}^{-1}$ at 37°C ; data not shown). Several other S_{TAR}/R_{TAR^*} combinations have similar catalytic efficiency (within 5 fold; dark gray shading in Fig. 3.3B), and these combinations involve all five R_{TAR^*} ribozymes and four of the five S_{TAR} substrates.

For R_{TAR^*-0} , efficient substrate cleavage is obtained with S_{TAR-1} as well as with S_{TAR-0} , indicating limited substrate promiscuity. Other R_{TAR^*} ribozymes also display similar specificity, cleaving efficiently only two to three out of the five S_{TAR} substrates. Given that the highest k_{cat}/K_M values for each ribozyme is observed with the compensatory substrate or one that varies by one or two base pairs, it can be concluded that efficient substrate cleavage depends on helix-length compensation in stem V of the variant ribozyme. For example, the optimal ribozyme for S_{TAR-1} is R_{TAR^*-0} , whereas the optimal one for S_{TAR+2} is R_{TAR^*-3} . Thus, for the S_{TAR}/R_{TAR^*} family, there is limited substrate promiscuity for all ribozymes as well as helix-length compensation across the family, as previously observed for the reference S/R family [31].

Kinetic studies of the VS ribozyme S_{L88}/R_{L22} and S_{L22}/R_{L88} pairs. In a similar manner, multiple and single turnover kinetic studies were performed to evaluate the k_{cat}/K_M of the 30 S_{L88}/R_{L22} and 36 S_{L22}/R_{L88} combinations (Supplementary Figs. 3.3 and 3.4). For the S_{L88}/R_{L22} pairs, the most active combination is S_{L88+1}/R_{L22-1} with a k_{cat}/K_M of $0.39 \pm 0.03 \text{ min}^{-1} \mu\text{M}^{-1}$, which is 63-fold lower than that of S_0/R_0 . A few other S_{L88}/R_{L22} combinations produce similar k_{cat}/K_M (within 5 fold; dark gray shading in Supplementary Fig. 3.3B), and these combinations involve three of the five S_{L88} substrates and three of the six R_{L22} ribozymes. Each of these three ribozymes displays similar cleavage activity for one or two S_{L88} substrates, and helix-length compensation is observed given that the most active S_{L88}/R_{L22} combinations for these ribozymes are the S_{L88-0}/R_{L22-0} , S_{L88+1}/R_{L22-1} and S_{L88+2}/R_{L22-2} pairs. For the S_{L22}/R_{L88} pairs, the highest k_{cat}/K_M obtained is $0.23 \pm 0.02 \text{ min}^{-1} \mu\text{M}^{-1}$ for S_{L22+1}/R_{L88-1} , which is 107-fold lower than that of S_0/R_0 . In contrast to the S_{L88}/R_{L22} combinations, several S_{L22}/R_{L88} combinations display similar catalytic activity (within 5 fold; dark gray shading in Supplementary Fig. 3.4B), and these combinations involve four of the six S_{L22} substrates and five of the six R_{L88} ribozymes. For each of the five active ribozymes, comparable k_{cat}/K_M values are obtained with one to four of the six S_{L88} substrates, indicating different degrees of substrate promiscuity. Furthermore, helix-length compensation is observed for all variant ribozymes given that the highest k_{cat}/K_M values for each ribozyme is observed with the compensatory substrate or one that varies by one base pair. For example, the optimal substrate for R_{L88+1} is S_{L22+0} , whereas the optimal one for R_{L88-4} is S_{L22+4} . Thus, for both the S_{L88}/R_{L22} and S_{L22}/R_{L88} families, there is limited substrate promiscuity for all ribozymes as well as helix-length compensation across the family, as previously observed for the reference S/R family [31].

Comparison of the catalytic activity of the different S/R families. Given that k_{cat}/K_M values can be misleading when comparing the relative activity of the various S/R families, we also compared the k_{obs} values for the S/R pair of each family that displays the highest catalytic activity. These k_{obs} values, determined under identical single turnover conditions, indicate that the activities of the S_{TAR}/R_{TAR^*} , S_{L88}/R_{L22} and S_{L22}/R_{L88} family representatives are reduced by 159-fold, 53-fold and 84-fold, respectively, compared to the parental S_0/R_0 pair (Supplementary Table III.II). Thus, although both the k_{cat}/K_M and

k_{obs} values indicate that there is a significant level of catalytic activity for some variant S/R pairs, their activity is still lower than that of the parental S_0/R_0 pair.

Structural characterization of the variant substrates. To insure that the variant substrates examined in this study are properly folded, these substrates were analyzed by native gel electrophoresis. The migration of variant substrates on native gels was compared with that of two previously investigated VS ribozyme substrates [31], S_{-1bp} and S_{-2bp} , which respectively adopt a hairpin and a duplex conformation. Our results show that all variant substrates migrate on a native gel in a manner similar to the active hairpin conformation (Supplementary Fig. 3.5).

Structural characterization of the variant ribozymes. To verify the proper folding of variant ribozymes, the secondary and tertiary structures of these *trans* ribozymes were analyzed by SHAPE experiments. In these studies, the reactivity of each nucleotide to NMIA is quantified and interpreted as being directly linked to the local nucleotide flexibility [33]. For the R_{TAR^*-0} ribozyme (Fig. 3.4), the nucleotides reactive to NMIA are found in non-helical regions that are likely to be dynamic, including the A_{718} and $A_{725}A_{726}$ bulges, the A_{756} loop and the terminal loops of stem-loops IV, V and VI [38-40]. The reactivity patterns of the other R_{TAR^*} ribozymes are very similar to that of the R_{TAR^*-0} ribozyme (Fig. 3.4). In addition, the reactivity patterns of the R_0 , R_{L22} and R_{L88} ribozymes are similar to those of the R_{TAR^*} ribozymes in that known helical regions display lower reactivity to NMIA compared to non-helical regions (Supplementary Fig. 3.6). As an exception, residues within and adjacent to G-U base pairs (in stem-loops V and VI) produce idiosyncratic patterns of NMIA reactivity, as previously demonstrated [42]. Of note, the most NMIA-reactive nucleotide in all ribozymes is A_{726} , an extruded residue from the II-VI bulge-bulge interaction and the most disordered residue in the NMR structures of the II-III-VI junction [13,41,43]. Overall, the SHAPE results for all of the engineered ribozymes are in good agreement with the structural schematics of the R_0 ribozyme derived from NMR and X-ray structures (Fig. 3.1; [13,41]). Thus, all of the variant VS ribozymes investigated here adopt a ground state conformation that is compatible with cleavage activity.

Structural modeling of the S_0/R_0 , S_{L88-0}/R_{L22-0} and S_{TAR-0}/R_{TAR*-0} pairs. Once a stable KLI is formed between a well-folded VS ribozyme variant and its stem-loop substrate, the G₆₃₈ internal loop of the substrate must properly interact with the A₇₅₆ loop of the ribozyme to allow formation of a productive active site. In order to assess the effect of substituting the I/V KLIs on formation of the VS ribozyme active site, the S_{L88-0}/R_{L22-0} and S_{TAR-0}/R_{TAR*-0} substrate/ribozyme pairs were modeled by fragment assembly, combining either the X-ray structure of the L22/L88 KLI or the NMR structure of the TAR/TAR* KLI with the X-ray structure of the VS ribozyme (Fig. 3.5; [13,34,35]). As a control, the same exercise was performed with the NMR structure of the SLI/SLV complex to model the S_0/R_0 pair. In contrast with the close state observed in the X-ray structure (Fig. 3.5B), the model of the S_0/R_0 pair (Fig. 3.5C) adopts an open state with the G₆₃₈ loop of SLI apart from the A₇₅₆ loop of SLVI. This is not surprising, since the open state, previously described in the NMR-based model of the complete VS ribozyme, likely represents a ground-state conformation of the S/R complex [41]. In the S_0/R_0 model, the scissile phosphate within the G₆₃₈ loop is 26.6 Å from the corresponding phosphate in the crystal structure, but oriented directly towards A₇₅₆, such that a simple repositioning of the helical domains containing the G₆₃₈ and A₇₅₆ loops would allow formation of the active site. The models of both the S_{L88-0}/R_{L22-0} (Fig. 3.5D) and S_{TAR-0}/R_{TAR*-0} pairs (Fig. 3.5E) also represent an open state of the ribozyme with their scissile phosphates respectively 23.0 Å and 31.9 Å from the corresponding phosphate within the crystal structure. However, in both these models, the scissile phosphate is not oriented favorably with respect to the A₇₅₆ loop to facilitate formation of the active site. Overall, these models suggest that substitutions of the KLI in the variant S/R pairs may reduce the probability of encounter between the G₆₃₈ internal loop of the substrate with the A₇₅₆ loop of the ribozyme and thus likely impair both the formation of the active site and the subsequent cleavage reaction.

DISCUSSION

In this work, a comprehensive RNA engineering approach was developed, which demonstrates that the I/V KLI of the VS ribozyme is an RNA module that can be substituted by other KLIs while maintaining efficient substrate cleavage. Several KLIs

were initially identified through a bioinformatics search of the PDB, of which two were selected as being structurally similar to the I/V KLI of the VS ribozyme: the HIV-1 TAR/TAR* KLI (PDB ID code 1KIS; [34]) and the L22/L88 KLI from the large subunit of rRNAs (PDB ID code 4IOA; [35]). These two KLIs were substituted for the I/V KLI of the S_0/R_0 reference pair formed between a pre-shifted substrate and a *trans* version of the VS ribozyme. In addition, a combinatorial approach was used in which several variant S/R pairs were synthesized that contain different helix lengths in stems Ib and V. The resulting S/R pairs were kinetically evaluated, and several show substantial catalytic activity. Below, we compare the catalytic properties of the most active variant S/R pairs with that of the previously investigated S_0/R_0 pair in light of the thermodynamic and structural properties of the parental and surrogate KLIs.

Within the three families of S/R pairs investigated (S_{TAR}/R_{TAR^*} , S_{L88}/R_{L22} and S_{L22}/R_{L88}), several S/R combinations show substantial catalytic efficiency (dark gray shading in Fig. 3.3B and in Supplementary Figs. 3.3B and 3.4B). For each family, helix-length compensation was observed for the most active S/R pairs, and this is consistent with what was observed in previous studies of the parental S_0/R_0 pair [31]. For the S_{TAR}/R_{TAR^*} family, the *trans* VS ribozyme variants can efficiently cleave an SLI substrate with stem Ib lengths of 3 to 6 bp (Fig. 3.3B). Both the S_{L88}/R_{L22} and S_{L22}/R_{L88} families also allow an appreciable level of catalytic efficiency for several SLI substrates, with a stem Ib length of 4-7 bp for S_{L22}/R_{L88} and 4-6 bp for S_{L88}/R_{L22} (Supplementary Figs. 3.3 and 3.4). Within the three families of S/R pairs investigated, the S_{TAR-1}/R_{TAR^*-0} , S_{L88+1}/R_{L22-1} and S_{L22+1}/R_{L88-1} show the highest k_{cat}/K_M values, all within 2.5-fold of each other (0.2-0.5 $\text{min}^{-1}\mu\text{M}^{-1}$). Although these S/R pairs with surrogate KLIs behave like the reference S_0/R_0 pair in some respects, their k_{cat}/K_M is reduced by 50- to 100-fold compared to S_0/R_0 . Similarly, their k_{obs} values determined under identical conditions is reduced 50- to 160-fold compared to S_0/R_0 . The KLIs most likely form and assist catalysis for the highly-active variant S/R pairs, given that several other variant S/R pairs within the same family are inactive and that disruption of the KLI in the parental S_0/R_0 pair can lead to more than a 1,000-fold decrease in k_{cat}/K_M and more than a 6,000-fold decrease in k_{obs} [31]. In addition, we verified by native gel electrophoresis and SHAPE analysis that all substrates

and ribozymes of the variant S/R pairs adopt a stable fold equivalent to S_0 and R_0 , respectively, and compatible with high catalytic efficiency. Thus, although surrogate KLIs allow high catalytic efficiency for some variant S/R pairs, it is puzzling that they do not allow the same high level of activity as that observed for the parental S_0/R_0 pair.

We have previously shown that mutations that decrease the affinity of the I/V KLI also lead to lower k_{cat}/K_M values when monitoring the *trans* cleavage reaction as done here [17,30]. For example, deletion of U700 in SLV causes a 210-fold decrease in binding affinity for SLI and 140-fold decrease in k_{cat}/K_M [17,30]. Thus, it is likely that the lower relative affinity of a surrogate KLI may affect the level of catalytic efficiency that could be attained in the variant S/R pairs. Given that the L22/L88 KLI is 16-fold less stable than the I/V KLI (Fig. 3.2C; [17]), its higher K_d could explain in part the lower k_{cat}/K_M of the S_{L88}/R_{L22} (>60-fold) and S_{L22}/R_{L88} pairs (>100-fold) compared to S_0/R_0 . However, given that the HIV-1 TAR/TAR* KLI is 40-fold more stable than the I/V KLI (Fig. 3.2C; [17]), the difference in K_d of the KLI complex cannot explain the lower k_{cat}/K_M of the S_{TAR}/R_{TAR^*} pairs (>50-fold) compared to S_0/R_0 .

The VS ribozyme S/R pairs likely adopt both an open state and a close state that can be respectively represented by the recent NMR-based model and X-ray structure [13,41]. The KLI is stably formed in both states, but only the close state captures the active site formed by the association of the G₆₃₈ and A₇₅₆ loops. In order to assess the effect of substituting the I/V KLI on formation of the VS ribozyme active site, the S_0/R_0 , S_{L88-0}/R_{L22-0} and S_{TAR-0}/R_{TAR^*-0} pairs were modeled based on high-resolution structures of KLIs and the S/R complex derived from the X-ray structure of the VS ribozyme (Fig. 3.5). Since all the modeled S/R pairs adopt an open state with the G₆₃₈ loop apart from the A₇₅₆ loop, it is difficult to precisely evaluate the compatibility of surrogate KLIs for formation of the active site. Nevertheless, we find that the distance between the scissile phosphates of the model and the crystal structure is longer for the S_{TAR-0}/R_{TAR^*-0} pair than for the S_0/R_0 and S_{L88-0}/R_{L22-0} pairs. In addition, it appears from these models that formation of the active site from the open state would involve the simple packing of two helical domains for the S_0/R_0 pair, but possibly be more restrained for the S_{L88-0}/R_{L22-0} and

S_{TAR-0}/R_{TAR*-0} pairs. Thus, the HIV-1 TAR/TAR* KLI and possibly the L22/L88 interaction, which both contain more base pairs than the I/V KLI, may partially restrict formation of the active site due to topological constraints. We envision that additional mutations of the VS ribozyme could be introduced to help compensate for such structural restriction and thereby increase the catalytic efficiency of the variant S_{L88}/R_{L22} and S_{TAR}/R_{TAR*} pairs.

Previous NMR studies of an SLI/SLV complex indicate that both the SLI and SLV loops contain several dynamic residues that could confer a certain degree of flexibility to the I/V KLI [12]. Thus, the I/V KLI likely acts as a dynamic hinge that facilitates formation of the active site. It is not clear at this time if the surrogate KLIs allow the same degree of flexibility, and how their dynamics affect formation of the active site in the variant S/R pairs. In this regard, future studies investigating the dynamic behavior of these KLIs would provide valuable insights into how these KLIs can be further engineered to more efficiently assist formation of the active site in the VS ribozyme.

CONCLUDING REMARK

Despite the ability of the VS ribozyme to specifically cleave a stem-loop substrate, very few studies had previously explored the possibility of modifying the VS ribozyme for cleavage of alternate substrates. Using a novel RNA engineering approach, we demonstrated here that the KLI within the VS ribozyme could be substituted by at least two other known KLIs while maintaining substantial cleavage activity. Interestingly, the possibility of substituting the KLI broadens the substrate specificity of VS ribozyme derivatives, which can now target SLI variants in which the 7-nt terminal loop can be replaced by at least three other 6-nt loop sequences and the length of the adjoining stem can be varied to some extent. Future engineering efforts are now geared towards improving the cleavage efficiency of the VS ribozyme variants identified here as well as further broadening the substrate specificity of VS ribozyme derivatives.

SUPPLEMENTARY DATA

Supplementary data are available at NAR online: Supplementary Tables III.I and III.II as well as Supplementary Figures 3.1, 3.2, 3.3, 3.4, 3.5 and 3.6.

FUNDING

Canadian Institutes of Health Research (CIHR MOP-86502 to P.L.); CIHR and Université de Montréal Ph.D. scholarships (to J.L.-L.); FRQNT Ph.D. scholarship (to P.D.). Funding for open access charge: CIHR.

Conflict of interest statement. None declared.

The authors declare no competing financial interests.

ACKNOWLEDGEMENTS

The authors thank Dominique Chaussé for RNA preparation, Ryan Richter for computer support, as well as James G. Omichinski for discussions and critical reading of the manuscript. They also thank Sébastien Lemieux and Calcul Québec for providing computing facilities.

REFERENCES

1. Lilley, D.M., *Catalysis by the nucleolytic ribozymes*. Biochem Soc Trans, 2011. **39**(2): p. 641-6.
2. Jimenez, R.M., J.A. Polanco, and A. Luptak, *Chemistry and biology of self-cleaving ribozymes*. Trends Biochem Sci, 2015.
3. Weinberg, Z., et al., *New classes of self-cleaving ribozymes revealed by comparative genomics analysis*. Nat Chem Biol, 2015. **11**(8): p. 606-10.
4. Harris, K.A., et al., *Biochemical analysis of pistol self-cleaving ribozymes*. RNA, 2015. **21**(11): p. 1852-8.
5. Li, S., et al., *Biochemical analysis of hatchet self-cleaving ribozymes*. RNA, 2015. **21**(11): p. 1845-51.
6. Herschlag, D., *Learning from ribozymes*. RNA, 2015. **21**(4): p. 527-8.
7. Collins, R.A., *The Neurospora Varkud satellite ribozyme*. Bioch. Soc. Trans. Rev., 2002. **30**: p. 1122-1126.
8. Lilley, D.M., *The Varkud satellite ribozyme*. RNA, 2004. **10**: p. 151-158.
9. Wilson, T.J. and D.M. Lilley, *A mechanistic comparison of the Varkud satellite and hairpin ribozymes*. Prog Mol Biol Transl Sci, 2013. **120**: p. 93-121.
10. Guo, H.C.T. and R.A. Collins, *Efficient trans-cleavage of a stem-loop RNA substrate by a ribozyme derived from Neurospora VS RNA*. EMBO J., 1995. **14**(12): p. 368-376.
11. Rastogi, T., et al., *A long-range pseudoknot is required for activity of the Neurospora VS ribozyme*. EMBO J., 1996. **15**(11): p. 2820-2825.
12. Bouchard, P. and P. Legault, *Structural insights into substrate recognition by the Neurospora Varkud satellite ribozyme: importance of U-turns at the kissing-loop junction*. Biochemistry, 2014. **53**(1): p. 258-69.
13. Suslov, N.B., et al., *Crystal structure of the Varkud satellite ribozyme*. Nat Chem Biol, 2015. **11**(11): p. 840-6.
14. Andersen, A. and R.A. Collins, *Rearrangement of a stable RNA secondary structure during VS ribozyme catalysis*. Mol. Cell, 2000. **5**: p. 469-478.
15. Andersen, A.A. and R.A. Collins, *Intramolecular secondary structure rearrangement by the kissing interaction of the Neurospora VS ribozyme*. Proc. Natl. Acad. Sci USA, 2001. **98**(14): p. 7730-7735.
16. Zamel, R. and R.A. Collins, *Rearrangement of substrate secondary structure facilitates binding to the Neurospora VS ribozyme*. J Mol Biol, 2002. **324**: p. 903-915.
17. Bouchard, P. and P. Legault, *A remarkably stable kissing-loop interaction defines substrate recognition by the Neurospora VS Ribozyme*. RNA, 2014. **20**(9): p. 1451-64.
18. Hiley, S.L. and R.A. Collins, *Rapid formation of a solvent-inaccessible core in the Neurospora Varkud satellite ribozyme*. EMBO J., 2001. **20**(19): p. 5461-5469.
19. Jones, F.D. and S.A. Strobel, *Ionization of a critical adenosine residue in the Neurospora Varkud satellite ribozyme active site*. Biochemistry, 2003. **42**: p. 4265-4276.
20. McLeod, A.C. and D.M. Lilley, *Efficient, pH-dependent RNA ligation by the VS ribozyme in trans*. Biochemistry, 2004. **43**(4): p. 1118-25.
21. Zhao, Z.Y., et al., *Nucleobase participation in ribozyme catalysis*. J. Am. Chem. Soc., 2005. **127**(14): p. 5026-7.

22. Smith, M.D. and R.A. Collins, *Evidence for proton transfer in the rate-limiting step of a fast-cleaving Varkud satellite ribozyme*. Proc. Natl. Acad. Sci. USA, 2007. **104**(14): p. 5818-23.
23. Wilson, T.J., A.C. McLeod, and D.M. Lilley, *A guanine nucleobase important for catalysis by the VS ribozyme*. EMBO J., 2007. **26**(10): p. 2489-500.
24. Jaikaran, D., et al., *An important role of G638 in the cis-cleavage reaction of the Neurospora VS ribozyme revealed by a novel nucleotide analog incorporation method*. RNA, 2008. **14**(5): p. 938-49.
25. Smith, M.D., et al., *The ionic environment determines ribozyme cleavage rate by modulation of nucleobase pKa*. RNA, 2008. **14**(9): p. 1942-9.
26. Wilson, T.J., et al., *Nucleobase-mediated general acid-base catalysis in the Varkud satellite ribozyme*. Proc. Natl. Acad. Sci. USA, 2010. **107**(26): p. 11751-6.
27. Guo, H.C.T., et al., *Nucleotide sequence requirements for self-cleavage of Neurospora VS RNA*. J. Mol. Biol., 1993. **232**: p. 351-361.
28. Beattie, T.L., J.E. Olive, and R.A. Collins, *A secondary-structure model for the self-cleaving region of Neurospora VS RNA*. Proc. Natl. Acad. Sci. USA, 1995. **92**: p. 4686-4690.
29. Rastogi, T. and R.A. Collins, *Smaller, faster ribozymes reveal the catalytic core of Neurospora VS RNA*. J. Mol. Biol., 1998. **277**: p. 215-224.
30. Bouchard, P., et al., *Role of SLV in SLI substrate recognition by the Neurospora VS ribozyme*. RNA, 2008. **14**: p. 736-748.
31. Lacroix-Labonté, J., et al., *Helix-length compensation studies reveal the adaptability of the VS ribozyme architecture*. Nucleic Acids Res, 2012. **40**(5): p. 2284-93.
32. Petrov, A.I., C.L. Zirbel, and N.B. Leontis, *WebFR3D--a server for finding, aligning and analyzing recurrent RNA 3D motifs*. Nucleic Acids Res., 2011. **39**(Web Server issue): p. W50-5.
33. Merino, E.J., et al., *RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE)*. J Am Chem Soc, 2005. **127**(12): p. 4223-31.
34. Chang, K.Y. and I. Tinoco, Jr., *The structure of an RNA "kissing" hairpin complex of the HIV TAR hairpin loop and its complement*. J Mol Biol, 1997. **269**(1): p. 52-66.
35. Magee, T.V., et al., *Novel 3-O-carbamoyl erythromycin A derivatives (carbamolides) with activity against resistant staphylococcal and streptococcal isolates*. Bioorg Med Chem Lett, 2013. **23**(6): p. 1727-31.
36. Major, F., et al., *The combination of symbolic and numerical computation for three-dimensional modeling of RNA*. Science, 1991. **253**(5025): p. 1255-60.
37. Eissenthal, R., M.J. Danson, and D.W. Hough, *Catalytic efficiency and k_{cat}/K_M : a useful comparator?* Trends Biotechnol, 2007. **25**(6): p. 247-9.
38. Campbell, D.O., et al., *NMR structure of Varkud satellite ribozyme stem-loop V in the presence of magnesium ions and localization of metal-binding sites*. Biochemistry, 2006. **45**(35): p. 10591-605.
39. Desjardins, G., et al., *NMR structure of the A730 loop of the Neurospora VS ribozyme: insights into the formation of the active site*. Nucleic Acids Res, 2011. **39**(10): p. 4427-37.
40. Bonneau, E. and P. Legault, *Nuclear magnetic resonance structure of the III-IV-V three-way junction from the Varkud satellite ribozyme and identification of*

- magnesium-binding sites using paramagnetic relaxation enhancement.* *Biochemistry*, 2014. **53**(39): p. 6264-75.
41. Bonneau, E., et al., *The NMR structure of the II-III-VI three-way junction from the Neurospora VS ribozyme reveals a critical tertiary interaction and provides new insights into the global ribozyme structure.* *RNA*, 2015. **21**(9): p. 1621-32.
 42. Wilkinson, K.A., et al., *Influence of nucleotide identity on ribose 2'-hydroxyl reactivity in RNA.* *RNA*, 2009. **15**(7): p. 1314-21.
 43. McGinnis, J.L., et al., *The mechanisms of RNA SHAPE chemistry.* *J Am Chem Soc*, 2012. **134**(15): p. 6617-24.
 44. Leontis, N.B. and E. Westhof, *Geometric nomenclature and classification of RNA base pairs.* *RNA*, 2001. **7**(4): p. 499-512.
 45. Lavery, R., et al., *Conformational analysis of nucleic acids revisited: Curves+.* *Nucleic Acids Res.*, 2009. **37**(17): p. 5917-29.

TABLES

Table III.I. Filtered results of the bioinformatics search to identify KLIs from the PDB.

Interaction Class	Organism	PDB codes	Experimental method (Resolution in Å)	Structural characteristics of KLI ^a		
				WC Pairs	Size of loops	Inter-helical Angle
HIV-1 TAR/TAR*	Synthetic	2JLT	X-ray (2.90)	6	6 / 8	160°
		2RN1	NMR	6	6 / 8	125° ± 0.7°
		1KIS	NMR	6	6 / 6	136°
		2PN9 ^b	NMR	6	6 / 8	155° ± 5°
L6/L7 of rRNA large subunit	<i>Thermus thermophilus</i>	3V2F	X-ray (2.70)	4	8 / 7	94°
	<i>Escherichia coli</i>	2QBG	X-ray (3.30)	4	8 / 7	93°
	<i>Deinococcus radiodurans</i>	4IOA	X-ray (3.20)	4	8 / 7	n/a
	<i>Haloarcula marismortui</i>	1S72	X-ray (2.40)	4	8 / 7	78°
L22/L88 of rRNA large subunit	<i>Thermus thermophilus</i>	3V2F	X-ray (2.70)	4	6 / 8	137°
	<i>Escherichia coli</i>	2QBG	X-ray (3.30)	4	6 / 8	146°
	<i>Deinococcus radiodurans</i>	4IOA	X-ray (3.20)	4	6 / 8	152°
	<i>Haloarcula marismortui</i>	1S72	X-ray (2.40)	6	7 / 7	154°
	<i>Tetrahymena thermophila</i>	4A1B	X-ray (3.52)	6	8 / 7	139°
	<i>Saccharomyces cerevisiae</i>	3U5H	X-ray (3.00)	6	8 / 7	154°
L5/L13 of cobalamin riboswitch	Marine Metagenome	4FRN	X-ray (3.43)	5	6 / 8	n/a ^a
L3/L4 of SRP Alu domain	<i>Homo sapiens</i>	1E8O	X-ray (3.20)	3	6 / 8	68°
ColE1 RNAI/RNAII	Synthetic	2BJ2	NMR	7	7 / 7	119°

a. Each KL structure was analyzed with the PyMol Molecular Graphics System (Version 1.3, Schrödinger) to identify the number of *cis* WC/WC base pair and size of loops at the KLI and with Curves+ [46] to measure the interhelical angle between the stems connecting a specific KLI. Interhelical angles were measured only for KL where the adjoining stems contain a minimum of three base pairs and are not available (n/a) for the other KLIs.

b. Contains modified nucleotides at the KLI.

FIGURES

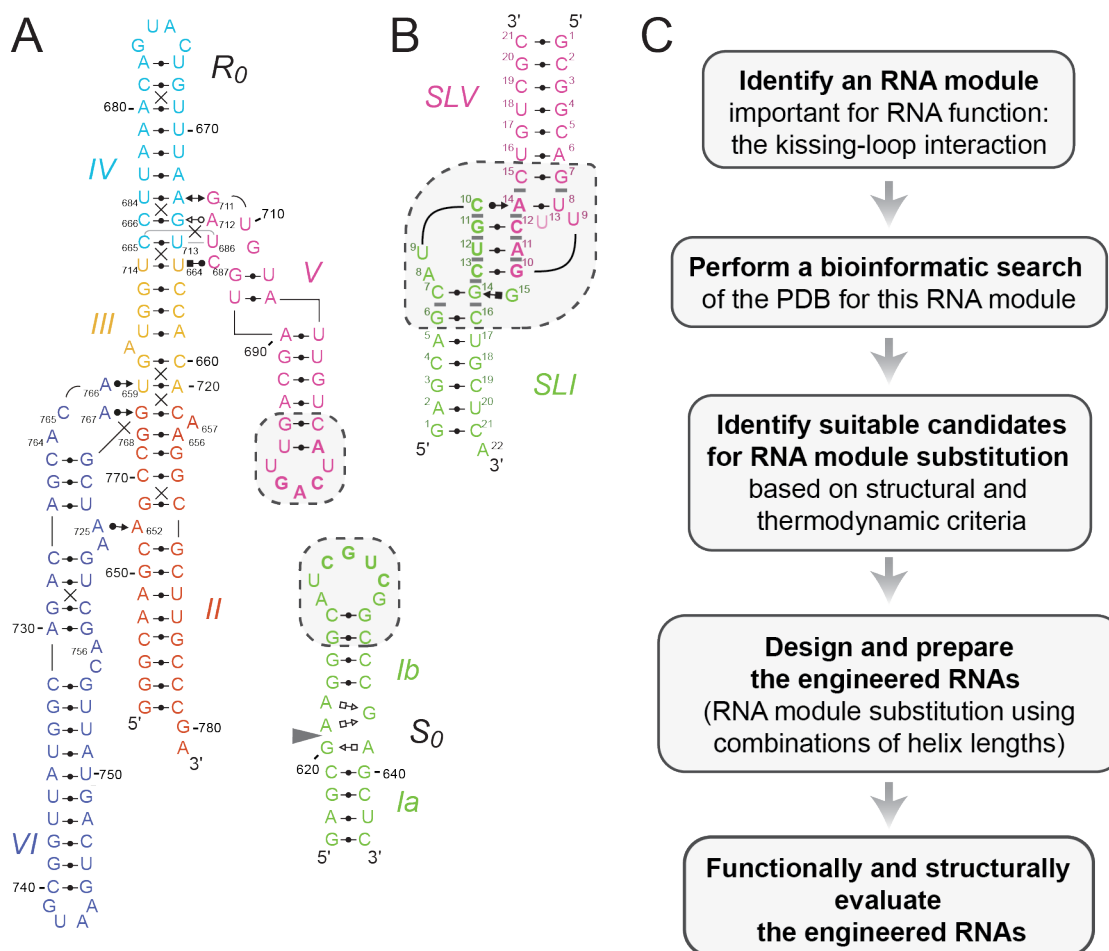


Figure 3.1 Rational engineering of KLIs for substrate recognition by the *Neurospora* VS ribozyme.

(A) Structural schematic of a *trans* VS ribozyme system formed by the S_0 substrate and the R_0 ribozyme (31). The schematic is derived from NMR and X-ray structures (13,41). The cleavage site is located between G620 and A621 as indicated by the gray arrowhead. (B) Structural characteristics of the VS ribozyme I/V KLI derived from the NMR structure of the SLI/SLV complex [PDB ID code 2MI0; (12)]. (C) Flowchart of the engineering strategy used in this study. In (A) and (B), the dashed boxes enclose residues that contribute to the KLI and the closing base pairs of the two stems, whereas base-pairing interactions are represented using the Leontis-Westhof notation (44). In addition, base-pairing residues at the KLI are in bold. In (B), gray bars represent base stacking at the KLI.

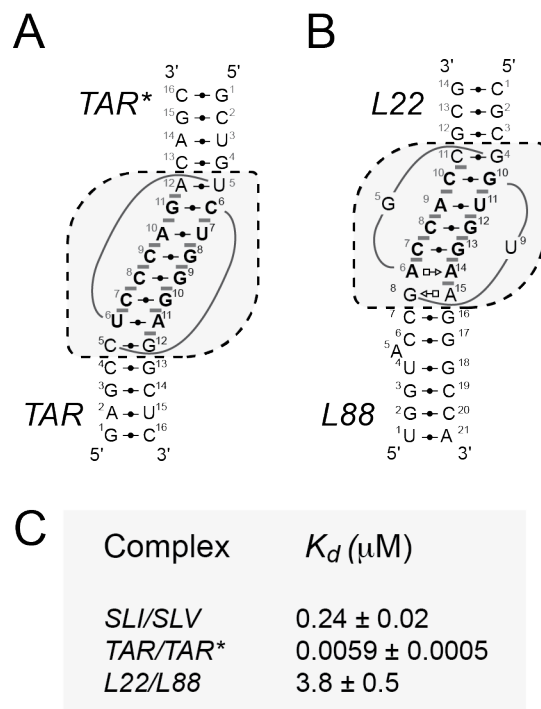


Figure 3.2 Schematics and thermodynamic stability of selected KLIs.

Structural characteristics of the selected KLIs: (A) the HIV-1 TAR/TAR* KLI [PDB ID code 1KIS; (34)] and (B) the large ribosomal subunit L22/L88 KLI from *Deinococcus radiodurans* [PDB ID code 4IOA; (35)]. (C) Dissociation constants (K_d) of the selected KLIs determined in a previous study for the SLI/SLV and TAR/TAR* complexes (17) and here for the L22/L88 complex (Supplementary Fig. 3.2). These K_d values were all determined at 25°C in 10 mM Tris pH 7.0 and 20 mM MgCl₂. In (A) and (B), the dashed boxes enclose residues of the KLI and the closing base pairs of the two stems and base-pairing interactions are represented using the Leontis-Westhof notation (44). In addition, base-pairing residues at the KLI are in bold and gray bars highlight base stacking at the KLI.

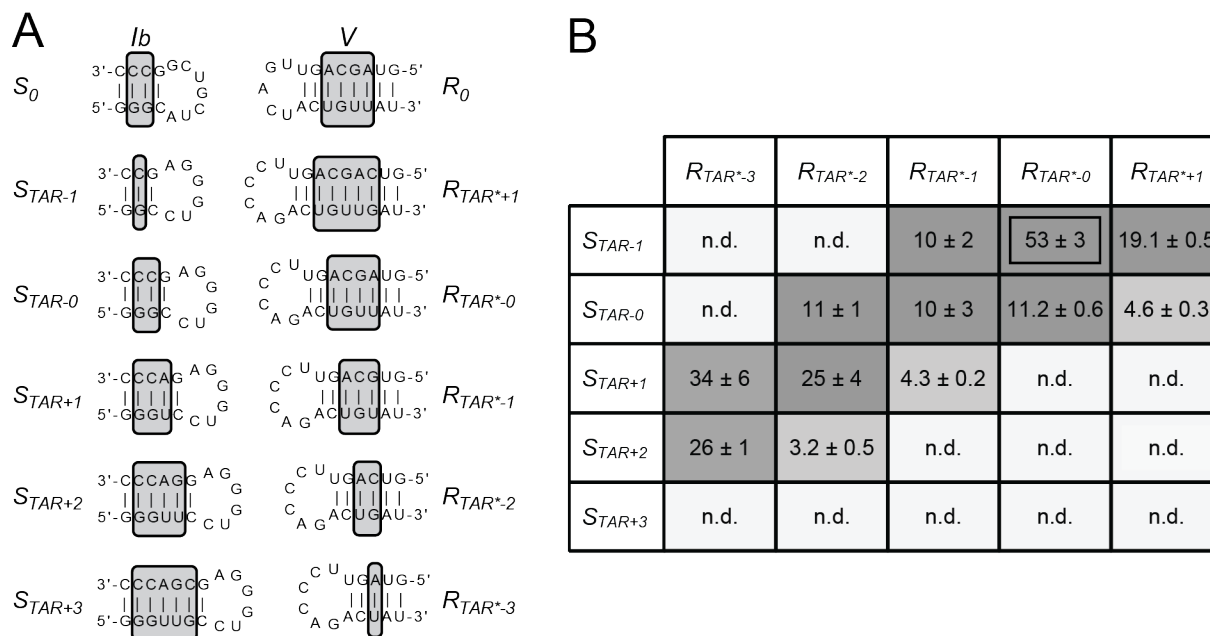


Figure 3.3 Sequences and kinetic data of the S_{TAR}/R_{TAR^*} variants.

(A) Sequences of the S_{TAR}/R_{TAR^*} variants. Only the regions of the S_0 substrate and R_0 ribozyme (stem-loops Ib and V shown above; Fig. 3.1A) that were substituted in the S_{TAR} and R_{TAR^*} variants are shown. The number of base pairs varies in stems Ib and V, as indicated by the gray boxes on the putative secondary structures. (B) Grayscale heat map of the k_{cat}/K_M values ($10^{-2} \text{ min}^{-1} \mu\text{M}^{-1}$) for the S_{TAR}/R_{TAR^*} variants. Dark gray shading corresponds to k_{cat}/K_M values within 5 fold of the most active S_{TAR}/R_{TAR^*} pair ($S_{TAR^{-1}}/R_{TAR^{-0}}$); medium gray shading corresponds to lower k_{cat}/K_M values (5-17 fold lower than the most active S_{TAR}/R_{TAR^*} pair), and light gray shading reflects the absence of detectable cleavage activity (n.d.: $k_{obs} \leq 0.002 \text{ min}^{-1}$) after 24 min at $[R] \leq 15 \mu\text{M}$.

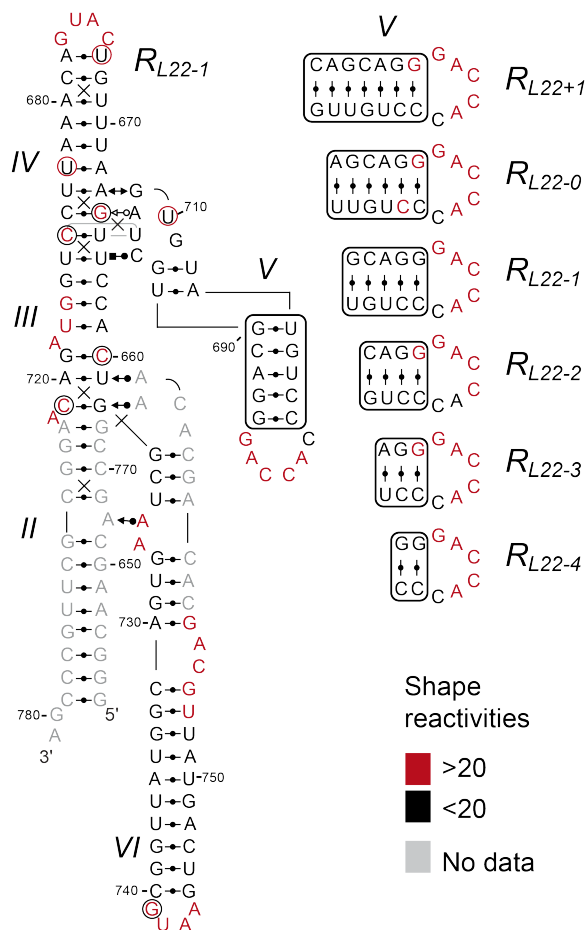


Figure 3.4 SHAPE analysis of the R_{TAR^*} ribozyme variants.

The normalized NMIA reactivity of each nucleotide within R_{TAR^*-0} is color-coded on its structural schematic as per the SHAPE reactivities key. For the other R_{TAR^*} variants, the normalized NMIA reactivity is within the same category as for R_{TAR^*-0} , except for a few nucleotides in the variable SLV region (right panel) and outside this region (black circles indicate inverted reactivity compared to R_{TAR^*-0} ; U₆₇₅ for R_{TAR-1} and G₇₄₁, U₇₄₆ and G₇₄₉ for R_{TAR-3}).

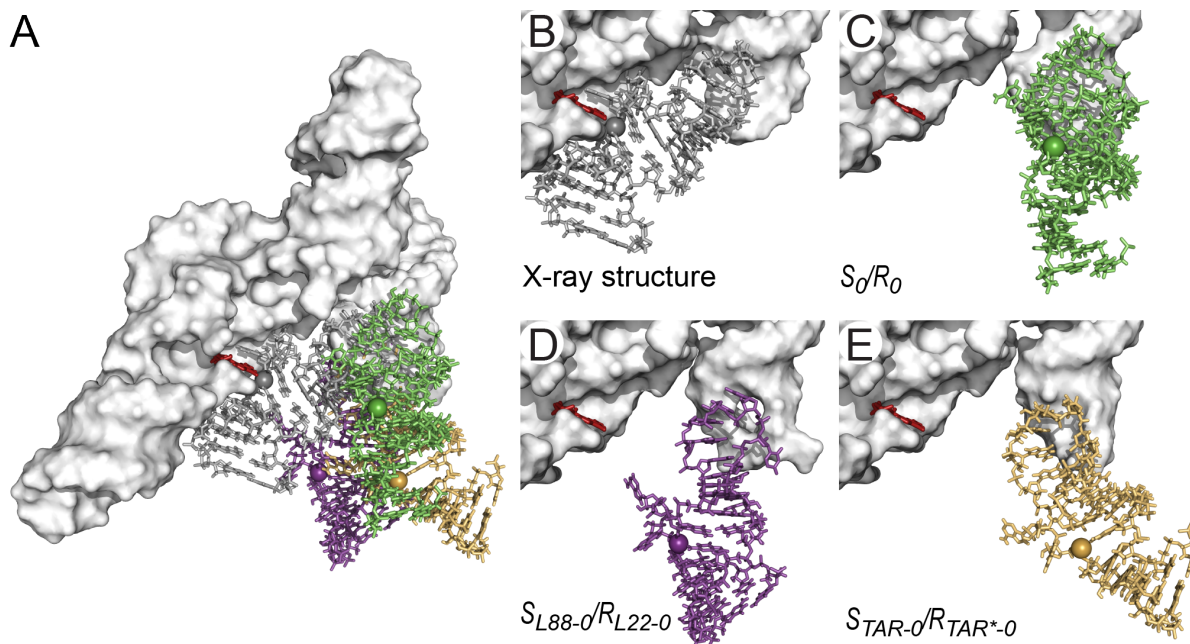


Figure 3.5 Three-dimensional models of substrate/ribozyme complexes.

(A) Superposition of three-dimensional models of the variant *S/R* pairs on the X-ray structure of the VS ribozyme. Individual representations from this superposition are zoomed in and described in details for the subsequent panels. (B) X-ray structure of the VS ribozyme [PDB ID code 4R4P; (13)]. In this representation, the substrate of one protomer is shown in dark grey sticks with the phosphorus at the scissile phosphate as a sphere and the *trans* VS ribozyme (helical domains II-VI) of the other protomer is shown as a white surface, except for residue A₇₅₆, which is shown as red sticks. (C) Model of the S₀/R₀ pair derived from the NMR structure of a VS ribozyme SLI/SLV complex [lime green; PDB ID code 2MI0; (12)]. (D) Model of the S_{L88-0}/R_{L22-0} pair derived from the X-ray structure of the ribosomal L22/L88 KLI [purple; PDB ID code 4IOA; (35)]. (E) Model of the S_{TAR-0}/R_{TAR*-0} pair derived from the NMR structure of the TAR/TAR* complex [mustard yellow; PDB ID code 1KIS; (34)]. In (A-E), the *trans* VS ribozyme is shown as a white surface with A₇₅₆ in red sticks and the individual substrates are shown as sticks with the phosphorus at the scissile phosphate as a sphere.

SUPPLEMENTARY INFORMATION

Supplementary Table III.I

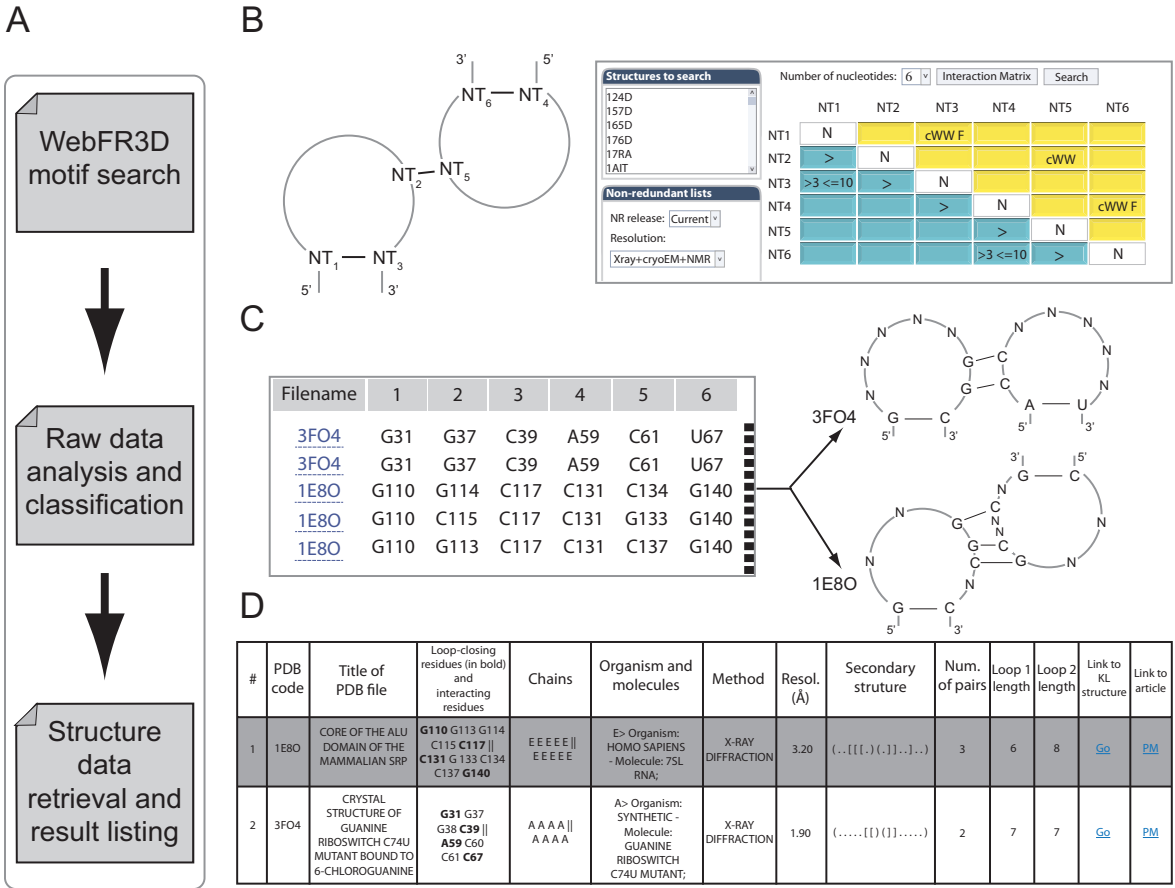
Note: Supplementary Table III.I is provided as a separate java/html file.

Supplementary Table III.II. Values of k_{obs} (min^{-1}) for selected S/R pairs measured with $[S] = 0.250 \text{ nM}$ and $[R] = 0.5 \text{ }\mu\text{M}$ under standard conditions.^a

	k_{obs} (min^{-1})	Relative decrease in k_{obs} value	n^b
S_0/R_0	12.2 ± 0.5	1	3
S_{TAR-1}/R_{TAR*-0}	0.0769 ± 0.0003	159	2
S_{L88+1}/R_{L22-1}	0.23 ± 0.02	53	3
S_{L22+1}/R_{L88-1}	0.145	84	1

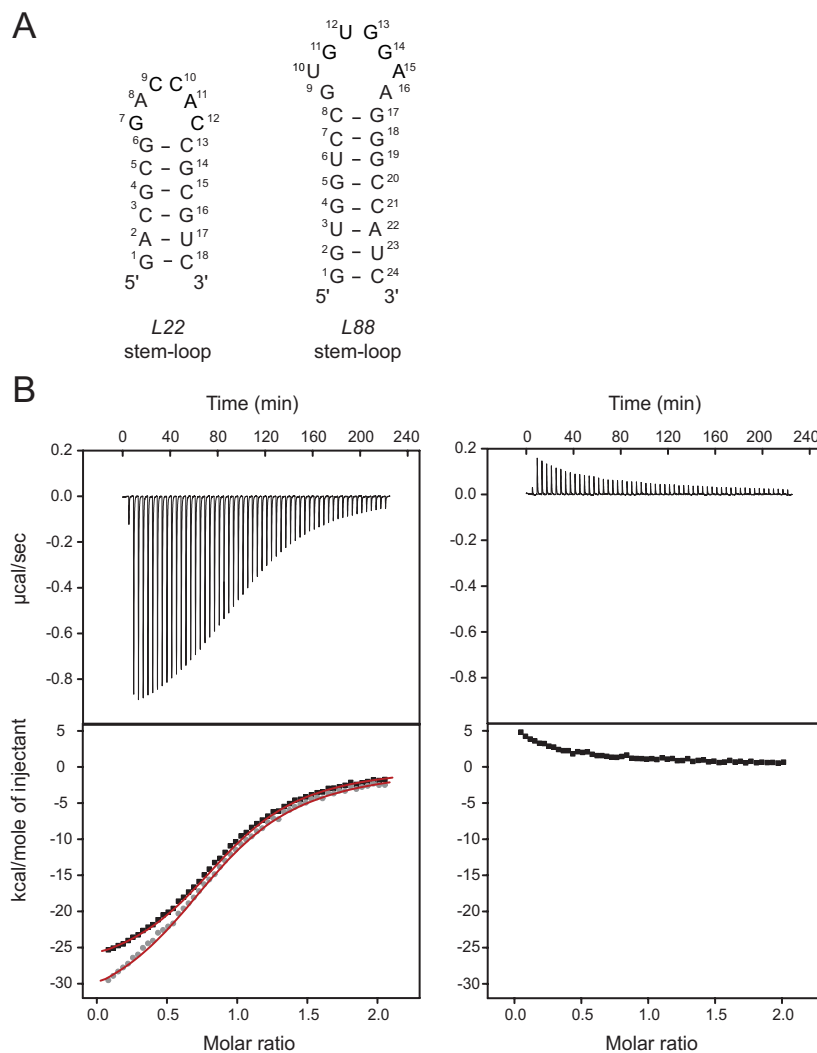
a. At 37°C in Kinetic Reaction buffer (50 mM Tris at pH 7.6, 25 mM KCl, 2 mM spermidine and 25 mM MgCl_2).

b. The value of n represents the number of replicates.



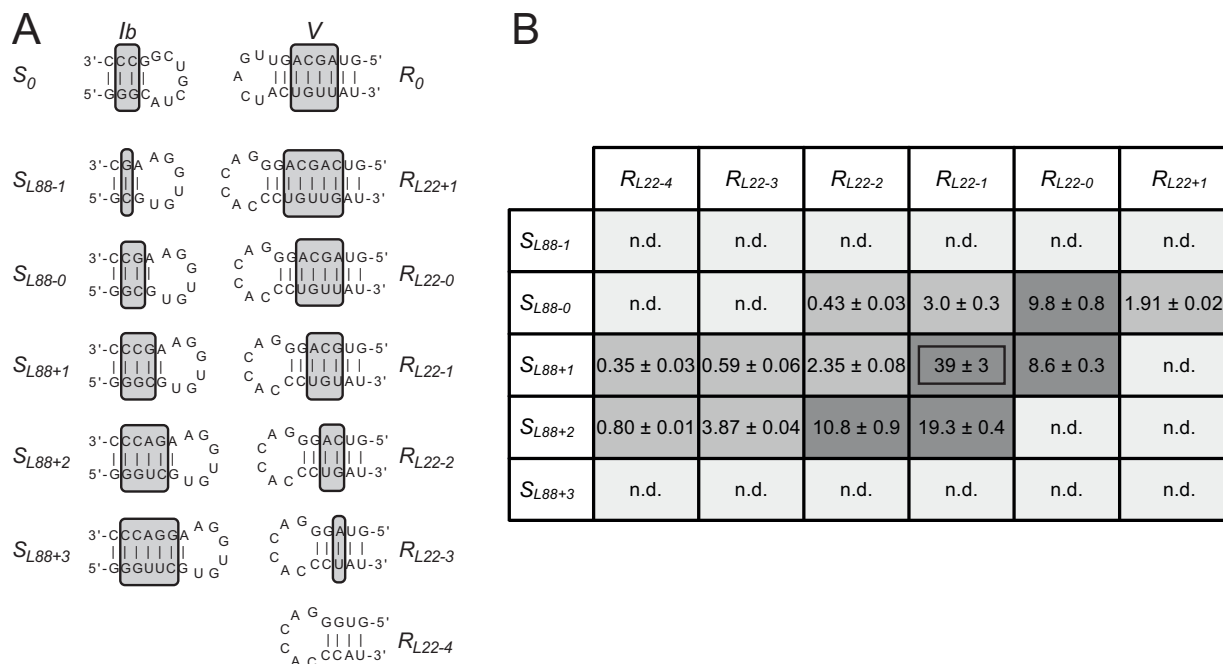
Supplementary Figure 3.1 Schematic representation of the computational workflow used to identify and characterize KLIs present in the PDB.

(A) The workflow contains three steps. (B) The first step is a motif search in WebFR3D. A secondary structure view of the motif to be queried is shown alongside the search input for WebFR3D. (C) The second step analyzes the raw data from the WebFR3D output and classifies the data to provide a list of the KLIs. For example, the two individual entries related to PDB file 3FO4 provide information about the same KLI because they report the same stem-loop closing base pairs (involving residues in columns 1, 3, 4 and 6), but different base pairs at the kissing-loop interface (involving residues listed in columns 2 and 5). (D) The final step involves retrieving additional data for each KLI and listing the results in an HTML/JavaScript table. Each KLI is listed with its index number, the PDB code of the structure in which it was found, the title of the PDB file, the residue numbers of the loop-closing and the interacting nucleotides between the two loops, the identifier(s) of the chain(s) in which these nucleotides were found, the organism and type of molecule involved, the method used for structure determination, the experimental resolution (n.a. = not available), the secondary structure in dot-bracket notation, the number of WC/WC base pairs at the kissing-loop interface, the sizes of the two loops, a hyperlink to a PDB file containing the isolated interaction, and a hyperlink to the pubmed reference.



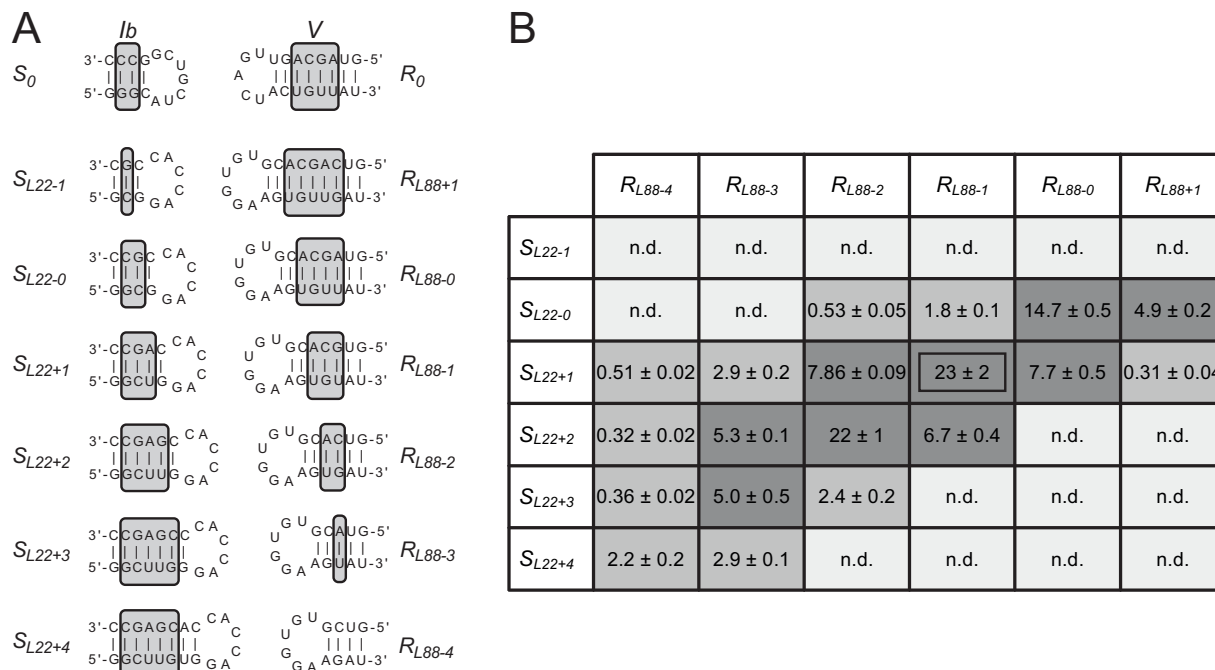
Supplementary Figure 3.2 ITC studies of the L22/L88 interaction.

(A) Sequence and proposed secondary structures of the L22 and L88 stem-loops derived from the L22 and L88 secondary structure elements of the 23S rRNA of *Deinococcus radiodurans* (1,2). (B) ITC data analysis for binding of the L22 stem-loop to the L88 stem-loop at 25°C in a buffer containing 10 mM Tris pH 7.0, 20 mM MgCl₂ and 0.05 mM NaN₃. The raw data of a representative titration experiment is shown on the top left panel and the corresponding dilution control (only buffer in the cell) is shown on the top right panel. The isothermal curves representing the integrated data of the titration experiments are shown below. On the left bottom panel, two isothermal curves of the titration experiment are shown, before (black) and after (gray) subtraction of the control experiment (right bottom panel). For the titration experiment, 200 µM L88 was used in the syringe and 20 µM L22 in the cell. The isothermal curve with subtraction of the control experiment (gray) was used to extract the thermodynamic parameters for this interaction: $K_d = 3.8 \pm 0.5 \mu\text{M}$; $N = 0.875 \pm 0.004$; $\Delta H = -35.2 \pm 0.3 \text{ kcal/mol}$; $\Delta S = -0.0933 \pm 0.0004 \text{ kcal/mol}$ and $\Delta G_{ITC} = -7.41 \pm 0.08 \text{ kcal/mol}$ ($\Delta G_{ITC} = -RT \ln(K_d)$ at $T = 298.15 \text{ K}$).



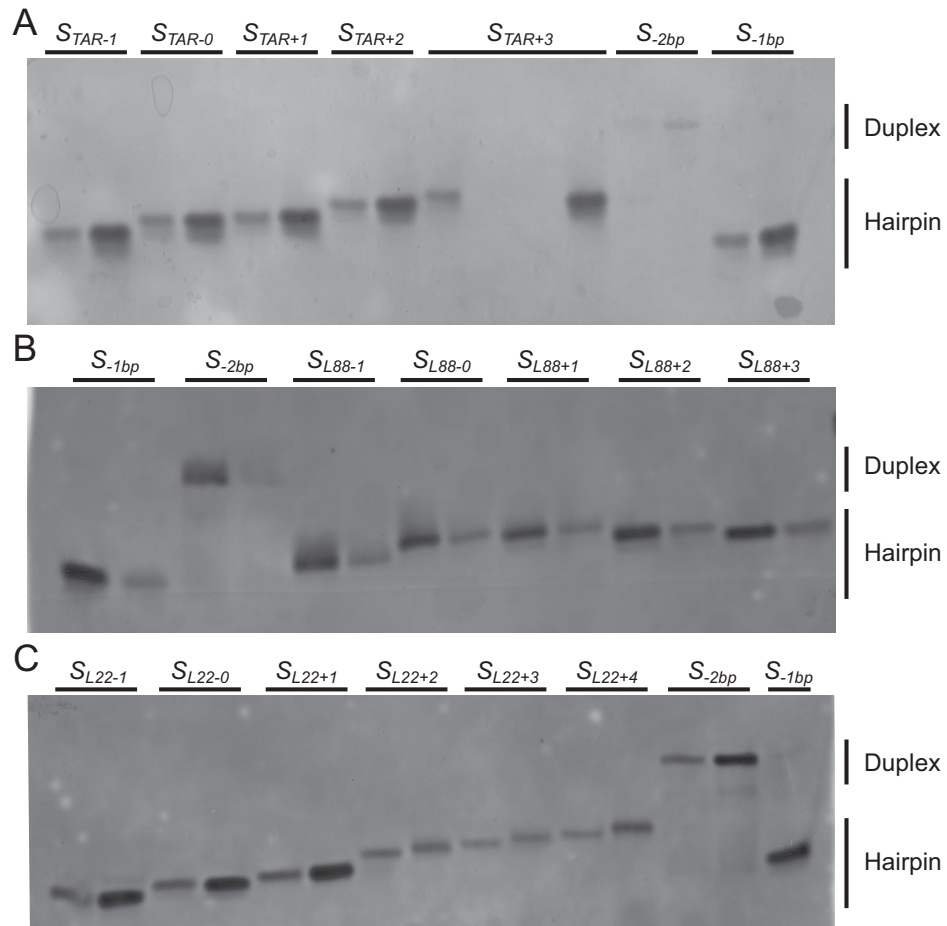
Supplementary Figure 3.3 Sequences and kinetic data of the S_{L88}/R_{L22} variants.

(A) Sequences of the S_{L88}/R_{L22} variants. Only the regions of the S_0 substrate and R_0 ribozyme (stem-loops Ib and V shown above; Fig. 3.1A) that were substituted in the S_{L88} and R_{L22} variants are shown. The number of base pairs varies in stems Ib and V, as indicated by the gray boxes on the putative secondary structures. (B) Grayscale heat map of the k_{cat}/K_M values ($10^{-2} \text{ min}^{-1} \mu\text{M}^{-1}$) for the S_{L88}/R_{L22} variants. Dark gray shading corresponds to k_{cat}/K_M values within 5 fold of the most active S_{L88}/R_{L22} pair (S_{L88+1}/R_{L22-1}); medium gray shading corresponds to lower k_{cat}/K_M values (5-111 fold lower than the most active S_{L88}/R_{L22} pair), and light gray shading reflects the absence of detectable cleavage activity (n.d.: $k_{obs} \leq 0.002 \text{ min}^{-1}$) after 24 min at $[R] \leq 15 \mu\text{M}$.



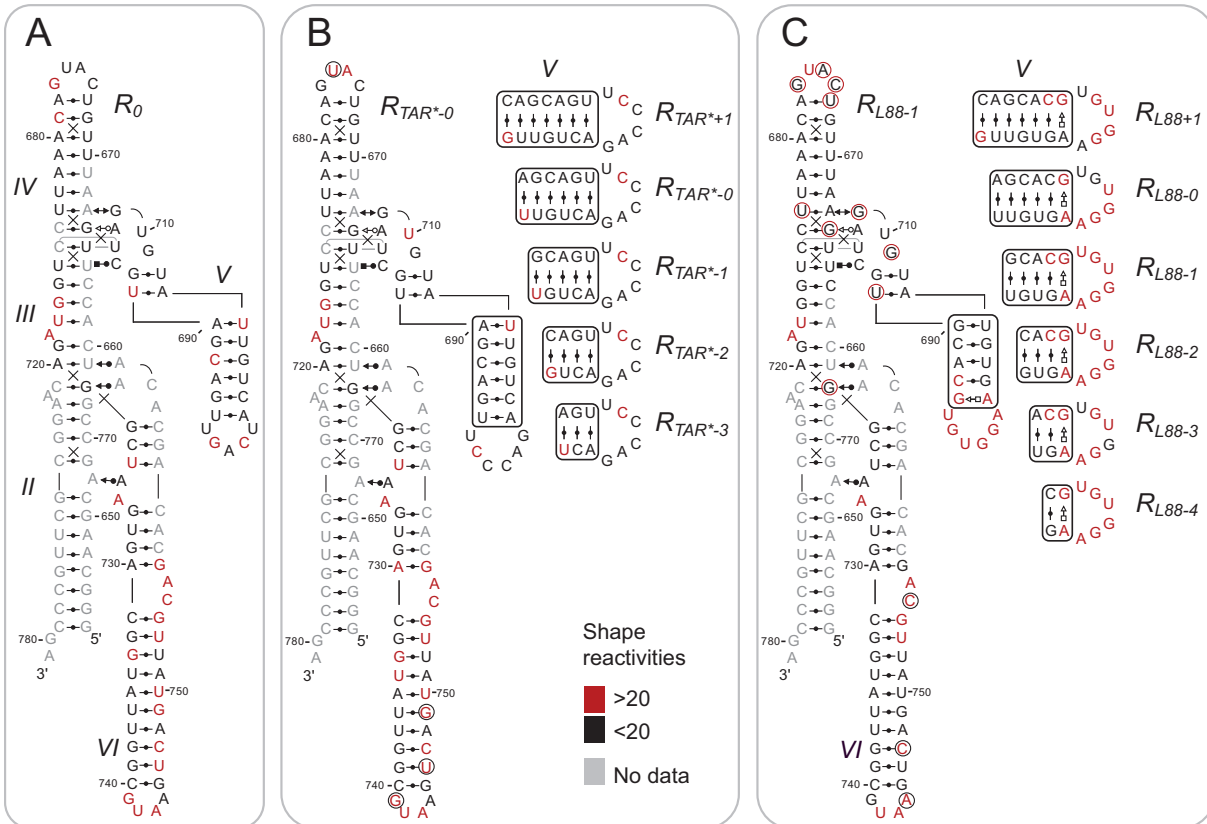
Supplementary Figure 3.4 Sequences and kinetic data of the S_{L22}/R_{L88} variants.

(A) Sequences of the S_{L22}/R_{L88} variants. Only the regions of the S_0 substrate and R_0 ribozyme (stem-loops Ib and V shown above; Fig. 3.1A) that were substituted in the S_{L22} and R_{L88} variants are shown. The number of base pairs varies in stems Ib and V, as indicated by the gray boxes on the putative secondary structures. (B) Grayscale heat map of the k_{cat}/K_M values ($10^{-2} \text{ min}^{-1}\mu\text{M}^{-1}$) for the S_{L22}/R_{L88} variants. Dark gray shading corresponds to k_{cat}/K_M values within 5 fold of the most active S_{L22}/R_{L88} pair (S_{L22+1}/R_{L88-1}); medium gray shading corresponds to lower k_{cat}/K_M values (5-75 fold lower than the most active S_{L22}/R_{L88} pair), and light gray shading reflects the absence of detectable cleavage activity (n.d.: $k_{obs} \leq 0.002 \text{ min}^{-1}$) after 24 min at $[R] \leq 15 \mu\text{M}$.



Supplementary Figure 3.5 Conformational analysis of the substrate variants used in this study.

The conformations of (A) the S_{TAR} substrates, (B) the S_{L88} substrates and (C) the S_{L22} substrates are analyzed by native gel electrophoresis. For each substrate, two quantities of RNA (1 and 4 μ g) were loaded on the gel. The hairpin and duplex controls are respectively the S_{1bp} and S_{2bp} substrates characterized in a previous study (3).



Supplementary Figure 3.6 SHAPE analysis of the R_0 ribozyme and related variants.

The normalized NMIA reactivity of each nucleotide is color-coded on the structural schematics of the (A) R_0 ribozyme, (B) R_{L22-1} variant, and (C) R_{L88-1} variant as per the SHAPE reactivities key. In (B), the normalized NMIA reactivity of the other R_{L22} variants is within the same category as for R_{L22-1} , except for a few nucleotides in the variable SLV region (right panel) and outside this region (red and black circles indicate inverted reactivity compared to R_{L22-1} ; C₆₅₈ for R_{L22+1} , R_{L22-2} , R_{L22-3} and R_{L22-4} , C₆₆₀ for R_{L22+1} , R_{L22-0} , R_{L22-3} and R_{L22-4} , C₆₆₅ for R_{L22+1} and R_{L22-3} , U₆₆₉ for R_{L22-2} , U₆₇₈ for R_{L22+1} , U₆₈₅ for R_{L22-3} , U₇₁₀ for R_{L22+1} and G₇₄₁ for R_{L22-0}). In (C), the normalized NMIA reactivity of the other R_{L88} variants is within the same category as for R_{L88-1} , except for a few nucleotides in the variable SLV region (right panel) and outside this region (red and black circles indicate inverted reactivity compared to R_{L88-1} ; G₆₇₄ and U₆₇₆ for R_{L88-4} , C₆₇₇ for R_{L88+1} , U₆₇₈ for R_{L88-2} , R_{L88-3} and R_{L88-4} , U₆₈₄ for R_{L88-4} , G₆₈₅ for R_{L88+1} and R_{L88-0} , U₆₈₉ for R_{L88-4} , G₇₀₉ for R_{L88+1} , G₇₂₁ for R_{L88+1} , A₇₄₄ for R_{L88-4} , C₇₄₇ for R_{L88-3} , R_{L88-4} , and C₇₅₅ for R_{L88+1}).

4.0 Études de SELEX pour la reconnaissance de l'ARN TAR du VIH-1 par le ribozyme VS

Mise en contexte

Au chapitre précédent, l'interaction *kissing-loop* I/V a été modifiée par des interactions *kissing-loop* sélectionnées dans des structures d'ARN et une activité de clivage est observée, mais elle est réduite comparativement à S_0/R_0 . Pour poursuivre dans la même voie que le chapitre précédent, l'objectif de ce chapitre est de modifier le ribozyme pour la reconnaissance et le clivage d'un substrat avec une boucle différente de son substrat naturel, mais cette fois-ci avec une activité comparable à celle de S_0/R_0 . Pour réaliser cet objectif, le SELEX a été utilisé dans l'optique d'identifier, parmi une vaste librairie de ribozymes, des séquences de ribozymes mutants avec un clivage efficace du substrat mutant. Pour ce dernier article, les nucléotides de la boucle V ont été changés aléatoirement par SELEX dans le but d'identifier un ribozyme mutant pour la reconnaissance et le clivage d'un substrat S_{TAR} comportant la boucle et les hélices Ia et Ib de l'ARN TAR du VIH-1. L'ARN TAR est un ARN de 59 nucléotides qui est essentiel à la transcription du génome viral du VIH-1.

Pour effectuer les expériences de SELEX, un ribozyme en deux parties a été utilisé. La réaction de clivage est amorcée par l'ajout de magnésium, après avoir préalablement associé les deux parties du ribozyme ensemble. Quatre cycles de SELEX ont été nécessaires pour obtenir le groupe de ribozymes le plus performant. De ce groupe, 21 séquences ont été retenues pour des analyses plus approfondies. Les séquences ont été transposées dans un système de ribozyme en *trans* avec le substrat SLI séparé du reste du ribozyme similaire à celui utilisé pour les études rapportées aux chapitres 2.0 et 3.0. Des cinétiques enzymatiques en *single-turnover* ont été effectuées pour déterminer les séquences de la boucle V qui permettent la meilleure activité de clivage. Les ribozymes sélectionnés ont été classés en 3 catégories, selon leur niveau d'activité. Le ribozyme R_{TAR14} est le meilleur candidat avec une activité similaire à celle du ribozyme R_0 pour son substrat S_0 (chapitre 2.0). Les résultats ont permis de faire ressortir différentes règles à respecter pour la reconnaissance du substrat S_{TAR} par des ribozymes dérivés du ribozyme VS et pour maintenir une bonne activité de clivage. La première règle est la présence d'une séquence minimale de quatre nucléotides (5'-YYCR-3) pour la reconnaissance de la boucle du substrat S_{TAR} . Les autres règles sont la présence de quatre

ou cinq paires de bases W-C dans l'interaction *kissing-loop* et la présence de nucléotides libres dans les boucles I et V une fois l'interaction *kissing-loop* en place. Les deux dernières règles sont plus générales et elles pourraient aussi servir à l'ingénierie de ribozymes qui clivent des substrats avec des boucles différentes. Cette étude démontre clairement l'adaptabilité du ribozyme VS pour la reconnaissance du substrat par le ribozyme VS, en plus de fournir des règles générales pour l'ingénierie de l'interaction *kissing-loop* de ribozyme VS mutant pour le clivage de substrats repliés en tige-boucles.

Pour cet article, je suis première auteure, car j'ai contribué au développement de l'approche expérimentale et j'ai effectué toutes les expériences décrites dans l'article. J'ai entrepris les expériences par des tests préliminaires de SELEX, qui consistent à vérifier le clivage du mutant positif $R_{SELEX}(S_0/R_0)$ et du mutant négatif $R_{SELEX}(S_0/R_{0-\Delta b/b})$. Une optimisation des différentes étapes du SELEX avec ces deux mutants a été effectuée, afin d'être capable d'effectuer un cycle complet avec un bon rendement pour chacune des étapes. Ces étapes sont : la transcription *in vitro* des ADN en ARN, la réaction de clivage, la purification des ARN clivés, la transcription inverse pour obtenir l'ADN et la réaction de PCR pour l'amplification de l'ADN. Pour faire suite au succès de l'optimisation, j'ai effectué les expériences avec la librairie d'ADN pour cinq cycles. J'ai réalisé la mutagenèse de plasmides et la transcription *in vitro* des ribozymes sélectionnés par SELEX (avec l'aide de Dominique Chaussé, technicienne au laboratoire), effectué les études de cinétiques en *single-turnover* et l'analyse de tous les résultats, ainsi que l'élaboration des règles pour l'interaction *kissing-loop*. J'ai participé à l'écriture du manuscrit en entier. Le projet a été conçu par la Dre Pascale Legault en collaboration avec le Dr Peter Unrau (Simon Fraser University). La Dre Pascale Legault a supervisé l'ensemble de l'étude, incluant la collecte des données et l'analyse des résultats. Elle a relu et corrigé ce manuscrit.

***In vitro* selection of the *Neurospora* VS ribozyme for
substrate recognition via a new HIV-1 TAR-TAR'
kissing-loop interaction**

Julie Lacroix-Labonté¹ Peter Unrau² and Pascale Legault¹

¹Département de Biochimie, Université de Montréal, C.P. 6128, Succursale Centre-Ville,
Montréal, QC, Canada, H3C 3J7

²Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888
University Drive, Burnaby, BC, Canada, V5A 1S6

Running head: SELEX of the *Neurospora* VS ribozyme kissing-loop

Keywords: *in vitro* selection, RNA engineering, *Neurospora* VS ribozyme, HIV-1 TAR
RNA

ABSTRACT

The *Neurospora* VS ribozyme has the unique ability to recognize its stem-loop substrate with high affinity and specificity via the formation of a kissing-loop interaction (KLI). To identify VS-derived ribozymes that target different substrates, we previously substituted the natural kissing-loop interaction of a VS ribozyme substrate/ribozyme complex by the previously characterized HIV-1 TAR/TAR* kissing-loop interaction. We identified several variant substrate/ribozyme complexes with substantial cleavage activity, but with observed cleavage rates at least 160-fold lower than the parental substrate/ribozyme complex. In the present study, we used *in vitro* selection to identify VS ribozyme variants that more efficiently cleave a stem-loop substrate containing the terminal loop of the HIV-1 TAR RNA (TAR-derived substrate). Several VS ribozyme variants were found to cleave this variant TAR-derived substrate with superior catalytic activity than the previously rationally-engineered ribozyme variants. In particular, one selected ribozyme cleaves the TAR-derived substrate with a rate ($k_{obs} = 5.8 \text{ min}^{-1}$) that is similar to that of the parental *trans* VS ribozyme for its cognate substrate ($k_{obs} = 12.2 \text{ min}^{-1}$). The selected ribozyme variants with the best cleavage activity display a minimal consensus sequence (5'-YYCR-3') complementary to the loop of the TAR-derived substrate (3'-GGGU-5') that allows the formation of 4-5 Watson-Crick base pairs at the kissing-loop interaction, leaving several unpaired nucleotides in both interacting loops. This study provides general rules for engineering of VS ribozyme kissing-loop variants that specifically cleave alternate hairpin substrates, and these rules may be applicable for the engineering of other functional RNAs.

INTRODUCTION

Small nucleolytic ribozymes, like the hairpin, HDV, hammerhead, *glmS*, and twister ribozymes, are ideal model systems for understanding the great number of non-coding RNAs in biology due to their small size and their easily detectable biological activity. Although these ribozymes play specific functional roles in diverse biological contexts, they also provide basic architectures from which RNAs with novel activities can be developed for biotechnology and synthetic biology applications. For example, the hammerhead ribozyme, one of the most studied ribozymes, has been modified to allow cleavage of exogenous substrates of different sequences by simply modifying the sequence of its substrate binding arms such that it can recognize a single-stranded substrate of interest through Watson-Crick base pairing. These simple substrate recognition properties of the hammerhead ribozyme and several other natural nucleolytic ribozymes have made it relatively easy to adapt these ribozymes for cleavage of alternate substrates with high catalytic activity. In contrast, only a few studies have addressed the possibility of exploiting the unique architecture of the *Neurospora* VS ribozyme to cleave alternate substrates. This may be due to the apparent complexity of modifying the substrate specificity of the *Neurospora* VS ribozyme, which, unlike other small ribozymes, specifically recognizes and cleaves a stem-loop substrate. Nevertheless, it is of interest to better understand how the VS ribozymes can be modified to target other RNA stem-loops as it may shed some light into how RNA stem-loops, which are widespread RNA structural elements that play fundamental role in biology, can be targeted by tertiary interactions with other RNAs. In addition, it may be useful to develop site-specific endonucleases that specifically recognize a stable stem-loop, as shown in recent applications of CRISPR-associated endoribonucleases, such as RNA purification and RNA pull-down studies [1-4].

Previous *in vitro* investigations have led to a detailed functional and structural understanding of the VS ribozyme in both its *cis* and *trans* forms, which vary according to the covalent attachment of its stem-loop I (SLI) substrate domain [for reviews see [5-8]]. The VS ribozyme recognizes its SLI substrate via formation of a kissing-loop interaction (KLI) with stem-loop V (SLV) of the catalytic domain of the ribozyme, also

known as the *trans* ribozyme [Fig. 4.1A; [9, 10]]. This I/V KLI is known to be essential for optimal cleavage activity since it allows binding of the substrate with high affinity [11], activates the substrate for catalysis [12, 13] and facilitates formation of the active site between the SLI internal loop (G₆₃₈ loop) and the SLVI internal loop (A₇₅₆ loop) [14]. Recently, the NMR structure of an isolated SLI/SLV complex [15] and the crystal structure of the dimeric form of the VS ribozyme [16] have confirmed the structural characteristics of the KLI previously inferred from mutagenesis data [10], namely the U-turn motif in both the SLI and SLV loops and the three Watson-Crick base pairs between these loops. In addition, it reveals an additional WC/sugar edge base pair between C₆₂₉ of SLI and A₇₀₁ of SLV as well as an extensive array of hydrogen bonds and stacking interactions [11, 15, 16]. The structural reorganization of the SLI terminal loop upon kissing-loop formation involves a helix shift that affects both stem Ib and the G₆₃₈ loop [12, 13, 15]. Variants of the SLI substrate that are locked in the unshifted conformation are inactive, whereas those locked in the shifted conformation, termed pre-shifted substrates, are active and bind with higher affinity to the *trans* ribozyme [11, 17]. For cleavage to occur, the shifted G₆₃₈ loop of SLI must interact with the A₇₅₆ loop of the catalytic domain. In the proposed general acid-base mechanism [18-22], the N1-H1 of the protonated A₇₅₆ and the N1 of the deprotonated G₆₃₈ function, respectively, as proton donor and acceptor. In the crystal structure of the VS ribozyme, G₆₃₈ and A₇₅₆ are both properly positioned to interact with the scissile phosphate and fulfill their putative roles in the cleavage reaction [16].

In a recent study, we used a rational engineering approach to identify KLIs that could substitute for the I/V KLI of a substrate/ribozyme (*S/R*) complex derived from the natural VS ribozyme (*S₀/R₀*; Fig. 4.1A), while maintaining substantial cleavage activity [23]. Among several KLIs present in the structural database, the previously-characterized HIV-1 TAR/TAR* KLI complex [24, 25] was identified as a good candidate for kissing-loop substitution. The HIV-1 TAR RNA hairpin is responsible for transcriptional activation of the viral RNA through the recruitment of the viral Tat protein [26], and thereby contributes to a transactivation mechanism that is apparently unique to HIV [27, 28]. In our previous study, the VS ribozyme KLI was substituted by respectively incorporating

in SLI and SLV the natural hexanucleotide loop sequence and one closing base pair of the HIV-1 TAR RNA and the complementary loop sequence with two closing base pairs derived from TAR*, a rationally-designed TAR RNA aptamer [24]. We kinetically characterized the resulting *S/R* pair (S_{TAR-0}/R_{TAR*-0} in Fig. 4.1A), as well as related *S/R* pairs with various combinations of helix length in stems Ib and V [23]. Using this approach, several *S/R* pairs were found that display substantial cleavage activity, including the S_{TAR-0}/R_{TAR*-0} pair and the related S_{TAR-1}/R_{TAR*-0} pair that contains one less base pair in stem Ib. Although the TAR/TAR* KLI [$K_d = 6-15$ nM; [15, 29, 30]] is more stable than the VS ribozyme I/V KLI [$K_d \sim 400-700$ nM; [15, 29, 30]], the observed cleavage rate of the R_{TAR*-0} ribozyme for its cognate TAR-derived substrates (S_{TAR-0} and S_{TAR-1}) was found to be reduced by more than 150-fold compared to that of the R_0 ribozyme for its cognate S_0 substrate [23]. The lower catalytic activity of the R_{TAR*-0} ribozyme for its cognate substrates indicates that the TAR/TAR* KLI may not favor the formation of a productive active site. However, it is not clear if the VS ribozyme sequence could be further modified to increase its cleavage activity of TAR-derived substrates.

In the present study, we used an *in vitro* selection approach to identify VS ribozyme variants with increased catalytic activity for a hairpin substrate carrying the TAR RNA terminal loop of HIV-1 (S_{TAR} ; Fig. 4.1B). For this selection, the 5 residues of loop V and 4 residues from the adjacent two closing base pairs were replaced with randomized sequences, each 10-nt long. Although high-affinity aptamers have been previously identified that stably associate with the HIV-1 TAR RNA loop [24, 29, 31, 32], the present selection was designed to identify ribozyme variants that allow substrate cleavage, which requires both substrate binding through a stable KLI and substrate cleavage via formation of a productive active site. Interestingly, we found several ribozyme variants that outperform the previously engineered R_{TAR*-0} ribozyme in terms of their observed cleavage rate for cleavage of the S_{TAR} substrate, including one that has a similar activity to the natural R_0 ribozyme for cleavage of its cognate substrate. By comparing the loop V sequences of the selected ribozymes with the highest catalytic activity, we defined a consensus sequence that is complementary to part of the HIV-1

TAR RNA loop. Moreover, we identified two additional requirements for active substitutions of the natural VS ribozyme I/V KLI that could help guide future engineering studies of VS-derived ribozyme that efficiently cleave other alternate substrates.

RESULTS AND DISCUSSION

A two-strand ribozyme for selection of VS ribozymes with new KLIs

For the *in vitro* selection experiment, we used the two-strand VS ribozyme illustrated in Fig. 4.1B and termed R_{SEL} . In addition, two variants of R_{SEL} containing modifications in loops I and V, $R_{SEL}(S_0/R_0)$ and $R_{SEL}(S_0/R_{0-\Delta b/b})$, were used for preliminary validation of this two-strand ribozyme for *in vitro* selection studies (Fig. 4.1B). In these RNA systems, the top strand contains the SLI substrate and an incomplete catalytic domain with an intact SLV. It requires the bottom strand to allow formation of a functional catalytic domain, as previously shown for a comparable *trans* VS ribozyme [33]. Thus, such VS ribozyme system makes it possible to initiate the cleavage reaction after RNA purification, by first annealing these two strands and then adding the Mg^{2+} ions required for cleavage. The 20-nt 5'-tail of the top strand was designed to facilitate its size-based separation from the largest cleavage product on a denaturing polyacrylamide gel.

To verify the validity of the R_{SEL} ribozyme system for the selection experiments, we first kinetically investigated the $R_{SEL}(S_0/R_0)$ ribozyme, a variant of the R_{SEL} ribozyme system containing the loops I and V sequences of the previously-characterized S_0/R_0 ribozyme [Fig. 4.1A; [34]]. The $R_{SEL}(S_0/R_0)$ ribozyme allows cleavage with a k_{obs} of $1.0 \pm 0.1 \text{ min}^{-1}$ at 37°C in Selection Reaction buffer supplemented with 20 mM $MgCl_2$ (data not shown), confirming the functionality of this two-strand ribozyme system. By comparison, a similar *cis* ribozyme cleaved with a k_{obs} of $\sim 0.2 \text{ min}^{-1}$ at 37°C and 20 mM Mg^{2+} [35]. As expected, modification of the $R_{SEL}(S_0/R_0)$ to create a defective KLI [$R_{SEL}(S_0/R_{0-\Delta b/b})$; Fig. 4.1B] significantly reduced the formation of cleavage products ($<1\%$ in 60 min, thus $k_{obs} < 0.0002 \text{ min}^{-1}$; data not shown). Thus, this two-strand ribozyme should allow for selection of KLIs that support efficient cleavage activity.

Selection of VS ribozymes with a TAR-TAR' KLI

For the *in vitro* selection experiment, the R_{SEL} ribozyme system incorporates the HIV-1 TAR RNA sequences in stem Ia and stem-loop Ib of the SLI substrate, but preserves the G_{638} loop necessary for formation of the active site (Fig. 4.1B). In addition, the 6 nucleotides (nt) of the SLV loop and the adjacent two closing base pairs (10 nt in total) were randomized to allow synthesis of an RNA library with approximately a million (4^{10}) different loop sequences.

Starting from this library, the *in vitro* selection protocol was employed to identify VS ribozymes that cleave with high catalytic activity, by allowing formation of a specific KLI interaction between the modified SLI substrate containing a TAR RNA loop and a modified SLV loop with a complementary sequence, termed TAR' to distinguish it from the previous TAR* RNA aptamer sequence [24]. For each round of selection, the cleavage reaction time was optimized to allow selection of the fastest-cleaving ribozymes. Consequently, the reaction time was reduced as the number of rounds increased, starting with 10 min for the 1st round and finishing with 10 sec for the 4th round. As expected, the percentage of products at the 10-min time point increased systematically at each round of selection, from 1% at round 1 to 69% at round 4 (Fig. 4.2). For all the selection rounds, the RNA pool had higher cleavage activity than the negative control $R_{SEL}(S_0/R_{0-Ab/b})$ (not shown). In addition, the percentage of cleavage products for rounds 3 and 4 was higher than that of the positive control $R_{SEL}(S_0/R_0)$ (Fig. 4.2). A fifth round did not further improve the cleavage activity (data not shown), which is not surprising given the limited sequence diversity of the starting library. The RNA pool of round 4 was therefore chosen for cloning and sequencing, with 21 individual sequences being randomly picked for further analysis.

Kinetic characterization of selected VS ribozymes with a TAR-TAR' KLI

To kinetically characterize selected VS ribozymes with a TAR-TAR' KLI, single-turnover kinetic studies were performed with S_{TAR}/R_{TAR} pairs generated by modifying the well-characterized *trans* S_0/R_0 pair (Fig. 4.1A) with the SLI and SLV sequences of the selected ribozymes. To characterize the R_{TAR} ribozymes, a single S_{TAR} substrate was used,

which is a modified version of the S_{TAR-0} substrate employed previously (Fig.1A) and that includes the HIV-1 TAR sequences in both stem Ia and stem-loop Ib (Fig. 4.1B). As expected, the catalytic activity of the previously-characterized R_{TAR^*-0} ribozyme is essentially identical for these two substrates, with a $k_{obs} \leq 0.03 \text{ min}^{-1}$ for S_{TAR} (Table IV.I) and S_{TAR-0} [23]. However, the cleavage activity of the selected R_{TAR} ribozymes for the S_{TAR} substrate vary greatly, with k_{obs} values that are similar to or larger than the R_{TAR^*-0} ribozyme (Table IV.I). Based on their catalytic activity with the S_{TAR} substrate, the 21 selected R_{TAR} ribozymes were classified in three classes. Class I includes the most active ribozymes (9 of 21) with a k_{obs} value within 100-fold of that obtained for the S_0/R_0 pair containing the natural KLI [k_{obs} of $12.2 \pm 0.5 \text{ min}^{-1}$; [23]]. Ribozymes in class II (8 of 21) are those with a k_{obs} value within 100- to 400-fold of that obtained for the S_0/R_0 pair ($0.12 \text{ min}^{-1} > k_{obs} > 0.03 \text{ min}^{-1}$). Only a few ribozymes (4 of 21), those of class III, have a lower activity ($k_{obs} \leq 0.03 \text{ min}^{-1}$) that compares to that of the rationally designed S_{TAR}/R_{TAR^*-0} and the negative control S_{TAR}/R_0 .

Interestingly, the most active ribozyme complex within class I, S_{TAR}/R_{TAR14} , has a k_{obs} of $5.8 \pm 0.5 \text{ min}^{-1}$ that is similar within two-fold to that previously determined for the S_0/R_0 pair containing the natural KLI [23]. These results clearly establish that the VS ribozyme provides an adaptable RNA architecture that allows substitution of the KLI for recognition and highly efficient cleavage of an alternate substrate.

Sequence analysis of the selected VS ribozymes with a TAR-TAR' KLI

The most active selected R_{TAR} ribozymes likely share sequence characteristics in loop V to allow formation of TAR/TAR' KLIs that are most favorable for cleavage of the S_{TAR} substrate. Sequence analysis of the randomized region from class 1 R_{TAR} ribozymes reveals a minimal 5'-YYCR-3' consensus sequence, compatible with formation of four *cis* WC/WC base pairs (C-G, U-G or A-U) with the 3'-GGGU-5' sequence of the S_{TAR} loop (Fig. 4.3). For some class 1 R_{TAR} ribozymes (5 of 9), there is potential for formation of a fifth *cis* WC/WC base pair with the S_{TAR} loop, extending the minimal consensus to a 5'-UYYCR-3' or a 5'-YYCRG-3' sequence. In contrast, the majority of the twelve loop V sequences of class 2 and class 3 R_{TAR} ribozymes do not likely form 4-5 *cis* WC/WC base

pairs with the S_{TAR} loop, but instead three (5 sequences), six (1 sequence) or even seven (1 sequence) *cis* WC/WC base pairs, which appear to be less favorable for cleavage of the S_{TAR} substrate. Of note, the position of the minimal consensus sequence within loop V is not conserved; it can start at the third, fourth or fifth residue of the randomized region. Also, there is no requirement for the first two and last two residues of the randomized region to form *cis* WC/WC base pairs that would stabilize stem V as found in the natural VS ribozyme [Fig. 4.1A; [15, 16]]. As a consequence, formation of 4-5 *cis* WC/WC base pairs at the KLI between the S_{TAR} substrate and the R_{TAR} ribozymes would leave several residues unpaired in both kissing loops: 1-2 in the S_{TAR} loop and 3-6 in loop V (Fig. 4.3 and Table IV.I). These free loop residues may help increase the flexibility of the KLI as likely necessary to assist the formation of the active site and efficient substrate cleavage.

General requirements for substitution of the natural VS ribozyme I/V KLI

The present *in vitro* selection study identifies three important requirements for substitution of the natural VS ribozyme I/V KLI by a TAR/TAR' KLI to allow efficient cleavage of the S_{TAR} substrate (Fig. 4.4A): (1) a 5'-YYCR-3' consensus sequence in loop V; (2) formation of 4-5 *cis* WC/WC base pairs; (3) and unpaired residues in both loops, at least 1-2 unpaired residues in loop I and 3-6 unpaired residues in loop V. Although the first requirement is specific to recognition of the S_{TAR} substrate investigated here, the last two requirements may be more generally applicable. Interestingly, both the NMR and the X-ray structures of near wild-type KLI reveal 3 *cis* WC/WC base pair and 1 *cis* WC/Sugar Edge base pair between loop I and loop V [Fig. 4.4B: [15, 16]]. The NMR structure, in which the loop sequences of SLI and SLV are equivalent to those of the S_0/R_0 pair, reveals several unpaired residues in loop I (3: A₆₂₇, U₆₂₈ and G₆₃₃) and loop V (2: U₆₉₆ and U₇₀₀) as well as residues forming unstable base pairs in both SLI (2: C₆₂₆, G₆₃₃ and G₆₃₄) and SLV (1: U₆₉₅) [Fig. 4.4B: [15]]. Moreover, the NMR studies indicate that these loop I and V residues that do not form stable base pairs display different dynamic behavior than most other residues in the complex [15]. In the X-ray structure, which includes a different loop I sequence, there are 3 residues in both loop I (G₆₂₇, U₆₂₈ and G₆₃₃) and loop V (U₆₉₅, U₆₉₆ and U₇₀₀) that do not form a stable base pair [16]. Thus, the two general requirements for substitution of the VS ribozyme I/V KLI defined from

our *in vitro* selection study appear consistent with both the NMR and X-ray structures of VS ribozyme KLI complexes [15, 16] and could help guide future design of VS-derived ribozymes that cleave alternate substrates.

The two general requirements described above are not met by the rationally-engineered S_{TAR}/R_{TAR^*} ribozymes previously investigated, including the S_{TAR-0}/R_{TAR^*-0} pair (Fig. 4.1A), for which the k_{obs} was found to be more than 400-fold smaller than that of the parental S_0/R_0 pair [23]. This R_{TAR^*-0} ribozyme likely forms 6 *cis* WC/WC base pairs with its S_{TAR-0} substrate (Fig. 4.4C), which is more than specified by the consensus defined here (Fig. 4.3). In addition, the formation of these 6 base pairs at the KLI of the S_{TAR-0}/R_{TAR^*-0} pair requires that all residues from loops I and V be paired, thus diverging from the requirements for unpaired nucleotides in these loops. We suggest that a minimum of 4 base pairs at the KLI interaction is necessary for formation of a sufficiently stable KLI to allow efficient SLI cleavage, but that more than 5 base pairs may restrict the flexibility of the KLI and thereby hinder SLI cleavage, at least with the defined loop sizes examined here. Furthermore, unpaired nucleotides in both kissing loops may be required for high-cleavage activity, possibly by allowing the necessary dynamics at the KLI to favor formation of a productive active site.

Comparison with previous *in vitro* selection studies to identify aptamers of the HIV-1 TAR RNA

It is interesting to compare the consensus sequence obtained here for the recognition and cleavage of the S_{TAR} substrate to those previously obtained for the selection of RNA aptamers that target the HIV-1 TAR RNA. In a first study [29], an 8-nt 5'-GUCCCAGA-3' consensus loop sequence was identified that is complementary to the entire 6-nt loop of the TAR RNA and allows formation of a loop-closing G-A base pair that contributes to the high affinity of the interaction [30, 36, 37]. In a second selection study, a 6-nt 5'-ACCCAG-3' consensus loop sequence was identified [31], with sequence complementary to only five of the six nt of the TAR RNA loop. In a subsequent genomic selection study, the aptamers with the best affinity for the TAR RNA share a 5'-ACCCAG-3' sequence [32], and it was found that this affinity is comparable to that of an aptamer with the 8-nt

5'-GUCCCAGA-3' consensus [29, 32]. In the context of the present in vitro selection study designed to identify R_{TAR} ribozymes that cleave S_{TAR} substrates containing an HIV-1 TAR RNA loop from a library containing a randomized region of 10 nt, a minimal 4-nt consensus loop sequence (5'-YYCR-3') that includes the 5'-CCCA-3' sequence was identified, which has sequence complementarity to four nt of the TAR RNA loop (3'-GGGU-5'). This consensus is not only shorter than consensus sequences obtained previously for the selection of TAR RNA aptamers [29, 31, 32], but it is also more permissive in terms of nucleotide identity. Indeed, the complementarity of previous consensus sequences for the TAR RNA is based on formation of G-C and A-U base pairs only, whereas that of the 4-nt 5'-YYCR-3' consensus defined here allows formation of G-U wobble pairs. It is noteworthy that the loop V of the most active R_{TAR14} likely forms only four base pairs with the terminal loop of the TAR-derived substrate. This indicates that this most active ribozyme was not solely selected for its high affinity for the terminal loop sequence of the HIV-1 TAR RNA, as for previous TAR RNA aptamers, but for other features that confers high cleavage activity of the TAR-derived substrate. There is no doubt that efficient cleavage by the VS ribozyme requires more than the formation of a high-affinity I/V KLI; the KLI must also allow docking of the substrate with the catalytic domain for formation of a productive active site.

CONCLUDING REMARKS

The present study demonstrates that the I/V KLI of the wild-type VS ribozyme can be substituted by another KLI and cleave with high catalytic activity a hairpin substrate that is very different from its natural SLI substrate. The S_{TAR} substrate targeted here by VS ribozyme variants contains the terminal loop as well as two stems derived from the HIV-1 TAR RNA, but its cleavage site internal loop is that of the wild-type VS ribozyme. Cleavage of the natural HIV-1 TAR RNA would require replacement of the internal loop of the S_{TAR} substrate by a three-nt bulge that does not contain a guanine, like G₆₃₈ of the natural VS ribozyme, that could act as a general base in the enzymatic mechanism. Further investigations will be necessary to determine if the VS ribozyme can be further modified to cleave such natural HIV-1 TAR RNA hairpin or other relevant stem-loop RNAs of biomedical interest.

MATERIALS AND METHODS

In vitro selection procedure

All DNA sequences were obtained from IDT (Integrated DNA Technologies, IA), including the DNA template library for the top strand of the R_{SEL} ribozyme (Figure 1B; 5'-CTA ATA CCG TCA CTT AGC CTC TAC CAA TCA CAT AAC ANN NNN NNN NNT GCT ACG ACA ATT TCA GTA CTG AAA TTG GAG GTG AGT TCC CTG CTT ACT TTC AGA TCG AGC TCC CAG GCT CTT CTC TGC ATG TCA TAT GTG CTG TGC CTA TAG TGA GTC GTA TTA-3'), the DNA template for the bottom strand of the R_{SEL} ribozyme (Figure 1B), the T7 promoter primer, the R-primer-23 (5'-CTA ATA CCG TCA CTT AGC CTC TA-3') and the F-primer-61 (5'-TAA TAC GAC TCA CTA TAG GCA CAG CAC ATA TGA CAT GCA GAG AAG AGC CTG GGA GCT CGA T-3'). All DNA oligos were gel purified. The top and bottom strands of the R_{SEL} RNA library (Figure 1B) were prepared by *in vitro* T7 RNA polymerase transcription, treated with DNase I (Roche, IN) and purified by denaturing gel electrophoresis, as previously described [38]. The top strand was ^{32}P -body-labeled as part of the transcription reaction by adding [α - ^{32}P] ATP to the transcription mix (Perkin Elmer, Ont.). The purified top and bottom strands were mixed in approximately equimolar amount in H_2O , heated at 90°C for 1 min and cooled down to 37°C for 5 min [19]. The buffer conditions were then adjusted to 40 mM Tris-HCl pH 8, 1 mM EDTA, 2 mM spermidine and 25 mM KCl (Selection Reaction buffer) and the mix was incubated for 5 min at 37°C . The cleavage reaction was initiated by the addition of 20 mM MgCl_2 . For each round of selection, the cleavage reaction was first performed on a small scale to determine the reaction time for the selection step. The top strand 3'-products containing the randomized region were purified by denaturing polyacrylamide gel [7.5% acrylamide:bisacrylamide (19:1) with 7 M urea] and then reverse transcribed with Superscript II reverse transcriptase and the R-primer-23 according to the manufacturer's protocol (Invitrogen, CA). The subsequent PCR reaction with the F-primer-61, the R-primer-23 and Platinum Taq DNA Polymerase (Invitrogen, CA) allowed DNA amplification of the selected top strand 3'-products, and the addition of the coding sequence for the T7 promoter and the top strand 5'-product. Between 9 and 13 PCR cycles were used for DNA amplification at each round. The resulting PCR products were

purified on a 3% agarose gel prior to the *in vitro* transcription of the next round of selection. PCR products from the fourth round were cloned for sequencing, since a fifth round did not improve the ribozyme activity of the RNA pool.

Preparation of S_{TAR} substrates and R_{TAR} ribozymes for kinetic studies

The S_{TAR} and R_{TAR} variants are derivatives of the S_0 and R_0 RNAs (Figure 1A), modified to fit the invariable sequence of stem Ia and stem-loop Ib (S_{TAR} ; Figure 1B), and selected sequences of stem-loop V (R_{TAR}) from the R_{SEL} top strand pool (Table IV.I). The R_{TAR} plasmids were prepared by QuickChange II site-directed mutagenesis (Stratagene, CA) from the pAvapl plasmid [34]. The 5'-end-³²P-labeled S_{TAR} and unlabeled R_{TAR} RNAs were prepared using established procedures [34].

Single-turnover kinetic studies of the S_{TAR}/R_{TAR} ribozyme complexes

Single-turnover kinetic studies were carried out at 37°C in 1X Kinetic Reaction buffer (50 mM Tris-HCl at pH 7.6, 25 mM KCl, 2 mM spermidine and 25 mM MgCl₂) to determine the values of the observed rate constant (k_{obs}) using a substrate concentration ([S]) of 250 pM and a ribozyme concentration ([R]) of 0.5 μM, as previously described [34]. Briefly, the percentage of remaining substrate [$F = S_t/(S_t+P_t) \times 100\%$; where S_t and P_t are the amount of substrate and product at time t] was plotted against time and the data fit to the equation $F = Ae^{-(k_{obs}xt)} + F_\infty$, where A is the amplitude, k_{obs} is the rate constant of the reaction and F_∞ represents the unreactive fraction of the substrate. The value of k_{obs} , the standard error for k_{obs} , and the F_∞ parameter were estimated by non-linear regression analysis within the OriginPro 8.1 software (OriginLab, MA). The F_∞ values were below 10% for all S/R combinations. The k_{obs} values are reported as average values with standard deviations obtained from at least three independent experiments.

ACKNOWLEDGMENTS

The authors thank Dominique Chaussé for RNA preparation. This work was supported by a grant from the Canadian Institutes of Health Research (CIHR MOP-86502 to P.L.); CIHR and the Université de Montréal Ph.D. scholarships (to J.L.-L.) and a Canada Research Chair In Structural Biology and Engineering of RNA (to P.L.).

REFERENCES

1. Lee, H.Y., et al., *RNA–protein analysis using a conditional CRISPR nuclease*. Proc Natl Acad Sci U S A, 2013. **110**(14): p. 5416-5421.
2. Salvail-Lacoste, A., et al., *Affinity purification of T7 RNA transcripts with homogeneous ends using ARiBo and CRISPR tags*. RNA, 2013. **19**(7): p. 1003-14.
3. Di Tomasso, G., et al., *Affinity purification of in vitro transcribed RNA with homogeneous ends using a 3'-ARiBo tag*. Methods Enzymol, 2014. **549**: p. 49-84.
4. Hochstrasser, M.L. and J.A. Doudna, *Cutting it close: CRISPR-associated endoribonuclease structure and function*. Trends Biochem Sci, 2015. **40**(1): p. 58-66.
5. Lilley, D.M., *Catalysis by the nucleolytic ribozymes*. Biochem Soc Trans, 2011. **39**(2): p. 641-6.
6. Lilley, D.M., *The Varkud satellite ribozyme*. RNA, 2004. **10**(2): p. 151-8.
7. Collins, R.A., *The Neurospora Varkud satellite ribozyme*. Biochem Soc Trans, 2002. **30**(Pt 6): p. 1122-6.
8. Wilson, T.J. and D.M. Lilley, *A mechanistic comparison of the Varkud satellite and hairpin ribozymes*. Prog Mol Biol Transl Sci, 2013. **120**: p. 93-121.
9. Guo, H.C. and R.A. Collins, *Efficient trans-cleavage of a stem-loop RNA substrate by a ribozyme derived from neurospora VS RNA*. EMBO J, 1995. **14**(2): p. 368-76.
10. Rastogi, T., et al., *A long-range pseudoknot is required for activity of the Neurospora VS ribozyme*. EMBO J, 1996. **15**(11): p. 2820-5.
11. Bouchard, P. and P. Legault, *A remarkably stable kissing-loop interaction defines substrate recognition by the Neurospora Varkud Satellite ribozyme*. RNA, 2014. **20**(9): p. 1451-64.
12. Andersen, A.A. and R.A. Collins, *Rearrangement of a stable RNA secondary structure during VS ribozyme catalysis*. Mol Cell, 2000. **5**(3): p. 469-78.
13. Andersen, A.A. and R.A. Collins, *Intramolecular secondary structure rearrangement by the kissing interaction of the Neurospora VS ribozyme*. Proc Natl Acad Sci U S A, 2001. **98**(14): p. 7730-5.
14. Hiley, S.L. and R.A. Collins, *Rapid formation of a solvent-inaccessible core in the Neurospora Varkud satellite ribozyme*. EMBO J, 2001. **20**(19): p. 5461-9.
15. Bouchard, P. and P. Legault, *Structural insights into substrate recognition by the Neurospora Varkud satellite ribozyme: importance of U-turns at the kissing-loop junction*. Biochemistry, 2014. **53**(1): p. 258-69.
16. Suslov, N.B., et al., *Crystal structure of the Varkud satellite ribozyme*. Nat Chem Biol, 2015. **11**(11): p. 840-6.
17. Zamel, R. and R.A. Collins, *Rearrangement of substrate secondary structure facilitates binding to the Neurospora VS ribozyme*. J Mol Biol, 2002. **324**: p. 903-915.
18. Hiley, S.L., et al., *4-thio-U cross-linking identifies the active site of the VS ribozyme*. EMBO J, 2002. **21**(17): p. 4691-8.
19. Lafontaine, D.A., et al., *Functional group requirements in the probable active site of the VS ribozyme*. J Mol Biol, 2002. **323**(1): p. 23-34.
20. Wilson, T.J., A.C. McLeod, and D.M. Lilley, *A guanine nucleobase important for catalysis by the VS ribozyme*. EMBO J, 2007. **26**(10): p. 2489-500.
21. Jaikaran, D., et al., *An important role of G638 in the cis-cleavage reaction of the Neurospora VS ribozyme revealed by a novel nucleotide analog incorporation method*. RNA, 2008. **14**(5): p. 938-49.

22. Wilson, T.J., et al., *Nucleobase-mediated general acid-base catalysis in the Varkud satellite ribozyme*. Proc Natl Acad Sci U S A, 2010. **107**(26): p. 11751-6.
23. Lacroix-Labonte, J.G., N.; Dagenais, P.; Lemieux, S.; Legault, P., *Rational and combinatorial design of kissing-loop interactions that allow substrate recognition by the Neurospora VS ribozyme*. NAR, 2015.
24. Chang, K.Y. and I. Tinoco, Jr., *Characterization of a "kissing" hairpin complex derived from the human immunodeficiency virus genome*. Proc Natl Acad Sci U S A, 1994. **91**(18): p. 8705-9.
25. Chang, K.Y. and I. Tinoco, Jr., *The structure of an RNA "kissing" hairpin complex of the HIV TAR hairpin loop and its complement*. J Mol Biol, 1997. **269**(1): p. 52-66.
26. Jones, K.A. and B.M. Peterlin, *Control of RNA initiation and elongation at the HIV-1 promoter*. Annu Rev Biochem, 1994. **63**: p. 717-43.
27. Arts, E.J. and D.J. Hazuda, *HIV-1 antiretroviral drug therapy*. Cold Spring Harb Perspect Med, 2012. **2**(4): p. a007161.
28. Karn, J. and C.M. Stoltzfus, *Transcriptional and posttranscriptional regulation of HIV-1 gene expression*. Cold Spring Harb Perspect Med, 2012. **2**(2): p. a006916.
29. Duconge, F. and J.J. Toulme, *In vitro selection identifies key determinants for loop-loop interactions: RNA aptamers selective for the TAR RNA element of HIV-1*. RNA, 1999. **5**(12): p. 1605-14.
30. Duconge, F., C. Di Primo, and J.J. Toulme, *Is a closing "GA pair" a rule for stable loop-loop RNA complexes?* J Biol Chem, 2000. **275**(28): p. 21287-94.
31. Ohuchi, S.P. and Y. Nakamura, *Slight sequence modifications unexpectedly alter the metal-dependency of a kissing-loop interaction*. Nucleic Acids Symp Ser (Oxf), 2007(51): p. 395-6.
32. Watrin, M., et al., *In vitro selection of RNA aptamers derived from a genomic human library against the TAR RNA element of HIV-1*. Biochemistry, 2009. **48**(26): p. 6278-84.
33. Lafontaine, D.A., et al., *The A730 loop is an important component of the active site of the VS ribozyme*. J Mol Biol, 2001. **312**(4): p. 663-74.
34. Lacroix-Labonte, J., et al., *Helix-length compensation studies reveal the adaptability of the VS ribozyme architecture*. Nucleic Acids Res, 2012. **40**(5): p. 2284-93.
35. Poon, A.H., et al., *Identification of separate structural features that affect rate and cation concentration dependence of self-cleavage by the Neurospora VS ribozyme*. Biochemistry, 2006. **45**(44): p. 13394-400.
36. Lebars, I., et al., *Exploring TAR-RNA aptamer loop-loop interaction by X-ray crystallography, UV spectroscopy and surface plasmon resonance*. Nucleic Acids Res, 2008. **36**(22): p. 7146-56.
37. Van Melckebeke, H., et al., *Liquid-crystal NMR structure of HIV TAR RNA bound to its SELEX RNA aptamer reveals the origins of the high stability of the complex*. Proc Natl Acad Sci U S A, 2008. **105**(27): p. 9210-5.
38. Bouchard, P., et al., *Role of SLV in SLI substrate recognition by the Neurospora VS ribozyme*. RNA, 2008. **14**(4): p. 736-48.
39. Leontis, N.B. and E. Westhof, *Geometric nomenclature and classification of RNA base pairs*. RNA, 2001. **7**(4): p. 499-512.

TABLE

Table IV.I Kinetic results for the S/R pairs¹

<i>S/R</i>	k_{obs} (min ⁻¹)	Relative k_{obs} value	Predicted secondary structure of <i>S/R</i> pairs				
			SLIb of substrate	SLV of ribozyme	Consecutive KLI WC/WC bp	Unpaired nt in the SLI loop	Unpaired nt in the SLV loop
S_0/R_0^2	12.2 ± 0.5	2.1	3'-CCGGCU 5'-GGCAUC	GUAGAC-5' AUCUG-3'	3	3	2
S_{TAR}/R_{TAR14}	5.8 ± 0.5	1.0	3'-UCGA 5'-AGCCUG	CGUAC-5' CAACUG-3'	4	2	6
S_{TAR}/R_{TAR4}	0.71 ± 0.05	0.12	3'-UCGA 5'-AGCCUG	CGAUAC-5' CA AUG-3'	4	2	3
S_{TAR}/R_{TAR34}	0.48 ± 0.04	0.083	3'-UCGA 5'-AGCCUG	CGUAGAC-5' CGACUG-3'	4	2	4
S_{TAR}/R_{TAR23}	0.45 ± 0.01	0.077	3'-UCGA 5'-AGCCUG	UACAC-5' CGAUG-3'	5	1	3
S_{TAR}/R_{TAR39}	0.40 ± 0.07	0.069	3'-UCGA 5'-AGCCUG	AGUAC-5' CAAGUG-3'	5	1	3
S_{TAR}/R_{TAR11}	0.34 ± 0.05	0.058	3'-UCGA 5'-AGCCUG	CUACAC-5' CAUAGUG-3'	5	1	3
S_{TAR}/R_{TAR10}	0.22 ± 0.03	0.037	3'-UCGA 5'-AGCCUG	AGCAC-5' CAGUG-3'	4	2	4
S_{TAR}/R_{TAR27}	0.16 ± 0.04	0.028	3'-UCGA 5'-AGCCUG	CAAC-5' CAAGUG-3'	5	1	5
S_{TAR}/R_{TAR28}	0.13 ± 0.02	0.022	3'-UCGA 5'-AGCCUG	CUACAC-5' CAAGUG-3'	5	1	3
S_{TAR}/R_{TAR35}	0.112 ± 0.008	0.019	3'-UCGA 5'-AGCCUG	CUCCAC-5' CAUAGUG-3'	6	2	2
S_{TAR}/R_{TAR3}	0.11 ± 0.01	0.019	3'-UCGA 5'-AGCCUG	CUAAC-5' CAUUG-3'	3	3	5
S_{TAR}/R_{TAR21}	0.10 ± 0.02	0.017	3'-UCGA 5'-AGCCUG	CAAC-5' CAUUG-3'	4	2	4
S_{TAR}/R_{TAR16}	0.057 ± 0.008	0.0098	3'-UCGA 5'-AGCCUG	ACCAC-5' CAUUG-3'	3	3	5
S_{TAR}/R_{TAR25}	0.051 ± 0.009	0.0088	3'-UCGA 5'-AGCCUG	CGUAC-5' CAUAGUG-3'	3	3	5
S_{TAR}/R_{TAR8}	0.037 ± 0.009	0.0064	3'-UCGA 5'-AGCCUG	CUACAC-5' CAUUG-3'	5	1	5
S_{TAR}/R_{TAR26}	0.035 ± 0.006	0.0060	3'-UCGA 5'-AGCCUG	CUAAC-5' CAUUG-3'	5	1	5
S_{TAR}/R_{TAR30}	0.031 ± 0.002	0.0053	3'-UCGA 5'-AGCCUG	AGCAC-5' CAUGUG-3'	4	2	2
S_{TAR}/R_{TAR33}	≤ 0.03	≤ 0.005	3'-UCGA 5'-AGCCUG	ACCAC-5' CAUUG-3'	3	3	5
S_{TAR}/R_{TAR7}	≤ 0.03	≤ 0.005	3'-UCGA 5'-AGCCUG	CGCAC-5' UUGUG-3'	4	2	0
S_{TAR}/R_{TAR31}	≤ 0.03	≤ 0.005	3'-UCGA 5'-AGCCUG	GAUAC-5' CAAGUG-3'	7	3	5
S_{TAR}/R_{TAR12}	≤ 0.03	≤ 0.005	3'-UCGA 5'-AGCCUG	GAAC-5' CGAUG-3'	3	3	7
S_{TAR}/R_{TAR*-0}	≤ 0.03	≤ 0.005	3'-CCGA 5'-GGCCUG	CUAGAC-5' CAACUG-3'	6	0	0
S_{TAR}/R_0	≤ 0.03	≤ 0.005	3'-UCGA 5'-AGCCUG	GUAGAC-5' AUCUG-3'	2	4	3

¹ Values of k_{obs} and errors are average values and standard deviations from at least three independent kinetic experiments collected with [S] = 250 pM and [R] = 0.5 μM at 37°C.

² This *S/R* pair was previously investigated [23] and the predicted secondary structure is based on the available NMR structure [15].

FIGURES

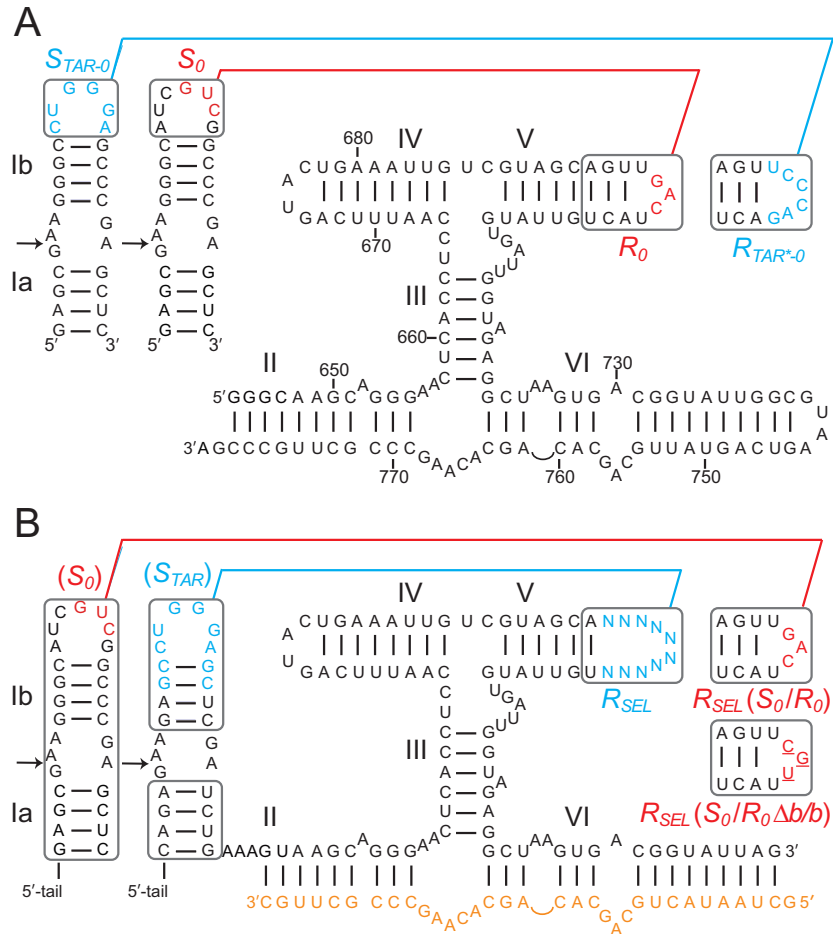


Figure 4.1. Sequences and secondary structures of the VS ribozyme variants used in this study.

((A) The previously-characterized S_0/R_0 and S_{TAR-0}/R_{TAR^*-0} substrate/ribozyme complexes [23] are shown with the nucleotides forming the WC base pairs in the KLI highlighted in red for R_0/S_0 and in blue for S_{TAR-0}/R_{TAR^*-0} . (B) The R_{SEL} ribozymes are formed by two strands: the top strand (black and blue) and the bottom strand (yellow). The 5'-tail (not shown in the figure) is composed of a 20-nt sequence (5'-GGC ACA GCA CAU AUG ACA UG-3'). The randomized nucleotides in loop V of the R_{SEL} ribozymes (blue) have the potential to base pair with the nucleotides of S_{TAR} loop (blue). The R_{SEL} variants $R_{SEL}(S_0/R_0)$ and $R_{SEL}(S_0/R_0\Delta b/b)$, containing modifications in loops I and V, were used as positive and negative controls, respectively. The specific cleavage site is located between G620 and A621 (black arrow).

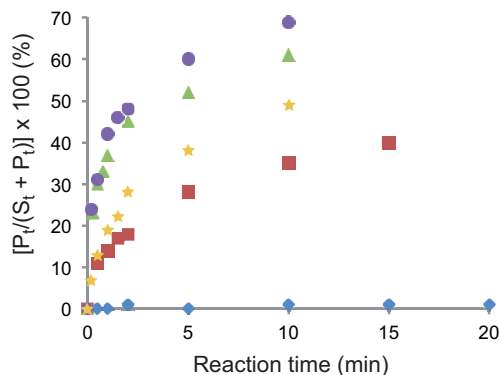


Figure 4.2 Evolution of the ribozyme cleavage activity for the different rounds of selection.

Evolution of the ribozyme cleavage activity for the different rounds of selection measured at 37°C in Selection Reaction buffer supplemented with 20 mM MgCl₂. The percentage of cleaved substrate $[(P_t/(S_t+P_t)) \times 100]$; where P_t and S_t are the amount of product and substrate at a time t is plotted against time for the first 4 rounds of selection [round 1 (blue diamond), 2 (red square), 3 (green triangle), and 4 (purple circle)] and for $R_{SEL}(S_0/R_0)$ (gold star).

Class 1	R_{PAR14}	<u>U G C C</u>	<u>C C C A</u>	<u>A C</u>
	R_{PAR4}	<u>U A G</u>	<u>C C C A</u>	<u>A A</u>
	R_{PAR34}	<u>G A</u>	<u>U C C G</u>	<u>C G A C</u>
	R_{PAR23}	<u>C C A U</u>	<u>C C C G</u>	<u>A G</u>
	R_{PAR39}	<u>U U G A</u>	<u>C C C A</u>	<u>G A</u>
	R_{PAR11}	<u>C A U</u>	<u>C C C A</u>	<u>U A G</u>
	R_{PAR10}	<u>C G A</u>	<u>C C C G</u>	<u>C A G</u>
	R_{PAR27}	<u>A C C A</u>	<u>C C C A</u>	<u>G A</u>
	R_{PAR28}	<u>C A U</u>	<u>C U C A</u>	<u>A A G</u>
minimal consensus			<u>Y Y C R</u>	
- with 5' extension			<u>U Y Y C R</u>	
- with 3' extension			<u>Y Y C R</u>	<u>G</u>
Class 2	R_{PAR35}	<u>C C U C C C A</u>	<u>U A G</u>	
	R_{PAR3}	<u>A G U C C G A</u>	<u>C A U</u>	
	R_{PAR21}	<u>A C A C C C G A A</u>	<u>U</u>	
	R_{PAR16}	<u>C C C A C C A C U</u>	<u>G</u>	
	R_{PAR25}	<u>U G C G C U C A</u>	<u>G</u>	
	R_{PAR8}	<u>C A U U C C A C G</u>	<u>U</u>	
	R_{PAR26}	<u>C A A U C C U A</u>	<u>U C</u>	
	R_{PAR30}	<u>C G U A C C C A</u>	<u>C G</u>	
	Class 3	R_{PAR33}	<u>C C A C C G A C A</u>	<u>G</u>
R_{PAR7}		<u>C C G C C C A</u>	<u>U G G</u>	
R_{PAR31}		<u>U A G A C U C A G G</u>	<u>U</u>	
R_{PAR12}		<u>A G A C C G C G G</u>	<u>A</u>	

Figure 4.3 Sequence alignment of the randomized region of the SLV loop for all the selected ribozymes characterized kinetically.

The loop V nucleotides with the potential to form a *cis* WC/WC base pair in the KLI are in red, whereas those likely to remain free are underlined. The sequences of the selected

ribozymes are separated in three classes according to the observed rate constant of their catalytic activity. For class 1, the boxed nucleotides are those that define the minimal consensus sequence for the recognition and cleavage of the S_{TAR} substrate. R_{TAR4} differs from other R_{TAR} ribozymes in that its loop contains 9 nt instead of 10. For R_{TAR31} in class 3, the U in italics is a non-randomized nucleotide potentially forming a *cis* WC/WC base pair in the KLI.

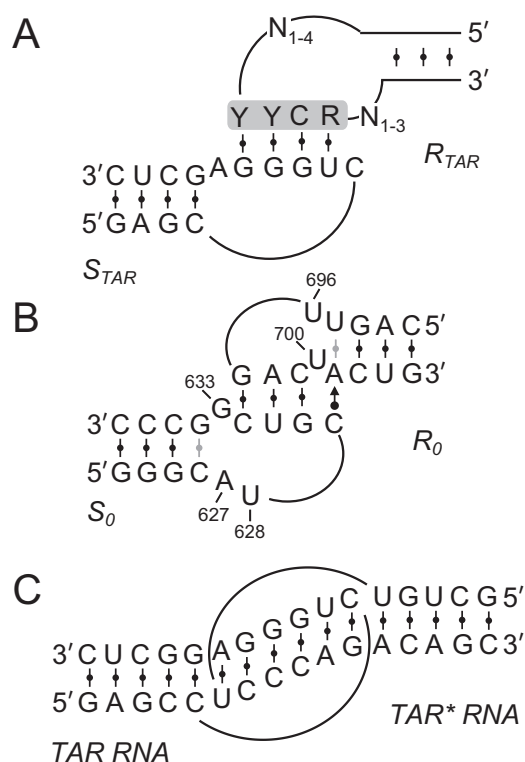


Figure 4.4 Schematics of kissing-loop interactions.

(A)(B) KLI as they likely form by (A) the class 1 R_{TAR} selected ribozymes with the S_{TAR} substrate and (B) by the R_0 ribozyme with the S_0 substrate [15]. (C) The KLI formed by the TAR RNA with the TAR* RNA as described by the NMR structure of the TAR/TAR* complex [25]. Base-pairing is indicated using the Leontis-Westhof notation [39], with black symbols for stable base pairs and grey symbols for unstable base pairs. In (A) and (B), only partial sequences of the ribozymes and the substrates are shown. In (A), the minimal consensus in the SLV loop of R_{TAR} is shaded in grey. A fifth *cis* WC/WC base pair involving the nucleotide on either the 5' or 3' side of the minimal consensus of R_{TAR} likely contributes to the KLI in some class 1 R_{TAR} ribozymes (Fig. 4.3).

5.0 Discussion et perspectives

5.1 Ingénierie du ribozyme VS

Pour le développement d'outils biochimiques et d'agents thérapeutiques basés sur l'ARN, c'est un avantage de pouvoir cibler des ARN repliés. Des motifs d'ARN particulièrement intéressants à cibler sont les tige-boucles d'ARN, puisqu'elles sont importantes pour le repliement global des molécules d'ARN complexes en permettant l'assemblage de modules d'ARN et en permettant les interactions longues distances [168-172]. Les tige-boucles peuvent aussi servir à la construction de motif d'ARN pour la formation de site catalytique ou encore, de motif de reconnaissance pour la liaison à des protéines. Des expériences de SELEX pour le ciblage de tige-boucles ont démontré que des aptamères repliés en tige-boucles étaient plus efficaces pour lier des tige-boucles cibles, que des aptamères simple brin [173]. Les aptamères repliés en tige-boucles peuvent faire la reconnaissance de tige-boucles cibles avec des interactions boucle apicale-boucle interne [166, 174, 175] ou avec des interactions *kissing-loop* entre les deux boucles apicales [176-181]. Les interactions *kissing-loop* sont très spécifiques, puisqu'elles combinent la contrainte de la spécificité d'une structure tridimensionnelle avec l'appariement de nucléotides complémentaires en paires de bases et il n'est pas nécessaire que tous les nucléotides des boucles soient complémentaires entre eux [66, 87, 182]. Jusqu'à maintenant, il n'y a pas de ribozymes autoclivants qui ont été modifiés pour la reconnaissance de tige-boucle par une interaction *kissing-loop*. Le but de cette thèse était d'explorer l'adaptabilité du ribozyme VS pour cette fonction. Des études d'ingénierie ont été entreprises pour modifier le ribozyme VS pour la reconnaissance de substrats modifiés en testant différents paramètres liés à la reconnaissance et au clivage du substrat. Un article préliminaire aux études d'ingénieries (Annexe 1) a démontré que la présence du *U-turn* canonique n'est pas essentielle à la réaction de clivage optimale et qu'il est possible de modifier plusieurs nucléotides de la boucle V et même d'ajouter un nucléotide dans la boucle sans diminuer l'activité. Cet article a déterminé que plusieurs modifications au niveau de l'interaction *kissing-loop* I/V sont tolérées et de multiples expériences de mutagenèse abondent aussi dans ce sens [65, 66, 80, 84, 85, 87, 89, 99, 105].

5.1.1 Ingénierie rationnelle

La première partie de cette thèse s'est concentrée sur l'ingénierie rationnelle du ribozyme VS aux chapitres 2.0 et 3.0. La première question à laquelle nous voulions répondre était si le ribozyme VS est capable de reconnaître et cliver des substrats, dont la longueur n'est pas la même que son substrat naturel SLI. Des études précédentes avaient démontré que des modifications de la longueur de la tige V inactivent le ribozyme, puisque ce changement de longueur a une répercussion sur la position du substrat SLI lors de l'interaction *kissing-loop* I/V et empêche probablement la formation du site actif [86]. Par conséquent, l'inactivation du ribozyme devrait aussi se produire si la longueur de la tige Ib du substrat SLI est modifiée. C'est pourquoi nous avons émis l'hypothèse que si l'on veut changer la longueur de la tige Ib, il doit se produire un changement compensatoire dans la tige V, pour conserver le même nombre de nucléotides total dans l'interaction *kissing-loop* I/V et permettre la bonne localisation du substrat au site actif. Les études de cinétiques enzymatiques sur toutes les combinaisons de substrat/ribozyme ont montré que le ribozyme R_0 peut cliver des substrats ayant jusqu'à trois longueurs différentes de tige Ib avec une activité similaire, ce qui a aussi été observé avec d'autres ribozymes mutants (R_{-1} et R_{-2} avec deux substrats chacun). De plus, cette étude d'ingénierie rationnelle supporte bien notre hypothèse de compensation d'hélices dans une certaine mesure, puisque chaque ribozyme a clivé de façon préférentielle le substrat qui le compense ou avec une paire de base de différence. Le k_{obs} lié à l'activité optimale de la paire de substrat/ribozyme de référence S_0/R_0 , qui ne comporte aucun changement dans les longueurs des tiges I et V, est de $6.5 \pm 0.1 \text{ min}^{-1}$ à 25 °C. Une activité similaire a été observée chez plusieurs ribozymes mutants pour le clivage d'un substrat plus court (S_{-1bp} par R_0), du substrat naturel (S_0 par R_{-1} et R_0) et de substrats plus longs (S_{+1bp} par R_0 , R_{-1bp} , R_{-2bp} , et S_{+2bp} par R_{-2bp} , R_{-3bp}) (voir figure 5.1B). Malgré les mutations dans les longueurs des tiges I et V, l'activité de clivage optimale est conservée seulement pour une gamme de longueurs de substrat avec des tiges Ib allant de 3 à 6 paires de bases. Un modèle de l'interaction *kissing-loop* I/V en accord avec les résultats des cinétiques enzymatiques a permis de constater que l'interaction entre les hélices est colinéaire et non pas coaxiale comme nous l'avions supposé avant cette étude. En résumé, cette étude démontre clairement l'adaptabilité de l'interaction *kissing-loop* I/V pour permettre la formation du site actif et le clivage du substrat.

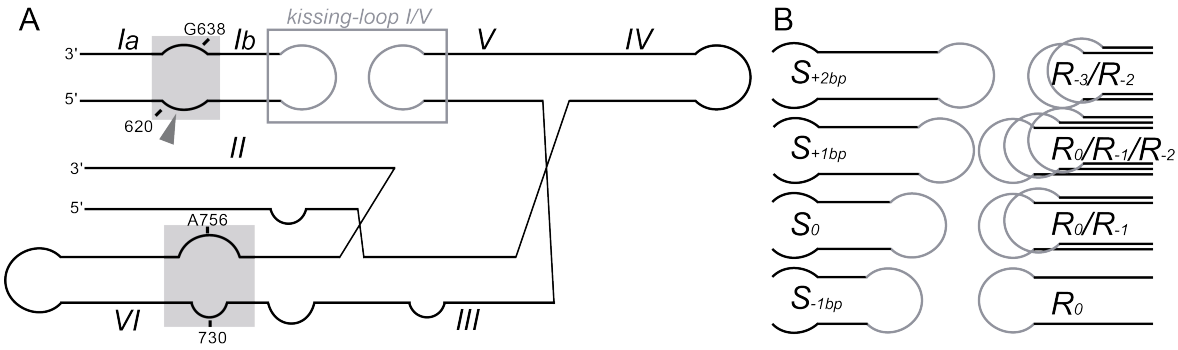


Figure 5.1 Résumé des expériences d'ingénierie du ribozyme VS.

Représentation schématique de la structure secondaire du ribozyme VS. Les boucles grises sont les boucles I et V qui forment l'interaction *kissing-loop* I/V dans l'encadré gris. Les régions ombragées représentent les boucles internes de SLI et SLVI impliquées dans la formation du site actif. (B) Résumé des meilleures combinaisons des longueurs des tiges I (à gauche) et V (à droite) formant l'interaction *kissing-loop* I/V obtenues lors des études de compensation de la longueur des hélices (chapitre 2.0).

La deuxième étape de l'ingénierie rationnelle a été de changer les boucles I et V au complet, pour voir si une substitution de l'interaction *kissing-loop* I/V permet de conserver l'activité de clivage du ribozyme VS. Le choix des interactions compatibles avec le ribozyme VS s'est effectué par recherche bio-informatique et les interactions choisies sont celles de TAR/TAR* du VIH-1 et L22/L88 de l'ARNr 23S de l'eubacterium *Deinococcus radiodurans*. Les mêmes mutants de compensation des longueurs des tiges Ib et V du chapitre 2.0 ont été utilisés pour déterminer les longueurs de tiges les plus appropriées pour accommoder chacune des interactions *kissing-loop*. La compensation des longueurs des hélices a été observée encore une fois dans cette étude. Les paires de substrat/ribozyme qui ont montré la meilleure activité pour chacune des substitutions des interactions *kissing-loop* ont des activités de clivage similaires entre elles (avec un facteur de moins de cinq) et ce sont les paires S_{TAR-1}/R_{TAR*-0} , S_{L88+1}/R_{L22-1} et S_{L22+1}/R_{L88-1} . Par contre, l'activité de clivage est diminuée comparativement à la séquence non modifiée de l'interaction *kissing-loop* I/V S_0/R_0 ($k_{obs} = 12.2 \pm 0.05 \text{ min}^{-1}$ à 37 °C) par un facteur de 159 pour l'interaction S_{TAR-1}/R_{TAR*-0} ($k_{obs} = 0.0769 \pm 0.0003 \text{ min}^{-1}$ à 37 °C), de 53 pour S_{L88+1}/R_{L22-1} ($k_{obs} = 0.23 \pm 0.02 \text{ min}^{-1}$) et de 84 pour

S_{L22+1}/R_{L88-1} ($k_{obs} = 0.145 \text{ min}^{-1}$ à 37 °C). Des modèles des interactions entre SLI et SLV avec les substitutions d'interactions *kissing-loop* ont été superposés sur le modèle de la structure cristalline du ribozyme VS complet. Les modèles montrent que le site de clivage des substrats mutants est mal positionné pour aller participer à la formation du site actif comparativement à S_0/R_0 , ce qui peut expliquer pourquoi ils ne sont pas aussi actifs que celui-ci.

L'ingénierie rationnelle a bien fonctionné pour les changements des longueurs des hélices I et V, mais elle est moins bien adaptée pour la substitution de l'interaction *kissing-loop*. Les interactions *kissing-loop* que nous avons sélectionnées dans des structures d'ARN de bases de données ne sont pas nécessairement les plus appropriées pour permettre l'activité, puisque l'interaction *kissing-loop* dans le contexte du ribozyme VS doit posséder une fonction essentielle, qui est de permettre la localisation du substrat pour la formation du site actif. En effet, il est difficile de prévoir comment le substrat est orienté lorsque l'interaction *kissing-loop* est changée, puisque la substitution a un effet sur l'angle adopté par les deux hélices et la dynamique de l'interaction. Par exemple, dans l'interaction S_{TAR-1}/R_{TAR*-0} , il n'y a aucun nucléotide libre, car ils sont tous impliqués dans les six pb de l'interaction, tandis que dans l'interaction S_0/R_0 , il n'y a que quatre pb et plusieurs nucléotides libres. La présence de nucléotides libres ou impliqués dans des paires de bases instables dans l'interaction *kissing-loop* I/V peut être un facteur important pour conférer l'adaptabilité nécessaire à la localisation du substrat. Une solution pour améliorer l'activité du ribozyme avec le substrat S_{TAR-1} , pourrait être de diminuer le nombre de paires de bases dans l'interaction *kissing-loop* à 4 bp, pour qu'il y ait plusieurs nucléotides libres et que l'interaction soit moins rigide. Une autre solution a été d'utiliser une autre forme d'ingénierie, comme le SELEX, pour identifier une interaction *kissing-loop* mieux adaptée pour la reconnaissance des substrats S_{TAR} .

5.1.2 Ingénierie par SELEX

L'ingénierie par SELEX possède l'avantage de tester de nouvelles interactions *kissing-loop* qui sortent de l'ordinaire. Dans le chapitre 4.0, les nucléotides de la boucle V ont été changés de façon aléatoire, afin d'identifier un ribozyme pour la reconnaissance et le clivage

d'un substrat S_{TAR} contenant les hélices et la boucle apicale de l'ARN TAR du VIH-1. Le SELEX a permis d'identifier plusieurs ribozymes qui sont compatibles avec le clivage du substrat S_{TAR} et le ribozyme R_{TAR14} démontre une activité de clivage de S_{TAR} similaire à celle de S_0/R_0 , avec une diminution de l'activité par un facteur de deux seulement ($k_{obs} = 5.8 \pm 0.5 \text{ min}^{-1} \mu\text{M}^{-1}$ à 37 °C). Ces expériences ont permis de déterminer que pour le clivage du substrat S_{TAR} , l'interaction *kissing-loop* doit contenir entre quatre et cinq paires de bases W-C et posséder un consensus minimal de quatre nucléotides avec la séquence 5'-YYCR-3' (dans laquelle les Y sont des pyrimidines et le R est une purine). De plus, il a été déterminé que la présence de nucléotides libres est importante autant dans la boucle I que dans la boucle V. Le contexte de nos expériences de SELEX a déterminé une séquence consensus différente de celles des études précédentes de SELEX pour cibler l'ARN TAR avec des aptamères repliés en tige-boucles [178-180]. Les séquences consensus de ces aptamères forment des interactions *kissing-loop* avec l'ARN cible, qui possèdent entre 5 et 6 paires de bases W-C et un seul ou aucun nucléotide libre dans les boucles. La séquence consensus déterminée pour le ribozyme VS est plus courte et plus permissive au niveau de l'identité des nucléotides, ce qui peut s'expliquer par des étapes de sélection du SELEX qui sont différentes. Dans le cadre du SELEX du ribozyme VS, la sélection est basée sur la vitesse du clivage du substrat, tandis que la sélection pour les expériences de SELEX des aptamères est basée seulement sur l'affinité de l'interaction entre les deux tige-boucles. Nos résultats indiquent que chez le ribozyme VS, la formation d'une interaction *kissing-loop* très stable n'est pas favorisée pour la réaction de clivage et qu'elle nécessite la présence de nucléotides libres pour lui conférer l'adaptabilité nécessaire à la formation du site actif. Ces caractéristiques de l'interaction *kissing-loop* I/V pourraient être appliquées au ciblage de substrats en tige-boucle, formant d'autres types d'interaction *kissing-loop* avec le ribozyme. L'ingénierie par SELEX a été un moyen efficace pour apporter des modifications au niveau de l'interaction *kissing-loop* du ribozyme VS pour la reconnaissance d'une nouvelle boucle de substrat, tout en permettant de conserver une activité de clivage efficace comparable à celle de S_0/R_0 .

5.1.3 Règles de l'ingénierie

Les différentes études de mutagenèse et d'ingénierie ont montré que le ribozyme VS peut être modifié de diverses façons sans nuire à son activité de clivage. Voici un résumé des changements qui peuvent être apportés, à partir de nos études et d'études antérieures, dans les domaines hélicoïdaux I, V et VI avec ou sans changements compensatoires et leur impact sur l'activité de clivage.

5.1.3.1 Identité des nucléotides dans les hélices Ia et Ib

Les études de mutagenèse ont montré que la plupart des nucléotides dans les tiges Ia et Ib du substrat peuvent être modifiés sans nécessiter de mutations compensatoires et même que la tige Ia peut être complètement enlevée sans nuire à l'activité de clivage [65, 80] (voir figure 5.2). Les études de compensation de longueurs des hélices (chapitre 2.0 et 3.0) ont montré que le nombre et l'identité des paires de bases de la tige Ib peuvent être modifiés et dans les études de SELEX (chapitre 4.0), les tiges Ia et Ib sont complètement remplacées par celle de l'ARN TAR du VIH-1 tout en conservant l'activité de clivage [163]. De plus, des études de SELEX sur l'ensemble de SLI ont démontré que plusieurs nucléotides pouvaient être modifiés dans les tiges Ia et Ib sans inhiber l'activité du ribozyme [84].

5.1.3.2 Longueurs des tiges I et V

Sans devoir apporter de changement compensatoire, le ribozyme VS peut reconnaître et cliver des substrats dont la tige Ib varie entre trois et cinq pb. Avec des mutations compensatoires, le ribozyme peut reconnaître et cliver des substrats SLI avec des tiges Ib variant entre trois et six pb. Par contre, ce chiffre pourrait être amené à changer si l'interaction *kissing-loop* est substituée pour une interaction qui occupe un espace différent. Par exemple, dans les études de substitution d'interaction *kissing-loop* (chapitre 3.0), les tiges Ib des substrats S_{TAR} varient entre trois et six pb et entre quatre et six pb pour les substrats S_{L88} . Pour les substrats S_{L22} , les tiges Ib sont encore plus longues, jusqu'à sept pb (entre quatre et sept pb) et ce malgré le fait que ces substrats possèdent la même interaction *kissing-loop* que les

substrats S_{L88} , mais avec les boucles I et V inversées. Les longueurs des tiges Ib qui soutiennent l'activité de clivage varient en fonction de l'interaction *kissing-loop* et par conséquent, c'est un facteur à tenir en compte dans les futures ingénieries.

5.1.3.3 Identité des nucléotides dans les boucles I et V

Dans les boucles I et V, il est possible de changer l'identité de plusieurs nucléotides sans affecter l'activité de clivage, même ceux impliqués dans les paires de bases de l'interaction *kissing-loop* I/V, à condition dans plusieurs cas de faire les mutations compensatoires dans la boucle opposée (voir figure 5.2B). Dans la boucle I, sans effectuer de mutations compensatoires, le nucléotide G627 peut être muté par n'importe quel nucléotide [65, 87], mais le nucléotide U628 ne peut pas être muté sans une perte d'activité par un facteur de plus de 25, puisqu'il est le U de la séquence UNR du motif *U-turn* de la boucle I [65]. Pour le nucléotide C629, des mutations vont diminuer l'activité par un facteur d'au moins dix et c'est cohérent avec les études de RMN et de cristallographie qui ont décelé une paire de base *sugar edge*/W-C entre C629 et A701 [65, 66, 87]. Les nucléotides G630, U631 et C632 sont impliqués dans l'interaction *kissing-loop* I/V et les mutations qui vont empêcher l'appariement d'une paire de base W-C dans l'interaction vont diminuer l'activité par un facteur d'au moins 10 [65, 80, 85, 86]. Si les trois paires de bases W-C sont absentes, c'est une diminution de l'activité de plus de 10 000 fois qui se produit. La seule mutation qui est tolérée est C632U, puisqu'elle remplace la paire de base G-C par une G-U dans l'interaction *kissing-loop* I/V. Dans la boucle V, sans effectuer de mutations compensatoires, il n'y a qu'un seul nucléotide qui peut être muté par n'importe quel nucléotide sans affecter l'activité et il s'agit de U700. Les autres nucléotides de la boucle V ne peuvent être mutés, car U696 est l'uridine du *U-turn* et les nucléotides G697, A698 et C699 sont impliqués dans l'interaction *kissing-loop* I/V [65, 89]. En incorporant les mutations compensatoires, il est possible de changer les nucléotides impliqués dans l'interaction *kissing-loop* I/V en utilisant des nucléotides qui permettent de former des paires de bases W-C dans l'interaction (réduction de l'activité par un facteur de moins de trois), pourvu que les séquences des motifs *U-turn* soient respectées dans les boucles I et V [65, 85]. Dans les études de substitution de l'interaction *kissing-loop* I/V au chapitre 3.0,

tous les nucléotides des boucles I et V sont changés et des baisses d'activité ont été remarquées, malgré la présence de plusieurs nucléotides complémentaires pouvant former plusieurs paires de bases W-C dans l'interaction *kissing-loop*. Par contre, les études de SELEX du chapitre 4.0 ont permis le remplacement de tous les nucléotides des boucles I et V, tout en conservant l'activité de clivage.

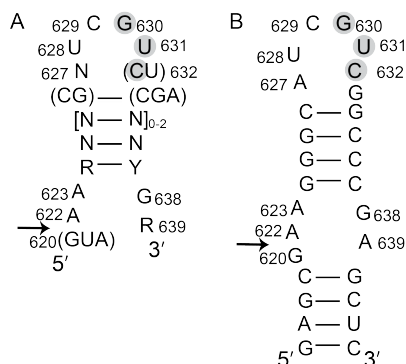


Figure 5.2 Consensus des modifications du substrat SLI sans changement compensatoire.

(A) Résumé des mutations qui peuvent être effectuées dans SLI de façon individuelle et sans mutations compensatoires pour conserver l'activité de clivage par le ribozyme VS naturel. Certaines combinaisons de ces mutations permettent aussi de conserver l'activité. (B) séquence du substrat SLI utilisé dans nos études. Le site de clivage est indiqué par la flèche noire et les nucléotides qui font des interactions W-C dans l'interaction *kissing-loop* I/V sont ombragés en gris. Les nucléotides entre parenthèses représentent les possibilités de nucléotides pour ces positions, les purines sont indiquées par R, les pyrimidines par Y et les N représentent les positions pour lesquelles tous les nucléotides sont possibles.

5.1.3.4 Nucléotides libres et nombre de paires de bases dans l'interaction *kissing-loop* I/V

Les études de SELEX du chapitre 4.0 ont mis en évidence l'importance d'avoir des nucléotides libres (ou impliqués dans des paires de bases instables) dans les boucles impliquées dans l'interaction *kissing-loop* du ribozyme VS. Dans l'interaction *kissing-loop* I/V, il y a trois paires de bases W-C et une paire de base non canonique observées dans les structures du cristal et de RMN, ce qui laisse plusieurs nucléotides libres (ou impliqués dans des paires de bases instables) dans chacune des boucles [66, 87]. Le nucléotide U700 est l'un

de ces nucléotides libres et il est celui avec le plus grand changement de conformation en réponse à l'ajout de magnésium [89]. Nos études ont démontré que U700 fait un pont hydrogène avec son phosphate et que la base de U700 n'est pas importante, contrairement au résidu [89, 90]. Les nucléotides libres peuvent être importants pour faire des ponts hydrogènes dans des paires de bases instables pour la stabilisation de l'interaction *kissing-loop* ou pour faire partie de motifs d'ARN comme le *U-turn* [87, 183, 184]. Lorsqu'il n'y a aucun nucléotide libre dans l'interaction, comme pour S_{TAR-1}/R_{TAR*0} (six pb W-C) dans les expériences de substitution de l'interaction *kissing-loop* (chapitre 3.0), l'activité de clivage est diminuée. Il y a aussi le nombre de paires de bases dans l'interaction *kissing-loop* qui peut s'avérer important, puisqu'il y a le même nombre de paires de bases dans l'interaction *kissing-loop* I/V (trois pb W-C et une pb non canonique) que pour le meilleur candidat du SELEX, S_{TAR}/R_{TAR14} (quatre pb W-C). Lorsqu'il y a des paires de bases supplémentaires, l'activité diminue comme pour S_{TAR-1}/R_{TAR*0} (six pb W-C), S_{L88}/R_{L22} et S_{L22}/R_{L88} (quatre pb W-C et une pb non canonique).

5.1.3.5 Site actif du ribozyme VS

Les travaux de cette thèse ont montré qu'il est possible pour l'instant de modifier complètement le substrat à l'exception de sa boucle interne. Dans les études de SELEX (chapitre 4.0), la séquence du substrat S_{TAR} contient la séquence de l'ARN TAR du VIH-1 pour la boucle apicale et les tiges Ia et Ib, mais nous avons changé le bulge de trois nucléotides de l'ARN TAR par la boucle interne du substrat SLI, pour faciliter la sélection des ribozymes mutants par SELEX. Afin de pouvoir cliver une nouvelle boucle interne, il faudrait faire des modifications au site actif du ribozyme dans la boucle interne de SLVI (boucle A730). La formation du site actif par l'interaction entre les boucles I et VI est une étape critique pour le clivage et pour faciliter cette étape, il faudrait choisir des ARN cibles avec des boucles internes qui ont des similarités avec la boucle interne du substrat SLI du ribozyme VS. La boucle interne de SLI contient le site de clivage, ainsi que la base de la catalyse G638 et elle est composée de cinq nucléotides dont trois d'entre eux, A621, A622 et G638 diminuent le clivage lorsqu'ils sont mutés [105] (voir figure 5.2). Par contre, les nucléotides G620 et

A639 peuvent tolérer quelques substitutions sans affecter l'activité [105]. La boucle A730 de SLVI contient plusieurs nucléotides importants pour l'activité, comme l'acide de la catalyse A756, ainsi que les nucléotides A730 et G757 [99]. Dans la boucle A730, il y a seulement le nucléotide C755 qui peut être substitué par A ou C sans réduire l'activité [99]. Malgré le fait que plusieurs nucléotides sont importants dans les boucles internes I et VI, le fait de changer entièrement la composition des deux boucles dans de futures études peut changer les prérequis pour la formation du site actif. C'est ce que nous avons observé dans les études de SELEX, qui ont permis de redéfinir les prérequis pour l'identité des nucléotides des boucles I et V impliqués dans l'interaction *kissing-loop* I/V.

5.2 Comparaison avec l'ingénierie des ribozymes autoclivants

Les ribozymes nucléolytiques normalement utilisés pour l'ingénierie comme les ribozymes hammerhead et le hairpin ciblent spécifiquement leur ARN cible à l'aide d'appariement W-C entre les bras du ribozyme et la région autour du site de clivage de l'ARN cible. Ensuite, la réaction de clivage s'effectue et les produits sont relâchés, en laissant le ribozyme libre d'agir avec une autre cible [133]. Dans le cas du ribozyme VS, la reconnaissance du substrat cible complet se fait en deux étapes : la première étape avec l'interaction *kissing-loop* I/V et la deuxième étape avec la formation du site actif, composé de l'interaction entre deux boucles internes. La reconnaissance du substrat et la formation du site actif sont différentes chez les ribozymes hammerhead et le ribozyme VS, puisqu'elle se fait par la formation de structures secondaires pour le ribozyme hammerhead et par la formation de deux interactions tertiaires pour le ribozyme VS. Par conséquent, les ribozymes s'adressent à des cibles différentes, puisque le ribozyme VS n'est pas adapté pour la reconnaissance des ARN simple brin et les ribozymes hammerhead et hairpin n'ont pas été modifiés jusqu'à maintenant pour la reconnaissance des ARN repliés en tige-boucle. Bien qu'il puisse être facile de déterminer une région simple brin à cibler avec le ribozyme hammerhead dans les nombreuses régions simple brin des ARNm ou des ARN viraux associés à des maladies, il peut être plus efficace pour la destruction complète de ces ARN de cibler plusieurs régions d'ARN en même temps. Il pourrait être intéressant pour ces fonctions d'utiliser les ribozymes hammerhead conjointement aux ribozymes VS, afin de pouvoir cibler un spectre plus large de régions à cliver contenant à la fois des régions simple brin et des régions repliées pour une plus grande efficacité dans la destruction complète d'un ARN cible.

5.3 Perspectives

Cette thèse a démontré que le ribozyme VS est un bon candidat pour l'ingénierie et qu'il devrait être possible de pousser l'ingénierie plus loin pour adapter le ribozyme au clivage de substrats entièrement modifiés. Ci-dessous, voici des propositions pour la suite des études d'ingénierie du ribozyme VS.

5.3.1 Ingénierie du ribozyme VS pour le clivage d'un ARN cible

Comme mentionné précédemment, la prochaine étape pour l'ingénierie du ribozyme VS est la reconnaissance et le clivage d'un substrat d'ARN entièrement modifié. Puisque de bons résultats ont été obtenus avec le SELEX pour la sélection d'une nouvelle interaction *kissing-loop*, la même technique pourrait être utilisée pour la formation d'un nouveau site actif entre la boucle A730 de SLVI du ribozyme VS et la boucle interne d'un substrat d'ARN cible. Pour la sélection des ARN cibles, il faut privilégier les tige-boucles qui ressemblent au substrat SLI, comme nous l'avons fait au chapitre 3.0 pour le choix des interactions *kissing-loop*. Des exemples de tige-boucles cibles sont donnés dans la prochaine section (5.3.2). Pour des expériences de SELEX dans le but de modifier la boucle A730 du ribozyme VS, le ribozyme en *trans* utilisé précédemment ne peut pas être utilisé, puisque les régions avec les nucléotides aléatoires dans la boucle A730 doivent se trouver dans le même brin d'ARN que le site de clivage pour permettre la sélection des ribozymes actifs. Le choix d'un ribozyme en *cis* s'impose pour ces expériences (voir figure 5.3) et la difficulté avec le ribozyme en *cis* est de déterminer des conditions de transcription qui vont restreindre le clivage au niveau le plus bas possible avant l'étape de sélection du SELEX. Puisque la sélection des ribozymes actifs lors du SELEX se fait avec la vitesse de la réaction de clivage, il est très important d'avoir un contrôle minutieux de la réaction afin de sélectionner seulement les ribozymes avec la meilleure activité.

Pour la reconnaissance d'un substrat entièrement modifié, il y a deux régions du ribozyme à modifier : la boucle V pour permettre l'interaction *kissing-loop* et la boucle A730

pour permettre la formation du site actif (représenté par les N en gris dans la figure 5.3). Il est possible d'effectuer le SELEX en deux étapes, c'est-à-dire d'effectuer la sélection de l'interaction *kissing-loop* tout d'abord et ensuite de faire la sélection de la séquence du site actif. Par contre, puisque la modification de l'interaction *kissing-loop* influence la localisation du substrat pour la formation du site actif du ribozyme, il peut être plus avantageux de faire les deux sélections dans une même étape de SELEX. Pour le SELEX avec le ribozyme en *cis*, le substrat SLI au complet est remplacé par l'ARN cible et les régions avec des nucléotides aléatoires se trouvent dans la boucle V et la boucle A730 de SLVI (figure 5.3). Il est possible d'élargir les régions de nucléotides aléatoires pour avoir une librairie de départ avec un plus grand nombre de séquences, pour augmenter les chances de sélectionner un ribozyme avec une bonne activité de clivage. La sélection des ribozymes actifs qui sont clivés peut s'effectuer de la même façon que les expériences précédentes, avec des gels dénaturants, ce qui s'est avéré une bonne technique pour l'isolation des ribozymes clivés.

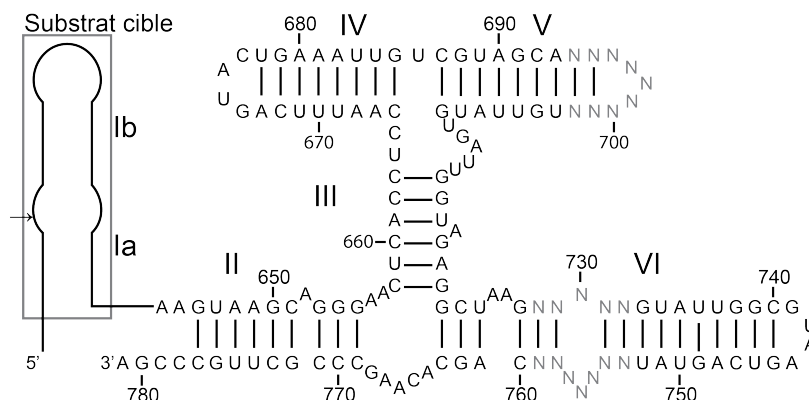


Figure 5.3 Expériences de SELEX pour la reconnaissance d'un substrat entièrement modifié.

Structure secondaire du ribozyme VS en *cis* pour des études de SELEX. La région de l'ARN cible est encadrée en gris. Le site de clivage est indiqué par une flèche noire et les lettres N en gris représentent les nucléotides qui sont changés de façon aléatoire dans la boucle V et la boucle A730.

5.3.2 ARN cibles intéressants pour le ribozyme VS

Pour la suite du projet, il y a plusieurs tige-boucles intéressantes à cibler qui sont importantes dans plusieurs systèmes biologiques et dont certaines sont impliquées dans des interactions *kissing-loop* [82, 185]. Les critères retenus pour le choix des cibles sont basés sur les règles établies au cours de nos études : (1) une boucle apicale entre quatre et dix nucléotides, (2) la présence d'une boucle interne ou d'un *bulge* et (3) une hélice entre trois et sept pb entre les deux boucles. Les tige-boucles suggérées dans cette section s'attardent sur deux maladies sur lesquelles il y a beaucoup de recherche, puisqu'elles infectent des millions de personnes autour du monde : le virus du VIH-1 et le virus de l'hépatite C.

5.3.2.1 VIH-1

Dans les études de SELEX au chapitre 4.0, l'ARN TAR du VIH-1 a été utilisé comme cible, mais la boucle interne avait été substituée par celle du substrat SLI du ribozyme VS afin de s'assurer que l'interaction *kissing-loop* du ribozyme VS pouvait être remplacée. Maintenant que nous avons cette information, une tentative de reconnaître et cliver l'ARN TAR complet est envisageable. L'ARN TAR (*transactivation response element*) est une tige-boucle composée de 59 nucléotides, localisé dans le 5'UTR de l'ARNm du VIH-1, près du site d'initiation de la transcription. Il a comme rôle principal le recrutement de la protéine tat au promoteur pour permettre la transactivation de la transcription [186, 187]. La structure de l'ARN TAR est importante pour sa reconnaissance par tat, car des études de mutagenèse ont démontré que la séquence des boucles et la structure secondaire de la tige-boucle sont nécessaires pour la transactivation de la transcription par tat [187-189]. Dans le but d'inhiber la réplication du virus en cellule, l'ARN TAR a été ciblé pour empêcher l'interaction entre TAR et tat par divers composés, comme de petites molécules organiques, des analogues de nucléotides polyamides, des peptides, des analogues de peptides ou des aptamères [190-195]. L'ARN TAR est une bonne cible pour le ribozyme VS, puisque l'on sait déjà que sa boucle apicale peut être liée par le ribozyme et il ne reste plus qu'à poursuivre l'ingénierie pour être capable de cliver dans son *bulge* de trois nucléotides (figure 5.4 A). Une deuxième tige-boucle pourrait être ciblée dans le 5'UTR de l'ARNm du VIH-1 et il s'agit du site d'initiation de la

dimérisation du VIH-1 (DIS) ou tige-boucle SL1 (figure 5.4 B). Il s'agit d'une tige-boucle très conservée de 35 nucléotides et elle accomplit une interaction *kissing-loop* qui représente un intermédiaire avant la formation d'un duplex étendu qui forme un homodimère [196, 197]. La tige-boucle SL1 a été ciblée par des aminoglycosides et bien que la liaison est bonne avec SL1 dans les cellules infectées, ce n'est pas efficace pour permettre l'inhibition de la réplication du VIH [198-201]. Par contre, des antisenses composés d'acides nucléiques peptidiques (PNA) qui ciblent la tige-boucle SL1 ont un effet inhibant très fort sur la dimérisation [202]. Cette tige-boucle SL1 est intéressante comme cible pour le ribozyme VS, car la tige-boucle SL1 comporte une hélice de sept paires de bases, qui relie la boucle apicale de neuf nucléotides et la boucle interne de quatre nucléotides [177, 203]. De plus, la tige-boucle SL1 possède elle aussi trois nucléotides libres, dans chacune de ses boucles après l'interaction *kissing-loop* et ces nucléotides sont importants pour la stabilité de l'interaction.

5.3.2.2 Virus de l'hépatite C

Le virus de l'hépatite C est un pathogène hématogène, qui infecte des millions de personnes à travers le monde et qui peut causer des infections du foie comme des cirrhoses ou des cancers [204, 205]. Le génome du virus de l'hépatite C est un ARN simple brin qui encode une grosse polyprotéine qui est clivée à de multiples sites pour produire au moins dix protéines [206]. Les régions 5'UTR et 3'UTR sont des régions conservées du virus et elles ont servis de cibles pour plusieurs ribozymes, aptamères, antisens et ARNi, afin de perturber le virus à différentes étapes de son cycle d'infection [207]. La région 5'UTR est essentielle pour la réplication du virus et la transcription de la polyprotéine. [208]. Cette région contient le domaine IRES (*internal ribosome entry site*), ainsi que quatre domaines d'ARN nommés SL1, SL2, SL3 et SL4 [209-211]. Un siARN qui cible la tige-boucle SL2 et une métallo-drogue avec un motif de liaison pour la tige-boucle SL2 permettent la dégradation spécifique de l'ARN du virus HCV [212, 213]. La tige-boucle SL2 est aussi une bonne cible pour le ribozyme VS, car elle possède une boucle apicale de sept nucléotides, une boucle interne de sept nucléotides, dont une séquence GAA dans le brin 5' de la boucle interne qui est identique à celle du ribozyme VS et l'hélice qui relie les deux boucles est composée de quatre paires de bases W-C

(voir figure 5.4 C). La région 3'UTR du génome ARN du HCV est, elle aussi, essentielle à la réplication du virus [214, 215]. Dans la région 3'UTR, il y a la région de l'ARN X qui contient trois tige-boucles : SL1, SL2 et SL3. L'ARN X contient aussi une région de nucléotides conservés à 100 % entre les souches de virus qui est nommée X55. La région X55 est suggérée d'être responsable de la dimérisation *in vitro* du virus, par la formation d'un complexe *kissing* qui est ensuite convertie dans une conformation duplex étendue qui est plus stable. Dans la région X55, il a la séquence DLS (*dimer linkage sequence*) qui est une séquence palindromique de 16 nucléotides, qui forme l'intermédiaire du complexe *kissing* et elle se trouve dans la tige-boucle SL2 [216, 217]. Cette tige-boucle SL2 du 3'UTR est elle aussi une bonne cible pour le ribozyme VS avec une boucle apicale de quatre nucléotides et une boucle interne de six nucléotides reliés par une hélice de six paires de bases W-C (figure 5.4 D).

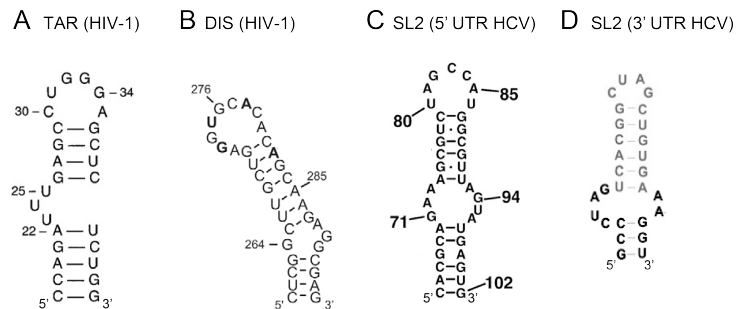


Figure 5.4 ARN intéressants pour le clivage par le ribozyme VS

Séquences des structures secondaires des tige-boucles d'ARN cibles : (A) tige-boucle TAR du VIH-1 (B) tige-boucle SL1 du DIS du VIH-1 (C) tige-boucle SL2 du 5'UTR de HCV et (D) la tige-boucle SL2 du 3'UTR de HCV. Les séquences sont raccourci pour laisser seulement les extrémités des tige-boucles. (Adaptée de [197, 216, 218, 219]).

5.3.3 Utilisation du ribozyme VS *in vivo*

Il n'y a pas beaucoup d'expériences qui ont été effectuées *in vivo* avec le ribozyme VS. Les expériences *in vivo* sont liées à la découverte du ribozyme VS dans les mitochondries de *Neurospora*, au cours desquelles les réactions de clivage et de ligation ont été observées *in vivo* [64, 76, 78]. Par la suite, les études de la réplication du ribozyme VS concluent le volet

des études *in vivo* [79]. Il existe différents problèmes liés à l'utilisation des ribozymes *in vivo* comme agents thérapeutiques, qui sont surtout liés à la stabilité de l'ARN dans les conditions cellulaires et à la livraison des ribozymes à proximité de leur cible.

5.3.3.1 Modifications des nucléotides

Pour augmenter la stabilité des ribozymes *in vivo*, il est possible d'apporter des modifications aux nucléotides pour les rendre plus résistants aux nucléases. Par exemple, la modification des 2'-hydroxyles des nucléotides permet d'augmenter la résistance des molécules d'ARN aux nucléases. Chez le ribozyme hammerhead, la modification des 2'-hydroxyles de la plupart des nucléotides par des 2'-O-méthyles permet une plus grande stabilité, en plus de ne pas conférer de toxicité [133]. Par contre, quelques nucléotides non modifiés doivent être conservés dans le ribozyme pour la formation du site actif et permettre de conserver une bonne activité de catalyse. Ces nucléotides sont déterminés en effectuant des tests de catalyse avec différentes combinaisons de nucléotides modifiés. D'autres modifications peuvent être réalisées sur les 2'-hydroxyles en combinaison avec les 2'-O-méthyle, comme des modifications 2'-amino, 2'-C-allyle, 2'-fluoro, 2'-méthylène et plus encore (voir figure 5.5) [220, 221]. La modification des nucléotides pour rendre les ribozymes résistants aux nucléases comporte un deuxième avantage, celui de minimiser les chances du ribozyme de faire des interactions non spécifiques pouvant causer de la toxicité. Pour augmenter la résistance aux nucléases, il est aussi possible d'utiliser les PNA, qui sont des acides nucléiques reliés par des liens peptidiques au lieu du squelette ribose-phosphate [222]. Ils sont résistants aux nucléases et aux protéases, en plus de posséder une grande affinité pour l'ARN et l'ADN [223, 224]. Une autre modification permettant d'augmenter la stabilité de l'ARN est l'introduction de quelques acides nucléiques bloqués (LNA, *locked nucleic acid*), qui contiennent une liaison méthylène dans le ribose entre le 2'-O et le 4'C [225-228]. Il est aussi possible d'utiliser des acides nucléiques avec un L-ribose (images miroirs des D-riboses) pour former des ribozymes appelés spiegelzymes, qui sont très stables et non toxiques [229, 230]. En plus de la modification des nucléotides, l'ajout d'un sucre inversé à l'extrémité 3' du ribozyme, ainsi que l'ajout de quelques phosphorothioates en 5' permettent d'augmenter la résistance aux nucléases [133, 231].

L'utilisation de plusieurs de ces stratégies permet d'obtenir la stabilité désirée *in vivo* pour augmenter la durée de vie des ribozymes.

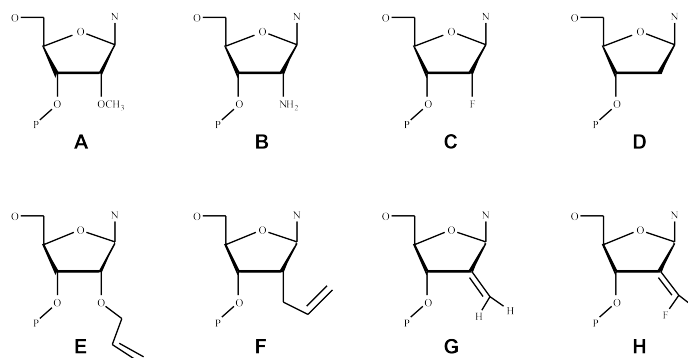


Figure 5.5 Modifications au sucre des nucléotides pour augmenter la résistance aux nucléases chez le ribozyme hammerhead.

Les groupements 2'-hydroxyyles des nucléotides sont remplacés par des groupes (A) 2'-méthyle, (B) 2'-amino, (C) 2'-fluoro, (D) 2'-déoxy, (E) 2'-O allyle, (F) 2'-allyle, (G) 2'-méthylène et (H) 2'-difluorométhylène. (Adaptée de [147]).

5.3.3.2 Livraison à proximité de la cible

Une des contraintes à l'utilisation des ribozymes comme agents thérapeutiques est de pouvoir faire une livraison efficace des ribozymes à proximité des ARN qu'ils ciblent. Il existe deux sortes de livraison, exogène et endogène [146]. La livraison exogène implique l'insertion de ribozymes préformés dans les cellules, tandis que la livraison endogène implique l'insertion d'un ribozyme encodé dans un gène pour produire le ribozyme dans la cellule. Dans la livraison exogène, les ribozymes sont plus exposés aux nucléases cellulaires et extracellulaires et il faut effectuer des modifications afin de les rendre plus tolérants aux nucléases, comme décrites à la section précédente (5.3.3.1). La livraison exogène peut aussi se faire à l'aide de véhicules à base de lipides ou de peptides, pour les protéger des différents environnements qui pourraient les dégrader. Cependant, l'utilisation de ces véhicules est facultative, puisque plusieurs exemples ont montré que les ribozymes peuvent bien fonctionner avec seulement une injection locale près de la cible [133, 146, 147].

Le deuxième mode de livraison est la livraison endogène dans laquelle le ribozyme est encodé dans le gène d'un vecteur d'expression, qui va par la suite être introduit dans la cellule par transfection [146]. Il y a plusieurs facteurs à tenir en compte avec la livraison endogène : la livraison du vecteur au bon endroit, la transcription efficace du ribozyme à partir du gène, la stabilisation du ribozyme contre la dégradation, le repliement du ribozyme et la localisation avec la cible. Les vecteurs viraux sont les plus utilisés pour la livraison endogène, car ils ont une grande efficacité de transfection et une intégration stable dans le génome pour les cellules divisibles. Pour les cellules post-mitotiques, la thérapie *ex vivo* est utilisée, c'est-à-dire, retirer les cellules et les infecter en culture cellulaire, pour ensuite les retourner dans le tissu d'origine. Les vecteurs d'expression doivent inclure un bon promoteur, un élément de stabilité et un élément de localisation. Le ribozyme doit aussi avoir un bon repliement, pour que les divers éléments du vecteur d'expression ne l'empêchent pas de se replier dans sa conformation active [146]. La livraison endogène a été utilisée avec le ribozyme anti-VIH OZ1, avec un vecteur qui est basé sur le virus *Moloney murine leukemia*, qui est inséré par transduction *ex vivo* dans des lymphocytes CD4⁺ T ou dans des cellules souches hématopoïétiques CD4⁺ et qui sont ensuite réinfusés chez les patients [132].

5.3.4 Utilisation du ribozyme VS comme outil biochimique

Dans les dernières sections, la discussion sur l'utilisation du ribozyme VS est orientée vers l'utilisation du ribozyme comme agent thérapeutique, mais il pourrait aussi être développé à d'autres fins, comme le développement d'outils biochimiques ou de diagnostics.

Le ribozyme VS est couramment utilisé comme enzyme de restriction pour enlever l'hétérogénéité en 3' des ARN transcrits *in vitro* [232-235]. Il suffit d'ajouter par PCR la séquence du substrat SLI à l'extrémité 3' de l'ARN à cliver et d'ajouter le ribozyme VS dans la transcription pour que le clivage s'effectue. Le ribozyme VS est idéal pour cette fonction puisqu'il n'y a pas de séquence minimale requise en 5' du site de clivage (mais le clivage est moins rapide avec une cytosine immédiatement en 5') [77]. Le ribozyme VS peut être ajouté à la transcription ou être transcrit en même temps que l'ARN à cliver.

Le ribozyme VS pourrait aussi être développé pour être utilisé comme système rapporteur, pour permettre le dépistage de composés ou pour suivre l'activité d'une enzyme. Les ribozymes sont appropriés pour ces fonctions, puisqu'ils peuvent être modifiés de sorte que leur activité soit dépendante de la liaison d'un ligand. Pour ce faire, le ribozyme autoclivant est couplé à un aptamère, qui va lier un ligand et la liaison va provoquer l'activité de clivage du ribozyme. Le ribozyme-aptamère peut être inséré dans l'ARNm d'une protéine rapporteuse et le clivage du ribozyme va empêcher la transcription du rapporteur ou à l'inverse la liaison du ligand va empêcher le clivage du ribozyme et permettre à la transcription de s'effectuer [236, 237]. Par exemple, il est possible de faire la synthèse d'un biosenseur, qui permet de dépister un métabolite avec un ribozyme-aptamère localisé dans le 3'UTR du gène *gfp* [238]. La présence du métabolite inhibe le clivage du ribozyme et permet la production de protéines fluorescentes vertes (*green fluorescence protein*). La détection de la fluorescence permet de déceler les petites variations dans l'activité enzymatique liée au métabolite. Le ribozyme VS pourrait lui aussi être modifié pour qu'il soit contrôlé par un ligand, en y ajoutant une séquence aptamère qui en absence du ligand empêche le substrat SLI de former sa conformation active. En présence du ligand, le changement de conformation de l'aptamère libère la séquence de SLI et permet le repliement de SLI en tige-boucle active pour le clivage par le ribozyme VS (voir figure 5.6).

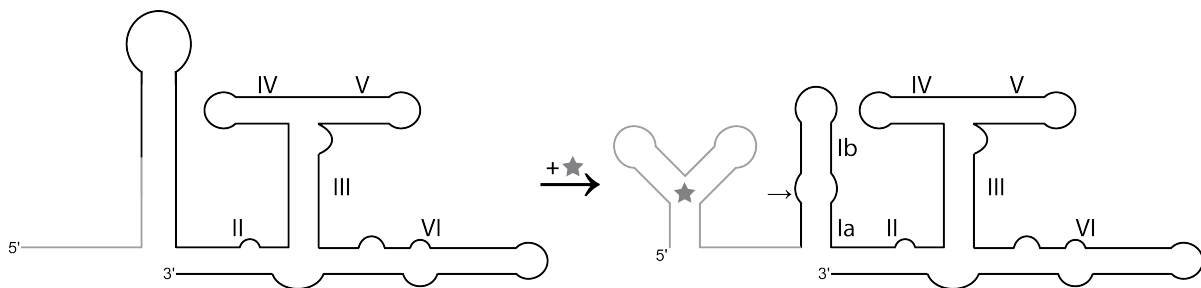


Figure 5.6 Exemple d'un ribozyme VS contrôlé par un ligand.

Structure schématique de la structure d'un ribozyme VS sous le contrôle d'un ligand. Les lignes noires représentent la séquence du ribozyme VS et les lignes grises représentent l'aptamère qui lie le ligand représenté par une étoile. Le ribozyme est inactif en absence du ligand et suite à la liaison du ligand, il y a un changement de conformation de l'aptamère, qui permet la libération du substrat SLI et peut maintenant être clivé par le ribozyme VS.

5.4 Conclusion générale

Cette thèse a démontré que le ribozyme VS est un bon candidat pour l'ingénierie de ribozyme et elle a permis d'établir des règles de base sur la façon d'adapter le ribozyme VS pour la reconnaissance de substrats modifiés, mais contenant toujours la boucle interne du site de clivage. Les études présentées dans cette thèse représentent les prémices pour des études plus approfondies d'ingénierie du ribozyme VS, pour la reconnaissance et le clivage d'ARN cibles complètement différents du substrat naturel, de façon spécifique et efficace. De plus, elles apportent des informations importantes sur l'adaptabilité de l'interaction *kissing-loop I/V* pour permettre une activité efficace du ribozyme VS. De façon plus générale, ces études fondamentales permettent de mieux comprendre les principes de base d'ingénierie de l'ARN. Nos études d'ingénieries permettent la possibilité de reconnaître et cliver des ARN repliés en tige-boucles qui sont difficilement reconnaissables par les autres ribozymes autoclivants, les antisens et les ARNi. Il serait intéressant d'intégrer le ribozyme VS à la panoplie d'outils exploités en ce moment en vue de faire de la thérapie génique, dans le but d'obtenir un effet synergique en les combinant pour cibler différentes régions des ARN à inactiver.

Bibliographie

1. Guerrier-Takada, C., et al., *The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme*. Cell, 1983. **35**(3 Pt 2): p. 849-57.
2. Crick, F.H., *The origin of the genetic code*. J Mol Biol, 1968. **38**(3): p. 367-79.
3. Nissen, P., et al., *The structural basis of ribosome activity in peptide bond synthesis*. Science, 2000. **289**(5481): p. 920-30.
4. Chow, J.C., et al., *Silencing of the mammalian X chromosome*. Annu Rev Genomics Hum Genet, 2005. **6**: p. 69-92.
5. Brown, C.J., et al., *The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus*. Cell, 1992. **71**(3): p. 527-42.
6. Autexier, C. and N.F. Lue, *The structure and function of telomerase reverse transcriptase*. Annu Rev Biochem, 2006. **75**: p. 493-517.
7. Cech, T.R., *Self-splicing of group I introns*. Annu Rev Biochem, 1990. **59**: p. 543-68.
8. Lehmann, K. and U. Schmidt, *Group II introns: structure and catalytic versatility of large natural ribozymes*. Crit Rev Biochem Mol Biol, 2003. **38**(3): p. 249-303.
9. Matera, A.G., R.M. Terns, and M.P. Terns, *Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs*. Nat Rev Mol Cell Biol, 2007. **8**(3): p. 209-20.
10. Bartel, D.P., *MicroRNAs: genomics, biogenesis, mechanism, and function*. Cell, 2004. **116**(2): p. 281-97.
11. Malone, C.D. and G.J. Hannon, *Small RNAs as guardians of the genome*. Cell, 2009. **136**(4): p. 656-68.
12. Hammond, S.M., *Dicing and slicing: the core machinery of the RNA interference pathway*. FEBS Lett, 2005. **579**(26): p. 5822-9.
13. Katayama, S., et al., *Antisense transcription in the mammalian transcriptome*. Science, 2005. **309**(5740): p. 1564-6.
14. Tucker, B.J. and R.R. Breaker, *Riboswitches as versatile gene control elements*. Curr Opin Struct Biol, 2005. **15**(3): p. 342-8.
15. Sorek, R., V. Kunin, and P. Hugenholtz, *CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea*. Nat Rev Microbiol, 2008. **6**(3): p. 181-6.
16. Seto, A.G., R.E. Kingston, and N.C. Lau, *The coming of age for Piwi proteins*. Mol Cell, 2007. **26**(5): p. 603-9.
17. Sloan, C.A., et al., *ENCODE data at the ENCODE portal*. Nucleic Acids Res, 2015. **44**(d1): p. d726-d732
18. Consortium, E.P., *The ENCODE (ENCyclopedia Of DNA Elements) Project*. Science, 2004. **306**(5696): p. 636-40.
19. Cech, T.R. and J.A. Steitz, *The noncoding RNA revolution--trashing old rules to forge new ones*. Cell, 2014. **157**(1): p. 77-94.
20. Mattick, J.S. and I.V. Makunin, *Non-coding RNA*. Hum Mol Genet, 2006. **15 Spec No 1**: p. R17-29.

21. Will, C.L. and R. Luhrmann, *Spliceosome structure and function*. Cold Spring Harb Perspect Biol, 2011. **3**(7).
22. Woese, C.R., *The genetic code: the molecular basis for genetic expression*. Vol. 55. 1967: Harper & Row.
23. Orgel, L.E., *Evolution of the genetic apparatus*. J Mol Biol, 1968. **38**(3): p. 381-93.
24. Kruger, K., et al., *Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena*. Cell, 1982. **31**(1): p. 147-57.
25. Chen, X., N. Li, and A.D. Ellington, *Ribozyme catalysis of metabolism in the RNA world*. Chem Biodivers, 2007. **4**(4): p. 633-55.
26. Martin, L.L., P.J. Unrau, and U.F. Muller, *RNA synthesis by in vitro selected ribozymes for recreating an RNA world*. Life (Basel), 2015. **5**(1): p. 247-68.
27. Cheng, L.K. and P.J. Unrau, *Closing the circle: replicating RNA with RNA*. Cold Spring Harb Perspect Biol, 2010. **2**(10): p. a002204.
28. Meyer, A.J., J.W. Ellefson, and A.D. Ellington, *Abiotic self-replication*. Acc Chem Res, 2012. **45**(12): p. 2097-105.
29. Breaker, R.R., *Riboswitches and the RNA world*. Cold Spring Harb Perspect Biol, 2012. **4**(2): p. 1-15.
30. Pressman, A., C. Blanco, and I.A. Chen, *The RNA World as a Model System to Study the Origin of Life*. Curr Biol, 2015. **25**(19): p. R953-63.
31. Penchovsky, R., *Computational design of allosteric ribozymes as molecular biosensors*. Biotechnol Adv, 2014. **32**(5): p. 1015-27.
32. Weinberg, Z., et al., *New classes of self-cleaving ribozymes revealed by comparative genomics analysis*. Nat Chem Biol, 2015. **11**(8): p. 606-10.
33. Jimenez, R.M., J.A. Polanco, and A. Luptak, *Chemistry and Biology of Self-Cleaving Ribozymes*. Trends Biochem Sci, 2015. **40**(11): p. 648-661.
34. Cochrane, J.C. and S.A. Strobel, *Catalytic strategies of self-cleaving ribozymes*. Acc Chem Res, 2008. **41**(8): p. 1027-35.
35. Fedor, M.J. and J.R. Williamson, *The catalytic diversity of RNAs*. Nat Rev Mol Cell Biol, 2005. **6**(5): p. 399-412.
36. Fedor, M.J., *Comparative enzymology and structural biology of RNA self-cleavage*. Annu Rev Biophys, 2009. **38**: p. 271-99.
37. Emilsson, G.M., et al., *Ribozyme speed limits*. RNA, 2003. **9**(8): p. 907-18.
38. Johnson-Buck, A.E., S.E. McDowell, and N.G. Walter, *Metal ions: supporting actors in the playbook of small ribozymes*. Met Ions Life Sci, 2011. **9**: p. 175-96.
39. Prody, G.A., et al., *Autolytic processing of dimeric plant virus satellite RNA*. Science, 1986. **231**(4745): p. 1577-80.
40. Hammann, C., et al., *The ubiquitous hammerhead ribozyme*. RNA, 2012. **18**(5): p. 871-85.
41. Jimenez, R.M., E. Delwart, and A. Luptak, *Structure-based search reveals hammerhead ribozymes in the human microbiome*. J Biol Chem, 2011. **286**(10): p. 7737-43.
42. Perreault, J., et al., *Identification of hammerhead ribozymes in all domains of life reveals novel structural variations*. PLoS Comput Biol, 2011. **7**(5): p. e1002031.
43. Martick, M., et al., *A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA*. Nature, 2008. **454**(7206): p. 899-902.

44. De la Pena, M., S. Gago, and R. Flores, *Peripheral regions of natural hammerhead ribozymes greatly increase their self-cleavage activity*. EMBO J, 2003. **22**(20): p. 5561-70.
45. Khvorova, A., et al., *Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity*. Nat Struct Biol, 2003. **10**(9): p. 708-12.
46. Martick, M. and W.G. Scott, *Tertiary contacts distant from the active site prime a ribozyme for catalysis*. Cell, 2006. **126**(2): p. 309-20.
47. Rubino, L., et al., *Nucleotide sequence and structural analysis of two satellite RNAs associated with chicory yellow mottle virus*. J Gen Virol, 1990. **71** (Pt 9): p. 1897-903.
48. Kaper, J.M., M.E. Tousignant, and G. Steger, *Nucleotide sequence predicts circularity and self-cleavage of 300-ribonucleotide satellite of arabis mosaic virus*. Biochem Biophys Res Commun, 1988. **154**(1): p. 318-25.
49. Hampel, A. and R. Tritz, *RNA catalytic properties of the minimum (-)sTRSV sequence*. Biochemistry, 1989. **28**(12): p. 4929-33.
50. Fedor, M.J., *Structure and function of the hairpin ribozyme*. J Mol Biol, 2000. **297**(2): p. 269-91.
51. Wilson, T.J., et al., *Nucleobase catalysis in the hairpin ribozyme*. RNA, 2006. **12**(6): p. 980-7.
52. Winkler, W.C., et al., *Control of gene expression by a natural metabolite-responsive ribozyme*. Nature, 2004. **428**(6980): p. 281-6.
53. McCarthy, T.J., et al., *Ligand requirements for glmS ribozyme self-cleavage*. Chem Biol, 2005. **12**(11): p. 1221-6.
54. Klein, D.J. and A.R. Ferre-D'Amare, *Structural basis of glmS ribozyme activation by glucosamine-6-phosphate*. Science, 2006. **313**(5794): p. 1752-6.
55. Ferre-D'Amare, A.R., *The glmS ribozyme: use of a small molecule coenzyme by a gene-regulatory RNA*. Q Rev Biophys, 2010. **43**(4): p. 423-47.
56. Gong, B., et al., *The glmS ribozyme tunes the catalytically critical pK(a) of its coenzyme glucosamine-6-phosphate*. J Am Chem Soc, 2011. **133**(36): p. 14188-91.
57. Viladoms, J. and M.J. Fedor, *The glmS ribozyme cofactor is a general acid-base catalyst*. J Am Chem Soc, 2012. **134**(46): p. 19043-9.
58. Kuo, M.Y., et al., *Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus*. J Virol, 1988. **62**(12): p. 4439-44.
59. Wu, H.N., et al., *Human hepatitis delta virus RNA subfragments contain an autocleavage activity*. Proc Natl Acad Sci U S A, 1989. **86**(6): p. 1831-5.
60. Salehi-Ashtiani, K., et al., *A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene*. Science, 2006. **313**(5794): p. 1788-92.
61. Shih, I.H. and M.D. Been, *Catalytic strategies of the hepatitis delta virus ribozymes*. Annu Rev Biochem, 2002. **71**: p. 887-917.
62. Chen, J.H., et al., *A 1.9 Å crystal structure of the HDV ribozyme precleavage suggests both Lewis acid and general acid mechanisms contribute to phosphodiester cleavage*. Biochemistry, 2010. **49**(31): p. 6508-18.
63. Ferre-D'Amare, A.R., K. Zhou, and J.A. Doudna, *Crystal structure of a hepatitis delta virus ribozyme*. Nature, 1998. **395**(6702): p. 567-74.
64. Collins, R.A. and B.J. Saville, *Independent transfer of mitochondrial chromosomes and plasmids during unstable vegetative fusion in Neurospora*. Nature, 1990. **345**(6271): p. 177-9.

65. Rastogi, T., et al., *A long-range pseudoknot is required for activity of the Neurospora VS ribozyme*. EMBO J, 1996. **15**(11): p. 2820-5.
66. Suslov, N.B., et al., *Crystal structure of the Varkud satellite ribozyme*. Nat Chem Biol, 2015. **11**(11): p. 840-6.
67. Guo, H.C. and R.A. Collins, *Efficient trans-cleavage of a stem-loop RNA substrate by a ribozyme derived from neurospora VS RNA*. EMBO J, 1995. **14**(2): p. 368-76.
68. Wilson, T.J., et al., *Nucleobase-mediated general acid-base catalysis in the Varkud satellite ribozyme*. Proc Natl Acad Sci U S A, 2010. **107**(26): p. 11751-6.
69. Roth, A., et al., *A widespread self-cleaving ribozyme class is revealed by bioinformatics*. Nat Chem Biol, 2014. **10**(1): p. 56-60.
70. Liu, Y., et al., *Crystal structure and mechanistic investigation of the twister ribozyme*. Nat Chem Biol, 2014. **10**(9): p. 739-44.
71. Ren, A., et al., *In-line alignment and Mg(2)(+) coordination at the cleavage site of the env22 twister ribozyme*. Nat Commun, 2014. **5**: p. 5534.
72. Eiler, D., J. Wang, and T.A. Steitz, *Structural basis for the fast self-cleavage reaction catalyzed by the twister ribozyme*. Proc Natl Acad Sci U S A, 2014. **111**(36): p. 13028-33.
73. Kosutic, M., et al., *A Mini-Twister Variant and Impact of Residues/Cations on the Phosphodiester Cleavage of this Ribozyme Class*. Angew Chem Int Ed Engl, 2015. **54**(50): p. 15128-15133.
74. Harris, K.A., et al., *Biochemical analysis of pistol self-cleaving ribozymes*. RNA, 2015. **21**(11): p. 1852-8.
75. Li, S., et al., *Biochemical analysis of hatchet self-cleaving ribozymes*. RNA, 2015. **21**(11): p. 1845-51.
76. Saville, B.J. and R.A. Collins, *A site-specific self-cleavage reaction performed by a novel RNA in Neurospora mitochondria*. Cell, 1990. **61**(4): p. 685-96.
77. Guo, H.C., et al., *Nucleotide sequence requirements for self-cleavage of Neurospora VS RNA*. J Mol Biol, 1993. **232**(2): p. 351-61.
78. Saville, B.J. and R.A. Collins, *RNA-mediated ligation of self-cleavage products of a Neurospora mitochondrial plasmid transcript*. Proc Natl Acad Sci U S A, 1991. **88**(19): p. 8826-30.
79. Kennell, J.C., et al., *The VS catalytic RNA replicates by reverse transcription as a satellite of a retroplasmid*. Genes Dev, 1995. **9**(3): p. 294-303.
80. Beattie, T.L., J.E. Olive, and R.A. Collins, *A secondary-structure model for the self-cleaving region of Neurospora VS RNA*. Proc Natl Acad Sci U S A, 1995. **92**(10): p. 4686-90.
81. Jones, F.D., S.P. Ryder, and S.A. Strobel, *An efficient ligation reaction promoted by a Varkud Satellite ribozyme with extended 5'- and 3'-termini*. Nucleic Acids Res, 2001. **29**(24): p. 5115-20.
82. Brunel, C., et al., *RNA loop-loop interactions as dynamic functional motifs*. Biochimie, 2002. **84**(9): p. 925-44.
83. Hiley, S.L. and R.A. Collins, *Rapid formation of a solvent-inaccessible core in the Neurospora Varkud satellite ribozyme*. EMBO J, 2001. **20**(19): p. 5461-9.
84. Andersen, A.A. and R.A. Collins, *Rearrangement of a stable RNA secondary structure during VS ribozyme catalysis*. Mol Cell, 2000. **5**(3): p. 469-78.

85. Andersen, A.A. and R.A. Collins, *Intramolecular secondary structure rearrangement by the kissing interaction of the Neurospora VS ribozyme*. Proc Natl Acad Sci U S A, 2001. **98**(14): p. 7730-5.
86. Rastogi, T. and R.A. Collins, *Smaller, faster ribozymes reveal the catalytic core of Neurospora VS RNA*. J Mol Biol, 1998. **277**(2): p. 215-24.
87. Bouchard, P. and P. Legault, *Structural insights into substrate recognition by the Neurospora Varkud satellite ribozyme: importance of U-turns at the kissing-loop junction*. Biochemistry, 2014. **53**(1): p. 258-69.
88. Campbell, D.O., et al., *NMR structure of varkud satellite ribozyme stem-loop V in the presence of magnesium ions and localization of metal-binding sites*. Biochemistry, 2006. **45**(35): p. 10591-605.
89. Bouchard, P., et al., *Role of SLV in SLI substrate recognition by the Neurospora VS ribozyme*. RNA, 2008. **14**(4): p. 736-48.
90. Bouchard, P. and P. Legault, *A remarkably stable kissing-loop interaction defines substrate recognition by the Neurospora Varkud Satellite ribozyme*. RNA, 2014. **20**(9): p. 1451-64.
91. Michiels, P.J., et al., *Structure of the ribozyme substrate hairpin of Neurospora VS RNA: a close look at the cleavage site*. RNA, 2000. **6**(12): p. 1821-32.
92. Flinders, J. and T. Dieckmann, *A pH controlled conformational switch in the cleavage site of the VS ribozyme substrate RNA*. J Mol Biol, 2001. **308**(4): p. 665-79.
93. Hoffmann, B., et al., *NMR structure of the active conformation of the Varkud satellite ribozyme cleavage site*. Proc Natl Acad Sci U S A, 2003. **100**(12): p. 7003-8.
94. Zamel, R. and R.A. Collins, *Rearrangement of substrate secondary structure facilitates binding to the Neurospora VS ribozyme*. J Mol Biol, 2002. **324**(5): p. 903-15.
95. Lilley, D.M., *The Varkud satellite ribozyme*. RNA, 2004. **10**(2): p. 151-8.
96. Hiley, S.L., et al., *4-thio-U cross-linking identifies the active site of the VS ribozyme*. EMBO J, 2002. **21**(17): p. 4691-8.
97. Sood, V.D., T.L. Beattie, and R.A. Collins, *Identification of phosphate groups involved in metal binding and tertiary interactions in the core of the Neurospora VS ribozyme*. J Mol Biol, 1998. **282**(4): p. 741-50.
98. Sood, V.D. and R.A. Collins, *Identification of the catalytic subdomain of the VS ribozyme and evidence for remarkable sequence tolerance in the active site loop*. J Mol Biol, 2002. **320**(3): p. 443-54.
99. Lafontaine, D.A., et al., *The A730 loop is an important component of the active site of the VS ribozyme*. J Mol Biol, 2001. **312**(4): p. 663-74.
100. McLeod, A.C. and D.M. Lilley, *Efficient, pH-dependent RNA ligation by the VS ribozyme in trans*. Biochemistry, 2004. **43**(4): p. 1118-25.
101. Jones, F.D. and S.A. Strobel, *Ionization of a critical adenosine residue in the neurospora Varkud Satellite ribozyme active site*. Biochemistry, 2003. **42**(14): p. 4265-76.
102. Zhao, Z.Y., et al., *Nucleobase participation in ribozyme catalysis*. J Am Chem Soc, 2005. **127**(14): p. 5026-7.
103. Suydam, I.T. and S.A. Strobel, *Fluorine substituted adenosines as probes of nucleobase protonation in functional RNAs*. J Am Chem Soc, 2008. **130**(41): p. 13639-48.

104. Lafontaine, D.A., et al., *Functional group requirements in the probable active site of the VS ribozyme*. J Mol Biol, 2002. **323**(1): p. 23-34.
105. Wilson, T.J., A.C. McLeod, and D.M. Lilley, *A guanine nucleobase important for catalysis by the VS ribozyme*. EMBO J, 2007. **26**(10): p. 2489-500.
106. Jaikaran, D., et al., *An important role of G638 in the cis-cleavage reaction of the Neurospora VS ribozyme revealed by a novel nucleotide analog incorporation method*. RNA, 2008. **14**(5): p. 938-49.
107. Wilson, T.J. and D.M. Lilley, *Do the hairpin and VS ribozymes share a common catalytic mechanism based on general acid-base catalysis? A critical assessment of available experimental data*. RNA, 2011. **17**(2): p. 213-21.
108. Collins, R.A. and J.E. Olive, *Reaction conditions and kinetics of self-cleavage of a ribozyme derived from Neurospora VS RNA*. Biochemistry, 1993. **32**(11): p. 2795-9.
109. Murray, J.B., et al., *The hammerhead, hairpin and VS ribozymes are catalytically proficient in monovalent cations alone*. Chem Biol, 1998. **5**(10): p. 587-95.
110. Maguire, J.L. and R.A. Collins, *Effects of cobalt hexammine on folding and self-cleavage of the Neurospora VS ribozyme*. J Mol Biol, 2001. **309**(1): p. 45-56.
111. Olive, J.E. and R.A. Collins, *Spermine switches a Neurospora VS ribozyme from slow Cis cleavage to fast trans cleavage*. Biochemistry, 1998. **37**(18): p. 6476-84.
112. Olive, J.E., et al., *Enhancement of Neurospora VS ribozyme cleavage by tuberactinomycin antibiotics*. EMBO J, 1995. **14**(13): p. 3247-51.
113. Lafontaine, D.A., D.G. Norman, and D.M. Lilley, *The global structure of the VS ribozyme*. EMBO J, 2002. **21**(10): p. 2461-71.
114. Lafontaine, D.A., D.G. Norman, and D.M. Lilley, *Structure, folding and activity of the VS ribozyme: importance of the 2-3-6 helical junction*. EMBO J, 2001. **20**(6): p. 1415-24.
115. Collins, R.A., *The Neurospora Varkud satellite ribozyme*. Biochem Soc Trans, 2002. **30**(Pt 6): p. 1122-6.
116. Lipfert, J., et al., *The complete VS ribozyme in solution studied by small-angle X-ray scattering*. Structure, 2008. **16**(9): p. 1357-67.
117. Bonneau, E., et al., *The NMR structure of the II-III-VI three-way junction from the Neurospora VS ribozyme reveals a critical tertiary interaction and provides new insights into the global ribozyme structure*. RNA, 2015. **21**(9): p. 1621-32.
118. Campbell, D.O. and P. Legault, *Nuclear magnetic resonance structure of the Varkud satellite ribozyme stem-loop V RNA and magnesium-ion binding from chemical-shift mapping*. Biochemistry, 2005. **44**(11): p. 4157-70.
119. Desjardins, G., et al., *NMR structure of the A730 loop of the Neurospora VS ribozyme: insights into the formation of the active site*. Nucleic Acids Res, 2011. **39**(10): p. 4427-37.
120. Flinders, J. and T. Dieckmann, *The solution structure of the VS ribozyme active site loop reveals a dynamic "hot-spot"*. J Mol Biol, 2004. **341**(4): p. 935-49.
121. Bonneau, E. and P. Legault, *Nuclear magnetic resonance structure of the III-IV-V three-way junction from the Varkud satellite ribozyme and identification of magnesium-binding sites using paramagnetic relaxation enhancement*. Biochemistry, 2014. **53**(39): p. 6264-75.

122. Poon, A.H., et al., *Identification of separate structural features that affect rate and cation concentration dependence of self-cleavage by the Neurospora VS ribozyme*. *Biochemistry*, 2006. **45**(44): p. 13394-400.
123. Zamel, R., et al., *Exceptionally fast self-cleavage by a Neurospora Varkud satellite ribozyme*. *Proc Natl Acad Sci U S A*, 2004. **101**(6): p. 1467-72.
124. Liang, J.C., R.J. Bloom, and C.D. Smolke, *Engineering biological systems with synthetic RNA molecules*. *Mol Cell*, 2011. **43**(6): p. 915-26.
125. Kobayashi, H., et al., *Safety and pharmacokinetic study of RPI.4610 (ANGIOZYME), an anti-VEGFR-1 ribozyme, in combination with carboplatin and paclitaxel in patients with advanced solid tumors*. *Cancer Chemother Pharmacol*, 2005. **56**(4): p. 329-36.
126. Mulhbacher, J., P. St-Pierre, and D.A. Lafontaine, *Therapeutic applications of ribozymes and riboswitches*. *Curr Opin Pharmacol*, 2010. **10**(5): p. 551-6.
127. Khan, A.U., *Ribozyme: a clinical tool*. *Clin Chim Acta*, 2006. **367**(1-2): p. 20-7.
128. Tang, J. and R.R. Breaker, *Rational design of allosteric ribozymes*. *Chem Biol*, 1997. **4**(6): p. 453-9.
129. Breaker, R.R., *Engineered allosteric ribozymes as biosensor components*. *Curr Opin Biotechnol*, 2002. **13**(1): p. 31-9.
130. Pavco, P.A., et al., *Antitumor and antimetastatic activity of ribozymes targeting the messenger RNA of vascular endothelial growth factor receptors*. *Clin Cancer Res*, 2000. **6**(5): p. 2094-103.
131. Sandberg, J.A., et al., *Pharmacokinetics and tolerability of an antiangiogenic ribozyme (ANGIOZYME) in healthy volunteers*. *J Clin Pharmacol*, 2000. **40**(12 Pt 2): p. 1462-9.
132. Mitsuyasu, R.T., et al., *Phase 2 gene therapy trial of an anti-HIV ribozyme in autologous CD34+ cells*. *Nat Med*, 2009. **15**(3): p. 285-92.
133. Usman, N. and L.M. Blatt, *Nuclease-resistant synthetic ribozymes: developing a new class of therapeutics*. *J Clin Invest*, 2000. **106**(10): p. 1197-202.
134. Macejak, D.G., et al., *Inhibition of hepatitis C virus (HCV)-RNA-dependent translation and replication of a chimeric HCV poliovirus using synthetic stabilized ribozymes*. *Hepatology*, 2000. **31**(3): p. 769-76.
135. Welz, R., et al., *Site-directed alteration of RNA sequence mediated by an engineered twin ribozyme*. *Angew Chem Int Ed Engl*, 2003. **42**(21): p. 2424-7.
136. Schmidt, C., R. Welz, and S. Muller, *RNA double cleavage by a hairpin-derived twin ribozyme*. *Nucleic Acids Res*, 2000. **28**(4): p. 886-94.
137. Vauleon, S., et al., *Site-specific fluorescent and affinity labelling of RNA by using a small engineered twin ribozyme*. *Chembiochem*, 2005. **6**(12): p. 2158-62.
138. Chen, C.J., et al., *Multitarget-ribozyme directed to cleave at up to nine highly conserved HIV-1 env RNA regions inhibits HIV-1 replication--potential effectiveness against most presently sequenced HIV-1 isolates*. *Nucleic Acids Res*, 1992. **20**(17): p. 4581-9.
139. Neupert, J., D. Karcher, and R. Bock, *Design of simple synthetic RNA thermometers for temperature-controlled gene expression in Escherichia coli*. *Nucleic Acids Res*, 2008. **36**(19): p. e124.
140. Saragliadis, A., et al., *Thermozymes: Synthetic RNA thermometers based on ribozyme activity*. *RNA Biol*, 2013. **10**(6): p. 1010-6.
141. Gold, L., et al., *Diversity of oligonucleotide functions*. *Annu Rev Biochem*, 1995. **64**: p. 763-97.

142. Ellington, A.D. and J.W. Szostak, *In vitro selection of RNA molecules that bind specific ligands*. Nature, 1990. **346**(6287): p. 818-22.
143. Tuerk, C. and L. Gold, *Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase*. Science, 1990. **249**(4968): p. 505-10.
144. Ng, E.W., et al., *Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease*. Nat Rev Drug Discov, 2006. **5**(2): p. 123-32.
145. Dotu, I., et al., *Complete RNA inverse folding: computational design of functional hammerhead ribozymes*. Nucleic Acids Res, 2014. **42**(18): p. 11752-62.
146. Bramlage, B., E. Luzi, and F. Eckstein, *Designing ribozymes for the inhibition of gene expression*. Trends Biotechnol, 1998. **16**(10): p. 434-8.
147. Citti, L. and G. Rainaldi, *Synthetic hammerhead ribozymes as therapeutic tools to control disease genes*. Curr Gene Ther, 2005. **5**(1): p. 11-24.
148. Mercatanti, A., et al., *A method for prediction of accessible sites on an mRNA sequence for target selection of hammerhead ribozymes*. J Comput Biol, 2002. **9**(4): p. 641-53.
149. Ojwang, J.O., et al., *Inhibition of human immunodeficiency virus type 1 expression by a hairpin ribozyme*. Proc Natl Acad Sci U S A, 1992. **89**(22): p. 10802-6.
150. Yamada, O., et al., *Activity and cleavage site specificity of an anti-HIV-1 hairpin ribozyme in human T cells*. Virology, 1994. **205**(1): p. 121-6.
151. Joseph, S., et al., *Substrate selection rules for the hairpin ribozyme determined by in vitro selection, mutation, and analysis of mismatched substrates*. Genes Dev, 1993. **7**(1): p. 130-8.
152. Muller, S., et al., *The many faces of the hairpin ribozyme: structural and functional variants of a small catalytic RNA*. IUBMB Life, 2012. **64**(1): p. 36-47.
153. Nomura, Y., et al., *Controlling mammalian gene expression by allosteric hepatitis delta virus ribozymes*. ACS Synth Biol, 2013. **2**(12): p. 684-9.
154. Kertsburg, A. and G.A. Soukup, *A versatile communication module for controlling RNA folding and catalysis*. Nucleic Acids Res, 2002. **30**(21): p. 4599-606.
155. Beaudoin, J.D. and J.P. Perreault, *Potassium ions modulate a G-quadruplex-ribozyme's activity*. RNA, 2008. **14**(6): p. 1018-25.
156. Laine, S., et al., *In vitro and in vivo cleavage of HIV-1 RNA by new SOFA-HDV ribozymes and their potential to inhibit viral replication*. RNA Biol, 2011. **8**(2): p. 343-53.
157. Levesque, M.V., et al., *Investigating a new generation of ribozymes in order to target HCV*. PLoS One, 2010. **5**(3): p. e9627.
158. Bergeron, L.J. and J.P. Perreault, *Target-dependent on/off switch increases ribozyme fidelity*. Nucleic Acids Res, 2005. **33**(4): p. 1240-8.
159. Bergeron, L.J., C. Reymond, and J.P. Perreault, *Functional characterization of the SOFA delta ribozyme*. RNA, 2005. **11**(12): p. 1858-68.
160. Paillart, J.C., et al., *Dimerization of retroviral RNA genomes: an inseparable pair*. Nat Rev Microbiol, 2004. **2**(6): p. 461-72.
161. Richter, S., H. Cao, and T.M. Rana, *Specific HIV-1 TAR RNA loop sequence and functional groups are required for human cyclin T1-Tat-TAR ternary complex formation*. Biochemistry, 2002. **41**(20): p. 6391-7.

162. McConnell, T.S., *Theoretical considerations in measuring reaction parameters*. Methods Mol Biol, 1997. **74**: p. 187-98.
163. Lacroix-Labonte, J., et al., *Helix-length compensation studies reveal the adaptability of the VS ribozyme architecture*. Nucleic Acids Res, 2012. **40**(5): p. 2284-93.
164. Merino, E.J., et al., *RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE)*. J Am Chem Soc, 2005. **127**(12): p. 4223-31.
165. Wilkinson, K.A., et al., *Influence of nucleotide identity on ribose 2'-hydroxyl reactivity in RNA*. RNA, 2009. **15**(7): p. 1314-21.
166. Saksmerprome, V., et al., *Artificial tertiary motifs stabilize trans-cleaving hammerhead ribozymes under conditions of submillimolar divalent ions and high temperatures*. RNA, 2004. **10**(12): p. 1916-24.
167. Jarvis, T.C., et al., *Ribozymes as tools for therapeutic target validation in arthritis*. J Immunol, 2000. **165**(1): p. 493-8.
168. Brion, P. and E. Westhof, *Hierarchy and dynamics of RNA folding*. Annu Rev Biophys Biomol Struct, 1997. **26**: p. 113-37.
169. Pyle, A.M. and J.B. Green, *RNA folding*. Curr Opin Struct Biol, 1995. **5**(3): p. 303-10.
170. Westhof, E., B. Masquida, and L. Jaeger, *RNA tectonics: towards RNA design*. Fold Des, 1996. **1**(4): p. R78-88.
171. Batey, R.T., R.P. Rambo, and J.A. Doudna, *Tertiary Motifs in RNA Structure and Folding*. Angew Chem Int Ed Engl, 1999. **38**(16): p. 2326-2343.
172. Hermann, T. and D.J. Patel, *Stitching together RNA tertiary architectures*. J Mol Biol, 1999. **294**(4): p. 829-49.
173. Darfeuille, F., et al., *Aptamers targeted to an RNA hairpin show improved specificity compared to that of complementary oligonucleotides*. Biochemistry, 2006. **45**(39): p. 12076-82.
174. Aldaz-Carroll, L., et al., *Apical loop-internal loop interactions: a new RNA-RNA recognition motif identified through in vitro selection against RNA hairpins of the hepatitis C virus mRNA*. Biochemistry, 2002. **41**(18): p. 5883-93.
175. Lunse, C.E., et al., *An aptamer targeting the apical-loop domain modulates pri-miRNA processing*. Angew Chem Int Ed Engl, 2010. **49**(27): p. 4674-7.
176. Scarabino, D., et al., *tRNA prefers to kiss*. EMBO J, 1999. **18**(16): p. 4571-8.
177. Lodmell, J.S., et al., *Convergence of natural and artificial evolution on an RNA loop-loop interaction: the HIV-1 dimerization initiation site*. RNA, 2000. **6**(9): p. 1267-76.
178. Duconge, F. and J.J. Toulme, *In vitro selection identifies key determinants for loop-loop interactions: RNA aptamers selective for the TAR RNA element of HIV-1*. RNA, 1999. **5**(12): p. 1605-14.
179. Ohuchi, S.P. and Y. Nakamura, *Slight sequence modifications unexpectedly alter the metal-dependency of a kissing-loop interaction*. Nucleic Acids Symp Ser (Oxf), 2007(51): p. 395-6.
180. Watrin, M., et al., *In vitro selection of RNA aptamers derived from a genomic human library against the TAR RNA element of HIV-1*. Biochemistry, 2009. **48**(26): p. 6278-84.
181. Boiziau, C., et al., *DNA aptamers selected against the HIV-1 trans-activation-responsive RNA element form RNA-DNA kissing complexes*. J Biol Chem, 1999. **274**(18): p. 12730-7.

182. Cho, B., et al., *Interacting RNA species identified by combinatorial selection*. Bioorg Med Chem, 1997. **5**(6): p. 1107-13.
183. Magee, T.V., et al., *Novel 3-O-carbamoyl erythromycin A derivatives (carbamolides) with activity against resistant staphylococcal and streptococcal isolates*. Bioorg Med Chem Lett, 2013. **23**(6): p. 1727-31.
184. Chang, K.Y. and I. Tinoco, Jr., *The structure of an RNA "kissing" hairpin complex of the HIV TAR hairpin loop and its complement*. J Mol Biol, 1997. **269**(1): p. 52-66.
185. Pleij, C.W., K. Rietveld, and L. Bosch, *A new principle of RNA folding based on pseudoknotting*. Nucleic Acids Res, 1985. **13**(5): p. 1717-31.
186. Jones, K.A. and B.M. Peterlin, *Control of RNA initiation and elongation at the HIV-1 promoter*. Annu Rev Biochem, 1994. **63**: p. 717-43.
187. Selby, M.J., et al., *Structure, sequence, and position of the stem-loop in tar determine transcriptional elongation by tat through the HIV-1 long terminal repeat*. Genes Dev, 1989. **3**(4): p. 547-58.
188. Garcia, J.A., et al., *Human immunodeficiency virus type 1 LTR TATA and TAR region sequences required for transcriptional regulation*. EMBO J, 1989. **8**(3): p. 765-78.
189. Jakobovits, A., et al., *A discrete element 3' of human immunodeficiency virus 1 (HIV-1) and HIV-2 mRNA initiation sites mediates transcriptional activation by an HIV trans activator*. Mol Cell Biol, 1988. **8**(6): p. 2555-61.
190. Kolb, G., et al., *Endogenous expression of an anti-TAR aptamer reduces HIV-1 replication*. RNA Biol, 2006. **3**(4): p. 150-6.
191. Kaushik, N., et al., *Anti-TAR polyamide nucleotide analog conjugated with a membrane-permeating peptide inhibits human immunodeficiency virus type 1 production*. J Virol, 2002. **76**(8): p. 3881-91.
192. Mei, H.Y., et al., *Inhibitors of protein-RNA complexation that target the RNA: specific recognition of human immunodeficiency virus type 1 TAR RNA by small organic molecules*. Biochemistry, 1998. **37**(40): p. 14204-12.
193. Huq, I., et al., *Controlling human immunodeficiency virus type 1 gene expression by unnatural peptides*. Biochemistry, 1999. **38**(16): p. 5172-7.
194. Klimkait, T., et al., *Rational optimization of a HIV-1 Tat inhibitor: rapid progress on combinatorial lead structures*. Biotechnol Bioeng, 1998. **61**(3): p. 155-68.
195. Turner, J.J., et al., *Cell-penetrating peptide conjugates of peptide nucleic acids (PNA) as inhibitors of HIV-1 Tat-dependent trans-activation in cells*. Nucleic Acids Res, 2005. **33**(21): p. 6837-49.
196. Haddrick, M., et al., *Evidence that a kissing loop structure facilitates genomic RNA dimerisation in HIV-1*. J Mol Biol, 1996. **259**(1): p. 58-68.
197. Lee, H.W., K.T. Briggs, and J.P. Marino, *Dissecting structural transitions in the HIV-1 dimerization initiation site RNA using 2-aminopurine fluorescence*. Methods, 2009. **49**(2): p. 118-27.
198. Ennifar, E., et al., *Targeting the dimerization initiation site of HIV-1 RNA with aminoglycosides: from crystal to cell*. Nucleic Acids Res, 2006. **34**(8): p. 2328-39.
199. Bernacchi, S., et al., *Aminoglycoside binding to the HIV-1 RNA dimerization initiation site: thermodynamics and effect on the kissing-loop to duplex conversion*. Nucleic Acids Res, 2007. **35**(21): p. 7128-39.
200. Ennifar, E., et al., *A structure-based approach for targeting the HIV-1 genomic RNA dimerization initiation site*. Biochimie, 2007. **89**(10): p. 1195-203.

201. Freisz, S., et al., *Binding of aminoglycoside antibiotics to the duplex form of the HIV-1 genomic RNA dimerization initiation site*. *Angew Chem Int Ed Engl*, 2008. **47**(22): p. 4110-3.
202. Parkash, B., et al., *Inhibition of 5'-UTR RNA conformational switching in HIV-1 using antisense PNAs*. *PLoS One*, 2012. **7**(11): p. e49310.
203. Clever, J.L., M.L. Wong, and T.G. Parslow, *Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA*. *J Virol*, 1996. **70**(9): p. 5902-8.
204. Choo, Q.L., et al., *Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome*. *Science*, 1989. **244**(4902): p. 359-62.
205. Qureshi, S.A., *Hepatitis C virus--biology, host evasion strategies, and promising new therapies on the horizon*. *Med Res Rev*, 2007. **27**(3): p. 353-73.
206. Miller, R.H. and R.H. Purcell, *Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups*. *Proc Natl Acad Sci U S A*, 1990. **87**(6): p. 2057-61.
207. Lee, C.H., J.H. Kim, and S.W. Lee, *Prospects for nucleic acid-based therapeutics against hepatitis C virus*. *World J Gastroenterol*, 2013. **19**(47): p. 8949-62.
208. Friebe, P., et al., *Sequences in the 5' nontranslated region of hepatitis C virus required for RNA replication*. *J Virol*, 2001. **75**(24): p. 12047-57.
209. Wang, C., P. Sarnow, and A. Siddiqui, *Translation of human hepatitis C virus RNA in cultured cells is mediated by an internal ribosome-binding mechanism*. *J Virol*, 1993. **67**(6): p. 3338-44.
210. Tsukiyama-Kohara, K., et al., *Internal ribosome entry site within hepatitis C virus RNA*. *J Virol*, 1992. **66**(3): p. 1476-83.
211. Honda, M., et al., *Structural requirements for initiation of translation by internal ribosome entry within genome-length hepatitis C virus RNA*. *Virology*, 1996. **222**(1): p. 31-42.
212. Prabhu, R., R.F. Garry, and S. Dash, *Small interfering RNA targeted to stem-loop II of the 5' untranslated region effectively inhibits expression of six HCV genotypes*. *Virology*, 2006. **3**: p. 100.
213. Bradford, S. and J.A. Cowan, *Catalytic metallodrugs targeting HCV IRES RNA*. *Chem Commun (Camb)*, 2012. **48**(25): p. 3118-20.
214. Kolykhalov, A.A., S.M. Feinstone, and C.M. Rice, *Identification of a highly conserved sequence element at the 3' terminus of hepatitis C virus genome RNA*. *J Virol*, 1996. **70**(6): p. 3363-71.
215. You, S. and C.M. Rice, *3' RNA elements in hepatitis C virus replication: kissing partners and long poly(U)*. *J Virol*, 2008. **82**(1): p. 184-95.
216. Shetty, S., et al., *Hepatitis C virus genomic RNA dimerization is mediated via a kissing complex intermediate*. *RNA*, 2010. **16**(5): p. 913-25.
217. Sharma, K.K., et al., *Analysis of the RNA chaperoning activity of the hepatitis C virus core protein on the conserved 3'X region of the viral genome*. *Nucleic Acids Res*, 2012. **40**(6): p. 2540-53.
218. Toroney, R., et al., *Regulation of PKR by HCV IRES RNA: importance of domain II and NS5A*. *J Mol Biol*, 2010. **400**(3): p. 393-412.
219. Kulinski, T., et al., *The apical loop of the HIV-1 TAR RNA hairpin is stabilized by a cross-loop base pair*. *J Biol Chem*, 2003. **278**(40): p. 38892-901.

220. Beigelman, L., et al., *Chemical modification of hammerhead ribozymes. Catalytic activity and nuclease resistance.* J Biol Chem, 1995. **270**(43): p. 25702-8.
221. Beigelman, L., et al., *Synthesis of 2'-modified nucleotides and their incorporation into hammerhead ribozymes.* Nucleic Acids Res, 1995. **23**(21): p. 4434-42.
222. Nielsen, P.E., et al., *Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide.* Science, 1991. **254**(5037): p. 1497-500.
223. Demidov, V.V., et al., *Stability of peptide nucleic acids in human serum and cellular extracts.* Biochem Pharmacol, 1994. **48**(6): p. 1310-3.
224. Knudsen, H. and P.E. Nielsen, *Antisense properties of duplex- and triplex-forming PNAs.* Nucleic Acids Res, 1996. **24**(3): p. 494-500.
225. Lebars, I., et al., *LNA derivatives of a kissing aptamer targeted to the trans-activating responsive RNA element of HIV-1.* Blood Cells Mol Dis, 2007. **38**(3): p. 204-9.
226. Darfeuille, F., et al., *LNA/DNA chimeric oligomers mimic RNA aptamers targeted to the TAR RNA element of HIV-1.* Nucleic Acids Res, 2004. **32**(10): p. 3101-7.
227. Elmen, J., et al., *Locked nucleic acid (LNA) mediated improvements in siRNA stability and functionality.* Nucleic Acids Res, 2005. **33**(1): p. 439-47.
228. Vester, B., et al., *Improved RNA cleavage by LNAzyme derivatives of DNazymes.* Biochem Soc Trans, 2004. **32**(Pt 1): p. 37-40.
229. Wyszko, E., et al., *Spiegelzymes(R) mirror-image hammerhead ribozymes and mirror-image DNazymes, an alternative to siRNAs and microRNAs to cleave mRNAs in vivo?* PLoS One, 2014. **9**(1): p. e86673.
230. Wyszko, E., et al., *Spiegelzymes: sequence specific hydrolysis of L-RNA with mirror image hammerhead ribozymes and DNazymes.* PLoS One, 2013. **8**(1): p. e54741.
231. Heidenreich, O., et al., *High activity and stability of hammerhead ribozymes containing 2'-modified pyrimidine nucleosides and phosphorothioates.* J Biol Chem, 1994. **269**(3): p. 2131-8.
232. Ferre-D'Amare, A.R. and J.A. Doudna, *Use of cis- and trans-ribozymes to remove 5' and 3' heterogeneities from milligrams of in vitro transcribed RNA.* Nucleic Acids Res, 1996. **24**(5): p. 977-8.
233. Delfosse, V., et al., *Riboswitch structure: an internal residue mimicking the purine ligand.* Nucleic Acids Res, 2010. **38**(6): p. 2057-68.
234. Salvail-Lacoste, A., et al., *Affinity purification of T7 RNA transcripts with homogeneous ends using ARiBo and CRISPR tags.* RNA, 2013. **19**(7): p. 1003-14.
235. Di Tomasso, G., et al., *The ARiBo tag: a reliable tool for affinity purification of RNAs under native conditions.* Nucleic Acids Res, 2011. **39**(3): p. e18.
236. Balke, D., et al., *Generation and selection of ribozyme variants with potential application in protein engineering and synthetic biology.* Appl Microbiol Biotechnol, 2014. **98**(8): p. 3389-99.
237. Gredell, J.A., C.S. Frei, and P.C. Cirino, *Protein and RNA engineering to customize microbial molecular reporting.* Biotechnol J, 2012. **7**(4): p. 477-99.
238. Michener, J.K. and C.D. Smolke, *High-throughput enzyme evolution in Saccharomyces cerevisiae using a synthetic RNA switch.* Metab Eng, 2012. **14**(4): p. 306-16.

**Annexe 1.0 Étude du rôle des nucléotides de la boucle V
dans la reconnaissance du substrat par le ribozyme VS**

Mise en contexte

Pour cet article, je suis co-premier auteur avec Patricia Bouchard. Les études RMN des mutants de la boucle V ont été effectuées par Patricia Bouchard et le modèle de l'interaction entre SLI et SLV a été réalisé par Geneviève Desjardins et Véronique Lisi. Dans la partie expérimentale, j'ai effectué la transcription *in vitro* des différents ribozymes mutants à l'aide de plasmides et du substrat à l'aide de gabarits d'ADN. J'ai ensuite mis au point les protocoles pour les cinétiques et analysé les résultats des expériences de cinétiques. J'ai participé à l'écriture des sections du manuscrit pertinentes à ma participation scientifique et Patricia a fait de même. Pascale Legault a conçu le projet et a supervisé l'ensemble de l'étude, incluant la collecte des données et l'analyse des résultats. Elle a relu et corrigé ce manuscrit.

Role of SLV in SLI Substrate Recognition by the *Neurospora* VS Ribozyme

Patricia Bouchard,^{1,4} Julie Lacroix-Labonté,^{1,4} Geneviève Desjardins,¹ Philippe Lampron,¹
Véronique Lisi,³ Sébastien Lemieux,^{2,3} François Major,^{2,3} and Pascale Legault¹

¹Département de Biochimie, Université de Montréal, C.P. 6128, Succursale Centre-Ville,
Montréal, QC, Canada, H3C 3J7

²Département d'Informatique et de Recherche Opérationnelle, Université de Montréal, C.P.
6128, Succursale Centre-Ville, Montréal, QC, Canada, H3C 3J7

³Institut de Recherche en Immunologie et en Cancérologie

⁴These authors contributed equally to this work.

This research was originally published in RNA journal:

RNA. 2008 Apr;14(4):736-48. doi: 10.1261/rna.824308. Epub 2008 Feb 26

ABSTRACT

Substrate recognition by the VS ribozyme involves a magnesium-dependent loop / loop interaction between the SLI substrate and the SLV hairpin from the catalytic domain. Recent NMR studies of SLV demonstrated that magnesium ions stabilize a U-turn loop structure and trigger a conformational change for the extruded loop residue U700, suggesting a role for U700 in SLI recognition. Here, we kinetically characterized VS ribozyme mutants to evaluate the contribution of U700 and other SLV loop residues to SLI recognition. To help interpret the kinetic data, we structurally characterized the SLV mutants by NMR spectroscopy and generated a three-dimensional model of the SLI / SLV complex by homology modeling with MC-Sym. We demonstrated that the mutation of U700 by A, C, or G does not significantly affect ribozyme activity, whereas deletion of U700 dramatically impairs this activity. The U700 backbone is likely important for SLI recognition, but does not appear to be required for either the structural integrity of the SLV loop or for direct interactions with SLI. Thus, deletion of U700 may affect other aspects of SLI recognition, such as magnesium ion binding and SLV loop dynamics. As part of our NMR studies, we developed a convenient assay based on detection of unusual ^{31}P and ^{15}N N7 chemical shifts to probe the formation of U-turn structures in RNAs. Our model of the SLI / SLV complex, which is compatible with biochemical data, leads us to propose novel interactions at the loop I / loop V interface.

Keywords : *Neurospora* VS ribozyme; substrate recognition; loop/loop interaction; NMR structural probing; kinetic studies; RNA homology modeling

INTRODUCTION

RNA hairpins play many essential roles as basic elements of RNA structure. They guide the folding of RNA, help maintain its structural integrity, and serve as recognition motifs in RNA-RNA and RNA-protein interactions. RNA hairpins have recently gained increasing interest given the central role played by microRNAs and various other non-coding RNAs in the regulation of gene expression [1, 2]. Therefore, it is important to understand how these small RNA building blocks assemble to form complex RNA-based architectures and support important biological functions.

Small ribozymes, such as the hammerhead, hairpin, HDV, and VS ribozymes, represent simple RNA architectures capable of enzymatic activity, and have been widely used as model systems to understand structure-function relationships in RNA. The VS ribozyme is one of these few naturally occurring ribozymes, which possesses a unique tertiary structure to perform its functions and has the distinct ability to recognize a hairpin substrate through formation of a loop / loop tertiary interaction [for recent reviews, see [3] and [4]]. Its catalytic activities are metal-dependent phosphodiester bond cleavage and ligation reactions, which both involve 5'-OH and 2',3'-cyclic phosphate termini [5, 6]. Although the natural VS ribozyme isolated from the mitochondria of certain strains of *Neurospora* performs self-cleavage activity [5], VS ribozyme fragments of ~100-160 nucleotides can perform cleavage *in trans* when incubated with small hairpin substrates *in vitro* [7].

There is currently no high-resolution structure of the VS ribozyme, but structural information is available from the proposed secondary structure (Fig. 1A; [8], tertiary structure models [9, 10], and NMR structures of individual stem-loops [11-16]. The catalytic domain of the VS ribozyme consists of five stem-loop subdomains (SLII-SLVI; Fig. 1A), and stem-loop I (SLI; Fig. 1B) defines the substrate domain [8]. The cleavage site is located within the internal loop of stem-loop I, between nucleotides G620 and A621 [5]. Our present understanding is that in order for cleavage to occur, the cleavage site internal loop must dock in a cleft formed by SLII and SLVI to allow its interaction with the active site formed by the A730 loop of SLVI [3, 4, 17, 18]. The proposed catalytic mechanism for site-specific cleavage

relies on the direct participation of G638 from the internal loop of SLI and A756 from the A730 loop of SLVI in general acid-base catalysis [19].

Substrate recognition by the VS ribozyme involves a loop / loop interaction between SLI and SLV, which is stabilized by magnesium ions (Mg^{2+}). Mutational studies have provided evidence for three Watson-Crick base pairs involving nucleotides 630-632 of loop I and nucleotides 697-699 of loop V [20]. Formation of this tertiary interaction is accompanied by a rearrangement of SLI from an unshifted to a shifted conformation (Fig. 1B), which affects the structure of the cleavage site internal loop [21, 22]. SLI mutants that cannot adopt the shifted conformation are not cleaved by the VS ribozyme, whereas those that can adopt the shifted conformation are active in the cleavage reaction (Figs. 1B-C; [21]. Three-dimensional structures of small hairpins containing the unshifted (inactive) and shifted (active) internal loop conformations have been determined by NMR spectroscopy [11-13]. It appears from these structures that the active conformation contains a unique Mg^{2+} -binding site and a unique tertiary interaction motif, both of which may be important for catalysis [13].

To better understand the role of the SLV receptor in SLI recognition, we have determined two NMR structures for an SLV fragment (Fig. 2), one in the absence (SLV^{free} ; [15] and one in the presence of magnesium ions (SLV^{Mg} ; [16]. The loop of SLV^{free} forms a loose noncanonical U-turn motif, whereas that of SLV^{Mg} forms a compact canonical U-turn motif [15, 16]. The U-turn of SLV^{free} was termed noncanonical because it lacks the stacking interaction between the U696 base and the A698 5'-phosphate and the hydrogen bond between the U696 H3 and the A698 3'-phosphate (Fig. 2) that are characteristic of canonical U-turn structures and found in SLV^{Mg} [15, 16]. Although Mg^{2+} ions affect the loop conformation, they do not significantly change the conformation of the three SLV bases (G697, A698, C699) that are proposed to base pair with SLI [20]. In both structures, these three bases of SLV are exposed and well positioned for binding to SLI [15, 16]. The most significant change induced by Mg^{2+} ions occurs at the extruded loop residue U700, which comes closer to the three interacting bases (G697, A698, and C699) of SLV (Fig. 2; [16]. Given this Mg^{2+} -dependent conformational change of U700 and the fact that Mg^{2+} ions stabilize the loop I / loop V interaction, we hypothesized that U700 could play an important role in SLI recognition [16].

In the present study, we explored the role of U700 and other SLV features (loop-closing base pair, nucleotide insertion) in SLI recognition. We first performed kinetic experiments with the VS ribozyme containing mutations in loop V to better understand the SLV loop requirements for catalysis. To analyze the effects of these mutations on the structural integrity of the U-turn structure in the SLV loop, we developed an NMR-based assay to allow quick structural mapping of small SLV fragments. Given that the structure of SLV^{Mg} is available and that structural characteristics of the SLI loop have been previously derived from biochemical data, we used homology modeling to obtain a structural model of the SLI / SLV interaction. This model helps explain our kinetic results and allows us to propose novel interactions at the loop I / loop V interface.

RESULTS

Using $k_{\text{cat}}/K_{\text{M}}$ values to investigate SLI recognition

To better understand the role of U700 and other SLV loop features in substrate recognition, we initiated kinetic studies of *trans* cleavage with the VS ribozyme [7]. We used the previously-characterized Aval VS ribozyme (Rz; [7], either the wild-type RNA (Fig. 1A) or RNAs containing a variety of mutations in the SLV loop (Fig. 3). Binding of this ribozyme to the wild-type SLI substrate involves the loop I / loop V interaction and a conformational change from the unshifted state to the shifted state, which activates the substrate for catalysis (Fig. 1B). The SLI(T) substrate (S) was selected for kinetic studies, because it can only adopt a preshifted active state (Fig. 1C; [21, 23]).

The second-order rate constant ($k_{\text{cat}}/K_{\text{M}}$) for the reaction of Rz with S was determined in single-turnover experiments [24], as illustrated for the U700G mutant Rz in Fig. 4. Cleavage experiments were performed with ^{32}P -labeled S and excess Rz, and the conversion of substrate into product was monitored by gel electrophoresis (Fig. 4A). For these experiments, the natural log of the fraction of remaining substrate was plotted against the time of the reaction, and the value of the first-order rate constant, k_{obs} , was derived from the slope of the graph, as shown in Fig. 4B. For the wild-type and mutant Rz, a linear relationship was observed when k_{obs} was plotted as a function of Rz concentration, and the values of catalytic efficiency ($k_{\text{cat}}/K_{\text{M}}$) were obtained from the slope of that graph (Fig. 4C). Table 1 provides a summary of the $k_{\text{cat}}/K_{\text{M}}$ values and of the relative catalytic efficiencies of the wild-type (WT) and mutant (MUT) ribozymes [$(k_{\text{cat}}/K_{\text{M}})_{\text{WT}}/(k_{\text{cat}}/K_{\text{M}})_{\text{MUT}}$].

The parameter $k_{\text{cat}}/K_{\text{M}}$ is an important kinetic constant that combines both the rate and the binding terms, and values of $k_{\text{cat}}/K_{\text{M}}$ of wild-type and mutant enzymes can be compared to address the role of specific residues in catalysis. For enzymes mutated at residues involved only in substrate binding and not in the reaction chemistry, $k_{\text{cat}}/K_{\text{M}}$ values of wild-type and mutant enzymes have been used to analyze the contribution of the mutated residues to the stabilization of the enzyme-substrate complex [25, 26].

Biochemical studies have implicated residues from the SLV loop in SLI substrate recognition [20], and there is no evidence that SLV residues participate in the reaction chemistry [19]. Therefore, $k_{\text{cat}}/K_{\text{M}}$ values measured for VS ribozymes carrying mutations in the SLV loop can be used to analyze the contribution of SLV loop residues to SLI recognition. SLI recognition is defined here as the binding of the SLI substrate to the *trans*-cleaving VS ribozyme and includes all ribozyme-substrate interactions.

Substrate recognition of the preshifted SLI(T) substrate used for our studies should involve formation of the loop I / loop V interaction [20] and docking of SLI at the active site [10, 27]. Current data are consistent with the role of the SLV loop being limited to the loop I / loop V interaction [20], although it is possible that mutations of the SLV loop also affect the proper docking of SLI at the active site, either because this docking depends on SLV interaction(s) with the rest of the ribozyme or as a consequence of a perturbed loop I / loop V interface. In the context where an SLV mutation would disrupt proper SLI docking at the active site, both SLI binding and the rate of chemistry could be affected, and a quantitative analysis of the effect of the mutation on SLI binding would not be possible [28]. In addition, mutations in SLV could also affect the folding of the ribozyme and/or change the energy landscape of the reaction. Thus, the $k_{\text{cat}}/K_{\text{M}}$ values obtained here can only be interpreted in a qualitative manner; they provide information on the loop I / loop V interaction, but could also reflect other aspects of the ribozyme cleavage reaction.

Kinetic properties of VS ribozymes carrying mutations in the SLV loop

The values of $k_{\text{cat}}/K_{\text{M}}$ indicate that mutation of U700 by A, C, or G does not significantly disrupt ribozyme activity, since the ratios of $(k_{\text{cat}}/K_{\text{M}})_{\text{WT}}/(k_{\text{cat}}/K_{\text{M}})_{\text{MUT}}$ are less than 4 (Table 1). The U700G mutation is the least disruptive, followed by U700A, and U700C. Deletion of U700, however, severely affects ribozyme activity, since the $k_{\text{cat}}/K_{\text{M}}$ of the Δ U700 mutant is reduced 140-fold compare to the wild-type value. The $k_{\text{cat}}/K_{\text{M}}$ value of Δ U700 is similar to that of the positive control C699G [$(k_{\text{cat}}/K_{\text{M}})_{\text{WT}}/(k_{\text{cat}}/K_{\text{M}})_{\text{MUT}} = 270$], a mutant which is expected to disrupt the formation of the loop I / loop V interaction [20].

Mutation of the loop-closing U695-A701 base pair by a C-G or a G-A base pair also does not significantly affect ribozyme activity, since $(k_{\text{cat}}/K_M)_{\text{WT}}/(k_{\text{cat}}/K_M)_{\text{MUT}}$ values of 4.1 and 2.7 are observed for the U695C/A701G and U695G mutants, respectively. However, deletion of U700 in the context of a loop-closing G-A base pair severely reduces ribozyme activity since a $(k_{\text{cat}}/K_M)_{\text{WT}}/(k_{\text{cat}}/K_M)_{\text{MUT}}$ value of 74 was observed for the U695G/ Δ U700 mutant, a value similar to that of the Δ U700 mutant (Table 1), which contains the wild-type loop-closing U-A base pair.

Insertion of a single U residue between C699 and U700 does not significantly affect ribozyme activity, since a $(k_{\text{cat}}/K_M)_{\text{WT}}/(k_{\text{cat}}/K_M)_{\text{MUT}}$ value of 3.9 was observed for the +U[^]U700 mutant. However, the insertion of the UC dinucleotide between C699 and U700 severely reduces ribozyme activity, as judged from the $(k_{\text{cat}}/K_M)_{\text{WT}}/(k_{\text{cat}}/K_M)_{\text{MUT}}$ value of 130 observed for the +UC[^]U700 mutant.

NMR studies of mutant SLV fragments

To help interpret the kinetic data of the mutant ribozymes, we characterized SLV fragments containing the equivalent mutations (Fig. 3) by NMR spectroscopy. In order to ascertain that SLV fragments formed a hairpin as in the folded ribozyme and not a duplex conformation (Fig. 5A), they were analyzed by 1D imino ¹H NMR spectroscopy (not shown) and native gel electrophoresis (Fig. 5B). The NMR data indicated that, except for U700A, wild-type SLV and all mutants adopted a single conformation in the absence of Mg²⁺. In the presence of Mg²⁺, the NMR data indicated the presence of a second conformation (>10%) for select mutants. As illustrated in Fig. 5B, the native gel data could be used to confirm that the hairpin conformation is favored for the majority of SLV mutants, but that a few mutants formed a mixed population of hairpin and duplex conformations. The presence of a significant population of duplex prevented further structural characterization of these mutants (U700A, U700G, and +UC[^]U700) by NMR. For the U700A and the U700G mutants, there is a clear possibility for base pairing involving residue 700 (A or G) and U696 (Fig. 5) that stabilizes the duplex conformation of the SLV fragment. It is important to note that the k_{cat}/K_M values of the ribozyme mutants U700A and U700G are not significantly lower than that of the wild-type

ribozyme (Table 1). Therefore the observation that certain mutations produce duplexes with SLV fragments does not suggest that these same mutations produce misfolded and functionally impaired ribozymes. On the contrary, since all mutations were localized to the loop of SLV and hydroxyl radicals footprinting data indicate that SLV does not interact with other parts of the ribozyme in the absence of SLI [27], it is likely that structural changes caused by our mutations would only affect the loop of SLV and not the general folding of the ribozyme. To obtain local structural information on the mutants, we compared their NMR data with that of the wild-type SLV.

Previous NMR studies of an SLV RNA fragment indicated that the SLV loop adopts a loose noncanonical U-turn fold in the absence of Mg^{2+} (SLV^{free}), which becomes a compact canonical U-turn fold in the presence of Mg^{2+} (SLV^{Mg} ; [15, 16]. In agreement with previous NMR studies of U-turn structures [29-31], unusually large chemical shift changes were associated with the formation of a canonical U-turn structure in SLV^{Mg} for two specific NMR signals, the ^{31}P resonance of U696 3'-phosphate (3'-P) and the ^{15}N resonance of A698 N7 [15, 16]. The detection of unusual ^{31}P and ^{15}N chemical shifts in SLV RNA fragments was therefore used to identify the formation of a canonical U-turn fold for the mutants being functionally characterized in this study (Fig. 3).

NMR structural probing of the wild-type and mutant SLV RNAs involved the collection and analysis of 1D 1H -decoupled ^{31}P spectra and 2D 1H - ^{15}N long-range HMQC spectra [32]. In the 1D ^{31}P spectra of the wild-type SLV RNA, distinct downfield-shifted signals were observed in the presence but not the absence of Mg^{2+} (Fig. 6A). Based on previous assignment, the ^{31}P signal of U696 3'-P is the most downfield-shifted signal, ~ 1.5 ppm downfield from the main cluster of ^{31}P signals [16]. The unusual chemical shift of U696 3'-P was previously associated with formation of a canonical U-turn structure for SLV^{Mg} . In this structure, the 3'-P of U696 acts as the turning phosphate of the U-turn fold and is part of a negatively-charged pocket that defines a Mg^{2+} -binding site (Fig. 2; [15, 16]. The ^{31}P NMR spectra of the C699G and $\Delta U700$ mutants provide examples of data where a Mg^{2+} -dependent ^{31}P downfield-shifted signal is either not observed (C699G) or less shifted from the main cluster of ^{31}P signals ($\Delta U700$) when

compared to the U696 3'-P resonance of the wild-type RNA (Fig. 6A). A summary of the ^{31}P chemical shift data for the mutant RNAs is provided in Table 2. For all mutant RNAs, no significantly downfield-shifted signal was observed in the absence of Mg^{2+} . In the presence of Mg^{2+} , the extent of the ^{31}P downfield shift varied depending on the mutant RNA, and results were separated in three classes based on their similarity with the U696 3'-P chemical shift of the wild-type SLV (Table 2); they were typical of (+; U695G and U695G/ Δ U700), not as significantly shifted as (+/-; U700C, Δ U700, and U695C/A701G), or very different from (-; +U[^]U700, C699G) values observed for the canonical U-turn structure of SLV^{Mg}.

In the 2D ^1H - ^{15}N long-range HMQC spectra of the wild-type SLV RNA, an upfield-shifted ^{15}N resonance, assigned to the A698 N7-H8 crosspeak based on previous NMR studies [16], was observed in the presence but not the absence of Mg^{2+} (Fig. 6B). This unusual N7 chemical shift, ~6 ppm upfield from the main cluster of N7 signals, was previously attributed to the stabilization of the hydrogen bond between A698 N7 and U696 2'-OH within the canonical U-turn structure for SLV^{Mg} [15, 16]. The 2D ^1H - ^{15}N long-range HMQC spectra of the Δ U700 mutant provide an example of data where a ^{15}N upfield-shifted H8-N7 crosspeak is observed in Mg^{2+} only, but where the N7 chemical shift is not as shifted from the main cluster of N7 signals as the A698 N7 from the wild-type SLV (Fig. 6B). A summary of the 2D ^1H - ^{15}N long-range HMQC data for the mutant RNAs is provided in Table 2. For all mutants being studied, no significantly upfield-shifted N7 signal was observed in the absence of Mg^{2+} . In the presence of Mg^{2+} , upfield-shifted N7 resonances were observed for some mutant RNAs, and results were separated in three classes based on their similarity with the chemical shift value of A698 N7 (Table 2); they were typical of (+; U695G and U695G/ Δ U700), not as significantly shifted as (+/-; U700C, Δ U700, and U695C/A701G), or very different from (-; +U[^]U700, C699G) values observed for the canonical U-turn structure of SLV^{Mg}.

It is interesting to note that in the presence of Mg^{2+} , the extent of the N7 upfield shift (typical of [+], not as significantly shifted as [+/-], or very different from [-]) always matched the extent of the ^{31}P downfield shift for any given mutant (Table 2). The fact that these independent observations correlate with each other supports the use of these ^{31}P and ^{15}N

chemical shifts to examine the U-turn structure of SLV mutants. For the wild-type, U695G, and U695G/ Δ U700 RNAs, their unusual ^{31}P and ^{15}N N7 chemical shifts indicate the formation of a canonical U-turn structure similar to that of SLV^{Mg} (Fig. 2). For the +U $^{\wedge}$ U700 and C699G mutants, the absence of unusual ^{31}P and ^{15}N N7 chemical shifts indicates that their loop does not form a canonical U-turn. In addition, the formation of a loose noncanonical U-turn fold in the +U $^{\wedge}$ U700 and C699G mutant RNAs cannot be evaluated, because the loose noncanonical U-turn fold of SLV^{free} did not yield unusual ^{31}P and ^{15}N N7 chemical shifts [15]. For the U700C, Δ U700, and U695C/A701G mutants, the extents of ^{31}P and ^{15}N N7 shifts in the presence of Mg^{2+} were not as significant as for the wild-type RNA, indicating that their U-turn structure is more compact than for SLV^{free} , but not as compact as for SLV^{Mg} (Fig. 2).

Kinetic and NMR studies of mutants containing a G-A loop-closing base pair

Hairpin loops containing U-turn structures are generally not closed by Watson-Crick base pairs as in SLV, but by noncanonical base pairs, such as the sheared G-A base pair. With the exception of SLV, the few U-turn structures with Watson-Crick loop-closing base pairs that have been structurally characterized form loose noncanonical U-turns [15]. In Watson-Crick base pairs, the C1'-C1' distance is larger than that of many noncanonical base pairs, and we previously suggested that this constrains the formation of a canonical U-turn fold [15]. We hypothesized that replacing the Watson-Crick U-A base pair by a G-A base pair in SLV would remove this constraint and allow for formation of a compact canonical U-turn fold in the context of the Δ U700 mutation. Consequently, if a canonical U-turn fold were sufficient for SLI recognition, this G-A mutant would allow SLI recognition even following deletion of U700.

In agreement with our prediction, the NMR data clearly indicate that the U695G and U695G/ Δ U700 mutants both adopt canonical U-turn structures in the presence of Mg^{2+} ions. In addition, mutation of the loop-closing U695-A701 base pair by a G-A [$(k_{\text{cat}}/K_{\text{M}})_{\text{WT}}/(k_{\text{cat}}/K_{\text{M}})_{\text{MUT}} = 2.7$] base pair has only a small effect on ribozyme activity. However, deletion of U700 in the context of a loop-closing G-A base pair has a damaging effect on ribozyme activity [$(k_{\text{cat}}/K_{\text{M}})_{\text{WT}}/(k_{\text{cat}}/K_{\text{M}})_{\text{MUT}} = 74$], similar to its effect in the wild-

type context. It therefore appears that formation of a canonical structure in this $\Delta 700$ mutant is not sufficient to allow a favorable interaction with SLI.

Homology modeling of the SLI / SLV interaction with MC-Sym

A model of the SLI / SLV interaction was generated using the constraint-satisfaction program MC-Sym [33, 34]. For these modeling studies, the previously determined NMR structure of SLV^{Mg} was used (PDB entry 1YN1) [16] as is, assuming that it would not be significantly changed by SLI binding. Although the structure of SLI with the wild-type terminal loop sequence has been previously determined by NMR [11], these studies were performed in the absence of Mg^{2+} and/or SLV and, under those conditions, the SLI loop structure was not well defined by the NMR data [11]. Previous biochemical studies indicated that SLI adopts a U-turn structure important for its recognition by SLV and for efficient ribozyme activity [20, 35, 36]. It is therefore likely that the loop of SLI adopts a U-turn structure when bound to SLV, and this SLI U-turn structure was constrained in our model of the SLI / SLV interaction. Previous mutational studies also indicated that SLI nucleotides G630, U631, and C632 form Watson-Crick base pairs with SLV nucleotides C699, A698, and G697, respectively [20]. It is therefore likely that these three Watson-Crick base pairs stack on each other to adopt a short right-handed helix, and these three Watson-Crick base pairs between SLI and SLV were also constrained in our model.

The model shown in Fig. 7 is a good representative of the seven models generated with MC-Sym and energy minimized with X-PLOR-NIH (heavy atom RMSD between 2.4-3.1 Å). This model indicates that the U-turn conformations of the SLI and SLV loops are compatible with formation of the three-predicted Watson-Crick base pair at the loop / loop junction, at least from a geometrical standpoint. In all models obtained, the point of closest approach between the phosphodiester backbones of the two loops is at the turning phosphate of their respective U-turn structure (3'-P of U628 and U696). The U700 nucleotide is the SLV loop residue located the farthest from SLI; it is generally not positioned to form a stable interaction with SLI, although a few models suggest the possibility for a hydrogen bonding interaction between the U700 O4 and the U631 2'-OH (Fig. 7). Interestingly, the models display interactions not

previously identified: the stacking of C629 on G630 and a hydrogen bond between C629 NH₂ and A701 2'-OH. Previous biochemical data support the involvement of C629 and A701 in catalysis [20, 35-37]. Mutation of C629 by A, G, or U leads to more than a 10-fold loss in the activity of a self-cleaving ribozyme [20] and modification interference data support the importance of the C629 base [37]. The importance of A701 2'-OH was also revealed from studies that analyzed the contribution of 2'-OH to the cleavage and ligation activities [35, 36]. Overall, our current model of the SLI / SLV interaction is consistent with existing biochemical data and allows us to propose new interactions at the loop I / loop V interface.

To more thoroughly address the role of U700, we removed residue U700 from our models and subjected them to energy minimization. We obtained structural models of the SLI / Δ U700SLV complex (not shown) that are fully compatible with all the restraints used for minimization of the SLI / SLV complex, including all NMR restraints previously used for structure calculation of SLV^{Mg} (except those involving U700; [16]). In the SLI / Δ U700SLV model, the three loop / loop Watson-Crick base pairs were maintained, as well as the position of C629 with respect to SLV, indicating that deletion of U700 is compatible with our three-dimensional model of the SLI / SLV interaction. This result is not surprising given that both the base and the ribose of U700 are completely extruded from the U-turn loop structure of SLV^{Mg} (Fig. 2).

DISCUSSION

Substrate recognition by the VS ribozyme involves an important and well-characterized Mg^{2+} -dependent loop / loop interaction between the SLI substrate and SLV from the catalytic domain. In the five-nucleotide loop of SLV, the sequence requirement and the role(s) of the first four nucleotides have been clearly defined; U696, G697, and A698 form a UNR sequence necessary for U-turn formation and G697, A698, and C699 form Watson-Crick base pairs with SLI residues C632, U631, and G630, respectively [20]. The roles of the fifth nucleotide, U700, and of the loop-closing base pair have not been clearly established, and the possibility for nucleotide insertion in the SLV loop has not been previously examined. Given the facts that U700 undergoes the largest conformational change between the NMR structures of SLV^{free} and SLV^{Mg} , and that Mg^{2+} ions stabilize the loop I / loop V interaction, we were particularly intrigued by the role of U700 in this interaction.

Formation of a canonical U-turn structure in SLV is neither necessary nor sufficient for SLI recognition

From previous NMR studies, we observed that the loop of SLV undergoes a conformational change from a loose noncanonical U-turn to a compact canonical U-turn structure in the presence of magnesium ions. Here, we developed an NMR-based assay specifically designed to structurally characterize U-turn formation in SLV mutants. From our study, we could not establish a correlation between the formation of a canonical U-turn structure and the k_{cat}/K_M values of the mutants. We observed that mutants with an activity similar to wild-type [$(k_{cat}/K_M)_{WT}/(k_{cat}/K_M)_{MUT} \leq 4$], U700C and +U^ΔU700 for example, did not adopt a compact canonical U-turn structure like the wild-type RNA. These mutants likely adopt a loose noncanonical U-turn structure that allows the correct positioning of the G697, A698, and C699 bases for pairing with loop I, but the details of these SLV loop structures are not available. Conversely, the U695G/ Δ U700 mutant adopts a canonical U-turn structure, but its k_{cat}/K_M is significantly reduced compared to the wild-type ribozyme value. Thus, formation of a canonical U-turn structure is neither necessary nor sufficient to allow a favorable interaction with SLI.

Many mutations in SLV are compatible with SLI recognition

Most mutations in our study yield k_{cat}/K_M values that are not significantly reduced when compared to the wild-type ribozyme [$(k_{\text{cat}}/K_M)_{\text{WT}}/(k_{\text{cat}}/K_M)_{\text{MUT}} \leq 4$], indicating that these loop mutations are compatible with the proper folding and catalytic activity of the VS ribozyme. We concluded that these mutations did not significantly affect SLI substrate recognition.

Mutation of the loop-closing U695-A701 base pair by a C-G base pair had only a small effect on k_{cat}/K_M . These results are in agreement with previous self-cleavage studies in which the U695G/A701C mutant was found to be slightly more active than the wild-type ribozyme [8]. Different types of Watson-Crick base pairs are therefore compatible with SLI recognition. In addition, mutation of the loop-closing U695-A701 base pair by a G-A base pair had only a small effect on ribozyme activity, which indicates that noncanonical base pair(s) can also be compatible with SLI recognition.

Given the importance of U700 for SLI recognition, we were interested to see if insertions in loop V were compatible with catalysis. It appears that addition of an extra U residue (+U^{U700} mutant) caused a minor decrease in k_{cat}/K_M , however, addition of a dinucleotide (+UC^{U700}) had a more dramatic effect. The NMR data for both of these mutants did not provide any evidence of U-turn structure formation. Although a single loop insertion in SLV supports SLI binding, larger loop insertions, like that of +UC^{U700}, could favor loop conformations that are not compatible with SLI recognition.

Substitutions of U700 by A, C, and G were performed to investigate the role of the base at position 700. The kinetic data indicate that these SLV mutants bind SLI similarly to the wild-type SLV, and, therefore, that the U700 base plays a negligible role in SLI recognition. These results are in agreement with modification interference experiments in which the U700 base was removed without significantly affecting the self-cleavage activity of the VS ribozyme [37]. In summary, we found that SLI recognition is compatible with many different loop-closing base pairs in SLV, a single nucleotide insertion in the SLV loop, and any of the four common bases at U700.

Deletion of U700 is likely detrimental to SLI recognition

Although base mutations of U700 have negligible effects on ribozyme activity, deletion of U700 substantially reduces k_{cat}/K_M . The reduction of k_{cat}/K_M for the Δ U700 mutant is similar to that of the C699G mutant, which was expected to abrogate the loop I / loop V interaction [20]. The NMR chemical shift changes of Δ U700 are similar to that of U700C, indicating the presence of a noncanonical U-turn fold for Δ U700 and suggesting that the reduced activity of Δ U700 is not due to a completely misfolded U-turn structure. Although we cannot eliminate the possibility that deletion of U700 affects other aspects of catalysis (e.g. the folding of the ribozyme and the rate of cleavage), it appears that deletion of U700 is detrimental to SLI recognition.

It is peculiar that U700 can be mutated by any other base, but cannot be deleted. This suggests a role for this nucleotide that does not involve interactions to the base. However, the backbone of U700 likely plays a role in substrate recognition. Our structural modeling of the SLI / SLV interaction suggests that the ribose and 5'-phosphate of U700 are too remote to allow for a direct interaction with SLI. Our modeling studies of a SLI / Δ U700SLV complex also indicate that *in silico* deletion of U700 does not significantly affect the structure of the complex. The remaining roles for the U700 backbone include providing direct interactions with a magnesium ion and/or other parts of the ribozyme, acting as a spacer, and/or providing a dynamic element that imparts an entropic advantage to substrate binding.

The backbone of U700 could participate in direct interactions with magnesium ions or other parts of the ribozymes. Previous studies have indicated that a 2'-deoxy at U700 and phosphorothioate substitutions at U700 (for both the 5'-phosphate and the 3'-phosphate) do not significantly affect cleavage and ligation activities [17, 35, 36]. In the phosphorothioate substitution experiments, the VS ribozyme was transcribed by T7 RNA polymerase, which produces only the Rp phosphorothioate isomer, but not the Sp isomer [17, 36]. It is therefore possible that the pro-Sp oxygens of the 5'- and 3'-phosphates of U700 are involved in binding a metal, and this is consistent with the large conformational change in the backbone of U700 upon addition of magnesium ions (see Fig.2; [16]). Four Mg^{2+} -binding sites were recently

identified in the SLV loop structure (Fig. 2) from paramagnetic NMR studies; one of them involves the pro-Sp oxygen of the 5'-phosphate of U700 [16]. Interaction of a magnesium ion with the backbone of U700 may therefore be important for stabilizing the loop I / loop V interaction and/or mediating interactions of the SLI / SLV junction with other parts of the ribozyme [27].

As a spacer, the backbone of U700 could provide sufficient flexibility in the SLV loop to allow an optimal loop I / loop V interaction and proper docking of SLI at the active site. By increasing the SLV loop dynamics, the backbone of U700 could also provide an entropic advantage for SLI recognition. Studies of protein / ligand interactions indicate that biomolecular associations often involve dynamic residues that can provide favorable entropic contribution to binding [38]. Further studies are required to determine if a similar entropic contribution is important for SLI recognition by the VS ribozyme and for RNA / RNA interactions in general.

CONCLUSION

Interpretation of structure-function studies of macromolecules is often complicated by the fact that it is difficult to obtain structural data on the mutants being functionally characterized. In this study, kinetic characterization of loop V mutants of the VS ribozyme was complemented by a convenient NMR-based assay specifically designed to structurally characterize U-turn formation in these SLV mutants. In addition, homology modeling of the SLI / SLV interaction with MC-Sym generated a three-dimensional model that is in agreement with previous biochemical data. The MC-Sym model was useful for the interpretation of our kinetic data and allowed us to propose novel interactions at the loop I / loop V interface that are supported from previous biochemical studies.

From a comparative analysis of the kinetic and NMR data, we conclude that formation of a canonical U-turn structure in SLV is neither necessary nor sufficient for SLI recognition. Our kinetic studies reveal that many SLV mutations are compatible with the VS ribozyme activity, including substitutions of the loop-closing base pair, a single nucleotide insertion, and U700

base substitutions. However, deletion of the extruded loop residue U700 significantly impairs the ribozyme activity. The reduced activity of the U700 deletion mutants is neither due to a completely misfolded U-turn structure in SLV nor to interactions involving the U700 base. Rather, the backbone of U700 likely plays a role in SLI substrate recognition. The backbone of U700 may provide a direct interaction with a magnesium ion and/or other parts of the ribozyme that contribute to SLI recognition. Alternatively, the backbone of U700 could act as a spacer that allows loop V to optimally recognize the SLI substrate and/or provide a dynamic element that imparts an entropic advantage to substrate binding. NMR structure determination and relaxation studies of an SLI / SLV complex are now in progress in our laboratory to gain a better structural understanding of the loop I / loop V interaction and provide insight into the role that entropy plays in this interaction.

MATERIALS AND METHODS

Plasmids

The plasmid A-3, which allows for production of the *Ava*I ribozyme (Rz), has been described (Guo et al. 1995). The stem-loop V mutants (Fig. 3) of Rz were prepared by using the QuickChangeII site-directed mutagenesis method (Stratagene, CA). All plasmids created for this study were validated by DNA sequencing.

Preparation of RNA for kinetic studies

For kinetic studies of *trans* cleavage, the *Ava*I ribozyme (Rz) and the SLI(T) substrate (S) were used. Rz contains a nine-nucleotide leader (5'-gggaaagcu) followed by the VS nucleotides 640-783 (Fig. 1). The wild-type and mutant sequences of Rz (Fig. 3) were prepared by *in vitro* T7 RNA polymerase transcription from plasmid A-3 (Guo et al. 1995) linearized at the *Ssp*I site. All RNAs were purified by electrophoresis in 5% (w/v) polyacrylamide gels containing 7 M urea. They were visualized by UV shadowing, eluted from the selected gel slices by crush and soak at 42 °C, ethanol precipitated and resuspended in water. The RNAs were applied to a semi-preparative DNAPac PA100 column (Dionex) heated at 65 °C [39] and equilibrated with buffer A (25 mM Tris pH 7.6 and 2 mM EDTA). They were eluted from the HPLC column using a gradient of buffer B (25 mM Tris pH 7.6, 2

mM EDTA and 2 M NaCl). The eluted RNAs were collected on ice, pooled, concentrated with a Centricon-3 ultrafiltration device (Millipore, MA) and exchanged into storage buffer (10 mM Tris pH 7.5 and 1 mM EDTA).

The SLI(T) RNA is a preshifted substrate with the sequence: 5'-GAGCGAAGGGCAUCGUCGGCCCGAGCUC-3'. It was synthesized in vitro using T7 RNA polymerase and a single-stranded synthetic oligonucleotide template (Integrated DNA Technologies, IA). SLI(T) was purified to single-nucleotide resolution by 20% denaturing gel electrophoresis, dephosphorylated at its 5'-end with calf alkaline phosphatase (Roche Molecular Biochemicals, NJ), and further purified by DEAE-Sephacel chromatography (Amersham Biosciences, NJ). The eluted RNA was concentrated with a Centricon-3 ultrafiltration device (Millipore, MA) and exchanged into 10 mM d₁₁-Tris pH 7.0, 50 mM NaCl and 0.05 mM NaN₃. The SLI(T) RNA was then 5'-end-labeled with γ -(³²P) ATP (PerkinElmer LAS Canada Inc., Ont. Canada) using T4 polynucleotide kinase (New England Biolabs, MA) according to the manufacturer's instructions and further purified by 20% denaturing gel-electrophoresis. After gel extraction, the ³²P-labeled SLI(T) RNA was resuspended in TE (10 mM Tris pH 7.6 and 1 mM EDTA) and desalted over a Micro Bio-Spin 6 chromatography column (Bio-Rad, CA).

Kinetic studies

All reactions were carried out at 30 °C in reaction buffer (50 mM Tris pH 7.5, 25 mM KCl, 2 mM spermidine and 25 mM MgCl₂). Prior to the cleavage reaction, the ³²P-labeled SLI(T) substrate was heated to 95 °C for 1 min in storage buffer and snap-cooled in iced water to promote hairpin formation. Cleavage reactions were initiated by addition of equal volume (20 μ L) of the ³²P-labeled SLI(T) substrate (S) and the *Ava*I ribozyme (Rz), either the wild-type (WT) or SLV loop mutants (Fig. 3), after a 5-min preincubation of S and Rz at 30 °C in reaction buffer. Typically, about 10 aliquots of 3 μ L were removed from the 40 μ L-reaction mixture at specific times, and the reaction was stopped by addition of 12 μ L of stop mix (50 mM EDTA pH 8.0, 0.04% xylene cyanol, 0.04% bromophenol blue and \geq 85% formamide) and transfer to 4°C. Reaction products were separated by electrophoresis on 20%

polyacrylamide / 7 M urea gels. For each time point, the amount of substrate and product were quantified using the Molecular Imager FX and the Quantity One 1D analysis software version 4.6.2 from Bio-Rad Laboratories (Ont., Canada).

Single-turnover kinetic experiments were typically performed with $[S] = 0.25$ nM and excess Rz. Reactions were first order for ~ 3 half-lives and end points were 90-99%. The extent of the reaction was calculated as $(S_t^*)/(S_{t=0}^*)$, where $S_t^* = S_t - S_{t \rightarrow \infty}$; $S_{t=0}^* = S_{t=0} - S_{t \rightarrow \infty}$; S_t is the amount of S at time t ; $S_{t \rightarrow \infty}$ is the amount of S at the end point; and $S_{t=0}$ is the initial amount of S. The natural log of $(S_t^*)/(S_{t=0}^*)$ was plotted against the time of the reaction and the value of k_{obs} was obtained from the slope by linear regression. The linear dependence of k_{obs} on $[Rz]$ was also analyzed by linear regression to derive the second order rate constant of the reaction, k_{cat}/K_M . The linear dependence of k_{obs} on $[Rz]$ was observed from 5 nM to 500 nM for the wild-type and U700G Rz, from 50 nM to 3 μM for U700A, U700C, U695C/A701G, U695G, and +U^{U700} Rz, and from 50 nM to 5 μM for all other mutant Rz. Kinetic experiments were all performed at least three times, and reported k_{cat}/K_M and their errors are average values and standard deviations, respectively, from these multiple measurements.

Preparation of ^{15}N -labeled SLV RNA fragments for native gel and NMR studies

A total of 10 SLV RNA fragments were chosen for NMR analysis, one containing the wild-type loop sequence (Fig. 4A) and nine others derived from this RNA, but containing selected loop mutations (Fig. 3). These SLV RNA fragments were synthesized in vitro using T7 RNA polymerase, a single-stranded DNA template, and ^{15}N -labeled NTPs, and they were purified as previously described [15]. The final concentration of the ^{15}N -labeled SLV RNA samples for NMR studies ranged from 0.8 to 1.2 mM. The sample buffer was exchanged by use of Centricon-3 ultrafiltration devices with either NMR buffer A (10 mM d_{11} -Tris pH 7.0, 50 mM NaCl, and 0.05 mM NaN_3), NMR buffer B (10 mM d_{11} -Tris pH 7.0 and 0.05 mM NaN_3), or NMR buffer C (10 mM d_{11} -Tris pH 7.0, 20 mM MgCl_2 , and 0.05 mM NaN_3), which contained 90% H_2O and 10% D_2O . For studies in D_2O , the RNA samples were transferred to 99.996% D_2O with multiple cycles of lyophilisation and resuspension in D_2O . Before exchanging in

NMR buffer C, the RNA samples were exchanged in NMR buffer A or NMR buffer B and heated to 95 °C for two minutes and then immediately cooled in iced water.

Native gel studies of SLV fragments

Non-denaturing polyacrylamide gel electrophoresis was carried out on 7.5% polyacrylamide gels, acrylamide : bis (37.5 : 1) in TB buffer (50 mM Tris borate pH 8.0 and 20 mM magnesium acetate). The gels were pre-run for 30 minutes at 200 V, loaded with the SLV RNA fragments (2 µg RNA at a concentration of 20 µM) and then run for 2 hours at 250 V and 4 °C. The gels were stained with Stains All (Sigma-Aldrich, Ont., CA).

NMR studies of SLV fragments

All NMR experiments were conducted at 25 °C on a Varian ^{Unity}INOVA 500 MHz spectrometer equipped with a pulse-field gradient unit and an actively-shielded z gradient probe, either a ¹H{¹³C/¹⁵N} triple resonance probe or a ¹H{¹⁵N-³¹P} indirect detection probe. The following experiments were recorded: a 1D ¹⁵N-decoupled ¹H flip-back watergate spectrum (in H₂O); a 1D ¹H-decoupled ³¹P spectrum (in H₂O); and a long-range 2D ¹H-¹⁵N HMQC spectrum [32] optimized for detection of adenine N1 and N3 and purine N7 and N9 (in D₂O, *J* = 21 Hz). The 2D spectra were processed and analyzed with the NMRPipe/NMRDraw package [40]. ¹H and ¹⁵N chemical shifts were referenced to an external standard of 2,2-dimethyl-2-silapentane-5-sulfonic acid [41] at 0.00 ppm [42] and ³¹P chemical shifts were referenced to an external standard of 85% phosphoric acid at 0.00 ppm.

Homology modeling of the SLI / SLV interaction

Three-dimensional models of the SLI / SLV interaction were built using MC-Sym version 4.2.0 [33, 34]. The modeling strategy was to generate ensembles of structures for both SLI and SLV hairpins, and merge all combinations of the SLI and SLV structures via the formation of Watson-Crick base pairing interactions for G630-C699, U631-A698, and C632-G697. For the ensemble of SLV structures, we used the 11 structural coordinates of SLV^{Mg} from PDB entry 1YN1 [16]. For the ensemble of SLI structures (wild-type residues 623-637), we generated structures that satisfied a number of base relations and inter-atomic distance restraints in order

to obtain an A-form helix for the stem (Watson-Crick base pairing for 623-637, 624-636, 625-635, and 626-633) with an extra helical C634 residue and a U-turn structure in the loop with stacking of C629, G630, U631, and C632. A total of 7 models of the SLI / SLV complex were obtained that fit the input restraints and that were structurally distinct from each other (heavy atom RMSD between 2.4-3.1 Å). These models were energy minimized using X-PLOR-NIH version 2.0.6 [43]. For the energy minimization, the force field consisted of bond, angle, improper (stereochemistry), and repulsive van der Waals energy terms. In addition, distance and torsion angle pseudo-energy terms were used to maintain the structural integrity of the complex. All NMR-derived distance restraints used for determination of the SLV^{Mg} structure [16] were included along with restraints to enforce A-form geometry in the stem of SLI, a U-turn structure in the SLI loop and Watson-Crick base-pairing at the loop I / loop V junction.

ACKNOWLEDGMENTS

We thank Richard A. Collins for providing plasmid A-3 and Ryan Richter for computer support. We also thank Richard A. Collins and James G. Omichinski for critical reading of the manuscript. This work was supported by a grant from the Canadian Institutes for Health Research (CIHR #117713). J.L.-L. holds a Ph.D. scholarship from the Université de Montréal. P.B. holds Ph.D. scholarships from CIHR and the Université de Montréal. G.D. holds scholarships from the Fonds de la Recherche en Nature et Technologies (FQRNT) and from the Luigi-Liberatore Foundation. P.L. (the eighth author) holds a Canada Research Chair in Structural Biology of RNA.

REFERENCES

1. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans**. *Nature*, 1998. **391**(6669): p. 806-11.
2. Svoboda, P. and A. Di Cara, *Hairpin RNA: a secondary structure of primary importance*. *Cell. Mol. Life Sci.*, 2006. **63**(7-8): p. 901-8.
3. Collins, R.A., *The *Neurospora Varkud* satellite ribozyme*. *Bioch. Soc. Trans. Rev.*, 2002. **30**: p. 1122-1126.
4. Lilley, D.M., *The *Varkud* satellite ribozyme*. *RNA*, 2004. **10**: p. 151-158.
5. Saville, B.J. and R.A. Collins, *A site-specific self-cleavage reaction performed by a novel RNA in *Neurospora* mitochondria*. *Cell*, 1990. **61**: p. 685-696.
6. Saville, B.L. and R.A. Collins, *RNA-mediated ligation of self-cleavage products of a *Neurospora* mitochondrial plasmid transcript*. *Proc. Natl. Acad. Sci. USA*, 1991. **88**: p. 8826-8830.
7. Guo, H.C.T. and R.A. Collins, *Efficient trans-cleavage of a stem-loop RNA substrate by a ribozyme derived from *Neurospora* VS RNA*. *EMBO J.*, 1995. **14**(12): p. 368-376.
8. Beattie, T.L., J.E. Olive, and R.A. Collins, *A secondary-structure model for the self-cleaving region of *Neurospora* VS RNA*. *Proc. Natl. Acad. Sci. U.S.A.*, 1995. **92**: p. 4686-4690.
9. Hiley, S.L. and R.A. Collins, *Rapid formation of a solvent-inaccessible core in the *Neurospora Varkud* satellite ribozyme*. *EMBO J.*, 2001. **20**(19): p. 5461-5469.
10. Lafontaine, D.A., D.G. Norman, and D.M. Lilley, *The global structure of the VS ribozyme*. *EMBO J.*, 2002. **21**(10): p. 2461-2471.
11. Flinders, J. and T. Dieckmann, *A pH controlled conformational switch in the cleavage site of the VS ribozyme substrate RNA*. *J. Mol. Biol.*, 2001. **308**: p. 665-679.
12. Michiels, P.J.A., et al., *Structure of the ribozyme substrate hairpin of *Neurospora* VS RNA: A close look at the cleavage site*. *RNA*, 2000. **6**: p. 1821-1832.
13. Hoffmann, B., et al., *NMR structure of the active conformation of the *Varkud* satellite ribozyme cleavage site*. *Proc. Natl. Acad. Sci. USA*, 2003. **100**(12): p. 7003-7008.
14. Flinders, J. and T. Dieckmann, *The solution structure of the VS ribozyme active site loop reveals a dynamic "hot-spot"*. *J. Mol. Biol.*, 2004. **341**: p. 935-949.
15. Campbell, D.O. and P. Legault, *NMR structure of the *Varkud* satellite ribozyme stem-loop V RNA and magnesium-ion binding from chemical-shift mapping*. *Biochemistry*, 2005. **44**: p. 4157-4170.
16. Campbell, D.O., et al., *NMR structure of varkud satellite ribozyme stem-loop V in the presence of magnesium ions and localization of metal-binding sites*. *Biochemistry*, 2006. **45**(35): p. 10591-605.
17. Sood, V.D., T.L. Beattie, and R.A. Collins, *Identification of phosphate groups involved in metal binding and tertiary interactions in the core of the *Neurospora* VS ribozyme*. *J. Mol. Biol.*, 1998. **282**: p. 741-750.
18. Lafontaine, D.A., et al., *The A730 loop is an important component of the active site of the VS ribozyme*. *J. Mol. Biol.*, 2001. **312**: p. 663-674.
19. Wilson, T.J., A.C. McLeod, and D.M. Lilley, *A guanine nucleobase important for catalysis by the VS ribozyme*. *EMBO J*, 2007. **26**(10): p. 2489-500.

20. Rastogi, T., et al., *A long-range pseudoknot is required for activity of the Neurospora VS ribozyme*. EMBO J., 1996. **15**(11): p. 2820-2825.
21. Andersen, A. and R.A. Collins, *Rearrangement of a stable RNA secondary structure during VS ribozyme catalysis*. Mol. Cell, 2000. **5**: p. 469-478.
22. Andersen, A.A. and R.A. Collins, *Intramolecular secondary structure rearrangement by the kissing interaction of the Neurospora VS ribozyme*. Proc. Natl. Acad. Sci. USA, 2001. **98**(14): p. 7730-7735.
23. Zamel, R. and R.A. Collins, *Rearrangement of substrate secondary structure facilitates binding to the Neurospora VS ribozyme*. J. Mol. Biol., 2002. **324**: p. 903-915.
24. Herschlag, D. and T.R. Cech, *Catalysis of RNA cleavage by the Tetrahymena thermophila ribozyme. I. Kinetic description of the reaction of an RNA substrate complementary to the active site*. Biochemistry, 1990. **29**(44): p. 10159-71.
25. Wilkinson, A.J., et al., *Site-directed mutagenesis as a probe of enzyme structure and catalysis: tyrosyl-tRNA synthetase cysteine-35 to glycine-35 mutation*. Biochemistry, 1983. **22**(15): p. 3581-6.
26. Fersht, A.R., et al., *Hydrogen bonding and biological specificity analysed by protein engineering*. Nature, 1985. **314**(6008): p. 235-8.
27. Hiley, S.L., et al., *4-thio-U cross-linking identifies the active site of the VS ribozyme*. EMBO J., 2002. **21**(17): p. 4691-4698.
28. Kraut, D.A., K.S. Carroll, and D. Herschlag, *Challenges in enzyme mechanism and energetics*. Annu Rev Biochem, 2003. **72**: p. 517-71.
29. Schweisguth, D.C. and P.B. Moore, *On the conformation of the anticodon loops of initiator and elongator methionine tRNAs*. J. Mol. Biol., 1997. **267**: p. 505-519.
30. Sundaram, M., P.C. Durant, and D.R. Davis, *Hypermethylated nucleosides in the anticodon of tRNA^{Lys} stabilize a canonical U-turn structure*. Biochemistry, 2000. **39**: p. 12575-12584.
31. Cabello-Villegas, J., I. Tworowska, and E.P. Nikonowicz, *Metal ion stabilization of the U-turn of the A₃₇ N⁶-dimethylallyl-modified anticodon stem-loop of Escherichia coli tRNA^{Phe}*. Biochemistry, 2004. **43**: p. 55-66.
32. Kay, L.E. and A. Bax, *New methods for the measurement of NH---CαH coupling constants in ¹⁵N-labeled proteins*. J. Magn. Reson., 1990. **86**(1): p. 110-126.
33. Major, F., et al., *The combination of symbolic and numerical computation for three-dimensional modeling of RNA*. Science, 1991. **253**(5025): p. 1255-60.
34. Major, F., *Building three-dimensional ribonucleic acid structures*. Comput. Sci. Eng., 2003. **5**: p. 44-53.
35. Sood, V.D., S. Yekta, and R.A. Collins, *The contribution of 2'-hydroxyls to the cleavage activity of the Neurospora VS ribozyme*. Nucleic Acids Res., 2002. **30**(5): p. 1132-1138.
36. Jones, F.D. and S.A. Strobel, *Ionization of a critical adenosine residue in the Neurospora Varkud satellite ribozyme active site*. Biochemistry, 2003. **42**: p. 4265-4276.
37. Beattie, T.L. and R.A. Collins, *Identification of functional domains in the self-cleaving Neurospora VS ribozyme using damage selection*. J. Mol. Biol., 1997. **267**: p. 830-840.
38. Homans, S.W., *Probing the binding entropy of ligand-protein interactions by NMR*. Chembiochem, 2005. **6**(9): p. 1585-91.

39. Shields, T.P., et al., *High-performance liquid chromatography purification of homogenous-length RNA produced by trans cleavage with a hammerhead ribozyme*. RNA, 1999. **5**(9): p. 1259-67.
40. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes*. J. Biomol. NMR, 1995. **6**: p. 277-293.
41. Rupert, P.B., et al., *Transition state stabilization by a catalytic RNA*. Science, 2002. **298**(5597): p. 1421-4.
42. Wishart, D.S., et al., *^1H , ^{13}C , ^{15}N chemical shift referencing in biomolecular NMR*. J. Biomol. NMR, 1995. **6**: p. 135-140.
43. Schwieters, C.D., et al., *The Xplor-NIH NMR molecular structure determination package*. J. Magn. Reson., 2003. **160**: p. 66-74.
44. Umlauf, F., et al., *Hepatitis C virus detection by single-round PCR specific for the terminal 3' noncoding region*. J Clin Microbiol, 1996. **34**(10): p. 2552-8.

TABLES

Table I. Summary of kinetic data for the VS ribozyme mutants.

Ribozyme	k_{cat}/K_M ($\text{min}^{-1}\mu\text{M}^{-1}$)	$(k_{\text{cat}}/K_M)_{\text{WT}}$
		$(k_{\text{cat}}/K_M)_{\text{MUT}}$
wild-type	3.8 ± 0.7	1.0
U700C	1.0 ± 0.2	3.7
U700A	2.1 ± 0.1	1.8
U700G	3.6 ± 0.4	1.0
ΔU700	0.028 ± 0.004	1.4×10^2
U695C/A701G	0.9 ± 0.1	4.1
U695G	1.4 ± 0.3	2.7
U695G/ ΔU700	0.051 ± 0.007	74
+U [^] U700	0.98 ± 0.09	3.9
+UC [^] U700	0.029 ± 0.006	1.3×10^2
C699G	0.014 ± 0.004	2.7×10^2

Table II. Summary of NMR data for the SLV fragments.

SLV fragment	Mg ²⁺ ^a	Hairpin ^b	Values of selected chemical shifts (ppm)			
			³¹ P downfield shift ^c	¹⁵ N N7 upfield shift ^c		
Wild type	-	Yes	-	-0.02 to -0.86	-	231.4
	+	Yes	+	1.52	+	224.0
U700C	-	Yes	-	0.09 to -0.83	-	> 229.3
	+	Yes	+/-	0.76	+/-	227.4
U700A	-	No	n.d.	n.d.	n.d.	n.d.
	+	No	n.d.	n.d.	n.d.	n.d.
U700G	-	Yes	-	0.38 to -0.87	-	> 229.5
	+	No	n.d.	n.d.	n.d.	n.d.
Δ U700	-	Yes	-	0.17 to -0.83	-	> 229.7
	+	Yes	+/-	0.61	+/-	227.7
U695C/A701G	-	Yes	-	-0.01 to -0.85	-	> 230.7
	+	Yes	+/-	1.23	+/-	225.8
U695G	-	Yes	-	-0.03 to -0.83	-	>230.4
	+	Yes	+	1.51	+	225.1
U695G/ Δ U700	-	Yes	-	0.32 to -0.85	-	> 228.2
	+	Yes	+	1.49	+	224.1
+U [^] U700	-	Yes	-	-0.03 to -0.86	-	> 229.8
	+	Yes	-	-0.42 to -1.21	-	> 228.9
+UC [^] U700	-	Yes	-	0.00 to -0.85	-	> 229.7
	+	No	n.d.	n.d.	n.d.	n.d.
C699G	-	Yes	-	0.28 to -0.85	-	> 229.2
	+	Yes	-	0.04 to -1.19	-	> 229.1

^a The buffer used was either NMR buffer A which does not contain MgCl₂ (-) or NMR buffer C which contains 20 mM MgCl₂ (+).

^b The conformation of the RNA was verified by native gels and NMR experiments.

^c ³¹P downfield shifts and ¹⁵N N7 upfield shifts were typical of (+), not as significantly shifted as (+/-), or very different from (-) values obtained for the canonical U-turn structure of SLV^{Mg}. Measured values are indicated. For ³¹P data, the range of observed values is indicated when no distinct downfield signal was observed.

FIGURES

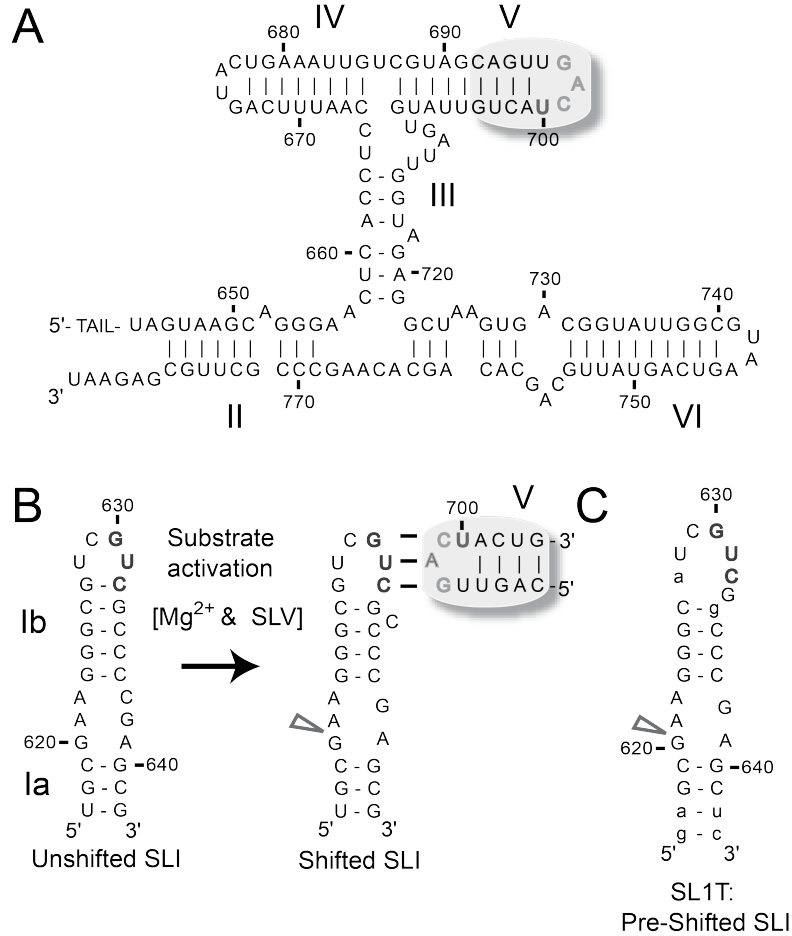


Figure 1. Substrate recognition by the *Neurospora* VS ribozyme. **A)** Primary and secondary structures of the catalytic domain of the VS ribozyme. For this study, the *AvaI* ribozyme was used; it includes a 5'-TAIL with the sequence: 5'-gggaaagcuGCGG-3' [7]. **B)** The catalytic domain recognizes the SLI substrate (subdivided in Ia and Ib) via a loop / loop interaction between residues 630-632 of SLI and residues 697-699 of SLV. Upon interaction with SLV, the SLI RNA with the wild-type sequence undergoes a structural change from an unshifted to a shifted conformation; only the later allows cleavage of the SLI internal loop at the site indicated by an arrowhead. **C)** Preshifted mutant substrates, such as SLI(T), are efficiently cleaved by the VS catalytic domain [21, 23]. Wild-type and mutant nucleotides are represented by upper and lower cases, respectively. Specific SLI and SLV nucleotides are color coded for easy identification in subsequent figures.

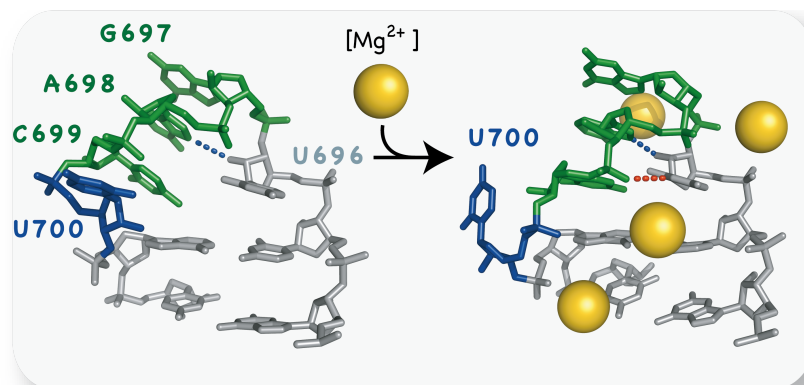


Figure 2. NMR structures of the SLV loop determined in the absence (left) and presence [44] of magnesium ions [15, 16], showing the large conformational change of U700 (blue). The dotted lines represent two hydrogen bonds that are characteristic of canonical U-turn structures, the A698 N7 to U696 2'OH (blue) and the U696 H3 to A698 3'-phosphate interactions (red). For simplicity, only heavy atoms are shown.

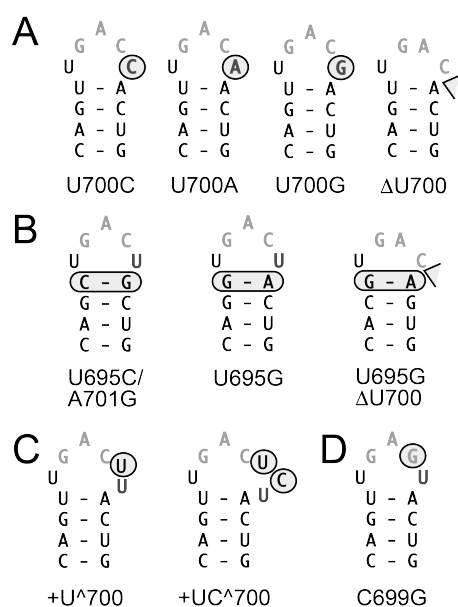


Figure 3. SLV loop mutants characterized kinetically and structurally in this study. Sequence and secondary structure of **A)** SLV with U700 mutations and deletion, **B)** SLV with mutations of the loop-closing base pair, **C)** SLV with insertions of nucleotides between residues C699 and U700, and **D)** SLV with a loop mutation that is expected to disrupt the loop I / loop V interaction [20] and also serves as an experimental control.

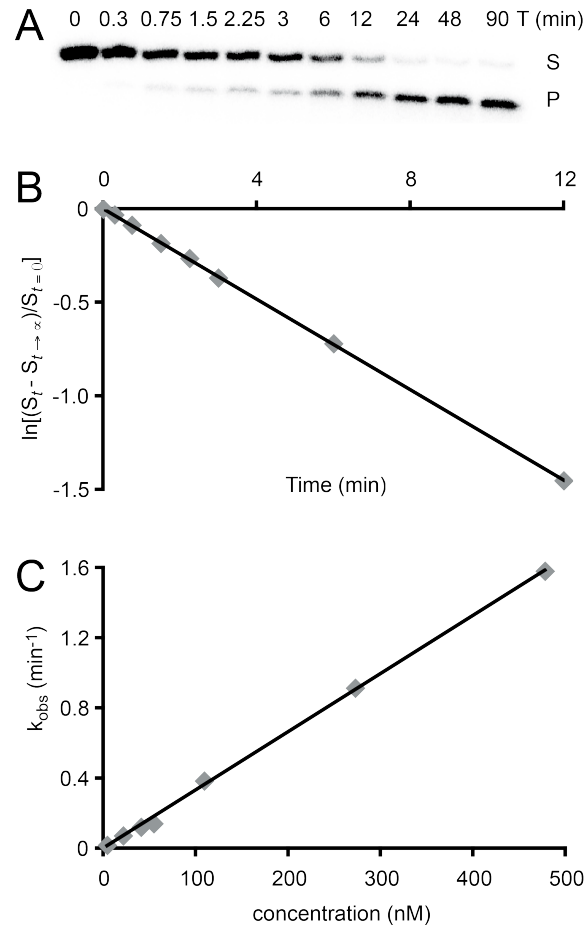


Figure 4. Kinetic analysis of the U700G Rz mutant.

A) Cleavage of ^{32}P -labeled SLI(T) by the U700G Rz is performed at 30 °C in the presence of 25 mM MgCl_2 under single turnover conditions [excess enzyme (37.5 nM) over substrate (0.25 nM)]. Denaturing polyacrylamide gel electrophoresis shows the amount of substrate (S) and product (P) at various time (T) points of the reaction. **B)** The natural log of the fraction of remaining substrate is plotted against time (see Materials and Methods). The data are fit to a single exponential and the first order rate constant is extracted by linear regression (k_{obs} = absolute value of the slope = 0.12 min^{-1} and $R^2 = 0.9999$) **C)** Cleavage reactions with the U700G Rz are performed at various ribozyme concentrations and the value of k_{obs} is plotted against ribozyme concentration. At low ribozyme concentration, there is a linear relationship between k_{obs} and ribozyme concentration. The data are fit to a single exponential and the k_{cat}/K_M is extracted by linear regression (k_{cat}/K_M = value of the slope = $3.3 \text{ min}^{-1}\mu\text{M}^{-1}$ and $R^2 = 0.9988$).

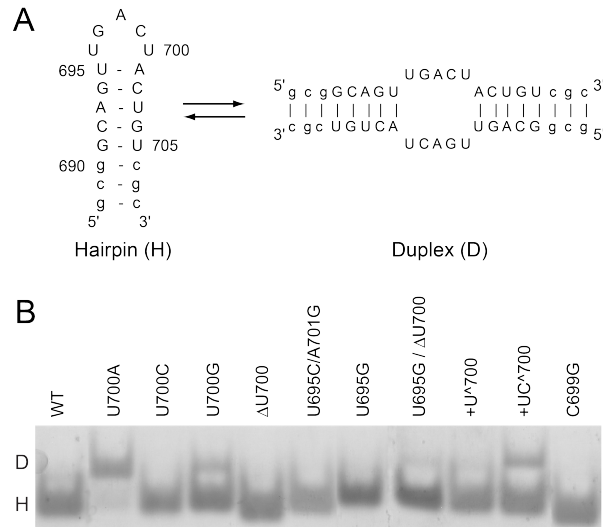


Figure 5. Conformation of SLV RNA fragments. **A)** Primary and secondary structures of the wild-type SLV RNA fragment used for NMR studies in its hairpin (H) and duplex (D) conformations. **B)** Native gel electrophoresis of the wild-type (WT) and mutant SLV RNAs. The RNAs were loaded at a concentration of 20 μ M in the presence of 20 mM $MgCl_2$.

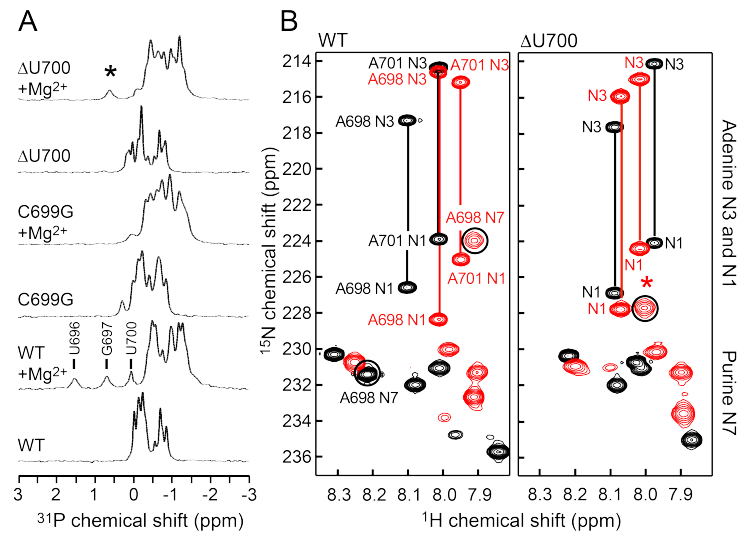


Figure 6. NMR studies of SLV loop mutants. **A)** 1D 1H -decoupled ^{31}P spectra of wild-type (WT), C699G, and $\Delta U700$ SLV RNA fragments in the absence and presence of 20 mM $MgCl_2$ ($+Mg^{2+}$). For the wild-type SLV in the presence of Mg^{2+} , the most downfield-shifted resonances correspond to the 3'-phosphate of the residues indicated [16]. **B)** Superpositions of

a selected region from 2D ^1H - ^{15}N long-range HMQC spectra recorded in the absence (black peaks) and presence (red peaks) of 20 mM MgCl_2 for the wild-type (WT; left panel) and ΔU700 RNAs (right panel). The adenine H2-N1 and H2-N3 correlations of a given adenine can be easily identified based on the common H2 frequency and are linked by a line in the spectra. For the wild-type SLV, the A698 N7-H8 crosspeaks are indicated. In **A**) and **B**) the asterisks point to a downfield-shifted ^{31}P or upfield-shifted N7 resonance of an SLV mutant that appears in the presence of Mg^{2+} .

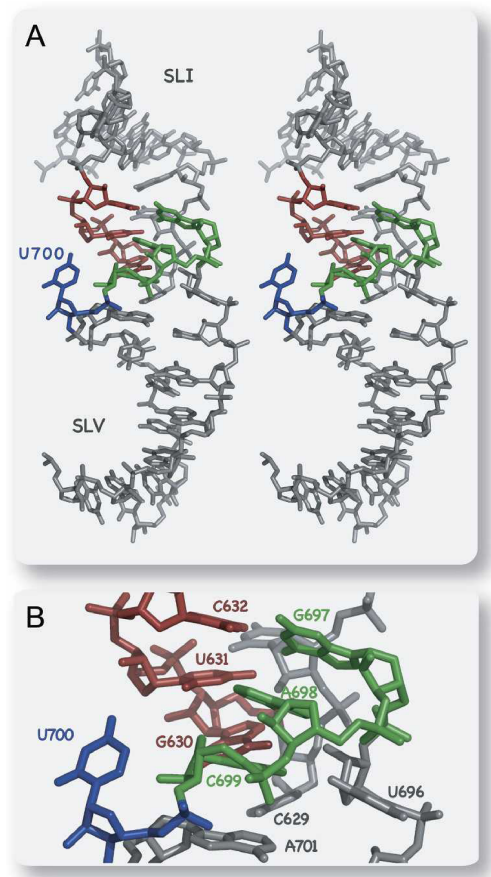


Figure 7. A) Stereo view of a structural model of the SLI / SLV complex. The selected model is a representative of the 7 structures obtained from homology modeling with MC-Sym. For simplicity, only heavy atoms are shown. **B)** Close up of the model shown in A.