

Université de Montréal

**Caractérisation structurale et thermodynamique de la
reconnaissance du substrat par le ribozyme VS de
*Neurospora***

par
Patricia Bouchard

Département de Biochimie et de 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

21 août 2014

© Patricia Bouchard, 2014

Université de Montréal
Faculté des études supérieures et postdoctorales

Cette thèse intitulée :

Caractérisation structurale et thermodynamique de la reconnaissance du
substrat par le ribozyme VS de *Neurospora*

Présentée par :
Patricia Bouchard

a été évaluée par un jury composé des personnes suivantes :

Normand Brisson, président-rapporteur
Pascale Legault, directeur de recherche
Serguei Chteinberg, membre du jury
Thorsten Dieckmann, examinateur externe
János G. Filep, représentant du doyen de la FES

Résumé

Les interactions ARN/ARN de type *kissing-loop* sont des éléments de structure tertiaire qui jouent souvent des rôles clés chez les ARN, tant au niveau fonctionnel que structural. En effet, ce type d'interaction est crucial pour plusieurs processus dépendant des ARN, notamment pour l'initiation de la traduction, la reconnaissance des ARN antisens et la dimérisation de génome rétroviral. Les interactions *kissing-loop* sont également importantes pour le repliement des ARN, puisqu'elles permettent d'établir des contacts à longue distance entre différents ARN ou encore entre les domaines éloignés d'un même ARN. Ce type d'interaction stabilise aussi les structures complexes des ARN fonctionnels tels que les ARNt, les riborégulateurs et les ribozymes.

Comme d'autres ARN fonctionnels, le ribozyme VS de *Neurospora* contient une interaction *kissing-loop* importante. Celle-ci est impliquée dans la reconnaissance du substrat et se forme entre la tige-boucle I (*stem-loop I*, *SLI*) du substrat et la tige-boucle V (*stem-loop V*, *SLV*) du domaine catalytique. Des études biochimiques ont démontré que l'interaction *kissing-loop* I/V, dépendante du magnésium, implique trois paires de bases Watson-Crick (W-C). De plus, cette interaction est associée à un réarrangement de la structure du substrat, le faisant passer d'une conformation inactive dite *unshifted* à une conformation active dite *shifted*. Les travaux présentés dans cette thèse consistent en une caractérisation structurale et thermodynamique de l'interaction *kissing-loop* I/V du ribozyme VS, laquelle est formée de fragments d'ARN représentant les tige-boucles I et V dérivées du ribozyme VS (*SLI* et *SLV*). Cette caractérisation a été réalisée principalement par spectroscopie de résonance magnétique nucléaire (RMN) et par titrage calorimétrique isotherme (*isothermal titration calorimetry*, *ITC*) en utilisant différents complexes *SLI/SLV* dans lesquels l'ARN *SLV* est commun à tous les complexes, alors que différentes variations de l'ARN *SLI* ont été utilisées, soit en conformation *shiftable* ou *preshifted*. Les données d'*ITC* ont permis de démontrer qu'en présence d'une concentration saturante de magnésium, l'affinité d'un substrat *SLI preshifted* pour *SLV* est extrêmement élevée, rendant cette interaction plus stable que ce qui est prédit pour un duplexe d'ARN équivalent. De plus, l'étude effectuée par *ITC* montre que des ARN *SLI preshifted* présentent une meilleure affinité pour *SLV* que des

ARN *SLI shiftable*, ce qui a permis de calculer le coût énergétique associé au réarrangement de structure du substrat. En plus de confirmer la formation des trois paires de bases W-C prédites à la jonction I/V, les études de RMN ont permis d'obtenir une preuve structurale directe du réarrangement structural des substrats *SLI shiftable* en présence de magnésium et de l'ARN *SLV*. La structure RMN d'un complexe *SLI/SLV* de grande affinité démontre que les boucles terminales de *SLI* et *SLV* forment chacune un motif *U-turn*, ce qui facilite l'appariement W-C intermoléculaire. Plusieurs autres interactions ont été définies à l'interface I/V, notamment des triplets de bases, ainsi que des empilements de bases. Ces interactions contribuent d'ailleurs à la création d'une structure présentant un empilement continu, c'est-à-dire qui se propage du centre de l'interaction jusqu'aux bouts des tiges de *SLI* et *SLV*. Ces études de RMN permettent donc de mieux comprendre la stabilité exceptionnelle de l'interaction *kissing-loop* I/V au niveau structural et mènent à l'élaboration d'un modèle cinétique de l'activation du substrat par le ribozyme VS. En considérant l'ensemble des données d'*ITC* et de RMN, l'étonnante stabilité de l'interaction I/V s'explique probablement par une combinaison de facteurs, dont les motifs *U-turn*, la présence d'un nucléotide exclu de la boucle de *SLV* (U700), la liaison de cations magnésium et l'empilement de bases continu à la jonction I/V.

Mots-clés : Ribozyme VS de *Neurospora*, motif *U-turn*, interaction *kissing-loop*, titrage calorimétrique isotherme (*ITC*), spectroscopie de résonance magnétique nucléaire (RMN), reconnaissance du substrat

Abstract

Kissing loops are tertiary structure elements that often play key roles in functional RNAs. Their formation is central to many RNA-mediated processes, such as translation initiation, antisense recognition and retroviral dimerization. Kissing loops are also involved in RNA folding as they form long-range interactions between different RNAs or remote domains within the same RNA and stabilize the complex architecture of functional RNA, such as tRNA, riboswitch aptamers and ribozymes.

Like several other functional RNAs, the *Neurospora* VS ribozyme contains an important kissing-loop interaction. The substrate recognition by the VS ribozyme depends largely on the formation of a magnesium-dependent kissing-loop interaction between stem-loop V (SLV) of the catalytic domain and stem-loop I (SLI) that defines the substrate domain. It has been shown from biochemical studies that the I/V kissing-loop interaction involves three Watson-Crick base pairs and is associated with a structural rearrangement of the SLI substrate from an unshifted and inactive to a shifted and active conformation. Here, we present a thermodynamic and structural characterization of the VS ribozyme I/V kissing-loop interaction using isolated stem-loop fragments (SLI and SLV). Both isothermal titration calorimetry (ITC) and nuclear magnetic resonance (NMR) spectroscopy studies were conducted with several SLI/SLV complexes using a common SLV, but either shiftable or preshifted SLI variants. From the ITC studies, we show that, under saturating amount of magnesium ions, the affinity of the preshifted SLI variants for SLV is remarkably high, the interaction being more stable than predicted for a comparable duplex. In addition, these ITC studies demonstrate that preshifted SLI variants have higher affinity for SLV than shiftable SLI variants, and these results allow us to evaluate the energetic cost of the conformational shift in SLI. From the NMR studies, we confirm formation of three Watson-Crick base pairs at the kissing-loop junction and provide direct evidence on the structural rearrangement of shiftable SLI variants in the presence of magnesium and SLV. The NMR structure of a high-affinity SLI/SLV complex demonstrates that both the SLI and SLV loops adopt U-turn structures, which facilitate intermolecular Watson-Crick base pairing. Several other interactions at the I/V interface, including base triples and base

stacking help create a continuously stacked structure. These NMR studies provide a structural basis for the high stability of the kissing-loop interaction and lead us to propose a kinetic model for substrate activation by the VS ribozyme. Taken together, our ITC and NMR data suggest that the remarkable stability of the I/V interaction is likely provided by a combination of several elements, especially the presence of the U-turn motif, the presence of an extruded nucleotide in SLV (U700), the binding of magnesium ions and the extensive base stacking interactions at the junction.

Keywords : *Neurospora* VS ribozyme, U-turn motif, kissing-loop interaction, isothermal titration calorimetry (ITC), nuclear magnetic resonance (NMR) spectroscopy, substrate recognition.

Table des matières

Résumé	v
Abstract.....	vii
Table des matières	ix
Liste des tableaux	xiii
Liste des figures.....	xv
Liste des abréviations	xix
Remerciements	xxii
1 Introduction.....	1
1.1 ARN non codants et ribozymes.....	2
1.1.1 Importance des ARN et versatilité des ARN non codants.....	2
1.1.2 Les ribozymes.....	5
1.1.3 Intérêt biochimique et médical des ARNnc.....	9
1.2 Le ribozyme VS de <i>Neurospora</i>	12
1.2.1 Structures secondaire et tertiaire.....	15
1.2.1.1 L'interaction <i>kissing-loop</i> I/V.....	16
1.2.1.1.1 Reconnaissance du substrat par l'interaction <i>kissing-loop</i> I/V	16
1.2.1.1.2 Activation du substrat par l'interaction <i>kissing-loop</i> I/V.....	17
1.2.1.1.3 Structures des domaines I et V seuls	21
1.2.1.2 Le domaine catalytique.....	27
1.2.2 Le site actif et mécanisme catalytique.....	29
1.2.2.1 Mécanisme catalytique	30
1.2.2.2 Éléments de structure de la boucle A730.....	34
1.2.2.3 Similarités mécanistiques entre les ribozymes VS et <i>hairpin</i>	36
1.2.3 Repliement du ribozyme VS.....	37
1.2.3.1 Rôle des cations divalents.....	38
1.2.3.2 Modèles tridimensionnels à basse résolution du ribozyme VS	43
1.3 Éléments de structure et repliement de l'ARN.....	46
1.3.1 Motif <i>U-turn</i>	47
1.3.2 Interaction <i>kissing-loop</i>	49

1.4 Étude de la structure de l'ARN par spectroscopie RMN.....	52
1.4.1 Spectroscopie RMN pour l'étude des macromolécules biologiques.....	53
1.4.2 Les difficultés et avantages de la RMN d'ARN.....	55
1.4.2.1 Les difficultés de la spectroscopie RMN	55
1.4.2.2 Les avantages de la spectroscopie RMN.....	60
1.4.3 Procédure pour la détermination de la structure d'un ARN.....	61
1.4.3.1 Préparation des échantillons.....	62
1.4.3.2 Utilisation des protons imino pour la détermination de structures secondaire et tertiaire	64
1.4.3.3 Stratégie d'attribution des signaux.....	65
1.4.3.4 Obtention des contraintes de RMN.....	70
1.4.3.5 Calcul de structure.....	71
1.5 Étude d'interactions ARN/ARN par titrage calorimétrique isotherme.....	72
1.6 Objectifs de la thèse.....	75
2 Étude du rôle des nucléotides de la boucle V dans la reconnaissance du substrat par le ribozyme VS	79
Mise en contexte.....	80
ABSTRACT	83
INTRODUCTION.....	84
RESULTS.....	87
DISCUSSION	95
CONCLUSION	99
MATERIALS AND METHODS	100
ACKNOWLEDGMENTS.....	104
REFERENCES.....	105
TABLES.....	109
FIGURES	112
3 Caractérisation de l'activation du substrat et de l'interaction <i>kissing-loop</i> I/V par spectroscopie RMN.....	119
Mise en contexte.....	120
ABSTRACT	123

INTRODUCTION	124
EXPERIMENTAL PROCEDURES.....	126
RESULTS	130
DISCUSSION.....	137
ASSOCIATED CONTENT.....	143
ACCESSION CODES.....	143
AUTHOR INFORMATION.....	143
ACKNOWLEDGMENT	144
ABBREVIATIONS	144
REFERENCES	144
TABLES	151
FIGURES.....	152
SUPPORTING INFORMATION.....	159
4 Caractérisation thermodynamique de l'interaction <i>kissing-loop</i> I/V et de l'activation du substrat.....	179
Mise en contexte	180
ABSTRACT.....	183
INTRODUCTION	184
RESULTS	186
DISCUSSION.....	194
MATERIALS AND METHODS.....	201
ACKNOWLEDGMENT	203
REFERENCES	203
TABLES	211
FIGURES.....	217
5 Discussion et perspectives	223
5.1 Construction d'un modèle RMN complet du ribozyme VS de <i>Neurospora</i>	227
5.2 Ingénierie rationnelle de l'interaction <i>kissing-loop</i> I/V	231
5.3 Prédiction de structures tridimensionnelles et bases de données d'ARN.....	233

5.4 Nanotechnologie à base d'ARN	236
5.5 Perspectives	238
Bibliographie.....	243

Liste des tableaux

1	Introduction.....	1
2	Étude du rôle des nucléotides de la boucle V dans la reconnaissance du substrat par le ribozyme VS.....	79
	Table 2.I : Summary of kinetic data for the VS ribozyme mutants.....	109
	Table 2.II : Summary of NMR data for the SLV fragments.....	110
3	Caractérisation de l'activation du substrat et de l'interaction <i>kissing-loop</i> I/V par spectroscopie RMN.....	119
	Table 3.I : Structural statistics.....	151
	Supplementary Table 3.I : Chemical shifts (ppm) of detectable G/U imino protons for SLI variants either in their free form or bound to SLV RNA. ^a	167
	Supplementary Table 3.II : Chemical shift (ppm) of detectable G/U imino protons for SLV either in its free form or bound to SLI variants. ^a	168
	Supplementary Table 3.III : Chemical shift assignments of exchangeable ¹ H and ¹⁵ N for the SLIDs2/SLV complex. ^a	169
	Supplementary Table 3.IV : Chemical shift assignments of non-exchangeable ¹ H and ¹³ C for the SLIDs2/SLV complex. ^a	173
	Supplementary Table 3.V : Summary of intermolecular NOEs at the kissing-loop junction of the SLIDs2/SLV complex.....	177
4	Caractérisation thermodynamique de l'interaction <i>kissing-loop</i> I/V et de l'activation du substrat.....	179
	Table 4.I : ITC data for binding of SLV variants to a preshifted SLI substrate (SLIps).....	211
	Table 4.II : Comparison of thermodynamic and kinetic data associated with the I/V kissing-loop interaction in the VS ribozyme.....	212
	Table 4.III : ITC data for binding of SLI variants to SLV.....	213
	Table 4.IV : Comparison between predicted ΔG and experimental ΔG values (in kcal/mol) for several kissing-loop complexes.....	214

Table 4.V : ITC data for binding of non-cleavable SLI substrates to SLV and VS Rz.....	216
5 Discussion et perspectives.....	223

Liste des figures

1	Introduction.....	1
	Figure 1.1 : Représentation schématique de l'importance de l'ARN dans la synthèse des protéines.	3
	Figure 1.2 : Structures secondaires et tridimensionnelles de plusieurs ribozymes.	6
	Figure 1.3 : Les réactions de clivage catalysées par les ribozymes.	8
	Figure 1.4 : Représentation schématique du mécanisme d'action du ribozyme <i>hammerhead</i> dans le contexte d'une application thérapeutique.	12
	Figure 1.5 : Séquence et structure secondaire de l'élément auto-clivant du ribozyme VS de <i>Neurospora</i>	13
	Figure 1.6 : Cycle de réplication du plasmide VS.	14
	Figure 1.7 : Séquences et structures secondaires des deux groupes représentant le substrat du ribozyme VS.	19
	Figure 1.8 : Réarrangement de la structure du substrat.	20
	Figure 1.9 : Effet du magnésium sur la structure de la boucle terminale du domaine V	23
	Figure 1.10 : Comparaison des conformations inactive et active de la boucle interne du substrat.	25
	Figure 1.11 : Représentation du domaine catalytique du ribozyme VS et de ses éléments structuraux importants.	27
	Figure 1.12 : Représentation schématique de la formation du site actif du ribozyme VS.	30
	Figure 1.13 : Mécanisme catalytique de type acide-base générale.	31
	Figure 1.14 : Mécanisme proposé utilisé par le ribozyme VS pour catalyser la réaction de clivage.	33
	Figure 1.15 : Éléments structuraux de la boucle A730 du domaine VI du ribozyme VS.	35

Figure 1.16 : Représentation schématique de l'organisation des sites actifs des ribozymes <i>hairpin</i> et VS.....	36
Figure 1.17 : Représentation des différents types d'interactions entre l'ARN et le magnésium.	39
Figure 1.18 : Résumé des expériences d'interférence à l'agent <i>ENU</i> et aux phosphorothioates.	40
Figure 1.19 : Interaction du magnésium avec la boucle interne A730 et la boucle terminale du domaine V.....	42
Figure 1.20 : Modèles à basse résolution de la structure tridimensionnelle du ribozyme VS.....	45
Figure 1.21 : Représentation de plusieurs éléments communs de structure secondaire.	46
Figure 1.22 : Représentation du motif de type <i>U-turn</i>	48
Figure 1.23 : Représentation du motif de type <i>kissing-loop</i>	50
Figure 1.24 : Composition chimique de l'ARN et caractéristiques spectrales des différents protons de l'ARN.....	56
Figure 1.25 : Représentation schématique de spectres RMN à une ou plusieurs dimensions.....	57
Figure 1.26 : Dispersion spectrale des différents atomes de carbone et d'azote qui composent l'ARN.....	58
Figure 1.27 : Couplage dipolaire résiduel (<i>RDC</i>).....	60
Figure 1.28 : Diagramme expliquant les étapes générales du protocole de détermination de structure par spectroscopie RMN.	62
Figure 1.29 : Diagramme décrivant les étapes du protocole de synthèse et purification d'échantillon pour une étude par spectroscopie RMN.....	63
Figure 1.30 : Attribution des protons imino.	66
Figure 1.31 : Attribution des protons non échangeables par la marche séquentielle.	68
Figure 1.32 : Différents types de contraintes RMN.	71
Figure 1.33 : Représentation des étapes générales du calcul de structure.....	72
Figure 1.34 : Titrage calorimétrique isotherme (<i>ITC</i>).....	73

2	Étude du rôle des nucléotides de la boucle V dans la reconnaissance du substrat par le ribozyme VS	79
	Figure 2.1 : Substrate recognition by the <i>Neurospora</i> VS ribozyme.....	112
	Figure 2.2 : NMR structures of the SLV loop determined in the absence (left) and presence (right) of magnesium ions [15, 16], showing the large conformational change of U700 (blue).	113
	Figure 2.3 : SLV loop mutants characterized kinetically and structurally in this study.	114
	Figure 2.4 : Kinetic analysis of the U700G Rz mutant.....	115
	Figure 2.5 : Conformation of SLV RNA fragments.	116
	Figure 2.6 : NMR studies of SLV loop mutants.	117
	Figure 2.7 : Model of the SLI/SLV interaction.....	118
3	Caractérisation de l'activation du substrat et de l'interaction <i>kissing-loop</i> I/V par spectroscopie RMN	119
	Figure 3.1 : Primary and secondary structures of RNAs used in this study.	152
	Figure 3.2 : NMR evidence for formation of the SLIsbΔIa/SLV complex.....	153
	Figure 3.3 : NMR evidence of structural rearrangements in SLI variants.....	154
	Figure 3.4 : NMR evidence for formation of the SLIds2/SLV complex.....	155
	Figure 3.5 : Stereoview of the NMR structures of the SLIds2/SLV kissing-loop complex.	156
	Figure 3.6 : U-turn structures in the SLIds2 and SLV loops of the SLIds2/SLV complex.....	157
	Figure 3.7 : Structural characteristics of the I/V kissing-loop junction.....	158
	Supplementary Figure 3.1 : Native gel electrophoresis of several RNAs investigated.....	159
	Supplementary Figure 3.2 : NMR evidence for formation of the SLIsbΔIa/SLV complex.....	160
	Supplementary Figure 3.3 : NMR evidence for formation of the SLIsbΔIa _{GU} /SLV complex.....	161
	Supplementary Figure 3.4 : NMR evidence for formation of the SLIps/SLV complex.....	162

Supplementary Figure 3.5 : NMR evidence for formation of the SLIps Δ Ia/SLV complex.....	163
Supplementary Figure 3.6 : NMR evidence for formation of the SLIds1/SLV complex.	164
Supplementary Figure 3.7 : Interhelical angle distribution for the 20 lowest-energy structures of the SLIds2/SLV complex.	165
Supplementary Figure 3.8 : Superposition of the loop structures of SLV free and bound to SLIds2.....	166
4 Caractérisation thermodynamique de l'interaction <i>kissing-loop</i> I/V et de l'activation du substrat	179
Figure 4.1 : Primary and secondary structures of the <i>Neurospora</i> VS ribozyme.	217
Figure 4.2 : Primary and secondary structures of SLI and SLV RNAs used in this study.	218
Figure 4.3 : ITC data analysis of wild-type SLV binding to SLIps at 20 mM MgCl ₂	219
Figure 4.4 : Magnesium-ion dependence of the I/V kissing-loop interaction.	220
Figure 4.5 : Primary and secondary structures of other kissing-loop interactions considered in this study.....	221
5 Discussion et perspectives.....	223
Figure 5.1 : Représentation schématique d'un modèle de la reconnaissance et de l'activation du substrat par le ribozyme VS de <i>Neurospora</i>	227
Figure 5.2 : Représentation de la structure secondaire du ribozyme VS entier avec plusieurs structures tridimensionnelles des domaines isolés connus.....	230
Figure 5.3 : Représentation des approches de modélisation sans (panneau du haut) et avec gabarit (panneau du bas).	234
Figure 5.4 : Représentation schématique des différents modules utilisés dans la nanotechnologie à base d'ARN.....	237

Liste des abréviations

ADN : acide désoxyribonucléique

ARN : acide ribonucléique

ARNas : ARN antisens

ARNase P : ribonucléase P

ARNlnc : long ARN non codant

ARNm : ARN messenger

ARNnc : ARN non codant

ARNr : ARN ribosomal

ARNsn : petit ARN nucléaire

ARNsno : petit ARN nucléolaire

ARNt : ARN de transfert

COSY : *correlation spectroscopy*

CoTC : *co-transcriptional cleavage* (clivage co-transcriptionnel)

CPEBP3 : *cytoplasmic polyadenylation element-binding protein 3* (protéine 3 liant l'élément cytoplasmique de polyadénylation)

EMSA : electrophoretic mobility shift assay (essai de mobilité sur gel d'électrophorèse)

ENU : *N-ethyl N-nitrourea* (N-nitroso N-éthylurée)

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

HDV : *hepatitis delta virus* (virus de l'hépatite D)

HSQC : *heteronuclear single quantum coherence*

Hz : hertz

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

kcal : kilocalorie

M : molaire

miARN : micro ARN

NOE : *nuclear Overhauser effect*

NOESY : *NOE spectroscopy*

nt : nucléotide

ppm : partie par million

RDC : *residual dipolar coupling* (couplage dipolaire résiduel)

RMN : résonance magnétique nucléaire

rmsd : *root-mean-square deviation* (racine de la moyenne des déviations au carré)

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

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

siARN : petit ARN interférent

SL : *stem-loop* (tige-boucle)

SPR : *surface plasmon resonance* (résonance plasmonique de surface)

TGF : *immunosuppressive transforming growth factor* (facteurs de croissance immunosuppresseurs endogènes)

TOCSY : *total correlation spectroscopy*

TTP : *thiamine pyrophosphate* (thiamine pyrophosphate)

VIH : virus de l'immunodéficience humaine

VS : Varkud satellite

W-C : Watson-Crick

*À ceux qui ont été patients et qui ont eu
confiance, ma famille!*

Remerciements

Je voudrais en premier lieu remercier ma directrice de recherche, le Dr Pascale Legault, pour m'avoir accueillie dans son laboratoire. J'ai eu la chance de faire partie de l'équipe originale du laboratoire, puisque je suis arrivée en tant que stagiaire d'été lors de son arrivée au département de biochimie. Je voudrais aussi exprimer ma reconnaissance envers Pascale pour tout le temps investi dans ma formation tant au niveau technique, pour les manipulations en laboratoire et la spectroscopie RMN, qu'au niveau des habiletés acquises en tant que scientifique. En effet, j'ai eu l'opportunité d'évoluer dans une équipe où les notions fondamentales de la rigueur scientifique et de l'analyse critique sont prioritaires et toujours mises de l'avant. Je voudrais ajouter un merci particulier pour les heures passées à m'enseigner l'importance des nombreux paramètres à ajuster pour chaque enregistrement de spectres de RMN, la calibration des instruments, les changements de sondes, etc. Je suis aussi très reconnaissante de la confiance qu'elle m'a montrée à travers des tâches telles que la formation d'étudiants, la revue de manuscrits et la coordination dans son projet d'implantation de plateforme de biologie structurale. De plus, avec Pascale je n'ai pas seulement appris à faire de la bonne science, mais aussi de la belle science, puisque pour chaque présentation de nos travaux, que ce soit oralement, par affiche ou par une publication dans un journal, il nous fallait toujours produire les plus belles figures.

Les années passées au laboratoire de Pascale Legault ont été très enrichissantes scientifiquement, mais ont aussi été agrémentées d'une vie «extra-labo» très intéressante. En effet, je voudrais remercier les membres des équipes Legault et Omichinski pour toutes les activités sportives (ski, escalade, canot camping) et sociales (5 à 7, soupers thématiques, etc). J'ai développé des amitiés toutes particulières avec Philippe Lampron et Geneviève Desjardins, avec qui j'ai débuté mes études graduées. Avec les années, plusieurs membres se sont ajoutés à l'équipe, notamment Alexandre Desjardins, Éric Bonneau, Julie Lacroix-Labonté, Pierre Dagenais, Geneviève Di Tomasso, Nicolas Girard, Jonathan Bouvette et Alix Salvail-Lacoste. J'ai eu le privilège de faire mon PhD en côtoyant ces gens, qui sont plus que de simples collègues de

travail, ce qui a permis de mettre du soleil dans les journées plus grises de la recherche. Des remerciements plus particuliers à Alexandre, Éric et Nico pour de nombreuses discussions scientifiques ou non scientifiques, des voyages quelque peu extrêmes, des marathons de gels trop tôt le matin et de multiple *brainstorm* pour mettre mes idées en ordre.

En terminant, je veux remercier ma famille, ma mère Andrée, mon père Doris et mon frère Pierre-Luc qui ont toujours su me donner le support nécessaire et qui n'ont jamais douté de ma réussite tout au long de mes études. Un merci particulier à mes amies Émilie, Valérie et Hélène qui, malgré leurs nombreuses questions sur la RMN et les ribozymes, m'ont toujours encouragée et n'ont jamais douté de mes capacités. Finalement, un merci tout spécial à Christopher, mon conjoint, celui qui a écouté toutes mes présentations et qui a su, tout au long de mon PhD, m'encourager, me motiver et me montrer chaque jour qu'il est très fier de ce que j'ai accompli.

1 Introduction

1.1 ARN non codants et ribozymes

1.1.1 Importance des ARN et versatilité des ARN non codants

Il est maintenant reconnu que les ARN non codants (ARNnc) jouent plusieurs rôles essentiels en biologie. En effet, l'ARN est considéré comme un régulateur important de l'expression génique. Cependant, cette reconnaissance de l'importance de l'ARN est relativement récente puisque pendant de nombreuses années, seulement trois types d'ARN étaient connus et ceux-ci participent à la synthèse des protéines (Figure 1.1). Tout d'abord, l'ARN messager (ARNm) sert de guide à la machinerie traductionnelle, les ARN de transfert (ARNt) servent de véhicule aux différents acides aminés, puis les ARN ribosomiaux (ARNr) représentent 60% de la composition du ribosome (Figure 1.1). Jusque vers le début des années 1980, les protéines étaient considérées comme les seules macromolécules pouvant participer aux différentes fonctions cellulaires, grâce à leur capacité à adopter une structure tridimensionnelle active permettant de catalyser des réactions chimiques diverses. Les ARN, de leur côté, jouaient simplement le rôle d'intermédiaire dans le transfert de l'information génétique. Cependant, cette vision de l'ARN a basculé au début des années 1980, lorsque des ARN ne codant pas pour des protéines, les ARN non codants, se sont révélés fonctionnellement importants.

Au début des années 1980, les équipes de recherche de Thomas R. Cech et de Sydney Altman [1-3] ont identifié les premiers ARN capables de catalyser une réaction chimique. Les découvertes de l'intron auto-clivant de l'ARN pre-ribosomal de *Tetrahymena thermophila* et de la ribonucléase P (*ribonuclease P*, RNase P) de *Escherichia coli* leur a valu le prix Nobel de chimie en 1989. Cette classe d'ARN avec activité catalytique a été nommée « ribozyme », une combinaison des termes *ribonucleic* et *enzyme*.

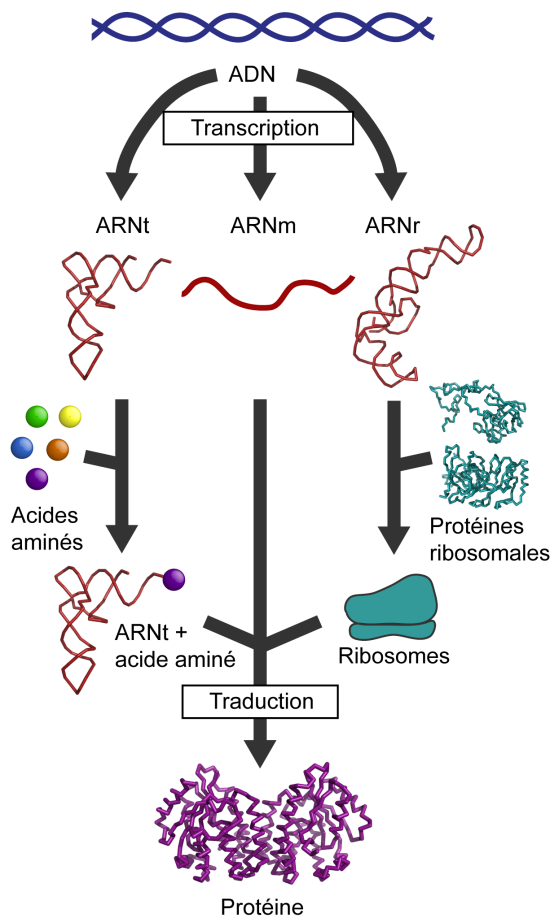


Figure 1.1 : Représentation schématique de l'importance de l'ARN dans la synthèse des protéines. La synthèse de l'ARN est un transfert d'information de l'ADN à l'ARN par un processus appelé transcription. Les trois types majeurs d'ARN synthétisés sont les ARNm, les ARNt et les ARNr. Les ARNt sont chargés avec un acide aminé et les ARNr s'associent avec les protéines ribosomales pour former le ribosome. Par un processus appelé traduction, le ribosome fait la lecture de l'ARNm et traduit la séquence d'acides nucléiques en séquence d'acides aminés à l'aide des ARNt pour faire la synthèse des protéines.

Quelques années plus tard, le groupe de recherche de Masayori Inouye [4] a démontré que la régulation de l'expression des gènes pouvait s'effectuer par des ARN simple brin. Ces ARN complémentaires à l'ARNm, les ARN antisens (*Antisense RNA*, ARNas) bloquent le processus de traduction après avoir lié leur cible. Dans les années 1990, les groupes de recherche d'Andrew Fire, Craig C. Mello et Victor Ambros [5-7] ont démontré que de petits ARN double brin interférents endogènes, les microARN

(miARN) et les petits ARN interférents (siARN) pouvaient réguler l'expression des gènes. Ces ARN interférents sont activés et lient ensuite spécifiquement une séquence cible sur l'ARNm, réprimant ainsi l'expression du gène. Cette découverte a d'ailleurs valu aux chercheurs Andrew Fire et Craig C. Mello le prix Nobel de médecine et physiologie en 2006.

Au début des années 2000, les groupes de recherche de Ronald R. Breaker et Evgeny Nudler [8-10] ont identifié les riborégulateurs (*riboswitch*) à la flavine (*flavin mononucleotide, FMN*) et à la thiamine pyrophosphate (*thiamine pyrophosphate, TTP*) fondant ainsi une nouvelle classe d'ARN non codants. Ces riborégulateurs, retrouvés principalement dans la partie 5' non traduite de certains ARNm bactériens, contiennent une région aptamère et une plateforme d'expression leur permettant d'agir comme interrupteur génique. La liaison d'un métabolite à la région aptamère induit un changement de conformation de l'ARN permettant de réguler l'expression du gène situé en aval.

De plus, de récentes recherches ont permis de mettre en évidence l'importance de nombreux autres ARNnc, incluant ceux impliqués dans l'organisation et le fonctionnement de larges complexes ribonucléoprotéiques tels le ribosome et le spliceosome. Ces grands complexes sont, entre autre, composés d'ARNr, d'ARNt, de petits ARN nucléolaires (ARNsno) et de petits ARN nucléaires (ARNsn). Puis, plus récemment, les longs ARN non codants (*long non-coding RNA, ARNlnc*) ont été identifiés, lesquels sont composés de plus de 200 nucléotides (nt) [11-13]. Ces derniers sont impliqués dans un large éventail de fonctions, telles que le remodelage de la chromatine et l'organisation structurale subcellulaire [13-16].

Les 30 dernières années de recherche sur l'ARN ont donc permis de dévoiler l'omniprésence et la diversité de fonction de l'ARN, de la catalyse de différents types de réactions jusqu'à la régulation des gènes en passant par l'organisation et l'activité des larges complexes ribonucléoprotéiques. Cette diversité de fonction de l'ARN semble être en accord avec l'hypothèse d'une forme de vie primitive basée sur l'ARN. Cette

hypothèse, dont les fondations ont été posées par les chercheurs Francis Crick, Leslie Orgel et Carl Woese [17-19] et ensuite précisée par Walter Gilbert [20], propose l'existence d'un monde ARN (*RNA world*) qui aurait existé sur la Terre avant l'apparition des premières formes de vie. Dans ce monde primitif, l'ARN aurait agi comme porteur de l'information génétique, ainsi que comme principal catalyseur biochimique responsable de sa propre réplication [20]. Bien que l'existence du monde ARN demeure hypothétique, l'importance de l'ARN dans le monde d'aujourd'hui est maintenant clairement établie.

1.1.2 Les ribozymes

Tel que mentionné précédemment, les ribozymes sont des ARN capables de catalyser une réaction chimique. Pour y arriver, ils doivent tout d'abord se replier, c'est-à-dire adopter une structure tridimensionnelle active. Par exemple, c'est une structure bien définie qui permet au ribozyme de reconnaître spécifiquement un substrat et d'effectuer la catalyse enzymatique [21]. La relation structure-fonction qui définit le fonctionnement des ribozymes en fait des systèmes modèles d'intérêt pour étudier la structure et la fonction de l'ARN en plus d'explorer leur diversité de fonction.

Depuis la découverte des premiers ribozymes, plusieurs classes de ribozymes naturels ont été identifiées (Figure 1.2) et de nombreux ribozymes artificiels ont été développés.

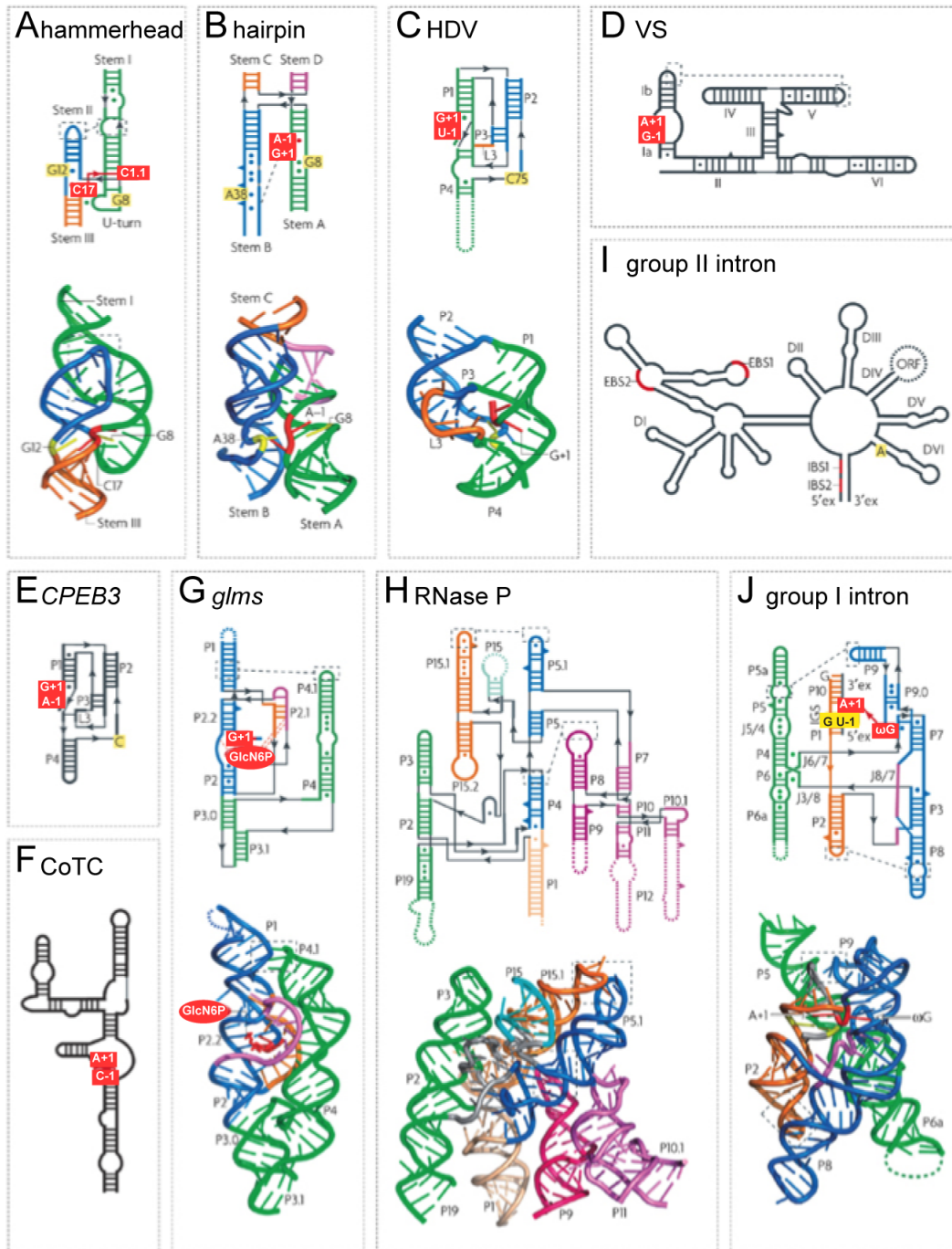


Figure 1.2 : Structures secondaires et tridimensionnelles de plusieurs ribozymes. Les paires de bases Watson-Crick et non Watson-Crick retrouvées dans la structure secondaire sont représentées par des traits pleins et des cercles respectivement. Les nucléotides non appariés (*bulge*) sont représentés par un triangle. Les lignes pointillées noires représentent des interactions tertiaires importantes. Les ribozymes dont la structure tridimensionnelle à haute

résolution était connue en 2007 sont représentés avec les mêmes couleurs que celles utilisées pour la structure secondaire. Les nucléotides situés de part et d'autre du site de clivage sont encadrés en rouge et ceux impliqués dans la catalyse sont encadrés en jaune. Les classes de ribozyme représentées sont celles (A) du *hammerhead*, (B) du *hairpin*, (C) du *HDV*, (D) du *VS*, (E) du *CPEB3*, (F) du *CoTC*, (G) du *glms*, (H) de l'ARNase P, (I) de l'intron du groupe II et (J) de l'intron du groupe I. (Tiré de [22]).

Les ribozymes naturels catalysent principalement le clivage réversible d'un lien phosphore-oxygène spécifique par une réaction de transestérification [23]. Ceux-ci peuvent être séparés en deux groupes selon les produits réactionnels générés. Le premier groupe est composé d'un ensemble de ribozymes appelé les ribozymes nucléolytiques auto-clivants qui se divise en six classes majoritaires : *hammerhead* [24], *hairpin* [25], du virus de l'hépatite D (*hepatitis delta virus, HDV*) [26, 27], Varkud satellite de *Neurospora* (*VS*) [28], *glms* activé par la glucosamine-6-phosphate (*glucosamine-6-phosphate activated*) [29] et *twister* [30, 31]. Deux autres ribozymes, similaires au *HDV* et au *hammerhead*, font aussi partie de ce groupe, celui du gène de la protéine 3 liant l'élément cytoplasmique de polyadénylation (*cytoplasmic polyadenylation element-binding protein 3 HDV-like, CPEBP3*) [32] et celui du gène de la protéine CLEC2 (*CLEC2 ribozyme*) [33], respectivement. La réaction catalysée par ces ribozymes implique l'attaque nucléophile du 2'-oxygène d'un ribose sur le groupement phosphate en 3' adjacent, ce qui résulte en un clivage de l'ARN générant deux produits, un avec une extrémité 2'-3' cyclique phosphate et l'autre avec une extrémité 5'-OH (Figure 1.3 A) [22, 23]. Le deuxième groupe est composé des ribozymes d'intron auto-clivant des groupes I et II effectuant l'épissage des ARN [34, 35] et de l'ARNase P qui est responsable de la maturation des ARNt [2]. Ce groupe contient également le ribozyme de clivage co-transcriptionnel (*co-transcriptional cleavage, CoTC*), impliqué dans la terminaison de la transcription des gènes de la β -globine [36] et les ribozymes similaires à l'intron du groupe I (*group I-like*) formant une structure en lasso (*lariat*) de branchement 2'-5' servant de *cap* pour protéger l'ARNm [37]. Les réactions catalysées par ce deuxième groupe sont initiées par une attaque nucléophile provenant généralement d'un groupement hydroxyle endogène (intron du groupe II [38] et similaire au groupe I [37]) exogène (intron du groupe I [39]) ou d'une molécule d'eau

(ARNase P [2], intron du groupe II [40]) (Figure 1.3 B). Néanmoins, pour ces deux groupes, la réaction de clivage s'effectue par une attaque nucléophile qui est suivie d'une inversion de la configuration stéréochimique des atomes d'oxygène du groupement phosphate à cliver. Ceci suggère que la réaction s'effectue par un mécanisme de substitution nucléophile bimoléculaire (S_N2) passant par la formation d'un intermédiaire réactionnel trigonal bipyramidal (Figure 1.3) [41].

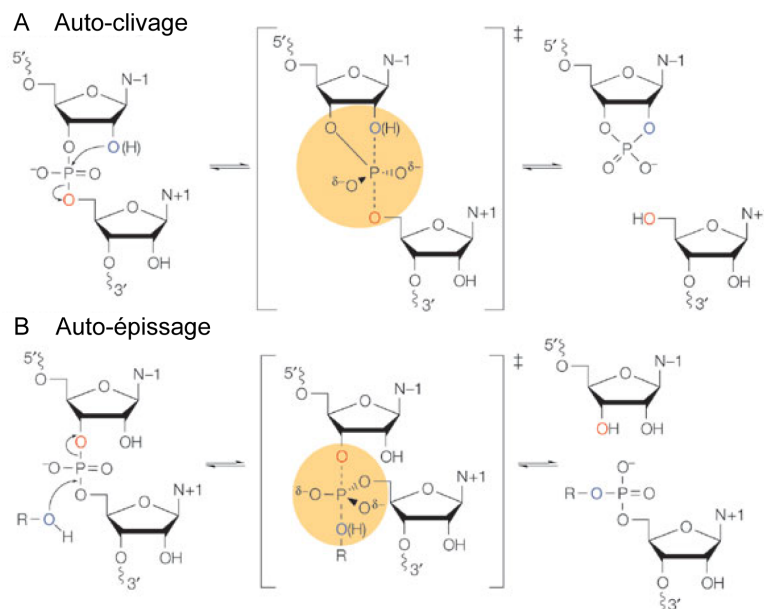


Figure 1.3 : Les réactions de clivage catalysées par les ribozymes. La réaction d'auto-clivage représentée en (A) est catalysée par les ribozymes nucléolytiques. La réaction représentée en (B) est catalysée par les ribozymes impliqués dans l'épissage. L'attaque nucléophile est effectuée par le groupement identifié en bleu et le groupe partant est représenté en rouge. L'intermédiaire réactionnel trigonal bipyramidal est identifié par un cercle jaune. (Tiré de [41]).

Outre les réactions de transestérification et d'hydrolyse, d'autres types de réactions peuvent être catalysés par les ribozymes. C'est le cas du transfert de groupement peptidique permettant la formation du lien peptidique lors de la synthèse protéique par le ribosome [42]. De plus, une panoplie d'autres réactions peuvent être catalysées par des ribozymes développés *in vitro* (ribozymes artificiels) [43], telles que la réaction de cyclo-addition Diels-Alder [44, 45] et la synthèse de pyrimidine [46].

Comparativement aux protéines qui bénéficient de vingt acides aminés aux caractéristiques structurales et chimiques variées comme unité structurale de base, les ribozymes ne comptent généralement que sur quatre bases azotées différentes pour composer toute leur séquence. Malgré cela, l'adoption de structures tridimensionnelles spécifiques rend ces ribozymes uniques et capables de catalyser plusieurs types de réactions.

1.1.3 Intérêt biochimique et médical des ARNnc

Tel que mentionné précédemment, les ARNnc se retrouvent au centre de plusieurs processus cellulaires essentiels tels que la synthèse protéique (ARNr, ARNt, RNase P, ARNsno), l'épissage de l'ARN (introns auto-clivants, ARNsn) et la régulation de l'expression des gènes (riborégulateurs, miARN, siARN, ARNas). De plus, plusieurs de ces ARNnc sont capables de reconnaître et lier des séquences d'ARN spécifiques ou encore de répondre à la liaison d'un métabolite. Il est donc peu surprenant que ces ARN soient une source d'intérêt pour plusieurs groupes de recherche visant à les utiliser comme base pour l'élaboration de différents outils biochimiques et approches thérapeutiques.

Tout d'abord, l'ingénierie de plusieurs ARNnc a permis la création de différents outils biochimiques utilisés régulièrement dans les laboratoires de recherche. Par exemple, la répression spécifique de certains gènes (*gene knockdown*) est couramment effectuée par l'utilisation de différents ARN interférents conçus pour cibler une séquence d'intérêt [47, 48]. De plus, l'utilisation de ribozymes en tant qu'enzymes de restriction s'est montrée fort utile dans la production d'ARN pour des études biochimiques et structurales. En effet, ces ribozymes permettent l'obtention d'extrémités 5' et 3' homogènes suite à une transcription *in vitro* [49-52]. Aussi, l'utilisation combinée de ribozymes et de riborégulateurs a permis de développer des protocoles de purification d'ARN par affinité qui sont rapides et permettent l'utilisation de conditions

non dénaturantes tout en permettant l'obtention d'ARN aux extrémités homogènes [53-56].

Ensuite, les ribozymes et riborégulateurs se sont avérés prometteurs pour la création de biosenseurs, notamment dans le diagnostic médical et l'analyse en laboratoire [57-60]. Par exemple, des ribozymes allostériques ont été développés pour permettre la détection d'une grande variété de ligands tels que des ions, des protéines, des bactéries, des nucléotides et différents métabolites dans des mélanges plus ou moins complexes [57, 59]. Ces ribozymes allostériques comprennent une partie aptamère et une partie ribozyme. La partie aptamère est une molécule d'ARN ou d'ADN naturelle ou évoluée artificiellement qui lie son ligand de façon spécifique. De façon générale, la liaison du ligand à l'aptamère cause un changement dans l'activité catalytique de la partie ribozyme [61, 62], ce qui permet la détection de ce ligand. Par contre, bien que l'utilisation de ces ribozymes allostériques soit intéressante, davantage de biosenseurs ont été développés avec uniquement une base d'aptamère d'ARN ou dans la plupart des cas d'ADN [60, 63, 64], sans nécessiter de ribozyme. Ces biosenseurs sont néanmoins couplés à différents systèmes de détection pour obtenir le signal de la liaison du ligand, comme par exemple de la fluorescence ou encore de l'électrochimie [64]. Bien que les aptamères d'ADN soient plus stables, la stabilité et la résistance aux ribonucléases des aptamères d'ARN peuvent être améliorées par l'introduction de groupements 2'-deoxy, 2'-NH₃, 2'-F ou 2'-OMe [60, 65]. Ainsi, avec une stabilité améliorée, une capacité à adopter une structure tertiaire et lier spécifiquement un ligand avec une grande affinité, l'ARN est une molécule prometteuse comme biosenseur et outil diagnostique. Plusieurs aptamères d'ARN font l'objet de recherches, notamment pour le diagnostic de différents cancers et maladies [60] et la détection de pathogènes infectieux et alimentaires [63, 66].

Aussi, plusieurs ARNnc ont été utilisés dans le développement de thérapies géniques, qui sont des stratégies visant à réparer ou modifier un gène ou encore en altérer son expression dans le but de traiter des maladies ou cibler des virus. Par exemple, le ribozyme *hammerhead*, un des ribozymes les plus étudiés et mieux

caractérisés, s'est avéré prometteur pour la conception d'agents thérapeutiques. Notamment, il a été la base dans la conception d'ANGIOZYME, un agent anti-cancer et d'OZ1, un agent antiviral contre le VIH-1 (virus de l'immunodéficience humaine de type 1, VIH-1), ayant tous deux fait l'objet d'essais cliniques [67-69]. Néanmoins, bien que ces derniers, ainsi que certains autres ribozymes se soient montrés prometteurs [67-71], il semble que les ARN interférents et antisens démontrent plus de succès dans le développement de thérapies géniques. Ceux-ci sont notamment utilisés pour inhiber ou encore réduire l'expression de gènes surexprimés dans les tumeurs [72-74]. Par exemple, le vaccin FANG par Gradalis Inc, à base d'ARN interférent, cible l'ARNm codant pour la protéine furine [75]. Ce vaccin s'est montré efficace pour diminuer l'expression des facteurs de croissance immunosuppresseurs endogènes TGFβ1 et β2 (*immunosuppressive transforming growth factor, TGF*) chez des patients atteints de cancer avancé, c'est pourquoi il se retrouve actuellement en essai clinique de phase II [75]. Plusieurs autres ARN interférents et antisens se retrouvent également en essai clinique pour traiter des cancers [74]. Même si les ribozymes et les ARN interférents semblent prometteurs indépendamment comme agents thérapeutiques, une approche combinée de plusieurs de ces ARN pourrait procurer des avantages additionnels. Par exemple, une approche combinant un ribozyme et des ARN interférents contre le VIH s'est montrée positive et a récemment fait l'objet d'essais cliniques [76-79]. De plus, un des défis actuels dans l'utilisation des ARN interférents, des ARN antisens et des ribozymes est leur mode d'administration systémique et intracellulaire. À cet égard, les aptamères d'ARN s'avèrent fort intéressants, puisqu'ils peuvent servir, par exemple, de véhicule à ces ARN interférents en les transportant à l'endroit approprié en reconnaissant un type de cellule spécifique [60].

De toute évidence, il est clair que les outils de recherche et les agents thérapeutiques à base d'ARN sont très prometteurs. Cependant, leur fonctionnement requiert encore des améliorations, notamment en minimisant les erreurs de ciblage (*off-target effect*) et en optimisant les modes d'administration et l'efficacité des effets recherchés. Bien que beaucoup de travail demeure nécessaire pour mettre au point ces agents thérapeutiques, il a déjà été démontré que l'utilisation de l'ARN en tant que

médicament est possible. En effet, un aptamère d'ARN appelé MACUGEN [80] est distribué par Pfizer (MACUGENTM) au Canada depuis 2006 pour traiter la dégénérescence maculaire liée à l'âge (*age-related macular degeneration, AMD*).

Jusqu'ici, plusieurs outils biochimiques et agents thérapeutiques ont été conçus pour reconnaître une séquence simple brin d'ARN par un appariement Watson-Crick (W-C). Par exemple, le site de reconnaissance du substrat du ribozyme *hammerhead* est une région simple brin qui va se lier à l'ARNm cible par un appariement W-C (Figure 1.4). Cependant, plusieurs ARN fonctionnels dans les cellules sont repliés, souvent en forme de tige-boucle, plutôt que linéaire et simple brin. Ainsi, une façon intéressante d'améliorer le ciblage de certains outils biochimiques et des agents thérapeutiques serait d'exploiter un mode de reconnaissance plus complexe du substrat. Par exemple le ribozyme VS de *Neurospora*, bien qu'il soit plus complexe que le *hammerhead*, a l'avantage de pouvoir reconnaître un ARN replié.

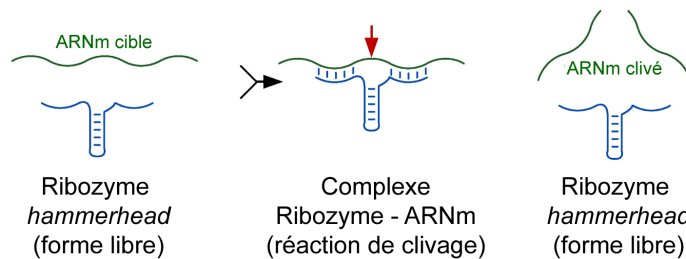


Figure 1.4 : Représentation schématique du mécanisme d'action du ribozyme *hammerhead* dans le contexte d'une application thérapeutique. Un ribozyme *hammerhead* est conçu pour cibler spécifiquement un ARNm cible. Le ribozyme lie l'ARNm en formant plusieurs appariements W-C. Le complexe ribozyme-ARNm est clivé par le ribozyme et celui-ci se dissocie en libérant les produits de clivage. Le ribozyme libre redevient alors disponible pour cliver une autre cible. (Adapté de [71]).

1.2 Le ribozyme VS de *Neurospora*

Le ribozyme Varkud satellite (VS) de *Neurospora* (Figure 1.5), qui fait partie du groupe des ribozymes nucléolytiques auto-clivants, a été découvert dans le laboratoire de Richard A. Collins en 1990 [28]. Il provient de l'ARN VS, un transcrit d'ADN

plasmidique retrouvé dans la mitochondrie de certaines souches du champignon *Neurospora*, dont Varkud-1c [28, 81].

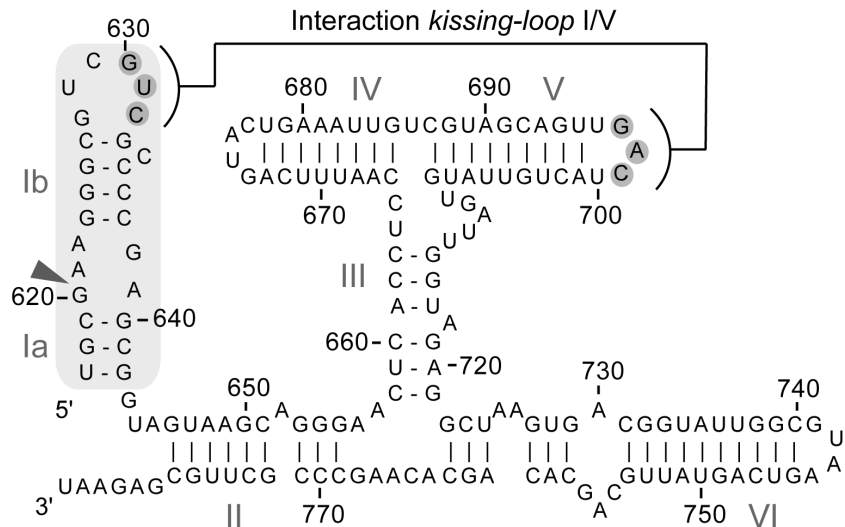


Figure 1.5 : Séquence et structure secondaire de l'élément auto-clivant du ribozyme VS de *Neurospora*. La partie encadrée en gris, contenant les tiges Ia et Ib, représente le substrat et la partie composée des domaines hélicoïdaux II à VI représente le domaine catalytique. La tête de flèche indique le site de clivage. Les nucléotides encerclés en gris foncé participent à l'interaction *kissing-loop* I/V. (Adapté de [82])

L'ARN VS est transcrit par une ARN polymérase mitochondriale qui génère un transcrit linéaire multimérique de cet ARN par un mécanisme de réplication circulaire similaire au *rolling circle* [83] (Figure 1.6 A) retrouvé chez d'autres classes de ribozymes [84]. Ce type de réplication est également utilisé pour les plasmides et les génomes circulaires de bactériophage [85]. L'ARN VS multimérique linéaire possède un élément auto-clivant lui permettant de passer de sa forme multimérique à une forme linéaire monomérique par une réaction de transestérification entre les nucléotides G620 et A621 [28] (Figure 1.6 B). Ensuite, l'ARN monomérique de 881 nt se circularise par une réaction d'auto-ligation et devient à ce moment un substrat pour l'enzyme *retroplasmid RT* qui, par transcription inverse, produira l'ADN complémentaire (ADNc) VS de polarité négative (Figure 1.6 C, D). L'ARN guide est ensuite déplacé ou dégradé pour permettre la synthèse du brin d'ADN de polarité positive, puis une réaction de ligation permet d'obtenir le plasmide VS initial double brin [83, 86] (Figure 1.6 E). Le

ribozyme VS catalyse donc des réactions de clivage et de ligation permettant sa propre réplication au sein de *Neurospora*.

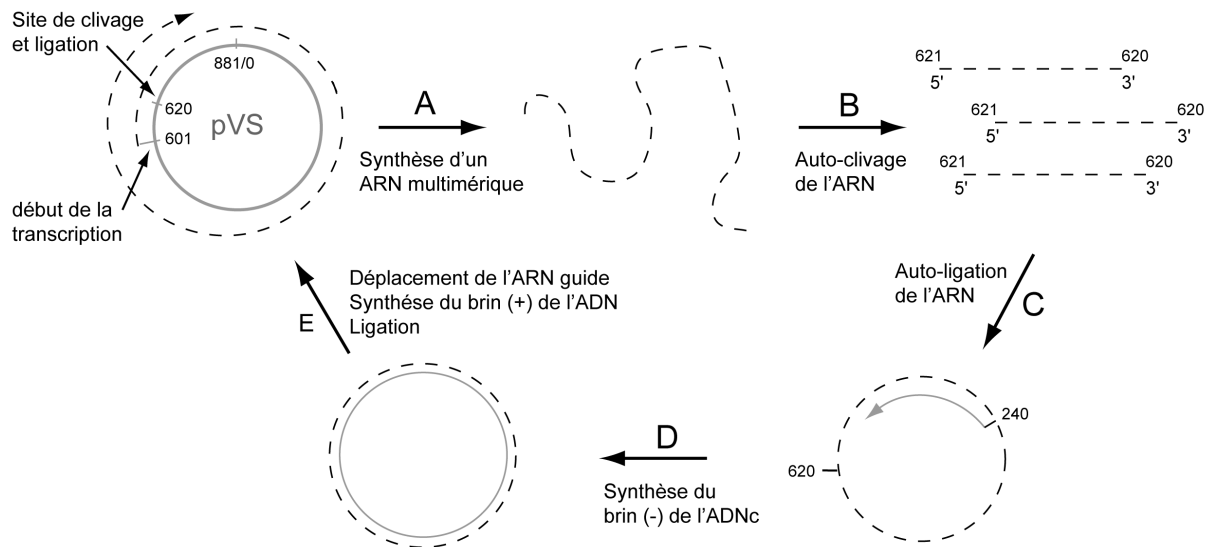


Figure 1.6 : Cycle de réplication du plasmide VS. (A) Le plasmide Varkud satellite (pVS) est transcrit en un ARN multimérique à partir de la position 601. (B) Cet ARN multimérique est composé de plusieurs monomères de 881 nt contenant un fragment capable d'auto-clivage (ribozyme VS). Le ribozyme catalyse l'auto-clivage du multimère, de façon à générer plusieurs monomères linéaires. (C) Le ribozyme catalyse l'auto-ligation de l'ARN monomérique pour générer un ARN monomérique circulaire. (D) L'enzyme *retroplasmid RT* utilise cet ARN circulaire pour faire la synthèse du brin (-) de l'ADNc. (E) L'ARN circulaire guide est ensuite déplacé, puis le brin (+) d'ADNc est synthétisé. Une étape de ligation va finalement permettre l'obtention du plasmide circulaire d'ADN VS initial. (Adapté de [83])

Un aspect particulièrement intéressant de la réaction de clivage du ribozyme VS est le fait qu'elle puisse non seulement s'effectuer de façon intramoléculaire, en *cis*, mais également de façon intermoléculaire, en *trans* [87]. En effet, le ribozyme peut être séparé en deux parties : un substrat composé d'une tige-boucle contenant le site de clivage et un domaine catalytique composé de cinq domaines hélicoïdaux [87] (Figure 1.5). Dans le ribozyme en *trans*, le domaine catalytique est capable de reconnaître le substrat et d'en catalyser le clivage [87]. Cette propriété de clivage en *trans* fait du ribozyme VS un modèle d'étude d'intérêt pour comprendre les éléments déterminants et essentiels à la reconnaissance du substrat. Une meilleure compréhension de son

fonctionnement est aussi importante pour la conception de cibles non naturelles pouvant être reconnues et clivées par le ribozyme VS.

1.2.1 Structures secondaire et tertiaire

La séquence minimale du ribozyme VS présentant une activité de clivage est d'environ 150 nt [88, 89]. Cette séquence est organisée en six domaines hélicoïdaux qui peuvent être divisés en deux sections principales : le substrat, représenté par le domaine I et le domaine catalytique, représenté par les domaines II à VI (Figure 1.5). Pour être capable de reconnaître son substrat et en catalyser le clivage ou la ligation, le ribozyme VS doit se replier pour adopter une structure tertiaire précise permettant à chaque domaine de s'orienter dans une configuration globale active. Cette structure tertiaire dépend de quatre éléments structuraux principaux : une interaction de type *kissing-loop* entre les boucles terminales des domaines I et V [82], deux jonctions à trois voies dans le domaine catalytique [87, 90], puis un site actif impliquant les boucles internes des domaines I et VI [91-105]. D'ailleurs, le repliement de ces éléments structuraux et l'activité de clivage sont favorisés par des cations divalents tels que le magnésium, le calcium ou le manganèse [82, 90, 106, 107]. Le repliement et l'activité du ribozyme sont également compatibles avec l'utilisation de certains cations monovalents en grande concentration, notamment le lithium [108].

Plusieurs études structure-fonction ont été effectuées *in vitro* à l'aide d'expériences biochimiques (mutagenèse et modifications chimiques) sur le ribozyme VS complet, ainsi qu'avec des expériences biophysiques de transfert d'énergie par résonance de type Förster (Förster *resonance energy transfert*, *FRET*), de spectroscopie de résonance magnétique nucléaire (RMN), de modélisation et de diffusion des rayons X aux petits angles (*small angle X-ray scattering*, *SAXS*) effectuées sur le ribozyme complet ou encore sur certains domaines isolés. Les sections suivantes présenteront plusieurs de ces travaux qui ont permis de caractériser les éléments de structure importants du ribozyme ainsi que son fonctionnement.

1.2.1.1 L'interaction *kissing-loop* I/V

Une interaction de type *kissing-loop* se définit par la formation de paires de bases entre les boucles terminales de deux tige-boucles. Dans le ribozyme VS, une interaction de ce type se forme entre les boucles terminales des domaines I et V et comprend trois paires de bases (Figure 1.5). Cette interaction dépendante du magnésium, appelée *kissing-loop* I/V, est la première interaction tertiaire à avoir été décrite pour le ribozyme VS. En plus de contribuer au repliement du ribozyme [92, 109], cette interaction permet la catalyse efficace des réactions de clivage et de ligation [82, 107]. De plus, l'interaction *kissing-loop* I/V est particulièrement intéressante, puisqu'elle est accompagnée d'un réarrangement structural du substrat qui l'active pour la catalyse [107, 109-112]. Tel qu'exposé en plus de détails ci-dessous, cette interaction est importante à plusieurs niveaux pour le fonctionnement du ribozyme, de la reconnaissance du substrat jusqu'à la catalyse en passant par le repliement et l'organisation du site actif.

1.2.1.1.1 Reconnaissance du substrat par l'interaction *kissing-loop* I/V

Dans le ribozyme VS, l'interaction *kissing-loop* I/V est caractérisée par trois paires de bases entre les nucléotides G630, U631 et C632 du domaine I et C699, A698 et G697 du domaine V (Figure 1.5) [82]. Des études exhaustives de mutagenèse sur les boucles terminales I et V démontrent que les trois paires de bases formant l'interaction *kissing-loop* sont importantes pour l'activité du ribozyme. Dans ces études, les mutations ponctuelles de chaque bases (G630, U631, C632, G697, A698 et C699) ont occasionné une diminution d'activité allant jusqu'à plus de 1000 fois [82]. De plus, chacune des trois paires de bases peut être individuellement substituée par au moins une autre paire de bases W-C non naturelle tout en conservant l'activité de clivage du ribozyme [82]. Bien qu'il y ait une certaine flexibilité dans les paires de base de l'interaction, quelques contraintes additionnelles s'imposent au niveau de la séquence des boucles I et V pour maintenir l'activité du ribozyme. En effet, seulement des purines en position 630 et 698,

des uridines en positions 628 et 696 et une cytosine en position 629 sont compatibles avec l'activité de clivage du ribozyme [82]. Ces contraintes particulières combinées aux différents patrons de protection chimique observés dans les boucles I et V ont permis de proposer la présence de motifs de type *U-turn* dans ces boucles [82, 90]. Ce type de motif est principalement caractérisé par : une séquence consensus UNR (U=uridine, N=adénine, guanine, cytidine ou uridine, R=adénine ou guanine), un changement de direction du squelette ribose-phosphate d'environ 180° entre le U et le N lui donnant une forme de « U » et un empilement des bases situées en 3' du U exposant leur face W-C dans le sillon mineur. Dans les boucles I et V, la séquence consensus serait composée des nucléotides U628-C629-G630 et U696-G697-A698, respectivement [82]. Bien que les données structurales actuelles ne soient pas suffisantes pour confirmer la présence de ces motifs *U-turn* dans l'interaction *kissing-loop* I/V, des études de spectroscopie RMN démontrent qu'un tel motif est présent dans le domaine V isolé [113, 114]. De plus, malgré le fait que toutes les données biochimiques et structurales actuelles définissent l'interaction *kissing-loop* I/V par la formation de trois paires de bases, il est probable que cette interaction soit plus complexe et implique davantage de nucléotides.

L'ensemble des résultats discutés ci-haut démontre clairement l'importance de l'interaction *kissing-loop* I/V pour le repliement et le fonctionnement du ribozyme VS. Malgré cela, dans des circonstances particulières où le substrat est bloqué dans une conformation active, le ribozyme peut demeurer fonctionnel même en absence de l'interaction *kissing-loop* [109, 112]. Par contre, bien que l'interaction *kissing-loop* puisse être accessoire dans cette circonstance, elle augmente tout de même considérablement la vitesse de réaction de clivage du ribozyme [109].

1.2.1.1.2 Activation du substrat par l'interaction *kissing-loop* I/V

La formation de l'interaction *kissing-loop* I/V est accompagnée d'un réarrangement de la structure du substrat, ce qui l'active pour la catalyse. Ceci implique que le substrat naturel du ribozyme VS peut adopter deux conformations : inactive et active, aussi appelées *unshifted* et *shifted* respectivement. Puisque le substrat naturel

peut passer d'une conformation *unshifted* inactive à une conformation *shifted* active, on réfère à une conformation dite *shiftable*. Les conformations *unshifted* et *shifted* sont toutes deux composées de quatre éléments structuraux : deux tiges (Ia et Ib) séparées par une boucle interne contenant le site de clivage entre les nucléotides G620 et A621, puis une boucle terminale qui ferme la tige Ib (Figure 1.5). De ces quatre éléments de structures, seulement la tige Ia demeure inchangée suite au réarrangement structural.

Les conformations adoptées par le substrat ont été démontrées initialement par des expériences de sélection *in vitro* (*systematic evolution of ligands by exponential enrichment, SELEX*), ainsi que des analyses de covariation et de modifications chimiques. Différentes séquences de substrat ont été obtenues par un protocole de sélection basé sur la réaction de ligation du ribozyme [107]. L'analyse de ces séquences démontre peu de covariation et moins d'appariement en paire de bases que le modèle initial (Ia et Ib représenté en Figure 1.5). L'analyse de la structure secondaire prédite pour ces séquences met en évidence deux groupes : le premier est caractérisé par des ARN composés des mêmes cinq paires de bases prédites dans le modèle initial (Figure 1.7 A) et le deuxième groupe est composé d'ARN formant une tige Ib différente dans laquelle les nucléotides G623, G624 et G625 sont appariés au C637, C636 et C635 plutôt qu'au C636, C635 et C634 tel qu'observé dans le modèle initial (Figure 1.7 B) [107]. Cette dernière configuration a été appelée *shifted*, étant donné le déplacement du brin 3' de l'hélice, créant une *shifted helix*. De plus, dans ces séquences, le nucléotide en position 626 est soit apparié au 634 ou encore non apparié, comparativement au modèle initial où le 626 est plutôt apparié au 633 [107]. L'ensemble de ces expériences a également montré que la conformation du deuxième groupe, contenant la *shifted helix*, est active et que la conformation du premier groupe, similaire à celle initialement prédite, est aussi active mais seulement parce qu'elle peut subir un changement de conformation lui permettant d'adopter la *shifted helix*. Par conséquent, cette conformation a été appelée *shiftable*. Le changement de conformation a été démontré par des expériences de modifications chimiques dans lesquelles différents agents chimiques ont été utilisés pour sonder l'accessibilité au solvant des nucléotides composant le substrat. En effet, le patron de protection chimique de la tige Ib de la conformation du

premier groupe change entre les conditions avec magnésium seulement et avec magnésium en plus du ribozyme, alors que celui de la conformation contenant la *shifted helix*, demeure inchangé [107]. En plus, il a été démontré que des substrats modifiés bloqués dans une conformation inactive (*unshifted*) ne permettent pas l'activité du ribozyme, alors que des substrats modifiés bloqués dans une conformation active (*shifted*) ou encore pouvant passer de la conformation inactive à active (*shiftable*) sont compatibles avec son activité [107]. La suite de la section décrira chacune des deux conformations (*shiftable* et *shifted*) du substrat, ainsi que le changement de conformation permettant l'adoption de la conformation *shifted* contenant la *shifted helix*.

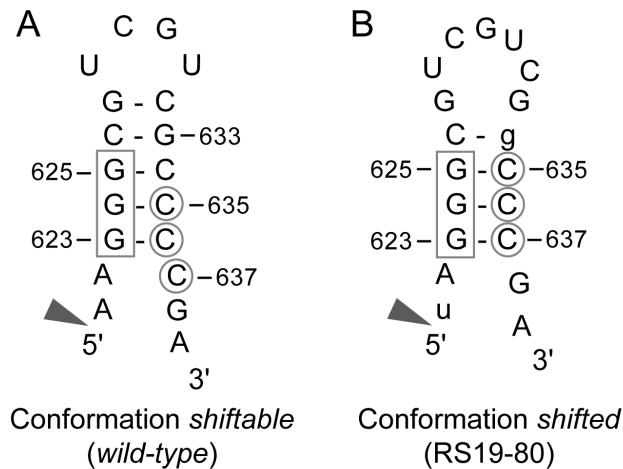


Figure 1.7 : Séquences et structures secondaires des deux groupes représentant le substrat du ribozyme VS. (A) Séquence représentant la conformation naturelle *shiftable* du substrat, telle qu'observée dans le modèle initial. (B) Séquence représentant la conformation *shifted* du substrat. Les guanines encadrées dans les conformations *shiftable* et *shifted* forment des paires de bases avec des cytosines différentes (cytosines encerclées) (Adapté de [107]).

Avant d'adopter la *shifted helix*, le substrat naturel *shiftable* est caractérisé par une conformation *unshifted* inactive, qui a été caractérisée par des expériences de mutagenèse et de modifications chimiques. Celle-ci est composée d'une boucle interne symétrique de six nucléotides comprenant les G620, A621, A622, C637, G638 et A639 [90] (Figure 1.8). De son côté, la tige Ib est caractérisée par cinq paires de bases W-C G-C impliquant les G623 à G627 avec les C636 à C632 respectivement [90]. Puis, la

boucle terminale est formée par quatre nucléotides : U628, C629, G630 et U631 [90] (Figure 1.8).

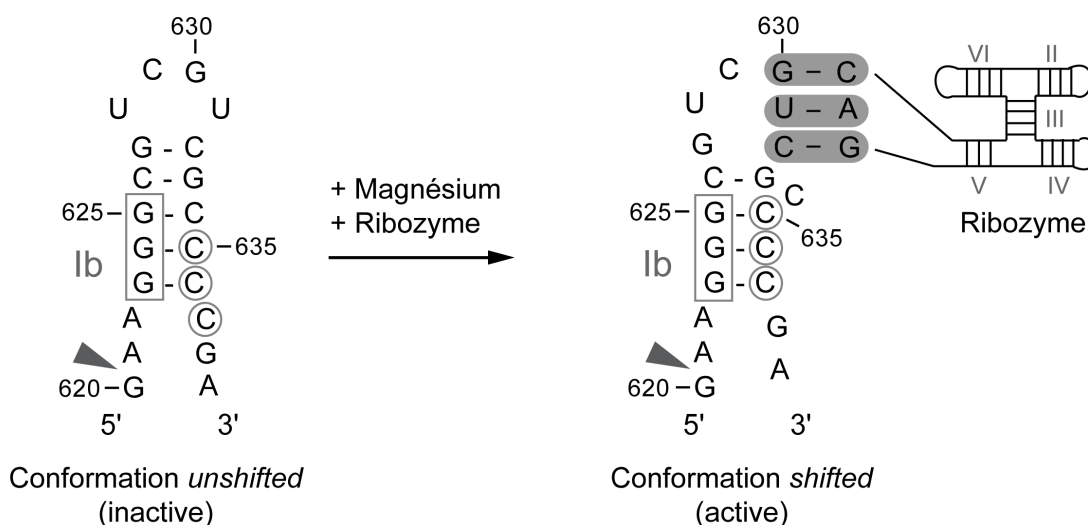


Figure 1.8 : Réarrangement de la structure du substrat. Le réarrangement du substrat d'une conformation *unshifted* à *shifted* est caractérisé par un *helix shift* dans lequel le brin 3' est déplacé en direction 5'. La conformation *shifted* résultante comprend une *shifted helix* dans laquelle les G623, G624 et G625 sont appariés aux C637, C636 et C635 respectivement. La cytosine 634 se retrouve non appariée et la boucle interne contenant le site de clivage (tête de flèche) passe d'une conformation initiale de six nucléotides à une conformation à cinq nucléotides. (Adapté de [107]).

Lors de la liaison du domaine V en présence d'ions magnésium, la conformation du substrat passe à une conformation *shifted* active qui a également été caractérisée par des expériences de mutagénèse et de modifications chimiques [107, 109]. Ce réarrangement comprend deux changements majeurs : un glissement des paires de bases de la tige Ib appelé *helix shift* et un réarrangement de la boucle interne [107, 109] (Figure 1.8). Premièrement, lors du *helix shift* les guanines 623, 624 et 625, initialement appariées aux cytosines 636, 635 et 634, subissent un changement de partenaire pour se retrouver appariées aux cytosines 637, 636 et 635, formant la *shifted helix* (Figure 1.8). Ceci crée un effet de glissement du brin 3' de la tige Ib en direction 5' qui s'accompagne de l'expulsion de la cytosine 634 de la tige [107, 109]. Deuxièmement, lors de la réorganisation de la boucle interne, le C637 quitte celle-ci en glissant en direction 5'

pour former une paire de base avec le G623 dans la tige Ib [107, 109]. Conséquemment, la boucle interne passe d'une conformation initialement symétrique de six nucléotides à une conformation asymétrique de cinq nucléotides (Figure 1.8). Ces changements ont été démontrés notamment par les patrons de protection chimique des nucléotides C634 et C637. Dans le contexte de la conformation *unshifted*, le C634 est protégé des modifications chimiques, tandis que le C637 est accessible. Ceci suggère que le C634, contrairement au C637, fait partie d'un empilement ou encore participe à une paire de bases [90, 107]. Dans le contexte de la conformation *shifted*, des patrons de protection chimique inverses sont observés pour ces deux cytosines, suggérant que le C637, contrairement au C634, fait partie d'un empilement ou encore participe à une paire de bases [90, 107]. Enfin, ce réarrangement comprend aussi un changement au niveau de la dernière paire de base de la tige Ib, entre le G627 et le C632. Lorsque le substrat est dans la conformation *shifted*, le C632 change de partenaire, délaissant le G627 de la tige Ib, pour s'apparier au G697 de la boucle V. Celui-ci participe donc directement à la formation de l'interaction *kissing-loop* I/V [82].

Le réarrangement du substrat est essentiel pour l'activité du ribozyme et la conformation *shifted* facilite sa liaison au ribozyme. Ceci a aussi été démontré par des études de liaison permettant de mesurer l'affinité de liaison avec des substrats adoptant différentes conformations [115].

1.2.1.1.3 Structures des domaines I et V seuls

La formation de l'interaction *kissing-loop* I/V, ainsi que le réarrangement du substrat sont deux événements dépendant du magnésium qui peuvent être reproduits avec les domaines I et V isolés, sans le reste du ribozyme [109]. De ce fait, dans le but de mieux comprendre la reconnaissance du substrat et l'interaction *kissing-loop* I/V, les domaines I et V ont été isolés du reste du ribozyme et étudiés indépendamment par spectroscopie RMN. Jusqu'à maintenant, ces études ont permis de caractériser la conformation du domaine V et les formes inactive et active du substrat.

Domaine V

Deux études RMN ont permis de caractériser le domaine V seul, aussi appelé *SLV* (*stem-loop V*, *SLV*) en absence (*SLV^{free}*, [113]) et en présence de magnésium (*SLV^{Mg}*, [114]). Ces études ont démontré que la boucle terminale V adopte un motif de type *U-turn*, dont la séquence consensus UNR implique les nucléotides U696, G697 et A698. De plus, la présence de magnésium entraîne deux effets principaux sur la conformation de la boucle V : un motif *U-turn* plus compact et un important changement de la position du nucléotide U700 [113, 114].

Tout d'abord, la boucle V en absence de magnésium adopte un motif de type *U-turn* comprenant plusieurs des caractéristiques principales d'un *U-turn* canonique : la séquence consensus UNR (U696-G697-A698), le changement de direction abrupte dans le squelette ribose-phosphate entre les nucléotides U et N (entre U696 et G697), l'empilement des nucléotides qui se trouvent en 3' du U (G697-A698-C699) et la présence d'un pont hydrogène entre le groupement 2'-OH du U et le N7 du nucléotide R (U696 2'-OH-A698 N7) [113] (Figure 1.9). Cependant, deux des caractéristiques importantes définissant un *U-turn* canonique sont manquantes : l'empilement de la base du U sur le groupement 5'-phosphate du nucléotide R (U696 et A698) et un pont hydrogène entre le proton imino du U et le groupement 3'-phosphate du nucléotide R (U696 H3-A698 3'-phosphate) [113]. Avec l'ajout d'ions magnésium, la boucle V adopte une structure globalement plus compacte avec un motif de type *U-turn* composé de toutes les caractéristiques principales du *U-turn* canonique [114]. En effet, cet arrangement plus compact est associé à un changement de direction du squelette ribose-phosphate plus drastique favorisant l'empilement de la base U696 sur le groupement 5'-phosphate du A698 et la formation du pont hydrogène manquant entre U696 H3-A698 3'-phosphate [114] (Figure 1.9). Globalement, ce motif *U-turn* permet aux nucléotides G697, A698 et C699 d'exposer leur face W-C dans le sillon mineur, créant une configuration idéale pour interagir avec la boucle terminale du substrat [114]. De plus, il a été démontré que les trois bases formant l'interaction *kissing-loop* adoptent une structure similaire à celle d'un brin d'hélice de forme A. De cette façon, elles se retrouvent dans une conformation compatible avec la formation de paires de bases [114].

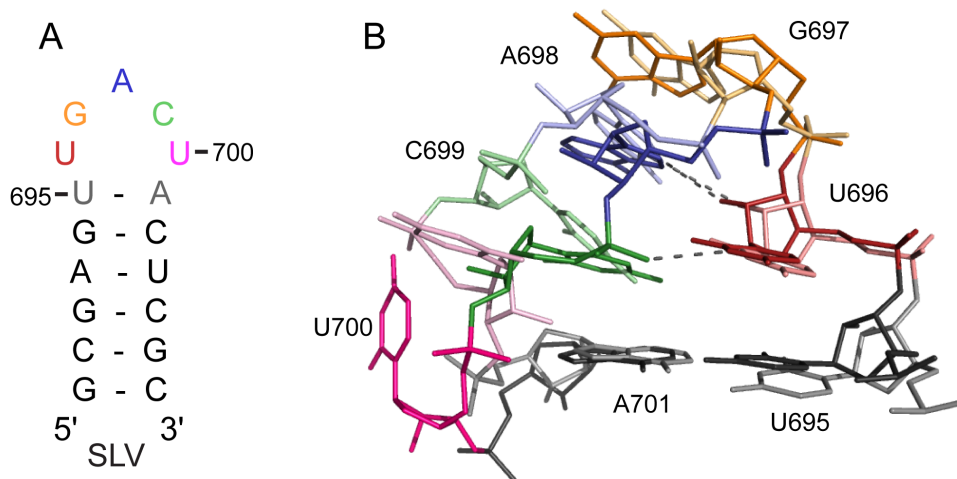


Figure 1.9 : Effet du magnésium sur la structure de la boucle terminale du domaine V (A) Séquence et structure secondaire de la tige-boucle V utilisées pour les deux études RMN [113, 114]. (B) Superposition des structures tridimensionnelles de la boucle V en absence (couleurs pâles) et en présence de magnésium (couleurs foncées). Les lignes pointillées représentent les ponts hydrogènes retrouvés dans la boucle V (U696 H3-A698 3'-phosphate et U696 2'-OH-A698 N7). Le code de couleur utilisé en (B) correspond à celui utilisé en (A).

La présence d'ions magnésium est associée à un autre changement important dans la structure de la boucle V : la position du nucléotide U700. En absence de magnésium, la base du nucléotide U700 est exclue de la boucle et exposée dans le sillon majeur où elle est empilée sur le groupement 5'-phosphate du C699 [113]. Le ribose du U700, quant à lui, se retrouve entre ceux des nucléotides C699 et A701 [114] (Figure 1.9). Avec l'ajout d'ions magnésium, la base et le ribose du U700 subissent un changement considérable de position pour se retrouver tout deux complètement exclus de la boucle du côté du sillon mineur [114] (Figure 1.9). Ce mouvement amène le U700 près des nucléotides G697, A698 et C699 impliqués dans l'interaction *kissing-loop* (Figure 1.9), suggérant qu'il pourrait lui aussi contribuer à la reconnaissance du substrat [114].

En résumé, ces deux études RMN démontrent clairement que la conformation libre de la boucle V prédispose plusieurs nucléotides dans le sillon mineur (G697, A698, C699 et U700), facilitant leur accessibilité pour l'interaction avec la boucle I du substrat. Toutefois, des études structurales supplémentaires sont nécessaires pour connaître les déterminants structuraux qui définissent précisément l'interface de l'interaction et la reconnaissance du substrat.

Domaine I

Le substrat naturel du ribozyme adopte soit une conformation *unshifted* inactive ou *shifted* active et toutes deux ont été étudiées par spectroscopie RMN. Tout d'abord, deux études de RMN ont été réalisées sur la conformation inactive du substrat. La première étude a été effectuée sur une tige-boucle dans laquelle la tige Ib et la boucle terminale ont été remplacées par une tige plus courte de trois paires de bases terminée par une boucle terminale stabilisatrice de quatre nucléotides [110], puis la deuxième étude a été réalisée sur le substrat naturel complet [111]. Quelques années plus tard, une troisième étude par RMN a été publiée sur la conformation active du substrat. Cette étude a été effectuée sur une tige-boucle dans laquelle seulement la boucle interne dans sa conformation active ainsi que les paires de bases adjacentes du ribozyme ont été préservées (G618 à G623 et C637 à C641) [112]. En effet, à des fins de simplification, la tige Ib a été raccourcie, puis la boucle terminale a été remplacée par une boucle terminale composée de quatre nucléotides [112].

Les deux études RMN de la conformation inactive du substrat ont démontré que la boucle interne est formée de trois paires de bases non canoniques: G620-A639, A621-G638 et A+622-C637 [110, 111] (Figure 1.10). De plus, la deuxième étude, effectuée sur le substrat complet, a montré que la tige Ib est caractérisée par quatre paires de bases W-C G-C: G623-C636, G624-C635, G625-C634 et C626-G633 [90, 111]. Bien qu'une cinquième paire de bases (G627-C632) ait initialement été proposée par des études biochimiques [90], celle-ci n'a pas été confirmée par les études RMN du substrat complet [111]. Cette étude par RMN a aussi permis de constater que la boucle terminale

du substrat est composée de six nucléotides, dont l'organisation est dynamique et désordonnée [111].

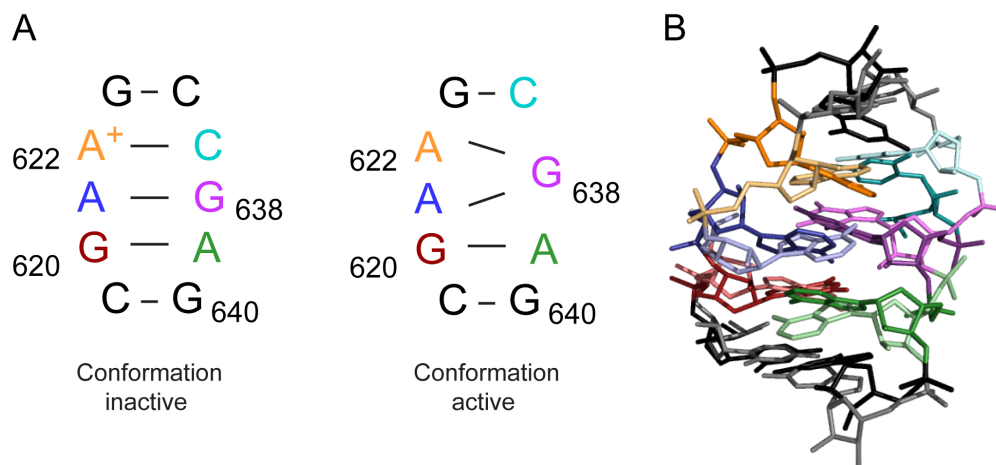


Figure 1.10 : Comparaison des conformations inactive et active de la boucle interne du substrat. (A) Séquences et structures secondaires de la boucle interne du substrat démontrant les différentes paires de bases (représentées par des traits noirs) de chaque conformation. (B) Superposition des structures de la boucle interne du substrat en conformation inactive (couleurs pâles) [110] et active (couleurs foncées) [112]. Le code de couleur utilisé en (A) correspond à celui utilisé en (B). (Adapté de [112]).

L'étude RMN de la conformation active du substrat a montré que la boucle interne du substrat forme un réseau de paires de bases différent de celui observé dans sa conformation inactive. En effet, les cinq nucléotides de la boucle interne sont organisés en trois paires de bases G-A où le G638 participe à deux de ces paires, avec le A621 et le A622. La troisième paire de base G-A se forme entre le G620 et le A639 et est similaire à celle observée dans la boucle interne inactive (Figure 1.10). De plus, ces paires de base G-A sont impliquées dans un motif de liaison au magnésium de type 5'-A R-3' / 5'-Y G-3' où R=adénine ou guanine et Y=cytosine ou uridine. En effet, deux sites de liaisons ont été identifiés à chaque extrémité de la boucle interne: 5'-A639 G640-3' / 5'-C619 G620-3' et 5'-A622 G632-3' / 5'-C637 G638-3'. Le premier site est commun aux conformations inactive et active du substrat, tandis que le dernier site est exclusif à la

conformation active [112]. Cette étude par RMN a aussi démontré que la boucle interne active du substrat, composée de ses trois paires de bases G-A flanquées de chaque côté par une paire de base G-C, constituent un motif d'interaction tertiaire [112]. Ce motif a d'ailleurs été identifié dans d'autres ARN, dont l'ARNr 16S où il participe à une interaction tertiaire de type *ribose zipper*. Ce type de motif permet à deux brins d'ARN d'interagir grâce à l'établissement d'un réseau de ponts hydrogènes entre les groupements 2'-OH des riboses de chacun des brins. De ce fait, il a été proposé que ce même motif dans la boucle active du substrat pourrait aussi participer à une interaction tertiaire avec le domaine catalytique pour la formation du site actif du ribozyme [112].

Dans l'ensemble, les structures à haute résolution des domaines I et V ont permis de caractériser plusieurs éléments de structure importants de leur forme libre. En effet, la boucle terminale V seule est structurée et prédispose ses nucléotides pour former l'interaction *kissing-loop* avec le substrat. Toutefois, le substrat présente une boucle terminale qui est plutôt désordonnée, mais à l'opposé, une boucle interne structurée. Les caractéristiques uniques de la configuration active de cette boucle interne comprennent un site supplémentaire de liaison au magnésium et l'adoption d'un motif de structure tertiaire probablement impliqué dans son positionnement au sein du ribozyme.

Enfin, bien que les études biochimiques précédentes semblent indiquer que l'interaction entre les domaines I et V soit définie par la formation de trois paires de bases W-C, d'autres déterminants structuraux sont possiblement impliqués dans cette interaction pour permettre au ribozyme de reconnaître spécifiquement son substrat. Effectivement, les études biochimiques précédentes ont montré que les nucléotides C629 et A701 jouent probablement un rôle dans l'interaction *kissing-loop* [82]. Une structure à haute résolution de l'interaction *kissing-loop* I/V est définitivement nécessaire pour mieux comprendre les éléments qui déterminent la reconnaissance du substrat et son activation.

1.2.1.2 Le domaine catalytique

Le domaine catalytique (domaines hélicoïdaux II-VI) est organisé par deux jonctions à trois voies : la jonction II-III-VI, qui joint les domaines hélicoïdaux II, III et VI formant le sous-domaine II-III-VI et la jonction III-IV-V qui connecte les domaines hélicoïdaux III, IV et V formant le sous-domaine III-IV-V (Figure 1.11 A). Ces jonctions permettent d'orienter les domaines hélicoïdaux du domaine catalytique et de créer une niche pour accueillir le substrat. De plus, chacun des sous-domaines du domaine catalytique démontre un rôle distinct, soit dans l'optimisation de liaison du substrat (sous-domaine III-IV-V) ou dans la création du site actif pour permettre la réaction catalytique (sous-domaine II-III-VI).

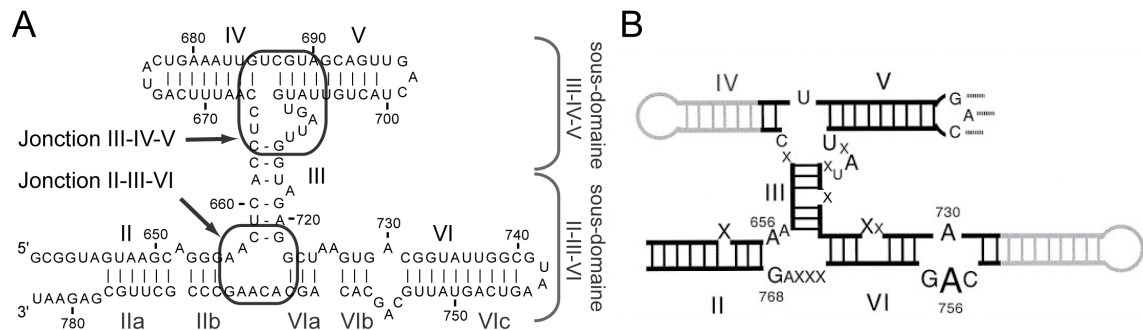


Figure 1.11 : Représentation du domaine catalytique du ribozyme VS et de ses éléments structuraux importants. (A) Séquence et structure secondaire du domaine catalytique. Les deux jonctions sont encadrées. (B) Représentation schématique du domaine catalytique. Les éléments en noir sont essentiels, alors que ceux en gris ne le sont pas. Les «X» indiquent que la position du nucléotide est importante, mais pas l'identité de la base. Les deux jonctions, ainsi que la boucle A730 contiennent des nucléotides importants représentés par un symbole dont la taille est proportionnelle à leur importance. [93, 116]

De substantielles études de mutagenèse ont montré que les domaines hélicoïdaux du domaine catalytique du ribozyme ne présentent pas tous la même importance fonctionnelle. Ceci a été mis en évidence par la variation du niveau de tolérance du ribozyme aux modifications de séquence. En effet, la quasi-totalité de la tige IV et la partie distale des domaines II et VI (IIa et VIc, Figure 1.11) peuvent être tronquées sans affecter significativement l'activité du ribozyme [88, 89, 93, 104] (Figure 1.11 B). Par

contre, bien que la grande majorité de la séquence des tiges III et V puisse être modifiée, le nombre de paires de bases qu'elles contiennent est crucial pour l'activité du ribozyme [89, 117, 118] (Figure 1.11 B).

Bien que certains éléments du domaine catalytique puissent être modifiés ou éliminés, d'autres sont plutôt cruciaux pour son repliement et pour l'activité catalytique du ribozyme. C'est le cas des adénines non appariées (*bulge*) interrompant les domaines II, III et VI, dont l'identité peut être modifiée (A652, A718, A725, A726, Figure 1.11). Par contre, leur présence ainsi que leur position sur la tige sont importantes [90, 93, 95, 117, 119]. D'un autre côté, la modification et la délétion des nucléotides non appariés situés dans les jonctions II-III-VI et III-IV-V diminuent grandement l'activité du ribozyme et interfèrent avec le repliement des jonctions [117, 120, 121] (Figure 1.11). Deux autres éléments du domaine catalytique semblent être cruciaux pour le fonctionnement optimal du ribozyme : la boucle interne A730 du domaine VI et la boucle terminale du domaine V, toutes deux impliquées dans des interactions avec le substrat [82, 89, 93, 97, 107] (Figure 1.11). La mutation ou la délétion individuelle de la majorité des nucléotides de ces éléments entraîne une diminution drastique d'activité [82, 93, 94, 122].

Davantage de modifications ont été effectuées dans le domaine catalytique dans le but de mieux comprendre l'importance et l'implication de chacun de ses sous-domaines II-III-VI et III-IV-V. D'une part, au niveau du sous-domaine II-III-VI, la jonction II-III-VI peut être remplacée par une autre jonction retrouvée dans l'ARNr 23S et le ribozyme demeure actif, mais montre une baisse d'activité de clivage de dix fois [120]. Lorsque la boucle interne A730 est enlevée et remplacée par une simple paire de bases W-C A-U, l'activité du ribozyme est perdue [95]. D'un autre côté, certaines séquences de boucles internes non naturelles peuvent remplacer la boucle A730 et sont compatibles avec l'activité de clivage du ribozyme. Cette activité est toutefois beaucoup plus lente qu'en présence de la séquence naturelle [95]. D'autre part, au niveau du sous-domaine III-IV-V, il a été démontré que la jonction à trois voies III-IV-V peut être remplacée par une jonction à quatre voies [121]. En effet, la région située entre les

domaines III et V, dans laquelle un motif de type *U-turn* a été prédit basé sur la séquence (U710-G711-A712), peut être entièrement remplacée par une structure en tige-boucle, créant ainsi une jonction à quatre voies [121]. Aussi, il a été démontré qu'en absence du sous-domaine III-IV-V en entier, le ribozyme demeure fonctionnel, bien qu'il soit beaucoup moins efficace que le ribozyme entier [95]. Ces différents résultats indiquent que les éléments essentiels à la formation du site actif du ribozyme sont situés dans le sous-domaine II-III-VI et dans le substrat [95]. De plus, malgré le fait que le sous-domaine III-IV-V ne soit pas essentiel pour la réaction catalytique, il contribue significativement au fonctionnement optimal du ribozyme. En effet, le sous-domaine III-IV-V favorise la formation de l'interaction *kissing-loop* entre les domaines I et V, ce qui permet l'activation du substrat et en facilite le positionnement dans le sous-domaine II-III-VI [89, 95, 109, 119, 121].

En terminant, les études sur le domaine catalytique du ribozyme VS ont permis de mettre en évidence le rôle distinct de chacun des deux sous-domaines : le rôle catalytique du sous-domaine II-III-VI, particulièrement de sa boucle A730 et le rôle structural de liaison et d'activation du substrat pour le sous-domaine III-IV-V. De plus, bien qu'on puisse lui enlever des domaines ou encore en modifier la séquence, le domaine catalytique naturel du ribozyme VS contribue considérablement à l'efficacité de son fonctionnement. Ses deux jonctions naturelles guident l'orientation de chacun de ses domaines hélicoïdaux et facilitent la formation des interactions tertiaires nécessaires au positionnement du substrat et à la création de son site actif.

1.2.2 Le site actif et mécanisme catalytique

La formation de l'interaction *kissing-loop* I/V, en plus d'activer le substrat pour la catalyse, facilite l'association (*docking*) du site de clivage avec la boucle interne du domaine VI [92, 96, 109]. En fait, d'après les données actuelles, il est entendu que le site actif du ribozyme est formé par une interaction de type boucle-boucle entre les boucles internes des domaines I et VI [91-105] (Figure 1.12).

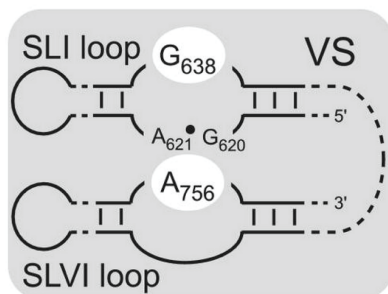


Figure 1.12 : Représentation schématique de la formation du site actif du ribozyme VS. Les nucléotides encadrés en blanc ont été proposés pour participer à la catalyse du ribozyme. Le point noir indique le site de clivage et les lignes pointillées représentent les parties manquantes de la représentation du ribozyme. (Tiré de [103]).

Le résultat le plus crucial ayant permis de mettre en évidence l'association directe entre ces deux boucles est le *cross-link* obtenu entre le nucléotide 621, en aval du site de clivage, et le nucléotide A756 de la boucle A730 [96]. En effet, un nucléotide analogue photo-actif de type 4-thio-uridine, une uridine dans laquelle l'atome d'oxygène en position 4 a été substitué par un atome de soufre, a été introduit en position 621 [96]. Suite à une exposition à des rayons UV, la 4-thio-uridine présente dans le ribozyme forme un lien covalent, un *cross-link*, avec un groupement situé à proximité dans la structure tertiaire, le A756. Bien que de nombreuses études de mutagenèse et de modifications chimiques aient proposé que la boucle A730 soit impliquée dans le site actif du ribozyme, l'obtention de ce *cross-link* est la première preuve directe d'une réelle proximité entre ces deux boucles internes [96]. Les prochaines sections détailleront la réaction catalysée au sein du site actif du ribozyme VS ainsi que les éléments structuraux importants de ce site actif, particulièrement ceux de la boucle A730.

1.2.2.1 Mécanisme catalytique

La réaction catalysée par le ribozyme VS est un auto-clivage réversible qui s'effectue par un mécanisme de transestérification entre les nucléotides G620 et A621 (Figure 1.13). Le ribozyme VS, ainsi que les autres membres de la classe des ribozymes nucléolytiques auto-clivants effectuent le clivage d'un lien phosphodiester par l'attaque

d'un groupement 2'-OH sur le groupement phosphate adjacent en 3' par un mécanisme de substitution nucléophile bimoléculaire (S_N2) [91, 99, 102, 123] (Figure 1.13). Pour ce faire, les nucléotides impliqués dans la réaction doivent tout d'abord être alignés avec le lien à cliver pour favoriser l'attaque nucléophile. Le groupement 2'-OH du ribose impliqué doit être déprotoné pour permettre son activation et agir en tant que groupement nucléophile (Figure 1.13). Puis, l'attaque de ce nucléophile sur le groupement phosphate en 3' est suivie d'une inversion de configuration de celui-ci et de la création d'un intermédiaire réactionnel de type trigonal bipyramidal (Figure 1.13). Enfin, la réaction se complète par le transfert d'un proton au groupe partant, le 5'-O, entraînant la rupture du lien 5'-O-P et la formation de deux produits de clivage dont un présentant une extrémité 2'-3' cyclique phosphate et l'autre une extrémité 5'-OH (Figure 1.13). Il a été proposé que des nucléotides du ribozyme participent activement à cette réaction comme accepteur du proton 2'-OH et donneur de proton pour former le 5'-OH dans une catalyse de type acide-base générale [99].

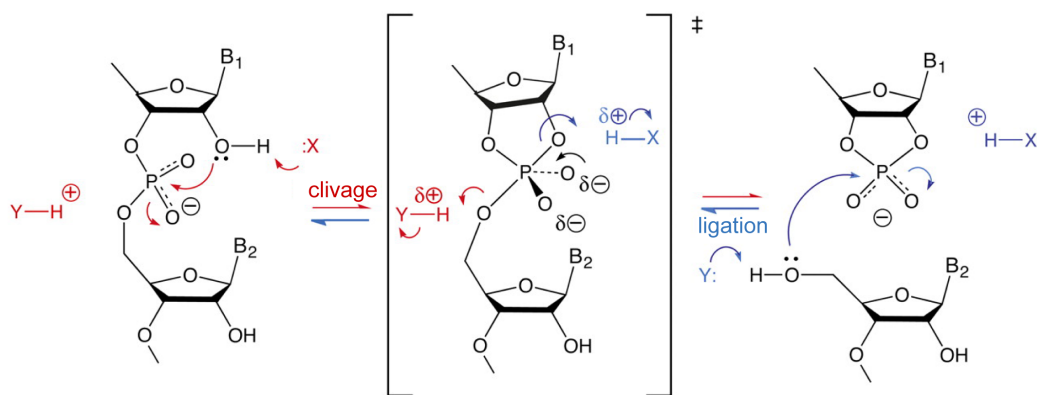


Figure 1.13 : Mécanisme catalytique de type acide-base générale. Les éléments de la réaction de clivage sont représentés en rouge alors que ceux de la réaction de ligation sont représentés en bleu. La réaction de clivage est catalysée par une base générale (X) et un acide général ($Y-H^+$). Dans la réaction inverse de ligation, le $X-H^+$ et le Y joueront les rôles d'acide et de base général respectivement. L'intermédiaire réactionnel trigonal bipyramidal est encadré de crochets. (Adapté de [124]).

Suite à une vaste exploration biochimique du ribozyme VS, notamment par des études de mutagenèse et l'utilisation d'interférence par des nucléotides analogues (*nucleotide analogue interference mapping, NAIM*), il a été possible de déterminer les nucléotides impliqués dans la catalyse par le ribozyme VS, ainsi que leurs groupements fonctionnels. Deux nucléotides ont ainsi été identifiés, soient le A756 dans la boucle A730 [93, 94, 96-98, 125] et le G638 dans la boucle interne contenant le site de clivage [99]. Leur mutation individuelle cause une diminution d'activité de 1000 à 10 000 fois sans toutefois affecter ni le repliement du ribozyme ni la liaison du substrat. Les études de mutagenèse et de *NAIM* ont aussi permis de déterminer que ce sont les faces W-C de ces nucléotides qui sont importantes pour la catalyse, le N1 pour le A756 et les groupements carbonyles en C6 et imino en N1 pour le G638. L'implication de chacun de ces nucléotides a également été démontrée par des expériences de complémentation effectuées avec un mélange de deux types de ribozymes dont l'un contenant une mutation du A756 (A756C) et l'autre une mutation de G638 (G638A) [126]. L'auto-clivage de chaque ribozyme mutant individuel est fortement diminué, alors que le mélange des deux types de ribozyme démontre un bon taux de clivage, indiquant que le substrat intact de l'un peut interagir avec le domaine catalytique intact de l'autre recréant un site actif fonctionnel en *trans* [126]. L'ensemble de ces résultats met en évidence l'implication des nucléotides A756 et G638 dans la catalyse du clivage par le ribozyme VS.

Le rôle des bases A756 et G638 dans la catalyse acide-base générale du ribozyme VS a été précisé par des études substantielles de *NAIM* et des mesures du taux de clivage du ribozyme en fonction du pH, appelées profils de dépendance au pH. Tout d'abord, pour que les bases A756 et G638 puissent participer à la catalyse comme accepteur et donneur de protons, les pK_a de leurs groupements fonctionnels doivent se retrouver sous leur forme protonée et déprotonée respectivement. Puisque les pK_a des N1 des guanosines et adénosines sont de 9.4 et 3.5 respectivement [127], les formes actives protonées des bases A et déprotonées des bases G sont peu représentées à pH neutre, les rendant peu susceptibles de participer à la catalyse. Les études de *NAIM* et les profils de dépendance au pH effectués sur le ribozyme ont démontré un profil caractérisé par une courbe en forme de cloche (*bell shaped pH profile*) correspondant à deux valeurs de pK_a

apparent de 5.2 et 8.4 pour un ribozyme en *trans* [99] et de 5.8 et 8.3 pour un ribozyme en *cis* [100]. Ces valeurs ont été associées aux groupements donneur et accepteur de proton des bases G638 et A756 [97, 99, 125]. Ces résultats sont en accord avec une catalyse de type acide-base générale, toutefois ils ne permettent pas de discriminer quels nucléotides jouent les rôles de base et d'acide. Ceci a toutefois été résolu en substituant le lien phosphate 5'-P-O à cliver par un lien phosphorothioate 5'-P-S [102]. L'insertion d'un atome de soufre crée un meilleur groupe partant qui ne nécessite pas de protonation pour compléter la réaction de clivage. Cette substitution a permis de rétablir complètement l'activité de clivage d'un ribozyme portant la mutation A756G dont la catalyse est normalement fortement diminuée, suggérant un rôle d'acide général pour le A756. De plus, le profil de dépendance au pH d'un ribozyme contenant la substitution 5'-P-S est en accord avec le G638 agissant comme base générale dans le mécanisme acide-base générale [102]. Ainsi, tous ces résultats indiquent que le ribozyme VS catalyse le clivage de son substrat par une catalyse de type acide-base générale dans laquelle le G638 agit comme base en déprotonant le groupement 2'-OH du ribose G620, puis le A756 agit comme acide en cédant un proton au groupe partant 5'-OH, complétant la réaction de clivage (Figure 1.14) [102].

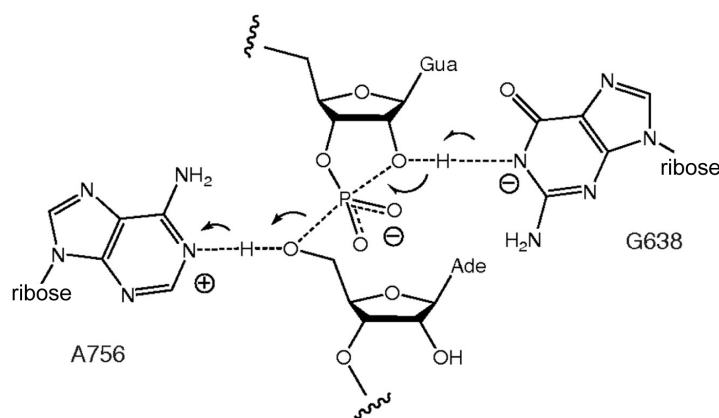


Figure 1.14 : Mécanisme proposé utilisé par le ribozyme VS pour catalyser la réaction de clivage. Le G638 et le A756 jouent les rôles de base et d'acide général respectivement. Les lignes pointillées représentent les liaisons qui seront formées ou défaites dans le clivage. (Adapté de [102]).

1.2.2.2 Éléments de structure de la boucle A730

Une fois la boucle A730 identifiée comme composante du site actif, des études supplémentaires ont été effectuées dans le but de comprendre les éléments structuraux permettant la formation du site actif ainsi que la réaction catalytique qui s'y déroule. En effet, plusieurs études par spectroscopie RMN effectuées sur le domaine VI isolé du reste du ribozyme, aussi appelé *SLVI* (*stem-loop VI*, *SLVI*), ont permis d'obtenir des données structurales à haute résolution de la boucle interne A730 [103, 105, 128]. La première étude montre une boucle interne symétrique composée de trois paires de bases non canoniques, A+730-C755, G729-A756 et U728-G757 [128]. Les deux autres études révèlent plutôt une boucle interne asymétrique caractérisée par un motif de type *S-turn* entre les nucléotides G754 et C758 et fermée par une paire de base G-A entre les A730 et G757 [103, 105] (Figure 1.15 A, B). La première étude RMN contraste avec la structure secondaire prédite et les données biochimiques [128]. À l'opposé, les deux autres études RMN, qui révèlent une structure complètement différente de la première, sont en accord avec la structure secondaire prédite et corrélient avec les données biochimiques actuelles [103, 105]. Le motif *S-turn* décrit dans ces deux études provoque un renversement local du squelette ribose-phosphate causé par la conformation inversée d'un ribose, celui du A756, introduisant une forme en « S » dans le squelette ribose-phosphate. De plus, ce motif, ainsi que la paire de base G757-A730, permettent l'exclusion des nucléotides C755 et A756 de la structure. Ces derniers sont donc positionnés de façon à exposer leur face W-C dans le sillon mineur élargi (Figure 1.15 B), ce qui est compatible avec le rôle central du A756 en tant qu'acide général dans le mécanisme catalytique [103, 105].

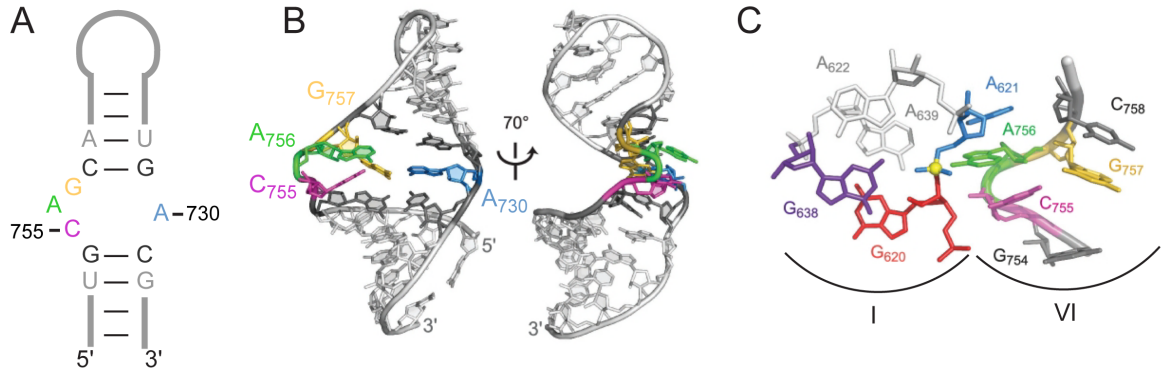


Figure 1.15 : Éléments structuraux de la boucle A730 du domaine VI du ribozyme VS. (A) Séquence et structure secondaire de la boucle A730 utilisées pour l'étude RMN [103]. (B) Structure RMN de la boucle A730 contenant le motif *S-turn* entre les nucléotides G754 et C758. (C) Modèle du site actif formé de la boucle A730 (domaine VI) et de la boucle interne du substrat (domaine I). La sphère jaune indique le site de clivage. Le code couleur utilisé en (A) est le même utilisé en (B) et (C). (Adapté de [103].)

Enfin, les données structurales des domaines I et VI isolés, en combinaison avec les données biochimiques, ont permis de proposer un mode d'association entre les deux boucles internes pour mieux comprendre la formation du site actif du ribozyme (Figure 1.15 C). Ce modèle montre une interaction entre les sillons mineurs des boucles internes I et VI dans laquelle le A621 subit un changement de conformation l'amenant à sortir de l'intérieur de la boucle de façon à adopter une structure dites *splayed* où les bases des nucléotides G620 et A621 pointent dans une direction opposée (Figure 1.15 C). Cette conformation stabiliserait une géométrie alignée (*in line*) entre le 2'-OH du G620 le lien phosphate à cliver et le groupement 5' du G620 en plus de favoriser le positionnement du A756 et G638. De plus, il a été suggéré qu'un motif de type *ribose zipper* se forme entre la boucle interne I et les domaines II ou VI du ribozyme [112]. Ce motif d'interaction tertiaire permettrait de stabiliser le substrat dans une crevasse formée par les domaines II et VI [92, 95, 96, 112]. En somme, l'association des boucles I et VI permet de rapprocher le site de clivage, situé entre les nucléotides G620 et A621, des nucléotides A756 et G638, impliqués dans la catalyse. D'ailleurs, il est intéressant de noter qu'un arrangement similaire, impliquant l'association de deux boucles internes, a été observé précédemment dans le ribozyme *hairpin* [91, 99, 102, 124].

1.2.2.3 Similarités mécanistiques entre les ribozymes VS et *hairpin*

Les ribozymes VS et *hairpin* font tous deux partie du groupe des ribozymes nucléolytiques auto-clivants, catalysant le clivage réversible d'un lien phosphodiester par une réaction de transesterification, telle que présentée plus haut. Bien que les structures secondaires et tertiaires de ces deux ribozymes soient très différentes (Figure 1.16), les dernières études effectuées sur le ribozyme VS ont révélé des similarités évidentes avec le ribozyme *hairpin*, quant à leur site actif et leur mécanisme catalytique.

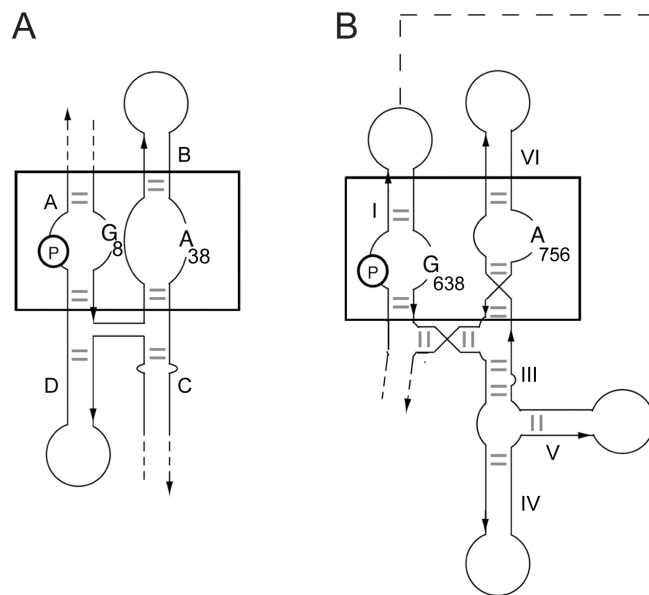


Figure 1.16 : Représentation schématique de l'organisation des sites actifs des ribozymes *hairpin* et VS. Le site actif des ribozymes *hairpin* (A) et VS (B) est encadré, puis le groupement phosphate situé au site de clivage est encerclé. L'orientation des brins est représentée par des flèches, dont la tête pointe en direction 3'. (Adapté de [124]).

La formation des sites actifs des deux ribozymes résulte de l'association de deux boucles internes. En effet, la structure du ribozyme *hairpin* obtenue par cristallographie montre une association des boucles internes A et B (Figure 1.16 A) [129-131]. Parallèlement, l'ensemble des études biochimiques et biophysiques effectuées sur le ribozyme VS supporte l'association des boucles internes I et VI (Figure 1.16 B) [91-105]. De plus, ces deux sites actifs contiennent deux nucléotides nécessaires à la

catalyse, une guanine et une adénine. Pour le ribozyme *hairpin*, ce sont le G8 dans la boucle A et le A38 de la boucle B qui sont importants [132-136] tandis que pour le ribozyme VS, ce sont les G638 de la boucle I et A756 de la boucle VI [93, 94, 96-99] qui sont utilisés. Aussi, l'organisation de ces nucléotides au sein du site actif est similaire : la guanine se retrouve sur le brin opposé au site de clivage, alors que l'adénine se retrouve dans l'autre boucle. Il a été proposé que la similitude observée pour l'organisation de ces deux sites actifs s'étende aussi au mécanisme catalytique des deux ribozymes [124]. En effet, les études présentées dans la section précédente proposent une catalyse de type acide-base générale pour le ribozyme VS, puis de façon similaire plusieurs études suggèrent un tel mécanisme pour le ribozyme *hairpin* [124, 137, 138]. À cet égard, il a été proposé que le G8 agisse en tant que base générale en déprotonant le A+1, pendant que le A38 de son côté, joue le rôle de l'acide général en cédant un proton au groupe partant 5'-O du G-1 [137, 138].

L'ensemble de ces résultats démontre que malgré les différences de structures secondaires et tertiaires globales des ribozymes VS et *hairpin*, l'organisation et la topologie de leur site actif, ainsi que leur mécanisme catalytique s'avèrent similaires. Pour ces ribozymes, la catalyse de type acide-base générale semble donc un moyen efficace pour permettre le clivage du lien phosphodiester. Ce type de catalyse a d'ailleurs été proposé pour d'autres classes de ribozymes auto-clivants, tels que les *hammerhead* [33, 139], *HDV* [140] et *glmS* [141-143]. Toutefois, chez ces ribozymes, les rôles d'acide et de base ne sont pas nécessairement effectués par des bases du ribozyme. Par exemple, pour le *HDV*, il a été proposé qu'un cation magnésium facilite la déprotonation du nucléophile et que la base C75 joue le rôle d'acide général [140, 144, 145], alors que pour le *glmS* c'est une molécule de glucosamine-6-phosphate qui joue ce rôle d'acide [142, 146].

1.2.3 Repliement du ribozyme VS

Le repliement du ribozyme VS, ainsi que son activité sont grandement influencés par les cations, notamment les cations divalents. En effet, le taux de clivage du ribozyme

s'est vu augmenté avec la concentration de cations divalents [106]. Parallèlement, plusieurs études biochimiques et biophysiques ont démontré que la structure du ribozyme change avec l'ajout de cations divalents [82, 90, 92, 107, 112, 114, 117, 120, 147]. Par exemple, différentes régions du ribozyme ont montré une protection contre des modifications par différents agents chimiques seulement en présence de cations divalents [90]. De plus, ces études révèlent que ces effets sont plus marqués avec le magnésium, par rapport au manganèse et au calcium [106].

1.2.3.1 Rôle des cations divalents

Les cations divalents jouent des rôles particulièrement importants dans la structure et la fonction des ARN, notamment au niveau de leur repliement [148-150]. En fait, le squelette ribose-phosphate de l'ARN lui confère un caractère anionique, donc il n'est pas surprenant que des cations divalents s'y lient. Ces cations peuvent ainsi faciliter l'assemblage des ARN en s'y associant, permettant au squelette de l'ARN d'adopter des conformations normalement interdites par la répulsion électrostatique.

Dans le cas du ribozyme VS, le cation divalent qui permet l'adoption d'une structure active et le meilleur taux de clivage est le magnésium (Mg^{2+}) [106]. Ce dernier est entouré d'une couche d'hydratation composée de six molécules d'eau. Lorsque le Mg^{2+} interagit avec l'ARN, il peut le faire de différentes façons : diffuse, indirecte (*outer-sphere*) ou encore directe (*inner-sphere*) [151] (Figure 1.17). Une interaction diffuse est de force faible et s'effectue via les couches d'hydratation du Mg^{2+} et de l'ARN (Figure 1.17 A). Les cations hydratés vont avoir tendance à s'accumuler près de l'ARN de façon à créer un environnement de contre-ions. Ce dernier permet de stabiliser l'ARN réduisant les forces de répulsion des groupements phosphates voisins chargés négativement [151]. Les deux autres types d'interaction Mg^{2+} /ARN sont plus fortes que la première et s'établissent à des endroits spécifiques où des charges négatives sont concentrées [151]. L'interaction indirecte se fait par l'intermédiaire d'une molécule d'eau liée aux Mg^{2+} qui interagit avec un groupement de l'ARN (Figure 1.17 B). Dans le cas de la liaison directe, c'est le Mg^{2+} qui interagit directement avec un groupement de

l'ARN, ce qui implique le déplacement d'une ou plusieurs de ses molécules d'eau [151] (Figure 1.17 C). Ces deux types d'interaction contribuent souvent à la formation des éléments de structure locale, ainsi qu'à l'adoption de la conformation globale de l'ARN [150].

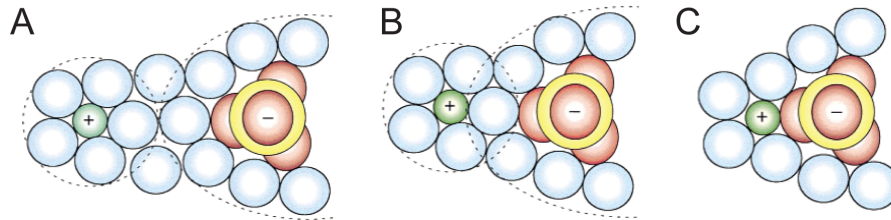


Figure 1.17 : Représentation des différents types d'interactions entre l'ARN et le magnésium. (A) La liaison diffuse s'effectue via les molécules d'eau des couches d'hydratation de l'ARN et du magnésium. (B) La liaison indirecte implique la couche d'hydratation du magnésium qui interagit avec un groupement de l'ARN. (C) La liaison directe s'effectue entre le magnésium et l'ARN sans impliquer de molécule d'eau. (Tiré de [151]).

Une fois l'importance du magnésium identifiée pour le ribozyme VS, différentes expériences biochimiques ont été utilisées pour identifier les groupements phosphates importants pour l'adoption de sa structure native ainsi que pour son fonctionnement. Ces expériences visent notamment à modifier les groupements phosphate de l'ARN pour créer une interférence avec la liaison du magnésium. Un premier type d'expérience nécessite l'utilisation de l'agent alkylant N-nitroso N-éthylurée (*N-ethyl N-nitrourea*, *ENU*) qui transfère son groupement éthyle sur un des oxygènes non pontants (*non-bridging oxygen*) d'un groupement phosphate de l'ARN, le rendant inapte à interagir avec le magnésium. Un deuxième type d'expérience consiste à créer une interférence avec la liaison du magnésium en substituant un des oxygènes non pontants des phosphates de l'ARN par un atome de soufre. Cette substitution crée un groupement phosphorothioate qui interfère avec la formation de pont hydrogène et la coordination du magnésium [152, 153]. En effet, le magnésium présente une affinité 30 000 fois plus élevée pour l'oxygène que pour le soufre [154, 155]. Ce dernier type d'interférence peut être restauré par l'ajout de cations manganèse. Le manganèse est aussi un cation divalent (Mn^{2+}), mais qui a la particularité de présenter une affinité similaire pour l'oxygène et le

soufre [154, 155]. Généralement, il est possible de remplacer une interaction magnésium-phosphate par une interaction manganèse-phosphorothioate. De ce fait, il est possible de récupérer l'activité perdue ou les défauts de repliement d'un ARN causés par l'interférence des modifications phosphorothioates par l'ajout de manganèse. Cette utilisation de manganèse est en fait une extension de l'expérience d'interférence aux phosphorothioates appelée restauration par le manganèse (*manganese rescue experiment*). De plus, cette dernière approche a l'avantage de permettre l'identification de liaison directe entre le cation et l'ARN [104, 154, 155]. L'utilisation de ces expériences d'interférences et de restauration a permis l'identification de plusieurs sites de liaisons pour le magnésium dans le ribozyme. Ces sites se retrouvent dans trois régions principales: la boucle interne du substrat (le site de clivage), la boucle A730 (le site actif) et l'interaction *kissing-loop* I/V [104, 156, 157] (Figure 1.18).

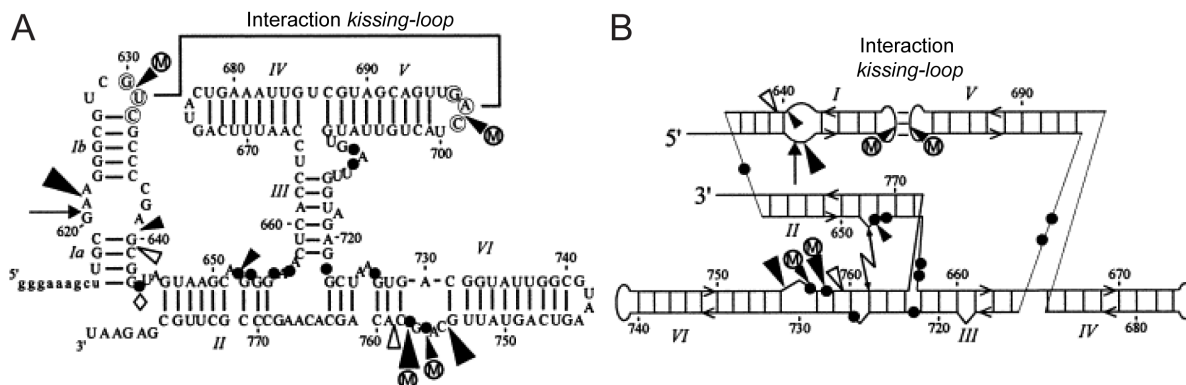


Figure 1.18 : Résumé des expériences d'interférence à l'agent ENU et aux phosphorothioates. (A) Structure secondaire du ribozyme VS dans laquelle les sites d'interférence à l'agent ENU et aux phosphorothioates sont indiqués : Les triangles noirs indiquent des sites où l'introduction de groupement phosphorothioates a eu un effet inhibiteur sur l'activité de clivage du ribozyme, alors que les triangles blancs indiquent un effet stimulateur sur l'activité. Les sites montrant une interférence aux phosphorothioates ayant été restaurée par l'utilisation de manganèse sont représentés par un M encadré. Les cercles noirs indiquent des sites protégés de l'interférence par l'agent ENU. (B) Représentation alternative du ribozyme VS accommodant l'interaction *kissing-loop* dans laquelle les sites d'interférence à l'agent ENU et aux phosphorothioates sont indiqués en utilisant les mêmes symboles qu'en A (Adapté de [104]).

Pour la boucle interne du substrat, un groupement phosphate important a été identifié, celui du résidu A622. Les expériences d'interférence aux phosphorothioates suggèrent que la liaison d'un cation magnésium à ce groupement phosphate est indirecte. Une étude RMN d'un substrat du ribozyme dans sa conformation active a aussi mis en évidence un site de liaison possible près du A622. En effet, le site 5'-A622-G623-3' est exclusif à la conformation active du substrat, suggérant que celui-ci est essentiel à la catalyse du ribozyme. La liaison du magnésium à ce site pourrait notamment permettre la liaison du substrat au site catalytique ou encore faciliter la réaction catalytique [104, 112].

Du côté de la boucle interne A730 du domaine VI, les groupements phosphates reconnus comme étant importants sont ceux des résidus C755, G757 et C758 [104, 105, 157] (Figure 1.18). Les études d'interférence à l'agent *ENU* et aux phosphorothioates ont montré que les groupements phosphates des résidus G757 et C758 sont impliqués dans des liaisons directes avec le magnésium, alors que celui du résidu C755 est impliqué dans une liaison indirecte. Une étude RMN d'un fragment de la boucle VI contenant la boucle interne A730 a permis d'identifier d'importants sites de liaison au magnésium composés des groupements phosphates de ces trois mêmes résidus (sites 1 et 2, Figure 1.19 A). Cette étude décrit les mêmes types de liaisons directes et indirectes entre ces groupements phosphates et le magnésium que ceux mis en évidence par les expériences d'interférence à l'agent *ENU* et aux phosphorothioates, à l'exception de celui impliquant le résidu G758 (site 2, Figure 1.19 A). En effet, le groupement phosphate de ce dernier, initialement associé avec une liaison directe par les expériences d'interférence avec le ribozyme VS actif [104], est plutôt décrit comme un ligand indirect par les données RMN obtenues sur le domaine VI isolé [105] (site 2, Figure 1.19 A). Enfin, différents rôles ont été proposés pour les cations s'associant aux boucles internes des domaines I et VI. Pour le cation près du A622, les rôles possibles sont notamment de faciliter la liaison du substrat et contribuer au positionnement du groupement phosphate à cliver [104, 112]. Pour les cations magnésium localisés à la boucle A730, près des nucléotides C755, G757 et C758, il a été proposé que ceux-ci participent à la neutralisation des charges négatives pour permettre le rapprochement du

domaine VI des domaines I ou II, facilitant ainsi la formation du site actif du ribozyme [104]. Il a aussi été suggéré qu'un cation magnésium localisé près de la boucle A730 joue un rôle structural au niveau de la catalyse du ribozyme, notamment en permettant la formation du motif *S-turn* (C755-A756-G757) et l'exclusion du nucléotide A756 [105]. De plus, il est possible que ce cation participe indirectement à la catalyse acide-base générale en contribuant à la modulation du pK_a des bases impliquées dans cette catalyse, soient l'adénine 756 et la guanine 638 [105, 158]. En effet, il a été démontré que la modulation des pK_a de ces bases dépend de l'identité du cation utilisé [158].

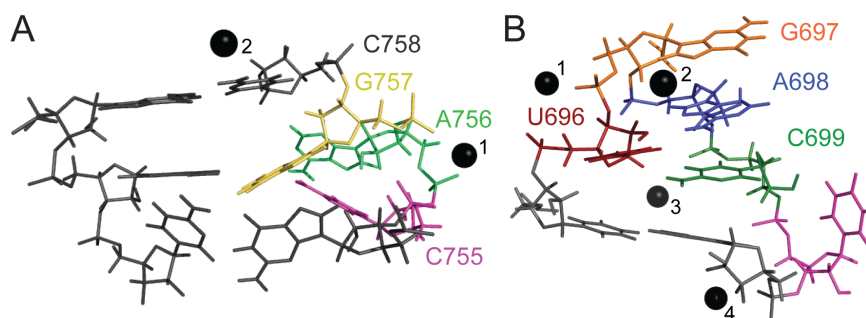


Figure 1.19 : Interaction du magnésium avec la boucle interne A730 et la boucle terminale du domaine V. (A) Structure RMN de la boucle A730 montrant deux sites de liaison au magnésium [105]. (B) Structure RMN de la boucle V contenant quatre sites de liaison au magnésium [114]. Le magnésium est représenté par une sphère noire.

Puisque la formation de l'interaction *kissing-loop* I/V est dépendante de la présence de cations magnésium [82, 90, 107, 109], il n'est pas étonnant que des groupements phosphates importants y aient été identifiés. En effet, il a été démontré que ceux des résidus U631 (domaine I) et C699 (domaine V) sont impliqués dans une liaison directe avec un ou plusieurs cations par des expériences d'interférence à l'agent *ENU* et aux phosphorothioates [104, 157] (Figure 1.18). Parallèlement, une étude par spectroscopie RMN a démontré que la boucle V seule, adoptant un motif de type *U-turn*, est associée à quatre cations magnésium (numérotés 1 à 4, Figure 1.19 B), dont un liant le groupement phosphate du résidu C699 [114] (site 3, Figure 1.19 B). Cependant, cette étude ne permet pas de déterminer si ces sites sont conservés une fois l'interaction *kissing-loop* formée [114]. Bien que les données RMN actuelles ne permettent pas d'identifier les ligands du magnésium dans l'interaction *kissing-loop* I/V, il demeure

clair que le magnésium est crucial pour sa formation. D'ailleurs, il a été proposé que celui-ci soit impliqué dans la stabilisation de la conformation des boucles I et V et dans la neutralisation des répulsions de charges négatives créées lors du rapprochement de ces deux boucles [114]. De cette façon, le magnésium favoriserait la formation de l'interaction I/V.

En résumé, le rôle général des cations magnésium pour le ribozyme VS semble s'orienter vers une implication structurale, plutôt que vers un rôle direct dans la réaction catalytique. En fait, les cations semblent être impliqués au niveau du repliement, facilitant le rapprochement de domaines initialement éloignés et la formation des différentes interactions nécessaires à la catalyse, notamment les interactions boucle-boucle I/VI et *kissing-loop* I/V. Chez d'autres ribozymes, tel que l'intron du groupe I, les cations divalents jouent non seulement un rôle structural, mais aussi un rôle catalytique en participant directement au mécanisme réactionnel du ribozyme [159]. Dans le cas du ribozyme VS, aucune preuve ne suggère un rôle direct des cations magnésium dans la réaction chimique. Cependant, il n'est pas possible d'exclure un rôle catalytique direct ou indirect pour ces cations. En effet, des cations localisés près du site actif pourraient stabiliser les charges négatives d'un intermédiaire réactionnel ou encore permettre à un ou plusieurs groupements fonctionnels d'avoir un pK_a modifié pour faciliter la catalyse. D'ailleurs, tel que mentionné précédemment, les bases des nucléotides A756 et G638, impliqués dans la fonction du ribozyme, présentent chacune un pK_a modifié [97, 99, 100, 125] et il a été démontré que la modification de pK_a de ces bases est associée à la liaison de cations [158]. De cette façon, la liaison du magnésium à la boucle A730 permet la définition structurale du site actif et contribue probablement à la catalyse acide-base générale de manière indirecte.

1.2.3.2 Modèles tridimensionnels à basse résolution du ribozyme VS

Comme mentionné précédemment, le ribozyme VS adopte une structure tertiaire active en présence de magnésium. Cette structure est caractérisée par le repliement des jonctions II-III-VI et III-IV-V, la liaison du substrat via l'interaction *kissing-loop* I/V et

l'association des boucles internes I et VI. Suite à plusieurs études biochimiques et biophysiques, différents modèles à basse résolution ont été présentés pour tenter de comprendre et visualiser l'organisation spatiale du ribozyme entier.

Tout d'abord, le premier modèle à basse résolution du ribozyme est entièrement basé sur les études de mutagénèse et de modifications chimiques [92]. Dans ce modèle, chaque domaine hélicoïdal est représenté sous la forme d'un cylindre (Figure 1.20 A). Dans cet arrangement, le substrat, domaine I, est orienté de manière parallèle par rapport au domaine V de façon à maintenir l'interaction *kissing-loop* I/V. Ce modèle montre également que le substrat est positionné dans un sillon formé par les domaines II et VI. Cette organisation entre le substrat et les autres domaines hélicoïdaux crée un environnement à l'abri du solvant et des modifications chimiques. Cette région protégée comprend la boucle interne A730 du domaine hélicoïdal VI, les nucléotides près des deux jonctions à trois voies, l'interaction *kissing-loop* I/V et une grande partie du domaine II [92]. Ce modèle suggère donc que l'interaction *kissing-loop* I/V s'installerait entre les domaines II et VI, positionnant le site du clivage du substrat à proximité la boucle interne du domaine VI de façon à former le site actif [92].

Par la suite, un deuxième modèle à basse résolution a été proposé, intégrant des contraintes angulaires entre les domaines hélicoïdaux formant les deux jonctions à trois voies (Figure 1.20 B) [117]. En effet, ce modèle a été élaboré à partir d'études effectuées par comparaison de mobilité sur gels d'électrophorèse et par *FRET* sur des ARN représentant chacun des deux sous-domaines du domaine catalytique (II-III-VI et III-IV-V). Ces expériences ont permis d'obtenir la grandeur relative des angles entre les hélices provenant de chaque jonction [117, 120]. Ce sont ces angles interhélicoïdaux qui ont permis d'élaborer ce modèle du domaine catalytique du ribozyme. Finalement, dans le but de compléter ce modèle, le substrat y a été placé manuellement, de façon à respecter l'interaction *kissing-loop* I/V et la proximité des boucles internes I et VI pour la création du site actif.

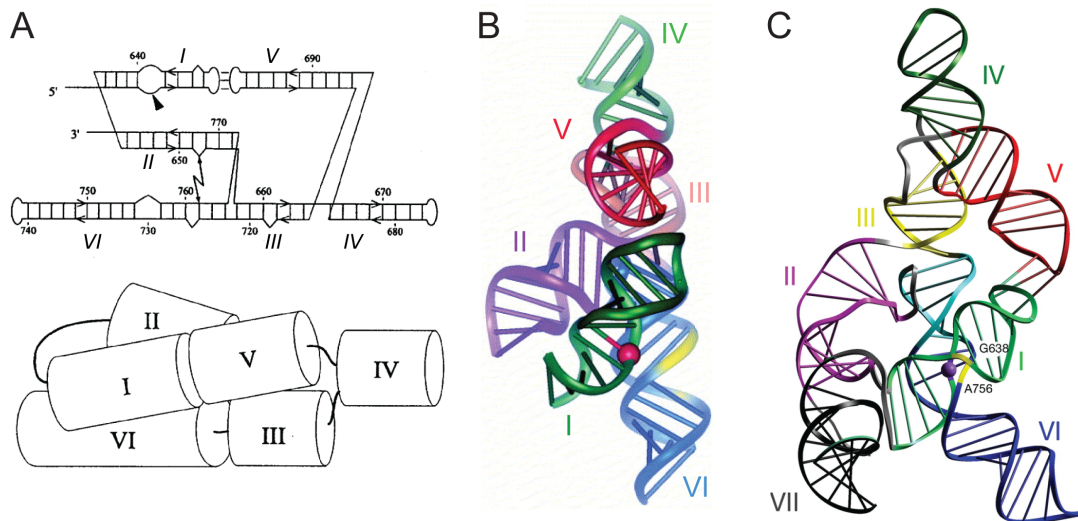


Figure 1.20 : Modèles à basse résolution de la structure tridimensionnelle du ribozyme VS.

(A) Modèle obtenu à partir d'études de mutagenèse et de modifications chimiques. Chaque domaine hélicoïdal est représenté par un cylindre [160]. (B) Modèle obtenu par des études de *FRET* et de mobilité sur gel. Le site de clivage est indiqué par la sphère rouge [116]. (C) Modèle obtenu par une étude de *SAXS*. Le site de clivage est indiqué par une sphère violette [147].

Plus récemment, un troisième modèle à basse résolution a été obtenu, cette fois à partir d'une étude de *SAXS* (Figure 1.20 C) [147]. Contrairement aux deux modèles précédents représentant un ribozyme à six domaines hélicoïdaux (Figure 1.20), celui-ci a été obtenu avec une séquence étendue contenant un domaine hélicoïdal supplémentaire, le domaine VII. Ce dernier fait partie d'une jonction à trois voies additionnelle joignant les domaines I, II et VII et favoriserait l'activité de ligation du ribozyme [161]. Cette étude de *SAXS* a permis d'obtenir l'enveloppe de densité électronique occupée par le ribozyme, à l'intérieur de laquelle les jonctions et les différents domaines hélicoïdaux ont été modélisés.

De façon générale, bien que ces modèles n'aient pas été conçus pour avoir une résolution atomique des différents éléments structuraux qui définissent le ribozyme VS, ils en illustrent bien le positionnement global. En effet, dans ces modèles, les domaines III, IV et VI sont globalement coaxiaux et forment l'axe le plus long de la structure, à partir duquel les domaines II et V forment chacun une ramification. Puis, tenant compte

de l'interaction *kissing-loop* entre les boucles I et V et du site actif créé par l'association des boucles internes I et VI, ces modèles proposent que le substrat soit posé dans une crevasse formée des domaines II et VI [92, 117, 147]. En terminant, il est nécessaire d'obtenir des structures à haute résolution pour mieux comprendre la reconnaissance du substrat et le mécanisme réactionnel du ribozyme VS.

1.3 Éléments de structure et repliement de l'ARN

La fonction de l'ARN est directement reliée à sa structure. Pour accomplir sa fonction, l'ARN doit se plier et adopter une forme active. Le repliement de l'ARN s'effectue à deux niveaux : secondaire et tertiaire. La structure secondaire est définie par les régions qui forment les hélices et les régions simples brins qui joignent celles-ci [162]. Les éléments de structure secondaire sont les tige-boucles, les boucles internes, les *bulges*, les jonctions multivoies et les hélices (Figure 1.21). La structure tertiaire, de son côté, est définie par des interactions à longue distance entre différents éléments de structures secondaires [162] et est stabilisée par deux éléments importants: les motifs structuraux et la liaison de cations divalents.

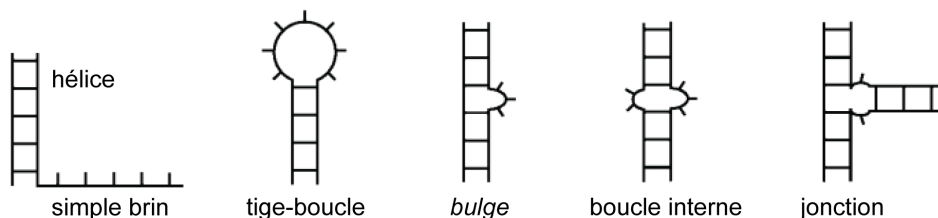


Figure 1.21 : Représentation de plusieurs éléments communs de structure secondaire. (Tiré de [163]).

Un motif structural est un élément structural récurrent qui peut se définir comme un réseau ordonné et orienté de paires de bases généralement non W-C qui façonne le squelette ribose-phosphate des brins d'ARN qui interagissent [164]. Par exemple, il y a des motifs de boucle terminale tels que le *U-turn*, les tétraboucles (boucles formées de quatre nt) de type GNRA (G=guanine, N= A, G, C ou U, R= A ou G et A= adénine) et CUYG (C=cytosine, U=uridine, Y= C ou U et G=guanine) et des motifs de boucle

interne tels que le *A-minor*, le *S-turn* ou le *K-turn* [165]. L'assemblage de plusieurs motifs structuraux permet notamment la formation et la stabilisation de la structure tertiaire de l'ARN. En effet, des motifs structuraux plus complexes tels que le *ribose zipper*, le récepteur à tétraboucle (*tetraloop receptor*) et l'interaction de type *kissing-loop* [165] résultent de l'association entre plusieurs motifs plus simples grâce à différents types d'interactions (des ponts hydrogènes impliquant des riboses, des bases ou des groupements phosphates, des empilements de bases ou encore la formation de paires de bases). Les sections suivantes traiteront de deux motifs pertinents à cette thèse: le *U-turn* et l'interaction de type *kissing-loop*.

1.3.1 Motif *U-turn*

Le motif de type *U-turn* est très répandu dans les structures d'ARN, notamment dans les boucles terminales des tige-boucles. Ce motif a été décrit pour la première fois dans la boucle terminale de l'anticodon de l'ARNt de la phénylalanine [166]. Il a par la suite été observé dans de nombreuses tige-boucles, boucles internes et jonctions chez plusieurs ARN fonctionnels [167-175]. Tel que mentionné précédemment (section 1.2.1.1.3), ce type de motif présente la séquence consensus UNR (U=uridine, N=adénine, cytosine, guanine ou uridine et R=adénine ou guanine) et est principalement caractérisé par un changement de direction du squelette ribose-phosphate d'environ 180° entre le U et N, lui donnant une forme de « U ». Le *U-turn* est stabilisé par deux ponts hydrogènes : un premier entre le H3 du U et le groupement 3'-phosphate du nucléotide R, puis un deuxième entre le groupement 2'-OH du U et le N7 du nucléotide R (Figure 1.22). Des interactions d'empilements de bases caractérisent aussi ce motif : l'empilement de la base du U sur le groupement 5'-phosphate du nucléotide R et un empilement des bases N et R situées en 3' du U exposant leur face W-C dans le sillon mineur (Figure 1.22).

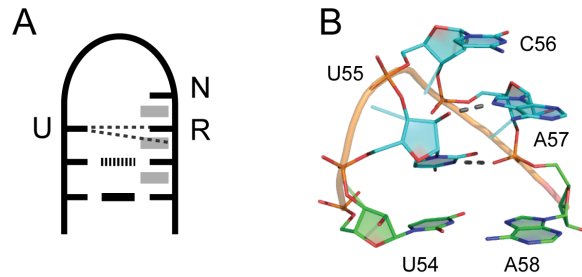


Figure 1.22 : Représentation du motif de type *U-turn*. (A) Représentation schématique du motif *U-turn* et (B) structure tridimensionnelle d'un motif *U-turn*, tel que retrouvé dans l'ARNt de l'isoleucine (PDB : 1FFY). Les rectangles gris représentent des interactions d'empilement, les rectangles noirs représentent des appariements de type W-C, alors que les rectangles pointillés représentent des appariements non W-C. Les lignes pointillées représentent les deux ponts hydrogènes : entre le proton imino du U et le 3'-P du nucléotide R et entre le groupement 2'-OH du U et le N7 du nucléotide R.

De façon générale, le motif *U-turn* favorise le repliement de l'ARN, notamment la formation des boucles terminales dans lesquelles il facilite le changement de direction abrupte du squelette ribose-phosphate. Ce motif stabilise aussi ces boucles, puisqu'il est caractérisé par un réseau de plusieurs interactions d'empilement et de ponts hydrogènes [166, 168, 176]. D'ailleurs, il a été démontré que les interactions stabilisatrices composant le *U-turn* permettent une affinité optimale de l'interaction codon/anticodon [176-178]. En effet, des études de liaison effectuées avec le *U-turn* de l'anticodon de l'ARNt de la phénylalanine ont permis de mettre en évidence l'importance des deux ponts hydrogènes décrits précédemment, en plus de deux interactions d'empilement [176, 177]: l'empilement entre la base du U de la séquence UNR et (1) le groupement 5'-phosphate du nucléotide R et (2) la base situé en 5' du U. Le motif *U-turn* sert également de charpente pour l'élaboration d'interactions tertiaires. En effet, l'exposition des bases localisées en 3' du U rend celles-ci accessibles pour des interactions avec d'autres ARN ou encore des protéines [166, 168-170, 172, 175, 179, 180]. D'ailleurs, le motif *U-turn* s'est vu impliqué dans l'initiation de l'appariement rapide entre deux ARN dans des processus tels que la reconnaissance codon/anticodon [176, 178] et le ciblage des ARNs [181]. En plus d'être fréquemment impliqué dans le repliement des boucles terminales, le motif *U-turn* favorise aussi le repliement des jonctions, telles que celles

retrouvées chez les ribozymes *HDV* et *hammerhead* et contribue à aider leur catalyse [174]. En effet, pour le ribozyme *HDV*, le *U-turn* crée une courbure qui expose le site de clivage au nucléotide catalytique C75, alors que pour le ribozyme *hammerhead*, le *U-turn* façonne le site actif en positionnant les nucléotides catalytiques autour de la liaison à cliver [167, 174, 182].

Il a été démontré que le motif *U-turn* est souvent associé à des cations divalents [114, 183]. En effet, une étude RMN caractérisant le *U-turn* de la boucle V du ribozyme VS a identifié des sites de liaison au magnésium qui sont communs dans plusieurs motifs *U-turn* [114]. Un de ces sites est composé des groupements 5'-phosphates des résidus U, N et R qui créent un regroupement compact de charges négatives dans le renversement du squelette ribose-phosphate. Un deuxième site de liaison au magnésium est situé dans le sillon mineur de la boucle et comprend le 2'-O du U, ainsi que les N7 des nucléotides N et R. Puis, un troisième site se situe dans le sillon majeur et est composé de l'atome O4 du résidu U et du groupement 3'-phosphate du résidu R [114].

1.3.2 Interaction *kissing-loop*

Une interaction de type *kissing-loop* résulte de la formation de paires de bases entre deux boucles terminales (Figure 1.23 A). Elle est cruciale dans de nombreux événements biologiques importants dépendant de l'ARN, tels que l'initiation de la traduction, le contrôle de la réplication de plasmides, la reconnaissance d'ARNas ou encore la dimérisation de génome de rétrovirus [184-186].

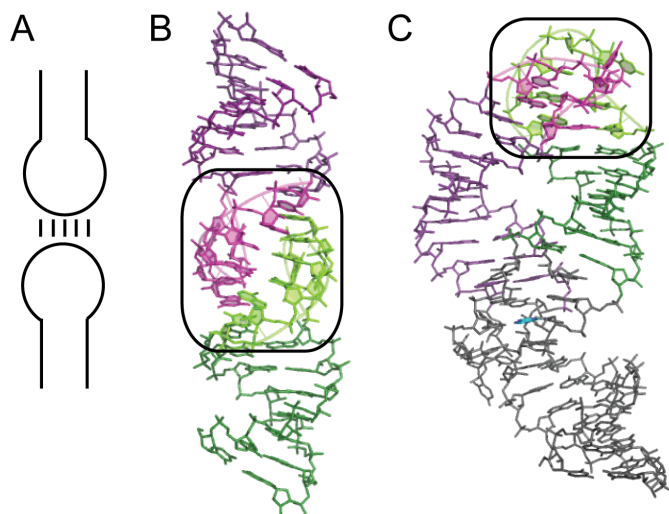


Figure 1.23 : Représentation du motif de type *kissing-loop*. (A) Représentation schématique d'une interaction *kissing-loop* entre deux tige-boucles. (B) Structures tridimensionnelles de l'interaction *kissing-loop DIS* retrouvée dans le VIH-1 (PDB : 2BS8) et (C) P2-P3 retrouvé dans le riborégulateur à adénine (PDB : 1Y26). Les nucléotides impliqués dans l'interaction *kissing-loop* sont encadrés et les deux molécules ou régions qui interagissent sont colorées de différente couleur (vert et violet).

Par exemple, dans le cas du virus du VIH-1, la formation d'une interaction *kissing-loop* implique une séquence nommée site d'initiation de la dimérisation (*dimerization initiation site, DIS*) et est importante pour la dimérisation du génome d'ARN rétroviral [187] (Figure 1.23 B). En effet, cette dimérisation est essentielle pour la survie du virus, au niveau de sa réplication, son empaquetage, sa transcription inverse et sa capacité à générer une infection [187, 188]. En plus de démontrer une importance au niveau fonctionnel, les interactions *kissing-loop* jouent aussi un rôle déterminant au niveau structural dans plusieurs ARN. Elles permettent l'interaction entre deux molécules d'ARN ou encore entre deux domaines éloignés dans une même molécule. Elles stabilisent également la charpente des ARN complexes comme les riborégulateurs [189-194], les ribozymes [195-198] et les ARNt [166]. Par exemple, dans le cas du riborégulateur à adénine, une interaction *kissing-loop* se forme entre les boucles des domaines P2 et P3 [191] (Figure 1.23 C). Cette interaction stabilise la structure native du riborégulateur et est également essentielle à la liaison de l'adénine en participant à la

création de son site de liaison. [190]. Les interactions *kissing-loop* s'avèrent donc particulièrement importantes au niveau fonctionnel et structural.

En plus d'être caractérisées par la formation de plusieurs paires de base W-C entre deux boucles, de nombreuses interactions *kissing-loop* présentent des interactions stabilisantes supplémentaires telles que des interactions non W-C ou encore des empilements avec les tiges adjacentes, aussi appelés empilement coaxial (*coaxial stacking*). L'ensemble de ces interactions fournit au motif *kissing-loop* une grande stabilité avec des ΔG variant de -6 à -15 kcal/mol dépendant du nombre de paires de bases de l'interaction [199-202]. Il a également été démontré que la présence de cations contribue à la stabilité de plusieurs interactions de type *kissing-loop* [200, 201, 203-208]. En plus de la nature des cations présents, l'identité des nucléotides des tiges près de la jonction influence aussi la stabilité des interactions *kissing-loop* [209].

Quelques interactions *kissing-loop* ont fait l'objet d'études approfondies dans le but d'en caractériser la structure et la stabilité. C'est notamment le cas de l'interaction *kissing-loop* de sept paires de bases entre l'ARN I et l'ARN II qui est importante pour la régulation du nombre de copie du plasmide *ColEI* dans la bactérie *E.coli* [210-214] et de l'interaction *DIS* du VIH-1 composée de six paires de bases [201, 215] mentionnée plus haut. Une autre interaction *kissing-loop* a aussi été beaucoup étudiée, celle de TAR/TAR* [216, 217]. Celle-ci implique l'élément de réponse de transactivation (*trans-activator responsive RNA, TAR RNA*) qui est situé dans les ARNm du VIH et qui contribue à l'efficacité de transcription du virus et est nécessaire à sa réplication [218, 219]. Cette interaction de six paires de bases se forme entre l'ARN TAR et un ARN TAR* obtenu par *SELEX* [220]. Les structures de ces interactions révèlent un arrangement globalement coaxial dans lequel la jonction entre les deux molécules forme une troisième hélice où des cations se lient (Figure 1.23 B) [208, 213, 215, 221].

Puisque les interactions de type *kissing-loop* démontrent une importance autant structurale que fonctionnelle, il est particulièrement intéressant de comprendre leur structure, la thermodynamique de leur formation, ainsi que leur dépendance aux cations.

Même si quelques interactions *kissing-loop* ont été bien caractérisées à ce jour, particulièrement celles de ARN I / II de *ColE1* et du *DIS* du VIH-1, du travail demeure nécessaire afin d'avoir une meilleure compréhension des règles gouvernant la formation de ce type d'interactions.

1.4 Étude de la structure de l'ARN par spectroscopie RMN

L'étude de la structure des macromolécules, telles que les protéines et les acides nucléiques (ARN et ADN), s'inscrit dans le domaine de la biologie structurale et peut s'effectuer par diverses méthodes biophysiques dont la cristallographie aux rayons X, la spectroscopie RMN, la cryomicroscopie électronique et la diffusion de rayon X à petits angles (*SAXS*). Ces méthodes comportent toutes leurs avantages et inconvénients et permettent d'obtenir de l'information sur la structure tridimensionnelle des macromolécules avec une résolution atomique plus ou moins grande.

La cryomicroscopie et le *SAXS* sont deux méthodes qui permettent de visualiser la forme globale des macromolécules, plutôt que la structure précise et bien résolue de chaque acide aminé ou acide nucléique. Un des intérêts particuliers de ces deux méthodes est l'étude de molécules de très grande taille, notamment des complexes macromoléculaires, sans nécessiter de cristallisation.

La cristallographie et la spectroscopie RMN sont deux méthodes qui permettent d'obtenir de l'information structurale à haute résolution. La cristallographie nécessite l'utilisation d'une molécule sous forme cristalline, mais présente l'avantage de n'avoir aucune limite de taille pour la molécule à étudier. De son côté, la RMN présente l'avantage d'utiliser une molécule en solution, mais comporte comme désavantage une limite de taille d'environ 30 kDa pour une protéine et de 15 à 30 kDa (50-100 nt) pour un acide nucléique [222]. De plus, un avantage additionnel de la RMN par rapport à la cristallographie est la possibilité d'obtenir de l'information structurale sur des régions désordonnées qui sont peu ou pas visibles par la cristallographie. Le reste de cette

section mettra l'emphase sur la spectroscopie RMN, car c'est cette méthode qui a été choisie pour réaliser les travaux présentés dans cette thèse.

1.4.1 Spectroscopie RMN pour l'étude des macromolécules biologiques

La spectroscopie RMN est une méthode idéale pour l'étude de la structure des macromolécules biologiques, notamment l'ARN. De toutes les structures d'ARN répertoriées en février 2014 dans la banque de données *PDB* (*protein data bank, PDB*), 43% (443/1022) d'entre elles ont été résolues par RMN. En plus de permettre la détermination de la structure tridimensionnelle, la spectroscopie RMN permet d'obtenir de l'information sur la dynamique, l'équilibre entre différentes conformations, les valeurs de pK_a de groupements ionisables, les sites de liaison aux cations et la cartographie des régions d'interaction intermoléculaire [223].

La spectroscopie RMN utilise les propriétés magnétiques de certains noyaux atomiques, notamment le spin nucléaire I . Le spin nucléaire est une propriété quantique des noyaux atomiques pouvant adopter une valeur nulle ($I=0$), entière ($I=1, 2$, etc) ou encore fractionnaire ($I=1/2, 3/2$, etc) [224]. Toutefois, seuls les noyaux possédant un spin non nul sont détectables par RMN. En fait, les noyaux qui sont particulièrement d'intérêt en spectroscopie RMN des macromolécules sont ceux possédant un spin $I=1/2$. Par contre, plusieurs des noyaux composant les macromolécules ne possèdent pas de spin de $1/2$. C'est notamment le cas du carbone (C) et de l'azote (N) pour lesquels les isotopes les plus abondants naturellement, le ^{12}C et le ^{14}N , présentent des spins $I=0$ et $I=1$ respectivement. De ce fait, lorsque des macromolécules sont étudiées en RMN, il est possible de procéder à leur enrichissement isotopique avec les isotopes moins abondants naturellement, mais présentant des spins I de $1/2$, tels que les ^{13}C et ^{15}N . À l'opposé, les noyaux de l'hydrogène (H) et du phosphore (P) pour lesquels les isotopes les plus abondants naturellement sont ^1H et ^{31}P possèdent chacun un spin I de $1/2$, un enrichissement n'est donc pas nécessaire pour ces atomes.

La spectroscopie RMN est basée sur le phénomène de résonance magnétique nucléaire qui est détectée chez certains noyaux atomiques lorsqu'ils sont en présence d'un champ magnétique externe et excités par un rayonnement électromagnétique [224]. En présence d'un champ magnétique, le spin I de $\frac{1}{2}$ se présente en deux états dont un est orienté avec le champ magnétique ($+\frac{1}{2}$) et l'autre orienté dans le sens opposé au champ ($-\frac{1}{2}$). L'excitation des noyaux à une radiofréquence précise permet aux spins de faire une transition d'un état à l'autre, ce qui correspond au phénomène de résonance [224]. Un de ces états est prépondérant par rapport à l'autre et cet excès (ΔE), représentant la magnétisation macroscopique résultante, permet la détection et est relié à l'intensité du signal produit [224]. Puisque la magnétisation résultante est aussi proportionnelle au champ magnétique appliqué, la puissance du champ magnétique du spectromètre utilisé influencera l'intensité des signaux et la résolution des spectres. La spectroscopie RMN mesure la fréquence de résonance des noyaux atomiques et celle-ci dépend de la nature du noyau et de son environnement chimique. La fréquence est mesurée en Hertz (Hz) et transformée en partie par million (ppm) par rapport à une valeur de référence pour obtenir une valeur indépendante de la puissance du spectromètre utilisé [224]. Cette valeur en ppm représente le déplacement chimique (δ) du noyau. Les différents noyaux qui composent les macromolécules sont principalement les carbones, les azotes et les hydrogènes. Ces derniers, aussi appelés les protons, sont les noyaux les plus utilisés dans la RMN des protéines et des acides nucléiques.

Les différents types d'expériences RMN permettent non seulement d'observer le déplacement chimique de chaque proton, mais aussi de détecter les corrélations proton-carbone (C-H) et proton-azote (N-H) grâce au couplage scalaire ou couplage J . Le couplage J reflète une interaction entre deux noyaux par l'intermédiaire des liaisons chimiques entre ces noyaux [224]. La force de ce couplage dépend notamment du nombre de liaisons qui séparent les deux noyaux et de la géométrie des liaisons chimiques. Par exemple, le couplage entre deux protons séparés par deux liaisons chimiques (H-C-H) est généralement plus fort que celui entre deux protons séparés par trois liaisons chimiques (H-C-C-H). Donc, le couplage J , observé à travers les liens chimiques, permet de définir les connections entre les différents atomes ainsi que les

angles de torsion. Pour déterminer si deux noyaux, particulièrement des protons, sont près l'un de l'autre dans l'espace, un autre type d'expérience basée sur l'effet nucléaire Overhauser (*nuclear Overhauser effect, NOE*) est utilisé. L'effet *NOE* est un transfert de magnétisation qui s'effectue entre deux noyaux lorsque ceux-ci sont près dans l'espace [224], sans nécessairement être liés l'un à l'autre par l'intermédiaire de liaisons chimiques. L'intensité de cet effet est inversement proportionnel à la distance entre les deux noyaux (r) à la puissance 6 ($NOE \propto 1/r^6$) [225]. Par conséquent, plus le signal observé est intense, plus la distance entre les deux noyaux est courte. Ainsi, l'analyse des intensités des signaux *NOE* interprotons permet d'obtenir de nombreuses contraintes de distances. Finalement, les signaux *NOE* interprotons observés à travers l'espace en combinaison avec les signaux résultant du couplage J à travers les liens chimiques permettent l'obtention de nombreuses contraintes de distances et d'angles de torsion qui seront utilisés pour le calcul de la structure de la molécule d'intérêt. Les sections suivantes présentent les stratégies spécifiques utilisées pour la détermination d'une structure d'ARN par RMN.

1.4.2 Les difficultés et avantages de la RMN d'ARN

1.4.2.1 Les difficultés de la spectroscopie RMN

Tel que mentionné plus haut, la spectroscopie RMN est une méthode idéale pour faire l'étude de la structure des ARN, mais celle-ci est également associée à quelques difficultés qui augmentent avec la taille des ARN. Une première difficulté est la dispersion restreinte des valeurs de déplacements chimiques des protons des quatre nucléotides qui composent l'ARN (Figure 1.24). Ceci constitue un problème particulièrement important pour l'attribution des signaux, car cette faible dispersion crée beaucoup de superposition de signaux et d'ambiguïtés (Figure 1.24 C). Puis, une deuxième difficulté concerne l'obtention de contraintes de distances et d'angles qui sont locales, plutôt que globales. En effet, l'effet *NOE* et le couplage J s'observent sur une distance limitée, notamment $\sim 5\text{\AA}$ pour l'effet *NOE*, ce qui permet de définir des environnements locaux. Ceci compromet la précision de la structure globale, car l'effet

NOE et le couplage *J* ne se mesurent pas sur une longue distance ou à travers des domaines éloignés. La rencontre de ces difficultés a eu comme avantage de stimuler l'avancement du domaine de la RMN et a permis à la méthode d'évoluer en développant différentes approches pour résoudre ou réduire ces difficultés.

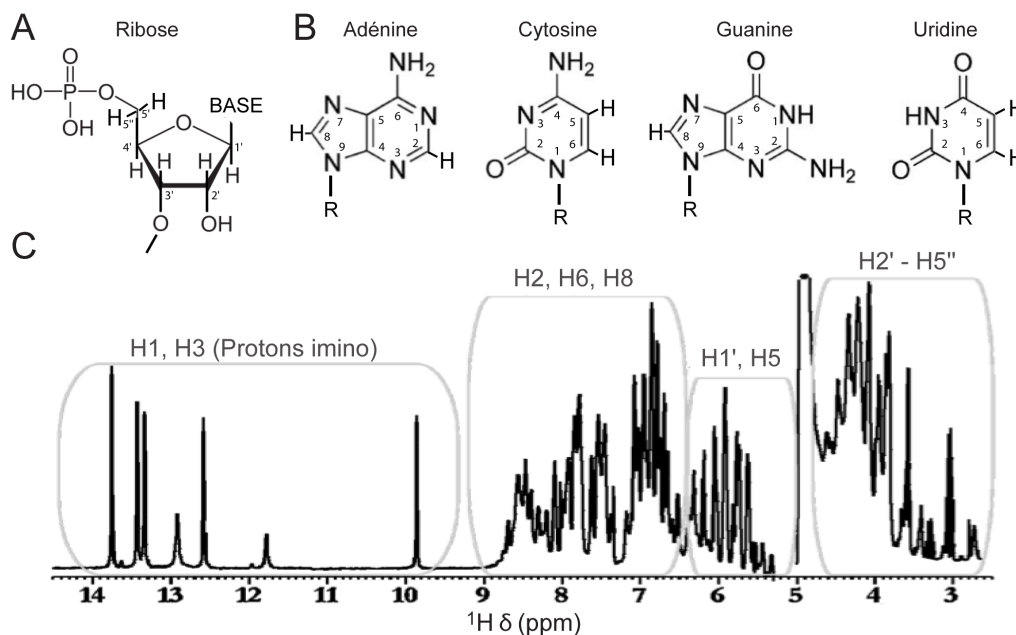


Figure 1.24 : Composition chimique de l'ARN et caractéristiques spectrales des différents protons de l'ARN. Chaque nucléotide est composé (A) d'une partie ribose-phosphate et (B) d'une base azotée. (C) Spectre ^1H 1D du proton d'un ARN de 14 nt contenant une boucle UUCG. Les différents protons sont attribués sur le spectre selon la même numérotation que celle utilisée en (A) et (B). (Adapté de [226]).

Dispersion limitée des déplacements chimiques des protons

La dispersion restreinte des valeurs de déplacement chimique des protons des nucléotides A, C, G et U crée des problèmes de superposition de signaux. En effet, ces nucléotides sont tous composés du même ribose et les bases azotées ont toutes le même anneau purine ou pyrimidine comme fondation (Figure 1.24 A, B). Néanmoins, les déplacements chimiques des hydrogènes des bases présentent un peu plus de dispersion grâce aux différents groupements fonctionnels qui différencient chaque pyrimidine et purine (Figure 1.24 C). Outre la nature chimique des nucléotides, un autre élément qui influence les déplacements chimiques des protons est l'environnement chimique et

électronique dans lequel ils sont situés. Puisque l'ARN est en grande partie composé d'éléments hélicoïdaux, de nombreux nucléotides se retrouvent dans un environnement relativement similaire, générant ainsi des déplacements chimiques semblables et par conséquent de la superposition de signaux. Le problème de dispersion peut être réduit significativement en passant de la RMN homonucléaire, qui se limite à l'analyse d'un seul noyau, le proton, à l'utilisation de la RMN hétéronucléaire dans laquelle plusieurs noyaux sont utilisés, le carbone, l'azote et le proton. La RMN hétéronucléaire nécessite le marquage isotopique des atomes de carbones et d'azotes avec les isotopes ^{13}C et ^{15}N pour que ceux-ci deviennent détectables par RMN. De cette façon, il est possible d'enregistrer des expériences plus complexes avec des dimensions additionnelles (2D ou 3D, Figure 1.25) pour faire la détection des corrélations H-C ou H-N.

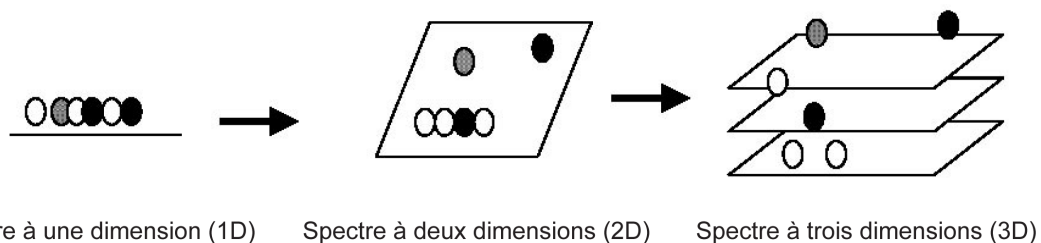


Figure 1.25 : Représentation schématique de spectres RMN à une ou plusieurs dimensions.

Le spectre à une dimension (1D) montre tous les signaux selon une seule dimension. L'ajout d'une dimension (2D) permet une dispersion de signaux selon la deuxième dimension. Un spectre en trois dimensions (3D) est un assemblage de spectres en 2D dispersés selon une troisième dimension. (Tiré de [227]).

Avec la RMN hétéronucléaire, il devient possible d'observer les signaux des protons via le carbone ou l'azote auquel ils sont liés. L'avantage de corréler les signaux ^1H à ceux du carbone ou de l'azote est leur grande dispersion de signal, ce qui permet de mieux résoudre les signaux ^1H . Ces corrélations permettent aussi d'utiliser la fréquence hétéronucléaire pour faire l'attribution des signaux. Par exemple, les H3 des uridines et les H1 des guanines présentent le même intervalle de déplacement chimique, de 9 à 15 ppm, mais il est possible de les distinguer par le déplacement chimique de leur azote, de 157 à 162 ppm pour le N3 des uridines et de 145 à 148 ppm pour le N1 des guanines [222] (Figure 1.26 C). L'attribution des protons est ainsi grandement facilitée.

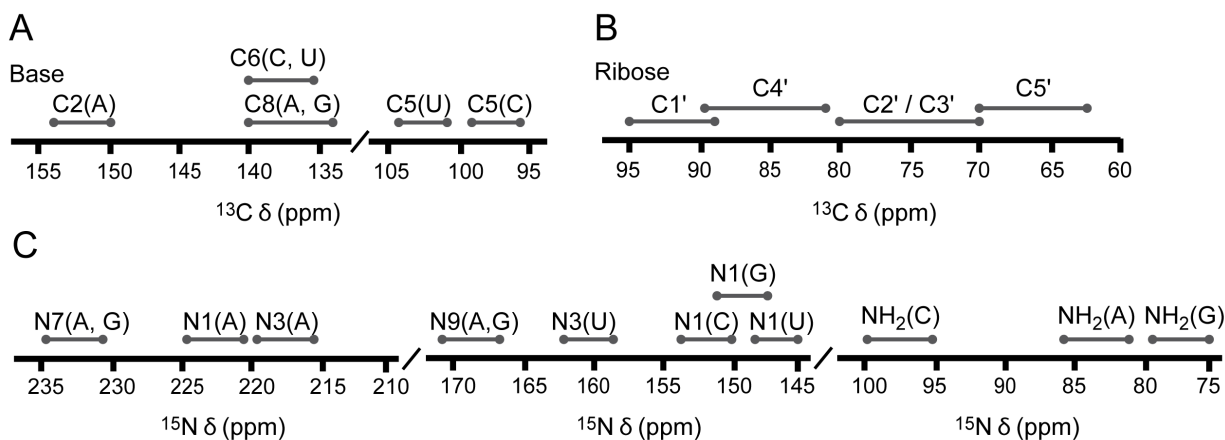


Figure 1.26 : Dispersion spectrale des différents atomes de carbone et d'azote qui composent l'ARN. Déplacements chimiques des carbones des (A) bases azotées et du (B) ribose. (C) Déplacement chimiques des atomes d'azotes des quatre types de bases. (Données tirées de [222]).

Ce type de marquage est idéal pour des ARN d'une cinquantaine de nucléotides. Cependant pour faire l'étude d'ARN de plus grande taille, d'autres stratégies peuvent être utilisées : le marquage sélectif ou par fragment et l'approche du *divide and conquer*. La stratégie de marquage isotopique sélectif permet la détection d'un seul type de nucléotide ou d'une combinaison de nucléotides, par exemple seulement les cytosines ou seulement les cytosines et les guanines sont marquées avec les isotopes ¹³C/¹⁵N, alors que les autres nucléotides demeurent non marqués. Il est aussi possible d'utiliser plusieurs types d'échantillons dans lesquels seulement un des domaines de la molécule est marqué avec des nucléotides ¹³C/¹⁵N. La stratégie du *divide and conquer* consiste à diviser la molécule entière en différents fragments qui seront par la suite caractérisés de façon individuelle et isolée du contexte naturel de la molécule entière. Cette approche nécessite toutefois de la prudence, car la structure des différents fragments isolés peut varier de celle des fragments dans le contexte de la molécule entière. Il est néanmoins possible d'utiliser cette approche en s'assurant que le repliement de chaque fragment est similaire à celui retrouvé dans la molécule entière. Le but ultime est d'utiliser les structures à haute résolution des fragments isolés et une structure à basse résolution de la molécule entière pour reconstruire un modèle RMN à haute résolution [223].

Contraintes de distances locales

Une autre difficulté de la RMN dans la détermination de structure provient du fait que l'information extraite des données *NOE* et du couplage *J* est généralement locale. Dans l'étude de molécules de petite taille, comme celles contenant uniquement une courte tige-boucle, le manque de données à longue distance n'est pas aussi important que pour un complexe de plusieurs ARN ou encore pour un ARN composé de plusieurs domaines. Dans de tels contextes, des données à longue distance deviennent nécessaires pour déterminer une structure globale précise. Ce type de données peut être obtenu par spectroscopie RMN grâce au phénomène de couplage dipolaire résiduel (*residual dipolar coupling, RDC*). Le couplage dipolaire est une interaction à travers l'espace qui survient entre deux noyaux, par exemple N-H ou C-H [228]. Toutefois, lorsque les molécules bougent de façon aléatoire, tel qu'observé en solution, ce phénomène de couplage dipolaire avoisine 0. À l'opposé, lorsque les molécules sont orientées, tel qu'observé à l'état solide, il est possible de mesurer un fort couplage dipolaire. Dans le but de mesurer un couplage dipolaire en solution, il suffit de créer un environnement dans lequel les molécules sont partiellement orientées dans le champ magnétique en ajoutant un agent responsable de créer cet arrangement, par exemple des bicelles ou des phages filamenteux [228]. Chacun des couplages mesurés dans les liaisons N-H ou C-H représente un vecteur, puis l'analyse et le traitement des données de *RDC* permettent de connaître l'orientation de chaque vecteur par rapport à un référentiel global (Figure 1.27). De cette façon les *RDC* permettent d'obtenir de l'information sur l'orientation des différents domaines de la molécule et en combinaison avec les données provenant des signaux *NOE*, ils permettent d'améliorer la précision de la structure globale.

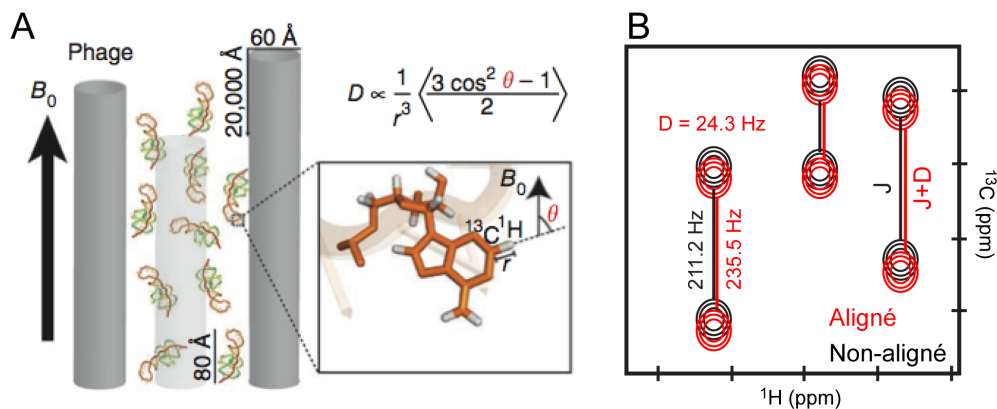


Figure 1.27 : Couplage dipolaire résiduel (RDC). (A) Représentation schématique de l'alignement partiel d'un ARN en présence d'un agent d'alignement, tel que le phage *Pfl* en présence d'un champ magnétique B_0 . (B) Mesure des couplages dipolaires résiduels à partir de la séparation des résonances observée en présence ($J+D$) et en absence (J) d'un agent d'alignement. D représente le RDC et J le couplage scalaire J (Adapté de [229]).

1.4.2.2 Les avantages de la spectroscopie RMN

En plus de permettre l'étude de la structure tridimensionnelle de l'ARN, la spectroscopie RMN permet de caractériser facilement la structure secondaire de l'ARN, particulièrement avec l'analyse des protons imino. En effet, l'enregistrement de quelques expériences seulement permet de déterminer une structure secondaire, ainsi que certains éléments de structures tertiaires. Cette stratégie est élaborée en section 1.4.3.2. La RMN permet également d'étudier des interactions entre un ARN et un ligand ou encore avec deux molécules d'ARN. En effet, des expériences simples permettent de suivre le titrage entre deux molécules, de cartographier la région d'interaction et d'évaluer la constante d'association entre ces deux molécules. La spectroscopie RMN est une technique très intéressante puisqu'elle permet d'accéder à de l'information structurale qui va au-delà de la structure tridimensionnelle de l'ARN. En effet, cette méthode permet d'étudier la dynamique de l'ARN, c'est-à-dire le mouvement intrinsèque de chaque élément structural à l'intérieur de la molécule. Ces données permettent de caractériser la flexibilité de l'ARN ou encore les transitions conformationnelles qu'elle subit [230]. De plus, l'analyse des déplacements chimiques des différents noyaux composant l'ARN permet de faire l'identification rapide de motifs

structuraux [114, 231]. C'est le cas notamment du motif *U-turn* qui peut être identifié facilement avec un nombre limité de données. La spectroscopie RMN permet aussi d'identifier des nucléotides présentant des pK_a modifiés via l'analyse de certains déplacements chimiques qui sont dépendant du pH [103, 232]. Une autre possibilité qu'apporte la spectroscopie RMN est l'étude du rôle des cations dans la structure et le fonctionnement de l'ARN [233]. En effet, il est possible de mesurer les changements de déplacements chimiques de chaque noyau dépendant de la présence des cations. Cette information permet ainsi de caractériser le repliement d'un ARN en fonction des concentrations de cations ou encore de localiser les sites de liaison de ces cations sur l'ARN [223].

1.4.3 Procédure pour la détermination de la structure d'un ARN

La détermination de la structure tridimensionnelle d'un ARN s'effectue en cinq étapes majeures [223] (Figure 1.28). Premièrement, il y a la préparation d'un ou plusieurs échantillons d'ARN. Deuxièmement, il est important d'identifier un ARN qui se replie en une seule conformation stable pour poursuivre avec les étapes ultérieures. Cela s'effectue en enregistrant des expériences préliminaires qui permettent de vérifier la structure secondaire, ainsi que certains éléments de structure tertiaire. Troisièmement, plusieurs types d'expériences RMN sont enregistrées sur l'ARN choisi dans le but de faire l'attribution des différents signaux, soient les H, C, N et P. Quatrièmement, une fois l'ensemble des signaux attribués aux atomes de l'ARN, plusieurs types d'expériences sont enregistrées et utilisées pour extraire les différentes contraintes de distances et d'angles de torsion. Ces contraintes peuvent ensuite être complétées par des contraintes à longue distance obtenues à l'aide de *RDC*. Finalement toutes ces contraintes structurales sont assemblées pour faire le calcul des structures qui permet d'obtenir un ensemble de structures en accord avec les contraintes RMN.

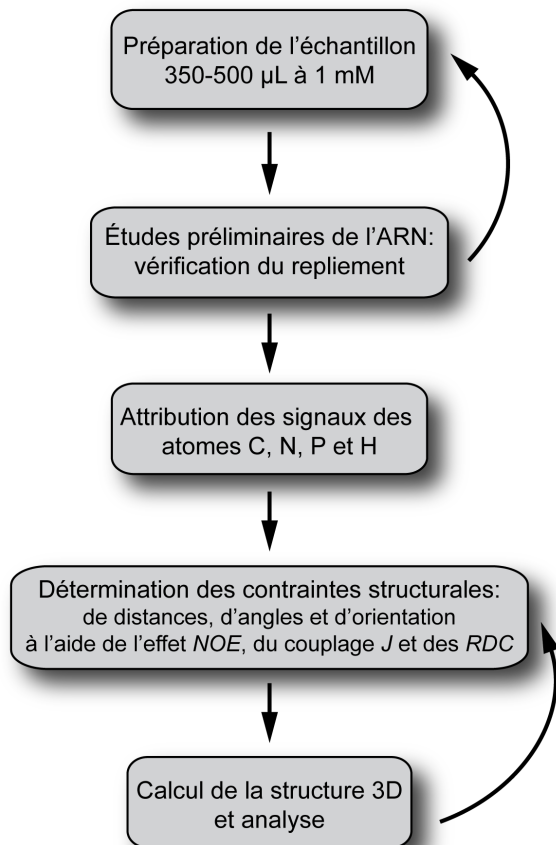


Figure 1.28 : Diagramme expliquant les étapes générales du protocole de détermination de structure par spectroscopie RMN. (Adapté de [223]).

1.4.3.1 Préparation des échantillons

Les échantillons nécessaires pour une étude de structure par RMN sont d'environ de 1 à 2 mM d'ARN dans un volume moyen de 350 à 500 µL. En plus de nécessiter des quantités relativement importantes d'ARN, celui-ci doit être pur. La méthode la plus commune pour produire de tels échantillons est illustrée en Figure 1.29 [223].

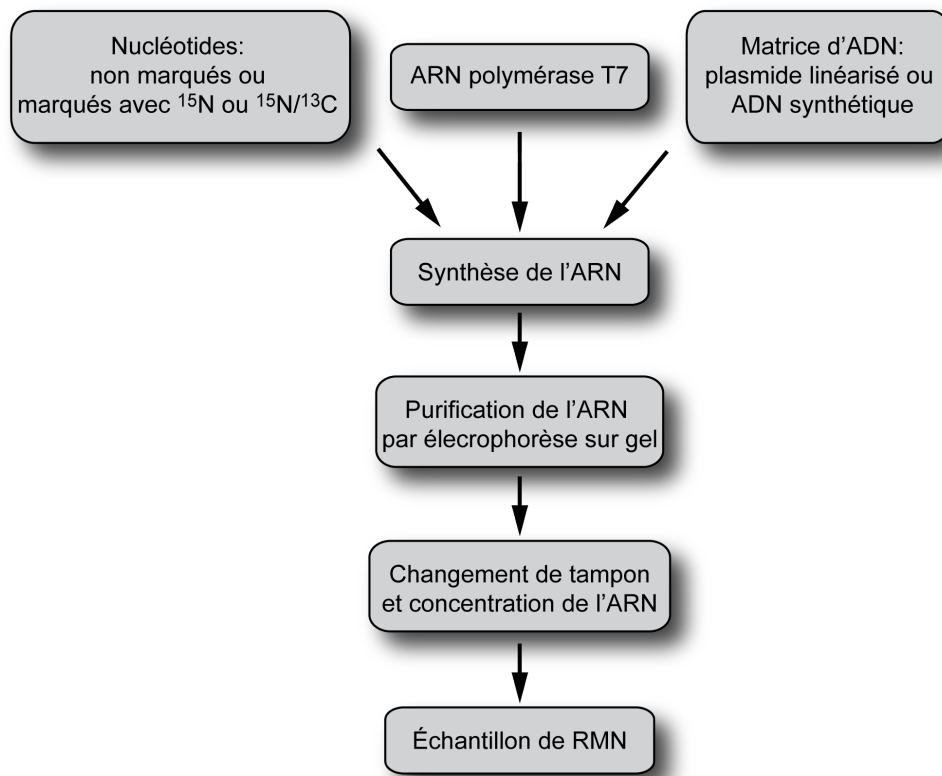


Figure 1.29 : Diagramme décrivant les étapes du protocole de synthèse et purification d'échantillon pour une étude par spectroscopie RMN. (Adapté de [223]).

La synthèse *in vitro* s'effectue à l'aide de nucléotides, de la polymérase à ARN du bactériophage T7 et d'une matrice d'ADN synthétique ou encore d'un plasmide linéarisé. Selon le type d'échantillon nécessaire, les nucléotides peuvent être non marqués, marqués uniformément en ^{15}N seulement ou en $^{15}\text{N}/^{13}\text{C}$. Si toutefois, un marquage sélectif est nécessaire, seulement un des nucléotides ou une combinaison de ceux-ci sont marqués (en ^{15}N seulement ou en $^{13}\text{C}/^{15}\text{N}$), alors que les autres nucléotides ne portent aucun marquage. Une fois l'ARN synthétisé, celui-ci doit être purifié. Une des principales méthodes utilisées est la purification par électrophorèse sur gel de polyacrylamide (*polyacrylamide gel electrophoresis*, *PAGE*). Cette méthode est idéale pour des ARN d'environ 35 nt ou moins et permet une purification avec une résolution d'un nucléotide. Suite à l'électrophorèse, l'ARN est extrait des gels et peut être purifié davantage par une étape de chromatographie échangeuse d'anion dans le but d'enlever

les particules d'acrylamide pouvant contaminer l'échantillon d'ARN. L'étape finale consiste à concentrer l'ARN et effectuer un changement de tampon pour la spectroscopie RMN [234].

1.4.3.2 Utilisation des protons imino pour la détermination de structures secondaire et tertiaire

Les protons imino sont les atomes d'hydrogène des groupements imino (NH) des guanines et des uridines (H1 et H3, respectivement) (Figure 1.24 B). Ces protons sont ceux qui présentent la plus grande dispersion de déplacement chimique sur le spectre RMN du proton (entre 9 et 15 ppm) et qui sont donc généralement les mieux résolus (Figure 1.24 C). Puisque les protons imino des bases libres sont en échange rapide avec les solvants aqueux, ils ne sont normalement pas détectables par spectroscopie RMN. Par contre, lorsque ces protons sont stabilisés dans des interactions, par exemple en participant à la formation de ponts hydrogènes dans une paire de bases, ils sont protégés de l'échange avec le solvant et deviennent ainsi détectables par RMN. L'analyse des protons imino est ainsi fort utile pour identifier les paires de bases de l'ARN et vérifier la structure secondaire prédite et même identifier certains éléments de structure tertiaire lorsqu'ils impliquent la formation de paires de bases [223]. Toutes les paires de bases W-C (G-C et A-U), ainsi que plusieurs paires de bases non W-C, telles que les G-U et les G-A impliquent la stabilisation d'au moins un proton imino dans un pont hydrogène $\text{NH}\cdots\text{N}$ ou $\text{NH}\cdots\text{O}$ (Figure 1.30 A) [223]. Par conséquent, chaque signal de proton imino sur le spectre correspond généralement à une paire de bases (exceptionnellement deux signaux imino sont obtenus pour une paire de bases G-U). Lorsque les paires de bases sont empilées, comme dans le contexte d'une hélice, les distances interprotons imino (NH-NH) dans les paires de bases adjacentes sont compatibles avec l'effet *NOE*, donc il est possible de définir une région hélicoïdale avec le patron de connections *NOE* NH-NH (Figure 1.30 B, C). L'utilisation des protons imino pour la détermination de structure est développée en plus de détails à la section suivante, 1.4.3.3. Il est également possible de vérifier l'homogénéité de conformation de l'ARN d'intérêt en utilisant les protons imino. Pour ce faire, il suffit d'enregistrer un spectre du proton en une

dimension, 1D ^1H , et évaluer si l'ARN adopte une ou plusieurs conformations par l'analyse du nombre et de l'intensité des signaux dans la région des protons imino.

1.4.3.3 Stratégie d'attribution des signaux

L'attribution des signaux représente le nœud dans le processus de détermination de structure. La stratégie utilisée comprend deux étapes principales : la première est l'attribution des protons échangeables, c'est-à-dire ceux qui peuvent s'échanger avec les protons du solvant, particulièrement les protons imino (NH) et amino (NH_2). La deuxième étape est l'attribution des protons non échangeables, notamment les protons du ribose ($\text{H1}'$, $\text{H2}'$, $\text{H3}'$, $\text{H4}'$, $\text{H5}'$ et $\text{H5}''$), les protons aromatiques des bases (H5 et H6 des pyrimidines et H8 des purines), puis le H2 des adénines qui n'échangent pas avec le solvant.

Attribution des protons échangeables

Le point de départ pour faire l'attribution des signaux provenant des protons échangeables est l'utilisation des protons imino, comme expliqué plus haut. En effet, la région entre 9 et 15 ppm d'un spectre 1D ^1H permet de vérifier le nombre de paires de bases de l'ARN. La stratégie utilisée pour faire l'attribution de ces signaux et identifier ces paires de bases est l'utilisation de l'effet *NOE* dans un type d'expérience en deux dimensions appelée 2D ^1H - ^1H *NOESY* (*NOE spectroscopy*, *NOESY*). Cette expérience homonucléaire effectuée sur un échantillon contenant un tampon composé de 90% H_2O permet de faire la corrélation à travers l'espace entre des protons qui sont à une distance de 5 Å ou moins. Cette expérience permet d'identifier les paires de bases A-U grâce à un fort signal *NOE* causé par la proximité du H3 de l'uridine avec le H2 de l'adénine à laquelle il est apparié (Figure 1.30 A) [222]. De façon similaire, les paires de bases G-C se caractérisent par un fort signal *NOE* entre le H1 de la guanine et les deux protons du groupement amino NH_2 de la cytosine à laquelle il est apparié (Figure 1.30 A) [222]. De plus, lorsque l'ARN adopte une structure en hélice de forme A, les protons imino des paires de bases adjacentes dans l'hélice sont situés à une distance variant de 3 à 5 Å, il est donc possible d'observer des signaux *NOE* entre ces protons (NH-NH, Figure 1.30

B, C) [222]. En plus de permettre l'identification des différentes paires de bases de la molécule d'ARN, ainsi que leur empilement dans les structures en hélice, les spectres homonucléaires 1D ^1H et 2D ^1H - ^1H *NOESY* des protons imino permettent non seulement l'attribution de plusieurs protons échangeables (GH1, UH3, CNH₂) mais également quelques protons non échangeables (AH2 et H1') [222].

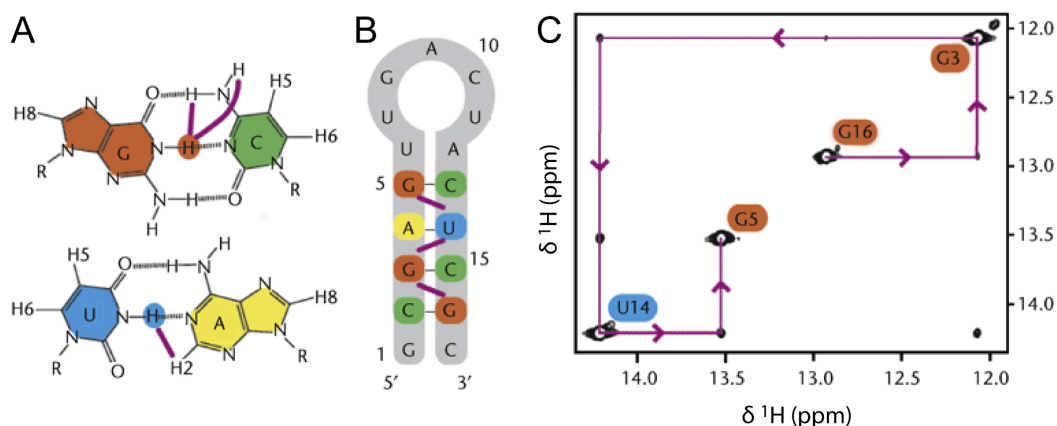


Figure 1.30 : Attribution des protons imino. (A) Protons imino des bases G et U (encerclés en orange et bleu respectivement) stabilisés par des ponts hydrogènes dans des paires de bases G-C et A-U. Lorsqu'impliqués dans ces paires de bases, les protons imino donnent de forts signaux *NOE* à d'autres protons (représentés par les traits roses). (B) Représentation schématique d'une tige-boucle et des connections *NOE* observées entre les protons imino (traits roses) de chaque paires de bases adjacentes dans la tige. (C) Région imino d'un spectre 2D *NOESY* montrant les connections entre les différents signaux *NOE* (flèches roses) provenant de chaque proton imino stabilisé dans une paire de bases. (Tiré de [223]).

Il est aussi particulièrement utile d'utiliser la RMN hétéonucléaire pour faciliter l'attribution des protons imino. L'utilisation d'une expérience de type *HSQC* (*heteronuclear single quantum coherence, HSQC*) permet de détecter les corrélations entre le proton et un autre type d'atome auquel il est lié, un hétéroatome tel que l'azote ou le carbone. Étant donné que cette expérience *HSQC* contient deux dimensions, par exemple 2D ^1H - ^{15}N *HSQC*, les protons sont identifiés grâce au déplacement chimique de l'azote, ce qui permet de différencier les H1 des guanines des H3 des uridines (Figure 1.26 C). D'autres expériences peuvent aussi être utilisées pour compléter l'attribution des protons, notamment le *HNN-COSY* [235] et le 3D ^{15}N -*edited NOESY-HSQC* [236].

Le HNN-COSY est une expérience en deux dimensions corrélant le ^1H et le ^{15}N et permet d'élucider les appariements de bases en confirmant les ponts hydrogène de type N-H...H [222, 235]. Par exemple, pour une paire de bases AU W-C, on retrouvera au déplacement chimique ^1H du UH3 deux corrélations à l'azote, une au UN3 et l'autre au AN1. L'expérience 3D ^{15}N -*edited NOESY-HSQC* contient la même information que le 2D *NOESY*, mais présente l'avantage d'utiliser la fréquence de l'azote pour faciliter l'attribution des protons, comme le 2D *HSQC*. Cette expérience permet ainsi d'analyser tous les *NOE* d'un proton imino, en particulier en utilisant le déplacement chimique de l'azote auquel il est attaché [222]. Un avantage additionnel de l'expérience 3D ^{15}N -*edited NOESY-HSQC* est qu'elle permet de filtrer les signaux (^{15}N -*edited*), générant ainsi des spectres où seulement les *NOE* donnés à un proton attaché à un azote marqués en ^{15}N sont visibles (imino (NH) et amino (NH₂)). De cette façon la troisième dimension (^{15}N) permet de résoudre significativement le problème de superposition de signaux.

Attribution des protons non échangeables

La deuxième étape dans l'attribution des signaux est ce qu'on appelle «la marche séquentielle» qui permet d'attribuer les protons non échangeables. La marche séquentielle permet de connecter chaque ribose à sa propre base et à celle du nucléotide situé immédiatement en 3' dans la séquence. Cette stratégie est basée sur les distances H1'-H6/H8 intra et internucléotide de l'hélice de forme A qui sont de l'ordre de 4 à 6 Å [222]. Dans une expérience de type *NOESY*, un signal *NOE* intranucléotide est ainsi observé entre le H1' d'un ribose et le proton aromatique de sa base, (n)H1' - (n)H6/H8 et un signal *NOE* internucléotide est observé entre le H1' du ribose et le proton aromatique du nucléotide situé en 3' dans la séquence, (n)H1' - (n+1)H6/H8 (Figure 1.31).

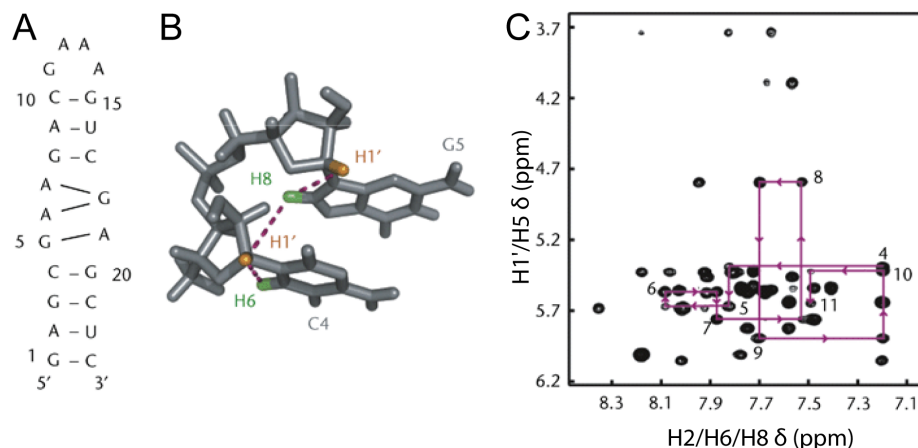


Figure 1.31 : Attribution des protons non échangeables par la marche séquentielle. (A) Séquence et structure secondaire de l'ARN utilisé dans l'étude RMN de la conformation active du substrat du ribozyme VS [112]. (B) Fragment de deux nucléotides (C4-G5) tiré de l'ARN en (A) montrant les distances H1'-H8/H6 intra et internucléotides utilisées pour la marche séquentielle (traits pointillés roses). (C) Région du spectre 2D *NOESY* de l'ARN représenté en (A) montrant les signaux *NOE* observés et leur connexions dans la marche séquentielle (flèches roses) (Tiré de [223]).

Pour procéder à la marche séquentielle, l'échantillon utilisé doit être dans un tampon composé de D₂O à 100% afin d'éliminer le signal du H₂O contenu dans le tampon. En effet, le déplacement chimique de celui-ci coïncide avec celui des protons H1' et H5 de l'ARN (~4.8 ppm). Pour faire l'attribution des autres protons non échangeables, c'est-à-dire les H5 des pyrimidines et les H2', H3', H4', H5', H5'' du ribose, des expériences RMN de corrélation appelées *COSY* et *TOCSY*, basée sur le couplage *J*, sont utilisées (*correlation spectroscopy*, *COSY* et *total correlation spectroscopy*, *TOCSY*) [222]. Plus le couplage est fort entre les deux noyaux, plus la corrélation à travers les liens qui les séparent est efficace. Par exemple, l'expérience homonucléaire de type 2D *COSY* permet d'attribuer le H2' à partir du H1' à l'aide du couplage *J* retrouvé dans les liaisons entre ces deux protons (³*J*_{H1'-H2'}). De façon similaire, il est possible d'attribuer le H3' à partir du H2' et ainsi de suite jusqu'au H5''. Il est aussi possible d'obtenir l'attribution de tous les protons du ribose à partir du H1' seulement. En effet, l'expérience de type *TOCSY* permet à la magnétisation de parcourir tout le ribose et ainsi à partir du H1' obtenir les signaux des H2', H3', H4', H5' et H5'' du même ribose. L'utilisation des expériences de

type *COSY* et *TOCSY* homonucléaires pour l'ARN est cependant limitée, notamment à cause de la faible constante de couplage $^3J_{H1'-H2'}$ de 1 Hz de la conformation 3'-endo du ribose, typiquement retrouvée dans les hélices de forme A. Par conséquent, les spectres obtenus ne permettent pas d'observer toutes les corrélations attendues. L'utilisation de la RMN homonucléaire pour l'attribution des protons non échangeables est donc utile pour des ARN de petite taille (< 15 nt). Toutefois, avec l'utilisation de la RMN de corrélation hétéronucléaire *HCCH-COSY* et *HCCH-TOCSY*, il devient possible d'observer les corrélations H1'-H2', puisque la magnétisation est transférée à partir du H1' jusqu'au H2' grâce aux couplages *J* entre les liaisons H-C et C-C du ribose (H1'-C1'-C2'-H2') qui sont plus fort. L'expérience de type *HSQC* est aussi utilisée pour l'attribution des différents protons non échangeables du ribose (H1', H2', H3', H4', H5' et H5'') et de la base (H2, H5, H6, H8). En effet, le 2D ^1H - ^{13}C *HSQC* permet d'identifier ces différents protons grâce au déplacement chimique du carbone (Figure 1.26 A, B). Par exemple, bien que les protons aromatiques H5 des pyrimidines et les H1' des riboses présentent un déplacement chimique très similaire, situés entre 4.4 et 6.5 ppm, il est possible de les séparer en utilisant le déplacement chimique des carbones auxquels ils sont attachés, qui eux sont bien séparés avec des δ de 95 à 105 ppm pour le C5 comparé à des δ de 89 à 95 ppm pour le C1'. Le couplage *J* hétéronucléaire permet également d'obtenir de nombreuses corrélations intrabases entre protons échangeables et non échangeables, facilitant l'attribution non ambiguë de plusieurs signaux [222].

De façon générale, l'attribution des signaux est grandement facilitée par l'utilisation de la RMN hétéronucléaire. En effet, elle permet une meilleure résolution des signaux en raison des corrélations ^1H - ^{13}C et ^1H - ^{15}N . Ainsi, l'utilisation de spectres hétéronucléaires multidimensionnels (2D, 3D ou 4D) permet de réduire la superposition de signaux. De plus, le couplage *J* hétéronucléaire permet des transferts de magnétisation plus efficaces par rapport au couplage *J* homonucléaire, ce qui facilite l'obtention des différentes corrélations nécessaires à l'attribution.

1.4.3.4 Obtention des contraintes de RMN

Une fois l'attribution des signaux complétée, une analyse détaillée de plusieurs expériences de RMN permet d'extraire des contraintes structurales dans le but de calculer la structure tridimensionnelle de l'ARN d'intérêt. Deux formes de contraintes sont initialement extraites : des distances et des angles de torsion [222]. Tout d'abord, une analyse de l'intensité des signaux *NOE* est effectuée. Chaque signal *NOE* est ensuite classé selon son intensité et ainsi associé à un intervalle de distances. En effet, tel que mentionné précédemment, l'intensité d'un signal *NOE* entre deux protons est inversement proportionnel à la distance qui sépare ces protons. D'autre part, les mesures de couplage J entre les différents atomes du ribose et du squelette ribose-phosphate sont importantes pour déterminer les angles de torsion (α , β , γ , δ , ε , ζ), la conformation des riboses en 2' endo ou 3' endo (*sugar pucker*), ainsi que l'angle glycosidique χ entre le ribose et la base (Figure 1.32 A, B). De plus, des expériences RMN basées sur le couplage J hétéronucléaire J_{NH} et J_{NN} de type HNN-COSY permettent d'obtenir une preuve directe de la présence de ponts hydrogènes dans l'ARN [235] (Figure 1.32 C). Ce type d'information fournit des contraintes de distances supplémentaires qui permettent de définir les paires de bases. Toutes ces contraintes de distances et d'angles définissent avec précision la structure locale de l'ARN. Dans le but de compléter ces contraintes locales et obtenir des contraintes à longue distance, des données de *RDC* sont utilisées. Tel que mentionné précédemment, ces données sont particulièrement intéressantes, notamment pour déterminer l'orientation des différents domaines hélicoïdaux au sein de la structure globale de l'ARN. Les *RDC* mesurés sur les liens N-H et C-H sont reliés à l'angle entre chaque lien et un référentiel. L'analyse de l'ensemble des *RDC* par rapport à ce même référentiel permet d'obtenir l'orientation entre chaque domaine de l'ARN.

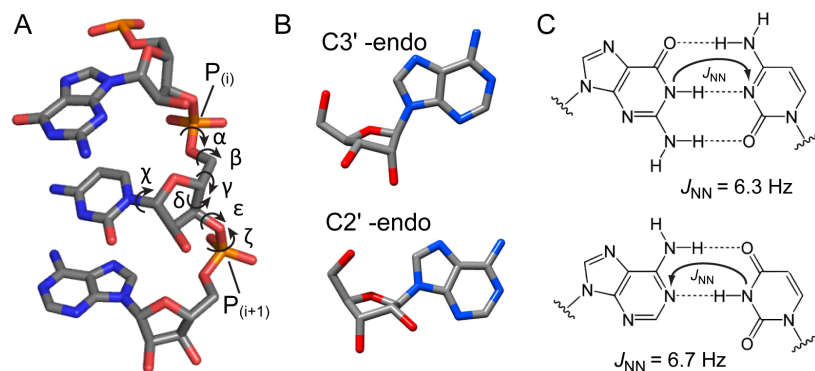


Figure 1.32 : Différents types de contraintes RMN. (A) Fragment de 3 nt dans lequel sont représentés des angles de torsion qui définissent la conformation du squelette ribose-phosphate et la conformation des riboses. (B) Les deux conformations du ribose retrouvées dans l'ARN, C3'-endo et C2'-endo (tiré du site web: <http://www.psc.edu/science/2012/inprog/>). (C) Couplage J_{NN} utilisé dans l'expérience de type HNN-COSY s'observant à travers un pont hydrogène.

1.4.3.5 Calcul de structure

Le calcul de structure est l'étape qui définit un ensemble de structures d'énergie minimale satisfaisant toutes les contraintes structurales obtenues par RMN. Le calcul s'effectue grâce à un protocole informatique basé sur une méthode appelée *simulated annealing* (SA) qui permet de simuler le repliement d'une molécule à partir de sa structure étendue en présence d'un champ de force et de contraintes de RMN [237]. Tout d'abord, la molécule d'ARN dans sa forme étendue est placée dans le champ de force (Figure 1.33), c'est-à-dire un ensemble d'équations représentant des fonctions d'énergie empirique et effective simulant les différentes forces qui s'appliquent sur tous les atomes d'un système moléculaire. Par exemple les fonctions d'énergie empirique représentent la structure covalente de l'ARN et correspondent notamment aux énergies des liaisons, des angles dièdres et des forces de Van Der Waals, alors que les fonctions d'énergie effective sont celles provenant des contraintes expérimentales, telles que les NOE obtenus par RMN [237]. Ensuite, de l'énergie est introduite dans le système sous forme de chaleur pour permettre aux atomes de bouger et d'explorer l'espace conformationnel. Durant le protocole de SA, les atomes se déplacent de manière à réduire l'énergie globale de la molécule [223], par exemple en essayant systématiquement de se positionner à une

distance optimale des atomes voisins pour éviter les superpositions de rayons de Van Der Waals (*clash*) (Figure 1.33). Le système est ensuite refroidi lentement pour permettre à l'énergie des structures adoptées d'atteindre un minimum [223]. Le protocole de calcul génère ainsi un ensemble de structures qui satisfont les contraintes de RMN fournies (Figure 1.33). En général, plus le nombre de contraintes extraites des données RMN est élevé, plus l'ensemble de structures généré par le protocole de calcul converge vers une structure précise. La précision de l'ensemble de structures est définie par la racine de la moyenne des déviations de distances interatomiques au carré (*root-mean-square deviation, rmsd*), qui est calculée pour chacune des structures par rapport à la structure moyenne. La valeur moyenne des *rmsd* permet d'évaluer la précision de l'ensemble de structures calculées.

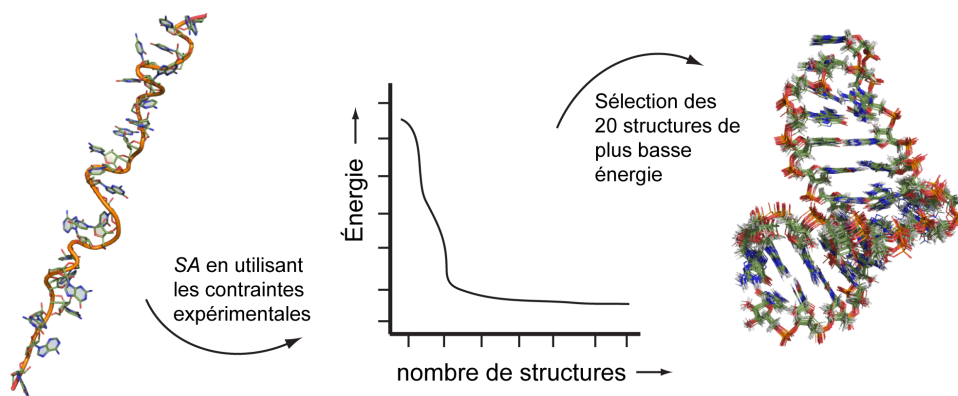


Figure 1.33 : Représentation des étapes générales du calcul de structure. La procédure de calcul débute avec une molécule de structure étendue. Le protocole de *simulated annealing* (*SA*) va simuler le repliement de cette molécule en considérant les contraintes expérimentales de distances et d'angles. Les structures de plus basses énergies seront sélectionnées et superposées pour former un ensemble représentatif de la molécule étudiée. (Adapté de [226]).

1.5 Étude d'interactions ARN/ARN par titrage calorimétrique isotherme

Le titrage calorimétrique isotherme (*isothermal titration calorimetry, ITC*) est une méthode qui permet d'étudier l'interaction entre deux molécules pour en définir les paramètres thermodynamiques. La formation d'une interaction entre deux molécules est

généralement associée à un dégagement (réaction exothermique) ou une absorption (réaction endothermique) de chaleur résultant notamment de la formation de nouvelles interactions, telles que des ponts hydrogène et des empilements. En utilisant l'*ITC*, il est possible de mesurer les changements de chaleur du système moléculaire tout au long d'un titrage [238]. Ces mesures s'effectuent à l'aide d'un microcalorimètre contenant deux cellules, dont l'une d'elle est la cellule expérimentale qui est reliée à une seringue d'injection (Figure 1.34 A). L'autre cellule sert de référence. La seringue permet de procéder au titrage en injectant des quantités précises d'une molécule (titrant) dans la cellule expérimentale où se trouve la molécule à titrer. Durant le titrage, chaque injection de titrant dans la cellule expérimentale est suivie d'un dégagement ou d'une absorption de chaleur suite à l'interaction entre les deux molécules, ce qui provoque un changement de température par rapport à la cellule de référence.

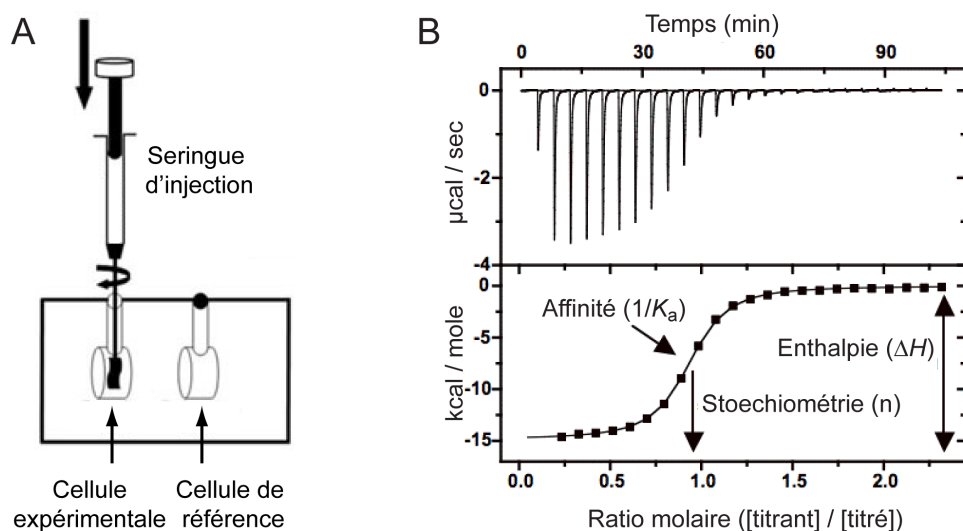


Figure 1.34 : Titrage calorimétrique isotherme (*ITC*). (A) Représentation schématique d'un microcalorimètre. La seringue d'injection contrôle les injections de titrant dans la cellule expérimentale tout en maintenant une agitation constante. (B) La partie supérieure montre le thermogramme obtenu suite à une expérience d'*ITC*, puis la partie inférieure montre la courbe de liaison isotherme obtenue après l'intégration de chaque pic d'injection du thermogramme. (Adapté avec permission de <http://www.ecu.edu/cs-cas/chem/Anne-Spuches-ITC.cfm>).

Les mesures enregistrées par le microcalorimètre correspondent à la puissance nécessaire à fournir à la cellule de référence pour maintenir la même température que

celle de la cellule expérimentale pour chaque ajout de titrant qui se fait à des temps précis. Le résultat brut d'une expérience d'*ITC* est représenté sous la forme d'un thermogramme, c'est-à-dire un graphique de la puissance en fonction du temps dans lequel il y a un pic de puissance pour chaque injections effectuées (Figure 1.34 B, graphique du haut). L'intensité des pics diminue en fonction du temps, puisque les sites de liaisons des molécules à titrer deviennent saturés. L'intégration de chaque pic d'injection correspond à l'énergie totale générée lors de la formation de l'interaction entre les deux molécules. La représentation de cette énergie en fonction du ratio molaire (concentration du titrant / concentration du titré) permet l'obtention d'une courbe de liaison isotherme (Figure 1.34 B, graphique du bas). L'analyse de cette courbe permet d'extraire trois paramètres thermodynamiques : la stœchiométrie (n), la constante d'association (K_A) et la différence d'enthalpie (ΔH) entre la forme libre et la forme liée (Figure 1.34 B, graphique du bas). À partir de ces trois paramètres, il est ensuite possible de compléter la caractérisation thermodynamique en dérivant la constante de dissociation (K_D), la différence d'entropie (ΔS) et d'énergie totale de Gibbs (ΔG) entre la forme libre et la forme liée à l'aide des équations de thermodynamiques suivantes : $\Delta G = \Delta H - T\Delta S$, $\Delta G = -RT\ln K_A$ et $K_D = 1/K_A$ où T est la température et R la constante des gaz parfaits.

Les paramètres thermodynamiques permettent de caractériser l'aspect énergétique de la formation d'une interaction. En effet, la constante de dissociation est directement reliée à l'énergie totale de Gibbs qui représente l'énergie impliquée dans l'interaction moléculaire. Cette énergie totale peut être séparée en deux termes principaux : l'enthalpie, correspondant notamment aux liaisons chimiques, et l'entropie qui réfère au niveau de désordre du système. En plus de déterminer l'affinité et l'énergie de l'interaction, il est aussi possible de définir si la formation de l'interaction est favorisée par la composante enthalpique ou entropique. De son côté, le paramètre de stœchiométrie correspond au nombre de sites de liaison de la molécule titrée pour son titrant. L'*ITC* permet la mesure directe de K_D variant de 10^{-9} à 10^{-2} M, alors que de manière indirecte, il est possible de déterminer des K_D allant de 10^{-12} à 10^{-2} M. En dehors

de ces valeurs, la qualité des courbes obtenues ne permet pas d'extraire tous les paramètres thermodynamiques.

L'avantage d'utiliser l'*ITC* pour faire l'étude d'une interaction est que cette méthode ne nécessite pas d'immobilisation, contrairement à la résonance plasmonique de surface (*surface plasmon resonance, SPR*), ni de marquage, contrairement à la méthode de retard sur gel (*electrophoretic mobility shift assay, EMSA*). L'*ITC* présente toutefois quelques désavantages : les expériences peuvent nécessiter une grande quantité de matériel, l'interaction étudiée doit être caractérisée par un $\Delta H \neq 0$ et il est important de s'assurer que les molécules utilisées adoptent un bon repliement homogène pour obtenir des paramètres thermodynamiques qui reflètent bien l'interaction étudiée.

En résumé, l'analyse de tous ces paramètres thermodynamiques permet de mieux comprendre la composante énergétique de la formation d'une interaction, ce qui est particulièrement intéressant pour faire le lien entre la structure et la fonction de l'ARN.

1.6 Objectifs de la thèse

L'importance fonctionnelle et structurale des interactions de type *kissing loop* a été mise en évidence pour plusieurs ARN, particulièrement chez les ARNt pour l'initiation de la traduction et chez plusieurs riborégulateurs et ribozymes pour leur repliement et la stabilisation de leur structure tridimensionnelle. D'ailleurs, une interaction *kissing-loop*, dépendante du magnésium, est présente chez le ribozyme VS de *Neurospora*. Celle-ci se forme entre la tige-boucle I (*stem-loop I, SLI*) du substrat et la tige-boucle V (*stem-loop V, SLV*) du domaine catalytique et est impliquée dans la reconnaissance et l'activation du substrat pour la catalyse.

Dans le but de mieux comprendre la reconnaissance du substrat par le ribozyme VS, j'ai entrepris l'étude de la structure et de la thermodynamique de l'interaction *kissing-loop* I/V. L'hypothèse globale de ces travaux est que l'interaction I/V du ribozyme VS doit impliquer des éléments structuraux plus complexes que la simple

formation de trois paires de bases W-C à l'interface pour permettre une reconnaissance spécifique de son substrat. Pour répondre à cette hypothèse, j'ai défini, conjointement avec mes collaborateurs, trois objectifs spécifiques correspondant chacun à un chapitre :

1. Au chapitre 2, nous nous sommes tout d'abord intéressé à la boucle V. En effet, suite à l'obtention des structures RMN de la boucle V en absence (SLV^{free} [113]) et en présence de magnésium (SLV^{Mg} [114]), nous avons noté qu'avec l'ajout de magnésium, le motif *U-turn* de la boucle V devenait plus compact et que le nucléotide U700 subissait un important changement de conformation. Nous avons donc procédé à une étude visant à comprendre l'importance du nucléotide U700 et à identifier les éléments importants de la boucle V dans la reconnaissance du substrat. Ceci a été effectué en vérifiant l'activité catalytique de ribozymes en *trans* contenant différentes variations de la boucle V par cinétique enzymatique et en caractérisant la conformation de ces différentes boucles par spectroscopie RMN.
2. Au chapitre 3, nous avons examiné la structure de l'interaction I/V pour déterminer précisément les éléments déterminants dans la reconnaissance du substrat. En effet, en plus des six nucléotides impliqués dans les trois paires de bases W-C (G630-C699, U631-A698 et C632-G697), d'autres nucléotides des boucles I et V ont montré une importance pour l'activité du ribozyme, notamment les U628 et C629 de la boucle I et les U696 et U700 de la boucle V. Il a d'ailleurs été suggéré que les boucles I et V adoptent chacune un motif de type *U-turn* impliquant les U628 et U696. Aussi, il a été démontré que cette interaction *kissing-loop* est accompagnée d'un important réarrangement structural du substrat, l'activant pour la catalyse. Toutes ces informations, obtenues à partir de méthodes biochimiques, mettent en évidence la complexité de la reconnaissance et l'activation du substrat. Dans le but de mieux comprendre ces deux événements, ainsi que la relation entre les éléments structuraux importants et la fonction du ribozyme, une caractérisation structurale est nécessaire. Pour ce faire, nous avons déterminé la structure tridimensionnelle de l'interaction *kissing-loop* I/V en

utilisant un complexe formé des tige-boucles I et V (*SLI/SLV*) du ribozyme VS par spectroscopie RMN. De plus, l'utilisation de plusieurs complexes *SLI/SLV* contenant différents types de substrats *SLI* a permis d'obtenir une preuve structurale par RMN de son changement de conformation lors de la formation de l'interaction *kissing-loop*.

3. Au chapitre 4, nous avons effectué une étude de thermodynamique pour avoir un profil plus complet de l'interaction I/V, c'est-à-dire connaître le coût énergétique du réarrangement structural du substrat et la contribution énergétique de l'interaction I/V au sein du ribozyme. Celle-ci a été faite par *ITC* en utilisant différents complexes *SLI/SLV* et *SLI/Ribozyme*.

Les études structurales et thermodynamiques de l'interaction *kissing-loop* I/V apporteront de nouvelles données essentielles à la compréhension de la reconnaissance et de l'activation du substrat par le ribozyme VS. Ces informations faciliteront notamment l'élaboration de projets d'ingénierie du ribozyme VS où celui-ci pourrait cliver différents ARN cibles. Dans un contexte plus global, la caractérisation structurale et thermodynamique d'une interaction *kissing-loop* permettra de mieux comprendre la grande stabilité généralement associée à ce type d'interaction. De plus, de petites interactions ARN/ARN stables dont la structure et la thermodynamique sont bien caractérisées, représentent des pièces modulaires indépendantes intéressantes pour la conception de différents types d'outils dans des domaines tels que l'ingénierie et la nanotechnologie de l'ARN.

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

Mise en contexte

Il est bien connu que le ribozyme VS reconnaît son substrat par la formation d'une interaction *kissing-loop* dépendante du magnésium qui implique les boucles terminales I et V localisées respectivement dans le substrat et le domaine catalytique. Les structures des domaines I et V sous leur forme libre ont précédemment été caractérisées, révélant une boucle terminale désordonnée pour le domaine I [111] et une boucle terminale contenant un motif de type *U-turn* pour le domaine V [113, 114]. Deux structures du domaine V ont été obtenues par RMN : la première en absence (SLV^{free} , [113]) et la deuxième en présence de magnésium (SLV^{Mg} , [114]). Ces deux structures ont révélé qu'en présence de magnésium, le motif *U-turn* de la boucle V devient plus compacte qu'en absence de magnésium avec un changement de conformation important du nucléotide U700 [113, 114]. En effet, en absence de magnésium, le U700 est situé dans le sillon majeur, alors qu'en présence de magnésium, celui-ci est positionné au sillon mineur. Le nucléotide U700 se retrouve ainsi sur la face de la boucle V qui interagit avec le substrat. Cette observation nous a mené à poser l'hypothèse suivante : le nucléotide U700 est impliqué dans l'interaction *kissing-loop* I/V.

Nous avons ainsi réalisé une étude visant à mieux comprendre le rôle de la boucle V dans la reconnaissance du substrat. Étant donné que les nucléotides U696, G697, A698 et C699 ont déjà été associés au motif *U-turn*, ainsi qu'aux trois paires de bases W-C de l'interaction *kissing-loop*, nous nous sommes penchés sur les nucléotides dont le rôle n'avait pas été étudié précédemment, notamment celui du U700 et de la paire de bases qui ferme la boucle (U695-A701). Pour se faire, nous avons étudié plusieurs variantes de la boucle de SLV dont certaines contenaient une modification du nucléotide U700. Ce dernier a été remplacé par les trois autres nucléotides, A, G ou C. Nous avons également étudié d'autres variantes de la boucle V contenant, par exemple, une plus grande boucle ou encore une paire de bases différente pour fermer la boucle. Nous avons évalué l'impact de ces différentes variations sur l'activité du ribozyme à l'aide d'expériences de cinétique enzymatique. De plus, ces variantes de SLV ont été

caractérisées par RMN pour vérifier leur impact sur la structure de la boucle V, particulièrement au niveau du motif *U-turn*. Cette étude a permis de démontrer que remplacer la base du U700 par un A, G ou C n'affecte pas significativement l'activité du ribozyme, alors que sa délétion entraîne une grande diminution de celle-ci. Ceci suggère que le squelette ribose-phosphate du nucléotide 700, et non sa base, est important pour l'interaction *kissing-loop*. De plus, nous avons utilisé la modélisation par homologie pour obtenir un modèle de l'interaction *kissing-loop* I/V qui nous a ensuite permis de proposer de nouvelles interactions entre les domaines I et V. Cette étude a également permis de développer un protocole RMN simple, basé sur la détection des déplacements chimiques inhabituels du ^{31}P et du ^{15}N N7, pour identifier des motifs *U-turn*.

Dans le cadre de cette étude, j'ai réalisé la caractérisation RMN de tous les ARN représentant chacune des variantes de *SLV*. Ceci inclut premièrement la synthèse *in vitro* et la purification de tous ces ARN. J'ai ensuite procédé à l'enregistrement de tous les spectres RMN du proton, du phosphore et de l'azote de chaque variante de *SLV* en absence et en présence de magnésium. J'ai ensuite analysé et attribué chacun des spectres RMN. Les expériences de cinétique enzymatique ont été effectuées par Julie Lacroix-Labonté, étudiante au Ph.D et co-auteure de cette étude. La modélisation a été effectuée par le Dr Pascale Legault, directrice du laboratoire, et Geneviève Desjardins, étudiante à la maîtrise, avec l'aide de Véronique Lisi, étudiante au Ph.D dans le laboratoire du Dr François Major. L'analyse des résultats s'est effectuée conjointement avec le Dr Pascale Legault. Pour la rédaction du manuscrit, j'ai rédigé une partie du matériels et méthodes, fait les figures de la partie RMN et le reste a été principalement rédigé par le Dr Legault.

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.

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. 2.1A; [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. 2.1A), and stem-loop I (SLI; Fig. 2.1B) 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. 2.1B), 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. 2.1B-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. 2.2), 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 non-canonical 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 non-canonical 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. 2.2) 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. 2.2; [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 manuscript, 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. 2.1A) or RNAs containing a variety of mutations in the SLV loop (Fig. 2.3). 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. 2.1B). The SLI(T) substrate (S) was selected for kinetic studies, because it can only adopt a preshifted active state (Fig. 2.1C; [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. 2.4. 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. 2.4A). 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. 2.4B. 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. 2.4C). Table 2.1 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 2.1). 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 ΔU700 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 $(k_{\text{cat}}/K_{\text{M}})_{\text{WT}}/(k_{\text{cat}}/K_{\text{M}})_{\text{MUT}}$ 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 $(k_{\text{cat}}/K_{\text{M}})_{\text{WT}}/(k_{\text{cat}}/K_{\text{M}})_{\text{MUT}}$ value of 74 was observed for the U695G/ ΔU700 mutant, a value similar to that of the ΔU700 mutant (Table 2.1), 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 $(k_{\text{cat}}/K_{\text{M}})_{\text{WT}}/(k_{\text{cat}}/K_{\text{M}})_{\text{MUT}}$ 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 $(k_{\text{cat}}/K_{\text{M}})_{\text{WT}}/(k_{\text{cat}}/K_{\text{M}})_{\text{MUT}}$ 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. 2.3) by NMR spectroscopy. In order to ascertain that SLV fragments formed a hairpin as in the folded ribozyme and not a duplex conformation (Fig. 2.5A), they were analyzed by 1D imino ¹H NMR spectroscopy (not shown) and native gel electrophoresis (Fig. 2.5B). 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. 2.5B, 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. 2.5) 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 2.1). 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 non-canonical 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 ³¹P resonance of U696 3'-phosphate (3'-P) and the ¹⁵N resonance of A698 N7 [15, 16]. The detection of unusual ³¹P and ¹⁵N 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. 2.3).

NMR structural probing of the wild-type and mutant SLV RNAs involved the collection and analysis of 1D ¹H-decoupled ³¹P spectra and 2D ¹H-¹⁵N long-range HMQC spectra [32]. In the 1D ³¹P spectra of the wild-type SLV RNA, distinct downfield-shifted signals were observed in the presence but not the absence of Mg^{2+} (Fig. 2.6A). Based on previous assignment, the ³¹P 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. 2.2; [15, 16]). The ^{31}P NMR spectra of the C699G and ΔU700 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 (ΔU700) when compared to the U696 3'-P resonance of the wild-type RNA (Fig. 2.6A). A summary of the ^{31}P chemical shift data for the mutant RNAs is provided in Table 2.2. 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 2.2); they were typical of (+; U695G and U695G/ ΔU700), not as significantly shifted as (+/-; U700C, ΔU700 , 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. 2.6B). 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 ΔU700 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. 2.6B). A summary of the 2D ^1H - ^{15}N long-range HMQC data for the mutant RNAs is provided in Table 2.2. 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 2.2); 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 Mg²⁺, the extent of the N7 upfield shift [typical of (+), not as significantly shifted as (+/-), or very different from (-)] always matched the extent of the ³¹P downfield shift for any given mutant (Table 2.2). The fact that these independent observations correlate with each other supports the use of these ³¹P and ¹⁵N chemical shifts to examine the U-turn structure of SLV mutants. For the wild-type, U695G, and U695G/ Δ U700 RNAs, their unusual ³¹P and ¹⁵N N7 chemical shifts indicate the formation of a canonical U-turn structure similar to that of SLV^{Mg} (Fig. 2.2). For the +U^{U700} and C699G mutants, the absence of unusual ³¹P and ¹⁵N N7 chemical shifts indicates that their loop does not form a canonical U-turn. In addition, the formation of a loose non-canonical U-turn fold in the +U^{U700} and C699G mutant RNAs cannot be evaluated, because the loose non-canonical U-turn fold of SLV^{free} did not yield unusual ³¹P and ¹⁵N N7 chemical shifts [15]. For the U700C, Δ U700, and U695C/A701G mutants, the extents of ³¹P and ¹⁵N N7 shifts in the presence of Mg²⁺ 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. 2.2).

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 non-canonical 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 non-canonical U-turns [15]. In Watson-Crick base pairs, the C1'-C1' distance is larger than that of many non-canonical 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_{cat}/K_M)_{WT}/(k_{cat}/K_M)_{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_{cat}/K_M)_{WT}/(k_{cat}/K_M)_{MUT} = 74], similar to its effect in the wild-type context. It therefore appears that formation of a canonical structure in this Δ 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. 2.7 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. 2.7). 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. 2.2).

DISCUSSION

Substrate recognition by the VS ribozyme involves an important and well-characterized Mg²⁺-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 Mg²⁺ 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 non-canonical 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 ($(k_{\text{cat}}/K_M)_{\text{WT}}/(k_{\text{cat}}/K_M)_{\text{MUT}} \leq 4$), U700C and +U^UU700 for example, did not adopt a compact canonical U-turn structure like the wild-type RNA. These mutants likely adopt a loose non-canonical 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 non-canonical 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^UU700 mutant) caused a minor decrease in k_{cat}/K_M , however, addition of a dinucleotide (+UC^UU700) 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^ΔU700, 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 non-canonical 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 phosphorothiate 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. 2.2; [16]). Four Mg^{2+} -binding sites were recently identified in the SLV loop structure (Fig. 2.2) 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 [7]. The stem-loop V mutants (Fig. 2.3) of Rz were prepared by using the QuikChangeII 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. 2.1). The wild-type and mutant sequences of Rz (Fig. 2.3) were prepared by *in vitro* T7 RNA polymerase transcription from plasmid A-3 [7] 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'-GAG CGA AGG GCA UCG UCG GCC CGA GCU C- 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 Aval ribozyme (Rz), either the wild-type (WT) or SLV loop mutants (Fig. 2.3), 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 \sim 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 $[\text{Rz}]$ 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 ¹⁵N-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. 2.4A) and nine others derived from this RNA, but containing selected loop mutations (Fig. 2.3). These SLV RNA fragments were synthesized in vitro using T7 RNA polymerase, a single-stranded DNA template, and ¹⁵N-labeled NTPs, and they were purified as previously described [15]. The final concentration of the ¹⁵N-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 d₁₁-Tris pH 7.0, 50 mM NaCl, and 0.05 mM NaN₃), NMR buffer B (10 mM d₁₁-Tris pH 7.0 and 0.05 mM NaN₃), or NMR buffer C (10 mM d₁₁-Tris pH 7.0, 20 mM MgCl₂, and 0.05 mM NaN₃), which contained 90% H₂O and 10% D₂O. For studies in D₂O, the RNA samples were transferred to 99.996% D₂O with multiple cycles of lyophilisation and resuspension in D₂O. 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 (DSS) at 0.00 ppm [41] 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 MCSym 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 XPLOR-NIH version 2.0.6 [42]. 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). JLL holds a Ph.D. scholarship from the Université de Montréal. P.B. holds Ph.D. scholarships from CIHR and the Université de Montréal. GD holds scholarships from the Fonds de la Recherche en Nature et Technologies (FQRNT) and from the Luigi-Liberatore Foundation. P. Legault 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 USA*, 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---C α H 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. Wishart, D.S., et al., *^1H , ^{13}C , ^{15}N chemical shift referencing in biomolecular NMR*. J Biomol NMR, 1995. **6**: p. 135-140.
42. Schwieters, C.D., et al., *The Xplor-NIH NMR molecular structure determination package*. J Magn Reson, 2003. **160**: p. 66-74.

TABLES

Table 2.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 2.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. ^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.
ΔU700	-	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/ΔU700	-	Yes	-	0.32 to -0.85	-	> 228.2
	+	Yes	+	1.49	+	224.1
+U ^Δ U700	-	Yes	-	-0.03 to -0.86	-	> 229.8
	+	Yes	-	-0.42 to -1.21	-	> 228.9
+UC ^Δ 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 with 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.

^d n.d.: not determined.

FIGURES

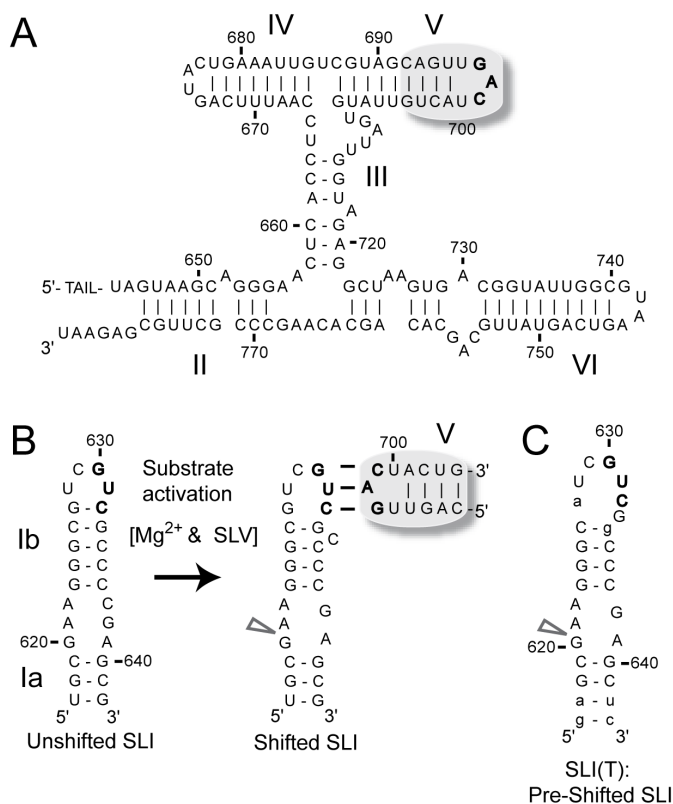


Figure 2.1 : Substrate recognition by the *Neurospora* VS ribozyme. A) Primary and secondary structures of the catalytic domain of the VS ribozyme. For this study, the Aval ribozyme was used; it includes a 5'-TAIL with the sequence: 5'-gggaaagcuGCGG-3' [7]. B) The catalytic domain recognizes the SLI substrate (subdivided in Ia and Ib) via a loop / loop interaction involving 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.

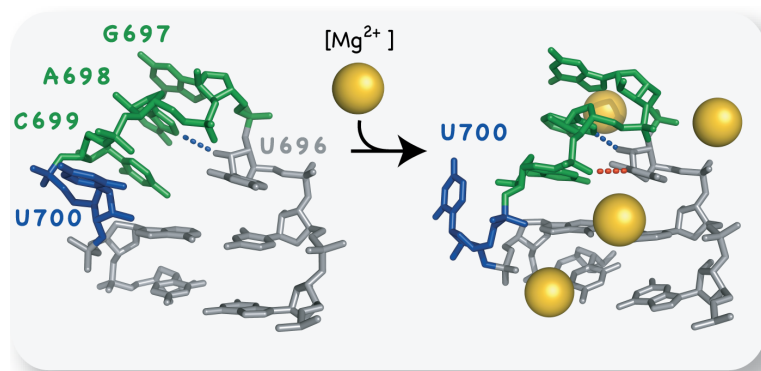


Figure 2.2 : NMR structures of the SLV loop determined in the absence (left) and presence (right) of magnesium ions [15, 16], showing the large conformational change of U700 (blue). The dotted lines represent two hydrogen bonds that are characteristic of canonical U-turn structures, the A698 N7 to U696 2'OH (blue) and the U696 H3 to A698 3'-phosphate interactions (red). For simplicity, only heavy atoms are shown.

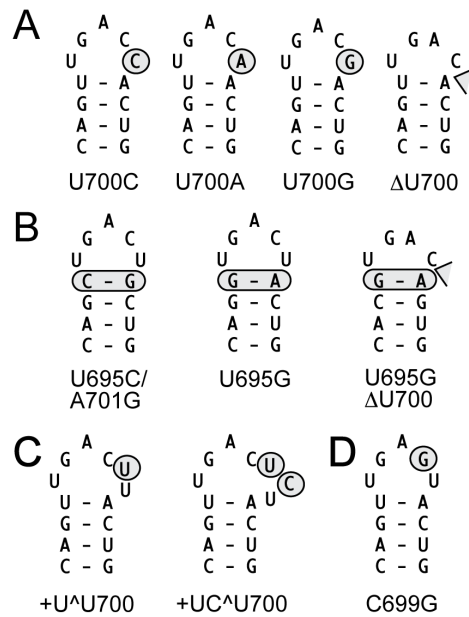


Figure 2.3 : SLV loop mutants characterized kinetically and structurally in this study.

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.

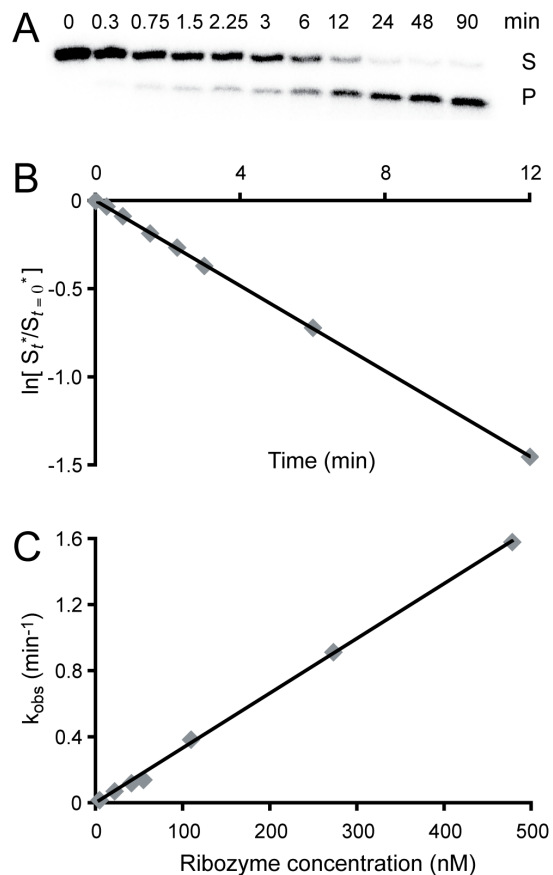


Figure 2.4 : Kinetic analysis of the U700G Rz mutant. A) Cleavage of ^{32}P -labeled SLI(T) by the U700G Rz is performed at $30\text{ }^{\circ}\text{C}$ in the presence of 25 mM MgCl_2 under single turnover conditions [excess enzyme (37.5 nM) over substrate (0.25 nM)]. Denaturing polyacrylamide gel electrophoresis 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 ($k_{\text{obs}} = \text{absolute value of the slope} = 0.12\text{ min}^{-1}$ 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 $k_{\text{cat}}/K_{\text{M}}$ is extracted by linear regression ($k_{\text{cat}}/K_{\text{M}} = \text{value of the slope} = 3.3\text{ min}^{-1}\mu\text{M}^{-1}$ and $R^2 = 0.9988$).

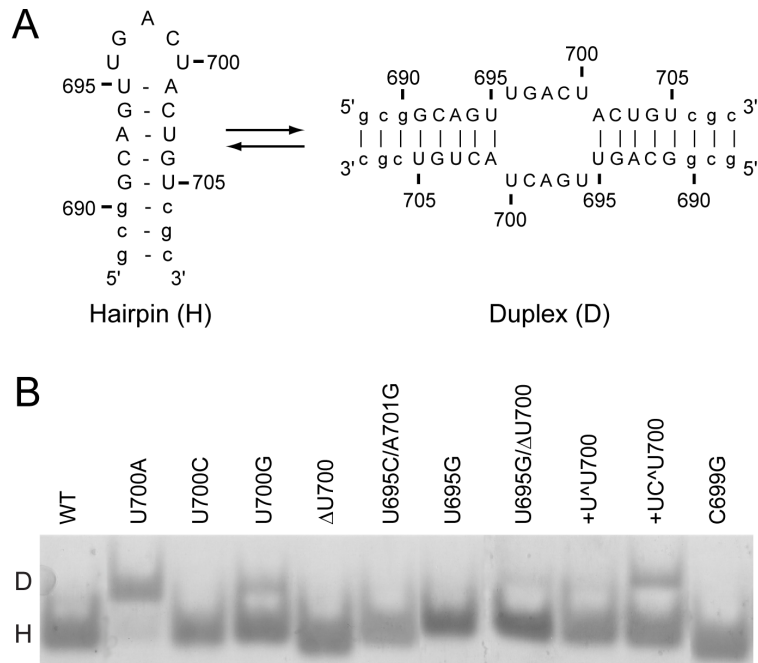


Figure 2.5 : Conformation of SLV RNA fragments.

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 $MgCl_2$.

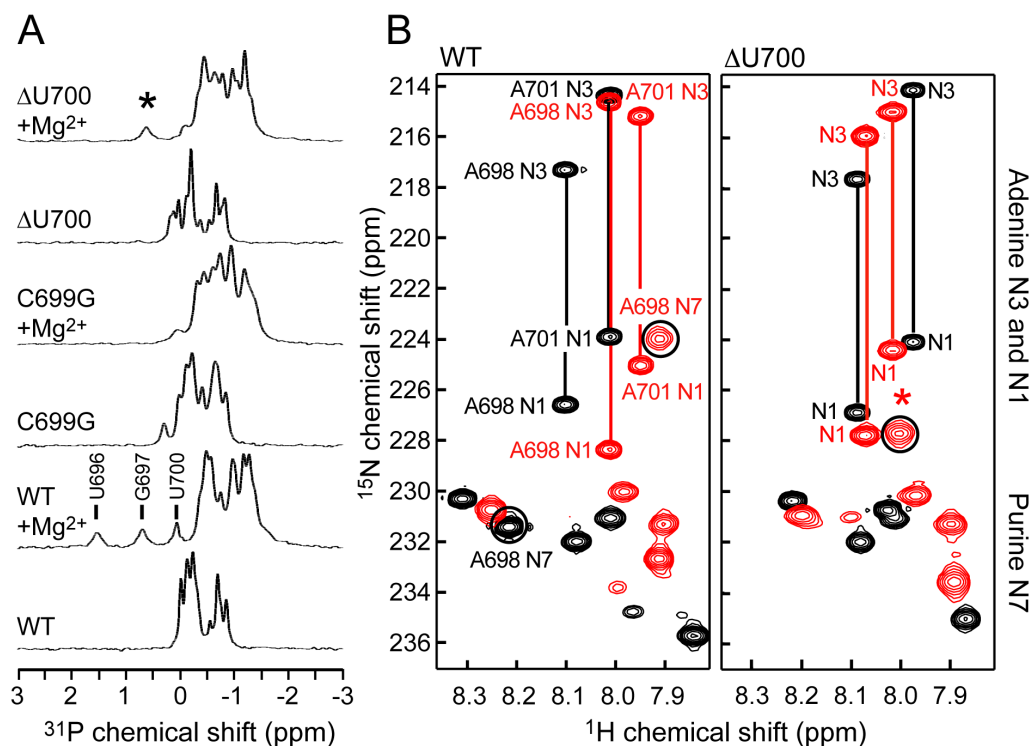


Figure 2.6 : NMR studies of SLV loop mutants.

A) 1D ^1H -decoupled ^{31}P spectra of wild-type (WT), C699G, and ΔU700 SLV RNA fragments in the absence and presence of 20 mM MgCl_2 ($+\text{Mg}^{2+}$). 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+} .

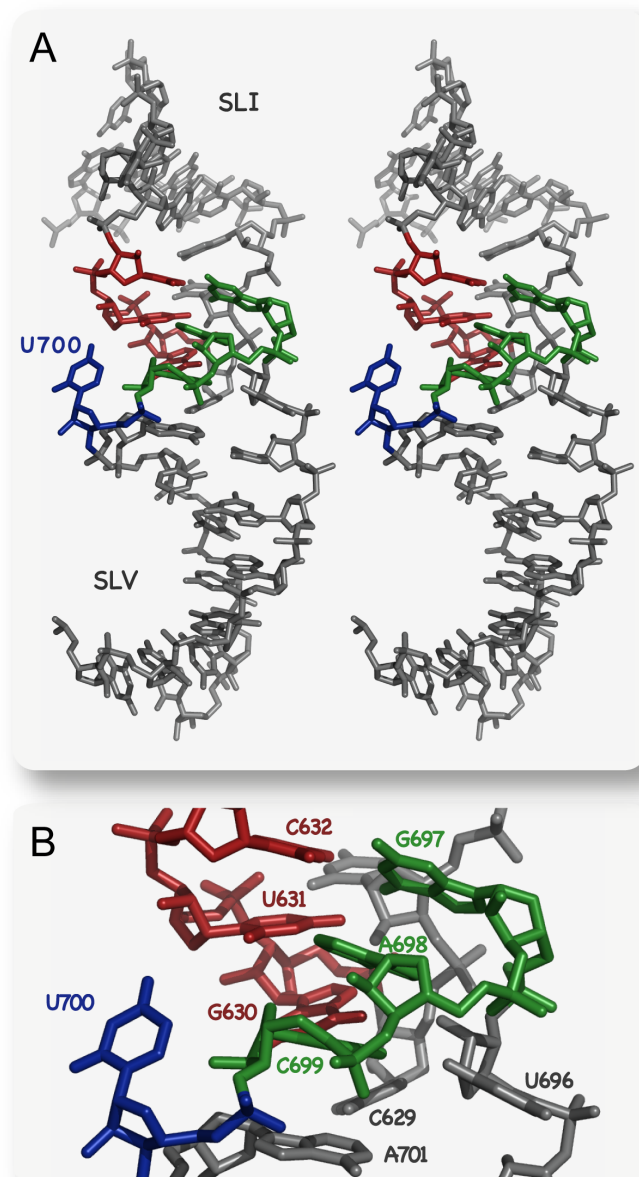


Figure 2.7 : Model of the SLI/SLV interaction. A) Stereo view of a structural model of the SLI/SLV complex. The selected model is a representative of the 7 structures obtained from homology modeling with MC-Sym. For simplicity, only heavy atoms are shown. B) Close up of the model shown in A.

3 Caractérisation de l'activation du substrat et de l'interaction *kissing-loop* I/V par spectroscopie RMN

Mise en contexte

Le ribozyme VS adopte une conformation native lui permettant de lier son substrat et d'en catalyser le clivage. La liaison au substrat est caractérisée par une interaction tertiaire de type *kissing-loop* entre les domaines I et V. En plus des six nucléotides impliqués dans cette interaction (G630-C699, U631-A698 et C632-G697), d'autres nucléotides des boucles I et V ont montré une importance pour l'activité du ribozyme, notamment les U628 et C629 dans le domaine I et les U696 et U700 dans le domaine V [82, 109]. En effet, il a été suggéré que les boucles I et V adoptent chacune un motif de type *U-turn* impliquant les U628 et U696 [82]. Nous avons donc émis l'hypothèse que l'interaction I/V est plus complexe que les trois paires de bases W-C décrites par les expériences biochimiques précédentes. En plus de permettre la reconnaissance du substrat, l'interaction *kissing-loop* I/V est responsable d'un important réarrangement structural permettant au substrat d'adopter une conformation active [90, 107, 109]. Ce réarrangement résulte en un *helix shift* dans la tige Ib et une réorganisation dans la boucle interne contenant le site de clivage. Ce réarrangement a d'ailleurs été démontré par des expériences biochimiques mais, jusqu'à maintenant, il n'y a pas de preuve structurale d'un tel changement de conformation. Bien que des structures à haute résolution des domaines I et V seuls ainsi que plusieurs modèles à basse résolution du ribozyme entier soient disponibles pour visualiser certains domaines ou le ribozyme de façon globale, ces données ne sont pas suffisantes pour comprendre la reconnaissance du substrat. En effet, ces données ne définissent pas les détails structuraux de l'interaction *kissing-loop* I/V. De ce fait, obtenir une structure à haute résolution de cette interaction pourrait permettre de mieux comprendre l'activation et la reconnaissance du substrat chez le ribozyme VS.

En utilisant la spectroscopie RMN, nous avons tout d'abord démontré le *helix shift* dans la tige Ib de substrat *shiftable*, puis avons résolu la structure de l'interaction *kissing-loop* I/V. Pour ce faire, nous avons utilisé des ARN représentant le substrat *SLI* (*stem-loop I*, *SLI*) et le domaine V *SLV* (*stem-loop V*, *SLV*) isolés, puisqu'il a été

démontré que le domaine *SLV* seul est suffisant pour activer et reconnaître le substrat *SLI*. Des données RMN ont été enregistrées sur différents complexes *SLI/SLV* composés d'un même *SLV* et d'un *SLI* en conformation *shiftable* ou *preshifted*. Ces données ont confirmé la présence des trois paires de bases W-C à l'interface I/V de chaque complexe et ont montré le *helix shift* dans les complexes *SLI/SLV* contenant un *SLI shiftable*. La structure RMN de l'interaction I/V a démontré que les boucles de *SLI* et *SLV* adoptent chacune un motif *U-turn* facilitant la formation des trois paires de bases intermoléculaires. De plus, la structure RMN a montré des interactions supplémentaires, notamment des interactions triples et des empilements de bases localisés de part et d'autre des trois paires de bases centrales. Globalement, cette étude RMN apporte une base structurale permettant de comprendre la stabilité de l'interaction *kissing-loop* I/V et également de proposer un modèle cinétique pour expliquer l'activation du substrat.

Pour ce deuxième article, j'ai essentiellement effectué l'ensemble des travaux, c'est-à-dire la synthèse et la purification des ARN, l'enregistrement et l'analyse des spectres RMN, le calcul et l'analyse des structures. J'ai bénéficié de l'aide de plusieurs membres de l'équipe du laboratoire pour la réalisation de cette étude : Élyse Benoit et Dominique Chaussé, techniciennes de laboratoire, ont fait la synthèse et la purification de plusieurs ARN. Pierre Dagenais, étudiant au Ph.D, et Dominique Chaussé ont fourni des nucléotides marqués en $^{13}\text{C}/^{15}\text{N}$ pour la préparation de certains ARN. Ryan Richter, agent de recherche en informatique, et Nicolas Girard, étudiant au Ph.D, ont contribué à l'optimisation de l'analyse de mes données RMN en créant différents programmes (*script*) d'analyse. De plus, Dr Pascale Legault, directrice du laboratoire, m'a accompagnée dans l'enregistrement des spectres RMN et dans le calcul de structure. L'analyse des résultats a été effectuée conjointement avec le Dr Legault. J'ai écrit la première version de l'article qui a ensuite été révisée par le Dr Legault.

Structural Insights Into Substrate Recognition by the *Neurospora Varkud* Satellite Ribozyme: Importance of U-Turns at the Kissing-Loop Junction

Patricia Bouchard and Pascale Legault

Département de Biochimie et Médecine Moléculaire, Université de Montréal, C.P. 6128,
Succursale Centre-Ville, Montréal, Quebec H3C1J7, Canada

BMRB entry code: 19662

PDB entry code: 2MI0

Reproduced with permission from *Biochemistry*, **2014**, 53 (1), pp 258–269, © 2014 American Chemical Society.

ABSTRACT

Substrate recognition by the *Neurospora* Varkud Satellite ribozyme depends on the formation of a magnesium-dependent kissing-loop interaction between the stem-loop I (SLI) substrate and stem-loop V (SLV) of the catalytic domain. From mutagenesis studies, it has been established that this I/V kissing-loop interaction involves three Watson-Crick base pairs and is associated with a structural rearrangement of the SLI substrate that facilitates catalysis. Here, we report the NMR structural characterization of this I/V kissing-loop using isolated stem-loops. NMR studies were performed on different SLI/SLV complexes containing a common SLV and shiftable, preshifted or double-stranded SLI variants. These studies confirm the presence of three Watson-Crick base pairs at the kissing-loop junction and provide evidence for the structural rearrangement of shiftable SLI variants upon SLV binding. NMR structure determination of an SLI/SLV complex demonstrates that both the SLI and SLV loops adopt U-turn structures, which facilitates intermolecular Watson-Crick base pairing. Several other interactions at the I/V interface, including base triples and base stacking, help create a continuously stacked structure. These NMR studies provide a structural basis to understand the stability of the I/V kissing-loop interaction and lead us to propose a kinetic model for substrate activation in the VS ribozyme.

INTRODUCTION

The U-turn fold is widespread in RNA structures. It was first identified in the anticodon hairpin loop of yeast phenylalanine tRNA [1] and later observed in several hairpin loops, internal loops and helical junctions of a variety of functional RNAs [2-11]. The UNR sequence (U = uracil, N = any base and R = purine) typical of this fold facilitates the formation of a sharp backbone turn at the UN step and promotes stacking of the bases after the turn in a favorable position for Watson-Crick base pairing. As a result, U-turn structures are often key contributors to RNA folding and tertiary interactions. In addition, U-turn structures are known to initiate rapid RNA/RNA pairing, as found for codon-anticodon recognition [12, 13] and antisense RNA targeting in prokaryotes [14], and such pairing is known to trigger conformational changes in macromolecular complexes involving RNA [14, 15]. However, despite the importance of U-turn structures in RNA function, few structural studies have thoroughly examined their contribution to RNA/RNA interactions and associated conformational changes.

The *Neurospora* Varkud Satellite (VS) ribozyme provides a simple model system for investigating the role of U-turn structures in an RNA/RNA interaction. The VS ribozyme is derived from the VS RNA, a circular single-stranded RNA satellite found in certain strains of *Neurospora* that depends on the cleavage and ligation activities of the VS ribozyme for its replication [16-18]. In vitro, the minimal region of the VS RNA that maintains catalytic activity contains six stem-loop domains (SLI-SLVI), with stem-loop I (SLI) corresponding to the substrate and stem-loops II to VI (SLII-SLVI) forming the catalytic domain (Figure 3.1A) [19, 20]. Of particular interest, a unique kissing-loop interaction between SLI and SLV (Figure 3.1A) plays a central role in substrate recognition and is very important for in vitro cleavage and ligation activities [21-26]. Mutagenesis and chemical probing data indicate that this I/V kissing-loop forms in the presence of magnesium (Mg^{2+}) ions and involves three Watson-Crick (W-C) base pairs [21, 23]. In addition, formation of the I/V kissing-loop is accompanied by a structural rearrangement in SLI, best described as a helix shift (Figure 3.1B), that activates the

substrate for catalysis [27-29]. Interestingly, the I/V kissing-loop and associated helix shift can be reproduced biochemically using isolated stem-loops (SLI and SLV) [29].

High-resolution NMR studies of stem-loop domains derived from the VS ribozyme have provided initial structural insight into the I/V kissing-loop and the associated helix shift. NMR structures of the SLI cleavage site internal loop were determined that correspond to both the unshifted and shifted conformations [30-32] and help explain the structural changes that likely occur at the cleavage site as a result of the helix shift. NMR structures of an isolated SLV domain indicate that the SLV loop forms a U-turn structure [33, 34], as predicted from the loop sequence and mutational studies [20]. Furthermore, this U-turn structure becomes more compact in the presence of Mg^{2+} ions [33, 34]. Although a U-turn structure was also predicted in SLI [20], NMR studies of an isolated SLI domain indicate that the SLI loop is mostly disordered in both the absence and the presence of Mg^{2+} ions [30]. On the basis of these NMR structures and biochemical data, three-dimensional models of the I/V kissing-loop interaction were derived [25, 35]. In addition, a model of the SLI/SLV interaction was incorporated in a low-resolution solution structure of the full VS ribozyme obtained by small angle X-ray scattering (SAXS) [36]. However, these models of the SLI/SLV complex lack the high level of precision that can only be achieved through high-resolution structural studies.

In this manuscript, we perform NMR structural studies of SLI/SLV complexes to gain additional structural insights into substrate recognition and activation by the VS ribozyme. We obtain structural evidence for formation of three W-C base pairs at the kissing-loop junction and for the helix shift in SLI as a result of SLV binding. In addition, our NMR data provide novel insights into the structure of the I/V kissing-loop that help explain the stability of this interaction and allow us to propose a kinetic model of helix shifting in SLI.

EXPERIMENTAL PROCEDURES

Preparation of RNA for NMR studies.

All RNAs used in this study (SLV, SLIps, SLIps Δ Ia, SLIsb, SLIsb Δ Ia, SLIsb Δ Ia_{GU}, SLIds1, and SLIds2; Figure 3.1B,C) were synthesized in vitro using T7 RNA polymerase (prepared in house), single-stranded DNA templates (IDT), and NTPs (¹³C/¹⁵N- or ¹⁵N-NTPs prepared in house or unlabeled NTPs from Sigma-Aldrich), and they were purified as previously described [25, 33]. The purified RNAs were first exchanged in NMR buffer (10 mM Tris-*d*₁₁ at pH 7.0, 50 mM NaCl, and 0.05 mM NaN₃ in 90% H₂O 10% D₂O) by use of Amicon Ultra-4 3K ultrafiltration devices and then refolded by heating to 95°C for two min and then cooling in ice water. Prior to complex formation, the RNA samples were transferred to NMR buffer supplemented with either 5 mM or 20 mM MgCl₂ 99.995% (Sigma-Aldrich) using the ultrafiltration device. In the case of SLIsb Δ Ia_{GU}, 20 mM MgCl₂ was added directly to the sample without the ultrafiltration device to ensure that it maintained its hairpin conformation. Using these procedures, the SLI and SLV RNAs formed hairpin structures, because only this conformation and not the duplex conformation was detected by native gels (Supplementary Figure 3.1 of Supporting Information).

Formation of SLI/SLV complexes was generally achieved by titrating the isotopically-labeled stem-loop with the unlabeled stem-loop in ratios of 1:0, 1:0.25, 1:0.5, 1:0.75 and 1:1, and the resulting complex was concentrated to a final volume of either 350 or 500 μ L. For imino NMR studies of complex formation, the following complexes were prepared at concentrations of 0.8-1.4 mM: SLIsb Δ Ia/¹⁵N-SLV, ¹⁵N-SLIsb Δ Ia/SLV, ¹⁵N-SLIsb Δ Ia/¹⁵N-SLV, ¹⁵N-SLIsb Δ Ia_{GU}/SLV, SLIps/¹⁵N-SLV, ¹⁵N-SLIps Δ Ia/SLV, ¹⁵N-SLIds1/SLV and ¹⁵N-SLIds2/SLV. For NMR structure determination, the following complexes were prepared at concentrations of 1.0-2.2 mM in NMR buffer supplemented with 5 mM MgCl₂: SLIds2/¹⁵N-SLV, SLIds2/¹³C¹⁵N-SLV, ¹⁵N-SLIds2/SLV, ¹³C¹⁵N-SLIds2/SLV, ¹³C¹⁵N-SLIds1/SLV, ¹⁵N-SLIsb Δ Ia/¹⁵N-SLV, SLIsb Δ Ia/¹⁵N-SLV and SLIsb Δ Ia/¹³C¹⁵N-SLV. Moreover, two complexes were similarly prepared in which ¹³C/¹⁵N-labeling of SLIds2 was restricted to either guanines or to

guanines and cytosines: $^{13}\text{C}^{15}\text{N}$ G-SLIDs2/SLV and $^{13}\text{C}^{15}\text{N}$ CG-SLIDs2/SLV. For NMR studies in D_2O , the samples were lyophilized and resuspended in 99.996% D_2O a minimum of four times.

Native gel electrophoresis studies of RNA hairpins.

The RNA samples were prepared by heating and snap-cooling 5 μg of each RNA in 20 μL of buffer containing 10 mM Tris pH 7.0, 50 mM NaCl and 20 mM MgCl_2 . Native polyacrylamide gel electrophoresis was carried out on 7.5% polyacrylamide gels (37.5:1 acrylamide:bisacrylamide) and run in TB buffer (20 mM Tris borate at pH 8). The gels were prerun for 30 min at 200 V and 4°C, loaded with RNA samples (SLI and SLV variants) and then run for 2 h at 250 V and 4°C. The gels were stained with Stains All (Sigma-Aldrich).

NMR spectroscopy.

NMR data were acquired on Varian ^{Unity}INOVA 500 and 600 MHz NMR spectrometers equipped with a pulse-field gradient unit and actively shielded z-gradient probes, either a $^1\text{H}\{^{13}\text{C}/^{15}\text{N}\}$ triple-resonance probe (standard or cold probe) or a $^1\text{H}\{^{15}\text{N}-^{31}\text{P}\}$ indirect detection probe. For imino NMR studies of complex formation, NMR data were acquired at either 25°C (for complexes of SLIps, SLIsb Δ Ia, SLIps Δ Ia, SLIds1 and SLIds2 with SLV) or 5°C (for the SLIsb Δ Ia_{GU}/SLV complex). The following NMR spectra were collected: 1D ^{15}N -decoupled ^1H watergate [37] and its ^{15}N -filtered and ^{15}N -edited versions, imino-optimized 2D ^1H - ^{15}N HSQC [38], 2D HNN-COSY [39], and 2D $^{15}\text{N}(t_1/t_2)$ -decoupled ^1H - ^1H flip-back watergate NOESY (2D NOESY) [37, 40] with mixing times of 150 ms or 250 ms. For NMR structure determination, NMR data were collected at 25°C on SLIsb Δ Ia/SLV and SLIds1/SLV, SLIds2/SLV complexes. For assignment of exchangeable protons and their attached nitrogens, the following experiments were collected in H_2O : imino and amino-optimized 2D ^1H - ^{15}N HSQC [38]; G-specific 2D H(NC)-TOCSY-(C)H [41] (for SLIsb Δ Ia/ ^{15}N -SLV only); 2D HNN-COSY [39]; 2D ^1H - ^{15}N CPMG-NOESY [42] (for SLIsb Δ Ia/SLV and SLIds2/SLV only); 2D NOESY [37, 40]; 3D ^{15}N -edited NOESY-HSQC [43] and 3D $^{15}\text{N}/^{13}\text{C}$ -edited NOESY-HSQC [44] with a mixing time of 150 ms (for SLIds2/SLV

only). For assignment of nonexchangeable protons and their attached carbons, the following experiments were collected in D₂O: 2D ¹H-¹³C constant-time-HSQC (CT-HSQC) [45, 46]; 3D CT-HCCH-COSY [47]; 3D HCCH-TOCSY [47] (for SLIsbΔIa/SLV and SLIds2/SLV only) and 3D ¹³C-edited HMQC-NOESY [48] collected with mixing times of 70 and/or 140 ms. In addition, 2D long-range ¹H-¹⁵N HMQC [49] spectra optimized for transfers via $J=21$ Hz were collected in D₂O for the assignment of purines N7 and N9. NMR data were processed using the NMRPipe/NMRDraw package [50] and analyzed with the CCPNMR suite [51].

Structural restraints.

Structural restraints were collected for structure determination of the SLIds2/SLV complex. Two-dimensional HNN-COSY spectra were collected to detect ² J_{NN} couplings across hydrogen bonds in W-C base pairs. Distance restraints were extracted from 2D NOESY [37, 40], 2D ¹H-¹⁵N CPMG-NOESY [42], 3D ¹⁵N-edited NOESY-HSQC [43] and 3D ¹⁵N/¹³C-edited NOESY-HSQC [44] spectra recorded in H₂O and 3D ¹³C-edited HMQC-NOESY [48] spectra collected in D₂O. The NOE-derived distance restraints were separated in four classes: very strong (1.8-3.3 Å), strong (1.8-4.5 Å), medium (1.8-5.5 Å) and weak (2.8-7.5 Å). Based on NMR evidence for specific base pairing obtained from NOESY and HNN-COSY spectra, canonical distance restraints were employed to define the hydrogen bonds and planarity of standard W-C and G-U wobble base pairs. In addition, cross-strand interphosphate distance constraints were introduced for each of these base pairs (≥ 18 Å) to help define A-form geometry in the SLIds2 and SLV stems. Based on NMR data, all glycosidic angles adopt an anti conformation, which was defined using H1'-H6/H8 distance restraints of 3.2-3.6 Å for pyrimidines and 3.4-3.8 Å for purines. Based on NMR evidence, backbone torsion angles of residues in helical regions (1^I - 6^I , 16^I - 22^I in SLIds2 and 1^V - 7^V , 15^V - 21^V in SLV) were restrained to A-form values ($\pm 15^\circ$), whereas a selected few were also restrained to A-form values ($\pm 30^\circ$) in the kissing-loop region (α and β of 13^I and 12^V , γ and δ of 12^I , 13^I , 11^V , 12^V and ϵ and ζ of 12^I and 11^V). Based on the N7 chemical shift values of $G11^I$ and $A11^V$, which are characteristic of the U-turn

motif, two specific distance restraints of 1.8-3.0 Å were defined (G11^I N7 to U9^I 2'-OH and A11^V N7 to U9^V 2'-OH). Finally, based on the C7^I NH₂ chemical shifts, which are characteristics of G-C base pairs, an ambiguous distance restraint (1.8-2.2 Å) was defined between C7^I NH₂ and the O6 of either G14^I or G15^I.

Residual dipolar coupling measurements.

NMR studies were performed at 25°C with an ¹⁵N-SLIDs2/¹⁵N-SLV sample (0.2-0.3 mM RNA in NMR buffer supplemented with 5 mM MgCl₂) titrated with a concentrated solution of Pfl filamentous phage to yield Pfl phage concentrations of 0, 14.6, 17.2, 18.9 and 20.5 mg/mL [52]. For each of these samples, values of the ²H splitting of D₂O were measured from 1D ²H NMR spectra, and imino ¹H-¹⁵N splittings were measured from 2D ¹H-¹⁵N HSQC spectra [38] collected with active J_{NH} coupling in t_1 . The ¹H-¹⁵N splittings were measured from fitted peaks using NMRWish [50] in the NMRPipe/NMRDraw package. Residual dipolar coupling (RDC) values were calculated for H1-N1 (G) and H3-N3 (U) signals detected at the different Pfl phage concentrations. For each H-N signal, the RDC values were plotted with respect to Pfl phage concentration in order to obtain a linear fit from which a precise RDC value was determined at 17.2 mg phage/mL.

Structure calculation.

Three-dimensional structures were calculated with restrained molecular dynamics and simulated annealing in X-PLOR-NIH version 2.19 using a two-stage protocol, as previously described [53]. Briefly, an initial set of 50 structures was calculated at stage one from structures with randomized backbone angles. At stage two, the structures were refined with RDC restraints using a single alignment tensor. The axial and rhombic components of the alignment tensor were explored with MODULE 1.0 and a grid search procedure to obtain reliable values of D_a and R . The grid search was repeated using RDC values from the SLIDs2 stem (residues 1-6, 16-21) and the SLV stem (residues 2-7, 15-20), independently. All grid search calculations converged on similar values of D_a and R [values of D_a (NH) = 17.7 Hz and R = 0.25 were used]. A final set of 500 structures was calculated, from which the 20 lowest-energy structures

that satisfied the experimental restraints (no distance violation $> 0.2 \text{ \AA}$, no torsion angle violation $> 5^\circ$ and no RDC violation $> 0.1 \text{ Hz}$) were selected for analysis. These 20 lowest-energy structures were used to calculate an average structure that was minimized against NOE and dihedral restraints. All structures were visualized with PyMOL Molecular Graphics System, Version 1.3 Schrödinger, LLC and analyzed with PyMOL and Curves+. Interhelical parameters were calculated as previously described [35].

RESULTS

NMR evidence for formation of the I/V kissing-loop interaction.

It was previously established that the SLV RNA used in this study (Figure 3.1C) adopts a stable hairpin that is compatible with SLI/SLV complex formation [25]. Thus, we next focused on identifying suitable SLI RNAs for our investigations of SLI/SLV complexes. Several SLI variants (Figure 3.1C) were characterized by imino NMR spectroscopy and native gel electrophoresis (Supplementary Figure 3.1 of Supporting Information) to identify those that form a stable hairpin to be tested for complex formation. Initially, we used shiftable SLI substrates with a stable stem Ia (e.g. SLIsb in Figure 3.1B) as found in the wild-type VS ribozyme, but such SLI variants do not form a stable SLI/SLV complex (not shown), in agreement with biochemical studies showing that stem Ia hinders formation of the I/V kissing loop interaction [27, 54-56].

Next, we investigated complex formation with a shiftable SLI substrate that does not contain stem Ia (SLIsb Δ Ia; Figure 3.2A). The SLIsb Δ Ia variant also contains a G627A mutation (A7^I in SLIsb Δ Ia) compared to the natural sequence of the VS ribozyme (Figure 3.2A). This variant was selected because a similar SLI substrate with the natural G at this position did not form a stable homogeneous complex with SLV for NMR studies (data not shown). Importantly, the G627A mutation is compatible with both cleavage and ligation activities of the VS ribozyme [24, 25, 35], and it slightly increases the affinity of the SLI substrate for the VS ribozyme [24]. Formation of the SLI/SLV complex with SLIsb Δ Ia was monitored by 1D imino proton NMR (Figure 3.2B). In addition, imino proton assignments of SLIsb Δ Ia and SLV were obtained from

2D NOESY spectra of both their free (not shown) and bound forms (Figure 3.2C). Interestingly, upon addition of SLIsbΔIa to SLV, three new imino proton signals are observed (Figure 3.2B), including one for SLV (G H1 at 12.11 ppm) and two for SLIsbΔIa (G H1 at 12.92 ppm and U H3 at 14.14 ppm; Figure 3.2C), which indicates that complex formation stabilizes three W-C base pairs, one A-U and two G-C base pairs. Interestingly, two NOE signals involving these new imino protons are observed (Figure 3.2C), which support stacking of the A-U base pair between the two G-C base pairs (Figure 3.2A). Further evidence for intermolecular W-C base pairing was obtained from 2D HNN-COSY[39] spectra collected on three SLIsbΔIa/SLV complexes prepared with differential ¹⁵N labeling (Supplementary Figure 3.2 of the Supporting Information). These spectra confirm W-C base pairing in the SLIsbΔIa stem (¹⁵N-SLIsbΔIa/SLV complex), the SLV stem (SLIsbΔIa/¹⁵N-SLV complex) and at the kissing-loop junction (¹⁵N-SLIsbΔIa/¹⁵N-SLV complex). Taken together, the imino NMR data support a secondary structure model for the SLIsbΔIa/SLV complex (Figure 3.2A) that includes the three proposed W-C base pairs (G11^I-C12^V, U12^I-A11^V and C13^I-G10^V) at the kissing-loop junction.

NMR evidence for the helix shift in SLI upon interaction with SLV.

It was previously demonstrated biochemically that SLI undergoes a structural rearrangement from an unshifted to a shifted conformation upon interaction with SLV, and this helix shift activates the substrate for catalysis [28, 29]. In SLIsbΔIa, this helix shift should involve three guanines (G3^I, G4^I and G5^I) that, upon SLV binding change their base pairing partners in stem Ib, thereby excluding C14^I from the helix (Figure 3.2A). To provide NMR evidence for this structural change, the 1D imino ¹H NMR spectrum of SLIsbΔIa was compared to the 1D ¹⁵N-filtered imino ¹H spectrum of the SLIsbΔIa/¹⁵N-SLV complex, which allows for selective detection of the SLI signals from the complex (Figure 3.3A,B). Four imino signals are observed for free SLI, in agreement with the four proposed G-C base pairs in the stem involving G3^I, G4^I, G5^I and G13^I. In contrast, five imino signals are observed for the SLV-bound SLI, two originating from U11^I and G10^I forming W-C base pairs at the kissing-loop junction and three from G3^I, G4^I and G5^I forming G-C base pairs in the SLI stem (Figure 3.3A,B).

The imino signal of G13^I, detected for free SLIsbΔIa, is absent in the spectrum of the SLIsbΔIa/SLV complex (Figure 3.3B), suggesting that SLV binding destabilizes the SLI loop closing base pair (C4^I-G13^I in Figure 3.3A). Importantly, there are significant chemical shift changes for the imino signals of G3^I and G5^I in the stem, which reflect important changes in their chemical environment upon complex formation (Figure 3.3B and Supplementary Table 3.1 of the Supporting Information). Taken together, changes in 1D ¹H imino spectra of free and SLV-bound SLI are compatible with the helix shift in SLI upon complex formation, although it does not provide evidence for the proposed C4^I-G13^I base pair closing the loop [28, 29].

To provide unambiguous evidence for base-pair shifting in the SLI stem, complex formation was investigated using a variant of SLIsbΔIa carrying a C15^I to U15^I change that transforms the G4^I-C15^I base pair of the free substrate to a G-U base pair (SLIsbΔIa_{GU}; Figure 3.3C). The imino NMR data of the SLIsbΔIa_{GU}/SLV complex confirm formation of the three W-C base pairs at the kissing-loop junction (Supplementary Figure 3.3 of the Supporting Information). Comparison of the imino ¹H spectra of SLIsbΔIa_{GU} in its free and SLV-bound forms is compatible with the helix shift in SLI (Figure 3.3CD). In particular, the imino protons of G13^I and G5^I observed in free SLIsbΔIa_{GU} are not observed for the SLV-bound SLIsbΔIa_{GU}, in agreement with destabilization of the loop-closing base pairs upon complex formation. Nevertheless, U15^I presents a large chemical shift change (from 11.69 to 10.91 ppm; Figure 3.3D and Supplementary Table 3.1 of the Supporting Information), consistent with formation of a G-U base pair closing the loop [57]. More importantly, NOE signals confirm that G4^I forms a G-U base pair in the free SLIsbΔIa_{GU} but forms a G-C base pair in the SLIsbΔIa_{GU}/SLV complex (Supplementary Figure 3.3 of the Supporting Information), providing definite NMR evidence for helix shifting in SLI as a result of complex formation.

Complex formation was also investigated with two preshifted SLI substrates, one with a stem Ia (SLIps) and one without a stem Ia (SLIpsΔIa; Figure 3.3E-H). The imino NMR data of the SLIps/SLV and the SLIpsΔIa/SLV complexes confirm formation of the

three W-C base pairs at the kissing-loop junction (Supplementary Figures 3.4 and 3.5 of the Supporting Information). For the stem residues of SLIps Δ Ia and SLIps, there are only minor changes in chemical shifts as a result of SLV binding (Figure 3.3E-H and Supplementary Table 3.1 of the Supporting Information), indicating that the structure of stem Ib is not largely affected by complex formation. However, we note the disappearance of G14¹ upon complex formation in both cases (Figure 3.3E-H), which is in agreement with destabilization of the SLI loop-closing base pair. In summary, imino NMR data obtained with shiftable versus preshifted substrates provide structural validation for helix shifting upon complex formation with the shiftable SLI substrates and the lack of such phenomenon with preshifted SLI substrates.

An ideal complex for NMR structure determination.

An ideal SLI/SLV complex for NMR structure determination should present a high affinity, a minimal size and stable stems associated with a large number of detectable imino protons. In this regard, we determined that the SLI variants considered thus far are suboptimal for NMR structure determination aimed at providing a high-resolution understanding of the I/V kissing-loop interaction. Thus, two additional small SLI variants were prepared, SLIds1 (Supplementary Figure 3.6 of the Supporting Information) and SLIds2 (Figure 3.4), that each contains a unique stem Ib that is extended compared to the wild-type SLI. Although these SLI variants are not substrates of the VS ribozyme, their extended stem Ib makes them particularly useful for NMR structure determination. Particularly, they should facilitate NMR resonance assignments and structure refinement using residual dipolar couplings (RDC). As observed with preshifted substrates, imino NMR studies indicate that there is little structural change in stem Ib of SLIds1 and SLIds2 upon complex formation with SLV (Figure 3.4 and Supplementary Figure 3.6 and Supplementary Table 3.1 of the Supporting Information). Of note, a weak G imino proton signal, assigned to G15¹ is observed in the 1D imino ¹H NMR spectra of free SLIds2, and this signal disappears completely upon complex formation (Figure 3.4B), in agreement with destabilization of the SLI loop-closing base pair. Furthermore, the SLIds1/SLV and SLIds2/SLV complexes both give high-quality imino NMR data, including the imino signals and NOE interconnectivities that are

characteristic of the I/V kissing-loop junction (Figure 3.4C and Supplementary Figure 3.6C of the Supporting Information). We pursued NMR structure determination with the SLIDs2/SLV complex, which provides superior imino proton dispersion for both SLI and SLV within a complex of reasonable size (43 nt) and of affinity similar to that of the two complexes formed with preshifted SLI substrates [$K_d = 0.4\text{-}0.7 \mu\text{M}$; (Bouchard, P and Legault, P., Département de Biochimie et Médecine Moléculaire, Université de Montréal, 2013, unpublished data)].

NMR structure determination of the SLIDs2/SLV complex.

Multidimensional heteronuclear NMR methods were employed to determine the structure of the SLIDs2/SLV complex using several combinations of unlabeled and uniformly-labeled (^{15}N or $^{13}\text{C}/^{15}\text{N}$) SLIDs2 and SLV RNAs. In addition, complexes incorporating ^{13}C and ^{15}N labeling of specific residue types within SLIDs2 and complexes containing different SLI variants (SLIsb Δ Ia/SLV and SLIDs1/SLV, Figure 3.1) were used to facilitate and confirm spectral analysis. The sequential resonance assignment mainly relied on 2D and 3D NOESY-type spectra of these complexes (see Experimental Procedures), allowing almost complete assignment of the observable resonances (^1H , ^{15}N and ^{13}C) from the bases and the C1'-H1' and C2'-H2' resonances from the ribose moieties (Supplementary Tables 3.3 and 3.4 of the Supporting Information).

Three-dimensional structures were calculated using distance, dihedral angle and RDC restraints (Figure 3.5 and Table 3.1). The overall structure is well defined by the NMR data, with a heavy-atom RMSD of $2.02 \pm 0.77 \text{ \AA}$ for the 20 lowest-energy structures (Figure 3.5 and Table 3.1). The stems of both SLI and SLV form well-defined A-form helices (RMSDs of 0.5-0.6 \AA). The SLV loop is also well defined (RMSD of $1.01 \pm 0.20 \text{ \AA}$), except for the extruded U13^V, whereas the SLI loop is not as well defined, mainly due to a limited number of restraints for residues A8^I, U9^I and G15^I (RMSD of $1.78 \pm 0.69 \text{ \AA}$). In terms of interhelical parameters (Supplementary Figure 3.7 of the Supporting Information), the majority of structures have an average interhelical

angle (θ_{I-V}) $\sim 130^\circ$ [$\theta_{I-V} = 131^\circ \pm 3^\circ$ with an axis displacement (d_{I-V}) = 15 ± 3 Å], although values $\sim 160^\circ$ are also observed in 6 of the 20 structures [$\theta_{I-V} = 161^\circ \pm 9^\circ$ with $d_{I-V} = 19 \pm 7$ Å].

Standard U-turns at the kissing-loop junction.

As previously proposed, both the SLI and SLV loops within the SLI/SLV complex adopt a U-turn structure (Figure 3.6) [21]. The consensus UNR sequence of the U-turn is represented in SLIDs2 by residues U9^I, C10^I and G11^I and in SLV by residues U9^V, G10^V and A11^V. These loop structures are characterized by a sharp turn in the ribose-phosphate backbone after the first U (U9^I or U9^V), which is supported by several NOEs for both SLIDs2 and SLV, namely NOEs between U9 H1'/H2' and R11 H8 and between U9 H1' and U/C12 H6/H5. As found in other U-turn structures, there are several base stacking interactions after the sharp turn, which are supported by sequential NOEs characteristic of A-form geometry. These stacking interactions involve C10^I, G11^I, U12^I, C13^I and G14^I in SLIDs2 and G10^V, A11^V, C12^V and A14^V in SLV (Figure 3.6).

U-turn structures are generally stabilized by two hydrogen bonds involving the required U residue: U 2'-OH to R N7 and U H3 to R 3'-phosphate [34]. For both the SLIDs2 and SLV loops, a downfield-shifted signal was observed for R N7 (G11^I at 228.5 ppm and A11^V at 221.5 ppm, Supplementary Table 3.2 of the Supporting Information) that has been previously associated with hydrogen bonding to U 2'-OH in the U-turn [25, 58]. On this basis, a distance restraint was defined between the U 2'-OH and R N7 (1.8-3.0 Å) that contribute to bring these residues within hydrogen-bonding distances in the ensemble of structures. The hydrogen bond between U H3 and R 3'-phosphate could not be similarly defined, nevertheless the distances of U N3 to R 3'-P in the ensemble of structures are between 3.88 and 6.19 Å for SLIDs2 and 4.02-5.65 Å for SLV, which is compatible with hydrogen bonding as observed in other U-turn structures [2, 5, 8, 15, 34, 59].

Both the SLIDs2 and SLV loops present an extruded nucleotide, G15^I in SLIDs2 and U13^V in SLV (Figure 3.6). The imino protons of G15^I and U13^V are not detected in

the 1D ^1H spectra of the complex and very few NOEs are observed with these bases as expected for an extruded conformation. For SLIDs2, the conformation of G15^I is not well defined, but this nucleotide is most often extruded from the loop fold and potentially forms a hydrogen bond between G15^I N7 and G14^I 2'-OH as well as perpendicular π stacking between G15^I and G14^I, in which either the NH₂ or the 2'-OH group of G14^I can form a hydrogen bond with the base of G15^I [60-62]. In contrast, the extruded U13^V of SLV is better defined with its W-C edge oriented toward the solvent in all 20 lowest-energy structures. Most of the U13^V signals in the 2D ^1H - ^{13}C CT-HSQC are observed at higher intensities and significantly different ^1H and ^{13}C chemical shifts compared to signals of other residues, which is indicative of fast dynamics for U13^V.

Previous NMR studies of the free SLI and SLV hairpins derived from the VS ribozymes have provided a wealth of structural information for the SLI and SLV loops in their free forms [30, 33, 34]. Although the free SLI loop is mostly disordered [30], the free SLV loop adopts a loose U-turn structure in the absence of Mg²⁺ ions (SLV^{free}) [33] that becomes more compact in the presence of Mg²⁺ ions (SLV^{Mg}) [34]. The superpositions of these SLV loops with the SLV loop from the SLIDs2/SLV complex demonstrate that in the complex, the structure of the SLV loop more closely matches the compact structure of the SLV^{Mg} loop (Supplementary Figure 3.8 of the Supporting Information). However, the base of U9^V is tilted in the complex such that it precludes its stacking with the 5'-phosphate of A11^V, as found in SLV^{Mg} and several U-turn structures [34]. In summary, upon formation of the I/V kissing-loop complex, loop I undergoes a significant structural change from a disordered state to a compact fold stabilized by a U-turn structure, whereas loop V undergoes only minor structural changes within its compact U-turn fold.

A kissing-loop Junction stabilized by base-pairing and stacking interactions.

U-turn structures typically expose the W-C edges of bases after the turn, making them available for base pairing. In the SLIDs2/SLV complex, these exposed bases from SLI and SLV interact with each other to create the kissing-loop junction (Figure 3.7A). The G11^I, U12^I and C13^I bases of SLIDs2 and C12^V, A11^V and G10^V bases of SLV

respectively form the three proposed W-C base pairs that stack on each other to adopt a short A-form-like helix that is well defined by several characteristic NOEs (Figures 3.3 and 3.7 and Supplementary Table 3.5 of the Supporting Information).

The kissing-loop junction is stabilized by additional hydrogen-bonding interactions, which contribute to formation of two base triples, one at each end of the kissing-loop helix (Figure 3.7). The C10^I-A14^V-U8^V base triple is observed in all of the 20 lowest-energy structures and involves a two hydrogen-bond base pair between the W-C edge of C10^I and the sugar edge of A14^V and a one hydrogen-bond pair between the W-C edges of U8^V and A14^V (Figure 3.7B). The C7^I-G14^I-G15^I base triple is defined by three hydrogen bonds observed simultaneously in the average structure (Figure 3.7C) and in 11 of the 20 lowest-energy structures: C7^I NH₂ to G14^I O6, C7^I N3 to G14^I H1 and G14^I 2'-OH to G15^I N7.

Stacking interactions also play an important role in stabilizing the kissing-loop interactions, since both C10^I and A14^V stack on the G11^I-C12^V base pair at one end of the kissing-loop helix, while at the other end both G14^I and G15^I can potentially stack on the C13^I-G10^V base pair (Figure 3.7D). Furthermore, the noncanonical A14^V-U8^V base pair is stacked on the W-C G7^V-C15^V base pair of stem V and the noncanonical G14^I-C7^I base pair is stacked on the W-C G6^I-C16^I base pair of stem I. Thus, stacking interactions at both ends of the kissing-loop helix propagate in both directions toward the stems of SLV and SLI to create a continuously stacked structure throughout the complex.

DISCUSSION

Structural characteristics of the kissing-loop interaction.

NMR studies of different SLI/SLV complexes performed in the presence of Mg²⁺ ions provide direct evidence that the SLI/SLV interaction of the VS ribozyme involves three consecutive W-C base pairs at the kissing-loop junction. In addition, the NMR structure of an optimal SLI/SLV complex formed with SLIDs2 and SLV clearly

demonstrates that the kissing-loop junction involves two U-turn structures, one in each loop, that enable formation of intermolecular W-C base pairs. In agreement with these NMR data, site-directed mutagenesis studies previously revealed that residues in the SLI and SLV loops must both fit the UNR consensus of U-turn motifs and accommodate W-C base pairing at the I/V kissing-loop junction [20, 21, 27]. Chemical modifications experiments are also in agreement with the observed U-turn structures in both SLI and SLV, in particular the two hydrogen bonds commonly found in U-turn structures: U 2'-OH to R N7 and U H3 to R 3'-phosphate [22, 27, 63-66].

In addition to the three W-C base pairs, other interactions contribute to the stability of the I/V kissing-loop interaction. The NMR structure of the SLI_{ids2}/SLV complex contains two previously uncharacterized base triples that provide additional hydrogen bonding and stacking interactions at the kissing-loop junction. The C10^I-A14^V-U8^V base triple (C629-A701-U695 in the VS ribozyme) is particularly well defined by the NMR data and helps rationalize the preference for a cytidine at the N position of the SLI U-turn [21]. Although several studies have highlighted the importance of C629 [20, 27-29] in the SLI/SLV interaction, its role had not been clearly established. As shown from the NMR structure and previous models of the SLI/SLV complex [25, 35], the U-turn fold in SLI brings C629 to stack with G630 in proximity of A701 [25]. A cytidine is likely optimal at this position because its small size prevents steric hindrance and its unique W-C edge allows hydrogen bonding with the sugar edge of A701 in SLV. Both deletion of C629 and 2'-deoxy modification at A701 significantly reduce the VS ribozyme self-cleavage and ligation activities [27, 63, 65], in agreement with the importance of these residues at the I/V kissing-loop interface. In contrast, the C7^I-G14^I-G15^I base triple (C626-G633-C634 in the VS ribozyme) is less well defined by the NMR data and the identity of the residues involved is not strictly required for cleavage and ligation activities [20, 21, 28].

Given the Mg²⁺-dependence for formation of the SLI/SLV complex [21, 23], it is likely that Mg²⁺ ions specifically associate with the kissing-loop junction. Previous phosphorothioate interference and manganese rescue experiments indicate that the 5'-

phosphates of U631 (SLI) and C699 (SLV) each coordinate a metal ion [22, 27, 63-66]. In addition, four Mg^{2+} -ion binding sites were previously identified within the free SLV loop [33, 34], and these may also be present in the SLI/SLV complex. In particular, the Mg^{2+} ion that associates with the 5'-phosphate of C699 in the free SLV is likely present in the complex. By analogy, a Mg^{2+} ion is likely associated with the 5'-phosphate of U631, which occupies a similar position in SLI as the 5'-phosphate of C699 in SLV. Additionally, Mg^{2+} ion(s) may help bridge phosphates from SLI and SLV that come in proximity at the kissing-loop junction, namely those of the two Us from the UNR motifs of SLI and SLV. Future studies are needed to precisely locate Mg^{2+} -ion binding sites at the I/V kissing-loop junction.

RNA loops containing U-turn structures often include an extruded nucleotide [4, 7, 9, 11], but the role of this residue is not always well understood. In the SLIDs2/SLV complex, both the SLI and SLV loops contain an extruded nucleotide. For SLV, the extruded U13^V (U700 in the VS ribozyme) can be substituted by any other standard base, but its phosphate backbone is important for cleavage activity and SLI binding [25]. U700 may play a dual role in the kissing-loop by providing a flexible linker to SLV as well as a Mg^{2+} -binding site, as previously proposed from kinetic studies of U700 variants [25]. For SLIDs2, the extruded G15^I (C634 in the VS ribozyme) can be substituted by any other standard base [20, 28] and its location is not precisely defined by the NMR data. Thus, C634 may be intrinsically dynamic in the VS ribozyme and thereby provide the necessary loop flexibility to optimize interactions at the kissing-loop interface. As discussed below, C634 likely plays a pivotal role in transducing part of the energy provided by formation of the kissing-loop interaction into the energy required for substrate activation.

Although the SLIDs2/SLV complex forms a continuously stacked structure that appears essentially linear, the stems of SLI and SLV are not coaxially stacked within the complex. Rather, the kissing-loop junction causes a helical displacement of 10-24 Å between stem I and stem V associated with an interhelical angle of 127-175°. The arrays of helical displacements and interhelical angles observed in the 20 lowest-energy

structures may reflect to some extent the relative range of motion of these stems. First of all, several residues in SLI (C7^I, A8^I, U9^I, C13^I, G14^I and G15^I) and SLV (U8^V and U9^V) give ¹H and ¹³C NMR signals of low relative intensity (data not shown), suggesting local dynamics at the kissing-loop junction. Moreover, it appears from the ensemble of structures that the observed range of interhelical angles may largely result from less-well defined local structure for residues C7^I-U9^I and C13^I-G15^I in SLI. Interestingly, we previously kinetically characterized SLI/ribozyme systems with helix-length variations in SLI and SLV and built SLI/SLV models compatible with the kinetic data that display helical displacements of 18-21 Å and interhelical angles of 165-171° [35]. The helical parameters of these models fall within the range detected in the ensemble of NMR structures [35], suggesting that within the spectrum of interhelical parameters observed for the SLI₂/SLV complex, some are compatible with an active VS ribozyme architecture. Similarly, the interhelical parameters previously reported for the SAXS model of the full VS ribozyme are more or less compatible with that of the SLI₂/SLV complex (helical displacement of ~25 Å and interhelical angle of ~154°) [35, 36].

Conformational changes in SLI and SLV upon formation of the kissing-loop interaction.

NMR studies of different SLI/SLV complexes also provide direct structural evidence that the SLI/SLV interaction involves a helix shift in stem Ib, as previously shown from biochemical experiments [28, 29]. As expected, this helix shift is observed only with shiftable SLI variants, but neither with preshifted variants nor with double-stranded variants. NMR studies of SLI/SLV complexes involving these SLI variants together with available NMR structures of the free SLI [30] and SLV [33, 34] allow us to examine structural changes that occur in SLI and SLV as a result of complex formation. Given that the formation of the SLI/SLV complex and the associated helix shift in SLI observed in the VS ribozyme can be reproduced with isolated hairpins [28, 29], it is likely that the structural changes observed with isolated stem-loops hold in the intact ribozyme. The SLV loop adopts a compact U-turn structure in its free form, and only a minor structural change is observed in the SLI/SLV complex [35]. In contrast,

SLI loop is mostly disordered in its free form [30], and this disordered state may help facilitate recognition by the compact SLV loop and trigger a structural transition in SLI upon interaction with SLV. Upon SLV binding, the SLI loop undergoes a disorder-to-order conformational transition that impacts on the loop closing base pairs of stem Ib. In particular, the NMR data indicate that the stem Ib of shiftable substrates contains four stable W-C base pairs in their free form, but only three stable W-C base pairs in their SLV-bound forms. A fourth base pair closing the loop was predicted to form in the SLV-bound form according to biochemical studies [28, 29], but the imino NMR data indicate that such a base pair is not stably formed in the complex. For SLIsbΔIa, for example, the C6^I-G13^I closing base pair (C626-G633 in the VS ribozyme) is destabilized upon complex formation such that the G13^I imino proton is no longer observable in the SLIsbΔIa/SLV complex.

Interestingly, destabilization of the SLI loop closing base pair is also observed upon formation of SLI/SLV complexes with preshifted substrates and the double-stranded substrate SL1ds2. In these cases, the imino proton of the closing base pair (G14^I for SLIpsΔIa and SLIps and G15^I for SL1ds2) is observed in the free SLI but not in the SLI/SLV complex. These NMR data obtained with shiftable and preshifted SLI substrates indicate that destabilization of the loop-closing G-C base pair is not a consequence of helix shifting in stem Ib, but instead results from formation of the kissing-loop interaction.

Detailed NMR studies of SL1ds2 demonstrate that the C7^I-G15^I loop-closing base pair is present in the free SLI, but that G15^I is preferentially extruded from the loop fold in the complex, with C7^I forming an unstable and distorted W-C base pair with G14^I (C626-G633 base pair in the VS ribozyme). Chemical modification data indicate that C626 remains mostly inaccessible to DMS modification upon binding of SLI to the VS ribozyme [28, 29], whereas mutational analysis indicate that disruption of the C626-G633 base pair is compatible with self-cleavage and ligation activities of the VS ribozyme [20, 21, 28]. Although these previous results appear contradictory, they both agree with the structure of the SL1ds2/SLV complex, where the C7^I-G14^I base pair

corresponding to C626-G633 is destabilized, but could still remain inaccessible to chemical modifications. Interestingly, the extrusion of G15^I is similar to the one occurring upon formation of the I/V kissing-loop complex with shiftable SLI substrates, where C634 (equivalent to G15^I) becomes extruded and accessible to DMS modification [28, 29]. Thus, the observation of a similar conformational change for residue G15^I in SLIDs2, which is not a shiftable SLI variant, strongly supports the concept that formation of the kissing-loop interaction is generally responsible for structural rearrangements involving G633 and G634 within stem Ib.

A kinetic model for substrate recognition and activation by the VS ribozyme.

Overall, the NMR data presented here are in agreement with previous biochemical studies, further supporting the value of using stem-loop domains to improve our understanding of substrate activation and recognition within the context of the full ribozyme. In depth analysis of these NMR data reveals interesting conformational changes that are consistent with a kinetic model in which formation of an initial contact between the loops of SLI and SLV in the natural VS ribozyme destabilizes the SLI loop closing base pairs and triggers the helix shift in SLI. In this model, residues G697, A698 and C699 of the prefolded U-turn of SLV initially target residues C632, U631 and G630 of the unshifted SLI and thereby induce formation of the U-turn fold in the SLI loop. Upon formation of the U-turn fold in SLI, other hydrogen-bonding and stacking interactions form at the I/V kissing-loop junction to further stabilize the complex and thereby induce the rearrangement of the C626-G633 base pair and the extrusion of C634 within stem Ib. These structural changes likely lead to helix shifting in stem Ib and structural rearrangement in the cleavage site internal loop. Although this model will need further experimental validation, it is compatible with the ability of U-turn structures to initiate rapid RNA/RNA pairing and trigger structural changes in RNA complexes [14, 15].

ASSOCIATED CONTENT

Supporting information

Supporting Figures: native gel electrophoresis of several RNAs investigated in this study (Supplementary Figure 3.1); NMR evidence of the SLIsb Δ Ia/SLV complex (Supplementary Figure 3.2); of the SLIsb Δ Ia_{GU}/SLV complex (Supplementary Figure 3.3), of the SLIps/SLV complex (Supplementary Figure 3.4), of the SLIps Δ Ia/SLV complex (Supplementary Figure 3.5), and of the SLIds1/SLV complex (Supplementary Figure 3.6); interhelical angle distribution for the 20 lowest-energy structures of the SLIds2/SLV complex (Supplementary Figure 3.7); superposition of the loop structures of SLV free and bound to SLIds2 (Supplementary Figure 3.8). Supporting Tables: chemical shifts of detectable G/U imino protons for SLI variants either in their free form or bound to SLV RNA (Supplementary Table 3.1); chemical shift of detectable G/U imino protons for SLV either in its free form or bound to SLI variants (Supplementary Table 3.2); chemical shift assignments of exchangeable ¹H and ¹⁵N for the SLIds2/SLV complex (Supplementary Table 3.3); chemical shift assignments of nonexchangeable ¹H and ¹³C for the SLIds2/SLV complex (Supplementary Table 3.4); summary of intermolecular NOEs at the kissing-loop junction of the SLIds2/SLV complex (Supplementary Table 3.5). This material is available free of charge via the Internet at <http://pubs.acs.org>.

ACCESSION CODES

The NMR chemical shifts, structural restraints and structural coordinates of the SLIds2/SLV complex have been deposited through the RCSB Protein Data Bank and given BMRB entry code 19662 and PDB entry code 2MI0.

AUTHOR INFORMATION

Corresponding Author

*E-mail: ; Phone: .

Funding Sources

This work was supported by the Canadian Institutes for Health Research (CIHR) to P.L. [MOP-86502]. P.B. was supported by a Frederick Banting and Charles Best Canada Graduate Scholarship Ph.D. scholarship from CIHR and graduate scholarships from the Université de Montréal. P.L. holds a Canada Research Chair in Structural Biology and Engineering of RNA. Funding for open access charge: CIHR.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

We thank Dominique Chaussé for sample preparation, Nicolas Girard for structural analysis and Ryan Richter for computer support. We also thank Alexandre Desjardins, Eric Bonneau and James G. Omichinski for critical reading of the manuscript.

ABBREVIATIONS

2D, two-dimensional; 3D, three-dimensional; bp, base pair; HSQC, heteronuclear single quantum coherence; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; PDB, protein data bank; SLI, stem-loop I; SLV, stem-loop V; RMSD, root-mean-square deviation; SAXS; small angle X-ray scattering; W-C, Watson-Crick.

REFERENCES

1. Quigley, G.J. and A. Rich, *Structural domains of transfer RNA molecules*. Science, 1976. **194**: p. 796-806.
2. Pley, H.W., K.M. Flaherty, and D.B. McKay, *Three-dimensional structure of a hammerhead ribozyme*. Nature, 1994. **372**: p. 68-74.
3. Jucker, F.M. and A. Pardi, *GNRA tetraloops make a U-turn*. RNA, 1995. **1**(2): p. 219-22.

4. Huang, S., Y.-X. Wang, and D.E. Draper, *Structure of a hexanucleotide RNA hairpin loop conserved in ribosomal RNAs*. J Mol Biol, 1996. **258**(2): p. 308-21.
5. Stallings, S.C. and P.B. Moore, *The structure of an essential splicing element: stem loop IIa from yeast U2 snRNA*. Structure, 1997. **5**(9): p. 1173-1185.
6. Gutell, R.R., et al., *Predicting U-turns in ribosomal RNA with comparative sequence analysis*. J Mol Biol, 2000. **300**(4): p. 791-803.
7. Huppler, A., et al., *Metal binding and base ionization in the U6 RNA intramolecular stem-loop structure*. Nature Struct Biol, 2002. **9**(6): p. 431-435.
8. Puglisi, E.V. and J.D. Puglisi, *HIV-1 A-rich RNA loop mimics the tRNA anticodon structure*. Nature Struct Biol, 1998. **5**(12): p. 1033-6.
9. Klosterman, P.S., et al., *Three-dimensional motifs from the SCOR, structural classification of RNA database: extruded strands, base triples, tetraloops and U-turns*. Nucleic Acids Res, 2004. **32**(8): p. 2342-52.
10. Sefcikova, J., et al., *The genomic HDV ribozyme utilizes a previously unnoticed U-turn motif to accomplish fast site-specific catalysis*. Nucleic Acids Res, 2007. **35**(6): p. 1933-46.
11. Harris, S. and S.J. Schroeder, *Nuclear magnetic resonance structure of the prohead RNA E-loop hairpin*. Biochemistry, 2010. **49**(29): p. 5989-97.
12. Dix, D.B., et al., *Effect of replacing uridine 33 in yeast tRNAPhe on the reaction with ribosomes*. J Biol Chem, 1986. **261**(22): p. 10112-8.
13. Ashraf, S.S., et al., *The uridine in "U-turn": contributions to tRNA-ribosomal binding*. RNA, 1999. **5**(4): p. 503-511.
14. Franch, T., et al., *Antisense RNA regulation in prokaryotes: Rapid RNA/RNA interaction facilitated by a general U-turn loop structure*. J Mol Biol, 1999. **294**(5): p. 1115-1125.
15. Ogle, J.M., et al., *Recognition of cognate transfer RNA by the 30S ribosomal subunit*. Science, 2001. **292**(5518): p. 897-902.
16. Collins, R.A., *The Neurospora Varkud satellite ribozyme*. Bioch Soc Trans Rev, 2002. **30**: p. 1122-1126.
17. Lilley, D.M., *The Varkud satellite ribozyme*. RNA, 2004. **10**: p. 151-158.

18. 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.
19. 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.
20. 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.
21. 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.
22. 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.
23. 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.
24. 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.
25. Bouchard, P., et al., *Role of SLV in SLI substrate recognition by the Neurospora VS ribozyme.* RNA, 2008. **14**: p. 736-748.
26. DeAbreu, D.M., J.E. Olive, and R.A. Collins, *Additional roles of a peripheral loop-loop interaction in the Neurospora VS ribozyme.* Nucleic Acids Res, 2011. **39**(14): p. 6223-8.
27. 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.
28. Andersen, A. and R.A. Collins, *Rearrangement of a stable RNA secondary structure during VS ribozyme catalysis.* Mol Cell, 2000. **5**: p. 469-478.

29. 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.
30. 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.
31. 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.
32. 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.
33. 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.
34. 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.
35. 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.
36. Lipfert, J., et al., *The complete VS ribozyme in solution studied by small-angle X-ray scattering*. Structure, 2008. **16**(9): p. 1357-67.
37. Piotto, M., V. Saudek, and V. Skleňár, *Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions*. J Biomol NMR, 1992. **2**: p. 661-665.
38. Kay, L.E., P. Keifer, and T. Saarinen, *Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity*. J Am Chem Soc, 1992. **114**: p. 10663-10665.

39. Dingley, A.J. and S. Grzesiek, *Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide $^2J_{NN}$ couplings*. J Am Chem Soc, 1998. **120**: p. 8293-8297.
40. Grzesiek, S. and A. Bax, *The importance of not saturating water in protein NMR. Application to sensitivity enhancement of NOE measurements*. J Am Chem Soc, 1993. **115**: p. 12593-12594.
41. Simorre, J.P., et al., *Correlation of the guanosine exchangeable and non-exchangeable base protons in $^{13}C/^{15}N$ -labeled RNA with an HNC-TOCSY-CH experiment*. J Biomol NMR, 1996. **7**: p. 153-156.
42. Mueller, L., P. Legault, and A. Pardi, *Improved RNA structure determination by detection of NOE contacts to exchange-broadened amino groups*. J Am Chem Soc, 1995. **117**: p. 11043-11048.
43. Zhang, O., et al., *Backbone 1H and ^{15}N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques*. J Biomol NMR, 1994. **4**(6): p. 845-58.
44. Pascal, S.M., et al., *Simultaneous acquisition of ^{15}N and ^{13}C -edited NOE spectra of proteins dissolved in H_2O* . J Magn Res Series B, 1994. **103**: p. 197-201.
45. Vuister, G.W. and A. Bax, *Resolution enhancement and spectral editing of uniformly ^{13}C -enriched proteins by homonuclear broadband ^{13}C decoupling*. J Magn Reson, 1992. **98**: p. 428-435.
46. Santoro, J. and G.C. King, *A constant-time 2D overbroadening experiment for inverse correlation of isotopically enriched species*. J Magn Reson, 1992. **97**: p. 202-207.
47. Pardi, A. and E.P. Nikonowicz, *Simple procedure for resonance assignment of the sugar protons in ^{13}C -labeled RNAs*. J Am Chem Soc, 1992. **114**: p. 9202-9203.
48. Ikura, M., et al., *Three-dimensional NOESY-HMQC spectroscopy of a ^{13}C -labeled protein*. J Magn Reson, 1990. **86**: p. 204-209.
49. Kay, L.E. and A. Bax, *New methods for the measurement of NH---C α H coupling constants in ^{15}N -labeled proteins*. J Magn Reson, 1990. **86**(1): p. 110-126.

50. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes*. J Biomol NMR, 1995. **6**: p. 277-293.
51. Vranken, W.F., et al., *The CCPN data model for NMR spectroscopy: development of a software pipeline*. Proteins, 2005. **59**(4): p. 687-96.
52. Hansen, M.R., L. Mueller, and A. Pardi, *Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions*. Nature Struct Biol, 1998. **5**: p. 1064-1074.
53. 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.
54. 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.
55. 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.
56. 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.
57. Theimer, C.A., et al., *Structural and functional characterization of human telomerase RNA processing and cajal body localization signals*. Mol Cell, 2007. **27**(6): p. 869-81.
58. 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.
59. Cabello-Villegas, J., M.E. Winkler, and E.P. Nikonowicz, *Solution conformations of unmodified and A37N6-dimethylallyl modified anticodon stem-loops of Escherichia coli tRNA^{Phe}*. J Mol Biol, 2002. **319**(5): p. 1015-1034.
60. Tewari, A.K. and R. Dubey, *Emerging trends in molecular recognition: utility of weak aromatic interactions*. Bioorgan Med Chem, 2008. **16**(1): p. 126-43.

61. Steiner, T. and G. Koellner, *Hydrogen bonds with pi-acceptors in proteins: frequencies and role in stabilizing local 3D structures*. J Mol Biol, 2001. **305**(3): p. 535-57.
62. Meyer, E.A., R.K. Castellano, and F. Diederich, *Interactions with aromatic rings in chemical and biological recognition*. Angew Chem Int Ed Engl, 2003. **42**(11): p. 1210-50.
63. 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.
64. Tzokov, S.B., I.A. Murray, and J.A. Grasby, *The role of magnesium ions and 2'-hydroxyl groups in the VS ribozyme-substrate interaction*. J Mol Biol, 2002. **324**: p. 215-226.
65. 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.
66. Kovacheva, Y.S., et al., *The role of phosphate groups in the VS ribozyme-substrate interaction*. Nucleic Acids Res, 2004. **32**(21): p. 6240-50.

TABLES

Table 3.I : Structural statistics

	SLIds2/SLV
Number of NOE-derived distance restraints	507
intranucleotide	181
internucleotide	291
intermolecular	32
ambiguous	3
hydrogen-bond restraints	89
base-pair planarity restraints	32
dihedral angle restraints	173
residual dipolar coupling ($^1D_{NH}$) restraints	16
P-P distance restraints	19
total number of restraints	836
RMSD from experimental restraints	
NOE (Å; none > 0.2 Å)	0.069 ± 0.008
dihedral (°; none > 5°)	0.008 ± 0.001
residual dipolar coupling (Hz; none > 0.1 Hz)	0.0018 ± 0.0005
RMSD from idealized geometry	
bonds (Å)	0.00441 ± 0.00005
angles (°)	0.995 ± 0.001
impropers (°)	0.385 ± 0.001
heavy-atom RMSD to the minimized average structure	
overall (SLIds2: 2-20, SLV: 2-20)	2.02 ± 0.77
overall SLI (SLIds2: 2-20)	1.82 ± 0.77
overall SLV (SLV: 2-20)	1.15 ± 0.27
stem I (SLIds2: 2-6, 16-20)	0.59 ± 0.10
stem V (SLV: 2-7, 15-20)	0.49 ± 0.07
loop I (SLIds2: 6-16)	1.78 ± 0.69
loop V (SLV: 7-15)	1.01 ± 0.20
kissing-loop stem (SLIds2: 11-13, SLV: 10-12)	0.73 ± 0.19
kissing-loop junction (SLIds2: 6-16, SLV: 7-15)	1.70 ± 0.51

FIGURES

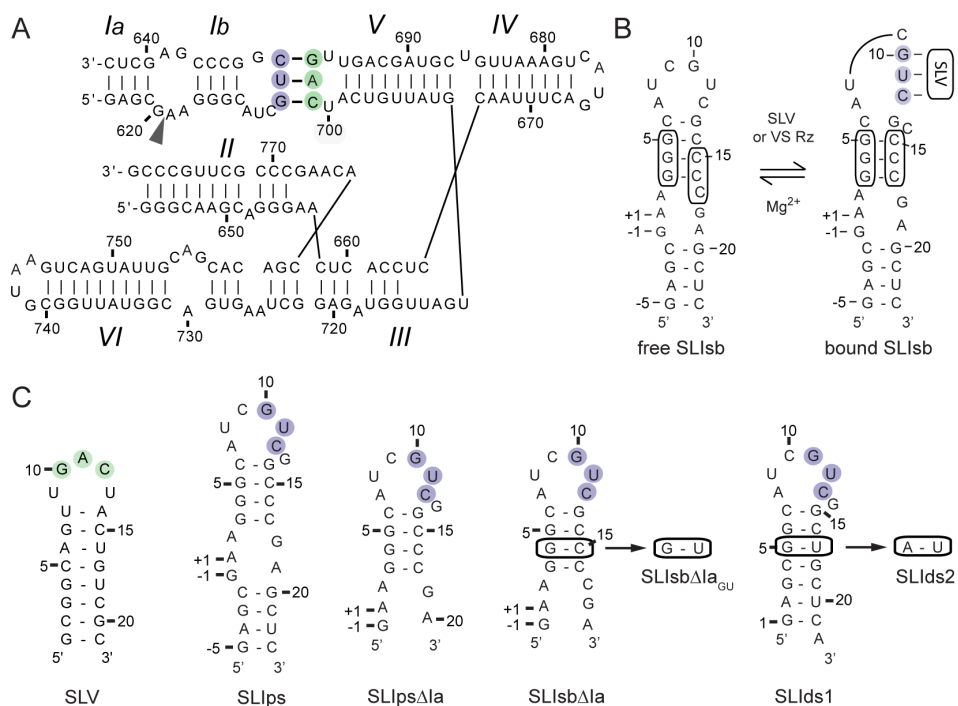


Figure 3.1 : Primary and secondary structures of RNAs used in this study. (A) The catalytic domain of the *Neurospora* VS ribozyme containing helical domains II-VI and an SLI substrate (SLIps) containing stems Ia and Ib. The cleavage site is indicated by an arrowhead. The I/V kissing-loop interaction involves Watson-Crick base pairs (thick black lines) between residues G630, U631 and C632 of SLI (shaded in purple) and residues C699, A698 and G697 of SLV (shaded in green). (B) Formation of the I/V kissing-loop is accompanied by a structural rearrangement of the SLI substrate from an unshifted (free SLIsb) to a shifted (bound SLIsb) conformation. (C) The SLV and SLI RNAs used for investigation of SLI/SLV complexes by NMR spectroscopy. In (B) and (C), the cleavage site of the SLI substrates (SLIsb, SLIps, SLIpsΔIa, SLIsbΔIa and SLIsbΔIa_{GU}) is between residues -1 and +1.

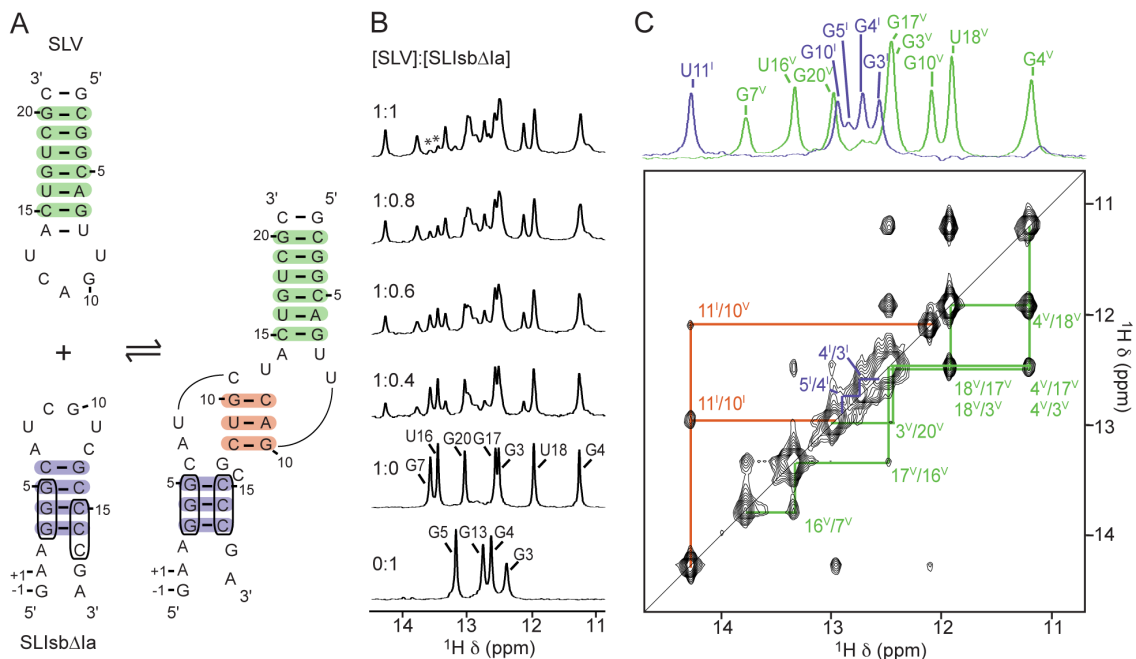


Figure 3.2 : NMR evidence for formation of the SLIsbΔIa/SLV complex. (A) Proposed secondary structure of SLIsbΔIa, a shiftable SLI variant, and SLV in their free (left) and bound (right) forms. The boxed nucleotides in SLIsbΔIa are those involved in the helix shift associated with SLV binding. (B) Imino region of 1D ^1H NMR spectra recorded for the titration of ^{15}N -labeled SLV with unlabeled SLIsbΔIa. 1D NMR spectra were collected under the same conditions and plotted on the same vertical scale but were adjusted to take into account sample dilution during the titration. Imino proton assignment is provided for both SLIsbΔIa and SLV in their free forms. The imino proton signals marked with an asterisk in the 1:1 SLIsbΔIa/SLV complex represent a small excess of free SLV. (C) Imino region of the 2D NOESY spectrum of the SLIsbΔIa/ ^{15}N -SLV complex. On top, 1D ^{15}N -filtered (purple; SLIsbΔIa) and ^{15}N -edited (green; SLV) spectra of the complex annotated with imino proton assignments. Annotations in (B) and (C) are used to identify imino signals within SLIsbΔIa (purple), within SLV (green) and at the kissing-loop junction (orange), and these imino signals provide evidence for the base pairs shaded with the corresponding colors in (A).

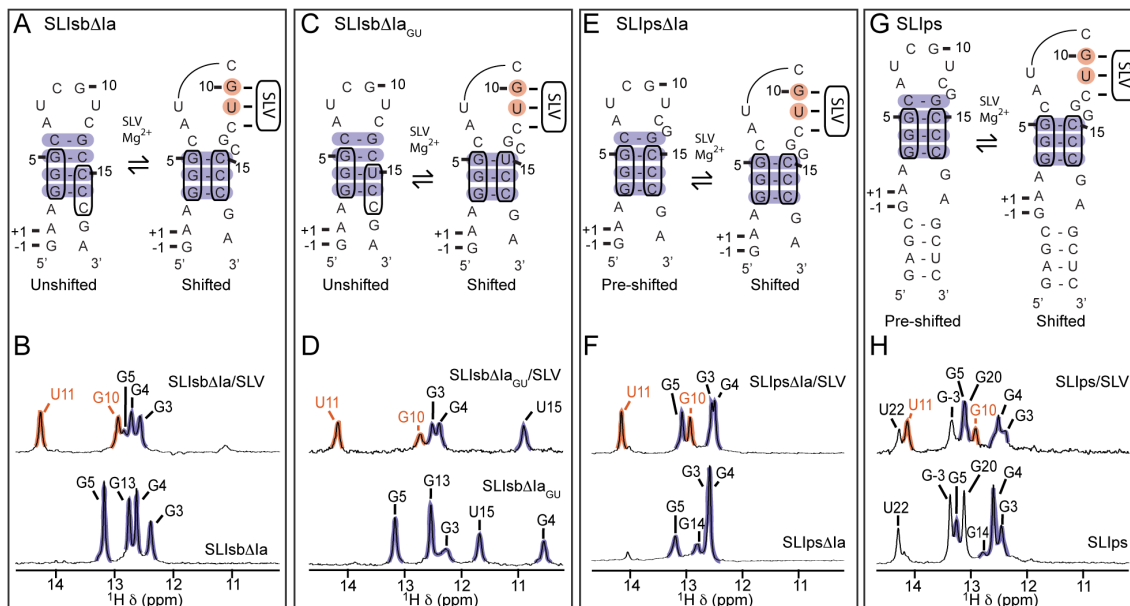


Figure 3.3 : NMR evidence of structural rearrangements in SLI variants. Upper panels: Proposed secondary structures of the free and SLV-bound forms of shiftable SLI [(A) SLIsbΔIa and (C) SLIsbΔIa_{GU}] and preshifted SLI [(E) SLIpsΔIa and (G) SLIps] variants. Lower panels: Imino region of 1D ¹H NMR spectra of (B) SLIsbΔIa, (D) SLIsbΔIa_{GU}, (F) SLIpsΔIa and (H) SLIps (bottom spectra), along with the 1D ¹H ¹⁵N-filtered or ¹⁵N-edited NMR spectra of the (B) SLIsbΔIa/¹⁵N-SLV, (D) ¹⁵N-SLIsbΔIa_{GU}/SLV, (F) ¹⁵N-SLIpsΔIa/SLV and (H) SLIps/¹⁵N-SLV complexes for detection of SLI imino proton signals only (top spectra). The shaded imino proton signals of residues from the kissing-loop interaction (orange) and the adjacent SLI stem (purple) provide evidence for the base pairs shaded with the corresponding colors in the proposed secondary structures shown in (A), (C), (E), and (G).

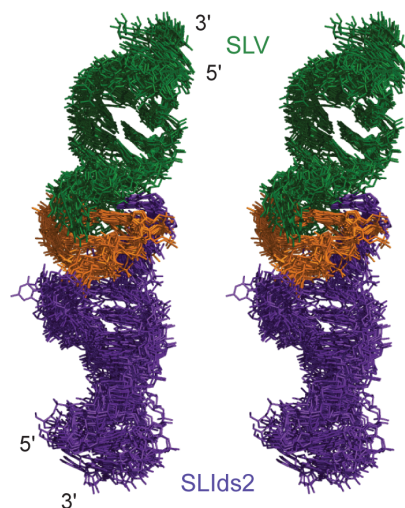


Figure 3.5 : Stereoview of the NMR structures of the SLIds2/SLV kissing-loop complex.

Superposition of the 20 lowest-energy structures on the minimized average structure. Only heavy atoms of SLIds2 and SLV residues 2-20 were used for the superposition and are shown. Residues are shown in the prevailing coloring scheme (SLIds2, purple; SLV, green; W-C base pairs at the kissing-loop junction, orange).

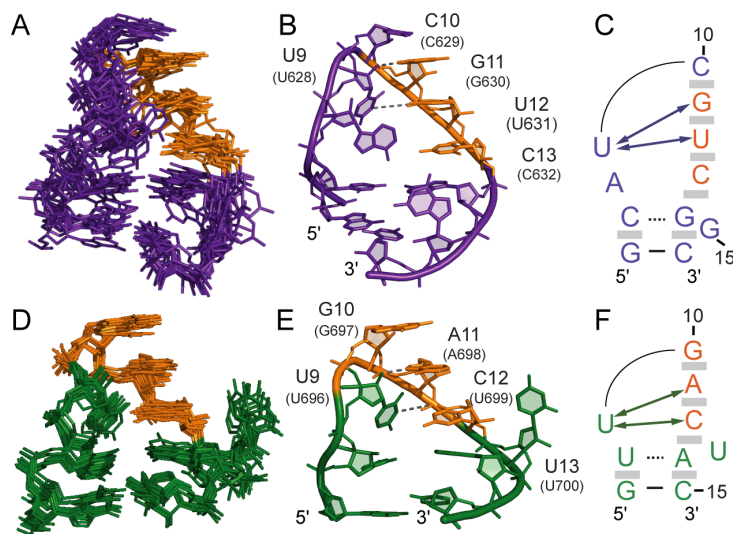


Figure 3.6 : U-turn structures in the SLIDs2 and SLV loops of the SLIDs2/SLV complex.

Superposition of (A) the SLIDs2 and (D) SLV loops of the 10 lowest-energy structures on the minimized average structure (not shown). Only heavy atoms of SLIDs2 residues 6-16 or SLV residues 7-15 were used for the superposition and are shown. Minimized average structure of the (B) SLIDs2 and (E) SLV loops showing base stacking after the sharp turn in the backbone, which is also represented by shaded gray boxes in (C) and (F). Dashed lines represent hydrogen bonds characteristic of a U-turn motif between U9^I 2'-OH and G11^I N7 and between U9^I H3 and G11^I 3'-phosphate in SLIDs2 as well as between U9^V 2'-OH and A11^V N7 and between U9^V H3 and A11^V 3'-phosphate in SLV. Schematic representation of the (C) SLIDs2 and (F) SLV loops in the complex. The arrows indicate observed NOEs that are characteristic of the U-turn motif, including those between U9^I H1'/H2' and G11^I H8 and between U9^I H1' and U12^I H6/H5 in SLIDs2 and the equivalent NOEs in SLV.

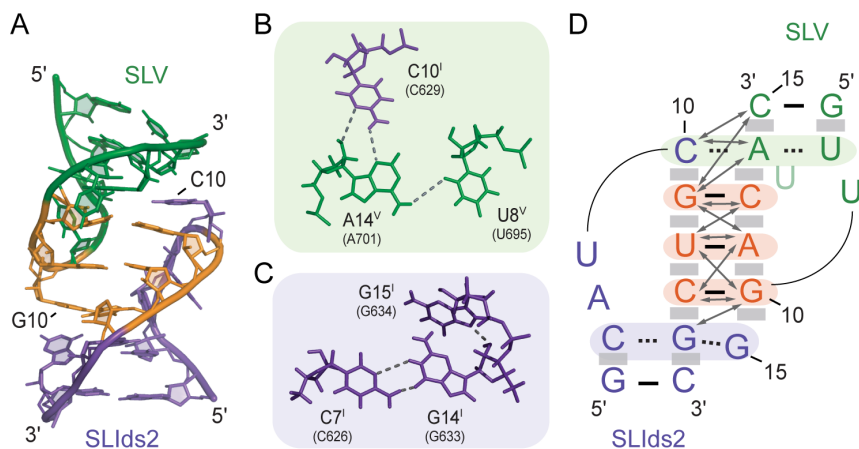
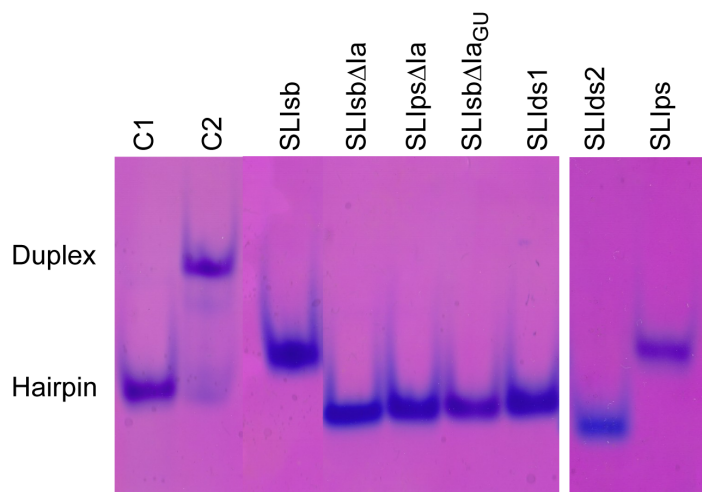
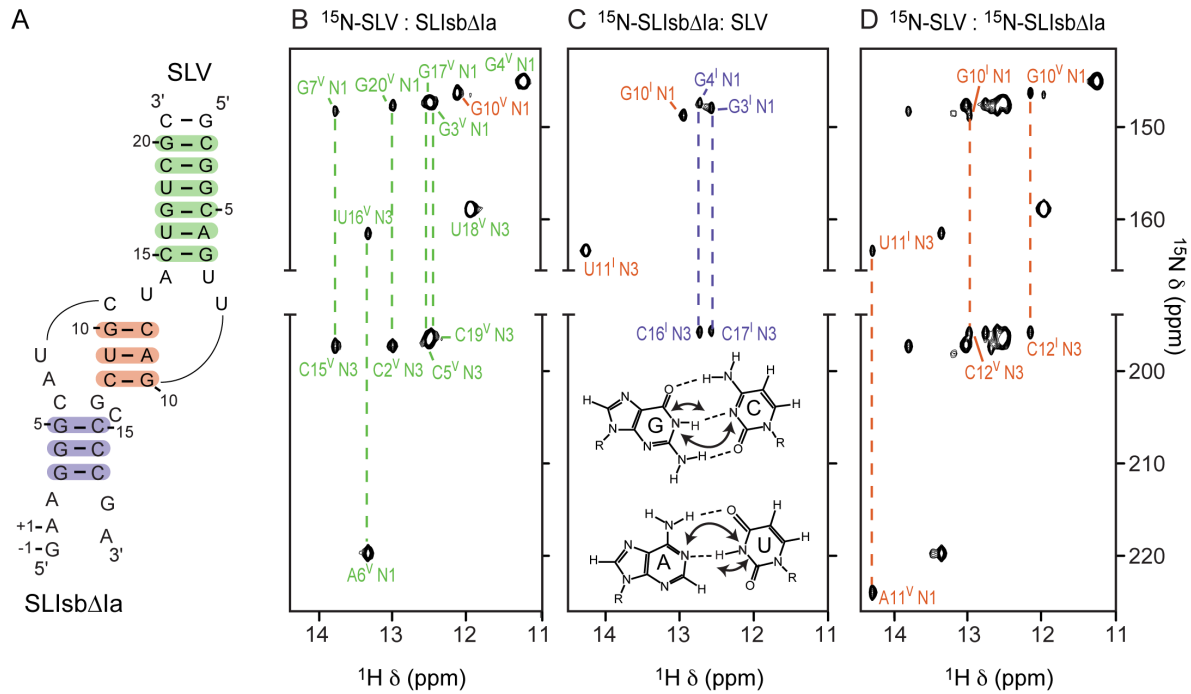


Figure 3.7 : Structural characteristics of the I/V kissing-loop junction. (A) Minimized average structure showing the kissing-loop interaction. (B)(C) Base triples in the minimized average structure. Dashed lines represent the following hydrogen bonds defined on the basis of short distances in the ensemble of structures (given in parentheses): A14^V H61 and U8^V O2 (2.85-4.88 Å), A14^V N3 and C10^I H41 (2.57-3.43 Å), A14^V O2' and C10^I N3 (3.06-4.56 Å), C7^I H41 and G14^I O6 (1.70-2.74 Å), C7^I N3 and G14^I H1 (2.85-4.52 Å in 19 structures). (D) Schematic representation of the kissing-loop interaction between SLids2 and SLV. The arrows indicate intermolecular NOEs observed between residues at the kissing-loop junction (Table S2 of the Supporting Information). Residues shaded in orange form Watson-Crick base pairs at the kissing-loop junction, those shaded in green and purple are involved in base triples, and shaded gray boxes illustrate base stacking. For clarity, U13^V was drawn in a lighter green color.

SUPPORTING INFORMATION

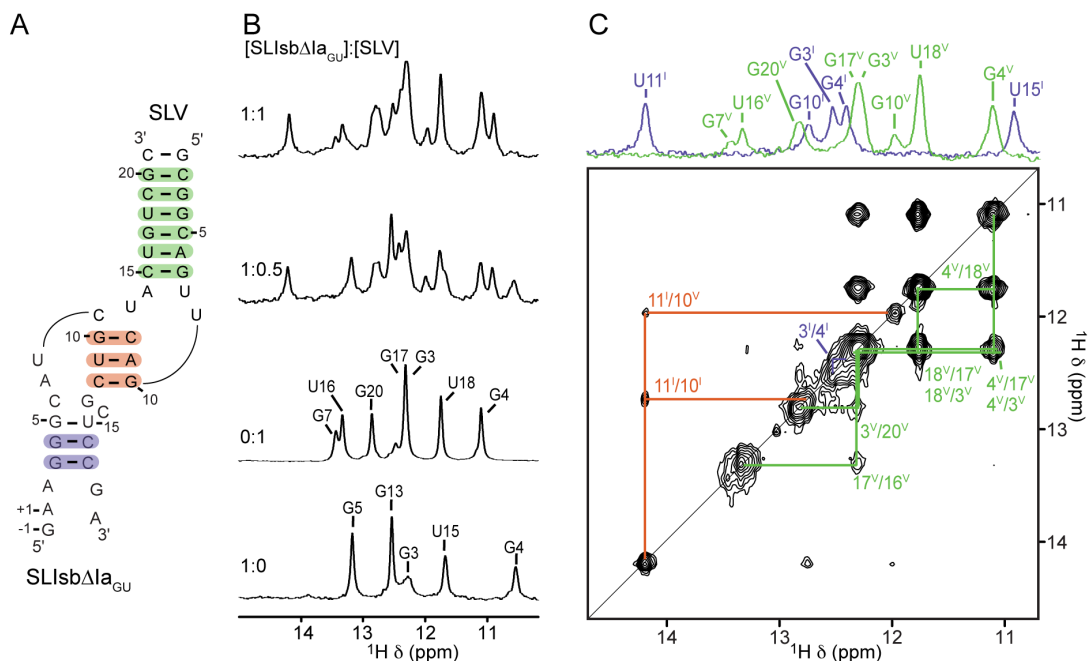


Supplementary Figure 3.1 : Native gel electrophoresis of several RNAs investigated. For each RNA, 5 μ g were loaded on the gel in the presence of 20 mM MgCl_2 . C1 and C2 are control RNAs that respectively correspond to S₁ (26-mer) and S₂ (24-mer) in a previous study (1). C1 and C2 mainly adopt hairpin and duplex conformations, respectively.



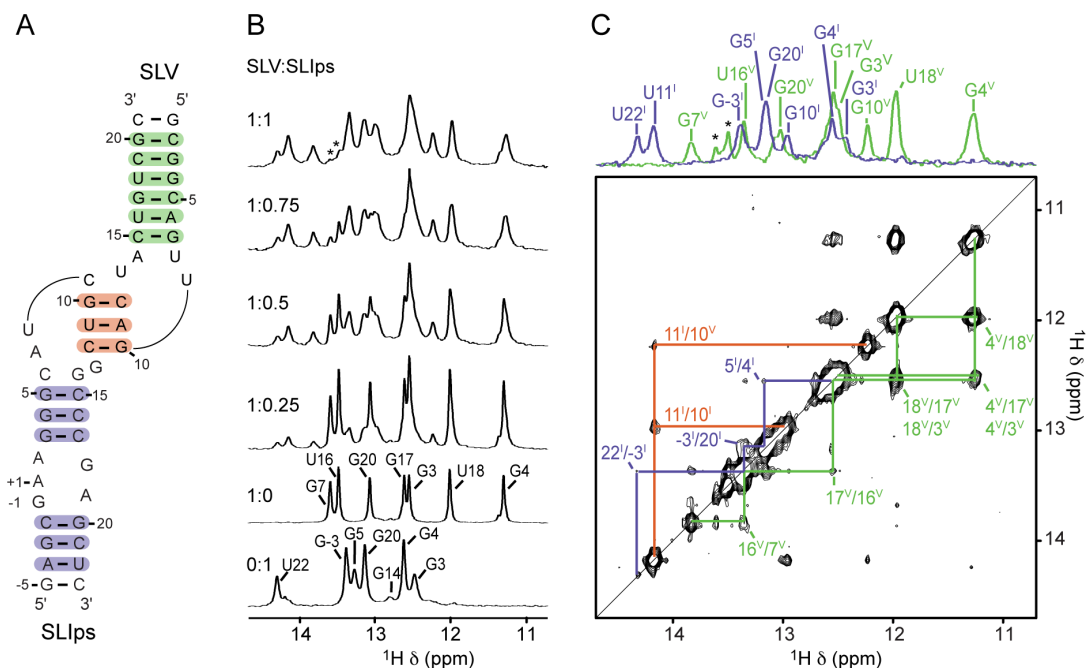
Supplementary Figure 3.2 : NMR evidence for formation of the SLIsb Δ Ia/SLV complex.

(A) Proposed secondary structure of the SLIsb Δ Ia/SLV complex. (B-D) Imino regions of 2D ^1H - ^{15}N HNN-COSY spectra at 25°C for (B) 1.0 mM SLIsb Δ Ia/ ^{15}N -SLV, (C) 2.3 mM ^{15}N -SLIsb Δ Ia/SLV and (D) 1.1 mM ^{15}N -SLIsb Δ Ia/ ^{15}N -SLV. Dotted lines and annotations are used to identify correlations between G N1 and C N3 and between U N3 and A N1 within SLIsb Δ Ia (purple), SLV (green) and at the kissing-loop junction (orange).

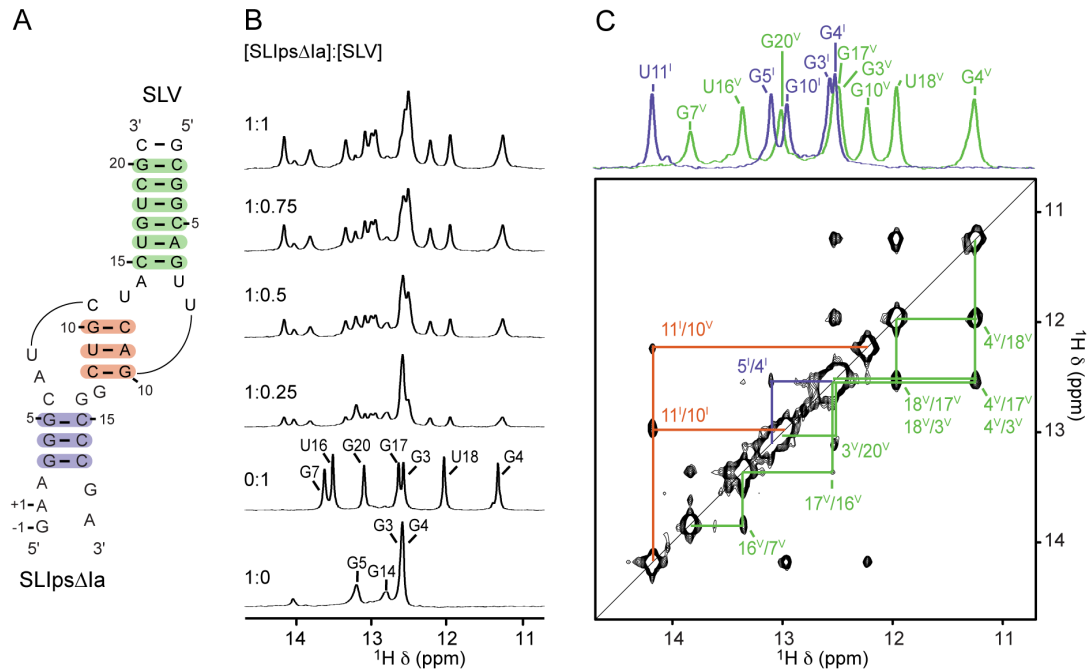


Supplementary Figure 3.3 : NMR evidence for formation of the SLIsbΔIa_{GU}/SLV complex.

(A) Proposed secondary structure of the complex. (B) Imino region of 1D ¹H NMR spectra recorded for the titration of ¹⁵N-labeled SLIsbΔIa_{GU} with unlabeled SLV. 1D NMR spectra were collected under the same conditions and plotted as described in Figure 3.2. Imino proton assignment is provided for both SLIsbΔIa_{GU} and SLV in their free form. (C) Imino region of the 2D NOESY spectrum of the ¹⁵N-SLIsbΔIa_{GU}/SLV complex. Lines and annotations are used to identify NOE interactions within SLIsbΔIa_{GU} (purple), SLV (green) and at the kissing-loop junction (orange). These NOE interactions provide evidence for the base pairs that are shaded with the corresponding colors in A. On top, 1D ¹⁵N-filtered (green; SLV) and 1D ¹⁵N-edited (purple; SLIsbΔIa_{GU}) spectra of the complex annotated with imino proton assignments.

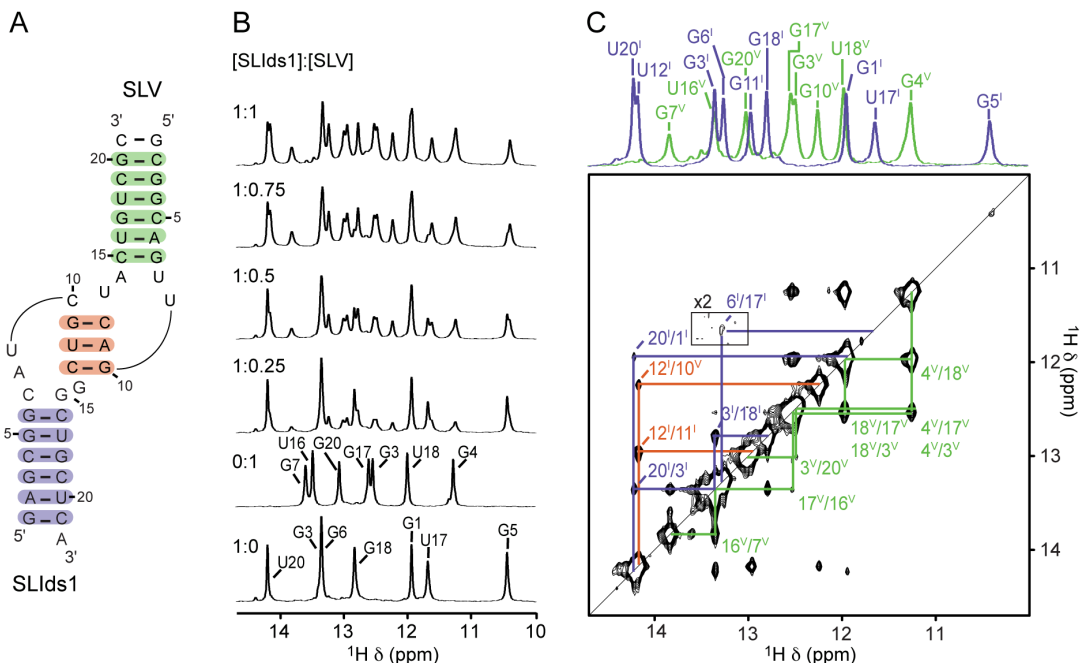


Supplementary Figure 3.4 : NMR evidence for formation of the SLIps/SLV complex. (A) Proposed secondary structure of the complex. (B) Imino region of 1D ^1H NMR spectra recorded for the titration of ^{15}N -labeled SLV with unlabeled SLIps. 1D NMR spectra were collected under the same conditions and plotted as described in Figure 3.2. Imino proton assignment is provided for both SLIps and SLV in their free form. (C) Imino region of the 2D NOESY spectrum of the SLIps/ ^{15}N -SLV complex. Lines and annotations are used to identify NOE interactions within SLIps (purple), SLV (green) and at the kissing-loop junction (orange). These NOE interactions provide evidence for the base pairs that are shaded with the corresponding colors in A. On top, 1D ^{15}N -filtered (purple; SLIps) and 1D ^{15}N -edited (green; SLV) spectra of the complex annotated with imino proton assignments. The asterisks in B and C indicate 1D imino proton signals from the small excess of SLV in the 1:1 complex.

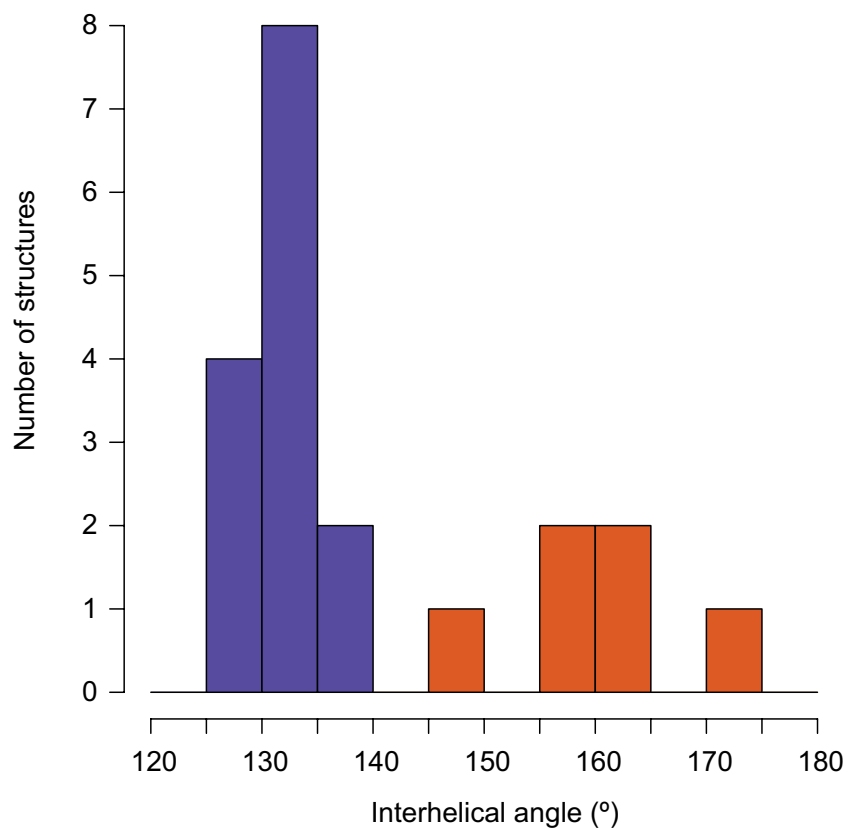


Supplementary Figure 3.5 : NMR evidence for formation of the SLIpsΔIa/SLV complex.

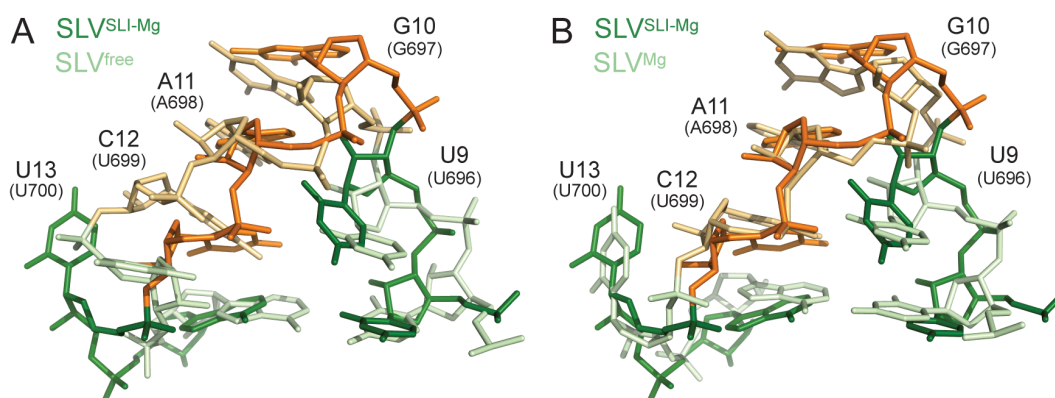
(A) Proposed secondary structure of the complex. (B) Imino region of 1D ^1H NMR spectra recorded for the titration of ^{15}N -labeled SLIpsΔIa with unlabeled SLV. 1D NMR spectra were collected under the same conditions and plotted as described in Figure 3.2. Imino proton assignment is provided for both SLIpsΔIa and SLV in their free form. (C) Imino region of the 2D NOESY spectrum of the ^{15}N -SLIpsΔIa/SLV complex. Lines and annotations are used to identify NOE interactions within SLIpsΔIa (purple), SLV (green) and at the kissing-loop junction (orange). These NOE interactions provide evidence for the base pairs that are shaded with the corresponding colors in A. On top, 1D ^{15}N -filtered (green; SLV) and 1D ^{15}N -edited (purple; SLIpsΔIa) spectra of the complex annotated with imino proton assignments.



Supplementary Figure 3.6 : NMR evidence for formation of the SLIds1/SLV complex. (A) Proposed secondary structure of the complex. (B) Imino region of 1D ^1H NMR spectra recorded for the titration of ^{15}N -labeled SLIds1 with unlabeled SLV. 1D NMR spectra were collected under the same conditions and plotted as described in Figure 3.2. Imino proton assignment is provided for both SLIds1 and SLV in their free form. (C) Imino region of the 2D NOESY spectrum of the ^{15}N -SLIds1/SLV complex. Lines and annotations are used to identify NOE interactions within SLIds1 (purple), SLV (green) and at the kissing-loop junction (orange). These NOE interactions provide evidence for the base pairs that are shaded with the corresponding colors in A. On top, 1D ^{15}N -filtered (green; SLV) and 1D ^{15}N -edited (purple; SLIds1) spectra of the complex annotated with imino proton assignments.



Supplementary Figure 3.7: Interhelical angle distribution for the 20 lowest-energy structures of the SLIDs2/SLV complex. The distribution shows two categories: one represented by 14 structures with an average interhelical angle of $131 \pm 3^\circ$ and an average axis displacement of $15 \pm 3^\circ$ (purple bars) and the other represented by 6 structures with an average interhelical angle of $161 \pm 9^\circ$ and an average axis displacement of $19 \pm 7^\circ$ (orange bars).



Supplementary Figure 3.8 : Superposition of the loop structures of SLV free and bound to SLIDs2. The heavy atom superposition of the loop residues (U695-A701) from the average structures of SLV^{SLI-Mg} with the average structures of (A) SLV^{free} determined in the absence of magnesium ions [1TBK.pdb; (2)] and (B) SLV^{Mg} determined in the presence of magnesium ions [1YN1.pdb; (3)]. The structure of SLV^{free} and SLV^{Mg} are represented in light colors and that of SLV^{SLI-Mg} in dark colors, either orange for residues involved in the three W-C base pairs at the kissing-loop junction or green for other residues. The heavy atom rmsd for the superposition in A and B are respectively 3.20 Å and 1.75 Å.

Supplementary Table 3.I : Chemical shifts (ppm) of detectable G/U imino protons for SLI variants either in their free form or bound to SLV RNA.^a

imino proton	SLIps ^b		SLIpsΔIa ^b		SLIsbΔIa ^c		SLIsbΔIa _{GU} ^d		imino proton	SLIds1 ^b		SLIds2 ^b	
	free	bound	free	bound	free	bound	free	bound		free	bound	free	bound
G-3	13.39	13.37	-	-	-	-	-	-	G1	11.95	11.94	11.93	11.9
G3	12.48	12.42	12.57	12.53	12.40	12.58	12.29	12.53	G3	13.37	13.35	13.32	13.3
G4	12.62	12.53	12.56	12.48	12.64	12.73	10.56	12.40	G5	10.45	10.41	-	-
G5	13.28	13.14	13.18	13.06	13.19	12.86	13.18	n.o.	G6	13.36	13.26	13.48	13.3
G10	n.o.	12.94	n.o.	12.92	n.o.	12.96	n.o.	12.74	G11	n.o.	12.97	n.o.	12.9
U11	n.o.	14.16	n.o.	14.14	n.o.	14.28	n.o.	14.19	U12	n.o.	14.18	n.o.	14.1
G13/14	12.81	n.o.	12.77	n.o.	12.77	n.o.	12.55	n.o.	U17	11.70	11.64	13.55	13.5
U15	-	-	-	-	-	-	11.69	10.91	G18	12.85	12.79	12.50	12.4
G20	13.15	13.12	-	-	-	-	-	-	U20	14.21	14.22	14.15	14.1
U22	14.32	14.30	-	-	-	-	-	-					

a. In this table, residues from different SLI variants that have equivalent positions in the secondary structure are aligned horizontally to facilitate the comparison of their chemical shifts. The shaded lines highlight the residues involved in Watson-Crick base-pairs at the kissing-loop junction. The symbol "-" indicates that the assignment could not be obtained because the listed nucleotide is not part of this SLI variant. The abbreviations "n.o." indicates that the imino signal is not observed.

b. The NMR data for SLIps, SLIpsΔIa, SLIds1 and SLIds2 were collected at 25°C under standard buffer conditions (10 mM d₁₁-Tris pH 7.0, 50 mM NaCl, 5 mM MgCl₂, 0.05 mM NaN₃, 10% D₂O)

c. The NMR data for SLIsbΔIa were collected at 25°C under standard buffer conditions, but with 20 mM MgCl₂ instead of 5 mM MgCl₂.

d. The NMR data for SLIsbΔIa_{GU} were collected at 5°C under standard buffer conditions, but with 20 mM MgCl₂ instead of 5 mM MgCl₂.

Supplementary Table 3.II : Chemical shift (ppm) of detectable G/U imino protons for SLV either in its free form or bound to SLI variants.^a

imino proton	Condition 1 ^b					Condition 2 ^c		Condition 3 ^d	
	SLV	SLV with				SLV	SLV with	SLV	SLV with
	free	SLIps	SLIpsΔIa	SLIds1	SLIds2	free	SLIsbΔIa	free	SLIsbΔIa _{GU}
G3	12.55	12.49	12.44	12.49	12.46	12.52	12.47	12.32	12.30
G4	11.29	11.25	11.21	11.25	11.24	11.26	11.21	11.11	11.12
G7	13.61	13.81	13.79	13.84	13.81	13.59	13.79	13.44	13.44
G10	n.o.	12.21	12.18	12.25	12.22	n.o.	12.11	n.o.	11.98
U16	13.50	13.34	13.31	13.36	13.34	13.47	13.35	13.34	13.34
G17	12.62	12.52	12.48	12.53	12.50	12.58	12.48	12.33	12.33
U18	12.01	11.95	11.92	11.97	11.94	11.97	11.93	11.76	11.76
G20	13.08	13.00	12.97	13.02	12.99	13.05	13.00	12.87	12.83

a. The shaded line highlights the residue involved in Watson-Crick base-pairing at the kissing-loop junction. The abbreviations "n.o." indicates that the imino signal is not observed.

b. The NMR data were collected at 25°C under standard buffer conditions (10 mM d₁₁-Tris pH 7.0, 50 mM NaCl, 5 mM MgCl₂, 0.05 mM NaN₃, 10% D₂O).

c. The NMR data were collected at 25°C under standard buffer conditions, but with 20 mM MgCl₂ instead of 5 mM MgCl₂.

d. The NMR data were collected at 5°C under standard buffer conditions, but with 20 mM MgCl₂ instead of 5 mM MgCl₂.

Supplementary Table 3.III : Chemical shift assignments of exchangeable ^1H and ^{15}N for the SLIDs2/SLV complex.^a

SLIDS2 BASE EXCHANGEABLE PROTON CHEMICAL SHIFTS (ppm)								
	H1 ^b	H21	H22	H3 ^b	H41 ^b	H42 ^b	H61 ^b	H62 ^b
G1	11.92							
A2								
G3	13.31							
C4					8.32	6.65		
A5								
G6	13.39							
C7					7.79	6.45		
A8								
U9								
C10					7.24	7.24		
G11	12.94							
U12				14.17				
C13					8.13	6.40		
G14								
G15								
C16					8.45	6.81		
U17				13.55				
G18	12.43							
C19					8.48	6.65		
U20				14.14				
C21					8.10	6.85		
A22							6.44	6.44

Supplementary Table 3.III (continued)

SLIDS2 BASE NITROGEN CHEMICAL SHIFTS (ppm)								
	N1^c	N2	N3^c	N4^d	N6	N7^e	N9^e	
G1	146.7					231.5	169.9	
A2	220.9		212.4			228.2	170.8	
G3	148.0					233.4	169.6	
C4			196.9	98.43				
A5	220.2		213.9			230.3	170.8	
G6	148.1					233.4	169.1	
C7				98.47				
A8								
U9								
C10				95.78				
G11	148.8					228.3	169.2	
U12			163.3					
C13				98.76				
G14								
G15								
C16			196.8	98.88				
U17			162.0					
G18	147.3					233.4	169.8	
C19			196.9	99.38				
U20			162.4					
C21			196.3	97.60				
A22	225.4		215.6			230.8		

Supplementary Table 3.III (continued)

SLV BASE EXCHANGEABLE PROTON CHEMICAL SHIFTS (ppm)								
	H1^b	H21	H22	H3^b	H41^b	H42^b	H61	H62
G1								
C2					8.49	6.53		
G3	12.48							
G4	11.24							
C5					8.21	6.74		
A6								
G7	13.82							
U8								
U9								
G10	12.23							
A11								
C12					8.21	6.35		
U13								
A14								
C15					8.13	6.94		
U16				13.35				
G17	12.52							
U18				11.95				
C19					8.39	6.58		
G20	13.00							
C21					8.16	6.72		

Supplementary Table 3.III (continued)

SLV BASE NITROGEN CHEMICAL SHIFTS (ppm)							
	N1 ^c	N2	N3 ^c	N4 ^d	N6	N7 ^e	N9
G1							
C2			197.0	97.92			
G3	147.3						
G4	145.0						
C5			196.5	98.92			
A6	219.8		214.4				
G7	148.4						
U8							
U9							
G10	146.5						
A11	223.8		212.4			221.5	
C12				97.26			
U13							
A14	225.7		217.9			233.5	
C15			197.3	98.02			
U16			161.5				
G17	147.2						
U18			158.9				
C19			196.2	98.32			
G20	147.6						
C21			197.0	97.53			

a. The NMR data were collected at 25°C under standard buffer conditions (10 mM d₁₁-Tris pH 7.0, 50 mM NaCl, 5 mM MgCl₂, 0.05 mM NaN₃, 10% D₂O).

b. ± 0.01 ppm

c. ± 0.1 ppm for uridines and guanines, ± 0.5 ppm for cytosines, ± 0.2 ppm for adenines

d. ± 0.06 ppm

e. ± 0.2 ppm

Supplementary Table 3.IV : Chemical shift assignments of non-exchangeable ^1H and ^{13}C for the SLIDs2/SLV complex.^a

SLIDS2 NON-EXCHANGEABLE PROTON CHEMICAL SHIFTS (ppm)										
(ppm +/- 0.02)										
	H1'	H2'	H3'	H4'	H5'	H5''	H2	H5	H6	H8
G1	5.87	4.83	4.64	4.45	3.97	4.13				8.13
A2	5.99	4.67	4.77				7.40			7.99
G3	5.66	4.37	4.48							7.30
C4	5.47	4.50						5.18	7.62	
A5	5.92	4.71	4.68				6.99			7.97
G6	5.53	4.36	4.37							7.12
C7	5.39	4.32	4.42					4.78	7.16	
A8	6.04	3.95	4.88				8.23			8.40
U9	5.73	4.66						5.11	7.25	
C10	5.50	4.82						5.98	7.94	
G11	5.92	4.58	4.90							7.43
U12^b	5.52							5.49	7.98	
C13	5.63	4.71						5.42	7.86	
G14	5.53	4.50								7.57
G15	5.23	4.40								7.16
C16	5.52	4.26						5.23	7.83	
U17	5.51	4.56						5.41	7.92	
G18	5.79	4.45	4.64							7.75
C19	5.45	4.31						5.19	7.78	
U20	5.50	4.48						5.38	7.92	
C21	5.52	4.36	4.47					5.58	7.74	
A22	5.91	4.03	4.28	4.23			7.30			8.02

Supplementary Table 3.IV (continued)

SLIDS2 CARBON CHEMICAL SHIFTS (ppm)										
(ppm +/- 0.2)										
	C1'	C2'	C3'	C4'	C5'	C2	C5	C6	C8	
G1	93.0	75.2	73.9	84.5	62.4				138.7	
A2	92.8	75.5	72.6			153.3			139.8	
G3	92.5	75.2	72.5						135.7	
C4	93.3	75.4					97.4	140.4		
A5	92.9	75.6	72.8			152.4			139.2	
G6	92.7	75.3	72.6						135.5	
C7	94.4	75.4	72.0				96.8	140.0		
A8	92.0	77.0	72.2			156.0			139.9	
U9	93.7	74.7					102.1	142.5		
C10	94.4	75.7					98.7	142.1		
G11	92.7	75.2	72.1						136.7	
U12^b	93.4-94						103.7	142.3		
C13	93.3	75.5					97.9	141.6		
G14	92.5	75.3							136.0	
G15	93.6	74.6							135.8	
C16	94.1	75.3					96.6	141.5		
U17	93.5	75.1					103.4	141.7		
G18	92.7	75.4	72.7						136.3	
C19	93.8	75.5					96.8	141.2		
U20	93.4	75.2					103.1	141.6		
C21	93.8	75.5	72.4				97.7	141.0		
A22	91.8	77.8	70.2	83.9		154.3			140.1	

Supplementary Table 3.IV (continued)

SLV NON-EXCHANGEABLE PROTON CHEMICAL SHIFTS (ppm)										
(ppm +/- 0.01)										
	H1'	H2'	H3'	H4'	H5'	H5''	H2	H5	H6	H8
G1	5.61	4.48		4.18						7.90
C2	5.62	4.63						5.19	7.88	
G3	5.81	4.56	4.62							7.60
G4	5.77	4.60	4.37							7.21
C5	5.46	4.46						5.33	7.65	
A6	5.95	4.75					7.05			7.98
G7	5.56	4.40								7.16
U8	5.57	4.11	4.36	4.35				4.82	7.60	
U9	5.51	4.69						5.95	8.01	
G10	5.76	4.99	4.75	4.50						8.06
A11	6.03	4.61	4.94	4.46			7.73			7.57
C12	5.25	4.46	4.60	4.40				4.90	6.95	
U13	6.26	4.77	5.18	4.61	4.24	4.34		6.02	8.07	
A14	5.41	4.93		4.60			7.62			8.36
C15	5.28	4.11	4.44	3.81				5.19	7.78	
U16	5.52	4.63						5.37	7.86	
G17	5.81	4.63								7.61
U18	5.42	4.08						5.47	7.81	
C19	5.59							5.61	7.96	
G20	5.69	4.03								7.57
C21	5.66	4.47	4.11					5.26	7.52	

Supplementary Table 3.IV (continued)

SLV CARBON CHEMICAL SHIFTS (ppm)										
(ppm +/- 0.2)										
	C1'	C2'	C3'	C4'	C5'	C2	C5	C6	C8	
G1	93.2	73.5		84.2					138.6	
C2	93.7	75.1					97.1	141.3		
G3	92.4	75.8	73.1						136.1	
G4	93.3	75.4	72.9						136.6	
C5	93.6	75.3					97.6	140.2		
A6	92.7	75.6				152.6			139.3	
G7	92.5	75.2							135.6	
U8^c	93.8	76.6	72.4	82.8			101.8	141.4		
U9	93.9	74.8					103.7	143.2		
G10	94.3	75.1	72.1	83.6					139.9	
A11	93.2	75.8	71.9	82.4		153.4			140.7	
C12	94.8	75.6	74.3	82.0			98.4	140.3		
U13	89.2	75.2	78.6	87.1	67.8		105.6	144.8		
A14	95.3	73.3		84.4		153.5			142.1	
C15	94.3	75.2	72.0	81.6			96.8	141.7		
U16	93.2	75.0					103.2	141.6		
G17	92.9	75.3							136.2	
U18	93.6	75.6					103.8	140.9		
C19	93.4						97.8	142.0		
G20	93.2	75.4							136.0	
C21	93.0	75.4	69.3				97.4	141.1		

a. The NMR data were collected at 25°C under standard buffer conditions (10 mM d₁₁-Tris pH 7.0, 50 mM NaCl, 5 mM MgCl₂, 0.05 mM NaN₃, 10% D₂O).

b. The U12¹H1' resonance was assigned with an error of ± 0.07 ppm and the U12¹C1' resonance is within 93.4-94.0 ppm due to signal overlap.

c. The U8^VC5 and H5 resonances were assigned with errors of ± 0.01 ppm for H5 and ± 0.4 ppm for C5.

Supplementary Table 3.V : Summary of intermolecular NOEs at the kissing-loop junction of the SLIDs2/SLV complex.

SLIDs2/SLV	G10 ^V	A11 ^V	C12 ^V	A14 ^V	C15 ^V
C10 ^I				NH ₂ / H2	NH ₂ / H1'
				NH ₂ / H1'	H5 / H1'4'
				H5 / H2	H6 / H1'2'4'
					H1' / H4'
G11 ^I		H1 / H2	H1 / NH ₂	H1 / H2	NH ₂ / H1'
			H1 / H5	NH ₂ / H1'	
				H1' / H1'	
U12 ^I	H3 / H1	H3 / H2	H3 / NH ₂		
			H3 / H1'		
C13 ^I	NH ₂ / H1	H1' / H2			
	H5 / H1	NH ₂ / H2			
G14 ^I	H8 / H1				
	H1' / NH ₂				

REFERENCES:

1. Lacroix-Labonte, J., Girard, N., Lemieux, S. and Legault, P. (2012) Helix-length compensation studies reveal the adaptability of the VS ribozyme architecture. *Nucleic Acids Res*, **40**, 2284-2293.
2. Campbell, D.O. and Legault, P. (2005) NMR structure of the Varkud satellite ribozyme stem-loop V RNA and magnesium-ion binding from chemical-shift mapping. *Biochemistry*, **44**, 4157-4170.
3. Campbell, D.O., Bouchard, P., Desjardins, G. and Legault, P. (2006) NMR structure of Varkud satellite ribozyme stem-loop V in the presence of magnesium ions and localization of metal-binding sites. *Biochemistry*, **45**, 10591-10605.

4 Caractérisation thermodynamique de l'interaction *kissing-loop* I/V et de l'activation du substrat

Mise en contexte

Les deux études précédentes ont permis d'approfondir notre compréhension de la reconnaissance du substrat par le ribozyme VS, dont les détails structuraux définissant l'interaction *kissing-loop* I/V. Certaines données ont soulevé un intérêt particulier, notamment la diminution drastique d'activité du ribozyme observée suite à la délétion du U700 (chapitre 2) et le réseau d'interactions à l'interface de l'interaction I/V qui va au-delà des trois paires de bases prédites (chapitre 3). Nous avons d'ailleurs proposé à la fin du chapitre 2 que le squelette ribose phosphate du nucléotide U700 soit impliqué dans la liaison du substrat. De plus, la complexité des interactions caractérisées à l'interface de l'interaction I/V dans le chapitre 3 suggère que celles-ci soient associées à une grande stabilité. Une étude thermodynamique complémentaire a d'ailleurs été amorcée en parallèle avec l'étude structurale (chapitre 3) pour évaluer l'affinité de l'interaction *kissing-loop* I/V et comprendre l'effet des différentes variantes *SLV* (chapitre 2) et *SLI* (chapitre 3) étudiées précédemment.

Cette troisième étude, effectuée par *ITC*, a permis de mesurer les paramètres thermodynamiques (K_D , n , ΔH , ΔS et ΔG) caractérisant l'activation et la reconnaissance du substrat. Ceci a été réalisé en utilisant différentes variantes du substrat *SLI*, soit dans une conformation *shiftable*, comme le substrat naturel, *preshifted* ou encore contenant une hélice double brin et différentes variantes de *SLV*, notamment celles contenant des modifications du nucléotide U700. Cette étude a tout d'abord permis de montrer que des substrats adoptant une conformation *preshifted* présentent une meilleure affinité (K_D) pour *SLV* que des substrats *shiftable*. D'ailleurs, nous avons démontré que le coût énergétique impliqué dans le réarrangement structural du substrat, c'est-à-dire du *helix shift*, est de 1.8-3 kcal/mol. Nous avons aussi noté que l'affinité de l'interaction *kissing-loop* *SLI/SLV* formée avec un substrat *preshifted* est particulièrement élevée, environ 7-8 kcal/mol plus stable qu'un duplexe d'ARN formé de la même séquence. Nous avons également étudié l'effet de plusieurs variantes de *SLV*, notamment la délétion du U700, sur l'affinité de l'interaction I/V. Les effets de ces variations sur l'affinité de

l'interaction *SLI/SLV* sont similaires à ceux observés précédemment sur l'activité du ribozyme (chapitre 2). De plus, nous avons observé que la stabilité de l'interaction *kissing-loop* I/V en présence de la tige-boucle *SLV* seule est similaire à celle obtenue en présence du domaine catalytique entier (domaines II à VI). Ceci suggère que la contribution énergétique de la reconnaissance du substrat par le ribozyme VS repose principalement sur l'interaction *kissing-loop* I/V. Nous avons également mesuré les paramètres thermodynamiques de deux autres interactions *kissing-loop*, celles du *DIS* du VIH-1 et du H3-18 du virus de la leucémie murine de Moloney (*Moloney murine leukemia virus, MMLV*). Ensuite, nous avons comparé la stabilité de ces interactions (I/V du ribozyme VS, *DIS* du VIH-1 et H3-18 de *MMLV*) avec d'autres interaction *kissing-loop* dont la structure et les paramètres thermodynamiques sont connus afin de mieux comprendre les règles qui définissent la stabilité des interactions *kissing-loop*.

Dans ce troisième article, j'ai essentiellement réalisé toutes les expériences : la synthèse et la purification des ARN, la mise au point des expériences d'*ITC*, l'enregistrement, ainsi que l'analyse des données. J'ai bénéficié de l'aide d'Étienne Beaulieu, étudiant d'été au laboratoire, d'Élyse Benoit et Dominique Chaussé pour la préparation de certains ARN. L'analyse des résultats a été effectuée conjointement avec le Dr Legault. J'ai écrit la première version de l'article qui a ensuite été révisée par le Dr Legault.

A remarkably stable kissing-loop interaction defines
substrate recognition by the *Neurospora* Varkud
Satellite ribozyme

Patricia Bouchard and Pascale Legault*

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

This research was originally published in RNA journal:

RNA. 2014 Sep;20(9):1451-64. doi: 10.1261/rna.046144.114. Epub 2014 Jul 22.

ABSTRACT

Kissing loops are tertiary structure elements that often play key roles in functional RNAs. In the *Neurospora* VS ribozyme, a kissing-loop interaction between the stem-loop I (SLI) substrate and stem-loop V (SLV) of the catalytic domain is known to play an important role in substrate recognition. In addition, this I/V kissing-loop interaction is associated with a helix shift in SLI that activates the substrate for catalysis. To better understand the role of this kissing-loop interaction in substrate recognition and activation by the VS ribozyme, we performed a thermodynamic characterization by isothermal titration calorimetry using isolated SLI and SLV stem-loops. We demonstrate that preshifted SLI variants have higher affinity for SLV than shiftable SLI variants, with an energetic cost of 1.8-3 kcal/mol for the helix shift in SLI. The affinity of the preshifted SLI for SLV is remarkably high, the interaction being more stable by 7-8 kcal/mol than predicted for a comparable duplex containing three Watson-Crick base pairs. The structural basis of this remarkable stability is discussed in light of previous NMR studies. Comparative thermodynamic studies reveal that kissing-loop complexes containing 6-7 Watson-Crick base pairs are as stable as predicted from comparable RNA duplexes, however those with 2-3 Watson-Crick base pairs are more stable than predicted. Interestingly, the stability of the SLI/ribozyme complexes is similar to that of SLI/SLV complexes. Thus, the I/V kissing loop interaction represents the predominant energetic contribution to substrate recognition by the trans-cleaving VS ribozyme.

INTRODUCTION

Kissing-loop interactions are widespread elements of RNA tertiary structures that result from base pairing between the terminal loops of two hairpins. Their formation is central to many RNA-mediated processes, such as antisense recognition, plasmid replication control, retroviral dimerization, translation initiation and ribosomal frameshifting [1-5]. One of the best-studied kissing-loop interactions occurs during retroviral genome dimerization in HIV-1, where it forms between palindromic base sequences of the dimer initiation site [3, 6, 7]. Although kissing loops form long-range interactions between different RNAs or remote domains within the same RNA, they also play important roles in stabilizing the complex architecture of functional RNA elements, such as tRNAs [8] riboswitch aptamers [9-14], and ribozymes [15-18]. Thus, kissing-loop motifs play crucial roles in RNA structure and function; however only a few of these have been thermodynamically characterized.

Like several other functional RNAs, the *Neurospora* VS ribozyme depends on the formation of a kissing-loop interaction. Early work on the self-cleaving VS ribozyme revealed that the catalytic domain (VS Rz) recognizes its substrate (stem-loop I) through a kissing-loop interaction that is stabilized by magnesium ions [Figure 4.1A; [19, 20]]. This interaction involves three Watson-Crick (W-C) base pairs between residues 630, 631 and 632 of stem-loop I (SLI) and residues 697, 698 and 699 of stem-loop V (SLV). It is required for proper folding [21, 22] as well as for the efficient phosphodiester bond cleavage between G620 and A621 and the reverse ligation reaction [20, 23]. Interestingly, formation of the I/V kissing-loop interaction involves a well-characterized conformational change in SLI from an unshifted to a shifted conformation (Figure 4.1B; [21, 23-26]) and either the catalytic domain or an isolated SLV can produce this helix shift within the SLI substrate [23]. Through the SLI helix shift, the I/V kissing-loop interaction activates the substrate for catalysis, since ribozymes containing SLI variants that stabilized the unshifted conformation are inactive, whereas those that allow or favor the shifted conformation (preshifted substrates) are active [23]. Formation of the I/V kissing-loop interaction also facilitates docking of the cleavage site internal loop of SLI

with the A730 loop of SLVI [21, 22, 27]. Our current understanding is that the internal loops of SLI and SLVI must precisely interact to create the active site, where the nucleobases of G638 and A756 play key roles in the general acid-base cleavage mechanism [28-36]. Thus, the I/V kissing-loop interaction is important for several steps of the VS ribozyme cleavage reaction, including substrate recognition, activation and docking for catalysis.

Until recently, structural insights into the I/V kissing-loop interaction were provided by three-dimensional models of the full ribozyme [22, 37-39] as well as high-resolution NMR structures of free SLV [40, 41] and of the unshifted and shifted conformations of SLI [24-26]. As part of our efforts to characterize the I/V kissing-loop interaction, we previously reported a structure-function study that addressed the role of loop V residues in SLI substrate recognition [42]. One intriguing observation is that, although any base can replace the extruded U700 loop residue, deletion of this nucleotide substantially reduces ribozyme cleavage activity [42], possibly by affecting substrate binding. More recently, we performed NMR studies of several SLI/SLV complexes formed with RNA hairpins derived from the VS ribozyme and determined the NMR structure of an SLI/SLV kissing-loop complex [43]. These NMR studies provide clear structural evidence for the helix shift in SLI upon SLV binding. In addition, the NMR structure of an SLI/SLV complex confirms that both the SLI and SLV loops adopt U-turn structures, which facilitate formation of the three expected W-C base pairs between SLI and SLV and additional hydrogen bonding and base stacking interactions at the kissing-loop junction (Figure 4.1C-D). Complementary thermodynamic investigations of the I/V kissing-loop interaction are needed to evaluate the affinity of the SLI/SLV interaction in light of the NMR structure and to help understand the effect of SLI and SLV sequence variations on substrate recognition and catalysis.

Binding studies have been previously carried out using various SLI substrates and a trans-cleaving VS ribozyme [44]. Although these studies revealed the importance of the shifted SLI conformation in facilitating binding to the ribozyme [44], it used a

ribozyme system stabilized by an artificial intermolecular stem and thus yielded K_D values that do not directly reflect the stability of the I/V kissing loop [44]. In the present work, we use isothermal titration calorimetry (ITC; [45, 46]) to thermodynamically characterize the I/V kissing-loop interaction formed between isolated SLI and SLV hairpins. As a result, our study provides K_D values as well as energetic contributions directly associated with formation of the I/V kissing-loop interaction. In addition, since most of the SLI/SLV complexes investigated here have been structurally characterized by NMR spectroscopy [42, 43], the thermodynamic parameters can be directly related to structural features of SLI/SLV complexes. In particular, we examine the effect of kinetically-characterized SLV sequence variations on the thermodynamic parameters associated with the I/V kissing-loop interaction. We also investigate the thermodynamic contribution of the helix shift in SLI on the stability of SLI/SLV complexes using several shiftable, preshifted and double-stranded SLI variants. Finally, using non-cleavable SLI substrates, we compare the affinity of SLI/SLV complexes with that of SLI/ribozyme complexes to evaluate the relative contribution of the I/V kissing-loop interaction to substrate recognition in the VS ribozyme.

RESULTS

Thermodynamic characterization of the VS ribozyme I/V kissing-loop interaction

To better understand substrate recognition by the VS ribozyme, we determined the thermodynamic parameters associated with the I/V kissing-loop interaction by ITC using two isolated stem-loops derived from the VS ribozyme: the SLI substrate and the SLV receptor (Figure 4.2). Biophysical characterization of the I/V kissing-loop interaction using isolated RNA stem-loops is justified by the fact that an isolated SLV receptor hairpin is sufficient to bind an isolated SLI substrate and cause the same helix shift in SLI that is observed with the full VS ribozyme catalytic domain [21, 42, 43]. To further simplify the thermodynamic characterization of the I/V kissing-loop interaction, the initial SLI substrate was designed to adopt only an active preshifted conformation (SLIps, Figure 4.2A; [23, 44]). The same SLIps sequence was previously employed for kinetic characterization of VS ribozyme variants [42]. Using NMR spectroscopy and

native gel electrophoresis, we previously verified that both SLIps and SLV RNAs form stable hairpins under conditions used for ITC studies [42, 43]. NMR spectroscopy studies also indicate that SLIps and SLV form a stable 1:1 kissing-loop complex in the presence of MgCl₂ [43]. ITC experiments were first carried out at 25°C in a buffer containing 20 mM MgCl₂ (Figure 4.3). Analysis of duplicate ITC experiments for formation of this 1:1 SLIps/SLV complex yields the following average thermodynamic values: $K_D = 0.24 \pm 0.01 \mu\text{M}$; $\Delta H = -22.0 \pm 0.1 \text{ kcal/mol}$; $-T\Delta S = 12.9 \pm 0.1 \text{ kcal/mol}$; and $\Delta G_{ITC} = -9.04 \pm 0.03 \text{ kcal/mol}$ (Table 4.1).

Magnesium dependence of the I/V kissing-loop interaction

Since Mg²⁺ ions are essential for formation of the I/V kissing-loop interaction, ITC experiments were performed at 25°C under varying concentrations of MgCl₂ to examine the Mg²⁺-ion dependence of this interaction and determine conditions where the Mg²⁺-ion concentration is saturated. A plot of the dissociation constants (K_D) against the MgCl₂ concentration indicates that the affinity of SLV for the SLIps substrate increases (decrease of K_D) with increasing Mg²⁺-ion concentration and reaches a plateau around 20 mM MgCl₂ (Figure 4.4). Thus, subsequent ITC experiments were carried out in a buffer containing 20 mM MgCl₂.

The Mg²⁺ dependence of SLIps/SLV binding can also be analyzed as:

$$\left(\frac{\partial \ln(K_A)}{\partial \ln[Mg^{2+}]} \right) = \Delta\Gamma_{2+},$$

where $\Delta\Gamma_{2+}$ is the apparent interaction coefficient that represents the net uptake or release of divalent cations associated with the interaction [47, 48]. Although the plot of $\ln(K_A)$ versus $\ln[Mg^{2+}]$ is generally linear within a certain range of Mg²⁺ concentration, it is not expected over the full range of Mg²⁺ concentration [47-49], as observed here for the SLIps/SLV complex (Figure 4.4B). Thus, our data is best fit by a second order polynomial equation:

$$\ln(K_A) = -0.1651(\ln[MgCl_2])^2 - 0.8162\ln[MgCl_2] + 14.748$$

The slope of this plot evaluated at the midpoint of the curve ($[\text{MgCl}_2] = 4.5 \text{ mM}$) was used to estimate an apparent $\Delta\Gamma_{2+}$ value of 0.97 [48]. Maximum and minimum values of 1.46 and 0.71, respectively, for $\Delta\Gamma_{2+}$ were derived by linear regression analysis using either the first four points or the last four points of the $\ln(K_A)$ versus $\ln[\text{MgCl}_2]$ plot (Figure 4.4B). The derived $\Delta\Gamma_{2+}$ values are termed "apparent" $\Delta\Gamma_{2+}$ because the experimental conditions, particularly the use of Tris buffer and the absence of monovalent salts, were selected to be compatible with those used for NMR structural characterization [41-43], but do not allow determination of accurate $\Delta\Gamma_{2+}$ values [48]. Nevertheless, this analysis indicates that there is an approximate net uptake of one Mg^{2+} ion associated with formation of the SLI/SLV complex in the concentration range between 0.5 mM and 40 mM MgCl_2 .

Effect of SLV sequence variations on the stability of the SLI/SLV complex

As part of a previous study, VS ribozyme variants were kinetically characterized to evaluate the contribution of U700 and other SLV loop residues to SLI recognition [42]. This study revealed that replacing U700 by A, C, or G does not significantly affect the ribozyme cleavage activity, whereas U700 deletion dramatically impairs this activity [42]. In order to further address the role of U700 in SLI substrate recognition, we investigated the effect of U700 deletion and other SLV sequence modifications on the thermodynamic stability of the SLI/SLV complex. ITC experiments were performed at 25°C in a buffer containing 20 mM MgCl_2 with SLIps and different SLV variants (Figure 4.2B). It was previously verified by native gel electrophoresis and NMR spectroscopy that under these conditions, these SLV variants adopt a stable hairpin conformation [42]. As expected, all SLV variants bound to the SLIps substrate with an approximate stoichiometry (n) of 1 (Table 4.1), except for the C699G variant for which no binding could be detected under our ITC conditions.

Changing the U700 base to a C (U700C) or changing the U-A closing base pair to a G-A (U695G) yields K_D values of 0.81 μM and 0.29 μM , respectively, which are on the same order of magnitude than that obtained with the wild-type SLV ($K_D = 0.24 \mu\text{M}$,

Table 4.1). In contrast, an SLV variant containing a U700 deletion (Δ U700) displays a substantial decrease in affinity for SLIps with a K_D value of 50 μ M (Table 4.1). For the C699G variant, thermodynamic parameters could not be derived even at high concentrations of RNA (1 mM C699G SLV in the syringe and 100 μ M of SLIps in the cell), and a $K_D > 200$ μ M was estimated, assuming that the poor quality of the titration curve reflects a situation where $c < 1$ ($c = (\text{concentration of RNA in the cell})/K_D$) [50, 51]. Accordingly, this SLV variant was not expected to stably bind to the SLIps substrate, since the C699G variation disrupts one of the three W-C base pairs at the kissing-loop junction [20, 43].

To compare the effect of SLV sequence variations on SLI binding with their effect on SLI cleavage by the VS ribozyme, we tabulated the relative K_A values $[(K_A)_{WT}/(K_A)_{VAR} = (K_D)_{VAR}/(K_D)_{WT}]$ obtained in the present study with the relative k_{cat}/K_M values $[(k_{cat}/K_M)_{WT}/(k_{cat}/K_M)_{VAR}]$ obtained previously [Table 4.2; [42]]. In addition, the contributions of modified residues to the free energy of the I/V kissing-loop interaction were derived from the relative K_A values $\{\Delta\Delta G_{ITC} = RT\ln[(K_A)_{WT}/(K_A)_{VAR}];$ Table 4.2}. For the U700C and U695G variants, the $(K_A)_{WT}/(K_A)_{VAR}$ ratios (3.4 and 1.2, respectively) are similar to the $(k_{cat}/K_M)_{WT}/(k_{cat}/K_M)_{VAR}$ ratios (3.7 and 2.7, respectively). In both cases, the small ratios (< 4) indicate that neither the affinity of the I/V kissing-loop interaction nor ribozyme activity is significantly affected by these modifications. In addition, the $\Delta\Delta G_{ITC} (\leq 0.7$ kcal/mol) indicates that no substantial loss of free energy of binding resulted from these SLV modifications (Table 4.2). For the Δ U700 variant, the large decrease in ribozyme activity corresponds to a $(k_{cat}/K_M)_{WT}/(k_{cat}/K_M)_{VAR}$ ratios of 140, which is on the same order of magnitude than the respective $(K_A)_{WT}/(K_A)_{VAR}$ ratio of 210 (Table 4.2). The $\Delta\Delta G_{ITC}$ value of this variant (3.2 kcal/mol) indicate that the deletion of U700 results in an energetic cost of ~ 3 kcal/mol for SLI binding (Table 4.2).

Effect of SLI sequence variations on the stability of the SLI/SLV complex

In the VS ribozyme, the kissing-loop interaction between SLI and SLV is accompanied by a structural rearrangement in SLI [21, 23], in which SLI changes from an inactive (unshifted) to an active (shifted) conformation. The cytosines C635 to C637

(C15-17, Figure 4.1B) in helix Ib shift their position in the stem, excluding C634 (C14, Figure 4.1B) from the helix and rearranging the cleavage site internal loop [23]. To better understand the importance of this helix shift in substrate recognition, we evaluated its energetic contribution by ITC using six different SLI variants (Figure 4.2A). We selected shiftable variants that convert from an inactive to an active conformation upon SLV binding in the presence of Mg^{2+} ions (SLIsb and SLIsb Δ Ia, Figure 4.2A). We also selected preshifted variants that adopt a single active shifted conformation in their free form (SLIps and SLIps Δ Ia, Figure 4.2A). Similar shiftable and preshifted variants that lack stem Ia (Δ Ia) were also investigated to evaluate the energetic cost of stem Ia in the formation of the SLI/SLV complex, since it was previously shown that removal of stem Ia increases ribozyme cleavage activity [52-55]. Finally, we also investigated SLI variants containing a non-cleavable double-stranded stem (SLIds1 and SLIds2, Figure 4.2A). The secondary structure of these six SLI variants in their free form was confirmed by native gel electrophoresis and NMR spectroscopy [43]. NMR spectroscopy was also used to demonstrate formation of stable kissing-loop complexes between several SLI variants and SLV (SLIsb Δ Ia/SLV, SLIps/SLV, SLIps Δ Ia/SLV, SLIds1/SLV and SLIds2/SLV; [43]), which include the three proposed W-C base pairs. Moreover, these studies provide direct evidence that shiftable SLI variants but not the preshifted or double-stranded variants undergo the proposed helix shift upon SLV binding [43].

The ITC results reveal higher affinities (lower K_D) of SLV for preshifted SLI variants compared to shiftable SLI variants (SLIps versus SLIsb and SLIps Δ Ia versus SLIsb Δ Ia, Table 4.3). Among the shiftable SLI variants, a lower affinity for SLV is observed in the presence of stem Ia (SLIsb) than in its absence (SLIsb Δ Ia, Table 4.3). In contrast, the presence of stem Ia does not significantly decrease the affinity of SLV to a preshifted SLI variant (SLIps versus SLIps Δ Ia, Table 4.3). Not surprisingly, SLI variants containing a double-stranded stem bind SLV with affinities that closely match those of the preshifted SLI variants (SLIds1 and SLIds2 versus SLIps and SLIps Δ Ia, Table 4.3). In terms of free energy (Table 4.3), these results show that SLV forms more stable complexes with preshifted SLI variants than with shiftable SLI variants. An energetic

cost of 1.8 kcal/mol is associated with complex formation of shiftable versus preshifted SLI variants in the absence of stem Ia, and this can be directly attributed to the structural rearrangement of the shiftable substrate upon complex formation. The presence of stem Ia further reduces by 1 kcal/mol the thermodynamic stability of the SLI/SLV complex formed with a shiftable substrate, but has essentially no effect on the stability of complexes formed with preshifted substrates. These results also indicate that the stability of SLI/SLV complexes formed with several non-shiftable SLI variants (SLIps, SLIpsΔIa, SLIids1, SLIids2) is influenced neither by the sequence nor the structure of the helical domain adjoining the terminal loop.

The VS ribozyme I/V kissing-loop interaction is more stable than predicted

In this study, we obtained K_D values ranging from 0.24 to 0.71 μM for binding of non-shiftable SLI variants to SLV (Tables 4.1 and 4.3), indicating that small variations in experimental conditions (e.g. switching the titrant, addition of 50 mM NaCl) do not significantly affect the measured K_D value (< 3 fold). The observed K_D values reflect the thermodynamic stability of the I/V kissing-loop interaction, which appears to be remarkably high given that only three W-C base pairs are formed at the kissing-loop junction [20, 43]. We compared the free energy values estimated at 37°C obtained for formation of SLI/SLV complexes with these non-shiftable SLI variants (ΔG_{ITC} of -7.7 to -8.5 kcal/mol; Table 4.4) to the predicted ΔG value for formation of a comparable RNA duplex at 37°C according to nearest-neighbor calculations [56, 57]. The SLI/SLV kissing-loop complex was investigated under high-salt conditions (20 mM MgCl_2) that are comparable to those used for establishing nearest-neighbor calculations. The three proposed W-C base pairs at the I/V kissing-loop junction form a short A-form helix, as shown in the NMR structure [43]. According to the nearest-neighbor thermodynamic calculation [56], the predicted ΔG (ΔG_{pre}) for the short RNA duplex $\frac{\text{GUC}}{\text{CAG}}$ composed of the same sequence as the I/V kissing-loop interaction $\frac{\text{G}_{630} \text{U}_{631} \text{C}_{632}}{\text{C}_{699} \text{A}_{698} \text{G}_{697}}$ is -0.5 kcal/mol under standard conditions (Table 4.4). The NMR structure of the SLIids2/SLV complex also reveals 5'- and 3'-stacking interactions, which are known to contribute to helix stability

(Figure 4.1C). They include a 5'-stacking interaction between C629 and G630 as well as 3'-stacking interactions between C632 and G633 in SLI and between C699 and A701 in SLV [43]. If we incorporate these stacking interactions to our thermodynamic calculations using nearest-neighbor interactions involving 3'- and 5'-unpaired nucleotides [58], a ΔG_{pre} value of -4.2 kcal/mol is obtained for the resulting duplex (Table 4.4). Nonetheless, this ΔG_{pre} value still differs by 3.5-4.3 kcal/mol from the range of ΔG_{ITC} values associated with formation of the I/V kissing-loop interaction. Thus, the I/V kissing-loop interaction is significantly more stable than a three-base-pair RNA duplex of the same sequence, even when stacking interactions with 5' and 3' nucleotides are included in the free energy calculation of the RNA duplex.

Comparison with thermodynamic parameters of other kissing-loop interactions

A few other known kissing-loop complexes have been thoroughly investigated both structurally and thermodynamically (Figure 4.5), and we examined their stability by comparing their experimental ΔG values obtained from previous studies (ΔG_{exp}) with their ΔG value predicted according to nearest-neighbor calculations (ΔG_{pre}). With exception of the Moloney Murine Leukemia virus (MMLV) H3-18 complex, the kissing-loop complexes that we considered have been thermodynamically investigated under high-salt conditions, either at 1 M NaCl or high Mg^{2+} concentrations (5-10 mM $MgCl_2$; Table 4.4), that are comparable to those used for establishing nearest-neighbor calculations [56, 57]. The ΔG_{pre} values were calculated as for the SLI/SLV complex based on available high-resolution structures (Figure 4.5). For the HIV-1 TAR/TAR*, ColE1 RNAIi/RNAIii, and HIV-1 DIS complexes, the experimental ΔG values (ΔG_{exp}) and the ΔG_{pre} can differ considerably (0.3-4.1 kcal/mol) when only base-pair stacking interactions are taken into account for calculation of the ΔG_{pre} value, but generally differ less (< 2 kcal/mol) when 5'/3'-stacking interactions are taken into account (Table 4.4). In contrast, for the MMLV complex, which contains only two W-C base pairs at the kissing-loop junction (Figure 4.5), $\Delta G_{pre} - \Delta G_{exp} = 7.7$ kcal/mol when 5'/3'-stacking interactions are taken into account (Table 4.4), indicating a more stable interaction than predicted by the thermodynamic model. In summary, except for MMLV H3-18 kissing-

loop interaction and the VS ribozyme SLI/SLV interaction, the values of ΔG_{exp} and ΔG_{pre} obtained for the other kissing-loop interactions investigated are similar.

In order to verify that our ITC conditions were not responsible for the unexpectedly large $\Delta G_{pre}-\Delta G_{exp}$ value obtained for the I/V kissing-loop interaction, we performed control ITC experiments with the HIV-1 TAR/TAR* and MMLV H3-18 complexes (Figure 4.5), under the same conditions than those used for the SLIps/SLV complex (Table 4.1). For the HIV-1 TAR/TAR* complex, we derived a ΔG_{ITC} of -9.8 kcal/mol (Table 4), which is similar to the value of ΔG_{exp} of -9.3 kcal/mol derived from a previous study (Table 4.4). For the MMLV H3-18 complexes, the use of our ITC conditions yielded a ΔG_{ITC} of -5.0 kcal/mol (Table 4.4), which indicates a weaker interaction than previously established under lower salt conditions [ΔG_{exp} of -6.5 kcal/mol; [59]]. However, with this new experimental ΔG_{exp} value (ΔG_{ITC}), the difference with ΔG_{pre} is still fairly high (6.2 kcal/mol). Thus, both the MMLV H3-18 and I/V kissing-loop interactions are substantially more stable than predicted from thermodynamic calculations, and their remarkable stability is not due to exceptional experimental conditions.

Thermodynamic characterization of SLI/ribozyme complexes

To characterize the thermodynamic parameters for SLI substrate recognition in the context of the *trans*-cleaving VS ribozyme (VS Rz, Figure 4.1), we selected two SLI substrate variants (Figure 4.2A) in which a 2'-O-methyl was introduced at nucleotide G-1 to prevent cleavage by the ribozyme, either a shiftable substrate with no stem Ia (SLIsb Δ Ia-1m) or a preshifted substrate with stem Ia (SLIps-1m). ITC experiments were performed at 25°C and 20 mM MgCl₂ using each of these non-cleavable SLI analogs with either SLV or the VS Rz to extract the thermodynamic parameters associated with the resulting complexes (Table 4.5). The K_D values associated with SLV binding to the non-cleavable SLI analogs are within 3 fold of those obtained for SLV binding to the corresponding unmodified SLI variants (Tables 4.3 and 4.4; SLIsb Δ Ia-1m/SLV versus SLIsb Δ Ia/SLV and SLIps-1m/SLV versus SLIps/SLV), indicating that the 2'-O-methyl

modification does not significantly affect the binding affinity. Such small changes in K_D values could be simply due to differences in experimental conditions (the titrant is an SLI derivative for Table 4.3, but SLV for Table 4.4) as described above for the SLIps/SLV interaction. More importantly, the K_D values for binding of the non-cleavable SLI analogs to the VS Rz are within 1.5 fold of those obtained for binding of the same SLI analogs to the free SLV (Table 4.5; SLI_{sb}ΔIa-1m/VS Rz versus SLI_{sb}ΔIa-1m/SLV and SLI_{ps}-1m/VS Rz versus SLI_{ps}-1m/SLV). These results indicate that binding of these SLI variants to either SLV or VS Rz involves similar free energies. Although the 2'-O-methyl modification used for our studies prevents interaction(s) with the 2'-OH at the cleavage site, SLI docking interactions with the VS ribozyme does not contribute significantly to the ground state of the SLI/ribozyme complexes tested here. As a result, the I/V kissing-loop interaction is the key determinant of SLI substrate recognition in the VS ribozyme.

DISCUSSION

The I/V kissing-loop interaction is remarkably stable

In the present study, ITC experiments were performed to determine the thermodynamic parameters associated with the I/V kissing-loop interaction of the VS ribozyme. Interestingly, the ITC data revealed that formation of the SLI/SLV complex with several SLI variants is remarkably high, more stable by 7-8 kcal/mol than predicted for a comparable RNA duplex with three Watson-Crick base pairs. As discussed below, this is likely due to extensive stacking at the kissing-loop junction (≥ 4 kcal/mol), the stabilizing effect of Mg^{2+} ions, the U-turn loop structures of SLI and SLV, and the ribose-phosphate of the extruded U700 in SLV.

The high stability of the SLI/SLV complex with three W-C base pairs at the kissing-loop junction can not be predicted from thermodynamic calculations even when both base-pair stacking and 5'/3'-stacking interactions are considered in estimating ΔG_{pre} ($\Delta G_{pre} - \Delta G_{exp} = \sim 4$ kcal/mol), and this is not a result of exceptional experimental conditions. For other kissing-loop interactions, namely the HIV-1 TAR/TAR*, the ColE1 RNAIi/RNAIi and the HIV-1 DIS complexes, the ΔG_{exp} and ΔG_{pre} values differ by less

than 2 kcal/mol when both base-pair stacking and 5'/3'-stacking interactions are considered in estimating ΔG_{pre} . Thus, the rules governing RNA duplex stability can be useful for estimating the stability of these kissing-loop complexes containing 6-7 W-C base pairs, when both base-pair stacking and 5'/3'-stacking interactions are taken into account. In contrast, similar thermodynamic calculations were not successful in predicting the high stability of the MMLV H3-18 complex ($\Delta G_{pre} - \Delta G_{exp} = 7-8$ kcal/mol), which contains only two W-C base pairs at the kissing-loop junction [59-61]. Thus, it appears that kissing-loop complexes with 2-3 Watson-Crick base pairs are generally more stable than predicted. Clearly, other factors than those computed by the thermodynamic model must contribute to the high stability of both the MMLV H3-18 complex and the VS ribozyme I/V kissing-loop complex.

Extensive stacking at the I/V junction

The compact network of hydrogen-bonding and stacking interactions observed at the kissing-loop junction in the NMR structure of the SLI_{ds}2/SLV complex [43] may not be justly accounted for by the thermodynamic model. In this structure, the kissing-loop junction contains the three predicted W-C base pairs forming a short helix, but also two base triples, one at each end of the short helix. For the two base triples, the thermodynamic model takes into account stacking interactions as if they were from unpaired nucleotides at the end of a helix, but does not consider the base triples as a whole. Both base triples are part of an intricate structure that extends from the kissing-loop junction toward the stems of SLI and SLV to create a continuously stacked structure. This continuous stacking and several hydrogen bonds not computed by the thermodynamic model likely contribute to the remarkable stability of the SLI/SLV complex.

Role of Mg²⁺ in the I/V kissing-loop interaction

Our study confirms that Mg²⁺ is required for formation of a stable I/V kissing-loop interaction, as previously inferred from chemical probing data [20, 22, 55, 62]. The K_D of the SLI_{ps}/SLV interaction could not be measured by ITC at MgCl₂ concentrations below 0.5 mM because of the lower affinity of the complex at lower MgCl₂

concentrations. Starting at 0.5 mM MgCl₂, we observed a decrease in K_D with increasing Mg²⁺-ion concentration to reach a plateau around 20 mM MgCl₂. In the interval between 0.5 mM and 40 mM MgCl₂, a net uptake of ~1 Mg²⁺ ion is associated with formation of the SLI/SLV complex.

It was previously estimated from the variation of T_m values with Mg²⁺ concentration that formation of kissing-loop complexes derived from the HIV-1 TAR-TAR*, the HIV-1 DIS and the ColE1 RNAI-RNAII interactions also involves net uptakes of 1-2 Mg²⁺ [49, 63, 64]. Such Mg²⁺ uptakes are likely necessary to neutralize the high charge density created by phosphate clusters at the loop-loop junction of these complexes [7, 47, 65-68]. In the I/V kissing-loop complex, Mg²⁺ ions likely play a similar role, since several phosphate clusters are found at the kissing-loop junction [43]. From detailed NMR structural characterization of the free SLI and SLV loops [25, 40, 41] as well as the SLI/SLV complex [43], we know that formation of the SLI/SLV complex leads to a more ordered and compact SLI loop, but creates only minor changes in the SLV loop structure [43]. Thus, there is a net increase in electronegative charge density upon complex formation that is due to both SLI folding into a more compact loop structure and formation of intermolecular interactions at the kissing-loop junction [43]. Our current structural understanding of Mg²⁺ interactions associated with formation of the I/V kissing-loop complex is limited to the free forms of each stem-loops; whereas Mg²⁺ do not grossly affect the disordered loop I [25], it induces a conformation change in the SLV loop, stabilizing a canonical U-turn fold with four specific Mg²⁺ binding sites [40, 41]. Three of these four preferential sites identified by NMR in the free SLV loop interact with phosphate groups in the major groove and likely play a similar role in the SLI/SLV complex. Future studies aimed at localizing Mg²⁺ binding sites in the SLI/SLV complex should help better understand how specific Mg²⁺ binding sites at the kissing-loop junction contribute to the stability of the I/V interaction.

Importance of U turns in the I/V kissing-loop interaction

U-turn loop structures facilitate other RNA interactions, such as codon-anticodon interactions [69, 70] and antisense RNA regulation [71, 72]. It has been known for some

times that trinucleotides bind more tightly to a complementary tRNA anticodon than would be expected for complementary RNA oligonucleotides [69]. In pioneering binding experiments with a kissing-loop complex formed between the complementary anticodons of yeast tRNA^{Phe} and *Escherichia coli* tRNA^{Glu}, a K_A of $3.6 \times 10^5 \text{ M}^{-1}$ and ΔH of -25 kcal/mol were measured at 25°C and 10 mM MgSO₄ (ΔG_{exp} at 37°C of -6.88 kcal/mol), which is more stable by 7.7 kcal/mol than predicted for the corresponding RNA duplex when only base-pair stacking interactions are taken into account [ΔG_{pre} of +0.81 kcal/mol; [69]]. Interestingly, this $\Delta G_{pre} - \Delta G_{exp}$ value is very similar to that of 7.2-8.0 kcal/mol obtained for the SLI/SLV complexes (Table 4.4). In the same study, three potential sources of affinity enhancement were identified: 1) the "loop constraint" or closure of the anticodon sequences into hairpin loops; 2) the stabilizing effect of dangling ends; and 3) the extra stability provided by modified nucleotides that contribute to 3'-stacking. This "loop constraint" that was identified is likely due to the U-turn loop conformation that may promote intermolecular binding in several ways. In the SLI/SLV complex, it promotes stacking of the interacting bases in a near A-form geometry compatible with W-C base pairing. Such stacking is present in free SLV, but not in free SLI and thus is increased upon formation of the SLI/SLV complex. In addition, the U turn conformation allows continuous stacking between bases of the kissing-loop and adjoining stems in the SLI/SLV complex. Furthermore, the U-turn structure may promote kissing-loop formation by retracting the phosphate backbone away from the interface, thereby reducing electrostatic repulsion. Binding of Mg²⁺ to the U-turn loop structure, as observed for the U-turn of SLV [41] and anticodon loops [73, 74], likely helps further reduce electrostatic repulsion at the loop-loop junction, thereby contributing to the remarkable stability of kissing-loop interactions involving U-turn structures.

Importance of U700 in the I/V kissing-loop interaction

Although the role of loop V nucleotides involved in the U-turn (U₆₉₆G₆₉₇A₆₉₈) and the I/V kissing-loop interaction (G₆₉₇A₆₉₈C₆₉₉) are well understood, that of U700 has remained more elusive [42]. NMR structural studies of SLV demonstrate that upon addition of Mg²⁺, the largest conformational change in loop V involves U700, which

becomes completely extruded from the SLV loop fold [40, 41]. In our previous enzymatic study of SLV loop variants [42], modification of U700 by a C, A, or G had negligible effects on ribozyme *trans* cleavage activity (k_{cat}/K_M), whereas deletion of U700 significantly reduced the k_{cat}/K_M . Similarly, ITC investigations of the I/V kissing-loop interaction using isolated stem-loops demonstrate that a U700C modification does not significantly affect the affinity of the SLI/SLV interaction, whereas deletion of U700 significantly reduces the affinity of this interaction. Thus, the role of U700 is confined to the I/V kissing-loop interaction and likely not other aspects of the cleavage reaction.

Both thermodynamic and kinetic data indicate that it is the ribose phosphate of the extruded U700 residue and not its nucleobase that contributes to the stability of the I/V kissing-loop interaction. In agreement with these results, the U700 base is disordered in the structure of the SLI/SLV complex and does not appear to form any stable interactions with SLI [43]. Interestingly, a comparison between the thermodynamic parameters of the wild-type SLV and the Δ U700 variant indicates that the free energy loss of ~ 3 kcal/mol is due to both unfavorable entropic (1.3 kcal/mole) and enthalpic changes (1.9 kcal/mol; Table 4.1). As previously suggested, U700 could provide an entropic advantage for substrate recognition by increasing the loop dynamics and/or by serving as a spacer to increase the flexibility of the loop [42]. Such loop dynamics may contribute the necessary plasticity within the SLV loop structure to optimize the intermolecular contacts upon formation of the kissing-loop interaction [43]. In addition, the ribose-phosphate backbone of U700 possibly contributes to the free enthalpy of the SLI/SLV complex through hydrogen bonding between one of the two non-bridging phosphate oxygens of U700 and an amino proton of A701, as these atoms are relatively close to another (3.0 ± 0.4 Å in the 20 lowest-energy structures) in the NMR structures [Figure 4.1D; [43]]. This interaction with the phosphate of U700 likely stabilizes the A14^V-U8^V-C10^I base triple at the kissing-loop junction (Figure 4.1D). The ribose-phosphate of U700 could also contribute to the enthalpic stability of the SLI/SLV complex through its interaction with a Mg²⁺ ion [42]. This possibility is supported by our previous observations that the structure of the free SLV loop is stabilized by Mg²⁺ ions and that a Mg²⁺ ion directly coordinates the 5'-phosphate of U700 [41]. Although

Mg²⁺-ion binding sites have not been mapped within the SLI/SLV complex, the SLV loop within this complex adopts a structure that resembles its free conformation in the presence of Mg²⁺ ions [43]. Taking into account these considerations, it is likely that the Mg²⁺-U700 5'-phosphate interaction somehow contributes to the stability of the complex. Thus, the ribose-phosphate of U700 likely contributes in several different ways to the remarkable stability of the I/V kissing-loop interaction.

Preshifting of the SLI substrate enhances its affinity for SLV

The SLI/SLV complex is highly stable when formed with preshifted substrates, but less so when formed with shiftable substrates. Using SLI substrates lacking stem Ia, we measured an increase of one order of magnitude in K_D for the SLI/SLV complex when replacing a preshifted SLI (SLIps Δ Ia) by a comparable shiftable SLI (SLIsb Δ Ia), indicating that shifting the SLI helix in this context entails an energetic cost of 1.8 kcal/mol. These results are consistent with a previous study in which K_D values were measured for several VS ribozyme/SLI complexes [44]. Two such complexes were formed using so-called "opened-loop" SLI products that are similar to SLIps Δ Ia and SLIsb Δ Ia because they are devoid of stem Ia and contain a G627A mutation in the loop (equivalent to A7 in SLIps Δ Ia and SLIsb Δ Ia; Figure 4.2A). Using these opened-loop SLI RNAs, an increase of one order of magnitude in K_D was measured for the VS ribozyme/SLI interaction when replacing the preshifted SLI by the comparable shiftable SLI [44]. The concordance with our results is particularly noteworthy given that these VS ribozyme/SLI complexes were artificially stabilized by an intermolecular helix [44]. Moreover, it confirms that only SLV and not other parts of the VS ribozyme contribute to shifting the SLI helix upon SLI binding.

The energetic cost for shifting the SLI helix further increases when forming the SLI/SLV complex in the presence of stem Ia. Using shiftable SLI substrates, we observed a 5-fold decrease in K_D ($\Delta\Delta G = 1.0$ kcal/mol) for the SLI/SLV complex when replacing a shiftable SLI containing stem Ia (SLIsb) by a comparable SLI lacking stem Ia (SLIsb Δ Ia). In contrast, no significant change in K_D values was observed for SLI/SLV complexes formed by preshifted SLI substrates with (SLIps) or without (SLIps Δ Ia) stem

Ia. These results are consistent with chemical modification and mutational studies in which the stem Ia was shown to be an inhibitory element of shiftable substrates for catalysis of self-cleaving ribozymes [53-55, 75]. However, it was not possible to conclude from these studies if the disruption of helix Ia affected formation of the SLI/SLV interaction or other aspect(s) of catalysis. Particularly, mutations that disrupt stem Ia also increase the size of the linker between SLI and the rest of the ribozyme, and longer linkers have been associated with faster cleavage of self-cleaving ribozymes [53]. Our thermodynamic results clearly indicate that the presence of stem Ia destabilizes the SLI/SLV interaction of shiftable SLI substrates, but has no effect on preshifted SLI substrates. With the SLI and SLV variants used here, formation of the SLI/SLV complex with a preshifted SLI substrate containing stem Ia (SLIps) is more energetically favorable by ~3 kcal/mol than with a shiftable SLI substrate containing stem Ia. In summary, stem Ia clearly contributes to stabilizing the unshifted inactive conformation and can affect cleavage rates of VS ribozymes when formation of the I/V interaction is rate limiting, as is likely the case for the wild-type self-cleaving ribozyme.

The I/V kissing-loop interaction defines substrate recognition by the VS ribozyme

The strong correlation observed between the thermodynamic and kinetic properties of several SLV variants supports the concept that the I/V kissing-loop interaction is a functionally modular unit for substrate recognition by the VS ribozyme. Structurally, the kissing-loop interaction forms an independent unit that is distinct from residues involved in cleavage chemistry [22, 37-39]. For cleavage to occur, the SLI internal loop must dock with the catalytic domain to form the active site [28-36], and we now know that such docking does not contribute significantly to ground-state substrate binding at least with the *trans* SLI substrate/ribozyme system investigated here. Rather, the I/V kissing-loop interaction is thermodynamically dominant for substrate recognition by the VS ribozyme. The remarkable stability of this interaction is modulated by Mg^{2+} concentration and SLI conformational shifting such that formation of the I/V kissing-loop can consequently directly regulate the VS ribozyme cleavage reaction.

MATERIALS AND METHODS

Preparation of RNA for ITC studies

The RNA hairpins used in this study, including the SLI and SLV variants (Figure 4.2) as well as the TAR, TAR* and Moloney Murine Leukemia virus (MMLV) H3-18 RNAs (Figure 4.5), were synthesized in vitro with the T7 RNA polymerase using synthetic DNA templates (Integrated DNA Technologies, IA) and purified as previously described [40]. Two SLI variants (SLIps and SLIsb Δ Ia; Figure 4.2A) were also chemically synthesized (Integrated DNA Technologies, IA) with a 2'-O-methyl modification at position G-1 (SLIps-1m and SLIsb Δ Ia-1m) and purified as previously described [40]. The Avapl ribozyme, herein named VS Rz (Figure 4.1A), was synthesized from plasmid pAvaplVS derived from plasmid pAvapl [76] to allow synthesis of the ribozyme with its substrate at the 3'-end. The transcription reaction was performed in vitro at 37°C using the pAvaplVS plasmid linearized with *EcoRI*, the T7 RNA polymerase, unlabeled NTPs and 25 mM MgCl₂. After 3 h, 15 mM MgCl₂ was added and the reaction mixture was incubated for 1 h to allow self-cleavage of the VS substrate. The cleavage reaction was stopped by addition of 40 mM EDTA, and then the VS Rz was purified as previously described [42]. Purified RNA samples were transferred to ITC buffers by use of Amicon Ultra-4 3K ultrafiltration devices: ITC buffer A (10 mM Tris pH 7.0) for experiments with SLIps and SLV variants (Table 4.1 and Figure 4.4), TAR, TAR* and MMLV H3-18 RNAs (Table 4.4) or ITC buffer B (10 mM Tris pH 7.0 and 50 mM NaCl) for experiments with SLV and SLI variants (Table 4.3) and VS Rz (Table 4.5). Prior to use, the RNAs were refolded by heating (2 min at 95°C for RNA hairpins or at 37°C for the VS Rz) and cooling in ice water. The RNA samples were subsequently transferred to ITC buffer supplemented with MgCl₂, generally 20 mM MgCl₂ although other MgCl₂ concentrations were used to investigate the Mg²⁺-ion dependence of the I/V kissing-loop interaction. The pH of the Tris buffer was adjusted at room temperature. RNA concentrations were estimated from UV absorbance at 260 nm (A_{260}) using extinction coefficients (ϵ_{260}) determined by taking the sum of ϵ_{260} values for each nucleotide [77] within the RNA sequence and multiplying

this sum by a hyperchromicity factor corresponding to the ratio of A_{260} values before and after Nuclease P1 digestion [78, 79].

Isothermal titration calorimetry (ITC) studies

ITC measurements were performed at 25°C using a VP-ITC or ITC₂₀₀ Microcalorimeter (GE Healthcare). All solutions used with the VP-ITC were degassed at 22°C for 10-15 min before each experiment. For ITC titrations with SLV variants (Table 4.1), the concentrations of the SLV titrants in the injection syringe were 100 μ M (wild-type, U700C and U695G SLV RNAs), 300 μ M (Δ U700 and C699G SLV RNAs) or 1 mM (Δ U700 and C699G SLV RNAs), and the concentration of SLIps in the cell was ten times lower than that of the titrants, except for the titration with 300 μ M of Δ U700 and C699G SLV RNAs in which 20 μ M of SLIps was used in the cell. For ITC titrations with SLI variants (Table 4.3), the concentrations of the SLI titrants in the syringe were 100 μ M (SLIps Δ Ia and SLIds1 RNAs), 200 μ M (SLIsb Δ Ia, SLIps, SLIsb and SLIds2 RNAs) or 600 μ M (SLIsb RNA) and the concentration of SLV in the cell was ten times lower than that of the titrants. For ITC titrations with the VS ribozyme catalytic domain (VS Rz; Table 4.5), the concentration of VS Rz in the syringe was 100 μ M and the concentration of non-cleavable SLI (SLIsb Δ Ia-1m and SLIps-1m; Table 4.5) in the cell was 10 μ M. For ITC titration studies of the TAR/TAR* kissing-loop interaction (Table 4.4), the concentration of TAR RNA in the syringe was 50 μ M and the concentration of TAR* RNA in the cell was 5 μ M. For ITC studies of the MMLV H3-18 kissing-loop dimer (Table 4.4), dilution experiments were performed with 500 μ M MMLV H3-18 hairpin in the syringe and buffer only in the cell.

Most ITC titrations were collected with c values ≥ 1 ($c = [\text{RNA in the cell}] / K_D$; c values between 2 to 850), from which accurate thermodynamic parameters were extracted [50, 51]. For titration data collected with c values < 1 ($c = 0.3$ and $c = 0.6$ for complexes SLIps/SLVC699G and SLIsb/SLV, respectively), the stoichiometry (n) was fixed to 1, such that only K_A and ΔG could be accurately determined [51]. For each titration experiment, a control experiment was performed under the same conditions,

except that only buffer was used in the cell. The control experiment was subtracted from the titration experiment in order to eliminate the effect of diluting the titrant. Data collected for each titration experiment were then fit to a single binding site model or to a dimer-monomer dissociation model (for MMLV H3-18 only) using equations from the ORIGIN software package version 7.0 (OriginLab Corporation). The quality of the fits was obtained from the square of the correlation coefficient (R^2), and in all cases R^2 was ≥ 0.99 . The values of n , K_A (or $K_D = 1/K_A$) and ΔH were directly obtained from these fits. The free energy (ΔG_{ITC}) was calculated as $\Delta G_{ITC} = -RT \ln K_A$, where R denotes the gas constant ($R = 1.987 \text{ cal}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$) and T is the temperature. The values of $T\Delta S$ were derived from that of ΔG and ΔH , given that $\Delta G = \Delta H - T\Delta S$. At least two ITC experiments were collected for each measurement. The reported values and errors on the thermodynamic parameters are, respectively, the average and standard deviations from these multiple measurements.

ACKNOWLEDGMENT

We thank Elyse Benoit, Dominique Chaussé, Étienne Beaulieu and Pierre Dagenais for RNA preparation, Ao Yang for help with ITC data analysis, and James G. Omichinski for critical reading of the manuscript. This work was supported by the Canadian Institutes of Health Research (CIHR) to P.L. [MOP-86502]. P.B. was supported by a Frederick Banting and Charles Best Canada Graduate Scholarship Ph.D. scholarship from CIHR and graduate scholarships from the Université de Montréal. The ITC instruments were purchased with funds from the Natural Sciences and Engineering Research Council of Canada. P.L. holds a Canada Research Chair in Structural Biology and Engineering of RNA. Funding for open access charge: CIHR.

REFERENCES

1. Wagner, E.G. and R.W. Simons, *Antisense RNA control in bacteria, phages, and plasmids*. Annu Rev Microbiol, 1994. **48**: p. 713-42.

2. Guo, L., E.M. Allen, and W.A. Miller, *Base-pairing between untranslated regions facilitates translation of uncapped, nonpolyadenylated viral RNA*. Mol Cell, 2001. **7**(5): p. 1103-9.
3. Brunel, C., et al., *RNA loop-loop interactions as dynamic functional motifs*. Biochimie, 2002. **84**(9): p. 925-44.
4. Baranov, P.V., et al., *Programmed ribosomal frameshifting in decoding the SARS-CoV genome*. Virology, 2005. **332**(2): p. 498-510.
5. Brantl, S., *Regulatory mechanisms employed by cis-encoded antisense RNAs*. Curr Opin Microbiol, 2007. **10**(2): p. 102-9.
6. Skripkin, E., et al., *Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro*. Proc Natl Acad Sci USA, 1994. **91**(11): p. 4945-9.
7. Ennifar, E., et al., *Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site*. Nat Struct Biol, 2001. **8**(12): p. 1064-8.
8. Quigley, G.J. and A. Rich, *Structural domains of transfer RNA molecules*. Science, 1976. **194**: p. 796-806.
9. Batey, R.T., S.D. Gilbert, and R.K. Montange, *Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine*. Nature, 2004. **432**(7015): p. 411-5.
10. Serganov, A., et al., *Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs*. Chem Biol, 2004. **11**(12): p. 1729-41.
11. Lemay, J.F., et al., *Folding of the adenine riboswitch*. Chem Biol, 2006. **13**(8): p. 857-68.
12. Blouin, S. and D.A. Lafontaine, *A loop-loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control*. RNA, 2007. **13**(8): p. 1256-67.
13. Garst, A.D., et al., *Crystal structure of the lysine riboswitch regulatory mRNA element*. J Biol Chem, 2008. **283**(33): p. 22347-51.

14. Serganov, A., L. Huang, and D.J. Patel, *Structural insights into amino acid binding and gene control by a lysine riboswitch*. Nature, 2008. **455**(7217): p. 1263-7.
15. 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.
16. Khvorova, A., et al., *Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity*. Nat Struct Biol, 2003. **10**(9): p. 708-12.
17. Chi, Y.I., et al., *Capturing hammerhead ribozyme structures in action by modulating general base catalysis*. PLoS Biol, 2008. **6**(9): p. e234.
18. Dufour, D., et al., *Structure-function analysis of the ribozymes of chrysanthemum chlorotic mottle viroid: a loop-loop interaction motif conserved in most natural hammerheads*. Nucleic Acids Res, 2009. **37**(2): p. 368-81.
19. 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.
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.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.
22. 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.
23. Andersen, A. and R.A. Collins, *Rearrangement of a stable RNA secondary structure during VS ribozyme catalysis*. Mol Cell, 2000. **5**: p. 469-478.
24. 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.

25. 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.
26. 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.
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. 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.
29. 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.
30. 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.
31. 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.
32. Zhao, Z.Y., et al., *Nucleobase participation in ribozyme catalysis*. J Am Chem Soc, 2005. **127**(14): p. 5026-7.
33. 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.
34. 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.
35. 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.
36. 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.

37. 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.
38. 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.
39. Lipfert, J., et al., *The complete VS ribozyme in solution studied by small-angle X-ray scattering*. Structure, 2008. **16**(9): p. 1357-67.
40. 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.
41. 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.
42. Bouchard, P., et al., *Role of SLV in SLI substrate recognition by the Neurospora VS ribozyme*. RNA, 2008. **14**: p. 736-748.
43. 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.
44. 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.
45. Reymond, C., M. Bisailon, and J.P. Perreault, *Monitoring of an RNA multistep folding pathway by isothermal titration calorimetry*. Biophys J, 2009. **96**(1): p. 132-40.
46. Salim, N.N. and A.L. Feig, *Isothermal titration calorimetry of RNA*. Methods, 2009. **47**(3): p. 198-205.
47. Draper, D.E., *RNA folding: thermodynamic and molecular descriptions of the roles of ions*. Biophys J, 2008. **95**(12): p. 5489-95.
48. Leipply, D., D. Lambert, and D.E. Draper, *Ion-RNA interactions: Thermodynamic analysis of the effects of mono- and divalent ions on RNA conformational equilibria*, in Methods Enzymol, 2009. p. 433-463.

49. Weixlbaumer, A., et al., *Determination of thermodynamic parameters for HIV DIS type loop-loop kissing complexes*. Nucleic Acids Res, 2004. **32**(17): p. 5126-33.
50. Wiseman, T., et al., *Rapid measurement of binding constants and heats of binding using a new titration calorimeter*. Anal Biochem, 1989. **179**(1): p. 131-7.
51. Turnbull, W.B. and A.H. Daranas, *On the value of c: can low affinity systems be studied by isothermal titration calorimetry*. J Am Chem Soc, 2003. **125**(48): p. 14859-66.
52. Zamel, R., et al., *Exceptionally fast self-cleavage by a Neurospora Varkud satellite ribozyme*. Proc Natl Acad Sci USA, 2004. **101**(6): p. 1467-72.
53. 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.
54. 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.
55. 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.
56. Xia, T., et al., *Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs*. Biochemistry, 1998. **37**(42): p. 14719-35.
57. Lorenz, C., N. Piganeau, and R. Schroeder, *Stabilities of HIV-1 DIS type RNA loop-loop interactions in vitro and in vivo*. Nucleic Acids Res, 2006. **34**(1): p. 334-42.
58. Serra, M.J. and D.H. Turner, *Predicting thermodynamic properties of RNA*. Methods Enzymol, 1995. **259**: p. 242-61.
59. Kim, C.H. and I. Tinoco, Jr., *A retroviral RNA kissing complex containing only two G-C base pairs*. Proc Natl Acad Sci USA, 2000. **97**(17): p. 9396-401.
60. Li, P.T., C. Bustamante, and I. Tinoco, Jr., *Unusual mechanical stability of a minimal RNA kissing complex*. Proc Natl Acad Sci USA, 2006. **103**(43): p. 15847-52.

61. Chen, A.A. and A.E. Garcia, *Mechanism of enhanced mechanical stability of a minimal RNA kissing complex elucidated by nonequilibrium molecular dynamics simulations*. Proc Natl Acad Sci USA, 2012. **109**(24): p. E1530-9.
62. 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.
63. 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.
64. Gregorian, R.S., Jr. and D.M. Crothers, *Determinants of RNA hairpin loop-loop complex stability*. J Mol Biol, 1995. **248**(5): p. 968-84.
65. 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.
66. 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 USA, 2008. **105**(27): p. 9210-5.
67. 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.
68. Lee, A.J. and D.M. Crothers, *The solution structure of an RNA loop-loop complex: the ColE1 inverted loop sequence*. Structure, 1998. **6**(8): p. 993-1005.
69. Grosjean, H., D.G. Soll, and D.M. Crothers, *Studies of the complex between transfer RNAs with complementary anticodons. I. Origins of enhanced affinity between complementary triplets*. J Mol Biol, 1976. **103**(3): p. 499-519.
70. Moras, D., et al., *Anticodon-anticodon interaction induces conformational changes in tRNA: yeast tRNA^{Asp}, a model for tRNA-mRNA recognition*. Proc Natl Acad Sci USA, 1986. **83**(4): p. 932-6.
71. Franch, T., et al., *Antisense RNA Regulation in Prokaryotes: Rapid RNA/RNA Interaction Facilitated by a General U-turn Loop Structure*. J Mol Biol, 1999. **294**(5): p. 1115-1125.

72. Franch, T. and K. Gerdes, *U-turns and regulatory RNAs*. Curr Opin Microbiol, 2000. **3**(2): p. 159-164.
73. Ogle, J.M., et al., *Recognition of cognate transfer RNA by the 30S ribosomal subunit*. Science, 2001. **292**(5518): p. 897-902.
74. Westhof, E., P. Dumas, and D. Moras, *Restrained refinement of two crystalline forms of yeast aspartic acid and phenylalanine transfer RNA crystals*. Acta Crystallogr A, 1988. **44 (Pt 2)**: p. 112-23.
75. 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.
76. 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.
77. Cavaluzzi, M.J. and P.N. Borer, *Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA*. Nucleic Acids Res, 2004. **32**(1): p. e13.
78. Zaug, A.J., C.A. Grosshans, and T.R. Cech, *Sequence-specific endoribonuclease activity of the Tetrahymena ribozyme: enhanced cleavage of certain oligonucleotide substrates that form mismatched ribozyme-substrate complexes*. Biochemistry, 1988. **27**(25): p. 8924-31.
79. Legault, P., *Structural studies of ribozymes by heteronuclear NMR spectroscopy*, 1995, University of Colorado at Boulder: Boulder.
80. Marky LA, Breslauer KJ., *Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves*. Biopolymers, 1987. **26**: p. 1601-1620.
81. Nair, T.M., D.G. Myszka, and D.R. Davis, *Surface plasmon resonance kinetic studies of the HIV TAR RNA kissing hairpin complex and its stabilization by 2-thiouridine modification*. Nucleic Acids Res, 2000. **28**(9): p. 1935-40.

TABLES

Table 4.I : ITC data for binding of SLV variants to a preshifted SLI substrate (SLIps).

Cell	RNAs Syringe	K_D (μM)	n	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)	ΔG_{ITC}^a (kcal/mol)
SLIps	WT SLV	0.24 ± 0.01	0.84 ± 0.01	-22.0 ± 0.1	12.9 ± 0.1	-9.04 ± 0.03
SLIps	U700C SLV	0.81 ± 0.04	0.83 ± 0.01	-21.5 ± 1.0	13.2 ± 1.0	-8.31 ± 0.03
SLIps	U695G SLV	0.29 ± 0.04	1.0 ± 0.1	-27.0 ± 0.5	18.0 ± 0.4	-8.93 ± 0.08
SLIps	ΔU700 SLV	50 ± 5	0.8 ± 0.1	-20.1 ± 0.09	14.19 ± 0.03	-5.87 ± 0.06
SLIps	C699G SLV ^b	> 200	1	-	-	> -5.05

a. $\Delta G_{ITC} = -RT\ln(K_A)$ at $T = 298.15$ K.

b. Given a low c value ($c < 1$), n was fixed to 1; in this case the ΔH and $-T\Delta S$ can not be accurately determined [51].

Table 4.II : Comparison of thermodynamic and kinetic data associated with the I/V kissing-loop interaction in the VS ribozyme.

Variant RNAs	Thermodynamic data		Kinetic data
	$\frac{(K_A)_{WT}^a}{(K_A)_{VAR}}$	$\Delta\Delta G_{ITC}^b$ (kcal/mol)	$\frac{(k_{cat}/K_M)_{WT}^c}{(k_{cat}/K_M)_{VAR}}$
WT SLV	1.0	0	1.0
U700C SLV	3.4	0.73	3.7
U695G SLV	1.2	0.10	2.7
Δ U700 SLV	2.1×10^2	3.2	1.4×10^2
C699G SLV	$\geq 8.4 \times 10^2$	> 4.0	2.7×10^2

a. Calculated from data reported in Table 4.1.

b. $\Delta\Delta G_{ITC} = RT\ln[(K_A)_{WT}/(K_A)_{VAR}]$ at $T = 298.15$ K; $\Delta\Delta G_{ITC}$ was used to compare the change in binding energy between various RNAs, these values reflect the difference in free energy observed as a result of each sequence variation.

c. Data taken from [42].

Table 4.III : ITC data for binding of SLI variants to SLV.

Cell	RNAs Syringe	K_D (μM)	n	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)	ΔG_{ITC}^a (kcal/mol)
SLV	SLIps	0.71 ± 0.04	0.84 ± 0.01	-24.9 ± 1.4	16.5 ± 1.4	-8.39 ± 0.04
SLV	SLIsb ^b	65 ± 4	1	-	-	-5.71 ± 0.03
SLV	SLIps Δ Ia	0.63 ± 0.02	1.13 ± 0.02	-24.2 ± 0.7	15.8 ± 0.7	-8.46 ± 0.02
SLV	SLIsb Δ Ia	12.5 ± 1.1	1.09 ± 0.08	-21.4 ± 2.7	14.7 ± 2.7	-6.69 ± 0.05
SLV	SLIds1	0.66 ± 0.04	1.03 ± 0.02	-25.2 ± 0.6	16.8 ± 0.5	-8.43 ± 0.03
SLV	SLIds2	0.41 ± 0.04	1.07 ± 0.01	-27.8 ± 0.8	19.1 ± 0.7	-8.71 ± 0.06

a. $\Delta G_{ITC} = -RT\ln(K_A)$ at $T = 298.15$ K.

b. Given a low c value ($c < 1$), n was fixed to 1; in this case the ΔH and $-T\Delta S$ can not be accurately determined [51].

Table 4.IV : Comparison between predicted ΔG and experimental ΔG values (in kcal/mol) for several kissing-loop complexes.

Kissing-loop complex	Sequence context	ΔG_{exp}	Experimental Conditions	No 5'/3' stacking interactions ^a		With 5'/3' stacking interactions ^b	
				ΔG_{pre}	$\Delta G_{pre} - \Delta G_{exp}$	ΔG_{pre}	$\Delta G_{pre} - \Delta G_{exp}$
VS Ribozyme I/V	$\begin{array}{c} \text{C G U C G} \\ \text{A C A G} \end{array}$	-8.5	c[SLIps/SLV]		8.0		4.3
		-7.9	d [SLIds2/SLV]		7.4		3.7
		-7.8	d [SLIds1/SLV]	-0.5	7.3	-4.2	3.6
		-7.8	d [SLIps Δ Ia /SLV]		7.3		3.6
		-7.7	d [SLIps/SLV]		7.2		3.5
HIV-1 TAR/TAR*	$\begin{array}{c} \text{C U G G G A G} \\ \text{A G A C C C U} \end{array}$	-9.8	c	-9.0	0.8	-10.9	-1.1
		-9.3	e		0.3		-1.6
CoIE1 RNAIi/RNAIii	$\begin{array}{c} \text{G G A U G G U U C} \\ \text{C C U A C C A A G} \end{array}$	-11.6	f	-7.9	3.7	-13.5	-1.9
HIV-1 DIS	$\begin{array}{c} \text{C C G A C C A} \\ \text{A G G C U G G} \end{array}$	-13.5	g	-9.4	4.1	-12.2	1.3
MMLV H3-18	$\begin{array}{c} \text{A C G} \\ \text{G C A} \end{array}$	-5.0	c	+2.2	7.2	+1.2	6.2
		-6.5	h		8.7		7.7

a. The predicted ΔG values (ΔG_{pre}) at 37°C were calculated based on nearest-neighbor calculations and do not include the effect of 5' and 3' stacking nucleotides (see text; [58]).

b. The ΔG_{pre} values were calculated based on nearest-neighbor calculations and include the effect of 5' and 3' stacking nucleotides (see text; [56, 58]).

c. The ΔG_{exp} value for the VS Ribozyme SLIps/SLV complex was derived herein from ITC data collected at 25°C in 10 mM Tris pH 7.0 and 20 mM MgCl₂ (Table 4.1). The same conditions were also used herein for the HIV-1 TAR/TAR* ($K_D = 5.89 \pm 0.48$ nM, $\Delta H = -47.73$ kcal/mol at 298.15 K) and the MMLV H3-18 ($K_D = 0.134 \pm 0.009$ mM, $\Delta H = -12.62$ kcal/mol at 298.15 K) complexes. The K_A values at 37°C were first derived using the Van't Hoff equation. The ΔG_{exp} value were then derived at 37°C using $\Delta G_{exp} = -RT\ln(K_A)$.

d. The ΔG_{exp} value for these VS Ribozyme SLI/SLV complexes were estimated as in (c) at 37°C using thermodynamic values obtained herein from ITC data collected at 25°C in 10 mM Tris pH 7.0, 50 mM NaCl and 20 mM MgCl₂ (Table 4.3).

e. The K_A value at 37°C for the HIV-1 TAR/TAR* complex was estimated by using the Van't Hoff equation, a measured K_D of 14.5 nM obtained by SPR at 23°C in 20 mM HEPES pH 7.3, 20 mM sodium acetate, 140 mM potassium acetate and 10 mM magnesium acetate [63], and a ΔH value of -39.2 kcal/mol obtained by SPR in 20 mM HEPES pH 7.3, 20 mM sodium acetate, 140 mM potassium acetate and 1 mM magnesium acetate [63]. The ΔG_{exp} value was then calculated using $\Delta G_{\text{exp}} = -RT\ln(K_A)$ at $T = 310.15$ K.

f. The K_A value at 37°C for the Cole1 RNAIi-RNAIii was estimated by using a modified Van't Hoff equation ($K_A = (4/C_T) \cdot \exp(-\Delta H/R \cdot (1/T - 1/T_m))$; [80]) and the following experimental values previously obtained from UV melting curve experiments [64]: melting temperature (T_m) of 40°C, $\Delta H = -47.5$ kcal/mol and a total strand concentration (C_T) of 2 μM . The ΔG_{exp} value was then calculated using $\Delta G_{\text{exp}} = -RT\ln(K_A)$ at $T = 310.15$ K.

g. The ΔG_{exp} value for the HIV-1 DIS complex is the standard free energy at 37°C derived from UV melting curves obtained in 10 mM sodium cacodylate pH=6.8 and 1 M NaCl [49].

h. The ΔG_{exp} value for the MMLV H3-18 complex is the standard free energy at 37°C derived from UV melting curves obtained in 10 mM sodium phosphate pH=6.5 and 50 mM NaCl [59].

Table 4.V : ITC data for binding of non-cleavable SLI substrates to SLV and VS Rz.

Cell	RNAs Syringe	K_D (μM)	n	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)	ΔG_{ITC}^a (kcal/mol)
SLIps-1m	SLV	0.21 ± 0.01	0.95 ± 0.02	-32.5 ± 0.1	23.4 ± 0.1	-9.11 ± 0.04
SLIsb Δ Ia-1m	SLV	3.6 ± 0.3	0.86 ± 0.01	-32.9 ± 1.1	25.5 ± 1.1	-7.43 ± 0.06
SLIps-1m	VS Rz	0.30 ± 0.02	1.07 ± 0.05	-28.0 ± 0.6	19.2 ± 0.6	-8.90 ± 0.03
SLIsb Δ Ia-1m	VS Rz	4.56 ± 0.01	0.82 ± 0.04	-42.6 ± 1.6	35.3 ± 1.6	-7.286 ± 0.001

a. $\Delta G_{ITC} = -RT\ln(K_A)$ at $T = 298.15$ K.

FIGURES

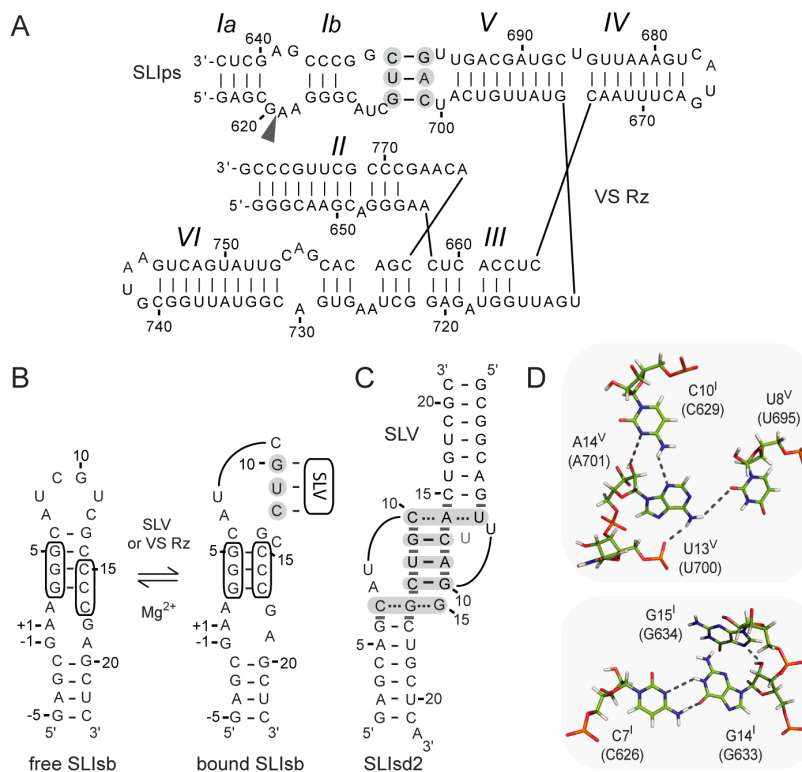


Figure 4.1 : Primary and secondary structures of the *Neurospora* VS ribozyme. (A) The catalytic domain of the VS ribozyme containing helical domains II-VI (VS Rz) and an SLI substrate (SLIps) containing stems Ia and Ib. The cleavage site is indicated by an arrowhead. The I/V kissing-loop interaction involves W-C base pairs (black lines) between shaded residues of SLI and SLV [19, 20]. (B) Formation of the I/V kissing-loop is accompanied by a structural rearrangement of the SLI substrate from an unshifted (free) to a shifted (bound) conformation. The cleavage site is between residues -1 and +1 of SLI_{sb}. (C)(D) Structural characteristics of the I/V kissing-loop based on the NMR structure of the SLI_{ds}2/SLV complex [43]. In (C), W-C and non-canonical base pairs are represented by solid and dashed lines, respectively, on the secondary structure of the complex, whereas stable base stacking at the kissing-loop junction is illustrated by grey rectangles. U13 is in grey to illustrate its extrusion from the loop V fold. In (D) hydrogen bonds within the base triples of the minimized averaged structure are represented by dotted lines.

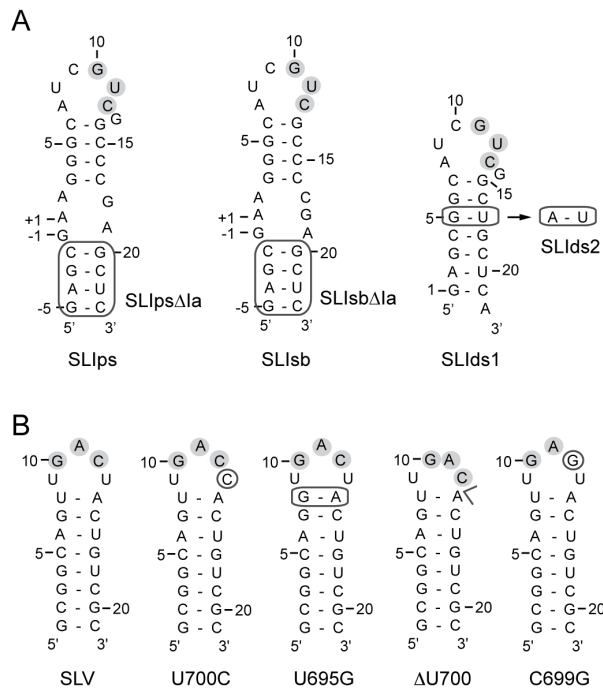


Figure 4.2 : Primary and secondary structures of SLI and SLV RNAs used in this study.

(A) The SLIps substrate RNA and related variants. SLIps Δ Ia and SLIsb Δ Ia are derivatives of SLIps and SLIsb, respectively, in which stem Ia is absent (boxed region). SLIsd2 is a G5A variant of SLIds1. The VS ribozyme cleavage site for SLIps, SLIps Δ Ia, SLIsb and SLIsb Δ Ia is between residues -1 and +1 of SLI. (B) The wild-type SLV RNA and related SLV variants. In (A) and (B), residues that participate in the three W-C base pairs at the I/V kissing-loop junction are shaded in grey.

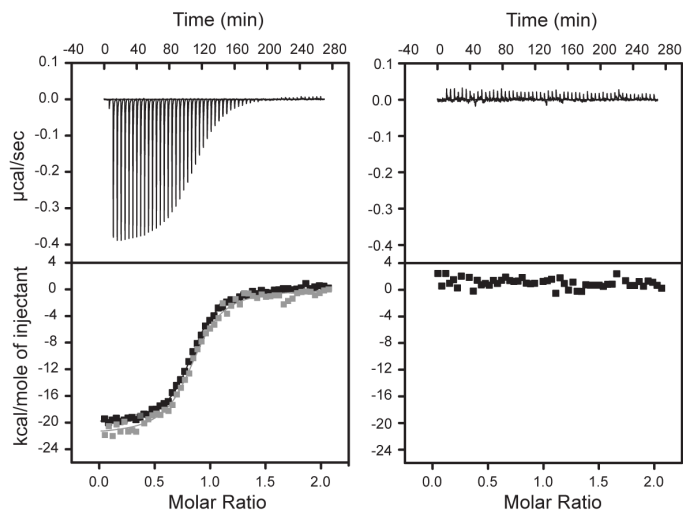


Figure 4.3 : ITC data analysis of wild-type SLV binding to SLIps at 20 mM MgCl_2 . The raw data of a representative titration experiment is shown on the left top panel and the corresponding control experiment (only buffer in the cell) is shown on the right top panel. The isothermal curves, which represent the integrated data of the above titration experiments, are shown on the bottom panels. On the left bottom panel, two isothermal curves of the titration experiment are shown, before (black) and after (grey) subtraction of the control experiment. They both yield the same K_D value ($0.24 \mu\text{M}$). In the titration experiment, $100 \mu\text{M}$ of wild-type SLV was used in the syringe and $10 \mu\text{M}$ of SLIps was used in the cell.

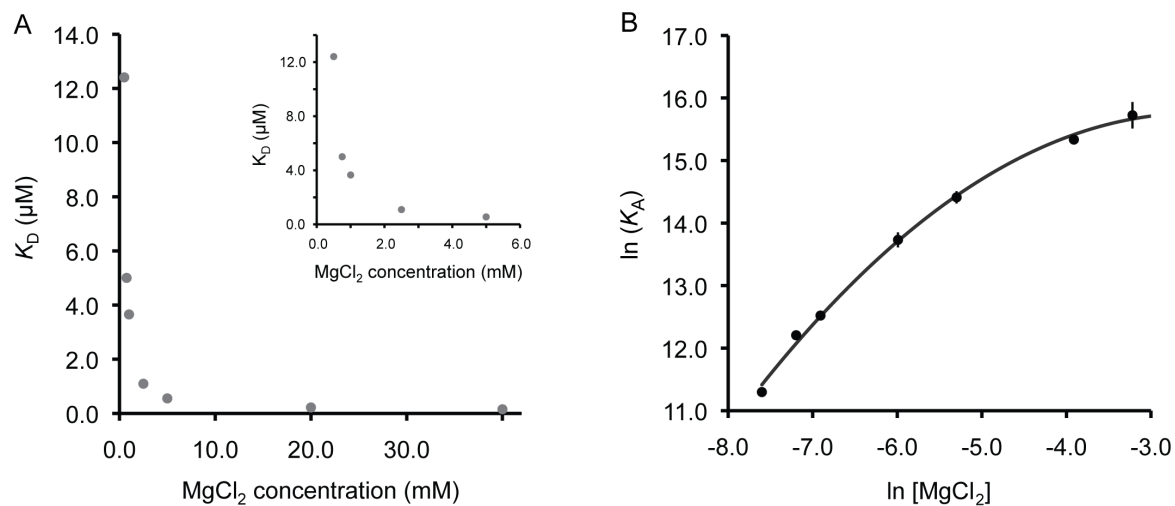


Figure 4.4 : Magnesium-ion dependence of the I/V kissing-loop interaction. (A) Dissociation constant (K_D) for the SLIps/SLV complex with respect to MgCl_2 concentration. The inset shows the same data on a different scale. (B) Dependence of $\ln(K_A)$ on $\ln[\text{MgCl}_2]$ for the SLIps/SLV complex. These data are best fit by a second order polynomial equation: $y = 0.1651x^2 - 0.8162x + 14.748$. In (A) and (B), error bars were calculated for each point from the standard deviations of two experiments, and are, in some cases, smaller than the data points.

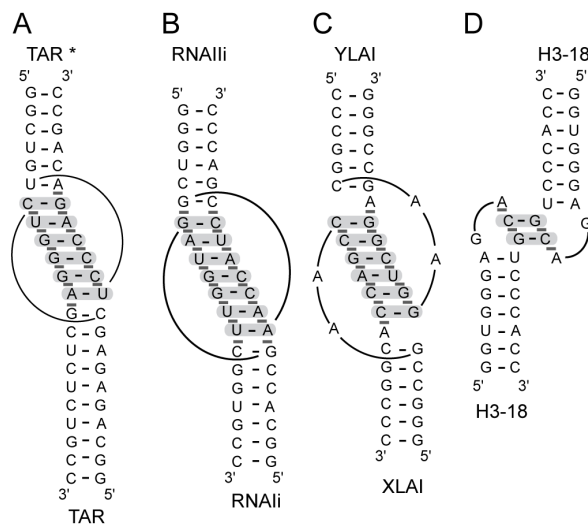


Figure 4.5 : Primary and secondary structures of other kissing-loop interactions considered in this study. (A) The HIV-1 TAR/TAR* kissing-loop complex used for the present study. Similar complexes were used in previous thermodynamic studies of the TAR/TAR* complex [63, 81]. (B) The ColE1 RNAIi/RNAIi complex previously thermodynamically characterized by UV melting curves [64]. (C) The HIV-1 DIS complex (XLAI/YLAI) previously investigated by UV melting studies [49]. (D) The MMLV H3-18 kissing-loop complex used here and in previous UV melting studies [59]. In A-D, interactions at the kissing-loop junction are highlighted by pale gray shading for W-C base pair and dark grey boxes for base stacking, according to previously determined three-dimensional structures of the HIV-1 TAR/TAR* (pdb code 1KIS.pdb [65]), ColE1 RNAIi/RNAIi (pdb code 1BJ2 [68]), HIV-1 DIS (pdb code 2B8R [7]), and MMLV H3-18 kissing-loop complexes (pdb code 1F5U [59]).

5 Discussion et perspectives

L'objectif global de cette thèse était de mieux comprendre la reconnaissance et l'activation du substrat chez le ribozyme VS de *Neurospora*. Pour ce faire, j'ai effectué une caractérisation structurale et thermodynamique de l'interaction *kissing-loop* I/V à l'aide de fragments d'ARN représentant les tige-boucles I et V dérivées du ribozyme VS.

Cette caractérisation, réalisée principalement par spectroscopie RMN et *ITC*, démontre que l'interaction *kissing-loop* I/V est un élément de structure complexe d'une stabilité étonnante qui est au centre de la reconnaissance du substrat. Tout d'abord, la complexité de la jonction I/V est représentée par l'organisation d'un réseau élaboré d'interactions qui va au-delà de la formation des trois paires de bases W-C prédites. En effet, l'interface I/V est caractérisée par l'interaction entre deux motifs *U-turn* et stabilisée par plusieurs couches d'interactions d'empilement de bases. Cet empilement provient, d'une part, des trois paires de bases W-C au centre de l'interaction I/V qui forment une courte hélice et, d'autre part, des deux triplets de bases localisés de chaque côté de ces trois paires de bases. Cet arrangement particulier à la jonction I/V permet de propager l'empilement de bases à partir du centre de l'interaction jusqu'aux bouts des tiges des domaines I et V. De cette façon, le complexe *SLI/SLV* présente une structure allongée dans laquelle l'empilement de bases est continu.

L'ensemble des ponts hydrogènes et de l'empilement de bases retrouvés à l'interface de l'interaction *kissing-loop* I/V contribue d'ailleurs à sa grande stabilité. En effet, nous avons démontré par *ITC* que la formation du complexe *SLI/SLV* est associée à une stabilité étonnante, représentant de 7 à 8 kcal/mol de plus que ce qui est prédit pour un modèle thermodynamique représentant un duplexe d'ARN équivalent formé de trois paires de bases. Bien que l'empilement de bases additionnel provenant des triplets de bases entourant les trois paires de bases W-C centrales explique une partie de la grande stabilité de l'interaction I/V (au moins 4 kcal/mol sur 8 kcal/mol), il ne permet pas d'expliquer la totalité de cette remarquable stabilité. Cette dernière s'explique clairement par plusieurs autres effets stabilisateurs ne pouvant être considérés par ce modèle thermodynamique, notamment la liaison de cations magnésium, les motifs *U-turn* des

boucles I et V, ainsi que de la contribution du squelette ribose-phosphate du nucléotide U700.

Le U700 est un nucléotide dont l'importance a été mise en évidence par nos données de cinétique enzymatique et d'*ITC*. Bien que ces données aient montré que l'importance du U700 repose sur son squelette ribose-phosphate, ce sont les données RMN qui ont permis de préciser le rôle de ce nucléotide. En effet, un des atomes d'oxygènes non pontants du U700 participe à un pont hydrogène avec un proton amino du A701, ce qui contribue probablement à la stabilité du triplet de bases A14^V-U8^V-C10^I (A701-U695-C629). Nous avons par la suite émis l'hypothèse que la partie ribose-phosphate du U700 soit également impliquée dans la liaison d'un cation magnésium important. Le groupement 5'-phosphate du U700 a d'ailleurs été identifié précédemment comme ligand pour un cation magnésium dans la boucle V seule [114]. Puis, étant donné que les structures de la boucle V seule et liée à *SLI* sont très similaires, il est possible qu'un cation magnésium lie aussi le groupement 5'-phosphate du U700 dans le contexte du complexe *SLI/SLV*.

L'ensemble des travaux présentés dans cette thèse a aussi mis en évidence le concept de modularité de l'interaction *kissing-loop* I/V. En fait, cette interaction forme un module indépendant dont les éléments structuraux sont distincts de ceux participant à la réaction enzymatique [92, 117, 120, 147]. Le concept de modularité a été renforcé grâce à l'excellente corrélation observée entre les données thermodynamiques obtenues avec les domaines I et V isolés et les données cinétiques obtenues avec le domaine I et le domaine catalytique entier du ribozyme. Notamment, l'effet des différentes variations de séquence de la boucle de *SLV* sur l'activité du ribozyme en *trans*, représenté par des ratio de l'activité catalytique ($(k_{cat}/k_M)_{WT}/(k_{cat}/k_M)_{VAR}$, chapitre 2) est similaire à l'effet observé sur l'affinité de l'interaction *SLI/SLV* isolée, représenté par des ratio des constantes d'association ($(K_A)_{WT}/(K_A)_{VAR}$, chapitre 4). De plus, l'étude de l'interaction I/V effectuée par *ITC* appuie ce concept de modularité : la contribution énergétique de l'interaction I/V calculée en utilisant le domaine I en présence du domaine V seul ou du domaine catalytique entier est similaire, démontrant que l'interaction *kissing-loop* I/V

représente la contribution énergétique dominante pour la reconnaissance du substrat par le ribozyme VS. Donc, bien que les données actuelles démontrent que le substrat se positionne dans le domaine catalytique au sein d'une crevasse formée des domaines II et VI [92, 95, 96, 117, 147], les contacts possibles entre ceux-ci ne contribuent pas significativement à la liaison du substrat.

En plus de permettre la reconnaissance du substrat, l'interaction I/V est aussi impliquée dans son activation pour la catalyse par le ribozyme VS. En effet, la formation de cette interaction s'accompagne d'un réarrangement de la structure du substrat, le faisant passer d'une conformation dite *unshifted* et inactive à une conformation *shifted* et active. Nous avons présenté une preuve structurale par RMN du *helix shift* de la tige Ib dans des substrats *SLI shiftable* et l'absence de cet événement dans des substrats *SLI preshifted* ou double brin. De plus, nous avons évalué à 1.8-3.0 kcal/mol le coût énergétique d'un tel *shift*. En plus d'apporter une preuve du *helix shift* dans les substrats *shiftable*, les études par RMN ont permis de montrer que la formation de l'interaction I/V est associée à la déstabilisation de la dernière paire de bases avant la boucle dans ces substrats *shiftable*, mais aussi dans les substrats *preshifted* et double brin. Ceci a mené à la proposition d'un modèle d'activation du substrat *shiftable* (Figure 5.1) dans lequel les nucléotides G697-A698-C699, préorganisés dans le *U-turn* de la boucle V, se lient aux nucléotides C632-U631-G630 de la boucle I qui est désordonnée dans sa forme libre. Ce premier contact entre *SLI* et *SLV* induit la formation du *U-turn* dans la boucle I, ce qui a pour conséquence la déstabilisation de la paire de bases G633-C626 (Figure 5.1). Suite à ce contact initial, les triplets de bases se forment pour maximiser l'empilement de bases à l'interface (Figure 5.1). Ces changements structuraux, particulièrement la déstabilisation de la paire de bases G633-C626, facilitent certainement l'initiation du *helix shift* de la tige Ib, puis la réorganisation de la boucle interne contenant le site de clivage (Figure 5.1). La conception de ce modèle permet non seulement d'intégrer les nouvelles données obtenues à travers cette thèse à celles présentées dans la littérature, mais permet aussi d'avoir une vision mieux définie de la reconnaissance et de l'activation du substrat chez le ribozyme VS.

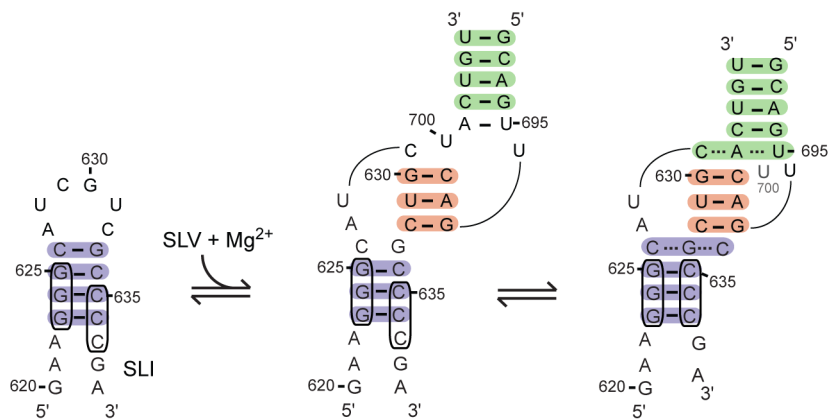


Figure 5.1 : Représentation schématique d'un modèle de la reconnaissance et de l'activation du substrat par le ribozyme VS de *Neurospora*.

L'intérêt des travaux présentés dans cette thèse, leur contribution ainsi que leur intégration scientifique seront présentés en trois aspects : Tout d'abord, au niveau de sa contribution à l'élaboration d'un modèle RMN tridimensionnel complet du ribozyme VS et de l'ingénierie de l'interaction, ensuite des bases de données et outils de prédiction, et finalement d'un point de vue plus technologique, au niveau de la nanotechnologie à base d'acides nucléiques.

5.1 Construction d'un modèle RMN complet du ribozyme VS de *Neurospora*

À ce jour, plusieurs modèles à basse résolution représentant la structure globale du ribozyme VS sont disponibles. De plus, la structure à haute résolution de certains domaines isolés est connue, c'est le cas pour les domaines hélicoïdaux I (en conformation inactive et active), V et la boucle interne A730 du domaine VI [103, 105, 111-114, 128]. Toute cette information structurale est considérable pour la compréhension du ribozyme VS au niveau de sa structure et de sa fonction. La structure de l'interaction I/V présentée dans cette thèse ne s'ajoutent pas simplement à celle des autres domaines isolés, mais représente un module clé qui définit la reconnaissance du substrat. Tel que mentionné dans l'introduction, l'adoption de la conformation active du ribozyme est guidée et facilitée par les jonctions II-III-VI et III-IV-V, mais aussi par la

formation de l'interaction I/V. Celle-ci constitue également une jonction importante du ribozyme, puisqu'elle guide le positionnement du substrat dans le domaine catalytique. Obtenir la structure d'un module contribuant à la structure active du ribozyme est un point de départ particulièrement intéressant pour entreprendre une reconstitution de la structure du ribozyme VS entier dans le cadre d'une approche de *divide and conquer*.

En biologie structurale, la stratégie *divide and conquer* consiste à diviser une molécule complexe en plusieurs sous-domaines plus simples, mais stables thermodynamiquement, qui seront étudiés indépendamment. L'information structurale obtenue à partir de ces sous-domaines est ensuite combinée pour être appliquée à la molécule entière originale. L'utilisation d'une telle stratégie pour une étude par spectroscopie RMN est particulièrement utile pour déterminer la structure à haute résolution d'ARN ayant une taille allant au-delà de 20 kDa (> 50 nt). En effet, l'utilisation d'ARN de plus petites tailles (< 50 nt) en RMN génère des données de meilleure qualité et facilite l'attribution des signaux, car les problèmes de superpositions de signaux et d'élargissement des raies spectrales augmentent avec la taille de la molécule. En plus d'avoir été utilisée pour déterminer des structures à haute résolution d'ARN de plus de 50 nt, la stratégie *divide and conquer* a aussi facilité l'attribution des signaux RMN et l'obtention de contraintes structurales [239-242]. Cette stratégie nécessite de la rigueur pour s'assurer que la structure des éléments isolés soit équivalente à celle adoptée dans le contexte du système intégral. Dans le cas de l'étude de la reconnaissance et de l'activation du substrat chez ribozyme VS, présentée dans cette thèse, j'ai utilisé un système composé des tige-boucles *SLI* et *SLV* isolées. La rigueur de l'approche *divide and conquer* pour ce système se traduit en plusieurs points. Il a été démontré, par des expériences biochimiques, que *SLV* seul est suffisant pour former l'interaction *kissing-loop* avec *SLI* et causer le même *helix shift* dans *SLI* que le domaine catalytique entier [109]. La structure RMN de l'interaction *SLI/SLV*, introduite au chapitre 3, présente plusieurs caractéristiques structurales qui correspondent au ribozyme entier mises en évidence par des études biochimiques: la dépendance de l'interaction I/V à la présence de cations divalents, la formation des trois paires de bases W-C à l'interface I/V et la formation d'un motif *U-turn* dans chaque boucle. De plus,

l'angle interhélicale (θ_{I-V}) et le déplacement axial (d_{I-V}) mesurés sur la structure RMN du complexe *SLI/SLV* isolé sont respectivement 127-175° et 10-24° et sont en accord avec ceux relevés sur deux modèles du ribozyme VS complet obtenus par *SAXS* ($\theta_{I-V}=154^\circ$ et $d_{I-V}=25^\circ$) et par modélisation contrainte par des données de cinétiques et de RMN ($\theta_{I-V}=165-171^\circ$ et $d_{I-V}=18-21^\circ$) [118, 147]. Les différents complexes formés entre *SLV* et les différents ARN *SLI* utilisés pour étudier le *helix shift* présentaient également ces mêmes signaux RMN. L'ensemble de ces caractéristiques indique que la structure de l'interaction *SLI/SLV* déterminée de manière isolée est très similaire à celle retrouvée dans le contexte du ribozyme entier.

Étant donné que l'étude du ribozyme VS entier par RMN occasionne des difficultés liées à sa taille (~150 nt), il devient intéressant d'utiliser l'approche du *divide and conquer* pour arriver à bâtir un modèle à haute résolution de ce ribozyme. D'ailleurs, un des projets du laboratoire porte sur l'obtention d'un modèle RMN à haute résolution du ribozyme VS. Ce modèle serait notamment le résultat d'un assemblage de ses modules les plus importants, c'est-à-dire l'interaction *kissing-loop* I/V présentée dans cette thèse, les boucles internes des domaines I et VI, ainsi que les jonctions II-III-VI et III-IV-V (Figure 5.2). À ce jour, plusieurs domaines du ribozyme ont été étudiés par RMN. Tout d'abord, trois structures du substrat *SLI* sont connues dont une du substrat entier en conformation inactive et une du site de clivage en conformation active contenant les sites de liaison au magnésium (Figure 5.2). Ensuite, pour ce qui est du domaine catalytique, la boucle terminale du domaine *SLV* et la boucle interne A730 du domaine *SLVI* ont été caractérisés et plusieurs structures sont disponibles, dont une du domaine *SLV* et une du domaine *SLVI* contenant chacune les sites de liaisons au magnésium (Figure 5.2). De plus, la structure RMN de l'interaction *SLI/SLV*, présentée dans cette thèse, constitue un module déterminant pour le repliement du ribozyme (Figure 5.2). Cette structure sera particulièrement utile pour permettre de contraindre le modèle et définir l'interface entre le substrat et le domaine catalytique.

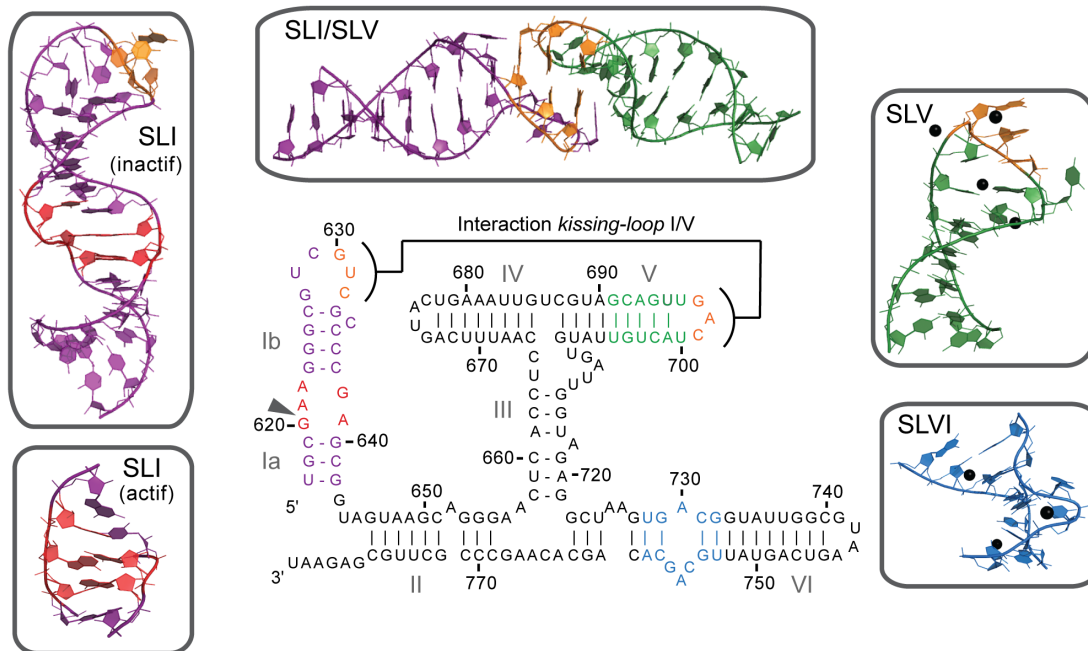


Figure 5.2 : Représentation de la structure secondaire du ribozyme VS entier avec plusieurs structures tridimensionnelles des domaines isolés connus. *SLI* inactif (PDB : 1HWQ [111]), *SLI* actif (PDB : 1OW9 [112]), *SLI/SLV* (PDB : 2MI0 [243]), *SLV^{Mg}* (PDB : 1YN2), *SLV^{Mg}* (PDB : 2MIS [114]).

Dans le but de faire un assemblage réaliste et rationnel, une approche incluant de la dynamique moléculaire s'avère intéressante. En effet, la structure RMN de la jonction I/V, présentée ici, permet de connaître l'organisation précise des nucléotides à l'interface et d'identifier les éléments structuraux rigides et ceux qui sont plutôt dynamiques. Par exemple, selon les données RMN du complexe *SLI/SLV*, les nucléotides C7^I, A8^I, U9^I, C13^I, G14^I, G15^I, U8^V et U9^V, localisés à la base des boucles terminales de *SLI* et *SLV*, présentent de faibles intensités de signaux (chapitre 3). Ceci suggère de la dynamique, créant ainsi des régions charnières dans *SLI* et *SLV*. Ce type d'information constitue une base essentielle pour l'élaboration d'un modèle dynamique. De plus, les contraintes provenant de la RMN sont très utiles pour créer un tel modèle puisqu'elles peuvent être utilisées pour restreindre l'espace conformationnel à explorer. De cette façon, la structure RMN de la jonction I/V, en combinaison avec les structures RMN des autres modules (les boucles internes I et VI et les jonctions II-III-VI et III-IV-V), pourrait être utilisée

dans la construction d'un modèle du ribozyme VS à haute résolution. De plus, ces structures RMN pourraient guider des études de dynamique moléculaire, permettant de définir les mouvements principaux impliqués dans la formation du site actif du ribozyme, c'est-à-dire l'association des boucles internes des domaines I et VI. Ainsi, ce modèle servirait à mieux comprendre les déterminants structuraux impliqués dans la reconnaissance du substrat et l'activité du ribozyme VS.

5.2 Ingénierie rationnelle de l'interaction *kissing-loop* I/V

Bien comprendre le fonctionnement d'un tel ribozyme est particulièrement intéressant au niveau de l'ingénierie de l'ARN. En effet, connaître les relations structure-fonction d'un système permet de pouvoir en faire l'optimisation en visant les éléments clés, comme la reconnaissance d'un substrat ou encore les éléments qui contribuent ou limitent son fonctionnement. Tout cela en vue de pouvoir utiliser le système en tant qu'outil biochimique ou encore comme agent ayant une cible d'intérêt thérapeutique. Dans le cas du ribozyme VS, son domaine catalytique entier présente un intérêt, mais aussi le module I/V en lui-même. L'interaction I/V du ribozyme VS permet la reconnaissance d'un substrat replié en tige-boucle, ce qui est particulièrement important au niveau du ciblage d'ARN *in vivo*, puisque ceux-ci se retrouvent principalement sous forme repliée.

Maintenant que les éléments déterminant la spécificité de l'interaction sont connus, il devient envisageable, par exemple, de faire des modifications rationnelles pour adapter l'interaction à reconnaître une cible différente. D'ailleurs, le module *SLI/SLV* présente de la flexibilité de séquence au niveau des tiges, mais aussi au niveau des boucles, ce qui permet un éventail de possibilité pour son ingénierie. En effet, des études biochimiques ont montré que l'identité des nucléotides dans la tige V pouvait être modifiée sans diminution significative de l'activité du ribozyme [89, 90, 117]. Pour la tige Ib d'un substrat *shiftable*, seulement certaines positions peuvent être modifiées en raison du *helix shift* et la conformation active du site de clivage, alors que pour un substrat *preshifted*, la tige Ib peut accommoder plusieurs modifications de séquence [82,

89, 90, 107, 115]. D'ailleurs, un exemple intéressant de l'exploitation de la flexibilité des tiges Ib et V est une étude dans laquelle des variations compensatoires de longueur d'hélices ont été effectuées entre les tiges Ib et V avec un ribozyme en *trans* [118]. Cette étude met notamment en évidence le niveau d'adaptabilité du ribozyme, pouvant cliver efficacement des substrats ayant une tige Ib variant de 3 à 6 paires de bases [118]. En ce qui concerne l'affinité uniquement de l'interaction *kissing-loop* I/V, nous avons montré par RMN et *ITC* que la structure et l'identité des nucléotides de la tige d'un substrat *preshifted* n'avait pas d'importance (chapitre 3 et 4). Les séquences des boucles I et V offrent également une certaine flexibilité qui est compatible avec l'activité du ribozyme, quoique modeste en raison de la nécessité de respecter la séquence consensus UNR du motif *U-turn* pour les nucléotides 628-629-630 dans *SLI* et 696-697-698 dans *SLV* [82]. En effet, les paires de bases W-C entre les nucléotides 630-631-632 de *SLI* et 699-698-697 de *SLV* ne peuvent être substituées que par certaines combinaisons pour notamment respecter la présence d'une purine (R) aux positions 630 et 698 [82]. Nous avons également montré que l'activité du ribozyme était compatible avec l'insertion d'un nucléotide dans la boucle V, la modification de la paire de bases fermant la boucle, la mutation du nucléotide U700, mais pas avec la délétion de ce dernier (chapitre 2). Du côté de *SLI*, il a été démontré qu'une plus grande boucle terminale (4-10 nt) était compatible avec l'activité du ribozyme [107, 109, 115](chapitre 4).

De plus, tel que présenté dans l'introduction, la formation de l'interaction I/V est associée à un changement de conformation de la tige-boucle I qui a été caractérisé plus en profondeur dans les chapitres 3 et 4. Ce réarrangement, qui permet à la tige-boucle I d'adopter une conformation favorable à la coupure par le ribozyme, est une propriété qui pourrait également être exploitée pour l'ingénierie. Puisque la formation et l'affinité du module *SLI/SLV* sont dépendantes des cations divalents, il est possible d'utiliser cette dépendance pour ajuster et contrôler la formation de l'interaction I/V. Il a aussi été démontré que le passage de la conformation *unshifted* à *shifted* dépendait aussi du pH [111]. En effet, une augmentation du pH mène à la déprotonation du A622 de la boucle interne de *SLI*, provoquant le bris de la paire de bases A+622-C637. Il a donc été proposé que la déstabilisation de cette paire de bases faciliterait l'adoption de la

conformation *shifted* [111]. L'avantage de connaître l'aspect structural et fonctionnel de ces caractéristiques est la possibilité de les exploiter dans un système différent. Par exemple, la conception d'un outil biochimique basé sur le ribozyme VS ou sur l'interaction *SLI/SLV* dont la fonction serait activable par une modification de pH de la solution ou sa concentration en cations divalents. En résumé, le module *SLI/SLV* du ribozyme VS présente plusieurs propriétés attrayantes pour en faire de l'ingénierie: flexibilité de séquence, petite taille, bonne affinité, activable et dépendance au pH et aux cations divalents.

5.3 Prédiction de structures tridimensionnelles et bases de données d'ARN

Étant donnée le nombre grandissant d'ARN fonctionnel, ainsi que de leur taille, il devient difficile d'obtenir la structure à haute résolution de ces ARN en utilisant uniquement la RMN ou la cristallographie. En effet, ces méthodes peuvent être coûteuses en temps et en échantillons. De plus, la taille de plusieurs ARN (par exemple les ARN^{Inc}) dépasse souvent largement ce qui est compatible avec la RMN, et au niveau cristallisation, les défis ne sont que grandissant pour obtenir des cristaux et/ou résoudre les problèmes de phases. Pour ces différentes raisons, la prédiction de structure tridimensionnelle devient une option intéressante pour faciliter l'obtention de structures d'ARN. La prédiction de structures tridimensionnelles d'ARN consiste à définir la structure native de l'ARN à partir de sa séquence et/ou de sa structure secondaire. Plusieurs outils ont été développés à cette fin et, de façon générale, sont basés sur deux approches principales : *ab initio* ou sans gabarit (*template free*) et *knowledge based* ou avec gabarit (*template based*) (Figure 5.3).

L'approche, avec gabarit (*template based*) part du principe que pour des ARN qui sont reliés évolutivement, les structures secondaires et tertiaires sont plus conservées que leurs séquences [244]. Les régions importantes pour la stabilité et la fonction de l'ARN sont aussi très conservées. Par conséquent, il est généralement possible d'utiliser

les alignements de séquence et les patrons de structures secondaires pour reconnaître cette conservation de structure tridimensionnelle [245, 246]. Cette approche est en fait une prédiction par homologie ou encore par analyse comparative, donc il est nécessaire d’avoir la structure d’un ARN similaire comme gabarit (Figure 5.3). De plus, puisque cette approche dépend des structures qui composent les bases de données, plus celles-ci sont riches et diversifiées, plus grande est la possibilité d’avoir une structure homologue comme modèle de comparaison. Le succès de cette méthode repose donc sur l’alignement de séquence et l’existence d’une structure gabarit [247].

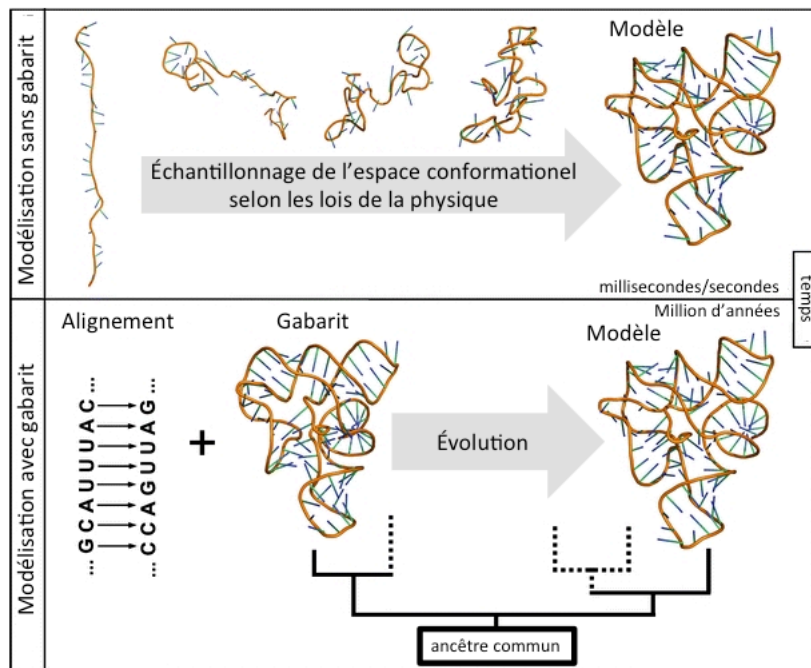


Figure 5.3 : Représentation des approches de modélisation sans (panneau du haut) et avec gabarit (panneau du bas). Adapté de [246].

Pour ce qui est de la deuxième approche, *ab initio* ou sans gabarit (*template free*), elle est basée sur une hypothèse thermodynamique selon laquelle l’état natif d’une macromolécule correspond à sa structure de plus basse énergie [248, 249]. La méthode *ab initio* utilise les lois de la physique et de la thermodynamique pour modéliser le repliement de l’ARN. En effet, les différents changements de conformation de l’ARN sont simulés pour arriver à déterminer sa structure native, étant définie comme la

structure de plus basse énergie (Figure 5.3). Contrairement à la méthode par homologie, l'approche *ab initio* s'effectue en absence de données initiales (*template free*).

Ces deux approches, avec ou sans gabarit, comportent des limites importantes, par exemple la dépendance d'un alignement ou d'une structure gabarit pour la prédiction par homologie et le vaste espace conformationnel à explorer ainsi que la complexité des fonctions d'énergie utilisées pour l'approche *ab initio* [246]. Conséquemment, la majorité des outils de prédiction utilisés sont dits hybrides et font appel à une combinaison des deux approches [246, 250-252]. C'est le cas notamment des outils de prédiction par assemblage de fragments qui assemblent la structure d'un ARN à partir d'une librairie de petits blocs d'ARN (FARNA [253, 254] et MCSYM [255]). Il y a également d'autres outils qui eux permettent l'incorporation de données expérimentales telles que le *SAXS* ou encore l'accessibilité au solvant (NAST, [256]).

Les travaux effectués dans cette thèse présentent un intérêt au niveau des bases de données de structure et aussi pour les outils de prédiction de structure tridimensionnelle d'ARN (avec ou sans gabarit et hybrides). En effet, la structure à haute résolution de l'interaction I/V représente un ajout modeste, mais important pour les bases de données, puisqu'à ce jour, celles-ci ne comprennent que peu d'éléments de structure tertiaire de type *kissing-loop*. Du côté des outils de prédiction, la structure d'une interaction *kissing-loop* est particulièrement intéressante, car le repliement de l'ARN est dicté par des éléments de structure tertiaires, tels les interactions *kissing-loop* et les jonctions multivoies. La compréhension de ce type de motif est donc nécessaire pour que les outils de prédiction puissent les incorporer dans l'assemblage de la structure tridimensionnelle de l'ARN. De plus, les études de structure et de thermodynamique présentées dans cette thèse ont montré que l'interaction I/V est un élément de structure tertiaire complexe avec une stabilité globale plus grande que la somme des éléments individuels qui la composent. Ceci est particulièrement important au niveau de l'évaluation de l'énergie des structures générées tout au long d'un calcul de prédiction. Parallèlement, bien que les outils de prédiction de structure tridimensionnelle d'ARN s'améliorent constamment, les études structurales et thermodynamiques de l'ARN

demeurent nécessaires, notamment pour améliorer les champs de forces, trouver de nouveaux assemblages de motifs tertiaires et mieux comprendre le rôle crucial des cations divalents dans le repliement et la fonction de l'ARN.

5.4 Nanotechnologie à base d'ARN

La nanotechnologie a comme objectif la manipulation et la conception de composantes, d'engins ou encore de systèmes de l'ordre de grandeur du nanomètre pour développer de nouveaux matériaux avec des propriétés particulières et, de façon plus globale, pour mieux comprendre la nature qui nous entoure [257]. Dans certains domaines de la nanotechnologie plus moderne, les acides nucléiques (ADN et ARN) sont utilisés comme matériel de base. En effet, les acides nucléiques forment des modules ou des blocs de constructions qui s'organisent notamment grâce à la formation de paires de bases W-C et d'interactions d'empilement de bases. L'ADN est utilisé depuis déjà une trentaine d'année [258], pour sa spécificité à former les appariements W-C A-T et G-C, pour adopter des doubles hélices robustes et pour sa stabilité physicochimique. De son côté, l'ARN est utilisé depuis une dizaine d'années pour ses propriétés qui le distinguent de l'ADN, par exemple pour sa capacité à former des hélices de type A qui sont plus compactes [259, 260] et plus stables thermodynamiquement que les hélices de type B formées par l'ADN [261-263] et pour sa capacité à adopter des motifs tertiaires stables facilitant son repliement, son assemblage et permettant une grande diversité de conformations [163]. Une approche appelée ARN tectonique (*tecto-RNA*), développée au début des années 2000 [264], consiste à accumuler des modules d'ARN, à les assembler pour construire des structures de forme désirée, à la manière d'un jeu de LEGO (Figure 5.4). De plus, puisque le repliement de l'ARN est souvent dépendant des cations, il est possible d'utiliser cette particularité pour manipuler la dynamique du repliement [149, 265, 266]. Ainsi, bien que l'ADN soit la molécule pionnière dans la conception de modules de construction pour la nanotechnologie à base d'acide nucléiques, l'ARN représente un matériel alternatif qui continue son émergence. La diversité de structure de l'ARN a permis de fournir un large éventail de modules différents avec des propriétés uniques à partir de

plusieurs motifs secondaires et tertiaires, tels que les jonctions multivoies, les boucles internes et les interactions *kissing-loop* (Figure 5.4).

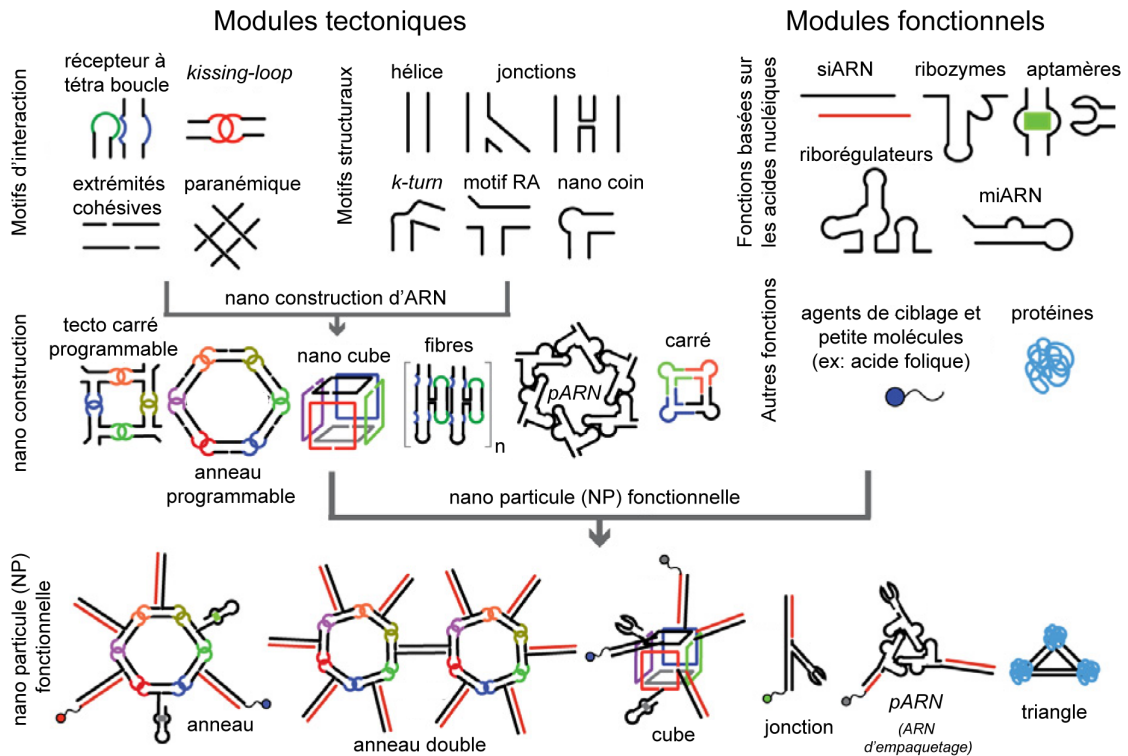


Figure 5.4 : Représentation schématique des différents modules utilisés dans la nanotechnologie à base d'ARN. Tiré de [267]

D'ailleurs les interactions *kissing-loop* sont particulièrement intéressantes pour faire l'assemblage de plusieurs modules. Par exemple, l'interaction *kissing-loop* de *DIS* a été utilisé pour obtenir des multimères qui ont permis de construire des structures en forme d'anneau [268, 269]. Puisque cette interaction *kissing-loop* est dépendante de la présence de cations magnésium, il a été démontré qu'il était possible de contrôler l'assemblage de l'interaction en modifiant les concentrations de magnésium utilisées [268]. Les interactions *kissing-loop* ont aussi été exploitées pour obtenir des modules comprenant des angles particuliers, comme par exemple celle du *DIS* permet d'introduire un angle de près de 180° [215], alors que l'interaction *ColE1* permet un angle de 120° [213].

Les défis actuels de la nanotechnologie basée sur les acides nucléiques sont, bien sûr, la stabilisation des molécules d'ARN pour élargir le potentiel des nanostructures d'ARN pour des applications telles que la conception de véhicules pour des agents thérapeutiques *in vivo* [65, 265, 267, 270, 271], mais aussi l'identification et la caractérisation de motifs structuraux plus dynamiques. De plus, il demeure préférable de travailler avec des modules de base de plus petite taille pour éviter les contraintes d'équilibre entre la cinétique et la thermodynamique de repliement, ce qui parfois mène à des conformations alternatives dans lesquelles l'ARN est piégé [265, 272, 273].

Ainsi, la structure de l'interaction *kissing-loop* I/V présentée dans cette thèse s'agence bien avec les besoins de la nanotechnologie basée sur l'ARN. En effet, l'interaction I/V est un élément modulaire extrait du ribozyme VS qui a été bien caractérisé en terme de structure et de thermodynamique, c'est une interaction *kissing-loop* de petite taille, comparée aux interactions *kissing-loop* du *DIS* et *ColE1*, mais qui est tout aussi stable. Les trois paires de bases centrales ainsi que la séquence des deux tiges-boucles (*SLI* et *SLV*) présentent une certaine flexibilité (présentée dans la section 5.2) et peuvent donc être facilement modifiées. L'interaction I/V offre également une certaine variabilité d'angle interhélice allant de 127° à 175°. De plus, étant donnée que l'interaction *kissing-loop* I/V est dépendante du magnésium, sa formation pourrait être contrôlée en modifiant les concentrations de cations. Donc les travaux présentés ici contribuent à un enrichissement du nombre de modules de construction disponibles pouvant, par exemple, mener à la construction de nanostructures à intérêt thérapeutique ou encore pour créer des outils biochimiques, ou même dans le développement de nouvelles composantes pour des domaines comme la nanomédecine.

5.5 Perspectives

Certains travaux complémentaires futurs seraient intéressants, tout d'abord pour pousser plus loin la compréhension des éléments responsables de la stabilité des motifs tertiaires, mais aussi pour faire l'ingénierie de l'interaction *kissing-loop* I/V.

Les travaux présentés ici montrent une interaction *kissing-loop* d'une grande stabilité et décrivent plusieurs éléments responsables de cette grande stabilité, par exemple l'empilement de bases de part et d'autre des trois paires de bases W-C centrales. À travers les études structurales et thermodynamiques de l'interaction I/V présentées dans cette thèse, il a notamment été montré que les cations magnésium ont un rôle important dans la stabilisation de cette interaction. Toutefois, la localisation de ces cations n'a pas été étudiée. Plusieurs facteurs indiquent que la liaison de cations magnésium contribue significativement à la grande stabilité de l'interaction : (1) ils facilitent le repliement des ARN en général [148-150], (2) ils sont essentiels et/ou stabilisent la formation de plusieurs éléments de structure tertiaire, par exemple le motif de type *U-turn* [114, 183], (3) ils stabilisent plusieurs interactions de type *kissing-loop* [200, 201, 203-208] et (4) il a été démontré qu'ils sont essentiels à la formation de l'interaction *kissing-loop* I/V du ribozyme VS [82, 90, 106, 107].

La localisation des sites de liaison des cations magnésium pourrait être étudiée par RMN en utilisant plusieurs méthodes complémentaires telles que la relaxation paramagnétique induite par le manganèse (*Mn²⁺-induced paramagnetic relaxation enhancement (PRE)*) [274] et la perturbation de déplacement chimique induite par le cadmium (*Cd²⁺-induced chemical-shift perturbation (CSP)*) [275, 276] sur des ARN contenant des modifications phosphorothioates dans les boucles terminales I et V. L'information extraite de ces expériences permettrait de localiser les cations magnésium dans l'interaction I/V en identifiant les ligands potentiels de ces cations, ainsi que le type de liaison (directe ou indirecte). L'utilisation de l'hexamine de cobalt ($\text{Co}(\text{NH}_3)_6^{3+}$) serait aussi intéressante, puisque ce cation interagit avec l'ARN en imitant la conformation du magnésium hexahydraté, ce qui permet d'étudier les interactions de type indirecte [277-279]. Une particularité intéressante de l'utilisation du $\text{Co}(\text{NH}_3)_6^{3+}$ est la possibilité de mesurer des signaux RMN de type *NOE* entre les protons du $\text{Co}(\text{NH}_3)_6^{3+}$ et l'ARN, ce qui permet d'extraire des distances précises entre eux [233]. La RMN du fluore 19 (^{19}F) pourrait également être utilisée pour l'étude de la liaison des cations à l'ARN, car le déplacement chimique du ^{19}F est très sensible à l'environnement électronique [280-283].

Jusqu'à maintenant, il n'y a que peu d'information disponible concernant la localisation des cations divalents à l'interaction I/V. En effet, il a été proposé, suite à des expériences avec l'agent alkylant *ENU*, que des cations magnésium étaient associés au 5'-phosphate des nucléotides U631 (*SLI*) et C699 (*SLV*) [104, 157]. De plus, une étude par RMN du domaine *SLV* seul a permis de définir quatre sites de liaison au magnésium, dont trois sont communs à plusieurs motifs de type *U-turn* : un s'associant aux phosphates dans le renversement du squelette ribose-phosphate (1), un dans le sillon mineur impliquant les N7 des positions N et R (2) et le troisième dans le sillon majeur impliquant notamment le O4' du U et le 3'-phosphate du R (3). En plus, un quatrième site implique le 5'-phosphate du U700 (4) [114]. Puisque les structures des boucles de *SLV* seul et en complexe avec *SLI* sont similaires (*rmsd* de 1.75 Å, voir chapitre 3), il est possible que certains sites retrouvés sur la forme libre soient conservés sur la forme en complexe. Ainsi, en analysant la structure de l'interaction I/V, il semble probable que les sites 1, 3 et 4 soient conservés dans *SLV*, contrairement au site 2 qui se retrouve dans le sillon mineur, là où l'interaction *kissing-loop* prend place. Étant donné que la boucle de *SLI* adopte aussi un motif de type *U-turn*, il est possible que les sites 1 et 3, retrouvés dans le *U-turn* de *SLV*, soient également compatibles avec le *U-turn* de *SLI*. En dernier lieu, il est probable qu'un site de liaison de magnésium se situe à l'endroit où les squelettes ribose-phosphate de *SLI* et *SLV* sont le plus près l'un de l'autre, c'est-à-dire entre les nucléotides U696 (*SLV*) et U628 (*SLI*), ce qui réduirait la densité de charges électronégatives à cet endroit.

En résumé, il est possible de spéculer sur la localisation de plusieurs sites de liaison de magnésium, dont deux dans chaque *U-turn* impliquant notamment les 5'-phosphates de C699 et U631 identifiés par *ENU*, un à l'interface où les squelettes ribose-phosphate de *SLI* et *SLV* se retrouvent à proximité et un impliquant le 5'-phosphate du U700.

Au niveau de l'ingénierie, si on avait accès à un modèle à haute résolution du VS dont on comprend la dynamique et la flexibilité, il serait possible de simuler des

changements optimaux au niveau de l'interaction pour l'optimiser ou encore la modifier pour cibler une séquence différente. Pour arriver à un tel modèle avec les données de RMN actuelles, il est nécessaire d'obtenir la structure à haute résolution des autres modules du VS, tels que les jonctions, puis de les assembler en un modèle complet. Il faut également valider le modèle en vérifiant que le repliement des modules individuels correspond à celui retrouvé dans contexte du ribozyme entier. Pour ce faire, il est nécessaire d'arriver à produire un ribozyme entier dont la conformation est stable et homogène pour que celui-ci soit soumis à des expériences de RMN, de SAXS et de cristallographie. L'information structurale résultante à plus ou moins haute résolution permettra de vérifier le repliement du ribozyme pour évaluer la validité du modèle.

Dans le but de mieux comprendre les relations structure-fonction de l'ARN et de compléter la caractérisation thermodynamique de l'interaction I/V, il serait particulièrement intéressant de caractériser la cinétique de sa formation en déterminant les constantes d'association (k_{on}) et de dissociation (k_{off}). Celles-ci représentent la vitesse à laquelle le complexe se forme (k_{on}) et se dissocie (k_{off}), alors que la constante de dissociation déterminée par le paramètre K_D au chapitre 4 représente la stabilité du complexe formé à l'équilibre $K_D = k_{\text{off}} / k_{\text{on}}$. Ces constantes de vitesse permettraient notamment de vérifier et quantifier la rapidité de l'association entre les deux molécules par rapport à la vitesse de diffusion ou encore le ralentissement causé par des changements de conformationnel. En terme de conception d'outil biochimique, connaître les constantes de vitesse et de dissociation à l'équilibre d'une interaction est intéressant notamment pour déterminer son potentiel à lier spécifiquement un ligand, mais aussi à le relarguer pour pouvoir se recycler et générer plusieurs événements de liaison. Ces constantes de cinétique pourraient, par exemple, être mesurées par résonance plasmonique de surface (*surface plasmon resonance, SPR*) [214, 284-286].

En résumé, avoir accès à une interaction *kissing-loop* dont la structure, ainsi que les paramètres thermodynamiques et cinétiques sont connus faciliterait grandement son utilisation. Cette interaction serait particulièrement utile dans des domaines comme l'ingénierie et la nanotechnologie d'ARN. En effet, avec toutes les caractéristiques

connues de l'interaction I/V, celle-ci pourrait être utilisée comme module de construction dans la conception d'un véhicule de livraison d'agents thérapeutiques dans une approche de nanomédecine.

Bibliographie

1. 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.
2. 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.
3. Guerrier-Takada, C. and S. Altman, *Catalytic activity of an RNA molecule prepared by transcription in vitro*. Science, 1984. **223**(4633): p. 285-6.
4. Mizuno, T., M.Y. Chou, and M. Inouye, *A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA)*. Proc Natl Acad Sci USA, 1984. **81**(7): p. 1966-70.
5. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans*. Nature, 1998. **391**(6669): p. 806-11.
6. Montgomery, M.K., S. Xu, and A. Fire, *RNA as a target of double-stranded RNA-mediated genetic interference in Caenorhabditis elegans*. Proc Natl Acad Sci USA, 1998. **95**(26): p. 15502-7.
7. Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14*. Cell, 1993. **75**(5): p. 843-54.
8. Winkler, W.C., S. Cohen-Chalamish, and R.R. Breaker, *An mRNA structure that controls gene expression by binding FMN*. Proc Natl Acad Sci USA, 2002. **99**(25): p. 15908-13.
9. Mironov, A.S., et al., *Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria*. Cell, 2002. **111**(5): p. 747-56.
10. Winkler, W., A. Nahvi, and R.R. Breaker, *Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression*. Nature, 2002. **419**(6910): p. 952-6.

11. Okazaki, Y., et al., *Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs*. Nature, 2002. **420**(6915): p. 563-73.
12. Kapranov, P., et al., *RNA maps reveal new RNA classes and a possible function for pervasive transcription*. Science, 2007. **316**(5830): p. 1484-8.
13. Mercer, T.R., M.E. Dinger, and J.S. Mattick, *Long non-coding RNAs: insights into functions*. Nat Rev Genet, 2009. **10**(3): p. 155-9.
14. Rinn, J.L., et al., *Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs*. Cell, 2007. **129**(7): p. 1311-23.
15. Mercer, T.R., et al., *Specific expression of long noncoding RNAs in the mouse brain*. Proc Natl Acad Sci USA, 2008. **105**(2): p. 716-21.
16. Amaral, P.P. and J.S. Mattick, *Noncoding RNA in development*. Mamm Genome, 2008. **19**(7-8): p. 454-92.
17. Crick, F.H., *The origin of the genetic code*. J Mol Biol, 1968. **38**(3): p. 367-79.
18. Orgel, L.E., *Evolution of the genetic apparatus*. J Mol Biol, 1968. **38**(3): p. 381-93.
19. Woese, C.R., *The genetic code; the molecular basis for genetic expression*. Modern perspectives in biology. 1967, New York,: Harper & Row. viii, 200 p.
20. Gilbert, W., *Origin of life: the RNA world*. Nature, 1986. **318**.
21. Talini, G., S. Branciamore, and E. Gallori, *Ribozymes: Flexible molecular devices at work*. Biochimie, 2011. **93**(11): p. 1998-2005.
22. Serganov, A. and D.J. Patel, *Ribozymes, riboswitches and beyond: regulation of gene expression without proteins*. Nat Rev Genet, 2007. **8**(10): p. 776-90.
23. Zhang, J., M.W. Lau, and A.R. Ferre-D'Amare, *Ribozymes and riboswitches: modulation of RNA function by small molecules*. Biochemistry, 2010. **49**(43): p. 9123-31.
24. Forster, A.C. and R.H. Symons, *Self-cleavage of plus and minus RNAs of a virusoid and a structural model for the active sites*. Cell, 1987. **49**(2): p. 211-20.
25. Buzayan, J.M., W.L. Gerlach, and G. Bruening, *Nonenzymatic cleavage and ligation of rnas complementary to a plant-virus satellite rna*. Nature, 1986. **323**(6086): p. 349-353.

26. Sharmeen, L., et al., *Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage*. J Virol, 1988. **62**(8): p. 2674-9.
27. Wu, H.N., et al., *Human hepatitis delta virus RNA subfragments contain an autocleavage activity*. Proc Natl Acad Sci USA, 1989. **86**(6): p. 1831-5.
28. 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.
29. Winkler, W.C., et al., *Control of gene expression by a natural metabolite-responsive ribozyme*. Nature, 2004. **428**(6980): p. 281-6.
30. Roth, A., et al., *A widespread self-cleaving ribozyme class is revealed by bioinformatics*. Nat Chem Biol, 2014. **10**(1): p. 56-60.
31. Liu, Y., et al., *Crystal structure and mechanistic investigation of the twister ribozyme*. Nat Chem Biol, 2014. **10**(9): p. 739-44.
32. 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.
33. Martick, M., et al., *A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA*. Nature, 2008. **454**(7206): p. 899-902.
34. Michel, F. and B. Dujon, *Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast- and nuclear-encoded members*. EMBO J, 1983. **2**(1): p. 33-8.
35. Michel, F., K. Umesono, and H. Ozeki, *Comparative and functional anatomy of group II catalytic introns--a review*. Gene, 1989. **82**(1): p. 5-30.
36. Teixeira, A., et al., *Autocatalytic RNA cleavage in the human beta-globin pre-mRNA promotes transcription termination*. Nature, 2004. **432**(7016): p. 526-30.
37. Nielsen, H., E. Westhof, and S. Johansen, *An mRNA is capped by a 2', 5' lariat catalyzed by a group I-like ribozyme*. Science, 2005. **309**(5740): p. 1584-7.
38. Pyle, A.M. and A.M. Lambowitz, *17 Group II Introns: Ribozymes That Splice RNA and Invade DNA*, in *The RNA world*. 2006, Cold Spring Harbor Monograph Archive.
39. Cech, T.R., A.J. Zaug, and P.J. Grabowski, *In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence*. Cell, 1981. **27**(3 Pt 2): p. 487-96.

40. Vogel, J. and T. Borner, *Lariat formation and a hydrolytic pathway in plant chloroplast group II intron splicing*. EMBO J, 2002. **21**(14): p. 3794-803.
41. Fedor, M.J. and J.R. Williamson, *The catalytic diversity of RNAs*. Nat Rev Mol Cell Biol, 2005. **6**(5): p. 399-412.
42. Nissen, P., et al., *The structural basis of ribosome activity in peptide bond synthesis*. Science, 2000. **289**(5481): p. 920-30.
43. Silverman, S.K., *Catalytic DNA (deoxyribozymes) for synthetic applications-current abilities and future prospects*. Chem Commun (Camb), 2008(30): p. 3467-85.
44. Seelig, B. and A. Jaschke, *A small catalytic RNA motif with Diels-Alderase activity*. Chem Biol, 1999. **6**(3): p. 167-76.
45. Tarasow, T.M., S.L. Tarasow, and B.E. Eaton, *RNA-catalysed carbon-carbon bond formation*. Nature, 1997. **389**(6646): p. 54-7.
46. Unrau, P.J. and D.P. Bartel, *RNA-catalysed nucleotide synthesis*. Nature, 1998. **395**(6699): p. 260-3.
47. Sugimoto, A., *High-throughput RNAi in Caenorhabditis elegans: genome-wide screens and functional genomics*. Differentiation, 2004. **72**(2-3): p. 81-91.
48. Dykxhoorn, D.M., C.D. Novina, and P.A. Sharp, *Killing the messenger: short RNAs that silence gene expression*. Nat Rev Mol Cell Biol, 2003. **4**(6): p. 457-67.
49. Walker, S.C., J.M. Avis, and G.L. Conn, *General plasmids for producing RNA in vitro transcripts with homogeneous ends*. Nucleic Acids Res, 2003. **31**(15): p. e82.
50. Price, S.R., et al., *Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies*. J Mol Biol, 1995. **249**(2): p. 398-408.
51. Dzianott, A.M. and J.J. Bujarski, *An in vitro transcription vector which generates nearly correctly ended RNAs by self-cleavage of longer transcripts*. Nucleic Acids Res, 1988. **16**(22): p. 10940.

52. 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.
53. Kieft, J.S. and R.T. Batey, *A general method for rapid and nondenaturing purification of RNAs*. RNA, 2004. **10**(6): p. 988-95.
54. Duss, O., et al., *A fast, efficient and sequence-independent method for flexible multiple segmental isotope labeling of RNA using ribozyme and RNase H cleavage*. Nucleic Acids Res, 2010. **38**(20): p. e188.
55. 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.
56. Batey, R.T. and J.S. Kieft, *Improved native affinity purification of RNA*. RNA, 2007. **13**(8): p. 1384-9.
57. Seetharaman, S., et al., *Immobilized RNA switches for the analysis of complex chemical and biological mixtures*. Nat Biotechnol, 2001. **19**(4): p. 336-41.
58. Robertson, M.P. and A.D. Ellington, *In vitro selection of nucleoprotein enzymes*. Nat Biotechnol, 2001. **19**(7): p. 650-5.
59. Breaker, R.R., *Engineered allosteric ribozymes as biosensor components*. Curr Opin Biotechnol, 2002. **13**(1): p. 31-9.
60. Germer, K., M. Leonard, and X. Zhang, *RNA aptamers and their therapeutic and diagnostic applications*. Int J Biochem Mol Biol, 2013. **4**(1): p. 27-40.
61. Robertson, M.P. and A.D. Ellington, *In vitro selection of an allosteric ribozyme that transduces analytes to amplicons*. Nat Biotechnol, 1999. **17**(1): p. 62-6.
62. Koizumi, M., et al., *Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP*. Nat Struct Biol, 1999. **6**(11): p. 1062-71.
63. Bora, U., A. Sett, and D. Singh, *Nucleic Acid Based Biosensors for Clinical Applications*. J Biosens, 2013. **1**(1).
64. Liu, J., Z. Cao, and Y. Lu, *Functional nucleic acid sensors*. Chem Rev, 2009. **109**(5): p. 1948-98.
65. Liu, J., et al., *Fabrication of stable and RNase-resistant RNA nanoparticles active in gearing the nanomotors for viral DNA packaging*. ACS Nano, 2011. **5**(1): p. 237-46.

66. Singh, A., S. Poshtiban, and S. Evoy, *Recent advances in bacteriophage based biosensors for food-borne pathogen detection*. *Sensors (Basel)*, 2013. **13**(2): p. 1763-86.
67. Weng, D.E. and N. Usman, *Angiozyme: a novel angiogenesis inhibitor*. *Curr Oncol Rep*, 2001. **3**(2): p. 141-6.
68. Morrow, P.K., et al., *An open-label, phase 2 trial of RPI.4610 (Angiozyme) in the treatment of metastatic breast cancer*. *Cancer*, 2012. **118**(17): p. 4098-104.
69. 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.
70. Nazari, R. and S. Joshi, *Exploring the potential of group II introns to inactivate human immunodeficiency virus type 1*. *J Gen Virol*, 2008. **89**(Pt 10): p. 2605-10.
71. Mulhbacher, J., P. St-Pierre, and D.A. Lafontaine, *Therapeutic applications of ribozymes and riboswitches*. *Curr Opin Pharmacol*, 2010. **10**(5): p. 551-6.
72. Izquierdo, M., *Short interfering RNAs as a tool for cancer gene therapy*. *Cancer Gene Ther*, 2005. **12**(3): p. 217-27.
73. Putral, L.N., W. Gu, and N.A. McMillan, *RNA interference for the treatment of cancer*. *Drug News Perspect*, 2006. **19**(6): p. 317-24.
74. Zhang, L., et al., *Potential use of nucleic acid-based agents in the sensitization of nasopharyngeal carcinoma to radiotherapy*. *Cancer Lett*, 2012. **323**(1): p. 1-10.
75. Senzer, N., et al., *Phase I trial of "bi-shRNAi(furin)/GM-CSF DNA/autologous tumor cell" vaccine (FANG) in advanced cancer*. *Mol Ther*, 2012. **20**(3): p. 679-86.
76. Li, M.J., et al., *Long-term inhibition of HIV-1 infection in primary hematopoietic cells by lentiviral vector delivery of a triple combination of anti-HIV shRNA, anti-CCR5 ribozyme, and a nucleolar-localizing TAR decoy*. *Mol Ther*, 2005. **12**(5): p. 900-9.
77. Anderson, J., et al., *Safety and efficacy of a lentiviral vector containing three anti-HIV genes--CCR5 ribozyme, tat-rev siRNA, and TAR decoy--in SCID-hu mouse-derived T cells*. *Mol Ther*, 2007. **15**(6): p. 1182-8.

78. DiGiusto, D.L., et al., *RNA-based gene therapy for HIV with lentiviral vector-modified CD34(+) cells in patients undergoing transplantation for AIDS-related lymphoma*. *Sci Transl Med*, 2010. **2**(36): p. 36ra43.
79. DiGiusto, D.L., et al., *Development of hematopoietic stem cell based gene therapy for HIV-1 infection: considerations for proof of concept studies and translation to standard medical practice*. *Viruses*, 2013. **5**(11): p. 2898-919.
80. Vinores, S.A., *Pegaptanib in the treatment of wet, age-related macular degeneration*. *Int J Nanomedicine*, 2006. **1**(3): p. 263-8.
81. 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.
82. 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.
83. 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.
84. Symons, R.H., *Small catalytic RNAs*. *Annu Rev Biochem*, 1992. **61**: p. 641-71.
85. Novick, R.P., *Contrasting lifestyles of rolling-circle phages and plasmids*. *Trends Biochem Sci*, 1998. **23**(11): p. 434-8.
86. Saville, B.J. and R.A. Collins, *RNA-mediated ligation of self-cleavage products of a Neurospora mitochondrial plasmid transcript*. *Proc Natl Acad Sci USA*, 1991. **88**(19): p. 8826-30.
87. 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.
88. Guo, H.C., et al., *Nucleotide sequence requirements for self-cleavage of Neurospora VS RNA*. *J Mol Biol*, 1993. **232**(2): p. 351-61.
89. 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.
90. 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**(10): p. 4686-90.

91. Lilley, D.M.J., *Ribozymes and RNA Catalysis*, ed. R.B. Sciences. 2007: The Royal Society of Chemistry.
92. 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.
93. 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.
94. 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.
95. 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.
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. 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.
98. Zhao, Z.Y., et al., *Nucleobase participation in ribozyme catalysis*. J Am Chem Soc, 2005. **127**(14): p. 5026-7.
99. 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.
100. 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.
101. 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.
102. 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.
103. 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.

104. 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.
105. Bonneau, E. and P. Legault, *NMR Localization of Divalent Cations at the Active Site of the Neurospora VS Ribozyme Provides Insights Into RNA-Metal Ion Interactions*. Biochemistry, 2013.
106. 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.
107. 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.
108. 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.
109. 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. 7730-5.
110. 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.
111. 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.
112. 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. 7003-8.
113. 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.
114. 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.

115. 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.
116. Lafontaine, D.A., D.G. Norman, and D.M. Lilley, *Folding and catalysis by the VS ribozyme*. Biochimie, 2002. **84**(9): p. 889-96.
117. 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.
118. 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.
119. 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**(4): p. 830-40.
120. 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.
121. Sood, V.D. and R.A. Collins, *Functional equivalence of the uridine turn and the hairpin as building blocks of tertiary structure in the Neurospora VS ribozyme*. J Mol Biol, 2001. **313**(5): p. 1013-9.
122. 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.
123. Lilley, D.M., *Mechanisms of RNA catalysis*. Philos Trans R Soc Lond B Biol Sci, 2011. **366**(1580): p. 2910-7.
124. 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.
125. Strobel, S.A., et al., *Biochemical detection of adenosine and cytidine ionization within RNA by interference analysis*. Nucleic Acids Res Suppl, 2003(3): p. 229-30.

126. Ouellet, J., M. Byrne, and D.M. Lilley, *Formation of an active site in trans by interaction of two complete Varkud Satellite ribozymes*. RNA, 2009. **15**(10): p. 1822-6.
127. Saenger, W., *Principles of nucleic acid structure*. 1984, New York: Springer-Verlag.
128. 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.
129. Rupert, P.B. and A.R. Ferre-D'Amare, *Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis*. Nature, 2001. **410**(6830): p. 780-6.
130. Rupert, P.B., et al., *Transition state stabilization by a catalytic RNA*. Science, 2002. **298**(5597): p. 1421-4.
131. Grum-Tokars, V., M. Milovanovic, and J.E. Wedekind, *Crystallization and X-ray diffraction analysis of an all-RNA U39C mutant of the minimal hairpin ribozyme*. Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 1): p. 142-5.
132. Grasby, J.A., et al., *Purine functional groups in essential residues of the hairpin ribozyme required for catalytic cleavage of RNA*. Biochemistry, 1995. **34**(12): p. 4068-76.
133. Shippy, R., A. Siwkowski, and A. Hampel, *Mutational analysis of loops 1 and 5 of the hairpin ribozyme*. Biochemistry, 1998. **37**(2): p. 564-70.
134. Pinard, R., et al., *Functional involvement of G8 in the hairpin ribozyme cleavage mechanism*. EMBO J, 2001. **20**(22): p. 6434-42.
135. Wilson, T.J., et al., *Importance of specific nucleotides in the folding of the natural form of the hairpin ribozyme*. Biochemistry, 2001. **40**(7): p. 2291-302.
136. Kuzmin, Y.I., C.P. Da Costa, and M.J. Fedor, *Role of an active site guanine in hairpin ribozyme catalysis probed by exogenous nucleobase rescue*. J Mol Biol, 2004. **340**(2): p. 233-51.
137. Kath-Schorr, S., et al., *General acid-base catalysis mediated by nucleobases in the hairpin ribozyme*. J Am Chem Soc, 2012. **134**(40): p. 16717-24.

138. Bevilacqua, P.C., *Mechanistic considerations for general acid-base catalysis by RNA: revisiting the mechanism of the hairpin ribozyme*. *Biochemistry*, 2003. **42**(8): p. 2259-65.
139. Han, J. and J.M. Burke, *Model for general acid-base catalysis by the hammerhead ribozyme: pH-activity relationships of G8 and G12 variants at the putative active site*. *Biochemistry*, 2005. **44**(21): p. 7864-70.
140. Nakano, S., D.M. Chadalavada, and P.C. Bevilacqua, *General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme*. *Science*, 2000. **287**(5457): p. 1493-7.
141. Klein, D.J., M.D. Been, and A.R. Ferre-D'Amare, *Essential role of an active-site guanine in glmS ribozyme catalysis*. *J Am Chem Soc*, 2007. **129**(48): p. 14858-9.
142. Cochrane, J.C., et al., *Structural and chemical basis for glucosamine 6-phosphate binding and activation of the glmS ribozyme*. *Biochemistry*, 2009. **48**(15): p. 3239-46.
143. Cochrane, J.C., S.V. Lipchock, and S.A. Strobel, *Structural investigation of the GlmS ribozyme bound to its catalytic cofactor*. *Chem Biol*, 2007. **14**(1): p. 97-105.
144. Chen, J., et al., *Identification of the catalytic Mg(2)(+) ion in the hepatitis delta virus ribozyme*. *Biochemistry*, 2013. **52**(3): p. 557-67.
145. Thaplyal, P., et al., *Thio effects and an unconventional metal ion rescue in the genomic hepatitis delta virus ribozyme*. *Biochemistry*, 2013. **52**(37): p. 6499-514.
146. 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.
147. Lipfert, J., et al., *The complete VS ribozyme in solution studied by small-angle X-ray scattering*. *Structure*, 2008. **16**(9): p. 1357-67.
148. Draper, D.E., *A guide to ions and RNA structure*. *RNA*, 2004. **10**(3): p. 335-43.
149. Draper, D.E., *RNA folding: thermodynamic and molecular descriptions of the roles of ions*. *Biophys J*, 2008. **95**(12): p. 5489-95.
150. Bowman, J.C., et al., *Cations in charge: magnesium ions in RNA folding and catalysis*. *Curr Opin Struct Biol*, 2012. **22**(3): p. 262-72.

151. Misra, V.K. and D.E. Draper, *On the role of magnesium ions in RNA stability*. Biopolymers, 1998. **48**(2-3): p. 113-35.
152. Eckstein, F., *Nucleoside phosphorothioates*. Annu Rev Biochem, 1985. **54**: p. 367-402.
153. Frey, P.A. and R.D. Sammons, *Bond order and charge localization in nucleoside phosphorothioates*. Science, 1985. **228**(4699): p. 541-5.
154. Jaffe, E.K. and M. Cohn, *Diastereomers of the nucleoside phosphorothioates as probes of the structure of the metal nucleotide substrates and of the nucleotide binding site of yeast hexokinase*. J Biol Chem, 1979. **254**(21): p. 10839-45.
155. Pecoraro, V.L., J.D. Hermes, and W.W. Cleland, *Stability constants of Mg²⁺ and Cd²⁺ complexes of adenine nucleotides and thionucleotides and rate constants for formation and dissociation of MgATP and MgADP*. Biochemistry, 1984. **23**(22): p. 5262-71.
156. 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.
157. Kovacheva, Y.S., et al., *The role of phosphate groups in the VS ribozyme-substrate interaction*. Nucleic Acids Res, 2004. **32**(21): p. 6240-50.
158. Smith, M.D., et al., *The ionic environment determines ribozyme cleavage rate by modulation of nucleobase pK_a*. RNA, 2008. **14**(9): p. 1942-9.
159. Stahley, M.R. and S.A. Strobel, *Structural evidence for a two-metal-ion mechanism of group I intron splicing*. Science, 2005. **309**(5740): p. 1587-90.
160. Collins, R.A., *The Neurospora Varkud satellite ribozyme*. Biochem Soc Trans, 2002. **30**(Pt 6): p. 1122-6.
161. 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.
162. Brion, P. and E. Westhof, *Hierarchy and dynamics of RNA folding*. Annu Rev Biophys Biomol Struct, 1997. **26**: p. 113-37.
163. 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.

164. Leontis, N.B. and E. Westhof, *Analysis of RNA motifs*. Curr Opin Struct Biol, 2003. **13**(3): p. 300-8.
165. Hendrix, D.K., S.E. Brenner, and S.R. Holbrook, *RNA structural motifs: building blocks of a modular biomolecule*. Q Rev Biophys, 2005. **38**(3): p. 221-43.
166. Quigley, G.J. and A. Rich, *Structural domains of transfer RNA molecules*. Science, 1976. **194**(4267): p. 796-806.
167. Pley, H.W., K.M. Flaherty, and D.B. McKay, *Three-dimensional structure of a hammerhead ribozyme*. Nature, 1994. **372**(6501): p. 68-74.
168. Jucker, F.M. and A. Pardi, *GNRA tetraloops make a U-turn*. RNA, 1995. **1**(2): p. 219-22.
169. Huang, S., Y.X. Wang, and D.E. Draper, *Structure of a hexanucleotide RNA hairpin loop conserved in ribosomal RNAs*. J Mol Biol, 1996. **258**(2): p. 308-21.
170. Stallings, S.C. and P.B. Moore, *The structure of an essential splicing element: stem loop IIa from yeast U2 snRNA*. Structure, 1997. **5**(9): p. 1173-85.
171. Huppler, A., et al., *Metal binding and base ionization in the U6 RNA intramolecular stem-loop structure*. Nat Struct Biol, 2002. **9**(6): p. 431-5.
172. Puglisi, E.V. and J.D. Puglisi, *HIV-1 A-rich RNA loop mimics the tRNA anticodon structure*. Nat Struct Biol, 1998. **5**(12): p. 1033-6.
173. Klosterman, P.S., et al., *Three-dimensional motifs from the SCOR, structural classification of RNA database: extruded strands, base triples, tetraloops and U-turns*. Nucleic Acids Res, 2004. **32**(8): p. 2342-52.
174. Sefcikova, J., et al., *The genomic HDV ribozyme utilizes a previously unnoticed U-turn motif to accomplish fast site-specific catalysis*. Nucleic Acids Res, 2007. **35**(6): p. 1933-46.
175. Harris, S. and S.J. Schroeder, *Nuclear magnetic resonance structure of the prohead RNA E-loop hairpin*. Biochemistry, 2010. **49**(29): p. 5989-97.
176. Ashraf, S.S., et al., *The uridine in "U-turn": contributions to tRNA-ribosomal binding*. RNA, 1999. **5**(4): p. 503-11.
177. Uhlenbeck, O.C., P.T. Lowary, and W.L. Wittenberg, *Role of the constant uridine in binding of yeast tRNAPhe anticodon arm to 30S ribosomes*. Nucleic Acids Res, 1982. **10**(11): p. 3341-52.

178. Dix, D.B., et al., *Effect of replacing uridine 33 in yeast tRNAPhe on the reaction with ribosomes*. J Biol Chem, 1986. **261**(22): p. 10112-8.
179. Lebars, I., et al., *Structure of 23S rRNA hairpin 35 and its interaction with the tylosin-resistance methyltransferase RlmAII*. EMBO J, 2003. **22**(2): p. 183-92.
180. Gutell, R.R., et al., *Predicting U-turns in ribosomal RNA with comparative sequence analysis*. J Mol Biol, 2000. **300**(4): p. 791-803.
181. Franch, T., et al., *Antisense RNA regulation in prokaryotes: rapid RNA/RNA interaction facilitated by a general U-turn loop structure*. J Mol Biol, 1999. **294**(5): p. 1115-25.
182. Doudna, J.A., *Hammerhead ribozyme structure: U-turn for RNA structural biology*. Structure, 1995. **3**(8): p. 747-50.
183. 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 tRNAPhe*. Biochemistry, 2004. **43**(1): p. 55-66.
184. Guo, L., E.M. Allen, and W.A. Miller, *Base-pairing between untranslated regions facilitates translation of uncapped, nonpolyadenylated viral RNA*. Mol Cell, 2001. **7**(5): p. 1103-9.
185. Brunel, C., et al., *RNA loop-loop interactions as dynamic functional motifs*. Biochimie, 2002. **84**(9): p. 925-44.
186. Wagner, E.G. and R.W. Simons, *Antisense RNA control in bacteria, phages, and plasmids*. Annu Rev Microbiol, 1994. **48**: p. 713-42.
187. Skripkin, E., et al., *Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro*. Proc Natl Acad Sci USA, 1994. **91**(11): p. 4945-9.
188. Paillart, J.C., et al., *Dimerization of retroviral RNA genomes: an inseparable pair*. Nat Rev Microbiol, 2004. **2**(6): p. 461-72.
189. Batey, R.T., S.D. Gilbert, and R.K. Montange, *Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine*. Nature, 2004. **432**(7015): p. 411-5.
190. Lemay, J.F., et al., *Folding of the adenine riboswitch*. Chem Biol, 2006. **13**(8): p. 857-68.

191. Serganov, A., et al., *Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs*. Chem Biol, 2004. **11**(12): p. 1729-41.
192. Blouin, S. and D.A. Lafontaine, *A loop loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control*. RNA, 2007. **13**(8): p. 1256-67.
193. Garst, A.D., et al., *Crystal structure of the lysine riboswitch regulatory mRNA element*. J Biol Chem, 2008. **283**(33): p. 22347-51.
194. Serganov, A., L. Huang, and D.J. Patel, *Structural insights into amino acid binding and gene control by a lysine riboswitch*. Nature, 2008. **455**(7217): p. 1263-7.
195. 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.
196. Khvorova, A., et al., *Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity*. Nat Struct Biol, 2003. **10**(9): p. 708-12.
197. Chi, Y.I., et al., *Capturing hammerhead ribozyme structures in action by modulating general base catalysis*. PLoS Biol, 2008. **6**(9): p. e234.
198. Dufour, D., et al., *Structure-function analysis of the ribozymes of chrysanthemum chlorotic mottle viroid: a loop-loop interaction motif conserved in most natural hammerheads*. Nucleic Acids Res, 2009. **37**(2): p. 368-81.
199. Kim, C.H. and I. Tinoco, Jr., *A retroviral RNA kissing complex containing only two G.C base pairs*. Proc Natl Acad Sci USA, 2000. **97**(17): p. 9396-401.
200. Lambert, D., et al., *The influence of monovalent cation size on the stability of RNA tertiary structures*. J Mol Biol, 2009. **390**(4): p. 791-804.
201. Weixlbaumer, A., et al., *Determination of thermodynamic parameters for HIV DIS type loop-loop kissing complexes*. Nucleic Acids Res, 2004. **32**(17): p. 5126-33.
202. Woodson, S.A., *Compact intermediates in RNA folding*. Annu Rev Biophys, 2010. **39**: p. 61-77.

203. Chang, K.Y. and I. Tinoco, Jr., *Characterization of a "kissing" hairpin complex derived from the human immunodeficiency virus genome*. Proc Natl Acad Sci USA, 1994. **91**(18): p. 8705-9.
204. Jossinet, F., et al., *Dimerization of HIV-1 genomic RNA of subtypes A and B: RNA loop structure and magnesium binding*. RNA, 1999. **5**(9): p. 1222-34.
205. Lorenz, C., N. Piganeau, and R. Schroeder, *Stabilities of HIV-1 DIS type RNA loop-loop interactions in vitro and in vivo*. Nucleic Acids Res, 2006. **34**(1): p. 334-42.
206. Chen, A.A., D.E. Draper, and R.V. Pappu, *Molecular simulation studies of monovalent counterion-mediated interactions in a model RNA kissing loop*. J Mol Biol, 2009. **390**(4): p. 805-19.
207. Tan, Z.J. and S.J. Chen, *Predicting ion binding properties for RNA tertiary structures*. Biophys J, 2010. **99**(5): p. 1565-76.
208. Singh, A., L. Sethaphong, and Y.G. Yingling, *Interactions of cations with RNA loop-loop complexes*. Biophys J, 2011. **101**(3): p. 727-35.
209. Sehdev, P., G. Crews, and A.M. Soto, *Effect of helix stability on the formation of loop-loop complexes*. Biochemistry, 2012. **51**(48): p. 9612-23.
210. Eguchi, Y. and J. Tomizawa, *Complex formed by complementary RNA stem-loops and its stabilization by a protein: function of ColE1 Rom protein*. Cell, 1990. **60**(2): p. 199-209.
211. Eguchi, Y. and J. Tomizawa, *Complexes formed by complementary RNA stem-loops. Their formations, structures and interaction with ColE1 Rom protein*. J Mol Biol, 1991. **220**(4): p. 831-42.
212. Gregorian, R.S., Jr. and D.M. Crothers, *Determinants of RNA hairpin loop-loop complex stability*. J Mol Biol, 1995. **248**(5): p. 968-84.
213. Lee, A.J. and D.M. Crothers, *The solution structure of an RNA loop-loop complex: the ColE1 inverted loop sequence*. Structure, 1998. **6**(8): p. 993-1005.
214. Di Primo, C., *Real time analysis of the RNAI-RNAIL-Rop complex by surface plasmon resonance: from a decaying surface to a standard kinetic analysis*. J Mol Recognit, 2008. **21**(1): p. 37-45.

215. Ennifar, E., et al., *Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site*. Nat Struct Biol, 2001. **8**(12): p. 1064-8.
216. 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.
217. Nair, T.M., D.G. Myszka, and D.R. Davis, *Surface plasmon resonance kinetic studies of the HIV TAR RNA kissing hairpin complex and its stabilization by 2-thiouridine modification*. Nucleic Acids Res, 2000. **28**(9): p. 1935-40.
218. Dayton, A.I., et al., *The trans-activator gene of the human T cell lymphotropic virus type III is required for replication*. Cell, 1986. **44**(6): p. 941-7.
219. Fisher, A.G., et al., *The trans-activator gene of HTLV-III is essential for virus replication*. Nature, 1986. **320**(6060): p. 367-71.
220. 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.
221. 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.
222. Furtig, B., et al., *NMR spectroscopy of RNA*. Chembiochem, 2003. **4**(10): p. 936-62.
223. Dieckmann, T. and P. Legault, *Biomolecular NMR Spectroscopy of Ribonucleic Acids*. eLS, 2008.
224. Friebolin, H., *Basic One- and Two-Dimensional NMR Spectroscopy*. 4, illustrated ed. 2005: John Wiley & Sons. 406.
225. Wüthrich, K., *NMR of proteins and nucleic acids*. 1986: Wiley.
226. Müller, S., *Nucleic Acids from A to Z*. 2008: Wiley.
227. Chary, K.V. and H.S. Atreya, *Structure determination of proteins in solution by nuclear magnetic resonance spectroscopy*. J Postgrad Med, 2002. **48**(1): p. 83-7.
228. Lipsitz, R.S. and N. Tjandra, *Residual dipolar couplings in NMR structure analysis*. Annu Rev Biophys Biomol Struct, 2004. **33**: p. 387-413.
229. Getz, M., et al., *NMR studies of RNA dynamics and structural plasticity using NMR residual dipolar couplings*. Biopolymers, 2007. **86**(5-6): p. 384-402.

230. Bothe, J.R., et al., *Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy*. Nat Methods, 2011. **8**(11): p. 919-31.
231. Legault, P. and A. Pardi, *³¹P chemical shift as a probe of structural motifs in RNA*. J Magn Reson B, 1994. **103**(1): p. 82-6.
232. Legault, P. and A. Pardi, *Unusual dynamics and pKa shift at the active site of a lead-dependant ribozyme*. J Am Chem Soc, 1997. **119**(28).
233. Gonzalez, R.L., Jr. and I. Tinoco, Jr., *Identification and characterization of metal ion binding sites in RNA*. Methods Enzymol, 2001. **338**: p. 421-43.
234. Batey, R.T., J.L. Battiste, and J.R. Williamson, *Preparation of isotopically enriched RNAs for heteronuclear NMR*. Methods Enzymol, 1995. **261**: p. 300-22.
235. Dingley, A.J. and S. Grzesiek, *Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide (²J(NN) couplings*. J Am Chem Soc, 1998. **120**(33): p. 8293-8297.
236. Zhang, O., et al., *Backbone ¹H and ¹⁵N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques*. J Biomol NMR, 1994. **4**(6): p. 845-58.
237. Schwieters, C.D., J.J. Kuszewski, and G.M. Clore, *Using Xplor-NIH for NMR molecular structure determination*. Prog Nucl Mag Res Sp, 2006. **48**(1): p. 47-62.
238. Salim, N.N. and A.L. Feig, *Isothermal titration calorimetry of RNA*. Methods, 2009. **47**(3): p. 198-205.
239. Lukavsky, P.J., et al., *Structure of HCV IRES domain II determined by NMR*. Nat Struct Biol, 2003. **10**(12): p. 1033-8.
240. Eichhorn, C.D. and H.M. Al-Hashimi, *Structural dynamics of a single-stranded RNA-helix junction using NMR*. RNA, 2014. **20**(6): p. 782-91.
241. Nikolova, E.N. and H.M. Al-Hashimi, *Preparation, resonance assignment, and preliminary dynamics characterization of residue specific ¹³C/¹⁵N-labeled elongated DNA for the study of sequence-directed dynamics by NMR*. J Biomol NMR, 2009. **45**(1-2): p. 9-16.

242. Ziegeler, M., et al., *NMR studies of HARI RNA secondary structures reveal conformational dynamics in the human RNA*. *Chembiochem*, 2012. **13**(14): p. 2100-12.
243. 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*, 2013.
244. Chothia, C. and A.M. Lesk, *The relation between the divergence of sequence and structure in proteins*. *EMBO J*, 1986. **5**(4): p. 823-6.
245. Dror, O., R. Nussinov, and H. Wolfson, *ARTS: alignment of RNA tertiary structures*. *Bioinformatics*, 2005. **21 Suppl 2**: p. ii47-53.
246. Rother, K., et al., *RNA and protein 3D structure modeling: similarities and differences*. *J Mol Model*, 2011. **17**(9): p. 2325-36.
247. Fiser, A., et al., *Evolution and physics in comparative protein structure modeling*. *Acc Chem Res*, 2002. **35**(6): p. 413-21.
248. Anfinsen, C.B., *Principles that govern the folding of protein chains*. *Science*, 1973. **181**(4096): p. 223-30.
249. Tinoco, I., Jr., O.C. Uhlenbeck, and M.D. Levine, *Estimation of secondary structure in ribonucleic acids*. *Nature*, 1971. **230**(5293): p. 362-7.
250. Magnus, M., et al., *Computational modeling of RNA 3D structures, with the aid of experimental restraints*. *RNA Biol*, 2014. **11**(5).
251. Leontis, N., et al., *RNA 3D Structure analysis and prediction*. *Nucleic Acids and Molecular Biology*. 2012: Springer Berlin Heidelberg.
252. Ya-Zhou, S., et al., *RNA structure prediction: Progress and perspective*. *Chinese Physics B*, 2014. **23**(7).
253. Das, R. and D. Baker, *Automated de novo prediction of native-like RNA tertiary structures*. *Proc Natl Acad Sci USA*, 2007. **104**(37): p. 14664-9.
254. Das, R., J. Karanicolas, and D. Baker, *Atomic accuracy in predicting and designing noncanonical RNA structure*. *Nat Methods*, 2010. **7**(4): p. 291-4.
255. Parisien, M. and F. Major, *The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data*. *Nature*, 2008. **452**(7183): p. 51-5.

256. Jonikas, M.A., et al., *Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters*. RNA, 2009. **15**(2): p. 189-99.
257. *Nanoscience and nanotechnologies: opportunities and uncertainties*, 2004, The Royal Society & The Royal Academy of Engineering.
258. Seeman, N.C., *Nucleic acid junctions and lattices*. J Theor Biol, 1982. **99**(2): p. 237-47.
259. Abels, J.A., et al., *Single-molecule measurements of the persistence length of double-stranded RNA*. Biophys J, 2005. **88**(4): p. 2737-44.
260. Kulic, I.M., et al., *Apparent persistence length renormalization of bent DNA*. Phys Rev E Stat Nonlin Soft Matter Phys, 2005. **72**(4 Pt 1): p. 041905.
261. Freier, S.M., et al., *Improved free-energy parameters for predictions of RNA duplex stability*. Proc Natl Acad Sci USA, 1986. **83**(24): p. 9373-7.
262. SantaLucia, J., Jr., *A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics*. Proc Natl Acad Sci USA, 1998. **95**(4): p. 1460-5.
263. Jaeger, L. and A. Chworos, *The architectonics of programmable RNA and DNA nanostructures*. Curr Opin Struct Biol, 2006. **16**(4): p. 531-43.
264. Jaeger, L. and N.B. Leontis, *Tecto-RNA: One-dimensional self-assembly through tertiary interactions*. Angew Chem Int Ed Engl, 2000. **39**(14): p. 2521-2524.
265. Grabow, W.W. and L. Jaeger, *RNA Self-Assembly and RNA Nanotechnology*. Acc Chem Res, 2014.
266. Leipply, D. and D.E. Draper, *Effects of Mg²⁺ on the free energy landscape for folding a purine riboswitch RNA*. Biochemistry, 2011. **50**(14): p. 2790-9.
267. Afonin, K.A., B. Lindsay, and B.A. Shapiro, *Engineered RNA nanodesigns for applications in RNA nanotechnology*. RNA Nanotechnol, 2013. **1**.
268. Horiya, S., et al., *RNA LEGO: magnesium-dependent formation of specific RNA assemblies through kissing interactions*. Chem Biol, 2003. **10**(7): p. 645-54.
269. Li, X., S. Horiya, and K. Harada, *An efficient thermally induced RNA conformational switch as a framework for the functionalization of RNA nanostructures*. J Am Chem Soc, 2006. **128**(12): p. 4035-40.

270. Afonin, K.A., et al., *Co-transcriptional assembly of chemically modified RNA nanoparticles functionalized with siRNAs*. Nano Lett, 2012. **12**(10): p. 5192-5.
271. Shu, Y., et al., *Fabrication of pRNA nanoparticles to deliver therapeutic RNAs and bioactive compounds into tumor cells*. Nat Protoc, 2013. **8**(9): p. 1635-59.
272. Grabow, W.W., et al., *Self-assembling RNA nanorings based on RNAI/II inverse kissing complexes*. Nano Lett, 2011. **11**(2): p. 878-87.
273. Behrouzi, R. and S.A. Woodson, *Rendering RNA in 3D*. Nat Methods, 2012. **9**(6): p. 552-3.
274. Feig, A.L., *The use of manganese as a probe for elucidating the role of magnesium ions in ribozymes*, CRC Press, 2000. **37**: p. 157-182.
275. Maderia, M., L.M. Hunsicker, and V.J. DeRose, *Metal-phosphate interactions in the hammerhead ribozyme observed by ³¹P NMR and phosphorothioate substitutions*. Biochemistry, 2000. **39**(40): p. 12113-20.
276. Maderia, M., T.E. Horton, and V.J. DeRose, *Metal interactions with a GAAA RNA tetraloop characterized by (³¹)P NMR and phosphorothioate substitutions*. Biochemistry, 2000. **39**(28): p. 8193-200.
277. Draper, D.E., *Strategies for RNA folding*. Trends Biochem Sci, 1996. **21**(4): p. 145-9.
278. Nixon, P.L., C.A. Theimer, and D.P. Giedroc, *Thermodynamics of stabilization of RNA pseudoknots by cobalt(III) hexaammine*. Biopolymers, 1999. **50**(4): p. 443-58.
279. Cowan, J.A., *Metallobiochemistry of RNA. Co(NH₃)₆(3+) as a probe for Mg²⁺(aq) binding sites*. J Inorg Biochem, 1993. **49**(3): p. 171-5.
280. Hardin, C.C., et al., *Fluorine-19 nuclear magnetic resonance studies of the structure of 5-fluorouracil-substituted Escherichia coli transfer RNA*. Biochemistry, 1986. **25**(19): p. 5699-709.
281. Chu, W.C., et al., *Fluorine-19 nuclear magnetic resonance as a probe of the solution structure of mutants of 5-fluorouracil-substituted Escherichia coli valine tRNA*. J Mol Biol, 1992. **227**(4): p. 1164-72.
282. Olejniczak, M., et al., *The bulge region of HIV-1 TAR RNA binds metal ions in solution*. Nucleic Acids Res, 2002. **30**(19): p. 4241-9.

283. Hammann, C., D.G. Norman, and D.M. Lilley, *Dissection of the ion-induced folding of the hammerhead ribozyme using ¹⁹F NMR*. Proc Natl Acad Sci USA, 2001. **98**(10): p. 5503-8.
284. Morton, T.A. and D.G. Myszka, *Kinetic analysis of macromolecular interactions using surface plasmon resonance biosensors*. Methods Enzymol, 1998. **295**: p. 268-94.
285. Palau, W. and C. Di Primo, *Single-cycle kinetic analysis of ternary DNA complexes by surface plasmon resonance on a decaying surface*. Biochimie, 2012. **94**(9): p. 1891-9.
286. Palau, W., et al., *Direct evidence for RNA-RNA interactions at the 3' end of the Hepatitis C virus genome using surface plasmon resonance*. RNA, 2013. **19**(7): p. 982-91.

