

Université de Montréal

**Étude de la relation entre structure, dynamique et fonction  
de l'ARN par l'ingénierie du ribozyme VS de *Neurospora***

par Nicolas Girard

Département de Biochimie et de Médecine Moléculaire  
Faculté de Médecine

Thèse présentée  
en vue de l'obtention du grade de PhD.  
en Bio-informatique

31 août 2018

© Nicolas Girard, 2018

## Résumé

L'activité et la fonction des ARN catalytiques sont étroitement liées à leur structure tridimensionnelle et à leur dynamique. Mieux comprendre la nature précise de ce lien permettrait potentiellement d'adapter ces ARN à de nouveaux substrats et modifier leurs caractéristiques cinétiques d'une manière rationnelle. Le ribozyme Varkud Satellite (VS) de *Neurospora* est un petit ribozyme nucléolytique caractérisé biochimiquement de manière approfondie. Son activité dépend d'une interaction boucle-boucle de type *kissing-loop* entre la boucle V du ribozyme et la boucle I de son substrat replié en tige-boucle. Ces caractéristiques en font un excellent modèle pour des études portant sur le lien entre structure, dynamique et activité des ARN catalytiques à des fins d'ingénierie.

Bien que notre compréhension de la structure du ribozyme VS se soit largement améliorée depuis le début de mes travaux de thèse, elle reste essentiellement d'une nature statique. Il est ainsi important d'une part d'approfondir les connaissances relatives au potentiel d'ingénierie du ribozymes VS et d'autre part de brosser un portrait dynamique de sa fonction de clivage dans le but de procéder à son ingénierie rationnelle. Dans cette optique, il a été entrepris de démontrer en premier lieu que le ribozyme VS était capable de reconnaître et cliver des substrats de longueur différentes mais également qu'il était possible de modifier la longueur de l'hélice V afin de cliver des substrats trop courts ou trop longs pour être clivés par le ribozyme naturel. Une fois cette démonstration faite, la possibilité de substituer l'interaction de type *kissing-loop* I/V, sous-domaine crucial pour la reconnaissance du substrat, par des interactions de même type et structurellement similaires a été confirmée. Finalement, les données cinétiques récoltées sur des ribozymes incorporant soit ces variantes, soit l'interaction naturelle ont été comparées aux capacités dynamiques intrinsèques de ces interactions *kissing-loop*, évaluées en dynamique moléculaire, afin de relier leur structure, leur dynamique et leur activité.

Ces travaux auront permis de mieux comprendre dans quelle mesure et de quelle façon le ribozyme VS pouvait être modifié pour être adapté à de nouveaux substrat. Mais plus généralement, ils nous ont offert une compréhension accrue des paramètres structuraux et dynamiques important à considérer afin de procéder à l'ingénierie rationnelle d'un ARN fonctionnel.

**Mots-clés :** Ribozyme VS de *Neurospora*, dynamique de l'ARN, ingénierie rationnelle, dynamique moléculaire

## Abstract

The activity and function of catalytic RNAs are closely related to their three-dimensional structure and dynamics. Better understanding the precise nature of this link would potentially allow to adapt these RNAs to cleave new substrates and modify their kinetic characteristics in a rational way. *Neurospora* Varkud Satellite (VS) ribozyme is a small nucleolytic ribozyme that has been very well characterized biochemically. Its activity depends on a kissing-loop interaction between loop V of the ribozyme and loop I of its substrate folded into a stem-loop. These features make it an excellent model for studying the link between structure, dynamics, and activity of catalytic RNAs for engineering purposes.

Although our understanding of the structure of the VS ribozyme has greatly improved since the beginning of my PhD., it remains essentially a static depiction. It is therefore important to know more about the engineering potential of the VS ribozyme and to draw a dynamic portrait of its cleavage function, in order to proceed to its rational engineering. From this perspective, efforts were made to demonstrate that the VS ribozyme could recognize and cleave substrates of different lengths and furthermore that it was possible to modify the length of helix V to cleave substrates that were too short or too long to be cleaved by the natural ribozyme. The possibility of substituting the I/V kissing-loop interaction, a subdomain crucial for the recognition of its substrate, for structurally similar interactions of the same type was then confirmed. Finally, the kinetic data collected on ribozymes containing either these variants or the natural interaction were compared to the intrinsic dynamics of the kissing-loop interactions, evaluated in molecular dynamics, in order to link their structure, dynamics and activity.

This work helped us better understand to what extent and how the VS ribozyme could be modified to be adapted to new substrates. More generally, it gave us an increased understanding of the important structural and dynamic parameters to take into account in order to rationally engineer a functional RNA.

**Keywords:** *Neurospora* VS ribozyme, RNA dynamics, rational engineering, molecular dynamics

# Table des matières

Résumé.....	i
Abstract.....	ii
Table des matières.....	i
Liste des tableaux.....	iv
Liste des figures.....	v
Liste des abréviations.....	viii
Remerciements.....	xii
1. Introduction.....	1
1.1. L'ARN.....	1
1.1.1. Les multiples rôles de l'ARN.....	1
1.1.2. Les structures de l'ARN.....	2
1.1.3. Les ARN catalytiques.....	8
1.1.4. Le ribozyme VS de <i>Neurospora</i> .....	11
1.2. La dynamique de l'ARN.....	17
1.2.1. Généralités.....	17
1.2.2. Une dynamique échelonnée dans le temps.....	18
1.2.3. L'étude de la dynamique de l'ARN.....	20
1.2.4. Dynamique et fonction des ARN catalytiques.....	21
1.3. La bio-informatique structurale.....	21
1.3.1. Principe général.....	21
1.3.2. L'annotation de structure et la recherche de motifs structuraux.....	22
1.3.3. La modélisation de structure 3D d'ARN.....	22
1.3.4. La dynamique moléculaire.....	23
1.3.5. L'analyse de la dynamique de l'ARN.....	29
1.4. L'ingénierie d'ARN.....	31
1.4.1. Généralités.....	31
1.4.2. Méthodes.....	31
1.4.3. Objectifs.....	33

1.5. Objectifs de la thèse .....	34
2. Étude de la compensation de longueur d'hélice dans le ribozyme VS de <i>Neurospora</i> .....	38
Mise en contexte .....	39
Abstract .....	41
Introduction.....	42
Material and Methods .....	43
Results.....	47
Discussion .....	51
References.....	54
Figures.....	59
Tables .....	64
Supporting Information.....	66
3. Ingénierie rationnelle du ribozyme VS de <i>Neurospora</i> par substitution de l'interaction kissing-loop I/V .....	69
Mise en contexte .....	70
Abstract.....	73
Introduction.....	74
Material and Methods .....	76
Results.....	79
Discussion .....	85
References.....	88
Figures.....	92
Tables.....	97
Supporting Information.....	98
4. Étude des liens entre structure et dynamique du complexe kissing-loop I/V et l'activité du ribozyme VS de <i>Neurospora</i> .....	108
Mise en contexte .....	109
Abstract.....	112
Introduction.....	113
Results and Discussion .....	115
References.....	122

Figures.....	125
Tables.....	134
Supporting Information.....	135
Experimental procedures .....	135
Supporting figures.....	140
References.....	153
5. Discussion.....	156
5.1. Conclusions majeures des articles présentés.....	156
5.1.1. La conformation active implique un angle entre les hélices I et V proche de la colinéarité.....	156
5.1.2. Le potentiel d'ingénierie du ribozyme VS et ses conséquences.....	156
5.1.3. La dynamique de l'interaction I/V est distinctive et suffisante pour la formation du site actif.....	157
5.1.4. L'activité du ribozyme est principalement dépendante de la dynamique de l'interaction I/V.....	158
5.1.5. Des différences structurales importantes reposent sur des mouvements locaux de faible amplitude .....	159
5.1.6. Le modèle dynamique du complexe I/V est prédictif de l'activité du ribozyme VS .....	160
5.2. Corollaires des résultats obtenus.....	160
5.2.1. Vers un modèle dynamique complet du ribozyme VS .....	160
5.2.2. Utilisation du ribozyme VS pour le clivage de substrats structurés .....	161
5.2.3. L'interaction boucle-boucle comme bloc de construction.....	162
5.2.4. Intégrer les modes dynamiques spécifiques à l'ARN dans l'évaluation des ARN fonctionnels.....	163
5.2.5. L'intégration de la dynamique dans la modélisation et l'ingénierie.....	163
5.3. Conclusion : passer de la structure native à l'ensemble natif.....	164
6. Bibliographie.....	i

## Liste des tableaux

Table 1.1. Liste des modèles et structures tridimensionnels du ribozyme VS. ....	15
Table 2.1. Summary of kinetic data for $R_0$ .....	64
Table 2.2. Summary of multiple-substrate kinetic data .....	65
Table 3.1. Filtered results of the bioinformatics search to identify KLIs from the PDB.....	97
Table S3.1. Values of $k_{obs}$ ( $\text{min}^{-1}$ ) for selected $S/R$ pairs measured with $[S] = 0.250$ nM and $[R] = 0.5$ $\mu\text{M}$ under standard conditions.....	98
Table 4.1. Characteristics of the first five PCs of motion for the I/V, L88/L22 and TAR/TAR* KL complexes. ....	134

## Liste des figures

Figure 1.1. Les ribonucléotides constitutifs de l'ARN. ....	2
Figure 1.2. Nomenclature et définition des paires de bases de l'ARN. ....	4
Figure 1.3. Les éléments de structure secondaire de l'ARN. ....	6
Figure 1.4. Exemple de structure secondaire, 3D et mécanismes catalytiques de petits ARN nucléolytiques. ....	10
Figure 1.5. Structures de l'état fermé du ribozyme VS en <i>trans</i> obtenu par cristallographie aux rayons X. ....	11
Figure 1.6. Structures de l'état ouvert du ribozyme VS en <i>trans</i> obtenu par RMN. ....	13
Figure 1.7. Mécanismes réactionnels de la transestérification par les ribozymes VS et hairpin. ....	14
Figure 1.8. Structures RMN et structures secondaires des sous-domaines du ribozyme VS impliqués dans la reconnaissance du substrat. ....	17
Figure 1.9. Hiérarchie des types de mouvements impliqués dans la dynamique de l'ARN. ....	19
Figure 1.10. Illustration de la fonction de potentiel d'un champ de force classique. ....	25
Figure 1.11. Schéma illustratif de la technique de T-REMD. ....	28
Figure 2.1. Primary and secondary structures of the substrate/ribozyme (S/R) pairs characterized here. ....	59
Figure 2.2. Single-substrate kinetic analysis of $R_0$ . ....	60
Figure 2.3. Simultaneous kinetic characterization of $R_{-3bp}$ with multiple substrates. ....	61
Figure 2.4. Identification of SLI/SLV models compatible with the kinetic data. ....	62
Figure 2.5. Model of an SLI/SLV complex compatible with the kinetic data. ....	63
Figure S2.1. Conformational analysis of the substrates used in this study by native gel electrophoresis. ....	66
Figure S2.2. Identification of a common cleavage site for all functioning S/R pairs. ....	67
Figure S2.3. Flowchart describing the stepwise procedure used to generate three-dimensional models. ....	68
Figure 3.1. Rational engineering of KLIs for substrate recognition by the <i>Neurospora</i> VS ribozyme. ....	92
Figure 3.2. Schematics and thermodynamic stability of selected KLIs. ....	93



Figure 3.3. Sequences and kinetic data of the $S_{TAR}/R_{TAR}$ variants. ....	94
Figure 3.4. SHAPE analysis of the $R_{TAR}$ ribozyme variants. ....	95
Figure 3.5. Three-dimensional models of substrate/ribozyme complexes. ....	96
Figure S3.1. Schematic representation of the computational workflow used to identify and characterize KLIs present in the PDB. ....	99
Figure S3.2. ITC studies of the L22/L88 interaction. ....	101
Figure S3.3. Sequences and kinetic data of the $SL88/RL22$ variants. ....	103
Figure S3.4. Sequences and kinetic data of the $SL22/RL88$ variants. ....	104
Figure S3.5. Conformational analysis of the substrate variants used in this study. ....	105
Figure S3.6. SHAPE analysis of the $R0$ ribozyme and related variants. ....	106
Figure 4.1. Substitution of the I/V kissing-loop interaction (KLI) in the <i>Neurospora</i> VS ribozyme. ....	125
Figure 4.2. Schematic representation of the parameters used to analyze the T-REMD trajectories. ....	127
Figure 4.3. Difference in conformational space between the I/V KL complex and surrogate KL complexes from Euler angle analysis. ....	128
Figure 4.4. Correlation between the interhelical angle $\theta$ and the $d_{pp}$ value for the I/V KL complex and surrogate KL complexes. ....	129
Figure 4.5. Cartoon representations of the first five PCs of motion for the I/V KL complex. ....	130
Figure 4.6. Identification of mobile and rigid residues in the I/V KL complex. ....	131
Figure 4.7. Improving the cleavage activity of a variant VS ribozyme for its substrate using a more dynamic KLI. ....	132
Figure S4.1. Summary of the T-REMD protocol. ....	141
Figure S4.2. Convergence analysis of the T-REMD simulations. ....	142
Figure S4.3. Correlation between the interhelical angle $\theta$ and the heavy-atom RMSD of the SLI internal loop for the I/V KL complex and surrogate KL complexes. ....	143
Figure S4.4. Scree plots of PC analysis for the T-REMD simulations. ....	144
Figure S4.5. Cartoon representations of the first five PCs of motion for the L88/L22 KL complex. ....	145

Figure S4.6. Cartoon representations of the first five PCs of motion for the TAR/TAR* KL complex.....	146
Figure S4.7. Identification of mobile and rigid residues .....	148
Figure S4.8. Results from T-REMD simulations of the TAR/TAR-S1 KL complex.....	149
Figure S4.9. Cartoon representations of the first five PCs of motion for the TAR/TAR-S1 KL complex.....	151

## Liste des abréviations

1D: une dimension

2D: deux dimensions

3D: trois dimensions

A: adénine

ACP: analyse en composante principale

ADN: acide désoxyribonucléique

ARN: acide ribonucléique

ARNi: ARN interférent

ARNm: ARN messenger

ARNnc: ARN non-codant

ARNr: ARN ribosomal

ARNt: ARN de transfert

C: cytosine

CGE: *comparative gel electrophoresis*

CP: *chemical probing*

DM: dynamique moléculaire

EM : *electron microscopy*

FRET: *Förster resonance energy transfer*

G: guanine

H: hoogsteen

HDV: *hepatitis delta virus*

HE: hoogsteen edge

L22/L88 : *loop 22/loop 88*

MCQ: *mean of circular quantities*

MI: *modification interference*

miARN: micro ARN

pARNi: petit ARN interférents

pARNn: petit ARN nucléaire

PDB : *protein database*

R: purine

REMD : *replica exchange molecular dynamics*

RMN : résonance magnétique nucléaire

RNaseP: ribonucléase P

S: sugar

SAXS: *small angles X-ray scattering*

SDM: *site directed mutagenesis*

SE: sugar edge

SELEX: *systematic evolution of ligand by exponential enrichment*

SHAPE: *selective 2' hydroxyl acylation by primer extension*

SL : *stem loop*

smFRET: *single molecule FRET*

TAR: *trans-activation response*

T-REMD: *temperature REMD*

U: uracile

VS : Varkud satellite

W: Watson

WC : Watson-Crick

Y : pyrimidine

*À Evelyne, mon amour, mon axe et mon socle.*

## Remerciements

Écrire une thèse est un travail extrêmement personnel qu'il est pourtant très difficile de réaliser sans le soutien de ses mentors, de ses collègues, de sa famille et d'un solide groupe d'amis.

Ma directrice de thèse, le Pr. Pascale Legault, a été un mentor dont l'exigence m'a permis de combattre mes penchants naturels pour le flou artistique et l'écriture un peu trop créative. Elle m'a aidé à révéler le scientifique qui dormait (profondément) en moi, m'a fait confiance plus que je ne me le faisais moi-même. Je sors de ces années de doctorat avec une précieuse collection d'expériences de recherche, parfaitement équipé pour la suite de ma carrière et surtout plus passionné que jamais. Pour cela, je tiens à la remercier sincèrement.

Le Pr. Sébastien Lemieux a été un co-directeur disponible, dont le mélange de décontraction et de précision m'a inspiré à considérer chaque problème et chaque incertitude comme un défi à relever plutôt que comme une difficulté insurmontable. Ses questions perçantes mais toujours posées de manière très calme m'ont également appris à prendre mon temps pour rassembler mes pensées et répondre précisément, même dans le feu de l'action d'un comité de thèse ou d'une présentation.

Ma mère et mon père n'ont jamais perdu espoir de me voir finir mes études un jour ou à tout le moins ont été très habiles pour dissimuler leurs doutes. Pour leur soutien sans faille et/ou leurs talents d'acteurs hors du commun, je les remercie chaleureusement.

L'excellence académique, c'est important. Rester psychologiquement équilibré tout au long d'un doctorat l'est aussi si on veut garder sa motivation lorsque la science n'est pas de notre côté. Pour m'avoir permis de traverser le creux des vagues, le calme plat, les cales sèches, les vents contraires et autres métaphores nautiques, sans ordre particulier, merci à Jobou, Alex, Dominique, Pat, Philippe Lalanne, Éric, Rico, les Rianes, Sam, Jérémie, Jen, Bruce, Éloi, Fred, Sylvain, Jocelyn, Phil R, Phil Lampron, Abraham L, Alix, Pete, DT, Carl, Kate, Hugh, Vanessa & Adrien, Mathieu, Matt, Caro et Julie. Vous connaissez beaucoup de moi et je ne pourrais pas en être plus heureux.

Je n'aurais probablement pas réussi à m'inscrire et à me maintenir au programme de bio-informatique si Elaine Meunier ne m'avais pas tenu la main et tapé sur la tête au bon moment pour m'aider dans mes démarches administratives. Son bon sens, sa gentillesse et sa connaissance surnaturelle des rouages du système me manqueront.

Enfin, je veux dédier cette thèse à la patience surhumaine et au cœur immense de la femme qui partage ma vie depuis maintenant plus de dix ans. Contrairement à moi, elle n'a jamais douté. Elle m'a poussé quand je n'avais plus d'énergie et porté quand je me laissais couler. Elle était là quand je me demandais si j'étais à ma place pour me dire d'arrêter mes niaiseries. Elle a parlé de mes quelques accomplissements à tout le monde comme si j'avais gagné le prix Nobel emballé dans le ruban de la médaille Fields. Et elle a relu mon introduction de thèse.

Eve, merci d'avoir été présente, solide et rassurante comme un phare dans la tempête. C'est sans doute un cliché mais c'est vrai : je t'aime et je n'y serais pas arrivé sans toi.

Je t'en dois quelques-unes.



# 1. Introduction

## 1.1. L'ARN

### 1.1.1. Les multiples rôles de l'ARN

L'acide ribonucléique ou ARN a longtemps été considéré comme une molécule principalement porteuse d'information génétique. L'existence du ribosome et, dans une moindre mesure de l'ARN de transfert, indiquait que le rôle de l'ARN ne pouvait complètement être restreint à la transmission d'information. L'absence d'autres exemples de molécules d'ARN fonctionnels semblait cependant confiner les ARN non-codants (ARNnc) à ces deux seuls exemples. À partir de la fin des années 60 et au fil de découvertes-clé, la polyvalence de l'ARN, autant dans ses capacités fonctionnelles que dans ses rôles au sein des processus cellulaires, s'est pourtant imposée comme un nouveau consensus.

En 1968, une nouvelle classe d'ARNnc, les petits ARN nucléaires (pARNn ou *snRNA*)[1, 2], est identifiée et son rôle dans l'épissage des ARNm reconnu dans les années 80 [3-5]. Vient ensuite la découverte des petits ARN nucléolaires (*snoRNA*) [6, 7], impliqué dans la maturation des ARN ribosomiaux (ARNr) et la modification chimique de nombreux ARNnc. Un champ de recherche très actif s'est ouvert en parallèle avec l'élucidation des mécanismes de régulation de l'expression génique au niveau traductionnel par l'interférence par ARN (ARNi) [8] puis la description du mécanisme d'action des microARN (miARN) et des petits ARN interférents (pARNi)[9].

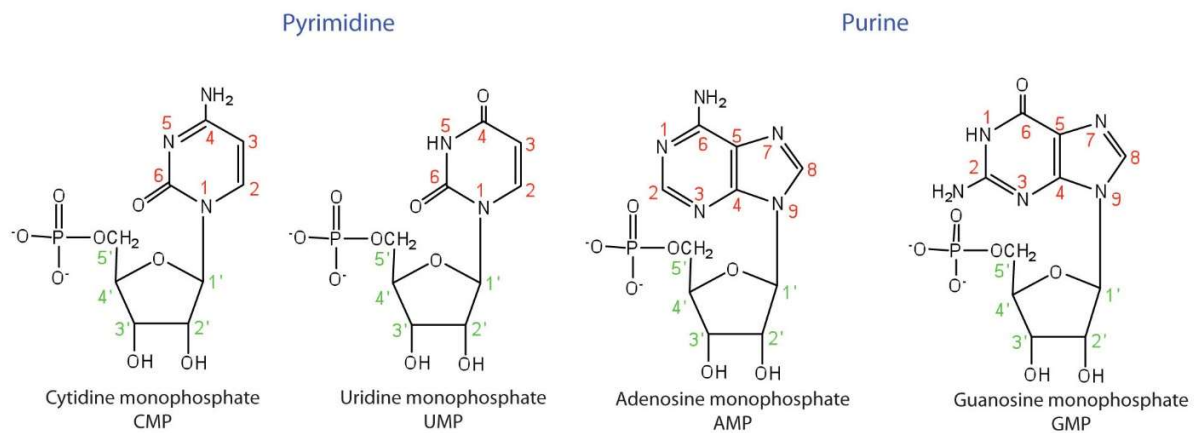
Ce sont la découverte des mécanismes de l'auto-épissage chez les introns du groupe I [10, 11] et du groupe II [12] ainsi que la mise en lumière du rôle de la ribonucléase P (RNase P)[13] dans le processus de maturation des ARNt qui ont réellement mis en lumière la capacité des ARN à catalyser des réactions chimiques sans l'aide de protéines. Découvert au milieu des années 50, le ribosome[14] est techniquement le premier ribozyme décrit mais la confirmation du statut de l'ARN ribosomique comme seul catalyseur de la formation de la liaison peptidique a nécessité plus d'un demi-siècle de recherches complémentaires[15].

Dans les dernières années, plusieurs nouveaux ARNnc ont été découverts à une fréquence très élevée et leur omniprésence a fait de l'étude de leur fonction un champ de recherche à part entière. Bien que le caractère fonctionnel de la plupart de ces ARN nouvellement découverts soit encore chaudement débattu [16], il est indéniable que les ARNnc sont indispensables dans de nombreux processus assurant la survie et la prolifération cellulaire et virale. En outre, leurs multiples fonctions semblent corroborer la possibilité d'une étape pré-protéique de l'évolution du vivant reposant principalement sur la molécule d'ARN ou hypothèse du « monde à ARN » [17-19].

## 1.1.2. Les structures de l'ARN

### 1.1.2.1. Composition chimique de L'ARN et structure primaire

L'unité de base de l'acide ribonucléique (ARN) est le ribonucléotide. Il est formé d'un ribose, le  $\beta$ -D-ribofuranose, phosphorylé en 5' et lié à une base azotée, soit une purine (adénine ou guanine), soit une pyrimidine (cytosine ou uracile), par son carbone C1' via une liaison N-glycosidique.



**Figure 1.1. Les ribonucléotides constitutifs de l'ARN.** Les chiffres en rouge désignent les atomes de la chaîne principale de la base azotée, les chiffres en vert ceux de la chaîne principale du ribose.

Adapté de <https://chem.libretexts.org/>

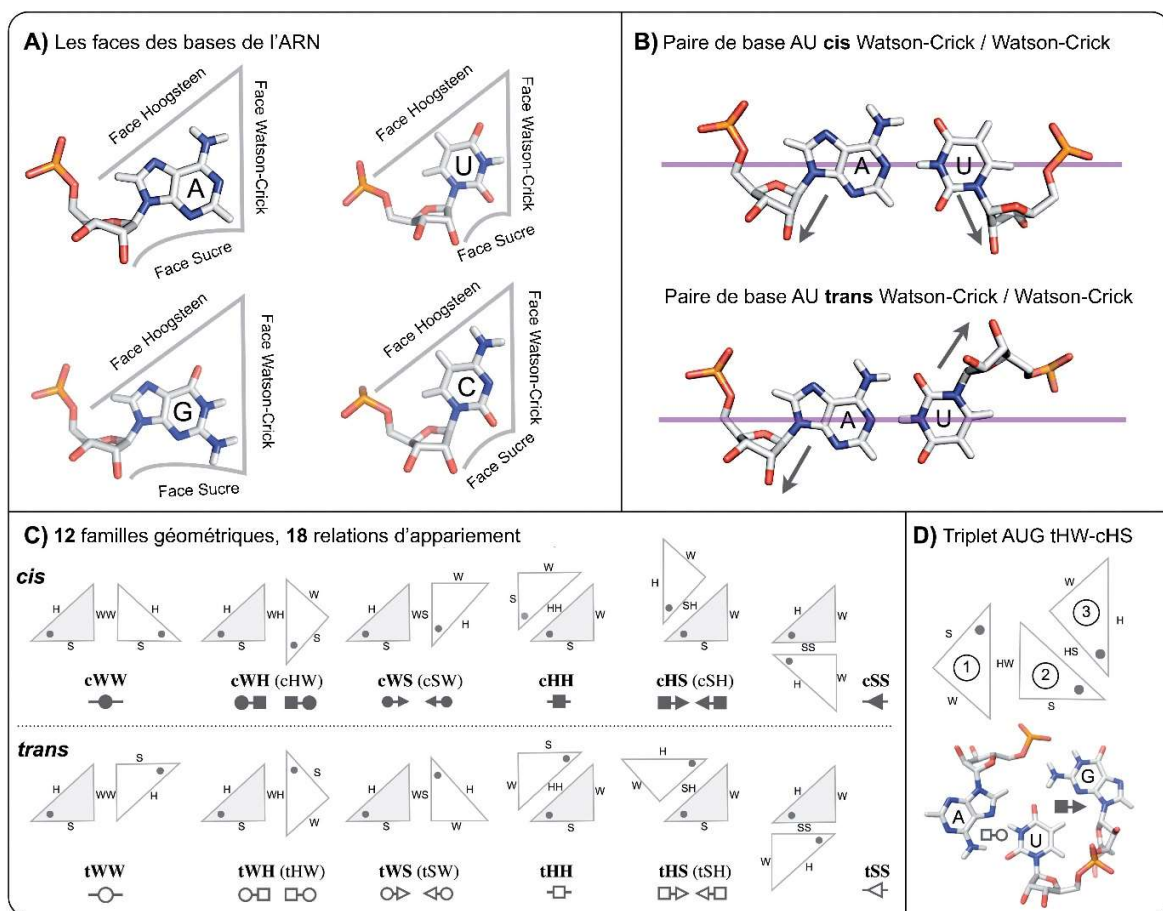
Pris individuellement, le ribonucléotide possède plusieurs caractéristiques chimiques et physiques notables qui vont avoir un impact important sur les niveaux d'organisation supérieurs dans lesquels il va être impliqué.

Tout d'abord, la conformation la plus favorable énergétiquement pour le ribose n'est pas un cycle dans lequel tous les atomes sont coplanaires. Les carbones C2' et C3' ne se trouvent pas dans le plan du cycle : cette propriété est appelée le « plissement du sucre », plus communément nommée par son appellation anglaise de *sugar pucker*. La forme prépondérante dans les structures hélicoïdales de l'ARN est la forme C3'-endo dans laquelle l'atome C3' est celui qui est le plus éloigné du plan. Le plissement va influencer la distance entre les groupements phosphates de deux nucléotides consécutifs au sein d'un polymère d'ARN ainsi que l'orientation relative de leurs riboses et bases. Il a été démontré que ce paramètre par nature très local pouvait avoir un impact global sur la fonction de certains ARN [20].

La plupart des atomes du phosphate et du ribose sont capables d'accepter et de donner des protons pour la création de liaisons hydrogènes [21, 22]. Le groupement hydroxyle en 2', qui distingue l'ARN de l'ADN, est un donneur de proton, ce qui explique à la fois la moindre stabilité chimique de l'ARN mais aussi certaines de ses capacités catalytiques qui seront détaillées en section 0. Ce groupement lui permet également de stabiliser des conformations particulières basées sur la création de ponts hydrogènes inter-nucléotides entre un ribose et un phosphate, une base ou un autre ribose [23, 24] et est la cause de la prépondérance de la forme C3'-endo du ribose au sein des structures hélicoïdales. Les donneurs et accepteurs de protons des bases azotées dépendent de l'identité de la base et sont répartis sur trois faces, appelées face ribose, face Hoogsteen et face Watson-Crick (Fig. 1.2A).

La base azotée a la capacité de tourner autour de la liaison N-glycosidique qui la relie au ribose. Elle peut par conséquent être orientée de deux façons différentes selon la valeur de l'angle de torsion  $\chi$  entre les atomes O4'-C1'-(N9-C4 ou N1-C2), appelées conformations *syn* lorsque les vecteurs C1-O4' et (N9-C4 ou N1-C2) sont orientés dans la même direction et conformation *anti* dans le cas contraire. Les deux conformations diffèrent également par l'orientation des donneurs et accepteurs de liaisons hydrogène de la base.

Enfin, le groupement phosphate situé en 5' est capable de lier chimiquement l'oxygène en 3' d'un autre ribonucléotide pour former une liaison phosphodiester afin de s'assembler en polymère. C'est l'identité des bases azotées constituant le polymère d'ARN, lue dans le sens du carbone 5' vers le carbone 3' de chaque nucléotide, qui constitue la structure primaire de l'ARN.



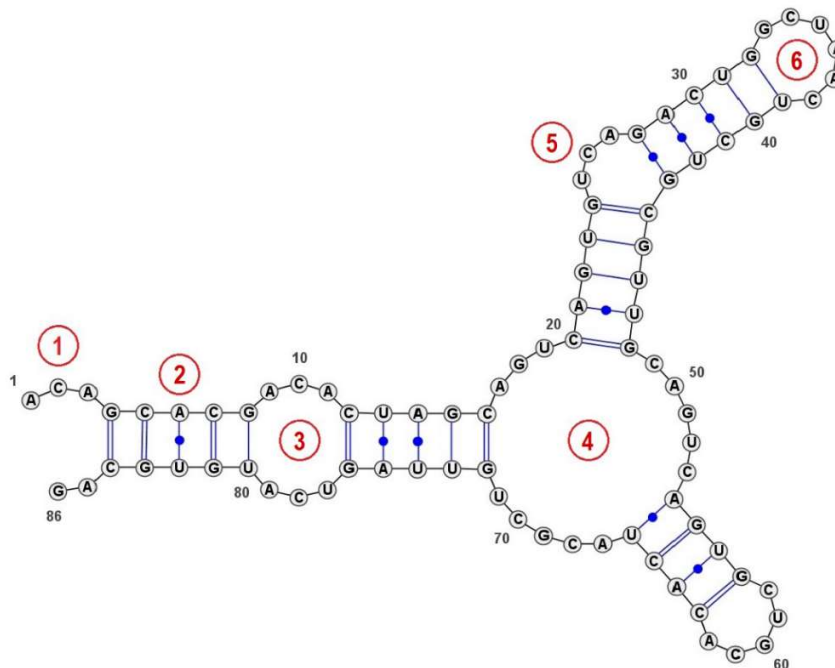
**Figure 1.2. Nomenclature et définition des paires de bases de l'ARN.** (A) Définition des faces Hoogsteen, Watson-Crick et Sucre pour les 4 bases de l'ARN. (B) Illustration des nomenclatures *cis* et *trans* pour une paire AU Watson-Crick / Watson-Crick. Les flèches représentent l'orientation de la liaison glycosidique. (C) Représentation schématique des paires de bases possibles dans l'ARN, avec leur nomenclature symbolique. (D) Exemple d'un triplet de bases d'ARN. Adapté de [25]

### 1.1.2.2. La structure secondaire de l'ARN

Les bases azotées ont tendance à s'apparier, c'est-à-dire à créer des liaisons hydrogènes entre leurs donneurs et accepteurs de protons respectifs. La nomenclature des paires ainsi créées dépend des faces impliquées [26] (Figure 1.2C). Une paire réunissant deux bases qui interagissent par leurs faces Watson-Crick est ainsi nommé paire Watson-Crick (WC). Les paires WC G-U, G-C et A-U sont appelées paires canoniques. L'appariement canonique entre les bases de deux brins d'ARN de sens contraire (Figure 1.2B) (paire de base *cis*) forme une structure en double hélice qui est la forme la plus commune de repliement de l'ARN. La structure secondaire de l'ARN représente la distinction entre ces hélices et les éléments au sein desquels les nucléotides ne sont pas appariés de manière canonique.

Bien que l'appariement de base soit un phénomène majeur lorsqu'il s'agit de considérer la formation et la stabilité de la structure secondaire de l'ARN, un autre phénomène appelé « empilement de bases » [27] doit être pris en compte. L'empilement de base est une interaction d'attraction non-covalente entre deux cycles aromatiques, de nature essentiellement électrostatique, qui stabilise les bases azotées de l'ARN lorsqu'elles se présentent parallèlement l'une à l'autre, à faible distance et avec un léger décalage axial ou perpendiculairement l'une à l'autre.

C'est au niveau de la structure secondaire que l'on voit apparaître les premiers exemples de motifs structuraux. Un motif est un arrangement tridimensionnel de ribonucléotides, récurrent et défini par l'identité de ses bases azotées, les ponts hydrogène et les interactions d'empilement de bases en son sein et la configuration de son squelette sucre-phosphate. Concrètement, la plupart des hélices d'ARN sont de la forme A. Les éléments non-appariés en revanche peuvent s'organiser en trois dimensions de multiples façons. Les motifs structuraux sont importants pour deux raisons : d'une part car ils renseignent sur l'organisation des hélices qui sont rattachées aux éléments non-appariés qui les contiennent [28]; d'autre part, parce qu'ils permettent de prédire l'organisation tridimensionnelle des séquences de nucléotides qui respectent leur définition. Cette organisation a un fort impact sur la disponibilité des divers groupes fonctionnels des nucléotides concernés et renseigne donc sur leur capacité à former des interactions tertiaires et/ou des interactions avec d'éventuels ions, protéines ou autres.



**Figure 1.3. Les éléments de structure secondaire de l'ARN.** (1) Extrémité en surplomb en 5'. (2) Hélice. (3) Boucle interne. (4) Jonction multi-voies, ici une jonction 3-voies. (5) Bulge. (6) Boucle terminale. Image libre de droit, adaptée d'une représentation générée par VARNA [29].

Les éléments non-appariés se classent en quatre catégories (Figure 1.3):

Les **extrémités en surplomb** qui sont les nucléotides situés en 5' ou 3' d'un ARN et ne sont impliqués dans aucun appariement canonique une fois la molécule repliée.

Les **boucles terminales** sont situées à l'extrémité d'une hélice formée d'un brin d'ARN replié sur lui-même. Au niveau de boucles terminales, deux motifs courants sont les *U-turns* [30, 31] et les boucles GNRA [32]. Dans les deux cas, il s'agit de boucles constituées d'au moins quatre nucléotides, suivant une séquence particulière (UNRN, N représentant n'importe quel base et R une purine pour le *U-turn* et GNRA), stabilisées par des liaisons hydrogène et des interactions d'empilement de bases et présentant un virage serré du squelette sucre-phosphate entre le premier et second nucléotide qui en inverse le sens. Les boucles terminales qui présentent l'un de ces deux motifs sont particulièrement stables [32, 33] et voient certaines bases rendues disponibles pour d'autres interactions.

Les **boucles internes** sont des structures non-appariées connectant deux hélices; les renflements sont des cas particuliers de boucle interne ne comprenant des nucléotides non-appariés que sur un brin. Les boucles internes présentent des motifs particuliers tels que le bulged-G [34] et le kink-turn [35]. Ces deux motifs entraînent une déformation importante du squelette sucre-phosphate qui provoque l'exposition de résidus au solvant et casse l'alignement des hélices connectées par le motif.

Les **jonctions multivoies** quant à elles connectent trois hélices ou plus et influencent fortement l'organisation des hélices qui y sont connectées. Au sein de certains ARN catalytiques, notamment le ribozyme hammerhead dont la séquence minimale se replie en une simple jonction trois voies, les caractéristiques structurales et dynamiques de la jonction sont cruciales pour en comprendre l'activité. Des jonctions multivoies de haut ordre, connectant jusqu'à 10 hélices ont été décrites et classifiées [36-38] sur la base de l'empilement coaxial (voir section suivante) de leurs hélices, leurs orientations respectives et leur capacité à se mouvoir dans l'espace.

### 1.1.2.3. La structure tertiaire de l'ARN

Ces structures secondaires sont à leur tour capables d'interagir entre elles ce qui permet à l'ARN d'adopter sa structure tertiaire. Les éléments de structure tertiaire de l'ARN peuvent ainsi être classés en trois catégories :

Les **interactions entre domaines hélicoïdaux**. Les domaines hélicoïdaux interagissent entre eux principalement par empilement coaxial [39, 40], phénomène basé sur les interactions  $\pi$  entre les bases des deux domaines hélicoïdaux séparés par une boucle, une jonction ou un motif plus complexe et qui provoque leur alignement selon un axe commun.

Les **interactions entre domaines hélicoïdaux et résidus non-appariés**. Ce type d'interaction intervient lorsque les résidus d'une boucle terminale ou interne se trouvent à proximité du sillon d'une hélice. Il nécessite des boucles prédisposées à exposer certains de leurs résidus et donc fréquemment organisées selon un motif structural au niveau de la structure secondaire. Le ribose zipper [40] est un motif représentatif de ce type d'interactions et associe une boucle terminale ou interne et une hélice au travers de liaisons hydrogène ribose/ribose ainsi

que des liaisons hydrogène base/ribose entre les deux brins de sens contraire. Ce motif permet à l'ARN d'être replié d'une manière plus condensée et est fréquemment retrouvé dans l'ARNr.

Les **interactions entre domaines non-appariés**. De même que pour le type d'interactions précédent, elles nécessitent des résidus exposés. Il s'agit principalement de combinaisons entre boucles, représentées par le motif d'interaction boucle-boucle (ou *kissing-loop*) [41] et entre boucle et extrémités en surplomb, appelé pseudonoeud [42]. Ces interactions sont particulièrement importantes pour l'activité de certains ribozymes comme les ribozyme VS, hairpin et twister, dont les activités de clivage dépendent respectivement d'une interaction boucle-boucle [43], d'une interaction entre boucles internes [44] et d'un pseudonoeud [45].

#### **1.1.2.4. L'ARN, molécule tridimensionnelle**

Les termes de structure tertiaire et tridimensionnelles sont souvent utilisés d'une manière interchangeable dans les publications relatives à la structure de l'ARN ; ce ne sera pas le cas dans la présente thèse. En effet, il existe une différence importante entre ces deux termes : la structure tertiaire est définie par la relation entre plusieurs éléments de structure secondaire, au travers de liaisons non-covalentes. La structure tridimensionnelle (ou 3D) de l'ARN est définie par la position précise des atomes constitutifs de la molécule étudiée, leurs coordonnées dans l'espace cartésien, position qui n'est pas complètement renseignée par la connaissance de la structure tertiaire.

La distinction est particulièrement importante lorsqu'il s'agit d'étudier les ARN catalytiques puisque la réaction chimique qu'ils promeuvent est la plupart du temps dépendante d'un positionnement extrêmement précis d'un groupe d'atome au sein du site catalytique. De plus, le caractère très dynamique de l'ARN implique qu'une structure tertiaire peut en fait signifier plusieurs structure 3D très différentes.

### **1.1.3. Les ARN catalytiques**

#### **1.1.3.1. Généralités**

Un ARN catalytique ou ribozyme est une enzyme composée de ribonucléotides, c'est-à-dire un ARN capable de catalyser une réaction chimique sans l'aide de protéines. Depuis la



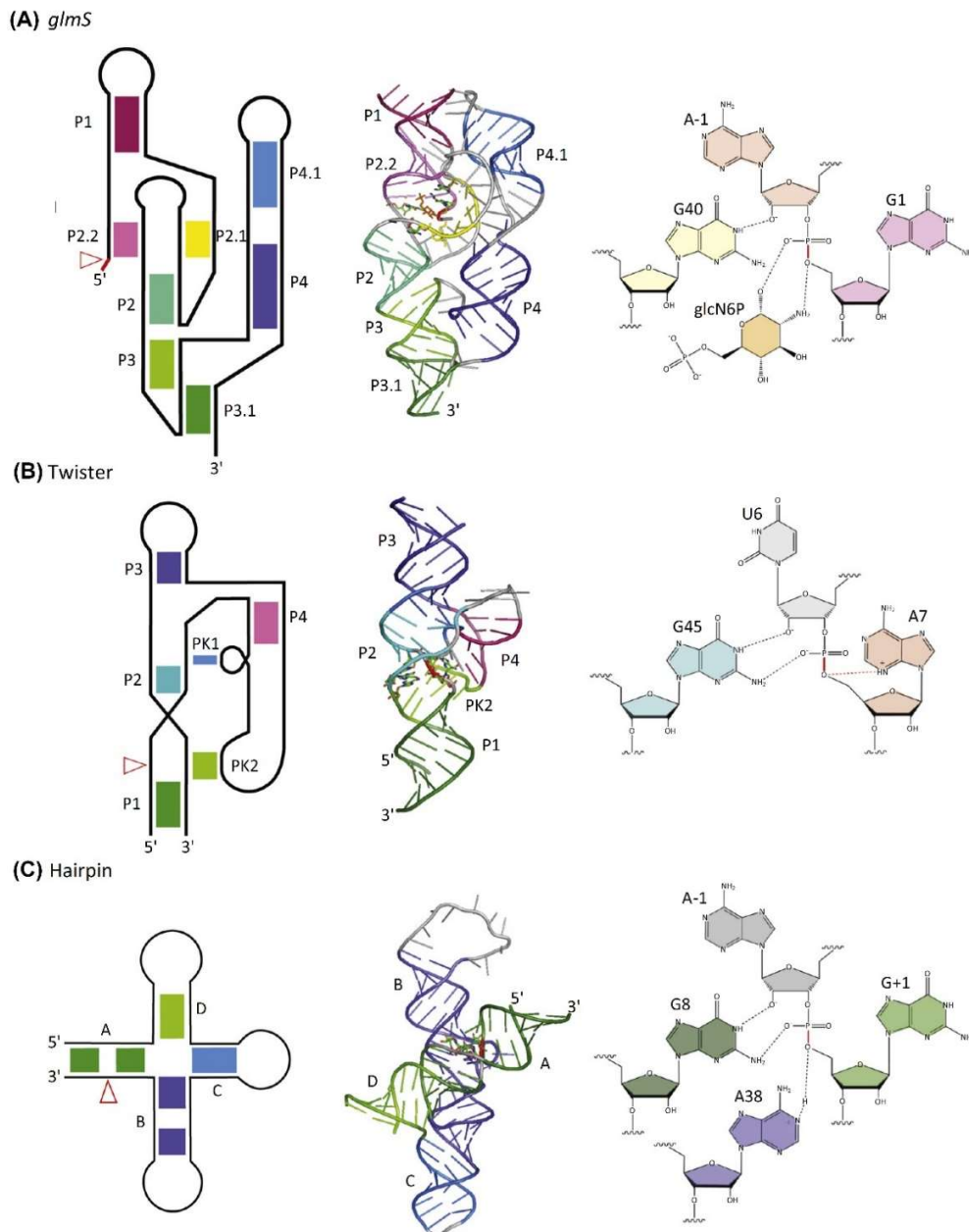
découverte des introns du groupe I et de la RNase P, plusieurs autres classes de ribozymes ont été mis au jour dans les génomes procaryotes, eucaryotes et viraux. Ces ribozymes naturels ont des fonctions variées au sein de leurs organismes; cependant, hormis le ribosome, ils ne possèdent qu'un seul type de fonction : le clivage d'une liaison phosphodiester, parfois réversible. Cette fonction repose sur la catalyse de deux types de réactions, soit l'hydrolyse par la RNase P et la transestérification par tous les autres.

Le regain d'intérêt apporté à l'origine de la catalyse chez les ARN et à leur rôle avant l'apparition des protéines a mené à une intense recherche sur le potentiel catalytique des ribozymes. De nombreux ribozymes artificiels ont ainsi été créés comme preuve de concept d'une forme de vie uniquement basée sur la molécule d'ARN, qui implique une capacité à s'auto-répliquer. Ces ribozymes sont ainsi capable de catalyser une vaste gamme de réactions chimiques qui vont de la création de lien glycosidique pour la synthèse de ribonucléotides [46] à la ligation 3'-hydroxyl / 5'-triphosphate dépendante d'une matrice d'ARN [47] utilisée pour l'amplification et la transcription inverse [48], en passant par la triphosphorylation en 5' [49].

### **1.1.3.2. Les petits ARN nucléolytiques**

La classe des petits ARN nucléolytiques est celle qui regroupe le plus de membres parmi les ribozymes naturels. Elle comprend les ribozymes glmS [50, 51], pistol [52], twister [53], twister sister [52], hatchet [52], hammerhead [54], Hepatitis Delta Virus (HDV) [55], hairpin [56] et Varkud Satellite (VS). Ces ribozymes catalysent tous la même réaction de transestérification, parfois de manière réversible, catalysée par un mécanisme acide-base général. Les rôles d'acides et de bases sont le plus souvent joués par des nucléotides, à l'exception du ribozyme glmS chez qui la glucosamine-6-phosphate sert d'acide général [57] et le ribozyme HDV, chez qui la base générale est un ion magnésium hydraté [58]; des ions magnésium sont également probablement impliqués dans la modification du micro-environnement chimique au site actif chez le ribozyme hammerhead [59]. Ces différences mettent en lumière la diversité du répertoire chimique employé par les petits ribozymes nucléolytiques pour catalyser leur réaction de transestérification. Ils sont tous de taille réduite (<200 nucléotides) et reconnaissent tous un substrat simple brin à l'exception notable du ribozyme VS qui reconnaît un substrat en tige-boucle.

Ces ribozymes sont particulièrement adaptés pour une analyse systématique du lien entre structure, dynamique et fonction des ARN en raison de leur petite taille, de la relative simplicité de leur site actif et de leurs interactions tertiaires.

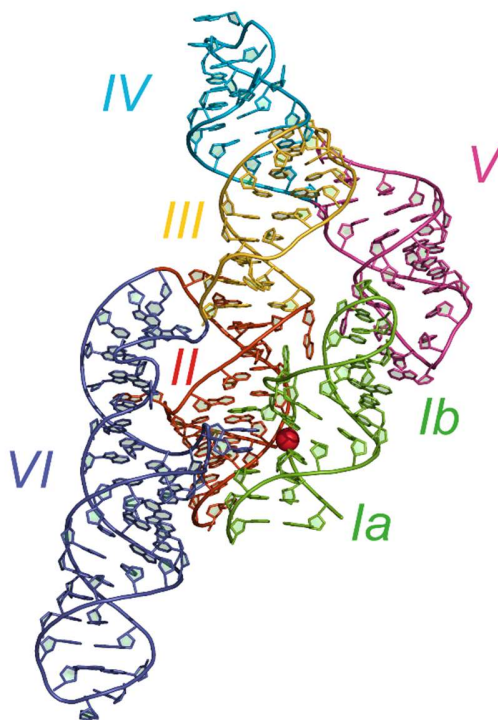


**Figure 1.4.** Exemple de structure secondaire, 3D et mécanismes catalytiques de petits ARN nucléolytiques. (A) Ribozyme *glmS*. (B) Ribozyme twister. (C) Ribozyme hairpin. Adapté de [60].

## 1.1.4. Le ribozyme VS de *Neurospora*

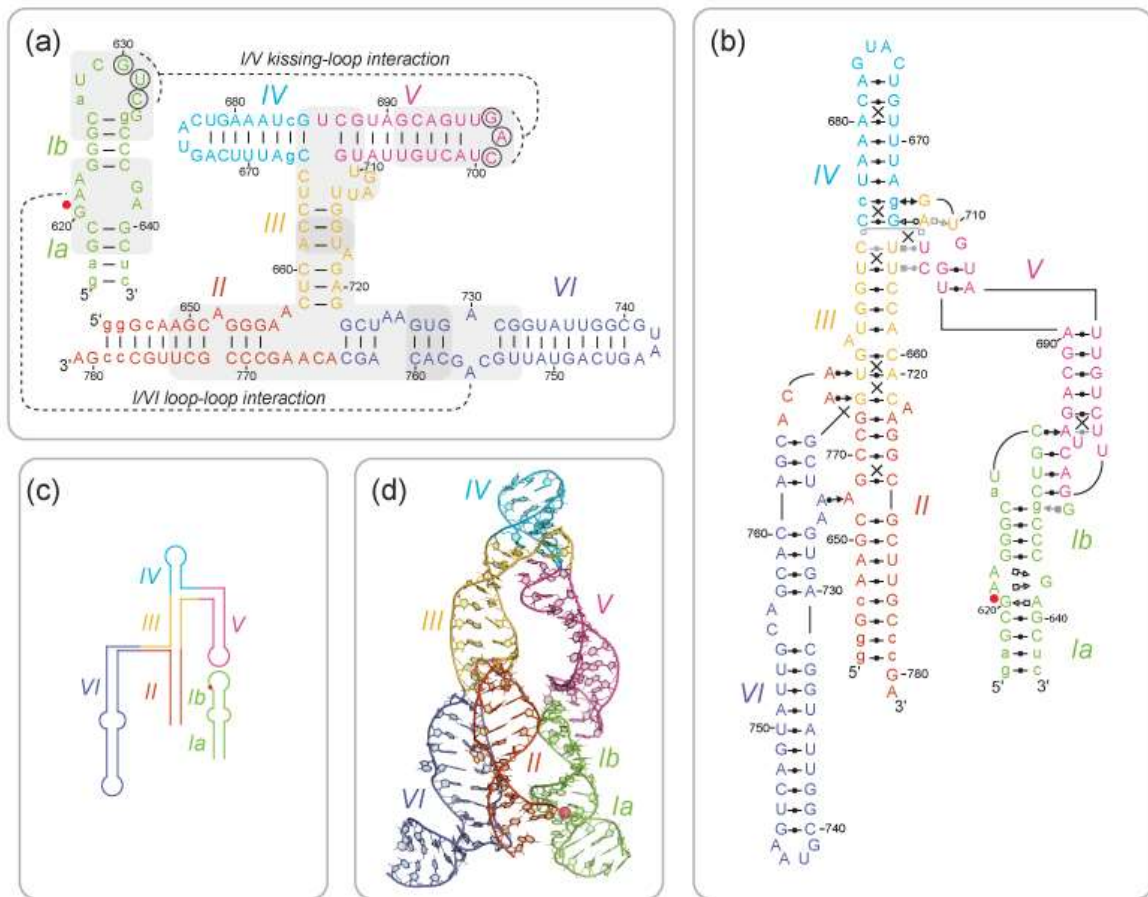
### 1.1.4.1. Vue d'ensemble

Le ribozyme VS a été découvert il y a plus de 25 ans au sein de l'ARN Varkud Satellite (VS), un ARN mitochondrial que l'on retrouve dans certains isolats du champignon filamenteux du genre *Neurospora* [61, 62]. C'est une molécule capable d'auto-clivage et d'auto-ligation; ces deux activités sont essentielles pour son cycle de réplication mais n'ont pas de rôle connu au sein de l'organisme hôte. *In vitro*, une séquence d'approximativement 160 nucléotides, organisée en six régions hélicoïdales (I–VI), est suffisante pour l'activité d'auto-clivage (Figure 1.5a)[63, 64].



**Figure 1.5. Structures de l'état fermé du ribozyme VS en *trans* obtenu par cristallographie aux rayons X.** Représentation tronquée d'une forme *trans* de la structure obtenue par cristallographie d'une variante G<sub>638</sub>A du ribozyme VS (code PDB 4R4V). Le phosphore du phosphate clivable est représenté par une sphère rouge. Les hélices II à V appartiennent à un protomère tandis que l'hélice I est celle du second protomère. Adapté de [65]

Plusieurs études ont fourni des modèles tridimensionnels à basse résolution [66-71] qui ont révélé des structures primaires, secondaires et tertiaires uniques parmi les ribozymes connus (Table 1.1). Ce n'est que récemment que des structures tridimensionnelles à haute résolution sont devenues disponibles et ont permis de se faire une meilleure idée d'ensemble de sa structure et sa dynamique (Table 1.1). Malgré sa nature hautement dynamique ainsi que de sa tendance à former de complexes multimériques en solution, un modèle obtenu par spectroscopie de résonance magnétique nucléaire (RMN) [72] ainsi que trois structures cristallines [65, 73] du ribozyme VS ont été publiées. Le modèle de RMN est celui d'un ribozyme en *trans* en complexe avec son substrat replié en tige-boucle et a été obtenu grâce à une stratégie dite de *divide-and-conquer* qui a permis de caractériser individuellement les sous-domaines cruciaux du ribozyme puis de les assembler en un modèle complet. Les deux structures cristallines ont été déterminées à partir de dimères de variantes du ribozyme en *cis* dans lesquels le substrat de l'un des deux protomères interagit avec le domaine catalytique de l'autre. Là où les structures cristallines présentent un ribozyme VS dans un état dit « fermé » (Figure 1.5), proche de la conformation active, les expériences de RMN ont produit un modèle représentant un ensemble d'états dit « ouvert » (Figure 1.6d), préalable à la formation du site actif et compatible avec les données biophysique et de *chemical probing* [70, 74].

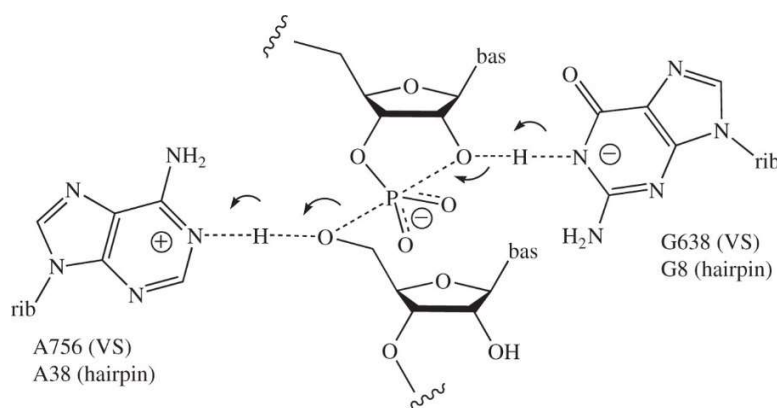


**Figure 1.6. Structures de l'état ouvert du ribozyme VS en *trans* obtenu par RMN. (a) Structure secondaire. Le site de clivage est indiqué par un point rouge. (b) Représentation schématique de la structure tertiaire du modèle de RMN du VS complet en *trans*. (c) Représentation schématique de l'orientation relative des domaines hélicoïdaux (d) Modèle tridimensionnel du ribozyme VS complet en *trans* obtenu par RMN. Adapté de [75].**

#### 1.1.4.2. La fonction de clivage

De nombreuses études biochimiques, biophysiques et cinétiques ont permis de caractériser de manière précise la fonction du ribozyme VS et l'effet des expériences de mutagenèse sur celle-ci [76-80]. Ces études supportent une catalyse basée sur un mécanisme acide-base général similaire à celui du ribozyme hairpin (Figure 1.6) qui requiert l'association intime de deux boucles internes afin de former le site actif au sein duquel deux bases, provenant

chacune de l'une des deux boucles, jouent un rôle-clé dans le mécanisme catalytique [78, 79]. L'une de caractéristiques distinctives du ribozyme VS est que l'une des deux boucles est située au sein du domaine catalytique (la boucle A756, aussi appelée boucle A730), alors que l'autre se trouve dans la structure du substrat; on peut ainsi voir le mécanisme de clivage comme une catalyse assistée par le substrat. Ce mécanisme repose sur deux résidus importants: le G638 d'une part, qui agit comme base générale et active l'oxygène en 2' pour lui permettre d'agir comme groupe nucléophile. Le résidu A756 d'autre part, qui agit comme acide général et donne un proton au groupe partant, soit l'oxygène en 5' [81].



**Figure 1.7. Mécanismes réactionnels de la transestérification par les ribozymes VS et hairpin.**

Les flèches indiquent les attaques par l'oxygène et le transfert de protons. Tiré de [80]

Les profils d'activité en fonction du pH de la réaction de clivage correspondent à un modèle de double ionisation avec des valeurs de pKa modifiées à 5.2–5.8 et 8.3–8.4 pour les résidus A756 and G638, respectivement.[81-84] De plus, la nécessité de la présence d'ions métalliques pour la modification de ces valeurs modifiées de pKa indiquent un rôle indirect des métaux dans la réaction.[84]. La réaction de clivage engendre des produits avec des extrémités 2',3'-phosphate cyclique et 5'-OH, comme on peut en retrouver à l'issue de réactions catalysées par d'autres petits ribozymes nucléolytiques comme le hairpin, le hammerhead ou le glmS [85, 86]. Un clivage efficace en *trans* du substrat est rendu possible grâce à un mécanisme complexe de reconnaissance (Figure 1.7), qui implique la formation d'une interaction boucle-boucle dépendante de la présence de magnésium, un changement de conformation du substrat d'un état *unshifted* à *shifted* (Fig 1.7a et 1.7b) et, comme mentionné précédemment, l'association intime de la boucle G638 avec la boucle A756 pour former le site actif [43, 87-89]

**Table 1.1. Liste des modèles et structures tridimensionnels du ribozyme VS. Adapté de [75].**

Sous-domaine	Approche Expérimentale <sup>1</sup>	Code PDB	Ion Métalliques Divalents dans le Tampon <sup>2</sup>	Ions Métalliques dans la Structure	Référence
<b>SLI unshifted</b>	RMN	1E4P	Aucun		[90]
	RMN	1HWQ	Aucun		[91]
<b>SLI shifted</b>	RMN	1OW9	Aucun		[92]
<b>SLV</b>	RMN	1TBK	Aucun		[93]
	RMN	1YN1	≥20 mM Mg <sup>2+</sup>		[94]
	RMN	1YN2	≥20 mM Mg <sup>2+</sup> et 20 μM Mn <sup>2+</sup>	4 Mn(H <sub>2</sub> O) <sub>6</sub> <sup>2+</sup>	[94]
	DM		Aucun; 40 mM Mg <sup>2+</sup>		[95]
<b>Interaction boucle-boucle I/V</b>	CP/MI/SDM		0–15 mM Mg <sup>2+</sup>		[66]
	RMN	2MI0	5 mM Mg <sup>2+</sup>		[96]
	DM				[97]
<b>Boucle A<sub>756</sub></b>	RMN	1TJZ	Aucun		[98]
	RMN	2L5Z	5 mM Mg <sup>2+</sup>		[99]
	RMN	2MIS	5 mM Mg <sup>2+</sup> et 10 μM Mn <sup>2+</sup> ; 4.75 mM Mg <sup>2+</sup> et 0.25 mM Cd <sup>2+</sup>	3 Mg(H <sub>2</sub> O) <sub>6</sub> <sup>2+</sup> 2 Mg(H <sub>2</sub> O) <sub>5</sub> <sup>2+</sup>	[100]
<b>Jonction III–IV–V</b>	FRET/CGE		3 mM Mg <sup>2+</sup>		[67]
	RMN	2MTJ	5 mM Mg <sup>2+</sup>		[101]
	RMN	2MTK	5 mM Mg <sup>2+</sup> et 10 μM Mn <sup>2+</sup>	5 Mg(H <sub>2</sub> O) <sub>6</sub> <sup>2+</sup> 1 Mg(H <sub>2</sub> O) <sub>5</sub> <sup>2+</sup>	[101]
<b>Jonction II–III–VI</b>	FRET		2 mM Mg <sup>2+</sup>		[68]
	RMN	2N3Q	5 mM Mg <sup>2+</sup>		[72]
	RMN	2N3R	5 mM Mg <sup>2+</sup> et 10 μM Mn <sup>2+</sup>	8 Mg(H <sub>2</sub> O) <sub>6</sub> <sup>2+</sup>	[72]
<b>Ribozyme VS (cis ou trans)</b>	CP/MI/SDM		0–20 mM Mg <sup>2+</sup>		[69]
	FRET/CGE		2–3 mM Mg <sup>2+</sup>		[67, 70]
	smFRET		35 mM Mg <sup>2+</sup>		[70]
	SAXS		10 mM Mg <sup>2+</sup>		[71]
	RMN		5 mM Mg <sup>2+</sup>		[72]
	X-ray (A <sub>756</sub> G)	4R4P	5 mM Mg <sup>2+</sup>	2 Mg <sup>2+</sup>	[65]
	X-ray (G <sub>638</sub> A)	4R4V	5 mM Mg <sup>2+</sup>	9 Mg <sup>2+</sup> , 5 K <sup>+</sup>	[65]
	X-ray	5V3I	5 mM Mg <sup>2+</sup>		[73]

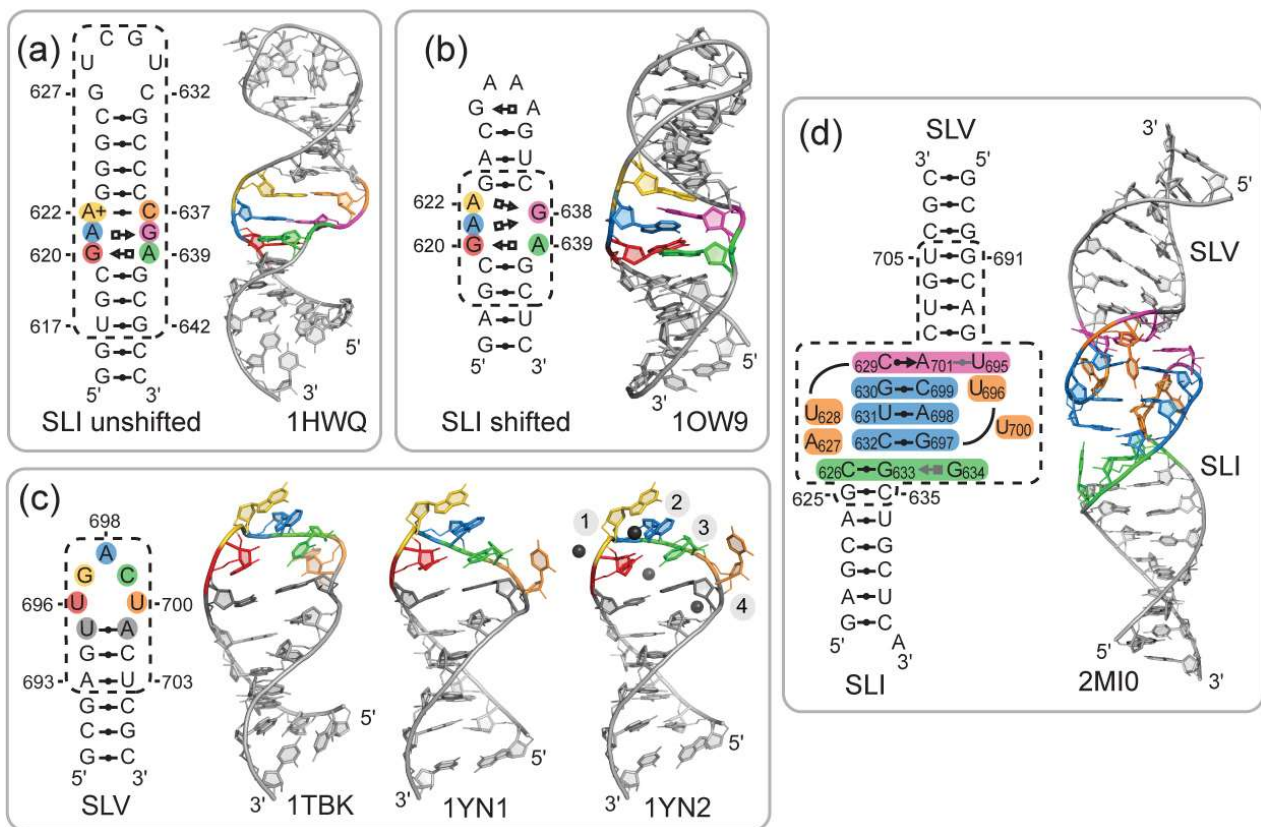
<sup>1</sup>Les approches expérimentales listées ici sont la spectroscopie RMN (RMN), les simulations en dynamique moléculaire (DM), le transfert d'énergie par résonance de type Förster (FRET), l'électrophorèse comparative sur gel (CGE), le *single-molecule* FRET (smFRET), le *chemical probing* (CP), la *modification interference* (MI), la mutagenèse dirigée (SDM), la diffusion des rayons X aux petits angles (SAXS), et la cristallographie aux rayons X (X-ray). Seules la RMN et la cristallographie ont donné des structures à hautes résolution.

<sup>2</sup>En terme de conditions expérimentales, seuls sont fournis la concentration et le type d'ion métallique utilisés pour la détermination de la structure (Aucun = aucun utilisé).

### 1.1.4.3. L'interaction boucle-boucle I/V

Les travaux présentés dans cette thèse se sont plus particulièrement concentrés sur l'interaction boucle-boucle entre la tige-boucle V du ribozyme et la tige-boucle I du substrat. Ce sous-domaine du ribozyme VS est le cœur du système de reconnaissance du substrat et l'une de ces caractéristiques les plus distinctives et les plus intéressantes du point de vue de l'ingénierie puisqu'il permet au ribozyme de reconnaître spécifiquement un ARN possédant une structure secondaire. La structure tridimensionnelle du complexe boucle-boucle I/V a été déterminée par RMN [96] et confirme la formation de motifs U-turn dans chacune des deux boucles ainsi que de trois paires de bases canoniques cis W/W entre elles (Figure 1.7d). Cette structure révèle également la présence de deux triplets de base : un triplet peu ordonné C626–G633–G634 et un triplet plus stable C629–A701–U695, lequel contient une paire non-canonique entre les deux boucles (cis WC/SE C629–A701). Au niveau de l'interaction boucle-boucle, les bases sont empilées de manière continue et cet empilement se prolonge dans les tiges adjacentes, ce qui contribue sans doute à la grande stabilité de l'interaction [102]. Malgré cette dernière, l'ensemble de structures obtenu en RMN présente une large gamme d'angle interhélicaux entre les hélices I et V, de 127° à 175°. Les données de RMN indiquent également une dynamique chez certains résidus des deux boucles (C626, A627, U628, C632, G633, G634 de la tige-boucle I et U695 et U696 de la tige-boucle V) à des échelles de temps allant de la  $\mu$ s à la ms. Considérées dans leur ensemble, ces informations semblent révéler une dynamique intrinsèque importante de l'interaction boucle-boucle I/V.





**Figure 1.8. Structures RMN et structures secondaires des sous-domaines du ribozyme VS impliqués dans la reconnaissance du substrat.** Structures secondaires et tridimensionnelles de la tige-boucle I, (a) non *shiftée* et (b) *shiftée*. (c) Structure secondaire et tridimensionnelle de la tige-boucle V. (d) Structure tertiaire et tridimensionnelle du complexe boucle-boucle I/V. Adapté de [75].

## 1.2. La dynamique de l'ARN

### 1.2.1. Généralités

Une molécule d'ARN en solution n'est pas présente dans une conformation unique mais sous forme d'un ensemble de structures à l'équilibre. Une représentation utile pour visualiser cet ensemble est celle du paysage énergétique [103, 104] : en représentant les variations d'énergie libre le long de l'espace conformationnel, il est possible d'obtenir une surface dont les vallées (les minima locaux) représentent des états métastables de la molécule étudiée. (Figure 1.9). Une molécule d'ARN donnée va passer spontanément d'une conformation à une autre au

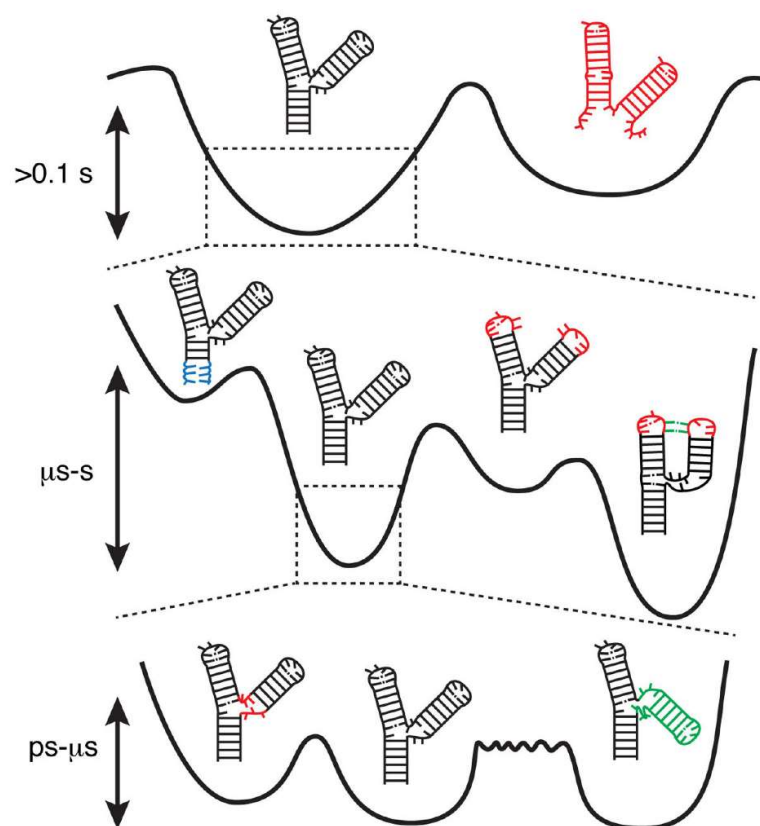
cours du temps à une fréquence et avec une rapidité dépendant directement de l'importance de la barrière énergétique qui les sépare. Connaître le paysage énergétique permet ainsi théoriquement de déterminer la population des différentes vallées et par conséquent la probabilité d'observer le système dans un état donné en solution.

### **1.2.2. Une dynamique échelonnée dans le temps**

Le paysage énergétique est très rugueux : chaque vallée contient à son tour des vallées, et ainsi de suite. Selon l'échelle à laquelle on observe le paysage énergétique d'une molécule, on observe ainsi des ensembles conformationnels différents, ce qui entraîne une hiérarchisation de la dynamique [105] (Figure 1.9). À l'échelle la plus large, chaque vallée représente une structure secondaire stable, séparée des autres par de hautes barrières énergétiques. Les transitions entre ces conformations sont lentes, de l'ordre de la seconde et plus. Au sein de chacune de ces vallées s'en trouvent d'autres qui représentent les variations dans l'organisation des paires de bases à un niveau local. Pour une structure secondaire donnée, il va être ainsi possible d'observer des bris de paires et des changements de partenaire d'appariement au sein des éléments de structure secondaire et tertiaire en des temps allant de la microseconde à la seconde. À une échelle encore plus réduite interviennent les changements de plissement des riboses, les fluctuations d'empilement de bases, les mouvements de résidus non-appariés et les mouvements inter-hélices dont les variations interviennent en quelques pico- à microsecondes. Finalement, les vibrations des atomes autour de la position d'équilibre de leur liaison covalente ont une fréquence de l'ordre de la picoseconde. Il est certain que ces différents niveaux de dynamique n'opèrent pas indépendamment. Le déplacement d'une hélice connectée à une jonction, par exemple, nécessite des déplacements locaux de résidus au niveau de l'interface entre l'hélice et la jonction. Il a aussi été montré que chez les riborégulateurs, des interactions tertiaires permettaient de stabiliser certain type de structure secondaire plus aptes à se lier à leur ligand [106, 107].

La hiérarchisation et l'interdépendance des niveaux de dynamique ont plusieurs conséquences remarquables. En effet, cela souligne l'importance de déterminer à quelle échelle de temps se déroulent les phénomènes d'intérêt et la résolution temporelle des protocoles utilisés pour leur observation afin de pouvoir obtenir des données pertinentes. En outre, s'il est possible

de modifier la dynamique à un niveau différent que celui auquel se passent ces phénomènes, par exemple en empêchant la formation de quelques interactions tertiaires pour changer radicalement les populations relatives des différentes structures secondaires d'un ARN, la connaissance précise du paysage énergétique d'un système devient alors cruciale pour son ingénierie rationnelle. Finalement, on peut constater que des mouvements importants de larges éléments de structure secondaire se font à des échelles de temps relativement rapides. Il est donc important de prendre en compte ces déplacements, qui peuvent radicalement changer l'organisation d'une molécule d'ARN et altérer sa fonction.



**Figure 1.9. Hiérarchie des types de mouvements impliqués dans la dynamique de l'ARN.** Les lignes pointillées représentent un agrandissement d'un minimum énergétique de l'échelon temporel supérieur. Le premier échelon représente les changements de structure secondaires. Le second échelon comprend les bris de paires de bases (en bleu), les changements de partenaires d'appariement (en rouge) et les appariements tertiaires (en vert). Le troisième échelon comprend les mouvements de résidus non-appariés (en rouge) et les mouvements inter-hélices (en vert). Adapté de [105].

### 1.2.3. L'étude de la dynamique de l'ARN

La dynamique de l'ARN peut être étudiée par des méthodes biophysiques, biochimiques ou bio-informatiques. Dans les deux premières catégories, on peut d'un côté trouver les techniques qui renseignent sur la dynamique moyenne d'un ensemble d'états métastables, telles la spectroscopie RMN [108] et les expériences de *small angles X-ray scattering* (SAXS) [109] et de *selective 2' hydroxyl acylation by primer extension* (SHAPE) [110] et de l'autre les techniques mono-moléculaires comme le *single molecule* FRET (smFRET) [111, 112] et la cryo-microscopie électronique (cryoEM) [113]. Ces méthodes apportent de précieuses informations mais il n'en existe pas une actuellement qui réunisse les conditions pour une observation complète de la dynamique de l'ARN, soit une méthode capable : 1) d'offrir une résolution atomique et apte à percevoir tous les types d'atomes, 2) d'échantillonner des mouvements à des échelles de temps allant de la pico à la milliseconde, 3) de discriminer les mouvements appartenant à chaque échelon temporel, 4) d'identifier et quantifier les différentes populations de l'ensemble conformationnel, 5) d'opérer dans des conditions expérimentales optimale pour l'observation de la dynamique du système étudié.

Les méthodes bio-informatiques basées sur la dynamique moléculaire, décrites dans la section 1.3, se sont fait une place parmi ces techniques car elles sont théoriquement capables de réunir ces conditions avec l'avantage supplémentaire qu'elles ne nécessitent pas la préparation d'échantillons, l'optimisation des paramètres de réaction et l'utilisation de matériel spécialisé. Bien entendu, des limitations existent. Les simulations de très longues durées (>1ms) demandent des temps de calculs et des ressources informatiques non triviales et sont soumises à des risques d'erreurs de calcul cumulatives. La simulation d'espèces chimiques peu communes comme des bases modifiées ou des ions complexes souffre d'imprécision. Malgré tout, la dynamique moléculaire permet d'obtenir des données inédites à résolution atomique et est maintenant une technique de choix pour l'étude de la dynamique de l'ARN.

## 1.2.4. Dynamique et fonction des ARN catalytiques

La dépendance de la fonction des ARN catalytiques à leur dynamique est un fait bien documenté. En particulier, le ribosome est une molécule dont l'intégralité de la fonction dépend non-seulement de mouvements inter- et intra sous-unités de grande amplitude centrés sur des mouvements locaux [114] mais également de la flexibilité des ARN de transfert [115]. Les petits ribozymes nucléolytiques sont des modèles très utiles pour comprendre la dynamique de l'ARN puisque leurs mouvements dépendent essentiellement de leurs capacités dynamiques intrinsèques en solution. Les études sur la dynamique de plusieurs petits ribozymes naturel et artificiels ont démontré l'importance de la dynamique à différentes échelles. À l'échelle atomique d'abord, comme dans le cas du *leadzyme* dont le passage du résidu C<sub>6</sub> d'une configuration C3' endo à C2' endo est cruciale pour l'activité de clivage [116]. À l'échelle locale ensuite, comme pour l'exploration d'états conformationnels rares par la boucle A du ribozyme hairpin [117]. À l'échelle globale enfin, comme dans le cas de l'exploration de conformation inactives causée par une formation précoce d'un pseudo-nœud par le ribozyme twister [118]. Dans tous ces cas de figure, la réaction de clivage par le ribozyme est très rapide mais le taux de réaction final est fortement influencé par la dynamique intrinsèque du système et sa capacité à explorer des régions de l'espace conformationnel dans lequel les différentes parties du site actif sont dans une conformation favorable à sa formation.

## 1.3. La bio-informatique structurale

### 1.3.1. Principe général

La biologie structurale étudie la structure tridimensionnelle des molécules biologiques, les modalités d'acquisition de cette structure et son impact sur la fonction de la molécule. La bio-informatique structurale a les mêmes objectifs mais repose sur la simulation informatique des molécules étudiées plutôt que sur l'observation expérimentale afin de les atteindre.

L'idée de base est la suivante : tenter de modéliser informatiquement les propriétés physiques et chimiques d'une molécule biologique d'intérêt puis dériver de ce modèle des propriétés plus complexes comme ses structures secondaire, tertiaire et tridimensionnelle, ses propriétés dynamiques et l'influence de ses interactions avec d'autres entités (autres molécules,

solvant, ions), elles aussi simulées. Idéalement, les molécules étudiées seraient simulées *ab initio*, c'est-à-dire que la connaissance de leur composition atomique serait suffisante pour déduire toutes les propriétés qui en découlent. En pratique, la bio-informatique structurale est loin d'avoir atteint cet idéal [119] et il reste difficile de passer d'un niveau de structure à l'autre. La discipline reste ainsi très dépendante des techniques de biologie structurales biochimiques et biophysiques classiques pour lui fournir des données complémentaires.

Les techniques et concepts employés en bio-informatique structurale sont nombreux. La suite du présent chapitre n'aura donc pas pour but d'être une revue exhaustive de la discipline mais développera les protocoles qui ont été employés dans les travaux détaillés dans les chapitres suivants.

### **1.3.2. L'annotation de structure et la recherche de motifs structuraux**

L'annotation de structure d'ARN consiste à établir la liste des ribonucléotides et des interactions intra- et inter-nucléotides d'une molécule sur la base d'un fichier de coordonnées atomiques tels que ceux stockés dans la base RCSB [120]. Ces interactions sont principalement des liaisons covalentes, des ponts hydrogènes et des empilements de bases, déterminées à partir de critères géométriques (orientation de bases et distance interatomiques) [121, 122]. À partir de ces annotations, il est alors possible d'établir et de cataloguer automatiquement la présence d'éléments de structure secondaire et tertiaire ainsi que de motifs structuraux [123].

Ces catalogues de motifs permettent d'effectuer des recherches de motifs. Ce volet de la bio-informatique structurale est important pour plusieurs autres aspects de la discipline, en particulier pour l'identification de nouveaux motifs et le raffinement de la définition de motifs existant ou encore la recherche de modules structuraux à des fins de prédiction et d'ingénierie.

### **1.3.3. La modélisation de structure 3D d'ARN**

Lorsque la structure 3D à haute résolution d'une molécule d'ARN est inconnue, incertaine ou incomplète, il faut pouvoir la modéliser, c'est-à-dire dans ce contexte, pouvoir la prédire grâce aux informations de niveaux de structure inférieurs connues et aux données structurales tridimensionnelles obtenues expérimentalement : c'est pour cette raison que les termes « modélisation » et « prédiction de structure » seront utilisés interchangeablement.

La prédiction peut théoriquement employer deux classes d'approches. Elle peut se baser sur la connaissance de structure tridimensionnelles déjà déterminées pour en tirer des fréquences d'interactions interatomiques ou pour prédire une structure 3D inconnue par homologie. Elle peut également se fonder sur la connaissance des principes physiques fondamentaux qui régissent le repliement de l'ARN. Ces deux extrêmes sont peu utilisés, d'une part car l'approche physique demanderait des ressources informatiques et un temps de calcul irréalistes, d'autre part parce que l'approche statistique ou par homologie exigerait des bases de données contenant toutes les conformations d'ARN possibles. Les logiciels de prédiction se servent donc de techniques qui empruntent aux deux classes d'approches en tentant de les simplifier et/ou de les compléter par d'autres méthodes : par exemple, l'homologie peut être utilisée sur des fragments de l'ARN à modéliser qui sont ensuite assemblés [124, 125] et les simulations de repliement basées sur les principes physiques peuvent utiliser une représentation réduite de l'ARN qui remplace des groupes d'atomes par un pseudo-atome unique (*coarse-grained*) [126, 127]. De plus, il est souvent possible d'ajouter des contraintes additionnelles afin de prendre en compte les données expérimentales biophysiques et biochimiques. Malgré ces simplifications, il est encore difficile d'obtenir des prédictions satisfaisantes sur des ARN de grande taille ou possédant une structure tertiaire complexe [119].

### 1.3.4. La dynamique moléculaire

#### 1.3.4.1. Principe général

La dynamique moléculaire se propose de résoudre l'équation du mouvement de Newton de l'ensemble des atomes d'un système simulé au cours du temps. Pour ce faire, le temps de simulation est découpé en intervalles de temps discrets au cours desquelles les coordonnées atomiques de chaque particule du système vont être mises à jour en fonction des forces qui s'appliquent sur elles et de leur vitesse. Ces forces sont dérivées grâce à un champ de force, qui est la combinaison d'une fonction mathématique permettant de calculer l'énergie potentielle en tout point du système et d'un jeu de paramètres utilisées pour résoudre cette fonction.

La fonction de potentiel est de la forme suivante :

$$E_{totale} = E_{covalente} + E_{non-covalente} + E_{contraintes}$$

Où,

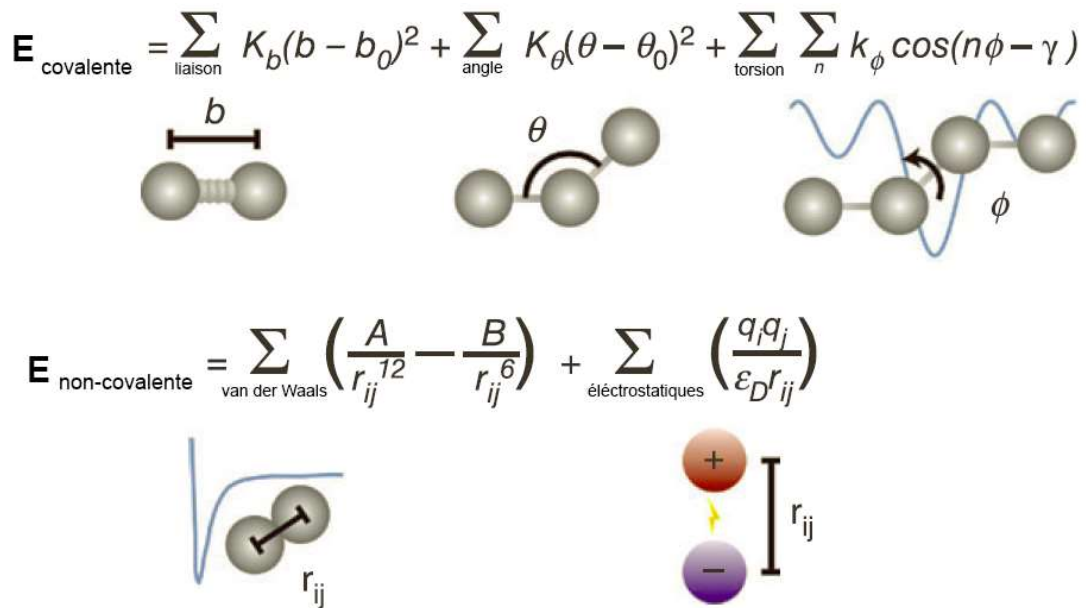
$$E_{\text{covalente}} = E_{\text{liaison}} + E_{\text{angle}} + E_{\text{angle dièdre}} + E_{\text{angle impropre}}$$

Et,

$$E_{\text{non-covalent}} = E_{\text{vanderWalls}} + E_{\text{électrostatiques}}$$

Concrètement, l'énergie potentielle de chaque atome du système dépend d'une part de sa position dans l'espace et d'autre part des deux grands types d'interactions auxquelles il participe : les interactions covalentes d'abord, avec une contribution de l'élongation de la liaison elle-même ( $E_{\text{liaison}}$ ) ainsi que des angles formés par les triplets ( $E_{\text{angle}}$ ) ou quadruplets ( $E_{\text{angle dièdre}}$ ) d'atomes liés auxquels appartient l'atome considéré. Un potentiel dièdre particulier ( $E_{\text{angle impropre}}$ ) peut parfois être ajouté afin, par exemple, de contraindre la planéité des noyaux aromatiques. Ensuite, sont considérées les interactions non-covalentes, en général les interactions de van der Waals et électrostatiques. Leurs contributions sont plus complexes à traiter puisqu'en théorie tous les atomes d'un système s'influencent par les biais de ces interactions, ce qui demande de considérer toutes les paires, les triplets, les quadruplets, etc... d'atomes possibles. En pratique, on procède à deux simplifications : les interactions non-covalentes ont non seulement un rayon d'influence limité mais on ne considère en outre que les paires d'atomes pour le calcul de l'énergie potentielle. La plupart des champs de force modernes autorisent en outre l'addition d'un terme ( $E_{\text{contraintes}}$ ) à cette fonction qui permet d'ajouter des interactions supplémentaires qui ne correspondent à aucune des deux catégories précédemment décrites, comme par exemple dans le but d'immobiliser des atomes, leur imposer une géométrie particulière ou encore d'incorporer des contraintes expérimentales.





**Figure 1.10. Illustration de la fonction de potentiel d'un champ de force classique.** Les fonctions individuelles sont représentées mathématiquement sous leur forme la plus usuelle.

Le jeu de paramètres associé à cette fonction de potentiel contient les variables qui permettent de caractériser chaque atome du système comme sa masse, sa charge, son rayon de Van Der Waals et les valeurs à l'équilibre de ses interactions covalentes. Ces paramètres sont en grande majorité dérivés empiriquement d'expériences menées sur de petites biomolécules et de simulations menées à un niveau quantique [128], exactes mais très coûteuses en temps et en ressources informatiques.

La simulation se déroule dans une enceinte virtuelle ou « boîte de simulation », une portion de l'espace au-delà de laquelle les forces ne sont plus calculées. Afin d'éviter les effets irréalistes sur les atomes simulés aux limites de l'enceinte, il est fréquent d'utiliser des conditions périodiques aux limites : lorsqu'un atome simulé dans la boîte quitte l'enceinte par une face, il réapparaît de l'autre côté par la face opposée. Les atomes de la structure simulée ne sont pas seuls dans la boîte de simulation. Habituellement, un solvant est aussi simulé, soit de manière explicite (les atomes sont présents et représentés par un jeu de paramètres au même titre que les atomes de la structure) ou implicite (les interactions sont représentées par une

équation globale). Dans le cas des simulations d'ARN, il est aussi fréquent d'ajouter des ions métalliques.

Le protocole de dynamique moléculaire permet d'obtenir la trajectoire des atomes simulés en fonction du temps; le fichier généré est la liste de coordonnées cartésiennes des atomes du système à chaque intervalle de temps, en général accompagnée des forces, vitesse et énergies associées. L'interprétation de ces données repose sur l'hypothèse d'ergodicité qui implique que pour des simulations suffisamment longues et générées grâce à des forces réalistes, la moyenne temporelle d'une valeur observée sur la trajectoire est égale à la moyenne d'ensemble de cette valeur à un temps donné. En d'autres termes, pour un système ergodique, il est possible d'échantillonner tous ses états possibles en ne simulant qu'un exemplaire de ce système durant une longue période.

#### **1.3.4.2. L'amélioration de l'échantillonnage**

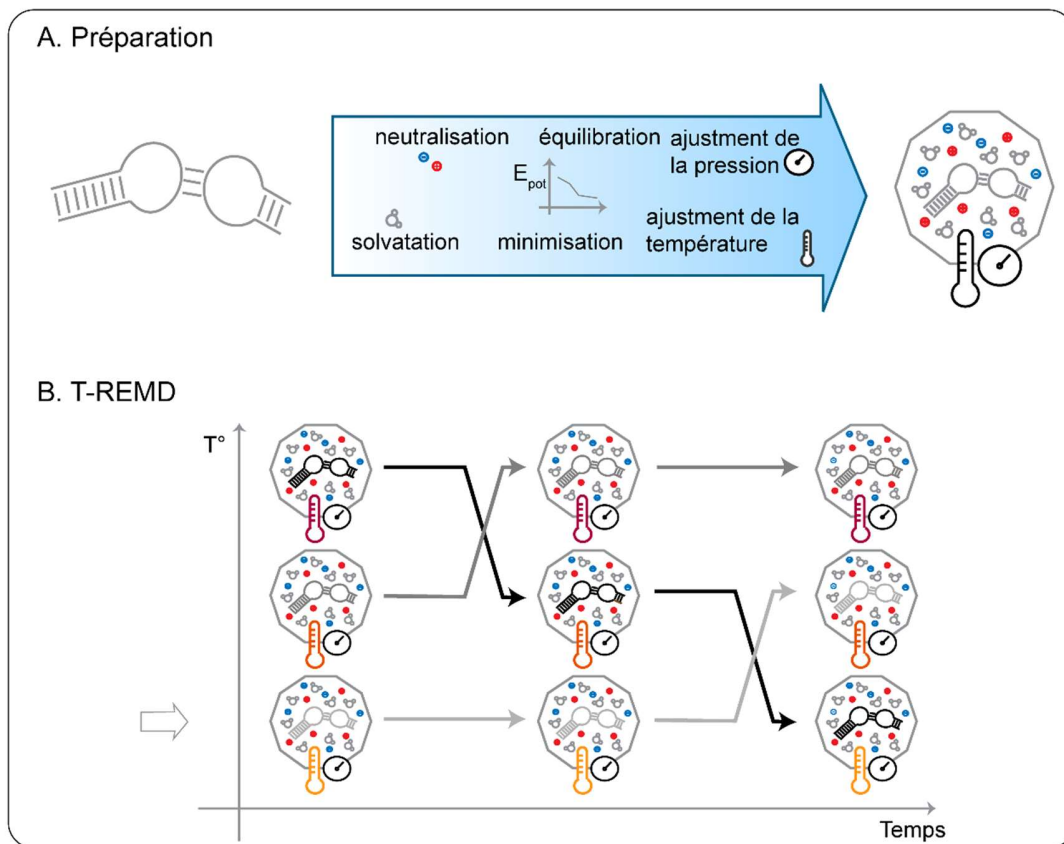
En pratique, la dynamique moléculaire se heurte à deux problèmes majeurs. D'une part, les minima locaux d'énergie peuvent piéger la molécule simulée dans une conformation particulière et l'empêcher d'explorer plus avant son espace conformationnel. D'autre part, si ce dernier est vaste, l'explorer ou à tout le moins l'échantillonner d'une manière exhaustive peut s'avérer impossible en un temps raisonnable. Afin de remédier à ces inconvénients, il est possible d'avoir recours à des techniques dites « d'amélioration de l'échantillonnage » qui ont pour but de permettre au système simulé d'accélérer le franchissement des barrières rencontrées dans son paysage énergétique.

Ces méthodes peuvent être regroupées en deux grandes catégories. La première catégorie est basée sur l'utilisation de forces externes artificielles (ou biais) appliquées sur une ou plusieurs variables collectives [129] afin d'aplanir le paysage énergétique et guider la simulation vers des portions de l'espace conformationnel rarement visité. Les variables collectives sont des fonctions des coordonnées cartésiennes des atomes du système, de moindre dimension et suffisantes pour décrire sa dynamique, et peuvent être par exemple des ensembles d'angles, de distances ou d'interactions interatomiques. Idéalement, le biais utilisé devrait être choisi pour annuler l'effet de l'intégralité des barrières énergétiques. Cela présuppose une connaissance préalable du paysage énergétique complet, ce qui n'est en pratique pas possible. Le biais doit

donc être soit un potentiel fixe, soit un potentiel adaptatif au cours du temps de simulation, ces deux méthodes étant représentées, respectivement, par l'*umbrella sampling* [130] et la métadynamique [131]. Les techniques non-biaisées profitent quant à elles de l'effet des hautes températures pour accélérer et faciliter le franchissement des barrières énergétiques. Elles ont l'avantage de ne nécessiter aucune information préalable sur le système et sur le type de dynamique que l'on désire observer. La dynamique moléculaire avec échange de répliques (*replica exchange molecular dynamics*, REMD)[132], utilisée dans les travaux exposés dans la présente thèse, et la trempe simulée (*simulated tempering*)[133] sont des exemples de méthodes non-biaisées.

#### **1.3.4.3. Le *temperature replica exchange molecular dynamics* (T-REMD)**

Le principe général de la T-REMD est le suivant : lancer plusieurs simulations de DM classiques en parallèle, chacune à une température plus élevée que la précédente. Les coordonnées de la structure simulée dans chaque simulation sont périodiquement échangées (Figure 1.11) avec une simulation voisine, selon un critère d'acceptation basé sur la différence d'énergie potentielle du système entre les simulations à échanger. L'échelle de températures est circonscrite entre la température à laquelle se déroulerait l'expérience en conditions réelles et une température élevée choisie pour être suffisante pour faire franchir au système les barrières énergétiques pertinentes sans le dénaturer irrémédiablement. Les températures successives sont fixées de manière à assurer un taux d'échange raisonnable. La trajectoire à plus basse température est finalement analysée : si la simulation a été assez longue, cette trajectoire discontinue présente alors en théorie un échantillonnage exhaustif des possibilités conformationnelles de la molécule étudiée.



**Figure 1.11. Schéma illustratif de la technique de T-REMD.** (A) Étape de préparation commune à tous les protocoles de dynamique moléculaire. (B) Illustration de l'échange de répliques au cours de la procédure de T-REMD proprement dite. Plusieurs simulations (répliques) sont lancées en parallèle à des températures différentes. Les molécules simulées sont périodiquement échangées avec les répliques voisines. La réplique à la température d'intérêt, indiquée par une flèche, sera analysée.

#### 1.3.4.4. La dynamique moléculaire pour l'étude de l'ARN

La dynamique moléculaire peut être utilisée pour étudier deux phénomènes distincts associés au mouvement la molécule d'ARN : son repliement et le comportement dynamique de ses éléments de structure secondaire et tertiaire une fois repliés.

Simuler le repliement d'ARN de plus de quelques dizaines de nucléotides de manière atomistique est aujourd'hui encore hasardeux à cause du grand nombre de degrés de liberté du squelette sucre-phosphate et du vaste répertoire d'interactions interatomiques possibles entre

phosphates, riboses et nucléobases qui entraînent un temps de calcul prohibitif. Avec la plus grande disponibilité du calcul parallèle et le développement des techniques d'amélioration de l'échantillonnage, il est possible pour l'instant d'étudier le repliement de petits motifs de type tetra-boucles [134].

L'étude des éléments de structures déjà repliés dont la structure 3D a été obtenue par des méthodes biochimiques ou biophysiques, peut se faire à plusieurs niveaux et a donné de très bons résultats. La dynamique moléculaire peut être appliquée au niveau de motifs spécifiques comme le T-loop [135] ou les boucles GNRA [134] et d'éléments de structure tertiaire comme les interactions boucle-boucle [136]. Ces études permettent de préciser leurs modalités de stabilisation, leurs capacités dynamiques intrinsèques et leurs paramètres thermodynamiques, ce qui est informatif à la fois du point de vue de la connaissance fondamentale de ces éléments mais aussi de la compréhension de leur rôle dans la formation du site actifs des ARN catalytiques [137]. À un niveau plus global, l'observation du mouvement des hélices connectées à ces éléments de structure peut permettre de mieux appréhender la dynamique globale d'un système complexe comme le ribosome [138]. Les études de dynamique moléculaire peuvent également aider à élucider les mécanismes d'interaction de l'ARN avec les espèces chimiques avec lesquelles il interagit fréquemment dans un contexte biologique, comme des protéines [139, 140] ou des ions [141].

En plus de ces applications, la dynamique moléculaire peut également être utilisée pour compléter des travaux menés par des méthodes biophysiques, que ce soit pour améliorer le raffinement d'ensembles de modèles de RMN obtenu avec des données structurales peu nombreuses [142], générer des ensembles à ajuster aux données de SAXS [143] ou de cryo-microscopie électronique (cryo-ME) [144].

### **1.3.5. L'analyse de la dynamique de l'ARN**

#### **1.3.5.1. Généralités**

La dynamique intrinsèque d'une molécule d'ARN en solution peut être vue comme un composite de mouvements plus simples représentant chacun le déplacement collectif d'un ensemble d'atomes le long d'une direction particulière. Afin d'identifier les composantes

importantes du mouvement global, il est important de distinguer les vibrations atomiques locales et de haute fréquence des mouvements de plus basse fréquence et de plus grande amplitude qui influencent réellement l'espace conformationnel exploré par l'ARN étudié et par extension, sa fonction. De plus, la comparaison de la dynamique de deux systèmes est plus aisée lorsque le mouvement de chaque système est découpé en ses composantes individuelles. L'identification de ces composantes peut être faite par l'entremise de deux techniques: l'analyse des modes normaux et la dynamique essentielle.

#### **1.3.5.2. L'analyse des modes normaux**

L'analyse des modes normaux (AMN) [145, 146] est une technique qui permet d'étudier le mouvement d'une biomolécule en prenant comme point de départ une structure énergétiquement minimisée et en simulant ses oscillations harmoniques autour de cette position d'équilibre. L'AMN ne nécessite pas de simuler la molécule étudiée en dynamique moléculaire ; les composantes du mouvement de la molécule sont déterminées à partir des coordonnées atomiques de la structure minimisée et d'une fonction de potentiel harmonique dont on calcule la dérivée seconde pour chaque combinaison de coordonnées du système. On obtient alors la matrice hessienne de la fonction de potentiel dont la diagonalisation donne les vecteurs propres, représentant chacun une direction selon laquelle les atomes du système se déplacent de concert autour de leur position dans la structure minimisée. Ces vecteurs sont les modes normaux et leurs valeur propres associées permettent de discriminer entre les simples vibrations et les mouvements de grande amplitude.

#### **1.3.5.3. La dynamique essentielle**

Une trajectoire de dynamique moléculaire est un jeu de donnée multidimensionnel. Chaque image est une observation et chaque coordonnée atomique (x, y, ou z) de chaque atome est une variable. Le jeu de donnée à analyser est donc une matrice de  $m$  par  $3n$ , où  $m$  est le nombre d'images et  $n$  le nombre d'atomes du système. Il est donc indispensable de recourir à une analyse multivariée pour extraire de l'information pertinente d'un jeu de donnée d'une telle complexité. Parmi les analyses multivariées disponibles, l'analyse en composante principale (ACP) est idéale pour l'étude de trajectoires. L'ACP est basée sur la matrice de covariance des variables, soit la matrice qui recense la covariance de chaque variable par rapport à toutes les

autres. Il est possible d'extraire les vecteurs propres et valeurs propres de cette matrice : chaque vecteur propre est une direction le long de laquelle les variables du jeu de données étudié covarient et est associé à une valeur propre qui indique sa contribution à la variabilité totale des données. En ne sélectionnant qu'un sous-ensemble des vecteurs propres qui explique suffisamment la variabilité des données, il est possible de les utiliser comme nouvelles variables afin de réduire la dimensionnalité du problème.

La dynamique essentielle [147] est une application de l'analyse en composantes principales (ACP) au cas particulier des trajectoires obtenues en dynamique moléculaire. Dans ce contexte, les variables étant des coordonnées dans l'espace, les vecteurs propres représentent littéralement la direction selon laquelle certains atomes se meuvent d'une manière concertée, tandis que les valeurs propres associées déterminent la contribution de ces mouvements à la dynamique globale du système étudié. Il est ainsi possible de découper le mouvement d'une molécule simulée en ses mouvements constitutifs, de ne garder que ceux qui ne représentent pas des vibrations atomiques aléatoires et de faible amplitude et d'identifier les atomes impliqués dans chaque composante.

## **1.4. L'ingénierie d'ARN**

### **1.4.1. Généralités**

L'ingénierie biomoléculaire a pour objectif la création d'outils, matériaux ou méthodes inédits basés sur des molécules biologiques. L'ARN est une molécule modulaire, polyvalente, omniprésente aussi bien dans la cellule eucaryote que procaryote et qui possède pourtant une structure chimique relativement simple par rapport aux protéines. Ces qualités en font une cible désirable pour cette discipline puisque les ARN modifiés peuvent potentiellement agir sur un grand nombre de processus cellulaires, dans divers organismes et que leur design repose sur un nombre relativement limité d'éléments constitutifs.

### **1.4.2. Méthodes**

Le processus de création d'une biomolécule inédite peut suivre deux voies : elle peut être construite en modifiant une molécule préexistante qui détient des propriétés proches de

celles recherchées; ou elle peut être construite *de novo*, en élaborant une séquence entièrement nouvelle. Dans les deux cas, on distingue deux types d'approche : d'abord les approches non rationnelles ou semi-rationnelles, dans lesquelles les modifications ou la construction se font aléatoirement, à la suite de quoi les molécules résultantes sont testées pour identifier des fonctions d'intérêt. Ensuite, les approches dites rationnelles, dans lesquelles la construction *de novo* ou la modification est réalisée d'après les prédictions émises sur la base d'un modèle permettant de relier la structure et/ou la dynamique à la fonction de la molécule.

Les techniques de sélection *in vitro* aussi appelées SELEX (*systematic evolution of ligands by exponential enrichment* ou évolution systématique de ligands par enrichissement exponentiel) [148, 149] font partie des approches non-rationnelles. Elles consistent à générer une librairie d'oligonucléotides de séquence aléatoire puis à les placer en présence d'une cible moléculaire d'intérêt. Les oligonucléotides qui n'ont pas réagi avec la cible (absence de liaison, de clivage, etc... dépendamment du critère de sélection choisi) sont éliminés et les autres sont amplifiés. Après plusieurs rondes de sélection dans des conditions de stringence croissante, seuls restent les oligonucléotides avec la plus forte réaction à la cible. Ces protocoles peuvent être rendus semi-rationnels si seule une partie de la séquence des oligonucléotides, choisie pour son influence sur la fonction de celui-ci, est générée de manière aléatoire.

Les approches rationnelles ne reposent pas sur un type de technique en particulier mais sur la capacité à concevoir un modèle prédictif de l'ARN étudié afin de déterminer quelles modifications pourraient conduire à l'effet souhaité. Ce modèle peut être extrêmement détaillé et permettre de prédire les variations du paysage énergétique du système selon les modifications apportées à sa structure. Un modèle partiel peut cependant être suffisant : grâce à la nature modulaire de l'ARN, le simple assemblage de plusieurs modules structuraux possédant les caractéristiques dynamiques et/ou fonctionnelles souhaitées isolément peut produire un hybride présentant toutes ces caractéristiques simultanément [150]. De même, lorsque la fonction de l'ARN visé dépend de phénomènes simples tel que la formation d'un duplex, le modèle prédictif peut simplement consister à prédire l'appariement de deux séquences d'ARN [151].



### 1.4.3. Objectifs

Il est possible de classer les objectifs des travaux d'ingénierie d'ARN catalytique selon trois grands axes en fonction de leur finalité pratique : la modification des cibles affectées, l'adoption de nouvelles fonctions ou l'adoption de nouvelles structures. Ces axes ne sont pas mutuellement exclusifs et les avancées les plus notables en termes d'ingénierie ont été obtenues par des ARN modifiés le long de plusieurs de ces axes, par exemple en créant une nouvelle fonction puis en améliorant son efficacité.

#### 1.4.3.1. Modification de la reconnaissance de la cible

Cet aspect de l'ingénierie est préoccupé principalement par la création de molécules d'ARN aptes à se lier à une cible spécifique avec une grande affinité et la modification d'ARN fonctionnels existants afin de leur faire reconnaître une cible différente. Les approches d'ingénierie non rationnelles de type SELEX ont connu de grands succès dans ce domaine, en particulier pour la production d'aptamères [152] et de riborégulateurs basés sur des aptamères produits *in vitro* [153, 154]. On peut également citer dans cette catégorie la modification de petits ribozymes nucléolytiques comme le ribozyme hairpin afin qu'ils clivent des substrats différents de leur substrat naturel [151].

Les travaux présentés ici sont apparentés à cette dernière stratégie puisqu'ils visent en partie à modifier le ribozyme VS de *Neurospora* pour qu'il puisse accommoder de nouveaux ARN repliés en tige-boucle comme substrats pour son activité de clivage.

#### 1.4.3.2. Modification de la fonction

Dans ce cas de figure, l'objectif est de modifier la fonction d'un ARN existant ou de concevoir une molécule nouvelle qui possède une fonction inédite. La modification de fonction peut impliquer de changer la nature de la réaction catalysée par un ribozyme, améliorer son taux d'activité, rendre une fonction dépendante à ou indépendante de la liaison d'un ligand ou la présence de cofacteurs.

La conception d'aptazymes [155], ribozymes fusionnés à des aptamères repose sur la nature modulaire de l'ARN afin d'obtenir de ribozymes dont le taux d'activité est contrôlé par la présence d'un ligand. À l'inverse, des expériences de sélection *in vitro* ont permis d'obtenir

des ribozymes *glmS* dont le clivage est indépendant de la glucosamine-6-phosphate [156] et porteurs d'une simple triple mutation. Comme mentionné précédemment, la création de ribozymes possédant des fonctions que l'on ne trouve pas chez les ribozymes naturels est un champ de recherche très actif dans le domaine de la théorie du « monde à ARN ». L'étape suivant la découverte d'un type de catalyse inédit est souvent de tenter d'identifier des formes plus rapides des ribozymes ainsi créés [49, 157].

#### **1.4.3.3. Modification de la structure**

Cet axe de recherche vise à modifier de manière importante la structure tridimensionnelle d'un ARN fonctionnel ou à créer une molécule adoptant un ensemble de conformations d'intérêt.

Dans le premier point, on peut trouver des recherches intéressées à réduire la taille d'ARN fonctionnels sans perdre leur fonction afin d'en obtenir la forme minimale active. On peut également considérer dans cette catégorie les travaux visant à rendre des ribozymes autocatalytiques capable d'être actifs en *trans*. Les formes en *trans* et minimales du hairpin [158, 159] et du ribozymes VS [64, 160] sont de bons exemples de ce type de travaux.

Une application plus intéressante de la modification de structure est la construction de nanostructures tridimensionnelles d'ARN. Cette branche des nanotechnologies s'intéresse à la conception de séquences d'ARN capables de se replier co-transcriptionnellement *in vivo* selon une conformation prévisible et de s'auto-assembler en structures plus larges, de taille comparable au ribosome [161]. Les formes adoptées par ces « origami d'ARN » dépendent des modalités de repliement, de la dynamique et des interactions entre les différentes pièces [162]. Les travaux portant sur la caractérisation structurale et dynamique des motifs et éléments de structure secondaire et tertiaire d'ARN sont donc cruciaux pour faire évoluer la discipline.

### **1.5. Objectifs de la thèse**

Le ribozyme VS de *Neurospora* possède un mode de reconnaissance de son substrat unique parmi les petits ribozymes nucléolytiques et cette reconnaissance autant que la formation de son site actif mettent en jeu la dynamique de sous-domaines riches en interactions secondaires et tertiaires. Ces particularités le rendent à la fois intéressant comme modèle pour l'explorations

des relations entre structure, dynamique et fonctions des ARN catalytiques et comme cible d'ingénierie. L'hypothèse de travail initiale était la suivante : la fonction du ribozyme VS repose sur un ensemble de paramètres structuraux et dynamiques finis qui, une fois connus, peuvent être modifiés rationnellement afin d'altérer la reconnaissance et le clivage de son substrat par le ribozyme. Nous avons donc cherché à en savoir plus à la fois sur sa structure tridimensionnelle et ses capacités dynamiques ainsi que sur son potentiel d'ingénierie.

Ces objectifs ont été remplis en suivant un protocole en trois étapes :

Premièrement, il a été nécessaire de démontrer que notre modèle, le ribozyme VS de *Neurospora*, pouvait être modifié sans perdre totalement sa structure et sa fonction. Il était important également d'en connaître plus sur sa conformation active puisqu'aucune structure complète à haute résolution du ribozyme VS n'était publiée à ce moment-là. Notre hypothèse pour ce premier article était la suivante : dans le modèle initial de l'interaction boucle-boucle, l'hélice I du substrat est positionnée de manière colinéaire par rapport à l'hélice V du ribozyme. Par conséquent, il devait être possible de retrancher des paires de bases à l'une de ces hélices pour allonger l'autre d'autant sans supprimer la fonction. Nous avons donc mis au point un procédé pour caractériser systématiquement la cinétique de combinaisons de ribozymes et de substrats présentant des hélices V et I de différentes longueurs. Nous avons ensuite produit différents modèles de l'interaction boucle-boucle compatibles avec les données biochimiques connues à ce moment-là puis, par comparaison statistique entre les critères structuraux de ces modèles et les données de cinétiques, nous avons sélectionné les modèles compatibles avec les activités observées. Ces travaux ont permis de déterminer que le ribozyme VS possède la capacité de reconnaître et cliver des substrats de longueurs différentes. Ils ont également permis de caractériser le phénomène de compensation de longueur d'hélice, par lequel des ribozymes dont les hélices sont allongées ou raccourcies peuvent cliver des substrats dont la longueur les rendaient non clivables par le ribozyme naturel. Les modèles de l'interaction *kissing-loop* sélectionnés par l'analyse statistique semblaient enfin indiquer qu'un angle proche de  $165^\circ$  entre les axes des hélices I et V expliquait au mieux les données cinétiques.

Ensuite, nous avons cherché à pousser plus loin la modification du ribozyme en modifiant un élément de structure tertiaire crucial à la reconnaissance de son substrat, l'interaction boucle-boucle I/V, dans l'idée de lui permettre d'accommoder et de cliver d'autres

substrats repliés tige-boucles. Dans cette optique, nous avons effectué une recherche d'interactions boucle-boucle de taille similaires dans la base de données RCSB puis nous avons filtré les candidats sur des critères structuraux afin de garder ceux dont les angles interhélicaux entre les hélices leur étant rattachées étaient compatibles avec les angles observés dans l'ensemble de conformations de l'interaction naturelle obtenu par RMN. Les cinétiques de ribozymes mutants pourvus des interactions alternatives ont été testées *in vitro* puis ont été structurellement caractérisés dans le contexte de la structure cristalline publiée peu de temps auparavant. Les deux candidats retenus pour la substitution, soit l'interaction TAR/TAR\* du VIH et l'interaction L22/L88 du ribosome de *D. radiodurens* ont démontré une activité de clivage substantielle, quoique 50-160 fois moindre que celle du ribozyme porteur de l'interaction naturelle.

La caractérisation structurale entreprise dans l'article précédent n'ayant pas permis d'expliquer l'activité plus basse des variantes, nous avons cherché à relier à la dynamique intrinsèque des différents complexes boucle-boucle à l'activité du ribozyme naturel et de ses variantes, selon l'hypothèse que, la dynamique du reste du ribozyme étant égale par ailleurs, ce paramètre était principalement responsable de la différence d'activité observée précédemment. Les espaces conformationnels des variantes de l'interaction ont été caractérisés par des expériences de *temperature-replica exchange molecular dynamics* associées à une analyse en dynamique essentielle. Ces protocoles nous ont permis de détailler la dynamique des trois complexes étudiés en leurs modes de mouvement essentiels, de comparer leurs espaces conformationnels et d'identifier les résidus impliqués dans leur dynamique. Ces travaux ont montré une différence importante entre les espaces conformationnels explorés par les trois interactions *kissig-loop* étudiées, différence due en partie à la disponibilité de résidus libres dans l'interaction boucle-boucle naturelle. Afin de confirmer l'importance de ces résidus libres, nous avons effectué une expérience de sélection *in vitro* visant à obtenir un ribozyme doté d'une boucle V modifiée, capable de reconnaître et cliver efficacement un substrat doté de la boucle de TAR. Les résultats de cette sélection ont révélé des interactions *kissing-loop* possédant moins de paires de base que l'interaction TAR/TAR\* originale et dotées de capacités dynamiques intrinsèques se rapprochant de celles de l'interaction I/V.

De la manipulation de quelques paires de base à l'observation atomistique de la dynamique d'une interaction tertiaire cruciale, nous avons exploré le potentiel d'ingénierie du ribozyme VS de *Neurospora* au travers de ses structures et de sa dynamique à haute résolution. Ces données nous ont permis de bâtir un cadre théorique pour une meilleur appréhension de la fonction du ribozyme et, par extension, des ARN catalytiques.

## **2. Étude de la compensation de longueur d'hélice dans le ribozyme VS de *Neurospora***

## Mise en contexte

Considérant la fonction de clivage du ribozyme VS de *Neurospora*, sa caractérisation biochimique approfondie et le mode de reconnaissance unique de son substrat, l'idée de le modifier rationnellement afin d'en faire un outil de clivage d'ARN repliés en tige-boucles s'est logiquement imposée. À l'époque de la publication de cet article, la structure tridimensionnelle du ribozyme VS et par extension les modalités d'association des deux parties du site actif n'étaient pas connues. En outre, bien que les informations de mutagenèse dirigée fussent nombreuses, aucune étude ne visait à modifier de manière rationnelle le ribozyme VS au niveau de sous-domaines cruciaux pour la reconnaissance de son substrat. Il était donc important d'en savoir plus sur la conformation active du ribozyme VS ainsi que sur la possibilité de le modifier pour pouvoir accommoder de nouveaux substrats.

L'article présenté ici vise trois objectifs. D'abord confirmer l'hypothèse selon laquelle les hélices V et I sont orientées de manière plus ou moins colinéaire lorsque le ribozyme VS est dans sa conformation active. Ensuite, vérifier que le ribozyme est bien capable d'accueillir des substrats de longueurs différentes. Finalement, vérifier que pour des longueurs de substrat modifiées, il est possible de récupérer une activité notable en modifiant la longueur de l'hélice V de manière compensatoire.

J'ai contribué à cet article en modélisant les complexes boucle-boucle de différentes longueurs. J'ai ensuite caractérisé ces modèles structurellement grâce à deux métriques, l'angle entre les axes des hélices I et V et le décalage axial entre ces deux mêmes axes puis je les ai regroupées par classe par une méthode de clustering. Je les ai évaluées dans une étude statistique visant à déterminer quel modèle expliquait le mieux les données cinétiques obtenues par ailleurs par Julie Lacroix-Labonté. L'article a été rédigé par Pascale Legault et révisé par Julie Lacroix-Labonté et moi-même.

# **Helix-length compensation studies reveal the adaptability of the VS ribozyme architecture**

*Julie Lacroix-Labonté<sup>1</sup>, Nicolas Girard<sup>1</sup>, Sébastien Lemieux<sup>2</sup> and Pascale Legault<sup>1, \*</sup>*

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

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

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

\*To whom correspondence should be addressed. Tel:+1 514 343 7326; Fax:+1 514 343 2210; Email: [pascale.legault@umontreal.ca](mailto:pascale.legault@umontreal.ca)

This work was originally published in Nucleic Acid Research:

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



## **Abstract**

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

## Introduction

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

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

In this work, we extend the concept of helix-length compensation for structure–function studies of RNA using the *Neurospora* Varkud satellite (VS) ribozyme as a model system [for reviews, see [8–10]; Figure 2.1A]. Like other nucleolytic ribozymes, the VS ribozyme adopts a

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

## Material and Methods

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

**Preparation of RNA for kinetic studies.** The R<sub>0</sub> ribozyme and variant sequences (Figure 2.1) were prepared by in vitro T7 RNA polymerase transcription from the pAvapl plasmid (or derivatives) linearized at the AvaI site. All ribozymes were purified as previously described [18]. The S<sub>0</sub> substrate and variant sequences (Figure 2.1) were synthesized in vitro using T7 RNA polymerase and a single-stranded synthetic oligonucleotide template (Integrated DNA Technologies, IA). They were purified by denaturing gel electrophoresis, dephosphorylated with calf alkaline phosphatase and 5'-end-labeled with  $\gamma$ -(<sup>32</sup>P) ATP (PerkinElmer, Ont.) using established procedures [18].

**Single-substrate kinetic studies.** All reactions were carried out at 30°C in reaction buffer (50 mM Tris pH 7.5, 25 mM KCl, 2 mM spermidine and 25 mM MgCl<sub>2</sub>). Prior to starting the cleavage reaction, the 5'-<sup>32</sup>P-labeled substrate was heated to 95°C for 1 min in storage buffer (10 mM Tris pH 7.6 and 1 mM EDTA) and snap-cooled to promote hairpin formation. Cleavage reactions were initiated by addition of equal volumes (20 µl) of the 5'-<sup>32</sup>P-labeled S and R, either the wild-type R (R<sub>0</sub>) or variants (Figure 2.1B), after a 5-min pre-incubation of S and R at 30°C in reaction buffer. Typically, about 10 aliquots of 3 µl were removed from the 40-µl reaction mixture at specific times, and the reaction was stopped by addition of 12 µl of stop mix (50 mM EDTA pH 8.0, 0.04% xylene cyanol, 0.04% bromophenol blue and ≥85% formamide) and transfer to 4°C. Samples were analyzed by denaturing gel electrophoresis [20% acrylamide:bisacrylamide (19:1) / 7 M urea gels]. For each time point, the amounts of substrate and product were quantified using a Molecular Imager FX and the Quantity One 1D analysis software version 4.6.2 (Bio-Rad Laboratories, Ont.).

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

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

**Multiple-substrate kinetic studies.** Multiple-substrate kinetic studies were carried out as single-turnover reactions under the same conditions as the single-substrate kinetic studies. The main difference is that the cleavage reactions were initiated by addition of equal volumes (20  $\mu$ l) of R (0.005–5  $\mu$ M of either  $R_0$  or a variant) and a mixture of 5'-<sup>32</sup>P-labeled nucleic acids containing all substrates ( $S_{-2}$  bp,  $S_{-1}$  bp,  $S_0$ ,  $S_{+1}$  bp,  $S_{+2}$  bp,  $S_{+3}$  bp and  $S_{+4}$  bp; 35 pM of each) and, in certain cases, a non-cleavable DNA hairpin control (5'- CGG CGA AGC ATC GTC GGC TTC GCC G -3'; 35 pM). In addition, the percentage of remaining substrate was expressed as  $F = (S_t/S_0) * (C_0/C_t) * 100\%$  and plotted against time, where  $S_0$  and  $C_0$  are respectively the amount of substrate and non-cleavable  $S_{-2}$ bp control at time  $t = 0$  and  $S_t$  and  $C_t$  represent the corresponding values at time  $t \neq 0$ . Values for the reaction rate  $k_{obs}$ , amount of uncleaved substrate  $F_\infty$  and specificity constant  $k_{cat}/K_M$  were derived as for the single-substrate kinetic studies. All multiple-substrate kinetic experiments were performed at least three times, and reported  $k_{cat}/K_M$  and their errors are average values and standard deviations, respectively, from these multiple measurements.

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

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

**Structural modeling of SLI/SLV complexes.** 3D models of the SLI/SLV complex were generated using a stepwise procedure (Supplementary Figure S2.3). As a first step, RNA fragments were modeled using the constraint-satisfaction software MC-Sym 3.3.2 [30]. Several models of the I/V kissing-loop junction core (residues 626–634 and 695–701) were first

obtained separately using the NMR structure of magnesium-bound SLV [31] and constraints derived from biochemical studies, as previously described [18]. A unique structural model was obtained for each type of SLI stems ( $S_{-1\text{bp}}$  to  $S_{+4\text{bp}}$ ; residues 619–625 and 635–640 for  $S_0$ ), in accordance with the secondary structure model (Figure 2.1) and the NMR structure of the active conformation of the cleavage site internal loop [32]. A unique structural model was also obtained for each type of SLV stems ( $R_{-4\text{bp}}$  to  $R_{+1\text{bp}}$ ; residues 689–694 and 702–707 for  $R_0$ ), in accordance with the secondary structure model (Figure 2.1). Before being grafted together, the core and stem models were energy minimized using the molecular dynamics software Sander [33] based on the Amber force field and including covalent, electrostatic, Van der Waals energies, as well as implicit solvation in the form of a Generalized Born model [34]. Several rounds of minimization (100 000) were first applied including restraints to maintain Watson-Crick base pairing and followed by fewer rounds of minimization (1000) without these restraints. The minimized core and stem models were then filtered using a custom C++ program based on the freely accessible C++ library MC-Core [30]. The filter-selected models that preserved global structural integrity (heavy-atom rmsd between the pre- and post-minimization model  $<2.0 \text{ \AA}$ ) and key structural features (covalent integrity of backbone, Watson-Crick base pairs at the I/V kissing-loop junction, and hydrogen bonds typical of U-turn structure in SLI between U628 2'-O and G630 N7 and between U628 N3 and U631) were kept. Each remaining core model was then assembled using MC-Sym with all possible combinations of stem I ( $S_{-1\text{bp}}$  to  $S_{+4\text{bp}}$ ) and stem V ( $R_{-4\text{bp}}$  to  $R_{+1\text{bp}}$ ), energy minimized and filtered to create 36 sub-models that form a common-core group.

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

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

## Results

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

**Single-substrate kinetic studies.** We first examined the specificity of  $R_0$  for each substrate individually by determining the second-order rate constant ( $k_{\text{cat}}/K_M$ ) under single-turnover conditions [18], as illustrated for  $S_0$  in Figure 2.2. Cleavage reactions were performed with  $^{32}\text{P}$ -labeled  $S_0$  and excess  $R_0$  and monitored by denaturing gel electrophoresis (Figure

2.2A). The percentage of remaining substrate was plotted against time to derive the value of the first-order rate constant  $k_{obs}$  (Figure 2.2B). For  $S_{-1}$  bp,  $S_0$  and  $S_{+1}$ bp a linear relationship is observed when  $k_{obs}$  is plotted as a function of  $R_0$  concentration, and the values of catalytic efficiency ( $k_{cat}/K_M$ ) were obtained from the slope of this graph (Figure 2.2C and Table 2.1). For all other substrates ( $S_{-2}$  bp,  $S_{+2}$  bp,  $S_{+3}$ bp and  $S_{+4}$ bp), the value of  $k_{obs}$  is below the detection limit ( $\leq 0.001 \text{ min}^{-1}$ ) at  $R_0$  concentrations up to  $5 \mu\text{M}$ , and therefore the value of  $k_{cat}/K_M$  cannot be determined (Table 2.1). To insure that the lack of reactivity of certain substrates is not due to major folding defects, they were analyzed by native gel electrophoresis (Supplementary Figure S2.1). All substrates, with exception of  $S_{-2}$  bp, migrate as a single band on the gel, at a position typical of small RNA hairpins, in agreement with the predicted secondary structure [38]. Most of the  $S_{-2}$ bp substrate migrates slower, indicating that it adopts a duplex conformation that is incompatible with cleavage by  $R_0$ .

For  $R_0$ , the highest value of  $k_{cat}/K_M$  is obtained with  $S_0$  ( $11 \pm 1 \text{ min}^{-1} \mu\text{M}^{-1}$ ). Surprisingly, the  $k_{cat}/K_M$  values for  $S_{-1}$  ( $5.2 \pm 0.8 \text{ min}^{-1} \mu\text{M}^{-1}$ ) and  $S_{+1}$  ( $2.7 \pm 0.3 \text{ min}^{-1} \mu\text{M}^{-1}$ ) are reduced by only 2- and 4-fold, respectively, with respect to  $S_0$  (Table 2.1). The similar substrate specificities of  $R_0$  for  $S_{-1}$ ,  $S_0$  and  $S_{+1}$  point to the adaptability of the VS ribozyme architecture to cleave a range of substrates with 3–5 bp in stem Ib, indicating a certain level of promiscuity in substrate selection.

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



normalization, we initially added a non-cleavable DNA control in the cleavage reaction, but later found that quantification is more reliable using the non-cleavable  $S_{-2\text{bp}}$  RNA that gives a better-defined band on the gel (Figure 2.3).

The multiple-substrate kinetic procedure was first performed with  $R_0$ , revealing essentially identical  $k_{\text{cat}}/K_M$  values to those obtained with the single-substrate kinetic method (Table 2.1). These results confirm the reliability of the multiple-substrate procedure and, thus, similar studies were conducted with the variant ribozymes, except for  $R_{+2\text{bp}}$  because its helix-length complementary substrate  $S_{-2\text{bp}}$  is not cleavable. Interestingly, substrate cleavage is observed for several S/R pairs, with many having similar  $k_{\text{cat}}/K_M$  values as the  $S_0/R_0$  pair (Table 2.2). To verify that substrate cleavage for all functional S/R pairs yields the same product, the reaction products were analyzed by comparing their migration with alkaline and T1 digests on denaturing gels. The results clearly demonstrate that cleavage occurs between G620 and A621 for all substrates cleaved by either  $R_0$  or variant ribozymes (Supplementary Figure S2.2). Thus, the helix-length mutations investigated here have no effect on cleavage site selection.

As expected, helix-length compensation is observed for all variant ribozymes, since the highest values of  $k_{\text{cat}}/K_M$  are observed with the compensatory substrate or one that differs by a single base pair. More precisely, the most active S/R combinations for each ribozyme are:  $S_{-1\text{bp}}/R_{+1\text{bp}}$ ,  $S_0/R_0$ ,  $S_{+1\text{bp}}/R_{-1\text{bp}}$ ,  $S_{+1\text{bp}}/R_{-2\text{bp}}$ ,  $S_{+2\text{bp}}/R_{-3\text{bp}}$  and  $S_{+3\text{bp}}/R_{-4\text{bp}}$ . However,  $k_{\text{cat}}/K_M$  values similar to that of  $S_0/R_0$  (within 4-fold; shaded in Table 2.2) are observed only for four of the six ribozymes ( $R_0$ ,  $R_{-1\text{bp}}$ ,  $R_{-2\text{bp}}$  and  $R_{-3\text{bp}}$ ). For the  $R_{+1\text{bp}}$  and  $R_{-4\text{bp}}$  ribozymes, the maximal  $k_{\text{cat}}/K_M$  values are reduced significantly, ~80- and ~30-fold, respectively, compared to  $S_0/R_0$ .

**Structural modeling of SLI/SLV complexes compatible with kinetic data.** Catalysis generally requires the precise positioning of chemically active groups within the active site, which in the VS ribozyme results from the precise docking of the cleavage site and A730 internal loops [10,19–27]. Thus, it is likely that the cleavage site loop is presented in a favorable position for productive docking in the most active S/R combinations, but not in the inactive or poorly active S/R combinations. To explore how the relative positioning of the cleavage site loop is affected by helix-length mutations, a structural modeling approach was used. We chose to model an SLI/SLV complex as opposed to models of complete S/R pairs to simplify the modeling

procedure and limit the structural consequences of helix-length mutations. Furthermore, the SLI/SLV models allow us to examine the effect of helix-length variations on the positioning of the scissile phosphate in SLI relative to a common base pair in SLV. This approach is consistent with our kinetic studies of S/R pairs in which ribozyme variants retain a common core sequence and interact with the free substrates primarily through the I/V kissing-loop interaction (Bouchard, P., Delfosse, V. and Legault, P., unpublished data). We assumed that realistic 3D models would be compatible with the kinetic data, such that the relative positioning of the cleavage site loop in active S/R combinations would be distinct from that found with inactive S/R combinations.

3D models of the SLI/SLV complexes were generated using a stepwise procedure (Supplementary Figure S2.3). First, several models of the I/V kissing-loop junction core were obtained using previously described constraints [18]. Of these core models, 56 were then selected because each could be used to build a common-core group of 36 sub-models that contain all combinations of SLI ( $S_{-1\text{bp}}$  to  $S_{+4\text{bp}}$ ) and SLV stems ( $R_{-4\text{bp}}$  to  $R_{+1\text{bp}}$ ) and remain consistent with input structural constraints after energy minimization. For all selected sub-models ( $56 \times 36$ ), the interhelical angle and axial displacement between stems Ib and V were computed and average values were determined for each common-core group. A wide range of axial displacements ( $d_{I-V} = 12\text{--}24 \text{ \AA}$ ), and interhelical angles ( $\theta_{I-V} = 108^\circ$  to  $175^\circ$ ; Figure 2.4) are observed among the common-core groups, pointing to the extensive conformational sampling of the modeling procedure.

We hypothesized that for common-core groups compatible with the kinetic data, the sub-models of the high-activity class [ $(k_{\text{cat}}/K_M)/(k_{\text{cat}}/K_M)_0 \geq 0.25$ ]; Table 2.2] should be significantly different than those of the low-activity class [ $(k_{\text{cat}}/K_M)/(k_{\text{cat}}/K_M)_0 \leq 0.01$ ]; Table 2.2] of S/R pairs. In order to compare the structures of the sub-models, pairwise rmsd's were calculated between the sub-models of each common-core group and their reference  $S_0/R_0$  sub-model. For each common-core group, the rmsd distributions for the high-activity and low-activity classes of S/R pairs were compared using a Welch's *t*-test [36]. Interestingly, we observe a good correlation between *P*-values and interhelical angles (Figure 2.4), where common-core groups with average interhelical angles ( $\theta_{I-V}$ ) closer to  $180^\circ$  show smaller *P*-values and are thus more likely to be consistent with the kinetic data. Group 13 ( $\theta_{I-V} = 166.0^\circ \pm 3.6$ ) was selected for

structural analysis, but similar conclusions are obtained for other common-core groups with  $P$ -values smaller than the  $\alpha$  threshold value of 0.02 (group 12:  $\theta_{I-V} = 165.4^\circ \pm 3.5^\circ$  and group 20:  $\theta_{I-V} = 171.4^\circ \pm 4.5^\circ$ ).

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

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

## Discussion

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

specificity of the VS ribozyme architecture, while revealing its adaptability for efficient substrate cleavage. Two interesting aspects of this adaptability are limited substrate promiscuity with respect to stem Ib length and helix-length compensation between stems Ib and V.

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

**Helix-length compensation between stems Ib and V.** We initially hypothesized that the reduced cleavage activity of defective stem Ib variants by  $R_0$  could be restored by adjusting the number of base pairs within stem V, such that the total number of base pairs within the two stems would be preserved. Examination of the kinetic data indicates that there is, as predicted, helix-length compensation between stems Ib and V. Indeed, substrates for which cleavage is not detected with  $R_0$  ( $S_{+2}$  bp,  $S_{+3}$  bp,  $S_{+4}$  bp) are cleaved by a ribozyme variant. However, the cleavage activities of the compensatory S/R pairs are not all restored to the reference level. In agreement with these results, 3D modeling of the I/V kissing-loop complex indicates that the global orientation of stems Ib and V significantly affects the position of the scissile phosphate in the variant S/R pairs.

An unexpected aspect of this helix-length compensation study is that it helps provide global structural information on the I/V kissing-loop interaction. From an ensemble of SLI/SLV models that display a wide range of interhelical angles ( $108^\circ \leq \theta_{I-V} \leq 175^\circ$ ) and axial displacements ( $d_{I-V} = 12\text{--}24 \text{ \AA}$ ), three models were selected that best fit the kinetic data. These selected models display a much smaller range of interhelical angles and axial displacements

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

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

An emerging view in RNA structure formation is that tertiary interactions form as a result of conformational sampling of helical domains within a distribution of conformations that is limited, in part, by the inherent flexibility of helical connectivities [for recent reviews see [40,41]]. It is likely that in the VS ribozyme, flexible elements such as its bulges, internal loops,

three-way junctions and even its I/V kissing-loop junction [42,43] allow conformational sampling of helical domains, including SLI and SLVI, through which a well-defined active site can emerge. According to this view, the ability to sample the active conformation may be conserved in high-activity S/R pairs but not in those with lower activity. Future investigations are needed to help identify the dynamic features responsible for the conformational adaptability of the VS ribozyme and determine if a similar adaptability exists in other functional RNAs.

**Prospects for RNA research.** Helix-length compensation represents an alternative approach for investigating the sequence and structural repertoire of functional RNAs that could be applied to other helix pairs. In principle, helix-length compensation could be observed for helices that are more or less coaxially stacked [1–6] or those that are not coaxial but nearly colinear as observed for the VS ribozyme. Although helix-length compensations have been previously identified by comparative sequence analysis [1–6], such analysis may not identify all possible helix-length mutants. Indeed, the genetic mechanism through which such complex mutation pattern evolves likely requires multiple mutational events with an overall lower probability than simple covariation of base-pairing residues. Moreover, sequence searches may not always take into account the possibility for helix-length compensation. Systematic helix-length compensation studies could help improve the consensus sequence and structural model for a given functional RNA. Subsequently, revised RNA consensus sequence could be used for bioinformatics database search of additional representatives of known functional RNAs [44–47]. Helix-length compensation studies could also help reveal the inherent adaptability of other complex RNA architectures and be useful for the design of RNAs with novel functions.

## References

1. Woese, C.R., et al., *Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids*. *Microbiol Rev*, 1983. **47**(4): p. 621-69.
2. Gutell, R.R. In: Zimmermann, R.A. and Dahlberg, A.E. (eds), *Ribosomal RNA: Structure, Evolution, Processing, and Function in Protein Biosynthesis*. CRC Press, Boca Raton, Florida, p. 109–126.

3. Ban, N., et al., *The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution*. Science, 2000. **289**(5481): p. 905-20.
4. Michel, F. and E. Westhof, *Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis*. J Mol Biol, 1990. **216**(3): p. 585-610.
5. Adams, P.L., et al., *Crystal structure of a self-splicing group I intron with both exons*. Nature, 2004. **430**(6995): p. 45-50.
6. Golden, B.L., H. Kim, and E. Chase, *Crystal structure of a phage Twort group I ribozyme-product complex*. Nat Struct Mol Biol, 2005. **12**(1): p. 82-9.
7. Steinberg, S., F. Leclerc, and R. Cedergren, *Structural rules and conformational compensations in the tRNA L-form*. J Mol Biol, 1997. **266**(2): p. 269-82.
8. Collins, R.A., *The Neurospora Varkud satellite ribozyme*. Biochem Soc Trans, 2002. **30**(Pt 6): p. 1122-6.
9. Lilley, D.M., *The Varkud satellite ribozyme*. RNA, 2004. **10**(2): p. 151-8.
10. Wilson, T.J. and D.M. Lilley, *Do the hairpin and VS ribozymes share a common catalytic mechanism based on general acid-base catalysis? A critical assessment of available experimental data*. RNA, 2011. **17**(2): p. 213-21.
11. Rastogi, T., et al., *A long-range pseudoknot is required for activity of the Neurospora VS ribozyme*. EMBO J, 1996. **15**(11): p. 2820-2825.
12. Andersen, A.A. and R.A. Collins, *Rearrangement of a stable RNA secondary structure during VS ribozyme catalysis*. Mol Cell, 2000. **5**(3): p. 469-78.
13. Andersen, A.A. and R.A. Collins, *Intramolecular secondary structure rearrangement by the kissing interaction of the Neurospora VS ribozyme*. Proc Natl Acad Sci U S A, 2001. **98**(14): p. 7730-5.
14. Hiley, S.L. and R.A. Collins, *Rapid formation of a solvent-inaccessible core in the Neurospora Varkud satellite ribozyme*. EMBO J, 2001. **20**(19): p. 5461-9.
15. 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.
16. 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.

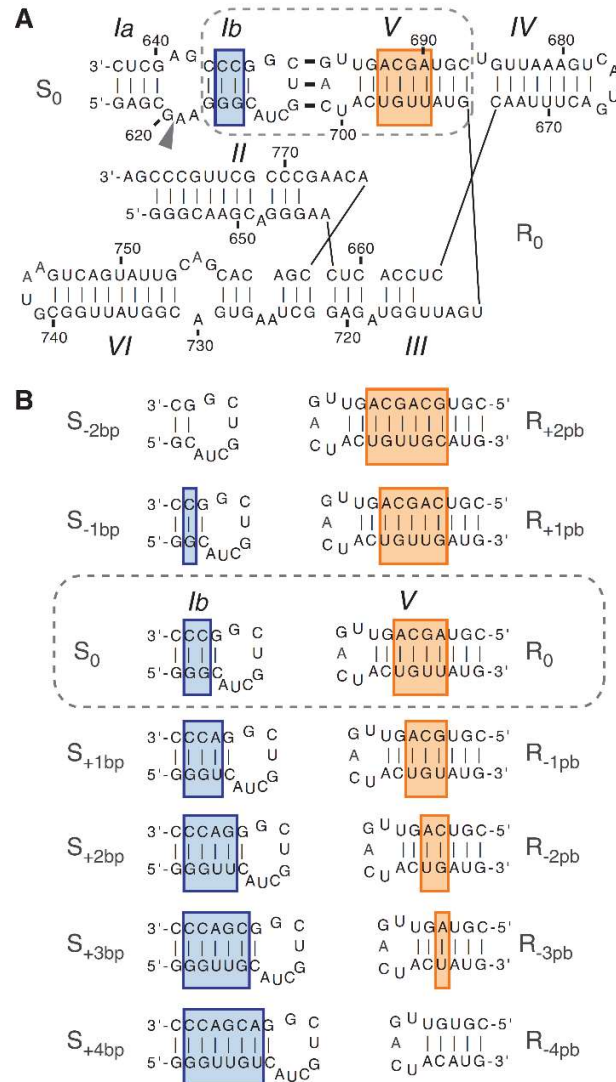
17. Lipfert, J., et al., *The complete VS ribozyme in solution studied by small-angle X-ray scattering*. Structure, 2008. **16**(9): p. 1357-67.
18. Bouchard, P., et al., *Role of SLV in SLI substrate recognition by the Neurospora VS ribozyme*. RNA, 2008. **14**(4): p. 736-48.
19. Lafontaine, D.A., et 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.
20. 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.
21. 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.
22. 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.
23. Zhao, Z.Y., et al., *Nucleobase participation in ribozyme catalysis*. J Am Chem Soc, 2005. **127**(14): p. 5026-7.
24. 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.
25. 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 U S A, 2007. **104**(14): p. 5818-23.
26. 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.
27. Wilson, T.J., et al., *Nucleobase-mediated general acid-base catalysis in the Varkud satellite ribozyme*. Proc Natl Acad Sci U S A, 2010. **107**(26): p. 11751-6.
28. 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.
29. 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.
30. Major, F., et al., *The combination of symbolic and numerical computation for three-dimensional modeling of RNA*. Science, 1991. **253**(5025): p. 1255-60.



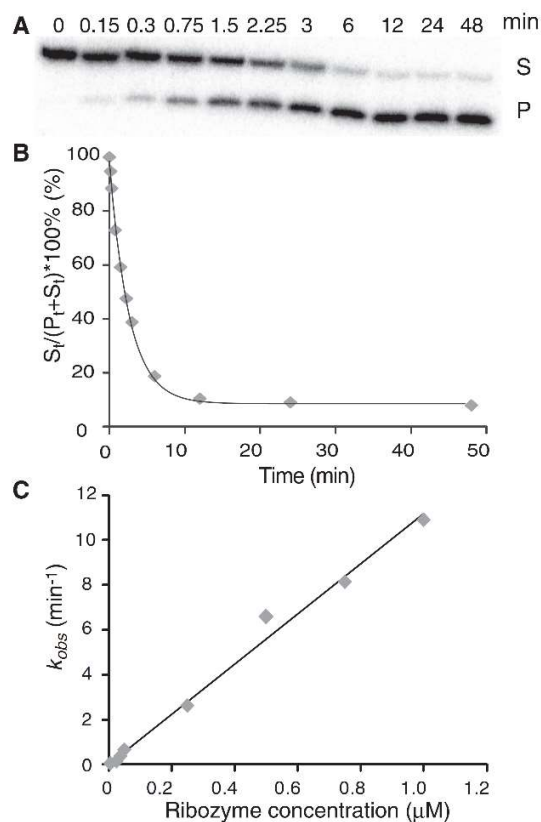
31. Campbell, D.O., et al., *NMR structure of varkud satellite ribozyme stem-loop V in the presence of magnesium ions and localization of metal-binding sites*. *Biochemistry*, 2006. **45**(35): p. 10591-605.
32. Hoffmann, B., et al., *NMR structure of the active conformation of the Varkud satellite ribozyme cleavage site*. *Proc Natl Acad Sci U S A*, 2003. **100**(12): p. 7003-8.
33. Pearlman, D.A., et al., *Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules*. *Computer Physics Communications*, 1995. **91**(1-3): p. 1-41.
34. Still, W.C., et al., *Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics*. *Journal of the American Chemical Society*, 1990. **112**(16): p. 6127-6129.
35. Lavery, R., et al., *Conformational analysis of nucleic acids revisited: Curves+*. *Nucleic Acids Res*, 2009. **37**(17): p. 5917-29.
36. Satterthwaite, F.E., *An approximate distribution of estimates of variance components*. *Biometrics*, 1946. **2**(6): p. 110-4.
37. Shapiro, S.S. and M.B. Wilk, *An Analysis of Variance Test for Normality (Complete Samples)*. *Biometrika*, 1965. **52**(3/4): p. 591-611.
38. Beattie, T.L., J.E. Olive, and R.A. Collins, *A secondary-structure model for the self-cleaving region of Neurospora VS RNA*. *Proc Natl Acad Sci U S A*, 1995. **92**(10): p. 4686-90.
39. Doudna, J.A., B.P. Cormack, and J.W. Szostak, *RNA structure, not sequence, determines the 5' splice-site specificity of a group I intron*. *Proc Natl Acad Sci U S A*, 1989. **86**(19): p. 7402-6.
40. Woodson, S.A., *Compact intermediates in RNA folding*. *Annu Rev Biophys*, 2010. **39**: p. 61-77.
41. Bailor, M.H., et al., *Topological constraints: using RNA secondary structure to model 3D conformation, folding pathways, and dynamic adaptation*. *Curr Opin Struct Biol*, 2011. **21**(3): p. 296-305.
42. Szewczak, A.A. and T.R. Cech, *An RNA internal loop acts as a hinge to facilitate ribozyme folding and catalysis*. *RNA*, 1997. **3**(8): p. 838-49.

43. Bajor, M.H., X. Sun, and H.M. Al-Hashimi, *Topology links RNA secondary structure with global conformation, dynamics, and adaptation*. Science, 2010. **327**(5962): p. 202-6.
44. Ferbeyre, G., J.M. Smith, and R. Cedergren, *Schistosome satellite DNA encodes active hammerhead ribozymes*. Mol Cell Biol, 1998. **18**(7): p. 3880-8.
45. Graf, S., et al., *A database search for hammerhead ribozyme motifs*. Biochem Soc Trans, 2005. **33**(Pt 3): p. 477-8.
46. Hammann, C. and E. Westhof, *Searching genomes for ribozymes and riboswitches*. Genome Biol, 2007. **8**(4): p. 210.
47. McCown, P.J., A. Roth, and R.R. Breaker, *An expanded collection and refined consensus model of glmS ribozymes*. RNA, 2011. **17**(4): p. 728-36.

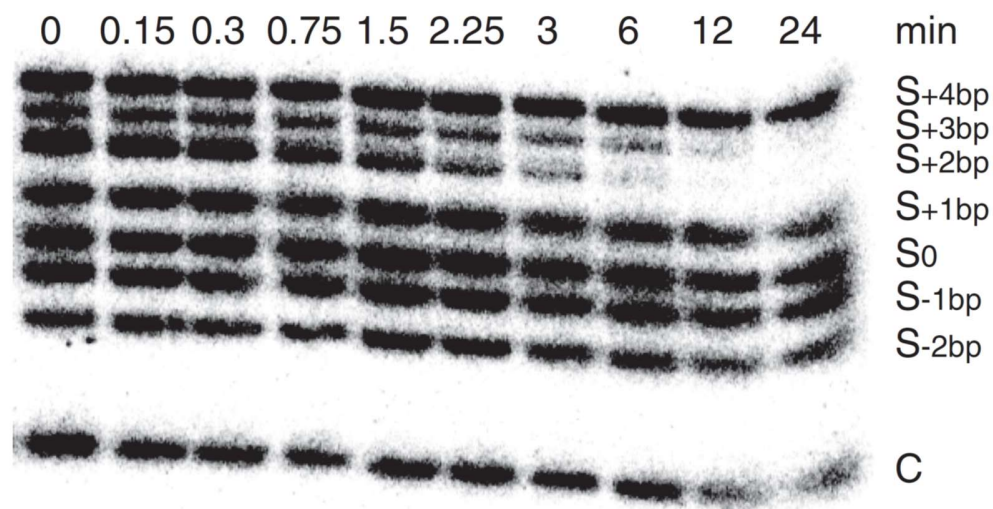
## Figures



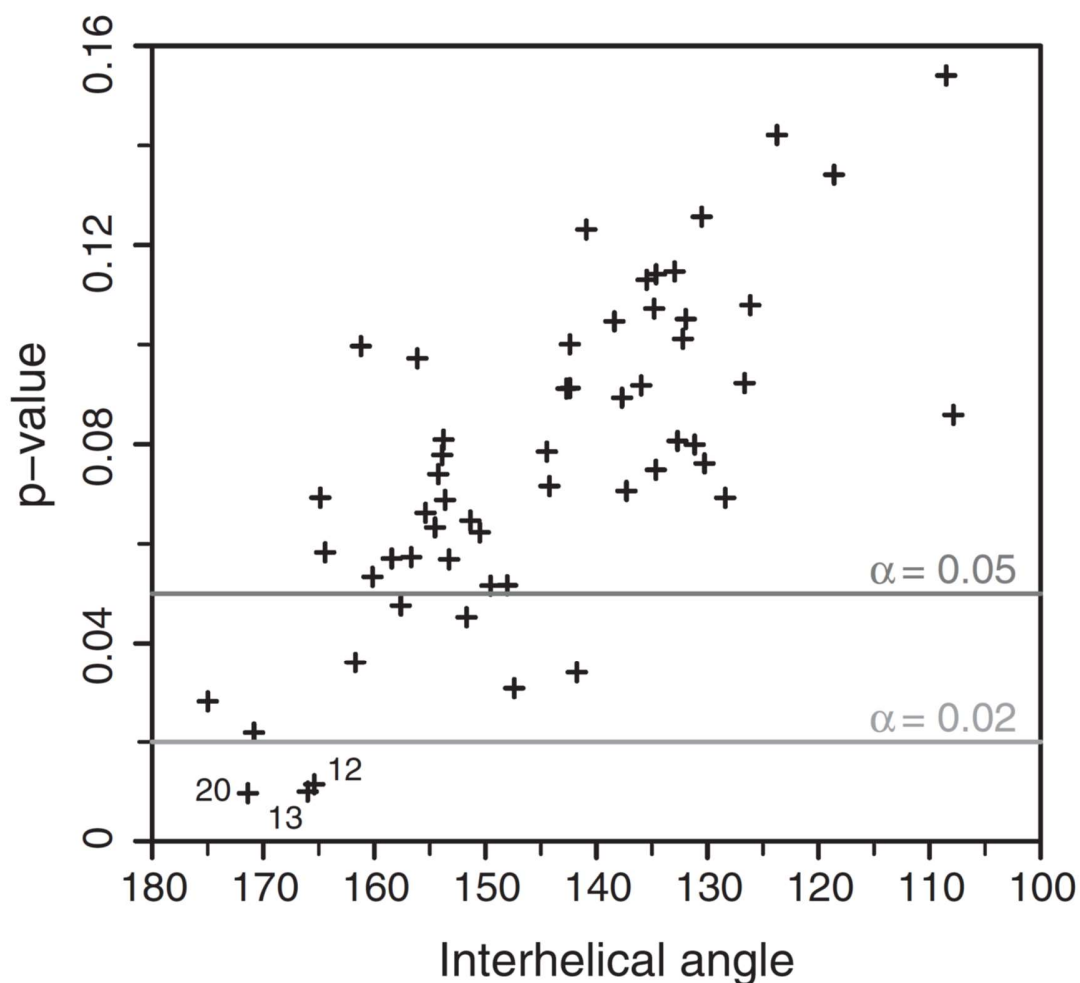
**Figure 2.1. Primary and secondary structures of the substrate/ribozyme (S/R) pairs characterized here.** (A) Parental VS ribozyme ( $R_0$ ) and substrate ( $S_0$ ). Substrate recognition involves a kissing-loop interaction between residues 630–632 of  $S_0$  and 697–699 of  $R_0$ . The arrowhead points to the cleavage site and the dashed box encloses the region under investigation. (B) Variant S/R pairs. Only the mutated regions are shown along with the nomenclature of the mutated substrates and ribozymes.



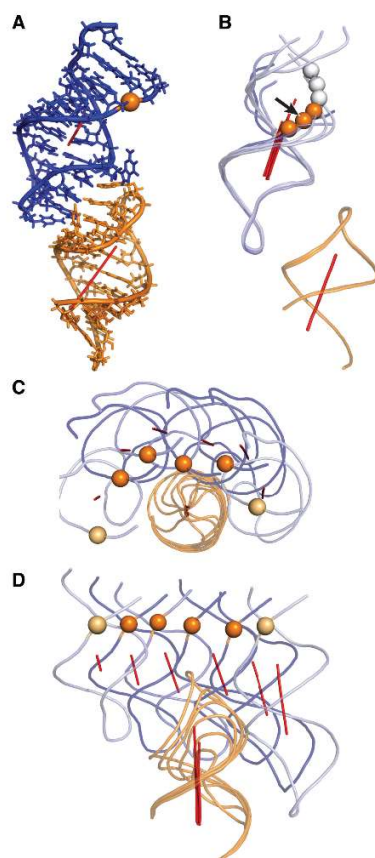
**Figure 2.2. Single-substrate kinetic analysis of  $R_0$ .** (A) Cleavage of 5'- $^{32}\text{P}$ -labeled  $S_0$  by  $R_0$  (at 37.5 nM) monitored by denaturing gel electrophoresis. (B) The percentage of remaining substrate was plotted against time. The data were fitted to the non-linear equation  $F = Ae^{-(k_{obs} * t)} + F_{\infty}$ , to extract the value of  $k_{obs}$  ( $k_{obs} = 0.332 \pm 0.008 \text{ min}^{-1}$ ;  $R^2 = 1.00$ ). (C) Cleavage reactions were performed at various concentrations of  $R_0$ , and the value of  $k_{obs}$  was plotted against ribozyme concentration. For all  $k_{obs}$  values, error bars are smaller than the data point on the graph. The data were fitted to a single exponential equation, and the  $k_{cat}/K_M$  is extracted by linear regression ( $k_{cat}/K_M = \text{value of the slope} = 11.2 \text{ min}^{-1}\mu\text{M}^{-1}$  and  $R^2 = 0.99$ ).



**Figure 2.3. Simultaneous kinetic characterization of  $R_{-3bp}$  with multiple substrates.** The cleavage reactions of  $5'$ - $^{32}P$ -labeled substrates (S) by  $R_{-3bp}$  (at 100 nM) were monitored by denaturing gel electrophoresis. A DNA hairpin was used as a normalization control (C).



**Figure 2.4. Identification of SLI/SLV models compatible with the kinetic data.** Plot of the  $P$ -value (Welch's  $t$ -test) against the mean interhelical angle. The  $P$ -value was obtained from a Welch's  $t$ -test that compared the rmsd distributions of the high-activity and low-activity classes of S/R pairs for a given common core group. The mean interhelical angle was obtained by computing the average interhelical angles of the 36 S/R pairs for each common-core group. All common-core groups showing a  $P$ -value  $< 0.02$  are numbered. Standard deviations on the interhelical angles vary between  $2^\circ$  and  $7^\circ$ , but were omitted from the graph for clarity.



**Figure 2.5. Model of an SLI/SLV complex compatible with the kinetic data** (common-core group 13). (A) Stick representation of the  $S_0/R_0$  sub-model. (B) Cartoon representation of a superposition of the five variant  $S/R_0$  sub-models on the  $S_0/R_0$  sub-model. The black arrow points at the scissile phosphate of  $S_0/R_0$ . The scissile phosphates of the following substrates are shown from bottom to top:  $S_{-1}$ ,  $S_0$ ,  $S_{+1}$ ,  $S_{+2}$ ,  $S_{+3}$ ,  $S_{+4}$ . (C) Top-down and (D) front views of the superposition of the most catalytically efficient sub-models of each ribozyme. The following S/R pairs are shown from left to right:  $S_{-1}/R_{+1}$ ,  $S_0/R_0$ ,  $S_{+1}/R_{-1}$ ,  $S_{+1}/R_{-2}$ ,  $S_{+2}/R_{-1}$ ,  $S_{+3}/R_{-4}$ . In (B), (C) and (D), the sub-models were superposed to the  $S_0/R_0$  sub-model based on heavy atom alignment of the first base pair of SLV (residues 689 and 707 in  $S_0/R_0$ ). SLI and SLV are represented in deep blue and orange, respectively. Helical axes are represented as red rods and the phosphorus atom at the cleavage site is shown as a sphere. The phosphorus atoms are color-coded according to the catalytic efficiency of the sub-models.  $[(k_{cat}/K_M)/(k_{cat}/K_M) \geq 0.25]$ : deep orange;  $[0.25 > (k_{cat}/K_M)/(k_{cat}/K_M) > 0.01]$ : light orange and  $[(k_{cat}/K_M)/(k_{cat}/K_M) \leq 0.01]$ : white.

## Tables

**Table 2.1. Summary of kinetic data for  $R_0^a$**

	Values of $k_{cat}/K_M$ ( $\text{min}^{-1}\mu\text{M}^{-1}$ )	
	Single-substrate kinetics	Multiple-substrate kinetics
S-2bp	n.d. <sup>a</sup>	n.d.
S-1bp	$5.2 \pm 0.8$	$5.9 \pm 0.7$
S <sub>0</sub>	$11 \pm 1$	$11 \pm 2$
S+1bp	$2.7 \pm 0.3$	$2.7 \pm 0.4$
S+2bp	n.d.	n.d.
S+3bp	n.d.	n.d.
S+4bp	n.d.	n.d.

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

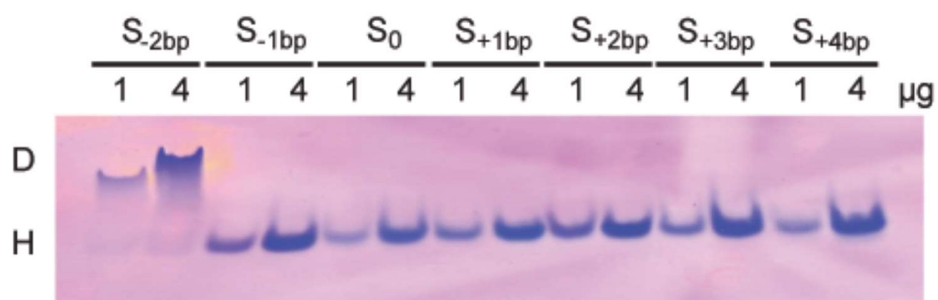


**Table 2.2. Summary of multiple-substrate kinetic data<sup>a</sup>**

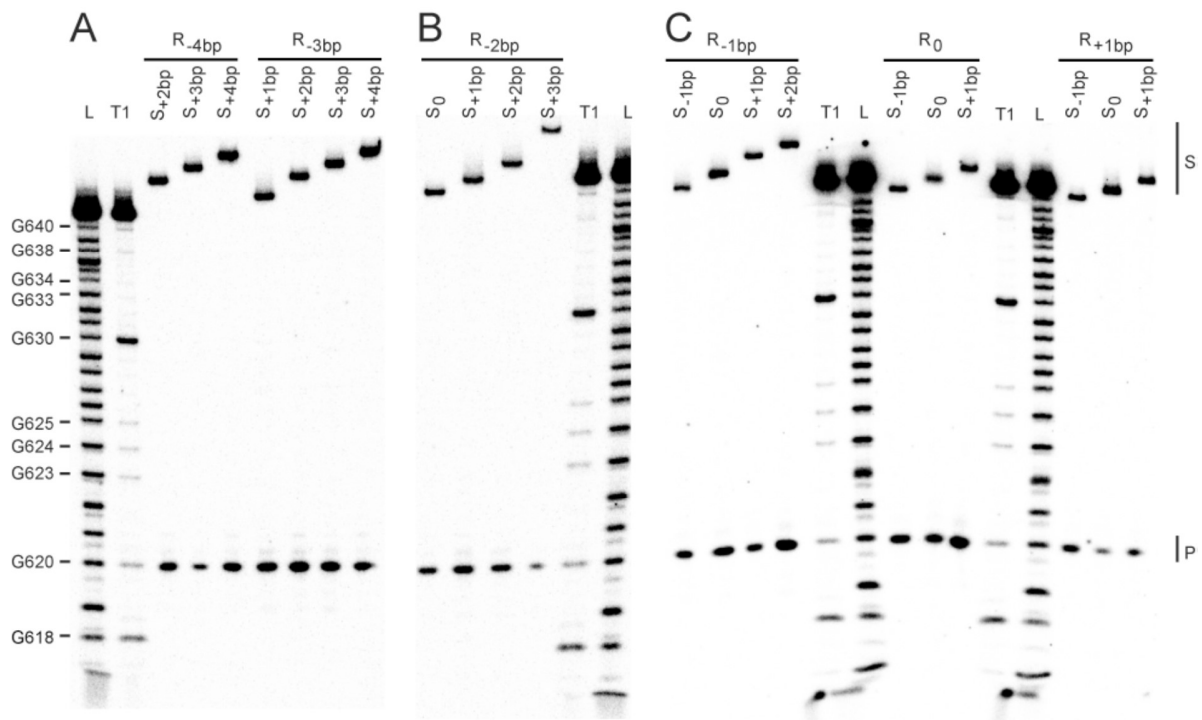
	R-4bp		R-3bp		R-2bp		R-1bp		R <sub>0</sub>		R+1bp	
	$k_{cat}/K_M$	ratio	$k_{cat}/K_M$	ratio	$k_{cat}/K_M$	ratio	$k_{cat}/K_M$	ratio	$k_{cat}/K_M$	ratio	$k_{cat}/K_M$	ratio
S-1bp	n.d.		n.d.		n.d.		0.11 ± 0.02	0.0095	5.9 ± 0.7	0.54	0.13 ± 0.02	0.012
S <sub>0</sub>	n.d.		n.d.		0.11 ± 0.02	0.01	12.0 ± 0.9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

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

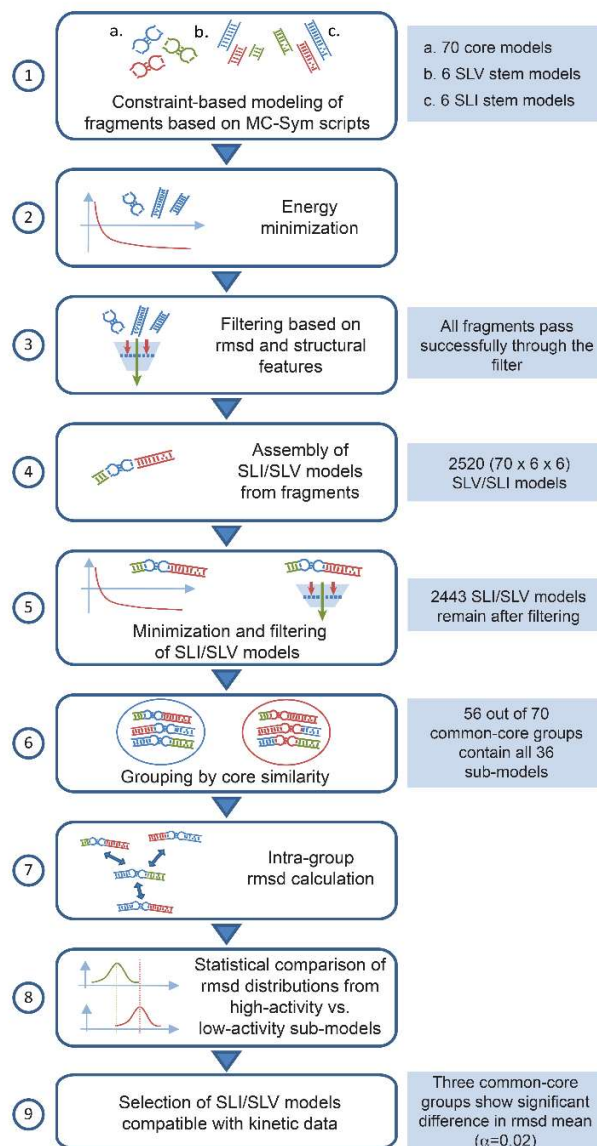
## Supporting Information



**Figure S2.1. Conformational analysis of the substrates used in this study by native gel electrophoresis.** For each substrate, two quantities of RNA (1 and 4  $\mu\text{g}$ ) were loaded on the gel. All substrates migrate at a similar position on the gel, except for S<sub>-2bp</sub> which migrates slower, indicating that all but one of the substrates are in a single hairpin (H) conformation, whereas S<sub>-2bp</sub> preferentially forms a duplex (D).



**Figure S2.2. Identification of a common cleavage site for all functioning S/R pairs.** Single substrate cleavage reactions of 5'-<sup>32</sup>P-labeled substrates by: (A) R<sub>-4bp</sub> and R<sub>-3bp</sub>; (B) R<sub>-2bp</sub>; and (C) R<sub>-1bp</sub>, R<sub>0</sub> and R<sub>+1bp</sub>. These reactions were analyzed by denaturing gels, which show the substrates (S) and the 5'-product (P). For comparison, S<sub>0</sub> was subjected to partial alkaline hydrolysis (L) and nuclease T1 digestion (T1).



**Figure S2.3. Flowchart describing the stepwise procedure used to generate three-dimensional models of the SLI/SLV complex and select those compatible with kinetic data. Complete details are provided in Materials and Methods.**

### **3. Ingénierie rationnelle du ribozyme VS de Neurospora par substitution de l'interaction kissing-loop I/V**

## Mise en contexte

L'article précédent a démontré à la fois la capacité du ribozyme VS à subir des modifications au niveau de la longueur des hélices I et V sans perdre son activité et sa flexibilité quant à sa reconnaissance de substrats différents. L'étape suivante dans le processus de modification rationnelle du ribozyme VS consiste à tenter de modifier le cœur du système de reconnaissance du substrat, soit l'interaction boucle-boucle elle-même, afin de permettre aux variantes d'accommoder des substrats d'intérêt, repliés en tige-boucle.

Pour ce faire, nous avons procédé au remplacement de l'interaction boucle-boucle naturelle par des candidats similaires identifiés à l'aide d'une recherche bio-informatique des bases de données structurales. Cette recherche a eu pour but d'isoler des interactions boucle-boucle structurellement similaires à l'interaction SLI/SLV, au niveau du nombre de paires de bases, de la taille des boucles et de l'orientation des hélices connectées. Les candidats potentiels ont été intégrés dans des variantes du ribozyme VS afin d'évaluer la cinétique de leurs réactions. Les études cinétiques ont révélé que les variantes du ribozyme VS dotées des interactions *kissing-loop* modifiées possédaient une activité de clivage substantielle mais environ 100 fois moindre que le ribozyme VS pourvu de l'interaction naturelle. Finalement, le complexe boucle-boucle naturel et celui des deux variantes retenues ont été modélisés et superposés sur la structure obtenue par cristallographie aux rayons X afin de déterminer les causes entraînant une activité moindre des ribozymes pourvus des variantes de l'interaction boucle-boucle.

J'ai contribué à cet article en développant un pipeline basé sur le serveur WebFR3D permettant de scanner la base de données RCSB à la recherche d'interactions boucle-boucle définies d'une manière symbolique (c'est-à-dire par les relations entre résidus plutôt que par homologie à un ensemble de positions dans l'espace) puis de les annoter et les classifier automatiquement. J'ai également effectué l'alignement sur le modèle de cristallographie. Tous les protocoles biochimiques ont été réalisés par Julie Lacroix-Labonté. L'article a été écrit conjointement par Pascale Legault, Julie Lacroix-Labonté et moi-même.

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

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

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

\*To whom correspondence should be addressed. Tel: +1 514 343 7326; Fax: +1 514 343 2210; Email: [pascale.legault@umontreal.ca](mailto:pascale.legault@umontreal.ca)

This work was originally published in Nucleic Acid Research:

Nucleic Acids Res. 2016 Aug 19;44(14):6924-34. doi: 10.1093/nar/gkw401. Epub 2016 May 10.





## Abstract

The *Neurospora* VS ribozyme is a catalytic RNA that specifically recognizes and cleaves a stem-loop substrate. Substrate recognition is achieved through formation of a highly stable kissing-loop interaction (KLI) between the stem-loop I (SLI) substrate and stem-loop V (SLV) of the *trans* VS ribozyme. In order to explore the engineering potential of the VS ribozyme to cleave alternate substrates, we previously successfully engineered VS ribozyme variants that cleave isolated SLI substrates in which the length of stem Ib is either increased or decreased. Here, we pursued this line of work by investigating the possibility of substituting the I/V KLI with other known KLIs using an innovative engineering approach that combines rational and combinatorial design. We first performed a symbolic search of the PDB using WebFR3D to identify alternative KLIs that are structurally similar to the I/V KLI of the VS ribozyme. Next, we kinetically characterized substrate/ribozyme pairs in which the VS ribozyme I/V KLI was substituted for alternative KLIs. Interestingly, several of the resulting substrate/ribozyme pairs were found to be considerably active at 37°C, clearly establishing that the KLI of the *trans* VS ribozyme can be adapted to cleave other folded RNA substrates.

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

## Introduction

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

The catalytic domain or *trans* VS ribozyme (helical domains II–VI; Figure 3.1A) recognizes its hairpin substrate (SLI) via the formation of a kissing-loop interaction (KLI) required for efficient cleavage activity [10,11]. In earlier work, mutagenesis and chemical modification studies provided evidence that this KLI contains three WC/WC [12] base pairs (bp) between loop I (G<sub>630</sub>, U<sub>631</sub> and C<sub>632</sub>) and loop V (G<sub>697</sub>, A<sub>698</sub> and C<sub>699</sub>), and that both loops likely form U-turn motifs [11]. More recently, the nuclear magnetic resonance (NMR) structure of an SLI/SLV complex [13] and the crystal structure of the complete VS ribozyme [14] confirmed formation of the U-turn motifs and three *cis* WC/WC bp, but they also revealed the presence of a *cis* WC/sugar edge bp between C<sub>629</sub> and A<sub>701</sub> (between C<sub>10</sub> of SLI and A<sub>14</sub> of SLV in the SLI/SLV complex; Figure 3.1B). *In vitro* selection and biochemical studies of the VS ribozyme [15,16] as well as NMR studies of several SLI/SLV complexes [13] have provided evidence that formation of the I/V KLI is associated with a helix shift in the SLI substrate that is essential for catalysis. SLI variants that adopt a pre-shifted conformation bind with higher affinity to both the *trans* ribozyme [17] and an isolated SLV [18]. Moreover, isothermal titration

calorimetry (ITC) studies of SLI/SLV complexes conducted with a pre-shifted SLI RNA demonstrated that the I/V KLI is remarkably stable ( $K_d = 0.24 \pm 0.02 \mu\text{M}$ ) and constitutes the predominant energetic contribution for substrate binding by the *trans* VS ribozyme [18]. The I/V KLI is also important for substrate docking with the catalytic domain, which allows for the formation of the active site [19]. In the proposed general acid-base mechanism of the VS ribozyme cleavage reaction, G<sub>638</sub> from the internal loop of SLI (G<sub>638</sub> loop) acts as the general base and A<sub>756</sub> from the internal loop of SLVI (A<sub>756</sub> loop) acts as the general acid [20–27]. The recent crystal structure of the VS ribozyme confirms that the active site is formed through the association of these two internal loops, in which G<sub>638</sub> and A<sub>756</sub> are properly positioned with respect to the scissile phosphate to fulfill their proposed roles in catalysis [14].

Only a limited number of studies have investigated the possibility of using the VS ribozyme to selectively cleave RNA stem-loops other than its natural SLI substrate. Although minor changes to the sequence of the natural SLI substrate are compatible with VS ribozyme cleavage [11,15,28–31], larger sequence changes in the substrate may only be possible by incorporating compensating changes in the catalytic domain. To test this possibility, we recently investigated the cleavage of variant SLI substrates, in which the number of base pairs in stem Ib was either increased or decreased, using variant *trans* ribozymes containing a variable number of base pairs in stem V [32]. We observed helix-length compensation between stem Ib and stem V, since the highest activity for each ribozyme variant was observed with a substrate that contains the compensatory number of base pairs in stem Ib (or one less). Moreover, this study establishes that VS ribozyme derivatives efficiently cleave SLI substrates that contain an alternate number of base pairs in stem Ib.

Given these encouraging results with the VS ribozyme, we were interested to gain further insights into the adaptability of this unique ribozyme for specific cleavage of folded RNA stem-loops. Previous studies have led to the idea that the I/V KLI of the VS ribozyme is an independent RNA module [13,16,18,31], and we hypothesized that this KLI can be substituted by other KLIs to create variant ribozymes that efficiently cleave stem-loop substrates with a modified terminal loop. In the present study, we developed a rational engineering approach to replace the I/V KLI of the *Neurospora* VS ribozyme by other KLIs and achieve efficient substrate cleavage (Figure 3.1C). This rational approach is supported by the recently available

NMR structure of an SLI/SLV complex [13] and the crystal structure of the complete VS ribozyme [14]. We first performed a bioinformatic search for KLIs in the protein data bank (PDB), which allowed us to identify two KLIs that share structural characteristics with the natural KLI of the VS ribozyme. Next, VS substrate/ribozyme (*S/R*) pairs, in which the I/V KLI was substituted by each of these two KLI candidates, were prepared using several combinations of helix length for stems Ib and V [32] in order to fine-tune the structural context of the surrogate KLIs necessary to obtain efficient cleavage of the modified substrates. The engineered *S/R* pairs were then kinetically characterized to identify the most efficient ones and structurally evaluated to validate the procedure. Interestingly, several novel *S/R* pairs were identified that display substantial catalytic activity, further illustrating the adaptability of the VS ribozyme architecture for cleavage of alternate stem-loop substrates.

## Material and Methods

**Search for KLIs with WebFR3D.** A bioinformatic search was performed with WebFR3D [33] to find KLIs in the PDB. The search query was defined as an interaction matrix between a set of 6 nt with no specific nucleotide identity (NT1 to NT6; Supplementary Figure S3.1B). Nucleotides NT1 and NT3 form a *cis* WC/WC base pair that closes the first loop, and this loop contains a non-self-pairing strand of 4–10 nt. Nucleotides NT4 and NT6 form a *cis*WC/WC base pair that closes the second loop, and this loop also contains a non-self-pairing strand of 4–10 nt. Nucleotides NT2 and NT5 are part of the first and second loops, respectively, and must form a *cis*WC/WC base pair. Finally, residues NT1-NT3 must be either sequentially located 5' respectively to residues NT4-NT6 or part of a different chain. For this search, we used the WebFR3D list of non-redundant RNA-containing structures determined by X-ray crystallography, Cryo-EM or NMR spectroscopy.

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

**Structural data retrieval and result listing.** A custom Python script was used to automatically gather information about each KLI. The size of each loop and the total number of *cis* WC/WC base pairs in the KLI were retrieved directly from the cluster of WebFR3D results and used to define the secondary structure of each KLI. Additional information was retrieved from the RCSB repository via the PDB ID of the KLI. The information gathered about each KLI structure was listed as an HTML page enriched with JavaScript (Supplementary Figure S3.1D and Table S3.1).

**Plasmids.** The Avapl ribozyme ( $R_0$ , Figure 3.1A) is a derivative of the AvaI ribozyme and its parental plasmid (pAvapl) was constructed by polymerase chain reaction from the A-3 clone [10]. The plasmids for transcription of the  $R_{TAR^*}$ ,  $R_{L22}$  and  $R_{L88}$  ribozyme variants (Figure 3.3A and Supplementary Figures S3.3A and 3.4A) were prepared by QuikChange II site-directed mutagenesis (Stratagene, CA, USA). All plasmids were validated by DNA sequencing.

**ITC studies.** Preparation of RNA and ITC studies were performed as previously described [18]. Relevant experimental details are provided in the legend of Supplementary Figure S3.2.

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

**Kinetic studies.** Multiple-substrate and single-substrate kinetic studies were carried out as single-turnover reactions at 37°C in 1× Kinetic Reaction buffer (50 mM Tris at pH 7.6, 25 mM KCl, 2 mM spermidine and 25 mM  $\text{MgCl}_2$ ), as previously described [32]. Briefly, multiple-substrate kinetic studies were typically performed with a mixture of 5'- $^{32}\text{P}$ -labeled nucleic acids containing all purified substrates for a given *S/R* pair (35 pM of each) and excess *R* (0.005–25

$\mu\text{M}$ ). These multiple-substrate kinetic studies were used to identify quickly the  $S/R$  pairs that lead to detectable substrate cleavage. All multiple-substrate kinetic experiments were performed at least three times.

Reactive  $S/R$  combinations were further investigated by single-substrate kinetic studies to derive the values of  $k_{\text{obs}}$  and  $k_{\text{cat}}/K_M$ . These studies were typically performed with  $[S] = 250$   $\mu\text{M}$  and excess  $R$  (0.005–15  $\mu\text{M}$ ). The percentage of remaining substrate [ $F = S_t/(S_t + P_t) * 100\%$ ; where  $S_t$  and  $P_t$  are the amount of substrate and product at time  $t$ ] was plotted against time. The resulting time courses were fit to the equation  $F = Ae^{-(k_{\text{obs}} * t)} + F_{\infty}$ , where  $A$  is the amplitude,  $k_{\text{obs}}$  is the rate of the reaction and  $F_{\infty}$  is a variable that represents the unreactive fraction of the substrate. The value of  $k_{\text{obs}}$ , the standard error for  $k_{\text{obs}}$  and the  $F_{\infty}$  parameter were estimated by non-linear regression analysis within the OriginPro 8.1 software (OriginLab, MA, USA). The  $F_{\infty}$  values are below 10% for all  $S/R$  combinations. A linear dependence of  $k_{\text{obs}}$  on  $[R]$  was observed for each  $S/R$  combination at the lowest values of  $[R]$  and analyzed by linear regression to derive the second order rate constant of the reaction,  $k_{\text{cat}}/K_M$ . The quality of the fit for the linear dependence of the  $k_{\text{obs}}$  on  $[R]$  was obtained from the square of the correlation coefficient ( $R^2$ ), and in all cases  $R^2$  was  $\geq 0.99$ . For the single-substrate kinetic experiments, three independent experiments were performed, and reported  $k_{\text{cat}}/K_M$  and their errors are average values and standard deviations, respectively, from these multiple measurements.

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

**SHAPE studies of ribozymes.** For SHAPE (Selective 2'-hydroxyl Acylation analyzed by Primer Extension) experiments, variant ribozymes (20 pmol) in 4  $\mu\text{l}$  of  $\text{H}_2\text{O}$  were first incubated at 37°C for 5 min with 5  $\mu\text{l}$  of 2 $\times$  Kinetic Reaction buffer and then treated with 1  $\mu\text{l}$  of 130 mM *N*-methylisatoic anhydride (NMIA; dissolved in anhydrous dimethyl sulfoxide). The

resulting mix was incubated for 54 min at 37°C and then transferred on ice. Reverse transcription reactions were performed on mutant ribozymes with 5'-<sup>32</sup>P-labeled primers and SuperScript III reverse transcriptase (ThermoFisher Scientific Inc.), as previously described [34]. The primer extension products were separated on an 8% denaturing polyacrylamide gel (7 M urea, 90 mM Tris/Borate, 2 mM EDTA). The gel was visualized by phosphorimager (Personal Molecular Imager™ System, Bio-Rad) and quantified using Image Lab Software 4.1.

**Three-dimensional modeling of S/R complexes with a substituted KLI.** Models of the  $S_0/R_0$ ,  $S_{L88-0}/R_{L22-0}$  and  $S_{TAR-0}/R_{TAR^*-0}$  complexes were constructed by fragment assembly in PyMOL (The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.) based on heavy-atom superpositions with the  $S_X/R_X$  complex obtained from the X-ray structure of the VS ribozyme [PDB ID: 4R4P; [14]] by taking the SLI substrate of one protomer ( $S_X$ ) and the *trans* VS ribozyme (helical domains II–VI) of the other protomer ( $R_X$ ). Starting with representative high-resolution structures of the studied KLIs [PDB ID: 2MI0—model 2 [13], 1KIS—model 2 [35] and 4IOA [36]], the hairpin stems within these complexes were modified by stems generated with MC-Sym [37] that match those of SLI (residues 623–626 and 634–637) and SLV (residues 688–694 and 702–708) from the  $S_0/R_0$  complex (Figure 3.1A), and the resulting hybrid structures retained the hairpin loop residues of the original high-resolution structures (2MI0: residues 4–17 and 104–118; 1KIS: residues 5–12 and 21–28; and 4IOA: residues 423–430 and 2383–2392). The hairpin stem corresponding to stem V within these hybrid structures was superposed to stem V of  $R_X$ . Subsequently,  $S_X$  was copied one time for each hybrid structure and its stem Ib was superposed to stem Ib of each hybrid structure. The final  $S/R$  complexes retain all residues from the hybrid structures, all residues of  $R_X$  except for residues 688–708 of SLV and only residues 616–622 and 638–643 of  $S_X$  (according to residue numbering in Figure 3.1).

## Results

**Search for KLIs in the PDB.** We searched for three-dimensional structures of KLIs in the PDB using a three-step computational workflow (Supplementary Figure S3.1A) that is based on the motif-search tool WebFR3D [33]. This online version of the FR3D program was chosen because it has the capacity to perform searches of RNA motifs composed of non-contiguous

nucleotides in a curated ensemble of non-redundant structures. In the first step, a search query was launched with WebFR3D to identify single *cis* Watson–Crick (WC) base pairings between any two residues belonging to two different terminal loops of 4–10 residues in length (Supplementary Figure S3.1B). This search used the WebFR3D list of non-redundant RNA-containing structures determined by X-ray crystallography, Cryo-EM or NMR spectroscopy. In the second step, the individual entries of the WebFR3D output were clustered to identify all existing WC pairings for a specific pair of terminal loops, and thereby provide a list of KLIs. In the third and final step, additional structural data and related information were retrieved for each KLI, and these data were listed in an interactive HTML/JavaScript table (Supplementary Figure S3.1D and Table S3.1). From this motif search of the PDB, 113 KLIs were identified and these form between one and ten *cis* WC/WC base pairs at the kissing-loop interface.

**Selection of KLIs that could substitute for the VS ribozyme I/V interaction.** To identify KLIs that could substitute for the VS ribozyme I/V KLI, we automatically filtered the list of 113 structures to eliminate all entries originating from X-ray and Cryo-EM structures with resolution  $\geq 3.6$  Å and only selected those that form heterodimeric KLI interactions with 3–10 *cis* WC/WC base pairs. Heterodimeric KLIs were selected because substitutions of SLI and SLV in the VS ribozyme *S/R* system (Figure 3.1A) by stem-loops that form an homodimeric KLI might lead to self-association of substrates and ribozymes, which could interfere with *trans* cleavage. In addition, KLIs with a minimum of three *cis* WC/WC base pairs were selected to prevent formation of an unstable KLI that could lead to inefficient *trans* cleavage [11,31]. This filtering step produced a list of 17 structures that represent 6 types of KLIs (Table 3.1): the TAR/TAR\* interaction derived from HIV-1; the L6/L7 and L22/L88 interactions from the large subunit of rRNAs; the L5/L13 interaction from the cobalamin riboswitch; the L3/L4 interaction from the human SRP Alu domain; and the Cole1 KLI. To identify suitable KLIs for our engineering studies, we selected those that share structural similarities with the VS ribozyme I/V KLI. More specifically, a suitable KLI must fit certain structural characteristics: (i) form consecutive *cis* WC/WC base pair with non-modified nucleotides, (ii) the adjoining stems must contain at least three consecutive *cis* WC/WC base pairs and (iii) the interhelical angle must fall between 125° and 175° [13,38]. Such structural criteria limited our choice to two classes of KLIs: the HIV-1 TAR/TAR\* interaction and the



L22/L88 (loop 22/loop 88) interaction from the large subunit of rRNAs, and those with PDB ID codes 1KIS [35] and 4IOA [36] were selected for our engineering studies.

**Thermodynamic stability of the selected KLIs.** To ascertain that the two selected KLIs are good candidates for our engineering studies, we verified their thermodynamic stability. Although the TAR/TAR\* and L22/L88 KLIs incorporate a larger number of WC base pairs than the VS ribozyme I/V KLI (Figures 3.1B and 3.2A and B), the latter is known to be exceptionally stable [18]. Based on ITC studies, the  $K_d$  of the TAR/TAR\* and L22/L88 complexes are 40-fold lower and 16-fold higher, respectively, than that of the SLI/SLV complex (Figure 3.2C and Supplementary Figure S3.2). These results confirm that the thermodynamic stability of the two selected KLIs is not considerably different from that of the I/V KLI. Thus, these two KLIs were considered suitable for our engineering studies.

**Kinetic characterization of VS ribozyme S/R pairs with a modified I/V KLI.** Next, kinetic studies were performed to test the hypothesis that the two selected KLIs can substitute for the natural I/V KLI interaction and yield active VS ribozyme variants. For these studies, we used the previously-characterized  $S_0/R_0$  pair formed between a pre-shifted SLI substrate and a *trans* VS ribozyme [32]. Since the selected KLIs may functionally replace the I/V KLI only when substituted in the proper orientation (e.g. either the L22 or the L88 loop replacing loop I) and with supporting helices of proper lengths, we prepared three combinatorial pools of  $S/R$  pairs that consist of 25  $S_{\text{TAR}}/R_{\text{TAR}^*}$  pairs, 30  $S_{\text{L88}}/R_{\text{L22}}$  pairs and 36  $S_{\text{L22}}/R_{\text{L88}}$  pairs (Figure 3.3A and Supplementary Figures S3.3A and 4A). Within each pool, helix-length variations were introduced in both stem Ib of the variant substrates ( $S_{\text{TAR}}$ ,  $S_{\text{L88}}$  and  $S_{\text{L22}}$ ) and stem V of the variant ribozymes ( $R_{\text{TAR}^*}$ ,  $R_{\text{L22}}$  and  $R_{\text{L88}}$ ). In this context, compensatory  $S/R$  pairs are those for which the total number of base pairs in stems Ib and V is identical to that found for the  $S_0/R_0$  reference pair (Figure 3.1). For all  $S/R$  combinations, the second-order rate constant  $k_{\text{cat}}/K_M$  was determined under single turnover conditions. It is important to note that the catalytic parameter  $k_{\text{cat}}/K_M$  is very useful when comparing the relative activity of a given enzyme for different competing substrates (e.g.  $R_{\text{TAR}^*-0}$  acting on the different  $S_{\text{TAR}}$  substrates), but may be misleading when comparing the relative activity of different enzymes [39], particularly those of different  $S/R$  families. Thus, to compare the catalytic activity of

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

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

For  $R_{\text{TAR}^*-0}$ , efficient substrate cleavage is obtained with  $S_{\text{TAR}-1}$  as well as with  $S_{\text{TAR}-0}$ , indicating limited substrate promiscuity. Other  $R_{\text{TAR}^*}$  ribozymes also display similar specificity, cleaving efficiently only two to three out of the five  $S_{\text{TAR}}$  substrates. Given that the highest  $k_{\text{cat}}/K_{\text{M}}$  values for each ribozyme is observed with the compensatory substrate or one that varies by one or two base pairs, it can be concluded that efficient substrate cleavage depends on helix-length compensation in stem V of the variant ribozyme. For example, the optimal ribozyme for  $S_{\text{TAR}-1}$  is  $R_{\text{TAR}^*-0}$ , whereas the optimal one for  $S_{\text{TAR}+2}$  is  $R_{\text{TAR}^*-3}$ . Thus, for the  $S_{\text{TAR}}/R_{\text{TAR}^*}$  family, there is limited substrate promiscuity for all ribozymes as well as helix-length compensation across the family, as previously observed for the reference  $S/R$  family [32].

**Kinetic studies of the VS ribozyme  $S_{\text{L88}}/R_{\text{L22}}$  and  $S_{\text{L22}}/R_{\text{L88}}$  pairs.** In a similar manner, multiple and single turnover kinetic studies were performed to evaluate the  $k_{\text{cat}}/K_{\text{M}}$  of the 30  $S_{\text{L88}}/R_{\text{L22}}$  and 36  $S_{\text{L22}}/R_{\text{L88}}$  combinations (Supplementary Figures S3.3 and 4). For

the  $S_{L88}/R_{L22}$  pairs, the most active combination is  $S_{L88+1}/R_{L22-1}$  with a  $k_{cat}/K_M$  of  $0.39 \pm 0.03 \text{ min}^{-1} \mu\text{M}^{-1}$ , which is 63-fold lower than that of  $S_0/R_0$ . A few other  $S_{L88}/R_{L22}$  combinations produce similar  $k_{cat}/K_M$  (within 5-fold; dark gray shading in Supplementary Figure S3.3B), and these combinations involve three of the five  $S_{L88}$  substrates and three of the six  $R_{L22}$  ribozymes. Each of these three ribozymes displays similar cleavage activity for one or two  $S_{L88}$  substrates, and helix-length compensation is observed given that the most active  $S_{L88}/R_{L22}$  combinations for these ribozymes are the  $S_{L88-0}/R_{L22-0}$ ,  $S_{L88+1}/R_{L22-1}$  and  $S_{L88+2}/R_{L22-2}$  pairs. For the  $S_{L22}/R_{L88}$  pairs, the highest  $k_{cat}/K_M$  obtained is  $0.23 \pm 0.02 \text{ min}^{-1} \mu\text{M}^{-1}$  for  $S_{L22+1}/R_{L88-1}$ , which is 107-fold lower than that of  $S_0/R_0$ . In contrast to the  $S_{L88}/R_{L22}$  combinations, several  $S_{L22}/R_{L88}$  combinations display similar catalytic activity (within 5-fold; dark gray shading in Supplementary Figure S3.4B), and these combinations involve four of the six  $S_{L22}$  substrates and five of the six  $R_{L88}$  ribozymes. For each of the five active ribozymes, comparable  $k_{cat}/K_M$  values are obtained with one to four of the six  $S_{L88}$  substrates, indicating different degrees of substrate promiscuity. Furthermore, helix-length compensation is observed for all variant ribozymes given that the highest  $k_{cat}/K_M$  values for each ribozyme is observed with the compensatory substrate or one that varies by one base pair. For example, the optimal substrate for  $R_{L88+1}$  is  $S_{L22+0}$ , whereas the optimal one for  $R_{L88-4}$  is  $S_{L22+4}$ . Thus, for both the  $S_{L88}/R_{L22}$  and  $S_{L22}/R_{L88}$  families, there is limited substrate promiscuity for all ribozymes as well as helix-length compensation across the family, as previously observed for the reference  $S/R$  family [32].

**Comparison of the catalytic activity of the different S/R families.** Given that  $k_{cat}/K_M$  values can be misleading when comparing the relative activity of the various  $S/R$  families, we also compared the  $k_{obs}$  values for the  $S/R$  pair of each family that displays the highest catalytic activity. These  $k_{obs}$  values, determined under identical single turnover conditions, indicate that the activities of the  $S_{TAR}/R_{TAR*}$ ,  $S_{L88}/R_{L22}$  and  $S_{L22}/R_{L88}$  family representatives are reduced by 159-, 53- and 84-fold, respectively, compared to the parental  $S_0/R_0$  pair (Supplementary Table S3.2). Thus, although both the  $k_{cat}/K_M$  and  $k_{obs}$  values indicate that there is a significant level of catalytic activity for some variant  $S/R$  pairs, their activity is still lower than that of the parental  $S_0/R_0$  pair.

**Structural characterization of the variant substrates.** To insure that the variant substrates examined in this study are properly folded, these substrates were analyzed by native gel electrophoresis. The migration of variant substrates on native gels was compared with that of two previously investigated VS ribozyme substrates [32],  $S_{-1}$  bp and  $S_{-2}$  bp, which respectively adopt a hairpin and a duplex conformation. Our results show that all variant substrates migrate on a native gel in a manner similar to the active hairpin conformation (Supplementary Figure S3.5).

**Structural characterization of the variant ribozymes.** To verify the proper folding of variant ribozymes, the secondary and tertiary structures of these *trans*ribozymes were analyzed by SHAPE experiments. In these studies, the reactivity of each nucleotide to NMIA is quantified and interpreted as being directly linked to the local nucleotide flexibility [34]. For the  $R_{TAR^*-0}$  ribozyme (Figure 3.4), the nucleotides reactive to NMIA are found in non-helical regions that are likely to be dynamic, including the  $A_{718}$  and  $A_{725}A_{726}$  bulges, the  $A_{756}$  loop and the terminal loops of stem-loops IV, V and VI [40–43]. The reactivity patterns of the other  $R_{TAR^*}$  ribozymes are very similar to that of the  $R_{TAR^*-0}$  ribozyme (Figure 3.4). In addition, the reactivity patterns of the  $R_0$ ,  $R_{L22}$  and  $R_{L88}$  ribozymes are similar to those of the  $R_{TAR^*}$  ribozymes in that known helical regions display lower reactivity to NMIA compared to non-helical regions (Supplementary Figure S3.6). As an exception, residues within and adjacent to G-U base pairs (in stem-loops V and VI) produce idiosyncratic patterns of NMIA reactivity, as previously demonstrated [44]. Of note, the most NMIA-reactive nucleotide in all ribozymes is  $A_{726}$ , an extruded residue from the II-VI bulge-bulge interaction and the most disordered residue in the NMR structures of the II-III-VI junction [14,43,45]. Overall, the SHAPE results for all of the engineered ribozymes are in good agreement with the structural schematics of the  $R_0$  ribozyme derived from NMR and X-ray structures [Figure 3.1; [14,43]]. Thus, all of the variant VS ribozymes investigated here adopt a ground state conformation that is compatible with cleavage activity.

**Structural modeling of the  $S_0/R_0$ ,  $S_{L88-0}/R_{L22-0}$  and  $S_{TAR-0}/R_{TAR^*-0}$  pairs.** Once a stable KLI is formed between a well-folded VS ribozyme variant and its stem-loop substrate, the  $G_{638}$  internal loop of the substrate must properly interact with the  $A_{756}$  loop of the ribozyme to allow formation of a productive active site. In order to assess the effect of substituting the I/V

KLIs on formation of the VS ribozyme active site, three-dimensional models of the  $S_{L88-0}/R_{L22-0}$  and  $S_{TAR-0}/R_{TAR*-0}$  pairs were produced. These models were built by fragment assembly, combining either the X-ray structure of the L22/L88 KLI or the NMR structure of the TAR/TAR\* KLI with the X-ray structure of the VS ribozyme [Figure 3.5; [14,35,36]]. As a control, the model of the  $S_0/R_0$  pair was similarly built from the NMR structure of the SLI/SLV complex. In contrast with the close state observed in the X-ray structure (Figure 3.5B), the model of the  $S_0/R_0$  pair (Figure 3.5C) adopts an open state with the G<sub>638</sub> loop of SLI apart from the A<sub>756</sub> loop of SLVI. This is not surprising, since the open state, previously described in the NMR-based model of the complete VS ribozyme, likely represents a ground-state conformation of the  $S/R$  complex [43]. In the  $S_0/R_0$  model, the scissile phosphate within the G<sub>638</sub> loop is 26.6 Å from the corresponding phosphate in the crystal structure, but oriented directly toward A<sub>756</sub>, such that a simple repositioning of the helical domains containing the G<sub>638</sub> and A<sub>756</sub> loops would allow formation of the active site. The models of both the  $S_{L88-0}/R_{L22-0}$  (Figure 3.5D) and  $S_{TAR-0}/R_{TAR*-0}$  pairs (Figure 3.5E) also represent an open state of the ribozyme with their scissile phosphates respectively 23.0 Å and 31.9 Å from the corresponding phosphate within the crystal structure. However, in both these models, the scissile phosphate is not oriented favorably with respect to the A<sub>756</sub> loop to facilitate formation of the active site. Overall, these models suggest that substitutions of the KLI in the variant  $S/R$  pairs may reduce the probability of encounter between the G<sub>638</sub> internal loop of the substrate and the A<sub>756</sub> loop of the ribozyme and thus likely impair both the formation of the active site and the subsequent cleavage reaction.

## Discussion

In this work, a comprehensive RNA engineering approach was developed, which demonstrates that the I/V KLI of the VS ribozyme is an RNA module that can be substituted by other KLIs while maintaining efficient substrate cleavage. Several KLIs were initially identified through a bioinformatics search of the PDB, of which two were selected as being structurally similar to the I/V KLI of the VS ribozyme: the HIV-1 TAR/TAR\* KLI [PDB ID: 1KIS; [35]] and the L22/L88 KLI from the large subunit of rRNAs [PDB ID: 4IOA; [36]]. These two KLIs were substituted for the I/V KLI of the  $S_0/R_0$  reference pair formed between a pre-shifted substrate and a *trans* version of the VS ribozyme. In addition, a combinatorial approach was

used in which several variant  $S/R$  pairs were synthesized that contain different helix lengths in stems Ib and V. The resulting  $S/R$  pairs were kinetically evaluated, and several show substantial catalytic activity. Below, we compare the catalytic properties of the most active variant  $S/R$  pairs with that of the previously investigated  $S_0/R_0$  pair in light of the thermodynamic and structural properties of the parental and surrogate KLIs.

Within the three families of  $S/R$  pairs investigated ( $S_{\text{TAR}}/R_{\text{TAR}^*}$ ,  $S_{\text{L88}}/R_{\text{L22}}$  and  $S_{\text{L22}}/R_{\text{L88}}$ ), several  $S/R$  combinations show substantial catalytic efficiency (dark gray shading in Figure 3.3B and in Supplementary Figures S3.3B and 4B). For each family, helix-length compensation was observed for the most active  $S/R$  pairs, and this is consistent with what was observed in previous studies of the parental  $S_0/R_0$  pair [32]. For the  $S_{\text{TAR}}/R_{\text{TAR}^*}$  family, the *trans* VS ribozyme variants can efficiently cleave an SLI substrate with stem Ib lengths of 3–6 bp (Figure 3.3B). Both the  $S_{\text{L88}}/R_{\text{L22}}$  and  $S_{\text{L22}}/R_{\text{L88}}$  families also allow an appreciable level of catalytic efficiency for several SLI substrates, with a stem Ib length of 4–7 bp for  $S_{\text{L22}}/R_{\text{L88}}$  and 4–6 bp for  $S_{\text{L88}}/R_{\text{L22}}$  (Supplementary Figures S3.3 and 4). Within the three families of  $S/R$  pairs investigated, the  $S_{\text{TAR}-1}/R_{\text{TAR}^*-0}$ ,  $S_{\text{L88}+1}/R_{\text{L22}-1}$  and  $S_{\text{L22}+1}/R_{\text{L88}-1}$  show the highest  $k_{\text{cat}}/K_{\text{M}}$  values, all within 2.5-fold of each other ( $0.2\text{--}0.5 \text{ min}^{-1}\mu\text{M}^{-1}$ ). Although these  $S/R$  pairs with surrogate KLIs behave like the reference  $S_0/R_0$  pair in some respects, their  $k_{\text{cat}}/K_{\text{M}}$  values are reduced by 50- to 100-fold compared to  $S_0/R_0$ . Similarly, their  $k_{\text{obs}}$  values determined under identical conditions are reduced 50- to 160- fold compared to  $S_0/R_0$ . The KLIs most likely form and assist catalysis for the highly-active variant  $S/R$  pairs, given that several other variant  $S/R$  pairs within the same family are inactive and that disruption of the KLI in the parental  $S_0/R_0$  pair can lead to more than a 1,000-fold decrease in  $k_{\text{cat}}/K_{\text{M}}$  and more than a 6000-fold decrease in  $k_{\text{obs}}$  [32]. In addition, we verified by native gel electrophoresis and SHAPE analysis that all substrates and ribozymes of the variant  $S/R$  pairs adopt a stable fold equivalent to  $S_0$  and  $R_0$ , respectively, and compatible with high catalytic efficiency. Thus, although surrogate KLIs allow high catalytic efficiency for some variant  $S/R$  pairs, it is puzzling that they do not allow the same high level of activity as that observed for the parental  $S_0/R_0$  pair.

We have previously shown that mutations that decrease the affinity of the I/V KLI also lead to lower  $k_{\text{cat}}/K_{\text{M}}$  values when monitoring the *trans* cleavage reaction as done here [18,31]. For example, deletion of U700 in SLV causes a 210-fold decrease in binding affinity for SLI

and 140-fold decrease in  $k_{\text{cat}}/K_M$  [18,31]. Thus, it is likely that the lower relative affinity of a surrogate KLI may affect the level of catalytic efficiency that could be attained in the variant  $S/R$  pairs. Given that the L22/L88 KLI is 16-fold less stable than the I/V KLI [Figure 3.2C; [18]], its higher  $K_d$  could explain in part the lower  $k_{\text{cat}}/K_M$  of the  $S_{L88}/R_{L22}$  (>60-fold) and  $S_{L22}/R_{L88}$  pairs (>100-fold) compared to  $S_0/R_0$ . However, given that the HIV-1 TAR/TAR\* KLI is 40-fold more stable than the I/V KLI [Figure 3.2C; [18]], the difference in  $K_d$  of the KLI complex cannot explain the lower  $k_{\text{cat}}/K_M$  of the  $S_{\text{TAR}}/R_{\text{TAR}^*}$  pairs (>50-fold) compared to  $S_0/R_0$ .

The VS ribozyme  $S/R$  pairs likely adopt both an open state and a close state that can be respectively represented by the recent NMR-based model and X-ray structure [14,43]. The KLI is stably formed in both states, but only the close state captures the active site formed by the association of the G<sub>638</sub> and A<sub>756</sub> loops. In order to assess the effect of substituting the I/V KLI on formation of the VS ribozyme active site, the  $S_0/R_0$ ,  $S_{L88-0}/R_{L22-0}$  and  $S_{\text{TAR-0}}/R_{\text{TAR}^*-0}$  pairs were modeled based on high-resolution structures of KLIs and the  $S/R$  complex derived from the X-ray structure of the VS ribozyme (Figure 3.5). Since all the modeled  $S/R$  pairs adopt an open state with the G<sub>638</sub> loop apart from the A<sub>756</sub> loop, it is difficult to precisely evaluate the compatibility of surrogate KLIs for formation of the active site. Nevertheless, we find that the distance between the scissile phosphates of the model and the crystal structure is longer for the  $S_{\text{TAR-0}}/R_{\text{TAR}^*-0}$  pair than for the  $S_0/R_0$  and  $S_{L88-0}/R_{L22-0}$  pairs. In addition, it appears from these models that formation of the active site from the open state would involve the simple packing of two helical domains for the  $S_0/R_0$  pair, but possibly be more restrained for the  $S_{L88-0}/R_{L22-0}$  and  $S_{\text{TAR-0}}/R_{\text{TAR}^*-0}$  pairs. Thus, the HIV-1 TAR/TAR\* KLI and possibly the L22/L88 interaction, which both contain more base pairs than the I/V KLI, may partially restrict formation of the active site due to topological constraints. We envision that additional mutations of the VS ribozyme could be introduced to help compensate for such structural restriction and thereby increase the catalytic efficiency of the variant  $S_{L88}/R_{L22}$  and  $S_{\text{TAR}}/R_{\text{TAR}^*}$  pairs.

Previous NMR studies of an SLI/SLV complex indicate that both the SLI and SLV loops contain several dynamic residues that could confer a certain degree of flexibility to the I/V KLI [13]. Thus, the I/V KLI likely acts as a dynamic hinge that facilitates formation of the active site. It is not clear at this time if the surrogate KLIs allow the same degree of flexibility, and

how their dynamics affect formation of the active site in the variant *S/R* pairs. In this regard, future studies investigating the dynamic behavior of these KLIs would provide valuable insights into how these KLIs can be further engineered to more efficiently assist formation of the active site in the VS ribozyme.

## References

1. Lilley, D.M., *Catalysis by the nucleolytic ribozymes*. Biochem Soc Trans, 2011. **39**(2): p. 641-6.
2. Jimenez, R.M., J.A. Polanco, and A. Luptak, *Chemistry and Biology of Self-Cleaving Ribozymes*. Trends Biochem Sci, 2015. **40**(11): p. 648-661.
3. Weinberg, Z., et al., *New classes of self-cleaving ribozymes revealed by comparative genomics analysis*. Nat Chem Biol, 2015. **11**(8): p. 606-10.
4. Harris, K.A., et al., *Biochemical analysis of pistol self-cleaving ribozymes*. RNA, 2015. **21**(11): p. 1852-8.
5. Li, S., et al., *Biochemical analysis of hatchet self-cleaving ribozymes*. RNA, 2015. **21**(11): p. 1845-51.
6. Herschlag, D., *Learning from ribozymes*. RNA, 2015. **21**(4): p. 527-8.
7. Collins, R.A., *The Neurospora Varkud satellite ribozyme*. Biochem Soc Trans, 2002. **30**(Pt 6): p. 1122-6.
8. Lilley, D.M., *The Varkud satellite ribozyme*. RNA, 2004. **10**(2): p. 151-8.
9. Wilson, T.J. and D.M. Lilley, *A mechanistic comparison of the Varkud satellite and hairpin ribozymes*. Prog Mol Biol Transl Sci, 2013. **120**: p. 93-121.
10. Guo, H.C. 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.
11. Rastogi, T., et al., *A long-range pseudoknot is required for activity of the Neurospora VS ribozyme*. EMBO J, 1996. **15**(11): p. 2820-2825.
12. Leontis, N.B. and E. Westhof, *Geometric nomenclature and classification of RNA base pairs*. RNA, 2001. **7**(4): p. 499-512.
13. 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.

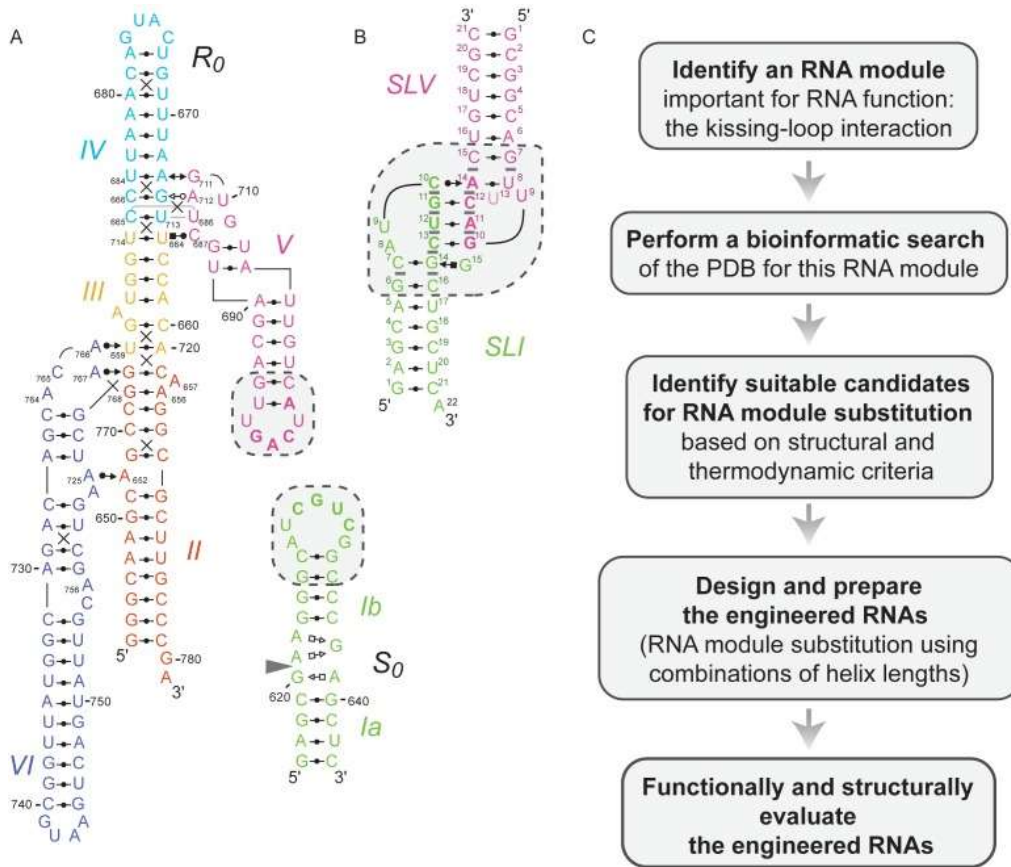


14. Suslov, N.B., et al., *Crystal structure of the Varkud satellite ribozyme*. Nat Chem Biol, 2015. **11**(11): p. 840-6.
15. Andersen, A.A. and R.A. Collins, *Rearrangement of a stable RNA secondary structure during VS ribozyme catalysis*. Mol Cell, 2000. **5**(3): p. 469-78.
16. Andersen, A.A. and R.A. Collins, *Intramolecular secondary structure rearrangement by the kissing interaction of the Neurospora VS ribozyme*. Proc Natl Acad Sci U S A, 2001. **98**(14): p. 7730-5.
17. 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.
18. Bouchard, P. and P. Legault, *A remarkably stable kissing-loop interaction defines substrate recognition by the Neurospora Varkud Satellite ribozyme*. RNA, 2014. **20**(9): p. 1451-64.
19. 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.
20. 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.
21. 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.
22. Zhao, Z.Y., et al., *Nucleobase participation in ribozyme catalysis*. J Am Chem Soc, 2005. **127**(14): p. 5026-7.
23. 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 U S A, 2007. **104**(14): p. 5818-23.
24. 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.
25. 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.
26. Smith, M.D., et al., *The ionic environment determines ribozyme cleavage rate by modulation of nucleobase pK<sub>a</sub>*. RNA, 2008. **14**(9): p. 1942-9.

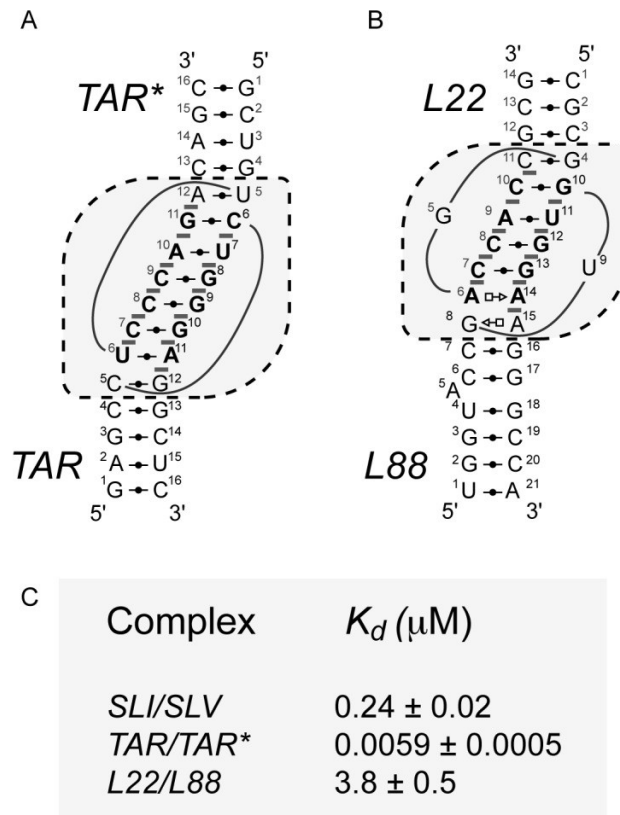
27. Wilson, T.J., et al., *Nucleobase-mediated general acid-base catalysis in the Varkud satellite ribozyme*. Proc Natl Acad Sci U S A, 2010. **107**(26): p. 11751-6.
28. Guo, H.C., et al., *Nucleotide sequence requirements for self-cleavage of Neurospora VS RNA*. J Mol Biol, 1993. **232**(2): p. 351-61.
29. Beattie, T.L., J.E. Olive, and R.A. Collins, *A secondary-structure model for the self-cleaving region of Neurospora VS RNA*. Proc Natl Acad Sci U S A, 1995. **92**(10): p. 4686-90.
30. 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.
31. Bouchard, P., et al., *Role of SLV in SLI substrate recognition by the Neurospora VS ribozyme*. RNA, 2008. **14**(4): p. 736-48.
32. 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.
33. Petrov, A.I., C.L. Zirbel, and N.B. Leontis, *WebFR3D--a server for finding, aligning and analyzing recurrent RNA 3D motifs*. Nucleic Acids Res, 2011. **39**(Web Server issue): p. W50-5.
34. Merino, E.J., et al., *RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE)*. J Am Chem Soc, 2005. **127**(12): p. 4223-31.
35. 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.
36. Magee, T.V., et al., *Novel 3-O-carbamoyl erythromycin A derivatives (carbamolides) with activity against resistant staphylococcal and streptococcal isolates*. Bioorg Med Chem Lett, 2013. **23**(6): p. 1727-31.
37. Major, F., et al., *The combination of symbolic and numerical computation for three-dimensional modeling of RNA*. Science, 1991. **253**(5025): p. 1255-60.
38. Lavery, R., et al., *Conformational analysis of nucleic acids revisited: Curves+*. Nucleic Acids Res, 2009. **37**(17): p. 5917-29.
39. Eienthal, R., M.J. Danson, and D.W. Hough, *Catalytic efficiency and kcat/KM: a useful comparator?* Trends Biotechnol, 2007. **25**(6): p. 247-9.

40. 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.
41. 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.
42. Bonneau, E. and P. Legault, *Nuclear magnetic resonance structure of the III-IV-V three-way junction from the Varkud satellite ribozyme and identification of magnesium-binding sites using paramagnetic relaxation enhancement*. *Biochemistry*, 2014. **53**(39): p. 6264-75.
43. Bonneau, E., et al., *The NMR structure of the II-III-VI three-way junction from the Neurospora VS ribozyme reveals a critical tertiary interaction and provides new insights into the global ribozyme structure*. *RNA*, 2015. **21**(9): p. 1621-32.
44. Wilkinson, K.A., et al., *Influence of nucleotide identity on ribose 2'-hydroxyl reactivity in RNA*. *RNA*, 2009. **15**(7): p. 1314-21.
45. McGinnis, J.L., et al., *The mechanisms of RNA SHAPE chemistry*. *J Am Chem Soc*, 2012. **134**(15): p. 6617-24.

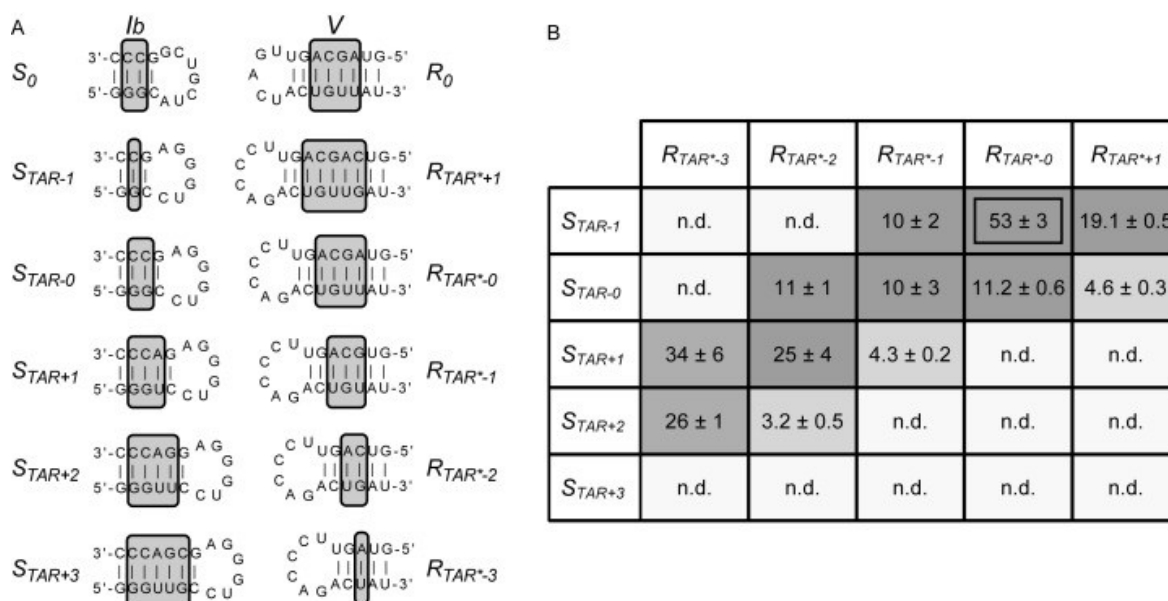
## Figures



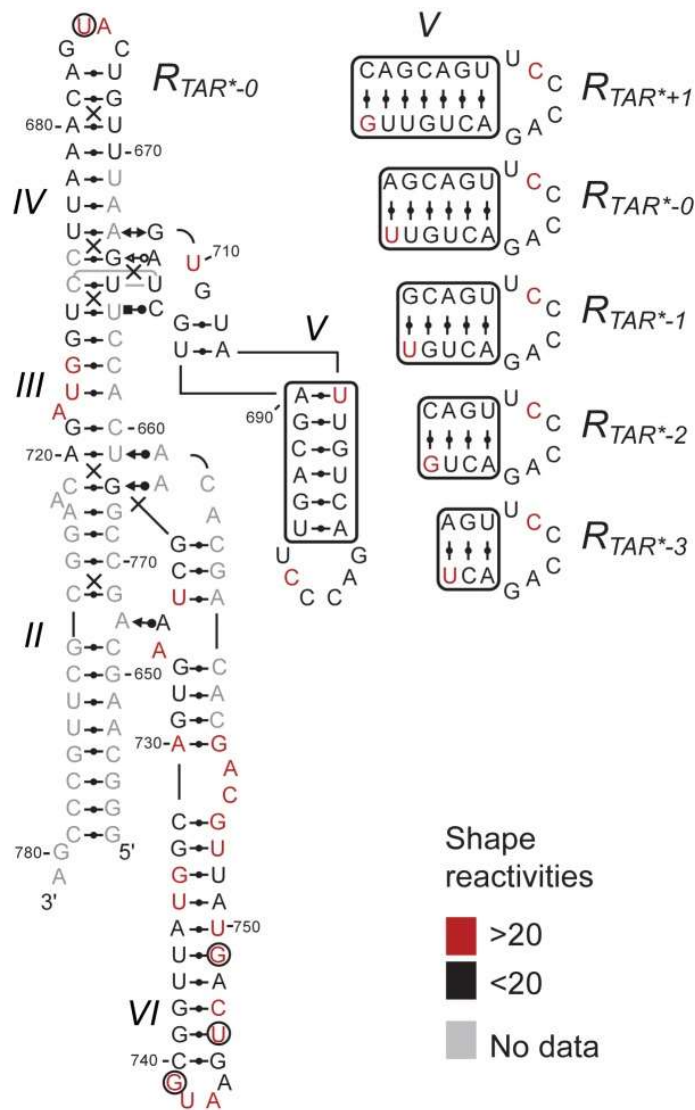
**Figure 3.1. Rational engineering of KLIs for substrate recognition by the *Neurospora* VS ribozyme.** (A) Structural schematic of a *trans* VS ribozyme system formed by the  $S_0$  substrate and the  $R_0$  ribozyme [32]. The schematic is derived from NMR and X-ray structures [14,43]. The cleavage site is located between G620 and A621 as indicated by the gray arrowhead. (B) Structural characteristics of the VS ribozyme I/V KLI derived from the NMR structure of the SLI/SLV complex [PDB ID: 2MI0; [13]]. (C) Flowchart of the engineering strategy used in this study. In (A and B), the dashed boxes enclose residues that contribute to the KLI and the closing base pairs of the two stems, whereas base-pairing interactions are represented using the Leontis-Westhof notation [12]. In addition, base-pairing residues at the KLI are in bold. In (B), gray bars represent base stacking at the KLI.



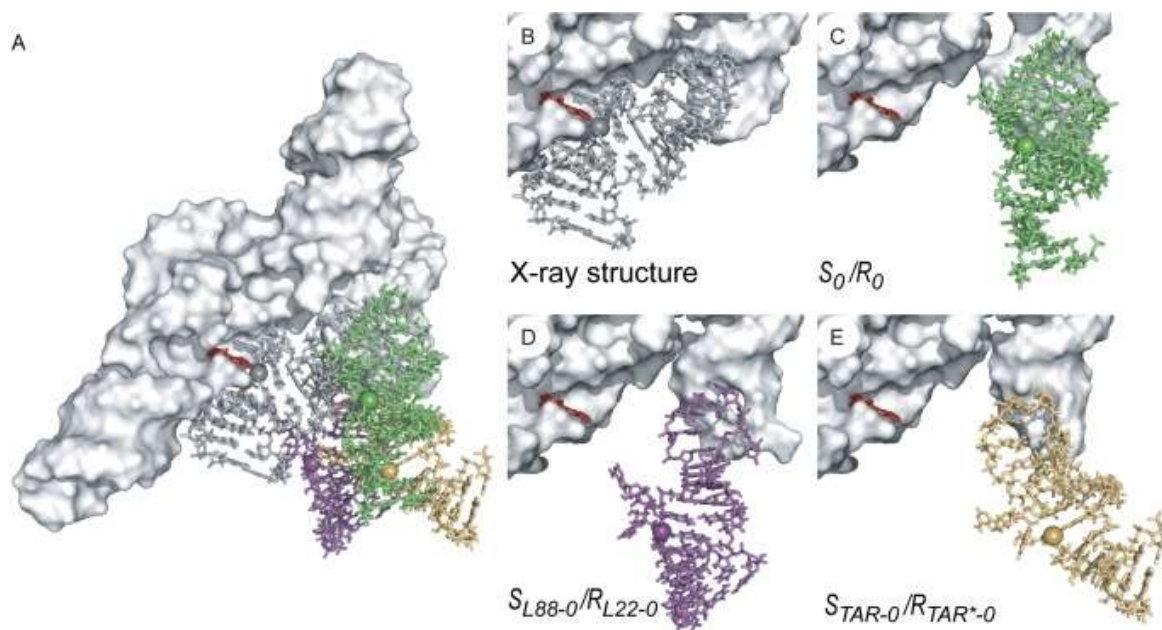
**Figure 3.2. Schematics and thermodynamic stability of selected KLIs.** Structural characteristics of the selected KLIs: (A) the HIV-1 TAR/TAR\* KLI [PDB ID: 1KIS; [35]] and (B) the large ribosomal subunit L22/L88 KLI from *Deinococcus radiodurans* [PDB ID: 4IOA; [36]]. (C) Dissociation constants ( $K_d$ ) of the selected KLIs determined in a previous study for the SLI/SLV and TAR/TAR\* complexes [18] and here for the L22/L88 complex (Supplementary Figure S2). These  $K_d$  values were all determined at 25°C in 10 mM Tris pH 7.0 and 20 mM MgCl<sub>2</sub>. In (A and B), the dashed boxes enclose residues of the KLI and the closing base pairs of the two stems and base-pairing interactions are represented using the Leontis-Westhof notation [12]. In addition, base-pairing residues at the KLI are in bold and gray bars highlight base stacking at the KLI.



**Figure 3.3. Sequences and kinetic data of the  $S_{TAR}/R_{TAR^*}$  variants.** (A) Sequences of the  $S_{TAR}/R_{TAR^*}$  variants. Only the regions of the  $S_0$  substrate and  $R_0$  ribozyme (stem-loops Ib and V shown above; Figure 3.1) that were substituted in the  $S_{TAR}$  and  $R_{TAR^*}$  variants are shown. The number of base pairs varies in stems Ib and V, as indicated by the gray boxes on the putative secondary structures. (B) Grayscale heat map of the  $k_{cat}/K_M$  values ( $10^{-2} \text{ min}^{-1} \mu\text{M}^{-1}$ ) for the  $S_{TAR}/R_{TAR^*}$  variants. Dark gray shading corresponds to  $k_{cat}/K_M$  values within 5-fold of the most active  $S_{TAR}/R_{TAR^*}$  pair ( $S_{TAR-1}/R_{TAR^*-0}$ ); medium gray shading corresponds to lower  $k_{cat}/K_M$  values (5–17-fold lower than the most active  $S_{TAR}/R_{TAR^*}$  pair), and light gray shading reflects the absence of detectable cleavage activity (n.d.:  $k_{obs} \leq 0.002 \text{ min}^{-1}$ ) after 24 min at  $[R] \leq 15 \mu\text{M}$ .



**Figure 3.4. SHAPE analysis of the  $R_{TAR^*}$  ribozyme variants.** The normalized NMIA reactivity of each nucleotide within  $R_{TAR^*-0}$  is color-coded on its structural schematic as per the SHAPE reactivities key. For the other  $R_{TAR^*}$  variants, the normalized NMIA reactivity is within the same category as for  $R_{TAR^*-0}$ , except for a few nucleotides in the variable SLV region (right panel) and outside this region (black circles indicate inverted reactivity compared to  $R_{TAR^*-0}$ ; U<sub>675</sub> for  $R_{TAR^*-1}$  and G<sub>741</sub>, U<sub>746</sub> and G<sub>749</sub> for  $R_{TAR^*-3}$ ).



**Figure 3.5. Three-dimensional models of substrate/ribozyme complexes.** (A) Superposition of three-dimensional models of the variant  $S/R$  pairs on the X-ray structure of the VS ribozyme. Individual representations from this superposition are zoomed in and described in details for the subsequent panels. (B) X-ray structure of the VS ribozyme [PDB ID: 4R4P; [14]]. In this representation, the substrate of one protomer is shown in dark gray sticks with the phosphorus at the scissile phosphate as a sphere and the *trans* VS ribozyme (helical domains II–VI) of the other protomer is shown as a white surface, except for residue A<sub>756</sub>, which is shown as red sticks. (C) Model of the  $S_0/R_0$  pair derived from the NMR structure of a VS ribozyme SLI/SLV complex [lime green; PDB ID: 2MI0; [13]]. (D) Model of the  $S_{L88-0}/R_{L22-0}$  pair derived from the X-ray structure of the ribosomal L22/L88 KLI [purple; PDB ID: 4IOA; [36]]. (E) Model of the  $S_{TAR-0}/R_{TAR^*-0}$  pair derived from the NMR structure of the TAR/TAR\* complex [mustard yellow; PDB ID: 1KIS; [35]]. In (A–E), the *trans* VS ribozyme is shown as a white surface with A<sub>756</sub> in red sticks and the individual substrates are shown as sticks with the phosphorus at the scissile phosphate as a sphere.



## Tables

**Table 3.1. Filtered results of the bioinformatics search to identify KLIs from the PDB.**

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

<sup>a</sup>Each structure was analyzed with the PyMOL Molecular Graphics System (Version 1.3, Schrodinger) to identify the number of *cis* WC/WC base pair and size of loops at the KLI and with Curves+ [38] to measure the interhelical angle between the stems connecting a specific KLI. Interhelical angles were measured only for the KLIs where the adjoining stems contain a minimum of three base pairs and are not available (n/a) for the other KLIs.

<sup>b</sup>Contains modified nucleotides at the KLI.

## Supporting Information

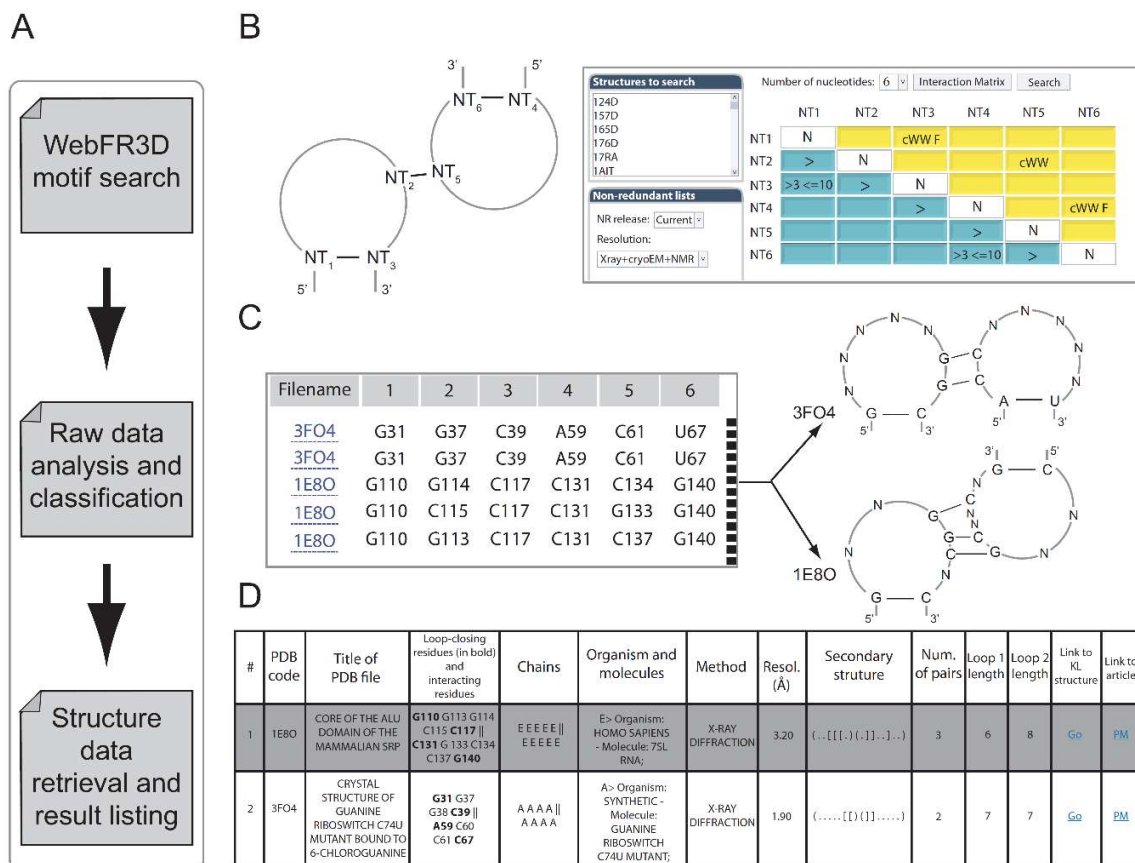
**Table S3.1. Values of  $k_{obs}$  ( $\text{min}^{-1}$ ) for selected  $S/R$  pairs measured with  $[S] = 0.250 \text{ nM}$  and  $[R] = 0.5 \text{ }\mu\text{M}$  under standard conditions.<sup>a</sup>**

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

a. At 37 °C in Kinetic Reaction buffer (50 mM Tris at pH 7.6,

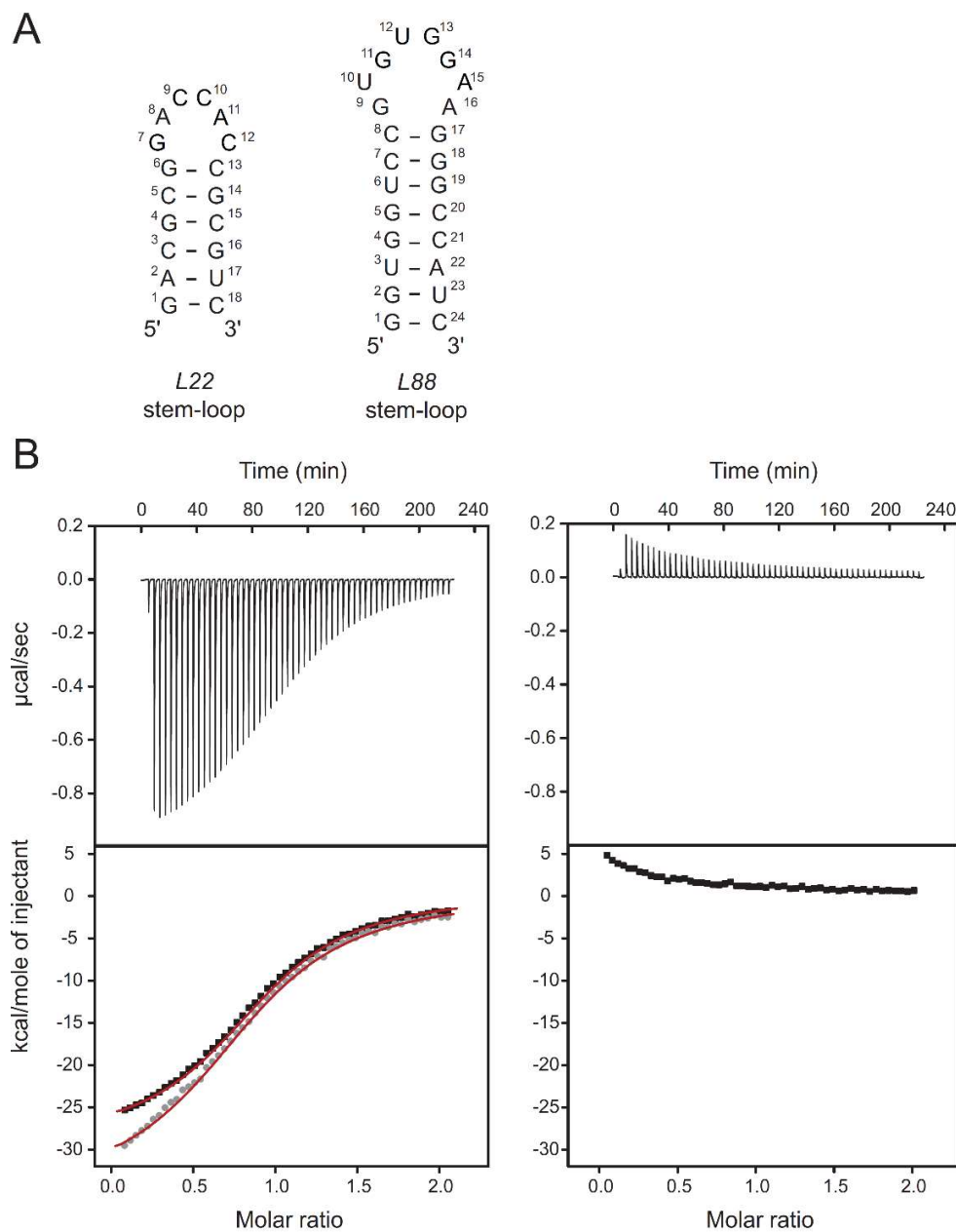
25 mM KCl, 2 mM spermidine and 25 mM  $\text{MgCl}_2$ ).

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



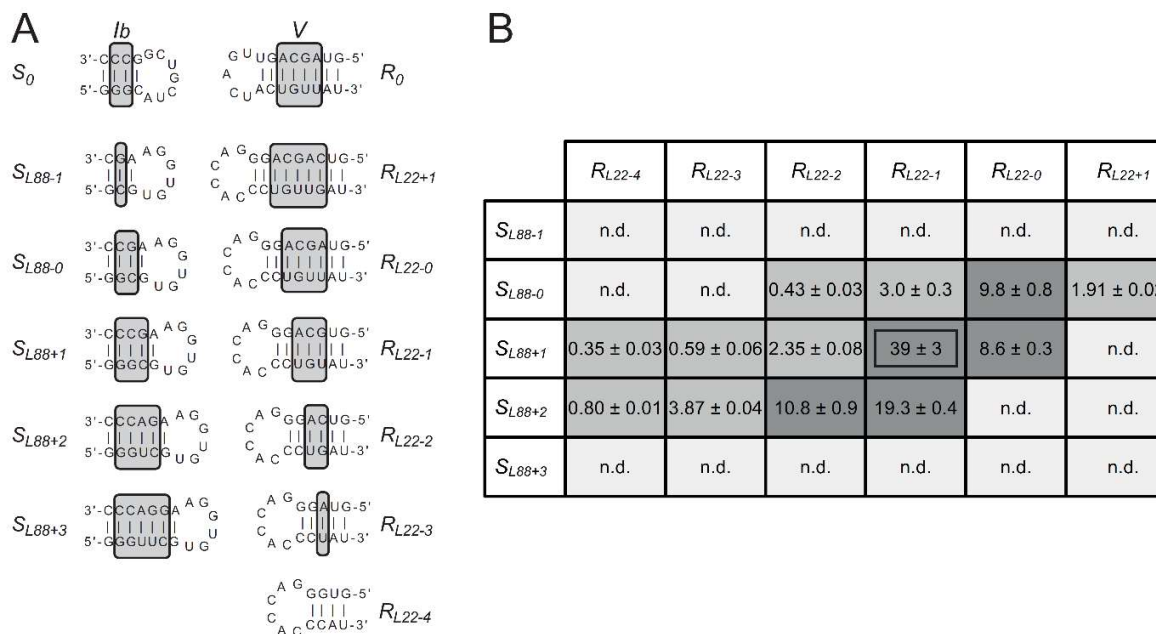
**Figure S3.1. Schematic representation of the computational workflow used to identify and characterize KLIs present in the PDB. (A)** The workflow contains three steps. **(B)** The first step is a motif search in WebFR3D. A secondary structure view of the motif to be queried is shown alongside the search input for WebFR3D. **(C)** The second step analyzes the raw data from the WebFR3D output and classifies the data to provide a list of the KLIs. For example, the two individual entries related to PDB file 3FO4 provide information about the same KLI because they report the same stem-loop closing base pairs (involving residues in columns 1, 3, 4 and 6), but different base pairs at the kissing-loop interface (involving residues listed in columns 2 and 5). **(D)** The final step involves retrieving additional data for each KLI and listing the results in an HTML/JavaScript table. Each KLI is listed with its index number, the PDB code of the structure in which it was found, the title of the PDB file, the residue numbers

of the loop-closing and the interacting nucleotides between the two loops, the identifier(s) of the chain(s) in which these nucleotides were found, the organism and type of molecule involved, the method used for structure determination, the experimental resolution (n.a. = not available), the secondary structure in dot-bracket notation, the number of WC/WC base pairs at the kissing-loop interface, the sizes of the two loops, a hyperlink to a PDB file containing the isolated interaction, and a hyperlink to the pubmed reference.

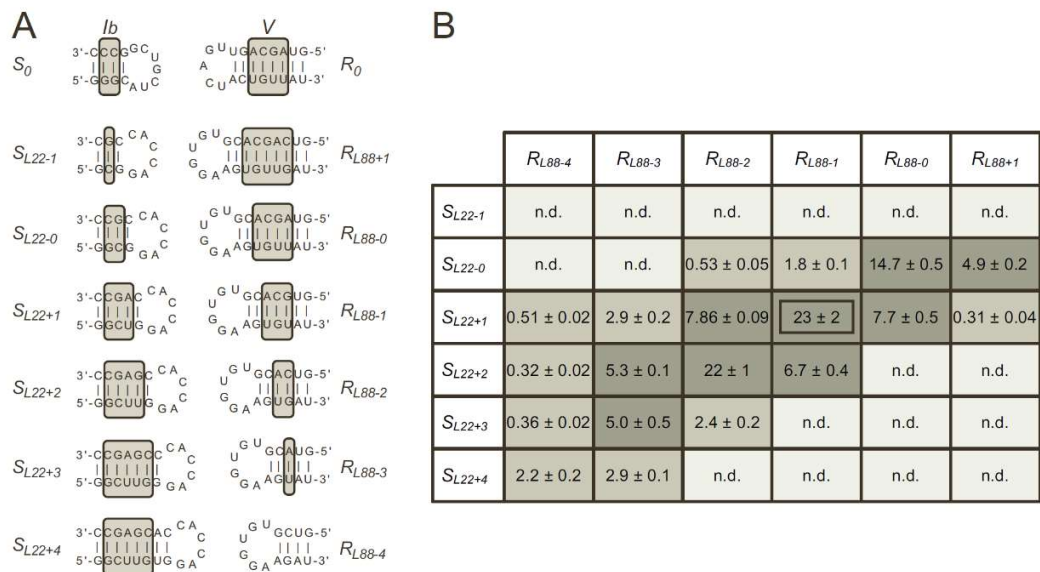


**Figure S3.2. ITC studies of the L22/L88 interaction.** (A) Sequence and proposed secondary structures of the L22 and L88 stem-loops derived from the L22 and L88 secondary structure elements of the 23S rRNA of *Deinococcus radiodurans* [1,2]. (B) ITC data analysis for binding of the L22 stem-loop to the L88 stem-loop at 25°C in a buffer containing 10 mM Tris pH 7.0, 20 mM MgCl<sub>2</sub> and 0.05 mM NaN<sub>3</sub>. The raw data of a representative titration

experiment is shown on the top left panel and the corresponding dilution control (only buffer in the cell) is shown on the top right panel. The isothermal curves representing the integrated data of the titration experiments are shown below. On the left bottom panel, two isothermal curves of the titration experiment are shown, before (black) and after (gray) subtraction of the control experiment (right bottom panel). For the titration experiment, 200  $\mu\text{M}$  L88 was used in the syringe and 20  $\mu\text{M}$  L22 in the cell. The isothermal curve with subtraction of the control experiment (gray) was used to extract the thermodynamic parameters for this interaction:  $K_d = 3.8 \pm 0.5 \mu\text{M}$ ;  $N = 0.875 \pm 0.004$ ;  $\Delta H = -35.2 \pm 0.3 \text{ kcal/mol}$ ;  $\Delta S = -0.0933 \pm 0.0004 \text{ kcal/mol}$  and  $\Delta G_{ITC} = -7.41 \pm 0.08 \text{ kcal/mol}$  ( $\Delta G_{ITC} = -RT\ln(K_A)$  at  $T = 298.15 \text{ K}$ ).

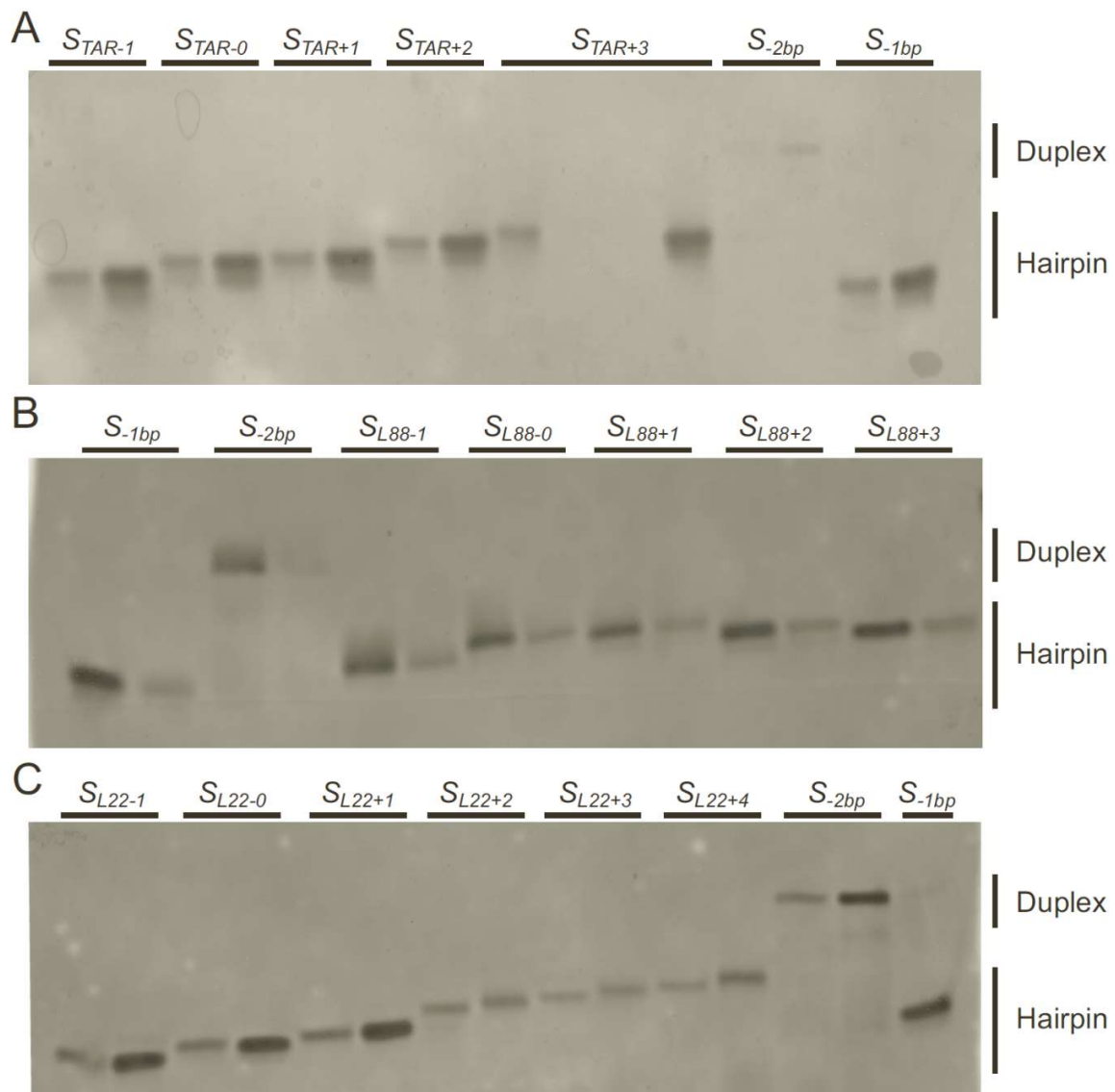


**Figure S3.3. Sequences and kinetic data of the  $S_{L88}/R_{L22}$  variants.** (A) Sequence of the  $S_{L88}/R_{L22}$  variants. Only the regions of the  $S_0$  substrate and  $R_0$  ribozyme (stem-loops Ib and V shown above; Fig. 3.1A) that were substituted in the  $S_{L88}$  and  $R_{L22}$  variants are shown. The number of base pairs varies in stems Ib and V, as indicated by the gray boxes on the putative secondary structures. (B) Grayscale heat map of the  $k_{cat}/K_M$  values ( $10^{-2} \text{ min}^{-1} \mu\text{M}^{-1}$ ) for the  $S_{L88}/R_{L22}$  variants. Dark gray shading corresponds to  $k_{cat}/K_M$  values within 5 fold of the most active  $S_{L88}/R_{L22}$  pair ( $S_{L88+1}/R_{L22-1}$ ); medium gray shading corresponds to lower  $k_{cat}/K_M$  values (5-111 fold lower than the most active  $S_{L88}/R_{L22}$  pair), and light gray shading reflects the absence of detectable cleavage activity (n.d.:  $k_{obs} \leq 0.002 \text{ min}^{-1}$ ) after 24 min at  $[R] \leq 15 \mu\text{M}$ .

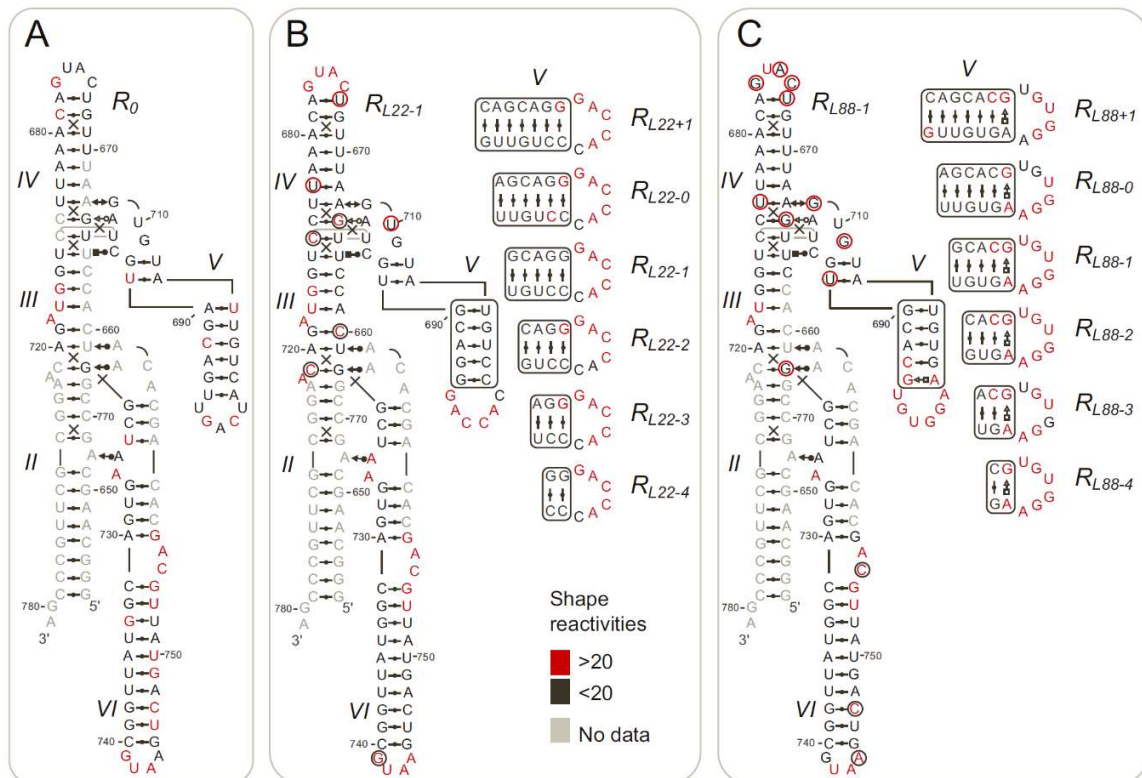


**Figure S3.4. Sequences and kinetic data of the  $S_{L22}/R_{L88}$  variants.** (A) Sequence of the  $S_{L22}/R_{L88}$  variants. Only the regions of the  $S_0$  substrate and  $R_0$  ribozyme (stem-loops Ib and V shown above; **Fig. S3.1A**) that were substituted in the  $S_{L22}$  and  $R_{L88}$  variants are shown. The number of base pairs varies in stems Ib and V, as indicated by the gray boxes on the putative secondary structures. (B) Grayscale heat map of the  $k_{cat}/K_M$  values ( $10^{-2} \text{ min}^{-1} \mu\text{M}^{-1}$ ) for the  $S_{L22}/R_{L88}$  variants. Dark gray shading corresponds to  $k_{cat}/K_M$  values within 5 fold of the most active  $S_{L22}/R_{L88}$  pair ( $S_{L22+1}/R_{L88-1}$ ); medium gray shading corresponds to lower  $k_{cat}/K_M$  values (5-75 fold lower than the most active  $S_{L22}/R_{L88}$  pair), and light gray shading reflects the absence of detectable cleavage activity (n.d.:  $k_{obs} \leq 0.002 \text{ min}^{-1}$ ) after 24 min at  $[R] \leq 15 \mu\text{M}$ .





**Figure S3.5. Conformational analysis of the substrate variants used in this study.** The conformations of (A) the  $S_{STAR}$  substrates, (B) the  $S_{L88}$  substrates and (C) the  $S_{L22}$  substrates are analyzed by native gel electrophoresis. For each substrate, two quantities of RNA (1 and 4  $\mu\text{g}$ ) were loaded on the gel. The hairpin and duplex controls are respectively the  $S_{-1bp}$  and  $S_{-2bp}$  substrates characterized in a previous study [3].



**Figure S3.6. SHAPE analysis of the *R0* ribozyme and related variants.** The normalized NMIA reactivity of each nucleotide is color-coded on the structural schematics of the (A) *R0* ribozyme, (B) *RL22-1* variant, and (C) *RL88-1* variant as per the SHAPE reactivities key. In (B), the normalized NMIA reactivity of the other *RL22* variants is within the same category as for *RL22-1*, except for a few nucleotides in the variable SLV region (right panel) and outside this region (red and black circles indicate inverted reactivity compared to *RL22-1*; C<sub>658</sub> for *RL22+1*, *RL22-2*, *RL22-3* and *RL22-4*, C<sub>660</sub> for *RL22+1*, *RL22-0*, *RL22-3* and *RL22-4*, C<sub>665</sub> for *RL22+1* and *RL22-3*, U<sub>669</sub> for *RL22-2*, U<sub>678</sub> for *RL22+1*, U<sub>685</sub> for *RL22-3*, U<sub>710</sub> for *RL22+1* and G<sub>741</sub> for *RL22-0*). In (C), the normalized NMIA reactivity of the other *RL88* variants is within the same category as for *RL88-1*, except for a few nucleotides in the variable SLV region (right panel) and outside this region (red and black circles indicate inverted reactivity compared to *RL88-1*; G<sub>674</sub> and U<sub>676</sub> for *RL88-4*, C<sub>677</sub> for *RL88+1*, U<sub>678</sub> for *RL88-2*, *RL88-3* and *RL88-4*, U<sub>684</sub> for *RL88-4*, G<sub>685</sub> for *RL88+1* and *RL88-0*, U<sub>689</sub> for *RL88-4*, G<sub>709</sub> for *RL88+1*, G<sub>721</sub> for *RL88+1*, A<sub>744</sub> for *RL88-4*, C<sub>747</sub> for *RL88-3*, *RL88-4*, and C<sub>755</sub> for *RL88+1*).

1. Magee, T.V., et al., *Novel 3-O-carbamoyl erythromycin A derivatives (carbamolides) with activity against resistant staphylococcal and streptococcal isolates*. *Bioorg Med Chem Lett*, 2013. **23**(6): p. 1727-31.
2. Cannone, J.J., et al., *The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs*. *BMC Bioinformatics*, 2002. **3**: p. 2.
3. 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.

**4. Étude des liens entre structure et dynamique du  
complexe kissing-loop I/V et l'activité du ribozyme VS  
de *Neurospora***

## Mise en contexte

L'article précédent a établi que la substitution de l'interaction boucle-boucle entre la boucle terminale de l'hélice V et la boucle terminale du substrat par des variantes structurellement similaires pouvait permettre d'obtenir des ribozymes VS modifiés actifs. Il a également montré que l'activité de ces variantes était sensiblement plus faible que celle du ribozyme pourvu de l'interaction boucle-boucle naturelle. Les questions à cette étape de notre étude étaient donc les suivantes : quels phénomènes, structuraux et/ou dynamiques, permettent d'expliquer la disparité de taux d'activité entre le ribozyme naturel et ses variantes? Et par extension, est-il possible de compenser ce phénomène afin de rétablir un taux d'activité au moins aussi bon que celui du ribozyme naturel? Pour répondre à ces interrogations, nous avons premièrement examiné la dynamique intrinsèque des trois interactions boucle-boucle étudiées et exécuté des expériences de sélection *in vivo* couplées à des caractérisations enzymatiques.

Afin d'explorer les capacités dynamiques intrinsèques de l'interaction I/V naturelle ainsi que des deux variantes étudiées précédemment (interaction TAR/TAR\* et L22/L88), nous avons simulé des constructions basées sur ces variantes grâce à un protocole de dynamique moléculaire par échange de répliques basé sur la température (T-REMD). Ce protocole nous a permis d'échantillonner l'espace conformationnel des interactions *kissing-loop* étudiées de manière plus étendue qu'une simulation en dynamique moléculaire classique. Les structures des trajectoires obtenues par simulation en T-REMD ont ensuite été analysées selon plusieurs métriques, soit par le calcul d'angles d'Euler entre les deux tiges des complexes et leur capacité à adopter une conformation fermée lorsque transposée dans le contexte de la structure cristalline du ribozyme VS. De plus, l'ensemble des structures de la trajectoire ont été analysées en composante principale afin d'extraire les modes de mouvement constitutifs du mouvement global de chaque variante au cours de la simulation. Muni de ces données, nous avons cherché à identifier les différences fondamentales entre les dynamiques des trois interactions *kissing-loop* et de localiser l'origine de leurs mouvements respectifs. L'examen de la dynamique du substrat naturel indiquent la grande importance de la présence de résidus libres dans l'une des boucles de l'interaction I/V et des expériences de sélection *in vitro* ont été entreprises pour le vérifier. Celles-ci ont permis d'améliorer l'activité du variant TAR/TAR\* et démontrent

l'importance de considérer l'échantillonnage conformationnel lors de l'ingénierie de ribozymes ayant de nouvelles activités.

Le projet a été développé par moi-même conjointement avec Pascale Legault et Pierre Dagenais. J'ai réalisé toutes les expériences à l'exception de la sélection *in vitro* et son analyse, réalisée par Pierre Dagenais et Julie Lacroix-Labonté. L'article a été écrit par moi-même conjointement avec Pascale Legault, Pierre Dagenais et Julie Lacroix-Labonté.

**A multi-axial RNA joint with a large range of motion promotes  
sampling of an active ribozyme conformation**

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

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

\*To whom correspondence should be addressed: Pascale Legault, Phone: 514-343-7326, Fax:  
514-343-2210, E-mail: [pascale.legault@umontreal.ca](mailto:pascale.legault@umontreal.ca).

This work was prepared for future submission in JACS

## Abstract

To fulfill their biological function, RNA molecules adopt complex structures that can be viewed as modular assembly of basic structural elements. Investigating the dynamics of these structural elements is important to better understand the mechanism of biological RNAs and for engineering RNAs with novel functions that could serve in system biology, nanotechnology and molecular medicine applications. Previously, we performed rational engineering studies of the Varkud satellite (VS) ribozyme to switch its specificity towards non-natural hairpin substrates via modification of a critical kissing-loop interaction (KLI). We identified highly-functional VS ribozyme variants with modified KLIs (ribosomal RNA L88/L22 and HIV-1 TAR/TAR\*), but they displayed ~100-fold reduction in cleavage activity. Here, we aimed to characterize the dynamics of KLIs to correlate dynamic properties with function and improve the activity of designer ribozymes. Using temperature replica-exchange molecular dynamics, we found that the natural KLI forms a multi-axial RNA joint that supports conformational sampling of the closed and active state of the ribozyme, whereas the surrogate KLIs display more restricted motions. Based on *in vitro* selection, we showed that the cleavage activity of the ribozyme variant with the TAR/TAR\* KLI could be markedly improved by modification of the TAR receptor loop (TAR\*), such that it partly destabilized its interaction with the TAR-derived substrate but added a critical twist motion and increased the overall dynamics of the KLI. This study offers a new perspective on the role of subdomain dynamics on ribozyme function and underlines the importance of considering dynamic sampling when designing RNAs with novel functions.



## Introduction

RNA molecules adopt complex three-dimensional structures to fulfill essential function in diverse cellular processes and in all living forms. RNA architecture is often recognized as being highly modular because RNA molecules contain a diverse but limited set of recurrent structural motifs within their secondary and tertiary structural elements, such as hairpin loops, multi-way junctions and kissing-loop interactions.[1] The folding of a structured RNA typically follows a hierarchical process, where the secondary structure forms first and is then compacted via tertiary contacts.[2,3] As part of the RNA secondary structure, paired regions form helical structures and unpaired regions adopt specific structural motifs, which impart orientations to the associated helical domains and contribute to formation of tertiary contacts.[4] It is becoming clear that local structural elements have inherent flexibility in solution that allows them to explore the conformational space. Consequently, each structural element should not simply be viewed as a static structure but as a sub-ensemble of structures with specific dynamic properties.[4-10] However, despite growing knowledge of how structural motifs define the structure of RNA and increased appreciation for the importance of RNA dynamics, much remains to be known regarding the dynamics that are inherent to specific structural motifs and how they contribute to RNA structure and function.

The modular and hierarchical organization of RNA structure has been made evident through successful efforts in RNA engineering, an emerging field of research that holds tremendous potential for synthetic biology, nanotechnology and molecular medicine applications.[11,12] Novel RNA function can be created by de novo design of self-assembled nanostructures or incorporation of structural elements into existing RNA architectures.[11-14] Our laboratory is exploring the latter approach to modify the substrate specificity of the *Neurospora* Varkud satellite (VS) ribozyme.[15,16] The wild-type VS ribozyme specifically recognizes and cleaves a folded hairpin substrate, and we are exploring the potential of this ribozyme to cleave non-natural hairpin substrates, a secondary structure elements that is abundantly found in a wide variety of cellular and viral RNAs.

The VS ribozyme is a nucleolytic self-cleaving ribozyme, of the same family as the *glmS*, hammerhead, hairpin, HDV, hatchet, pistol, twister and twister sister ribozymes.[17-19] Like

many other self-cleaving ribozymes, it represents a good model system for structure, function and engineering studies because of its relatively small size and its ability to cleave its substrates both in *cis* and in *trans*.<sup>[20,21]</sup> Its minimal sequence is comprised of approximately 160 nucleotides forming six paired domains. The *trans* ribozyme (domains II-VI) recognizes its hairpin substrate (stem-loop I or SLI) by way of a kissing-loop interaction (KLI) formed between loop I and loop V (Fig. 4.1a), and which is crucial for efficient cleavage activity.<sup>[22]</sup> This I/V KLI is highly stable and allows formation of the open state of the ribozyme, which likely predominates in solution (Fig. 4.1b).<sup>[23-26]</sup> Subsequently, the G<sub>638</sub> loop, known as the cleavage site internal loop of the SLI substrate, intimately associate with the A<sub>756</sub> loop to form the active site within the closed state of the ribozyme (Fig. 4.1ab).<sup>[27]</sup> The cleavage mechanism has been thoroughly investigated and involves the N1 atoms of guanine (G<sub>638</sub>) and adenine (A<sub>756</sub>) nucleobases as general base and acid, respectively.<sup>[18]</sup> However, the dynamic processes associated with formation of the active site from the open state to the closed state have yet to be examined.

In our most recent rational engineering studies, we substituted the I/V KLI by two other KLIs with similar thermodynamic and structural properties (Fig. 4.1c): the L88/L22 KLI from the large subunit of the *Deinococcus radiodurans* ribosomal RNA and the synthetic TAR/TAR\* KLI related to the human immunodeficiency virus (HIV-1).<sup>[16]</sup> The domain substitution was successful since the overall structure of the ribozyme remained intact as judged from SHAPE analysis and substantial levels of cleavage activity were obtained for variant substrate/ribozyme (*S/R*) complexes that incorporate either the L88/L22 KLI (*S<sub>L88</sub>/R<sub>L22</sub>*) or TAR/TAR\* KLI (*S<sub>TAR</sub>/R<sub>TAR\*</sub>*). However, although the catalytic activity of the variant ribozymes was optimized by fine-tuning the lengths of stems I and V, the *k<sub>obs</sub>* values for the most active variant *S/R* complexes (*S<sub>L88+1</sub>/R<sub>L22-1</sub>* and *S<sub>TAR-1</sub>/R<sub>TAR\*-0</sub>*) were 50-160 fold lower than that of the reference *S<sub>0</sub>/R<sub>0</sub>* system (Fig. 4.1c). These kinetic results suggest that the transition from the open state to the closed state is rate-limiting in these KLI variants of the VS ribozyme.

With this work, we investigated the dynamic properties of three KL complexes, the I/V, L88/L22 and TAR/TAR\* KL complexes, to correlate the dynamics of KLI with ribozyme function. We ran 50-ns unrestrained Temperature-Replica Exchange Molecular Dynamics (T-REMD), an enhanced atomistic sampling method that allows a better exploration of the

conformational space than standard molecular dynamics (MD) and has been applied to several RNA systems.[28] We analyzed the resulting trajectories to characterize the intrinsic dynamics of these KLIs and establish how they may contribute to formation of a closed state for substrate cleavage in the VS ribozyme. The T-REMD simulations support the hypothesis that the reduced cleavage activity of KLI variants of the VS ribozyme correlates with restricted motion of their KLI. Furthermore, to improve the cleavage activity of one of the  $S_{TAR}/R_{TAR}^*$  ribozyme variant, we performed *in vitro* selection and identified an  $S_{TAR}/R_{TAR}^*$  variant with increased dynamics at the KLI and for which the cleavage activity was markedly improved. With this structural dynamics-function study, we demonstrate the importance of dynamics at the KLI for active-site formation and cleavage activity of the VS ribozyme and highlight the remarkable dynamic potential of RNA.

## Results and Discussion

**T-REMD simulations of kissing-loop complexes.** We ran 50-ns T-REMD atomistic simulations to characterize the structural dynamics of three kissing-loops (KL) complexes in explicit solvent (Fig. S4.1). The T-REMD approach was used to enhance conformational sampling; MD simulations were run in parallel in ~60 replicas at evenly-distributed temperatures between 300 K and 375 K, with periodic attempts to exchange RNA coordinates between replicas of adjacent temperatures. Thus, the conformational space was sampled for a total MD simulation time of ~3  $\mu$ s for each KL complex.

To ensure that the T-REMD protocol allowed for a thorough exploration of the conformational space, the convergence of the demultiplexed trajectories was assessed.[29] The demultiplexed trajectories were first obtained by following the continuous trajectory of each starting structure through the temperature space, and for each demultiplexed trajectory an RMSD frequency profile was generated based on heavy atom RMSDs between each frame and the T-REMD starting structure (Fig. S4.2). The resulting RMSD profiles are all very similar with only a few outliers for the I/V KL complex (Fig. S4.2a), as well as for the L88/L22 and TAR/TAR\* KL complexes (Fig. S4.2bc). Moreover, the detection of a single peak in the overlaid profiles indicates that all the conformations are very similar in RMSD, and that the T-REMD protocol did not cause large structural changes. Thus, the T-REMD trajectories of the

I/V, L88/L22 and TAR/TAR\* KL complexes converge reasonably well, and this validates the analysis of their structural dynamics.

**Basic structural analysis of T-REMD trajectories.** For each KL complex, the 250,000 frames of the lowest-temperature T-REMD trajectory (300 K) were structurally characterized using metrics that are either intrinsic to the KL complex or evaluated in the context of the VS ribozyme crystal structure (Fig. 4.2). First, the distribution of Euler angles ( $\alpha$ ,  $\beta$  and  $\gamma$ ) between the two helical stems of the KL complexes were computed,[30] which helped derive the more intuitive interhelical angle  $\theta$  (Fig. 4.2a). Then, composite models were built starting from the KL complex of each simulated frame and then attaching the internal loop of SLI to one stem (I, L88 or TAR) and the *trans* VS ribozyme to the other stem (V, L22 or TAR\*) via superposition of common helical elements. These composite models allowed the analysis of individual frames in the context of the VS ribozyme crystal structure. From heavy atom superposition of central base pairs of stem V within the composite model and the crystal structure, we measured the distance between the phosphorus atoms at the scissile phosphate of the two SLI internal loops ( $d_{PP}$ ; Fig. 4.2b) as well as the RMSD between these two loops.

Two-dimensional (2D) heatmaps showing distributions of Euler angles in terms of pairwise combinations ( $\beta$ - $\alpha$ ,  $\gamma$ - $\alpha$  and  $\gamma$ - $\beta$ ) demonstrate that the I/V KL complex explores a wider region of the Euler angle space compared to the L88/L22 and TAR/TAR\* KL complexes (Fig. 4.3). The I/V KL complex access interhelical angles that the surrogate KL complexes never adopt during the simulation, notably the set of Euler angles measured for the I/V KL complex within the crystal structure (marked by a star in Fig. 4.3). Thus, the analysis of interhelical Euler angles indicate that the intrinsic global dynamics of the I/V KL complex are less constrained than those of the TAR/TAR\* and L88/L22 KL complexes. The restricted global mobility of the surrogate KL complexes is particularly important for the TAR/TAR\* KL complex, which is confined to a narrow range of  $\alpha$ ,  $\beta$  and  $\gamma$  angles.

This disparity in conformational space between the I/V KL complex and the surrogate KL complexes was also examined from 2D heatmaps showing the distribution of interhelical angle  $\theta$  as a function of  $d_{PP}$ .(Fig. 4.4) For the I/V KL complex, several frames of the T-REMD trajectory adopt  $d_{PP}$  values approaching zero Ångström, and small  $d_{PP}$  values are observed for a wide range of interhelical angles ( $d_{PP} < 2 \text{ \AA}$  for  $\theta = 136^\circ$ - $178^\circ$ ), matching the range observed in

the ensemble of 20 lowest-energy NMR structures of the I/V KL complex ( $\theta = 127^\circ$ - $175^\circ$ ).<sup>[24]</sup> This analysis clearly indicates that the trajectory of the I/V KL complex brings the internal loop of the substrate close to its position in the crystal structure (marked by a star in Fig. 4.4), which is compatible with formation of the active site. In contrast, for the L88/L22 and TAR/TAR\* KL complexes,  $d_{PP}$  values greater than 6 Å and 11.5 Å, respectively, are observed, and these  $d_{PP}$  values directly correlate with the interhelical angle  $\theta$ . Similar observations were made by examining 2D heatmaps showing the distribution of interhelical angle  $\theta$  as a function of RMSD (Fig. S4.3), where the smallest RMSD value attained by the I/V KLI (2.6 Å), is smaller than those for the L88/L22 KL complex (8.7 Å) and the TAR/TAR\* KL complexes (14.9 Å). Thus, the I/V KLI allows for a more complex range of motion and enables a large subpopulation of structures to approach the closed state in which the active site is formed.

**Principal component analysis of T-REMD trajectories.** To better understand the dynamics of the KL complexes, a principal component (PC) analysis was performed on the atomic coordinates of each T-REMD trajectory. This analysis extracts non-overlapping concerted atomic movements, or principal components (PCs), from the seemingly random vibrations of atomic positions observed in the T-REMD trajectories. For the three KL complexes, the first five PCs together represent more than 70% of the cumulative percentage of variance (Fig. S4.4 and Table 4.1), thus the main contributions of the global motion. Pseudo-trajectories for each of these first five PCs were produced, in which atomic movements are projected along a single PC. Although these pseudo-trajectories can give rise to some partially unrealistic local structures, they are very useful in illustrating the direction and relative amplitude of concerted motions. They are displayed here in the context of the VS ribozyme crystal structure (Fig. 4.5 and Movies S1-S15) to provide insights into how specific motions may contribute to formation of the active site.

For the I/V KL complex, the first five PCs contribute to bringing the scissile phosphate of the SLI substrate in proximity of the A<sub>756</sub> internal loop within the ribozyme, with the first two PCs having the greatest impact (Fig. 4.5 and Movies S1-S5). The first PC, which accounts for 34% of the variance (Table 4.1), is characterized by a twist motion of large amplitude of the SLI substrate around an axis close to its own principal axis (Fig. 4.5a). Interestingly, this twist motion appears to be correlated with an important structural transition in loop I between compact

and loose states (Movie S1). The second PC, with 16% of the variance, is also of large amplitude but is better described by a side-to-side bend motion that swings the substrate perpendicularly to the II-VI plane formed by helical domains II and VI (Fig. 5b). The third PC describes a twist motion of smaller amplitude than the first PC and accounts for 9.5% of the variance. Finally, the fourth and fifth PCs both describe back-to-front bend motions, during which the substrate travels parallel to the II-VI plane (Fig. 5de). Taken together these PCs of motion indicate that the I/V KLI can be viewed as a multi-axial "RNA joint" that resembles a ball-and-socket shoulder joint, both allowing circular movements in three orthogonal directions.

Analysis of the individual PCs for the surrogate KL complexes indicate that their global dynamics are similarly composed of a mixture of movements, which comprises a twist motion of the substrate around its main axis and different bend motions (Figs. S4.5 and S4.6; Movies S6-S15). The contribution and amplitude of these movements differ between the three KL complexes (Table 4.1). Only the PCs for the I/V KL complex have sufficient amplitudes to bring the G<sub>638</sub> loop in proximity of the A<sub>756</sub> loop, in agreement with the short  $d_{pp}$  values observed for this complex. Moreover, the I/V KL complex is the only one in which the preeminent motion is a twist of the substrate around an axis close to its principal axis. Whereas twist motions are observed for the L88/L22 and TAR/TAR\* KL complexes, their contribution and amplitude are far less important.

**Dynamics of KL complexes versus cleavage activity of S/R complexes.** In order for the VS ribozyme to cleave its substrate, the G<sub>638</sub> and A<sub>756</sub> loops must intimately associate such that the reactive atoms are precisely positioned at the active site. However, in solution the VS ribozyme is best represented by an ensemble of open conformations in which these two loops are apart.[25,26,31] Our analysis of T-REMD trajectories suggest that the higher activity of the  $S_0/R_0$  complex compared to the  $S_{L88}/R_{L22}$  and  $S_{TAR}/R_{TAR^*}$  complexes (Fig. 4.1c) is linked to the dynamics of the I/V KL complex, which is better suited to allow spatial closeness and proper positioning of the G<sub>638</sub> loop relative to the A<sub>756</sub> loop. The twist and side-to-side motions of large amplitude described by the first two PCs can bring the two internal loops in proximity, while other twist and bend motions of smaller amplitude can fine tune the relative position of the two internal loops.

Given that the  $S_{L88/R_{L22}}$  and  $S_{TAR/R_{TAR^*}}$  variant complexes retain a substantial level of cleavage activity (Fig. 1c), it is intriguing that the dynamics of the corresponding KL complexes characterized here does not allow the G<sub>638</sub> and A<sub>756</sub> loops to come in proximity. This is clearly shown both from the large  $d_{pp}$  values of the T-REMD frames (Fig. 4.4bc) and visualization of the main PCs of motions (Figs. S4.5 and S4.6; Movies S6-S15). The lack of frames with smaller  $d_{pp}$  values may be due to the limited timescale of the T-REMD simulations. MD/T-REMD simulations that extend to longer timescales could possibly capture closed conformations, possibly by exploring less stable KL structures with fewer base pairs. Indeed, we observed surrogate KL complexes with disrupted KL base pairs in frames with the shortest  $d_{pp}$  values. Alternatively, other dynamic hotspots of the VS ribozyme, potentially the stem III bulge and the two three way junctions,[26] could contribute to bringing the G<sub>638</sub> and A<sub>756</sub> loops in sufficient proximity to allow substrate cleavage.

#### **Identification of the local hotspots of motion responsible for the global dynamics.**

To identify local hotspots of motion, we computed for each KL complex the root-mean square fluctuation (RMSF) of every consecutive pair of residues along the sequence (Fig. 4.6). Remarkably, the I/V KL complex exhibits several clusters of residues with high RMSF values ( $> 1.5 \text{ \AA}$ ), defining three local hotspots of motion (residues 8-12 of loop I; residues 16-17 of loop I and residue 15 of loop V), in agreement with previous NMR studies.[24] In contrast, fewer residues with high RMSF values were identified for the L88/L22 and TAR/TAR\* KL complexes (Fig. S4.7), in agreement with more restricted motions associated with these surrogate KL complexes. It is not surprising that for all three KL complexes, the most dynamic residues are found at the junction between the hairpin stem and the KL stem. Unlike the surrogate KL complexes, the I/V KL complex contains several unpaired residues (I: A<sub>9</sub>, U<sub>10</sub> and G<sub>16</sub>; V: U<sub>11</sub> and U<sub>15</sub>) and unstable base pairs (I: C<sub>8</sub>-G<sub>15</sub>; V: U<sub>10</sub>-A<sub>16</sub>) at this junction (Fig. S4.1).[24]

Through their association with dynamic hotspots, the unpaired residues in loop I largely contribute to the conformational dynamics the I/V KL complex. Notably, they participate in the structural transition between compact and loose states of loop I that produces the large twist motion observed in the I/V KL complex (Movie S1). The two base pairs closing loop I (G<sub>7</sub>-C<sub>17</sub> and C<sub>8</sub>-G<sub>15</sub>) can form in the compact state, as observed in the NMR structure of the I/V KL

complex,[24] but are clearly disrupted in the loose state. Such helical unwinding is not observed for the surrogate KL complexes, which would explain why the preeminent twist motion observed for the I/V KL complex is not observed for these KL complexes.

**RNA engineering by increasing the dynamics of an RNA junction.** Given the importance of dynamics at the KLI for the high cleavage activity of the VS ribozyme complex ( $S_0/R_0$ ), we postulated that the cleavage activity of the variant  $S_{TAR}/R_{TAR^*}$  complex could be increased if one could add some dynamics at the TAR/TAR\* KLI. In order to test this idea, we first performed an *in vitro* selection looking for VS ribozyme variants that would more efficiently cleave the  $S_{TAR}$  substrate (Fig. 4.7a). For this selection, the  $S_{TAR}$  substrate was covalently bound to the ribozyme; it contains the HIV-1 TAR RNA sequence in stem Ia and stem-loop Ib, but preserved the natural G<sub>638</sub> loop necessary for formation of the active site. In addition, loop V (5 nt) and the adjacent two closing base pairs (4 nt) were replaced by a stretch of 10 randomized nt to form an RNA library with approximately one million ( $4^{10}$ ) different loop sequences. After six rounds of selection and amplification, next-generation sequencing of the selected ribozyme pool was performed, and enriched sequences (present 10 times or more in the pool) were analyzed in terms of the potential number of consecutive WC base pairs that could form with loop I (Fig. 4.7b). Almost all selected sequences could form at least 4 WC base pairs, confirming that the *in vitro* selection procedure allowed us to select variant VS ribozymes for which cleavage in *cis* depends on the formation of the KLI. Formation of 4 WC base pairs at the KLI appears to be necessary, but also sufficient for cleavage activity since 64% of the sequences could form a maximum of 4 WC base pairs. Since only 2% of the ribozyme pool could form the maximum of 6 WC base pairs, increasing the thermodynamic stability of the KLI does not appear to be beneficial, but instead may inhibit the activity. We found two main classes of loop V sequences with the preferred CCCA pattern starting either at position +4 or +5 of the randomized sequence (Fig. 4.7c). In both cases, formation of 4-5 WC base pairs at the KLI would leave several residues unpaired in both kissing loops, which would provide more dynamics for formation of the active site and efficient substrate cleavage.

Interestingly, when we modified the *trans* VS ribozyme  $R_0$  to incorporate one of the selected loop V sequence predicted to form 4 consecutive WC base pairs with  $S_{TAR}$ , the resulting ribozyme ( $R_{TAR-SI}$ ) cleaved  $S_{TAR}$  with a relatively high single-turnover rate ( $k_{obs} = 5.8 \text{ min}^{-1}$ ; Fig.



4.7d). This rate is much higher than that previously reported for the most active rationally-engineered  $S_{TAR}/R_{TAR^*}$  complex ( $k_{obs} = 0.077 \text{ min}^{-1}$ ; Fig. 4.1c) and similar to the rate of the parental  $S_0/R_0$  complex ( $k_{obs} = 12.2 \text{ min}^{-1}$ ; Fig. 4.1c).[16] In addition, the  $R_0$  and  $R_{TAR^*-0}$  ribozymes, which can respectively form 3 and 6 WC consecutive base pairs with  $S_{TAR}$ , both cleave this substrate with a much lower rate ( $\leq 200$ -fold;  $k_{obs} \leq 0.03 \text{ min}^{-1}$ ; Fig. 4.7d) than  $R_{TAR-S1}$ . These results further underline the importance of a KLI that is sufficiently stable yet appropriately dynamic for efficient substrate cleavage by the VS ribozyme.

We performed a T-REMD simulation to investigate the dynamics of the TAR/TAR-S1 KL complex (Figs. 4.7 and S4.8). Although this T-REMD simulation was initiated from a structural model and did not converge quite to the same extent as the other simulations reported here (Fig. S4.8a), broad comparisons with the original TAR/TAR\* KL complex can still be drawn. We found that the selected TAR/TAR-S1 KL complex explores a much larger region of the Euler angle space (Fig. S4.8b) than the original TAR/TAR\* KL complex (Fig. 4.3c), and include the values determined from the crystal structure (marked by a star in Fig. S4.8b). Several frames of the T-REMD trajectory adopt  $d_{pp}$  values (down to 5.6 Å; Fig. S4.8b) that are much smaller than the smallest values observed for the TAR/TAR\* KL complex (11.5 Å; Figs. 4.4c), indicating that the dynamics of TAR/TAR-S1 KL complex brings the internal loop of the substrate closer to its position in the crystal structure. Interestingly, PC analysis of the 300-K T-REMD trajectory for the TAR/TAR-S1 KL complex reveals a dominating twist motion of the substrate around an axis close to its principal axis (Fig. 4.7f; first PC with 38.4% variance and a 58-Å amplitude of motion), as previously observed for I/V KL complex, but not the surrogate TAR/TAR\* KL complex. Furthermore, higher RMSF values were obtained for the TAR/TAR-S1 KL (Fig. S4.8e) compared to TAR/TAR\* KL complex (Fig. S4.7), particularly for residues at the junction between the hairpin and the KL stems, in agreement with increased motions associated with the selected TAR/TAR-S1 KL complex. Taken together, our kinetic and T-REMD results demonstrate that the cleavage activity of the variant  $S_{TAR}/R_{TAR^*}$  complex could be significantly increased by changing the sequence of loop V to partly destabilize the KLI and thereby increasing the dynamics at the KL junction.

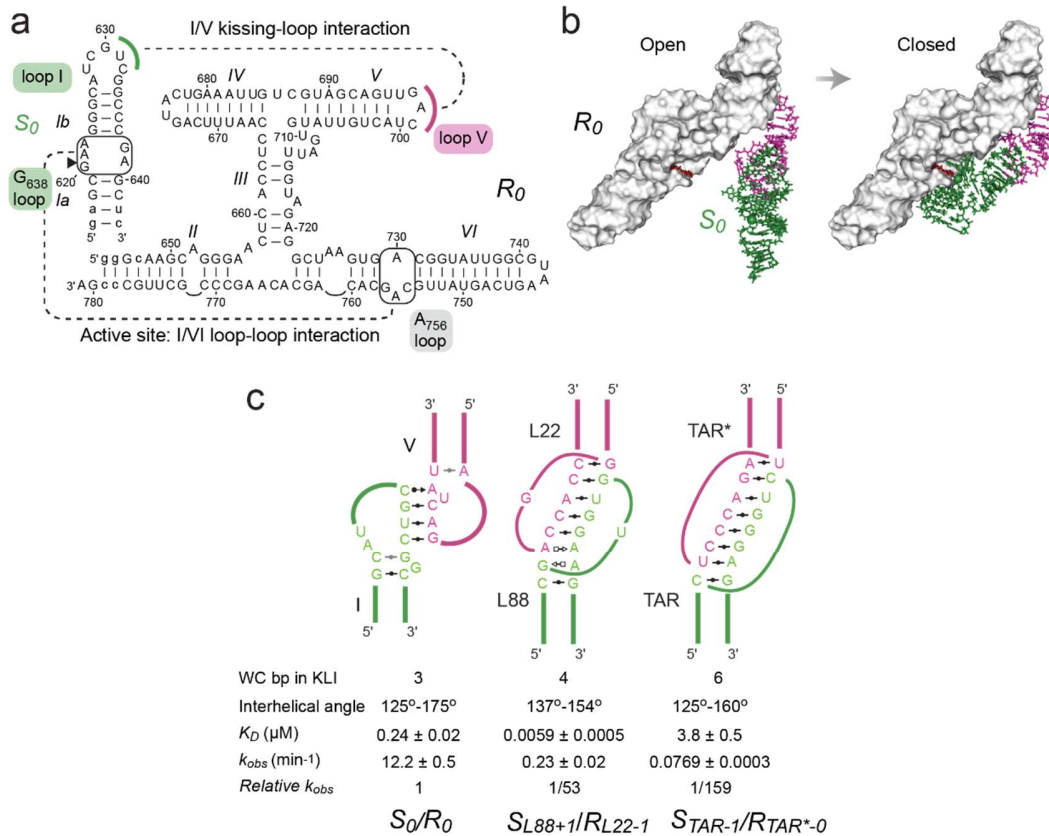
## References

1. Parlea, L.G., et al., *The RNA 3D Motif Atlas: Computational methods for extraction, organization and evaluation of RNA motifs*. *Methods*, 2016. **103**: p. 99-119.
2. Brion, P. and E. Westhof, *Hierarchy and dynamics of RNA folding*. *Annu. Rev. Biophys. Biomol. Struct.*, 1997. **26**(1): p. 113-137.
3. Tinoco, I., Jr. and C. Bustamante, *How RNA folds*. *J. Mol. Biol.*, 1999. **293**(2): p. 271-81.
4. Bailor, M.H., X. Sun, and H.M. Al-Hashimi, *Topology links RNA secondary structure with global conformation, dynamics, and adaptation*. *Science.*, 2010. **327**(5962): p. 202-6.
5. Mustoe, A.M., C.L. Brooks, and H.M. Al-Hashimi, *Hierarchy of RNA functional dynamics*. *Annu. Rev. Biochem.*, 2014. **83**: p. 441-66.
6. Herschlag, D., B.E. Allred, and S. Gowrishankar, *From static to dynamic: the need for structural ensembles and a predictive model of RNA folding and function*. *Curr. Opin. Struct. Biol.*, 2015. **30**: p. 125-133.
7. Daher, M., et al., *Tuning RNA folding and function through rational design of junction topology*. *Nucleic Acids Res.*, 2017. **45**(16): p. 9706-9715.
8. Bisaria, N., et al., *Quantitative tests of a reconstitution model for RNA folding thermodynamics and kinetics*. *Proc. Natl. Acad. Sci. U.S.A.*, 2017. **114**(37): p. E7688-e7696.
9. Ganser, L.R., et al., *High-performance virtual screening by targeting a high-resolution RNA dynamic ensemble*. *Nat. Struct. Mol. Biol.*, 2018. **25**(5): p. 425-434.
10. Denny, S.K., et al., *High-throughput investigation of diverse junction elements in RNA tertiary folding*. *Cell.*, 2018. **174**(2): p. 377-390.e20.
11. Geary, C., et al., *Composing RNA nanostructures from a syntax of RNA structural modules*. *Nano Lett.*, 2017. **17**(11): p. 7095-7101.

12. Jasinski, D., et al., *Advancement of the emerging field of RNA nanotechnology*. ACS Nano., 2017. **11**(2): p. 1142-1164.
13. Hallberg, Z.F., et al., *Engineering and in vivo applications of riboswitches*. Annu. Rev. Biochem., 2017. **86**(1): p. 515-539.
14. Balke, D., R. Hieronymus, and S. Muller, *Challenges and perspectives in nucleic acid enzyme engineering*. Adv. Biochem. Eng. Biotechnol., 2017.
15. Lacroix-Labonté, J., et al., *Rational engineering of the Neurospora VS ribozyme to allow substrate recognition via different kissing-loop interactions*. Nucleic Acids Res., 2016. **44**(14): p. 6924-34.
16. Lacroix-Labonté, J., et al., *Helix-length compensation studies reveal the adaptability of the VS ribozyme architecture*. Nucleic Acids Res., 2011. **40**(5): p. 2284-93.
17. Jimenez, R.M., J.A. Polanco, and A. Luptak, *Chemistry and biology of self-cleaving ribozymes*. Trends Biochem. Sci., 2015. **40**(11): p. 648-661.
18. Wilson, T.J., Y. Liu, and D.M.J. Lilley, *Ribozymes and the mechanisms that underlie RNA catalysis*. Front. Chem. Sci. Eng., 2016. **10**(2): p. 178-185.
19. Ren, A., R. Micura, and D.J. Patel, *Structure-based mechanistic insights into catalysis by small self-cleaving ribozymes*. Curr. Opin. Chem. Biol., 2017. **41**: p. 71-83.
20. Haseloff, J. and W.L. Gerlach, *Sequences required for self-catalysed cleavage of the satellite RNA of tobacco ringspot virus*. Gene., 1989. **82**(1): p. 43-52.
21. 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.
22. 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.
23. Bouchard, P. and P. Legault, *A remarkably stable kissing-loop interaction defines substrate recognition by the Neurospora VS Ribozyme*. RNA., 2014. **20**(9): p. 1451-64.

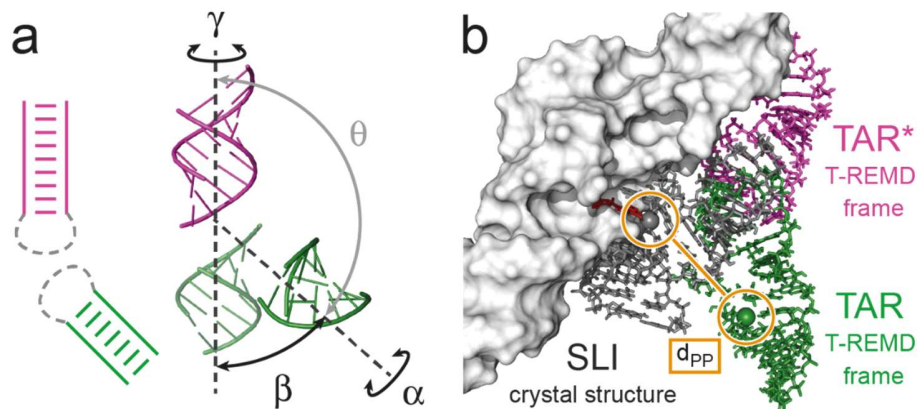
24. Bouchard, P. and P. Legault, *Structural insights into substrate recognition by the Neurospora Varkud satellite ribozyme: importance of U-turns at the kissing-loop junction*. *Biochemistry.*, 2013. **53**(1): p. 258-69.
25. Bonneau, E., et al., *The NMR structure of the II-III-VI three-way junction from the Neurospora VS ribozyme reveals a critical tertiary interaction and provides new insights into the global ribozyme structure*. *RNA.*, 2015. **21**(9): p. 1621-32.
26. Dagenais, P., et al., *Insights into RNA structure and dynamics from recent NMR and X-ray studies of the Neurospora Varkud satellite ribozyme*. *WIREs RNA.*, 2017. **8**(5).
27. Suslov, N.B., et al., *Crystal structure of the Varkud satellite ribozyme*. *Nat. Chem. Biol.*, 2015. **11**(11): p. 840-6.
28. Sponer, J., et al., *RNA structural dynamics as captured by molecular simulations: A comprehensive overview*. *Chem Rev*, 2018. **118**(8): p. 4177-4338.
29. Henriksen, N.M., D.R. Roe, and T.E. Cheatham, *Reliable oligonucleotide conformational ensemble generation in explicit solvent for force field assessment using reservoir replica exchange molecular dynamics simulations*. *J. Phys. Chem. B.*, 2013. **117**(15): p. 4014-4027.
30. Bailor, M.H., et al., *3D maps of RNA interhelical junctions*. *Nat. Protoc.*, 2011. **6**(10): p. 1536-45.
31. 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.

## Figures

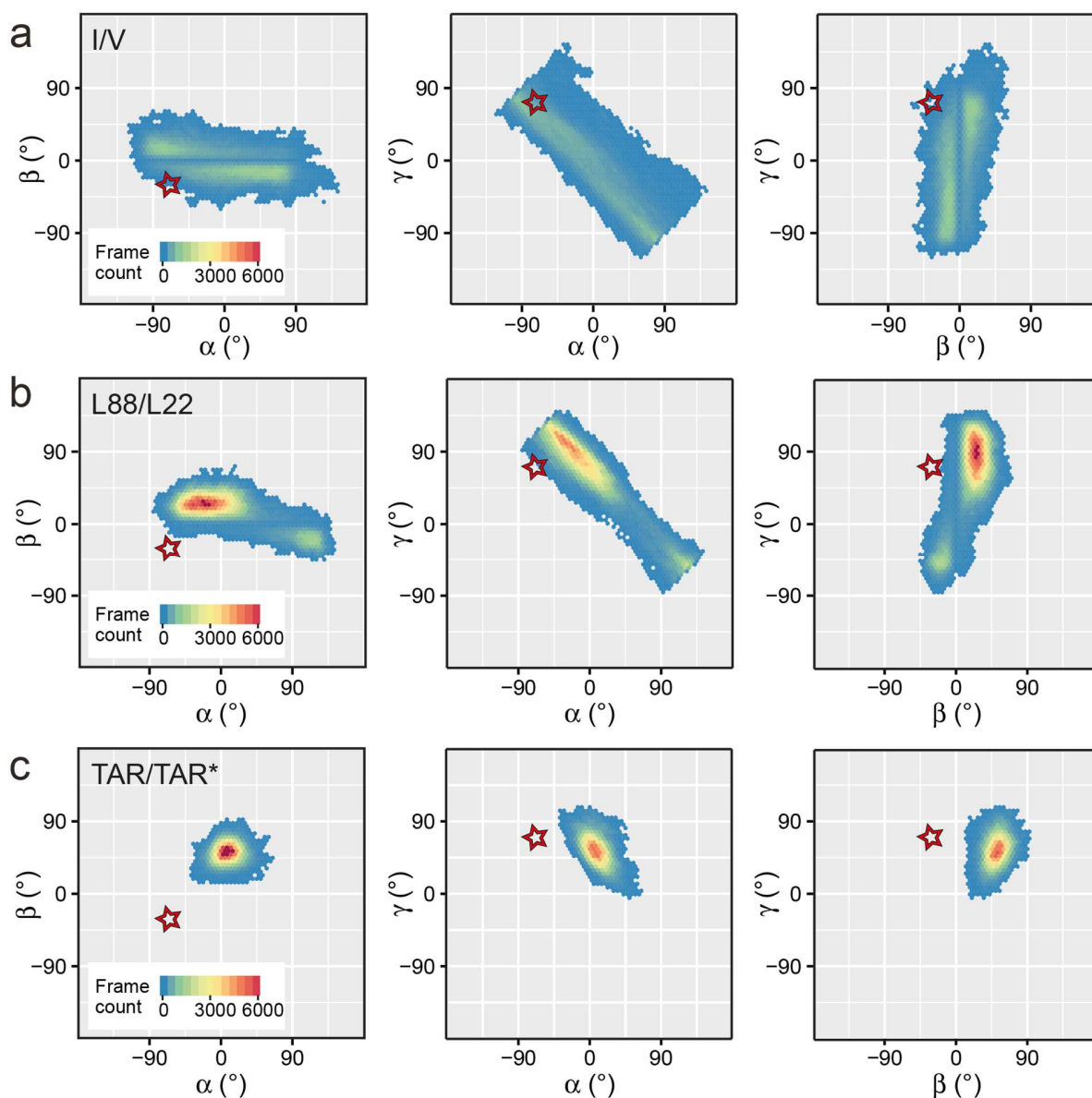


**Figure 4.1. Substitution of the I/V kissing-loop interaction (KLI) in the *Neurospora* VS ribozyme.** (a) Primary and secondary structures of the parental *trans* VS ribozyme ( $R_0$ ) and substrate ( $S_0$ ). Substrate recognition involves a KLI between loop I and loop V as well as a loop/loop interaction, which defines the active site, between the  $G_{638}$  loop and the  $A_{756}$  loop. (b) Representation of the open and closed states of the VS ribozyme. The closed state (right panel) is taken from the crystal structure of the *Neurospora* VS ribozyme substrate/ribozyme complex with SLI (green) and the SLV (pink) in sticks, helical domains II, III, IV and VI as a white surface,  $A_{756}$  in red sticks and the scissile phosphate as a green sphere. A representative of the many conformations that likely populate the open state (left panel) is illustrated similarly, but with the I/V KLI from the crystal structure replaced by the I/V KLI from the NMR structure of an SLI/SLV complex.[24] (c) The I/V KLI of the  $S_0/R_0$  ribozyme complex was previously substituted by rRNA L88/L22 KLI from the rRNA large subunit of *Deinococcus radiodurans* and by the HIV-1 TAR/TAR\* KLI.[16] Compared to the I/V KLI, these other KLI also display a small number of WC base pairs (bp) at the KLI, have similar

interhelical angles and are thermodynamically stable.[16] The  $k_{obs}$  values of the resulting VS ribozyme variants are given for the parental  $S_0/R_0$  pair as well as for the  $S_{L88+1}/R_{L22-1}$  and the  $S_{TAR-1}/R_{TAR*-0}$  variants.[16] Residues at the KLI, either free or involved in KL base pairs, are shown together with the loop closing base pairs.

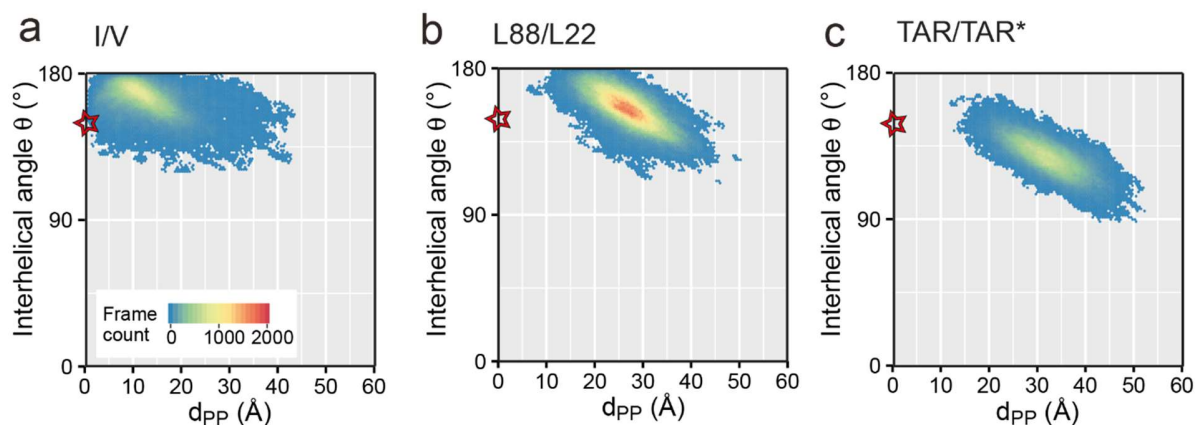


**Figure 4.2. Schematic representation of the parameters used to analyze the T-REMD trajectories.** (a) The Euler angles ( $\alpha$ ,  $\beta$  and  $\gamma$ ) were defined from the KL complexes as shown and the interhelical angle  $\theta$  is computed as  $180^\circ - |\beta|$ . (b) Each T-REMD frame of the trajectories (SLV in pink sticks and SLI in green sticks) was analyzed in the context of the VS ribozyme crystal structure (PDB entry 4R4P; *trans* ribozyme in white surface with A<sub>756</sub> in red stick and SLI in grey sticks), superposing the composite model of the frame on stem V as described in Experimental Procedures (Supplementary Information). In this context, the P-P distance ( $d_{PP}$ ) corresponds to the distance between the phosphorus atom of the scissile phosphates within the crystal structure (grey sphere) and the selected frame model (green sphere).

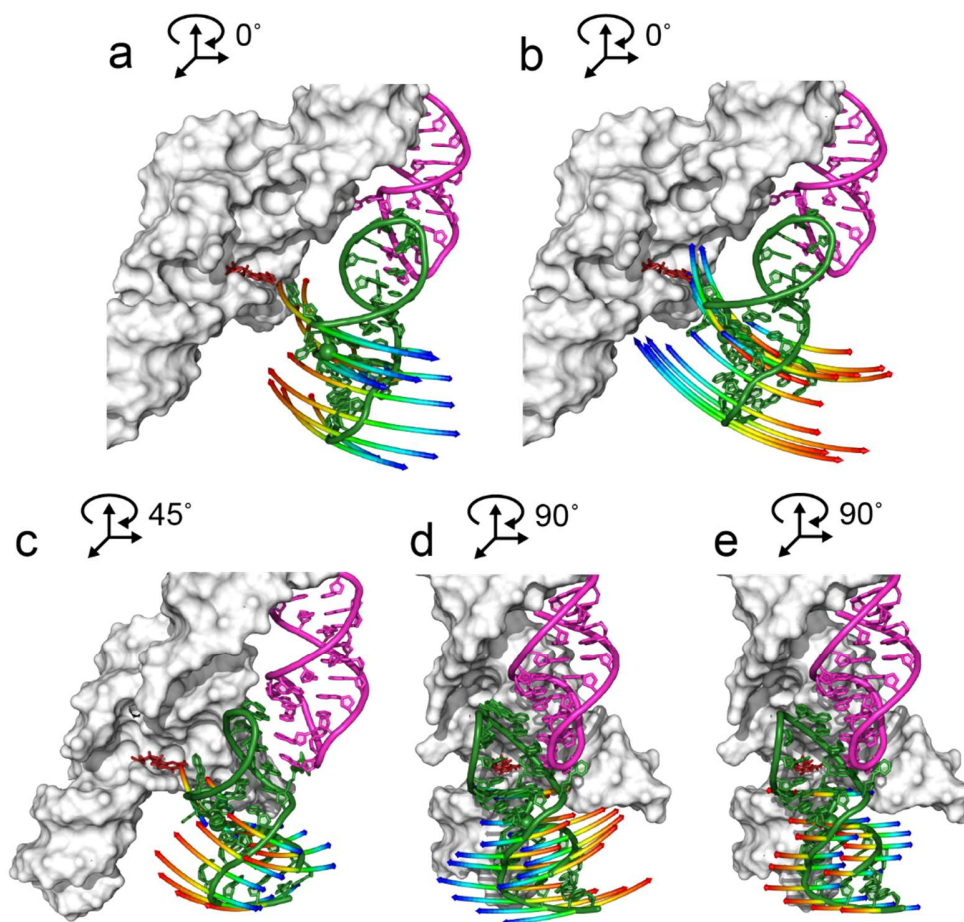


**Figure 4.3. Difference in conformational space between the I/V KL complex and surrogate KL complexes from Euler angle analysis.** 2D heat maps of Euler angle combinations representing all frames of the 300-K T-REMD trajectories are shown for the (a) I/V, (b) L88/L22 and (c) TAR/TAR\* KL complexes. In the 2D maps, the star points to the Euler angle values measured for the I/V KLI within the reference crystal structure (PDB entry 4R4P;  $\alpha = -69^\circ$ ;  $\beta = -31^\circ$ ; and  $\gamma = 71^\circ$ ).

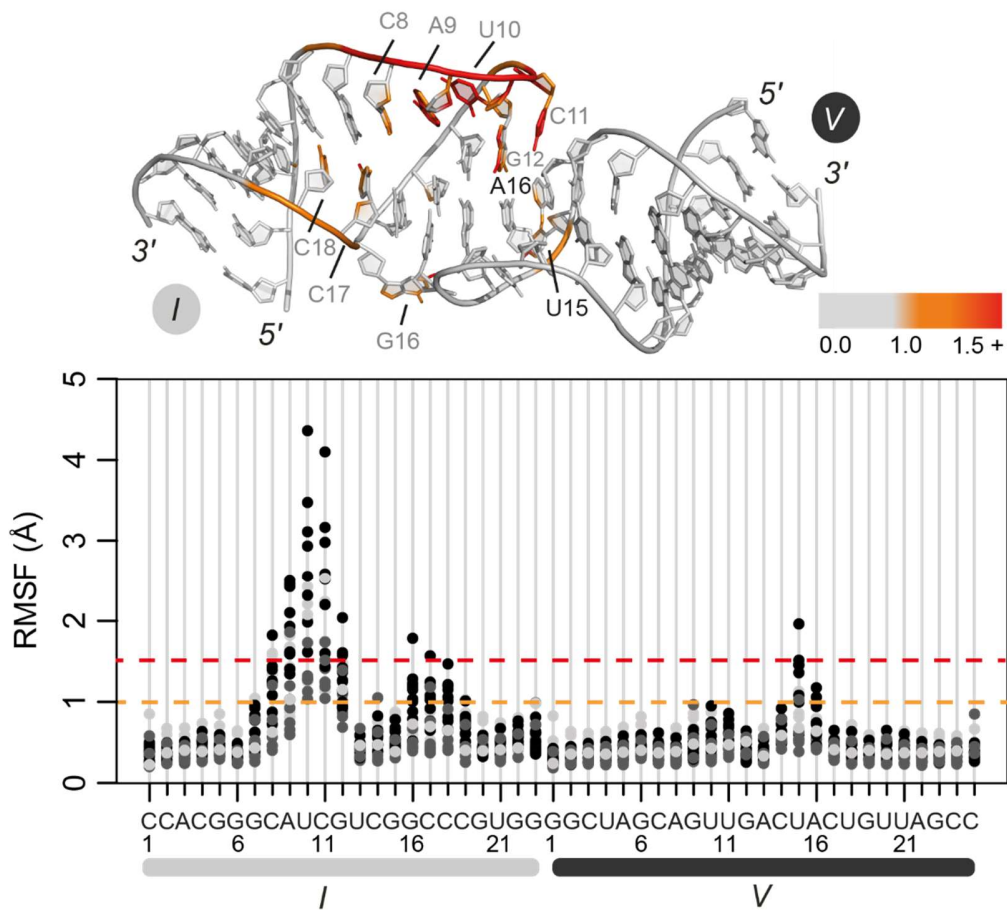




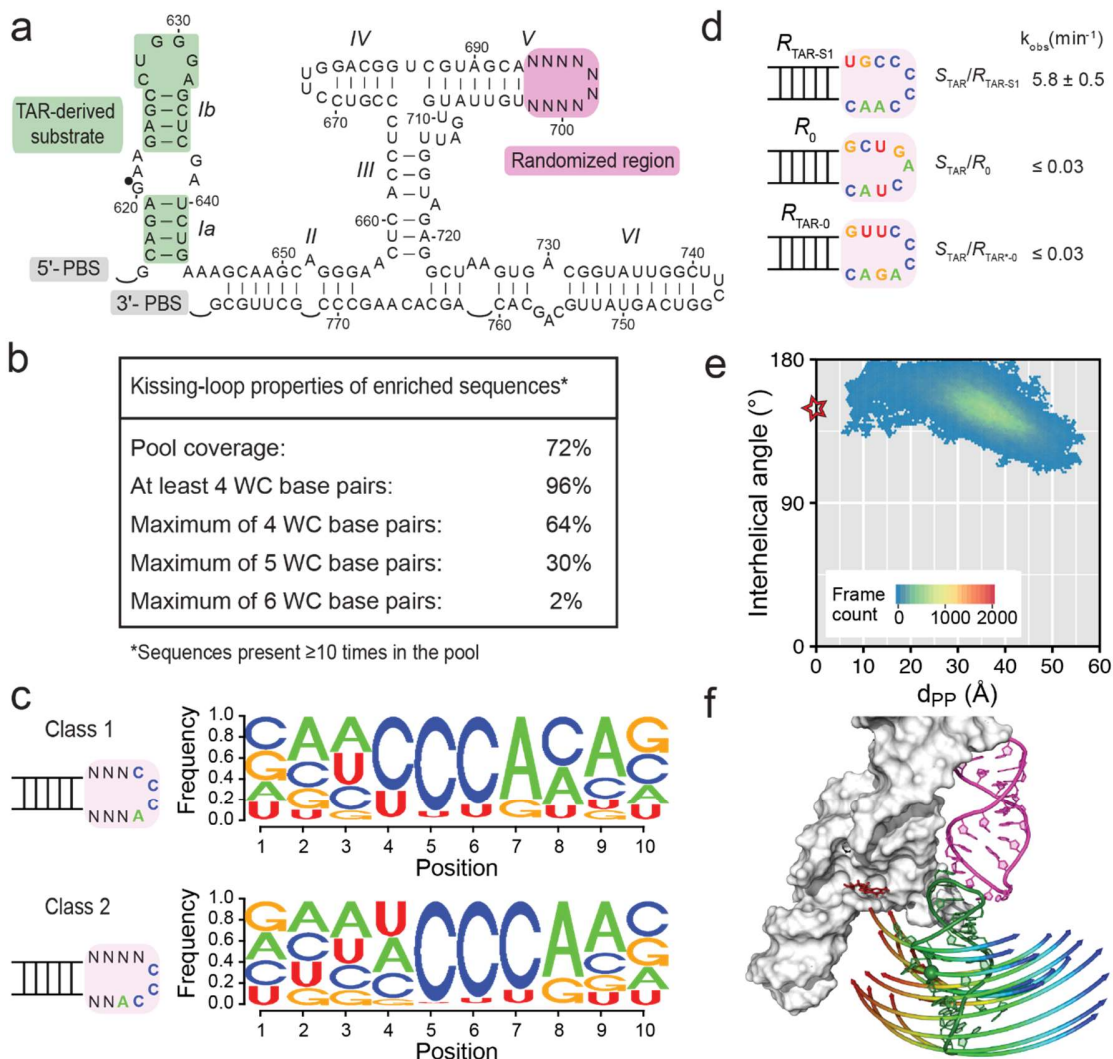
**Figure 4.4. Correlation between the interhelical angle  $\theta$  and the  $d_{PP}$  value for the I/V KL complex and surrogate KL complexes.** 2D heat maps of the interhelical angle  $\theta$  versus  $d_{PP}$  representing all frames of the 300-K T-REMD trajectories are shown for the (a) I/V, (b) L88/L22 and (c) TAR/TAR\* KL complexes. In the 2D maps, the star points to values measured for the I/V KLI within the reference crystal structure (PDB entry 4R4P;  $\theta = 149^\circ$  and  $d_{PP} = 0$ ).



**Figure 4.5.** Cartoon representations of the first five PCs of motion for the I/V KL complex. (a) The large twist of the first PC, (b) the large side-to-side bend of the second PC, (c) the small twist of the third PC and (d-e) the two small back-to-front bends of the fourth and fifth PCs. The PCs of motion were extracted from the 300-K T-REMD trajectory for the I/V KL complex and are represented in the context of the VS ribozyme X-ray structure (PDB entry 4R4P; in white surface with A<sub>756</sub> in red stick), by superposing interpolated frames from the pseudo-trajectory (SLV in pink and SLI in green) as described in Experimental Procedures (Supplementary Information). The movements of the backbone phosphorus atoms of SLI residues 616-623 and 637-643 are materialized as rainbow-colored arrows, blue representing one extreme position of the atom along the motion and red the other, along with a single interpolated frame taken at the halfway point of the pseudo-trajectory.



**Figure 4.6. Identification of mobile and rigid residues in the I/V KL complex.** RMSF values were computed for phosphate atoms (pale grey), ribose ring atoms (dark grey) and base heavy atoms (black) of the 300-K T-REMD trajectories. The RMSF values are mapped according to the displayed color scheme on the three-dimensional structure of the I/V KL complex (top panel) that was used to initiate the T-REMD calculation.



**Figure 4.7. Improving the cleavage activity of a variant VS ribozyme for its substrate using a more dynamic KLI.** (a) *In vitro* selection library to identify a *trans* VS ribozyme that efficiently cleaves a TAR-derived substrate. (b) Kissing-loop properties of enriched ribozyme sequences ( $\geq 10$  representatives in the selected pool) from the *in vitro* selection pool in terms of the number of potential Watson-Crick (WC; G-C, A-U and G-U) base-pairs between loop I and loop V. (c) Sequence preferences of the randomized region for the two classes of enriched ribozyme sequences. (d) SLV loop sequence of a selected ribozyme with high activity ( $R_{TAR-S1}$ ) and observed *trans* cleavage rate ( $k_{obs}$ ) of the  $S_{TAR}$  substrate. For comparison, the corresponding information is provided for control ribozymes. The  $k_{obs}$  values

were determined under the same conditions for the  $S_{TAR}/R_{TAR-S1}$  complex derived from *in vitro* selection as for control  $S/R$  complexes (shown here and in Fig. 4.1c). (e) 2D heat map of the interhelical angle  $\theta$  versus  $d_{PP}$  representing all frames of the 300-K T-REMD trajectory for the selected TAR/TAR-S1 KL complex. The star points to values measured for the I/V KLI within the reference crystal structure (PDB entry 4R4P;  $\theta = 149^\circ$  and  $d_{PP} = 0$ ). (f) Cartoon representation of the first PC of motion for the TAR/TAR-S1 KL complex represented in the context of the VS ribozyme X-ray structure as in Fig. 4.5.

## Tables

**Table 4.1. Characteristics of the first five PCs of motion for the I/V, L88/L22 and TAR/TAR\* KL complexes.**

		Principal components				
		1	2	3	4	5
I/V	% of variance	34.4	16.3	9.5	7.0	5.2
	Type of motion	Twist	Side-to-side	Twist	Back-to-front	Back-to-front
	Amplitude of	44	31	13	21	19
L88/L22	% of variance	24.4	20.8	13.5	8.9	5.3
	Type of motion	Bend <sup>b</sup>	Bend <sup>b</sup>	Twist	Back-to-front	Side-to-side
	Amplitude of	38	32	6	11	17
TAR/TAR*	% of variance	31.5	21.7	9.6	7.6	5.3
	Type of motion	Bend <sup>b</sup>	Side-to-side	Bend <sup>b</sup>	Twist	Twist
	Amplitude of	34	35	19	6	11

<sup>a</sup>The amplitude of motion was calculated as the distance travelled by the scissile phosphate in the pseudo-trajectory.

<sup>b</sup>It contains almost equivalent back-to-front and side-to-side components.

## Supporting Information

### Experimental procedures

**Computational tools and hardware.** All system minimization, equilibration and molecular dynamics (MD) simulations were carried out in GROMACS version 2016.3[1,2] on Compute Canada's Graham GPU server. For the equilibration and MD simulations, the Amber ff99SB[3] force field was used with the ParmBSC0 nucleic acid parameters[4,5] and chi<sub>ol</sub> corrections for glycosidic torsion angle chi[6] using explicit aqueous solvent. The Particle Mesh Ewald method was used to handle electrostatic interactions.[7,8] MC-Sym 3.3.2[9] and Pymol (The PyMOL Molecular Graphics System, Version 1.9.0 Schrödinger, LLC.) associated with custom Python scripts were used for system building and/or structural analysis.

**RNA system building.** Three RNA systems were initially built for molecular dynamic studies: the SLI/SLV, L88/L22 and TAR/TAR\* kissing-loop (KL) complexes. The starting conformation of the L88/L22 KL complex was obtained from PDB ID code 4IOA (chain A).[10] For the other two KL complexes, the starting conformation was obtained by extracting the lowest-energy structure of their representative NMR ensembles, namely the second model of PDB ID codes 2MI0[11] and 1KIS.[12] The hairpin stems within these three KL complexes were substituted by standard A-form stems generated with MC-Sym 3.3.2[9] through heavy-atom superposition of the first Watson-Crick base pair following the loop-closing base pair. The resulting hybrid structures (Supplementary Fig. S1a) retained the hairpin loop residues and closing base-pair [2MI0: residues 6-16 (I) and 108-114 (V); 1KIS: residues 5-12 (TAR) and 21-28 (TAR\*) and 4IOA: residues 423-430 (L22) and 2383-2392 (L88)]. Each RNA complex was solvated with approximately 20,000 TIP3P water molecules[13] in a periodic rhombic dodecahedral box and neutralized by addition of K<sup>+</sup> placed at randomized positions. In addition, K<sup>+</sup> and Cl<sup>-</sup> ions equivalent to a concentration of 0.15 M KCl were added to the solvent.[14] For T-REMD of the TAR/TAR-S1 KL complex derived from *in vitro* selection (Fig. 4.7), the RNA system was built by taking the hybrid structure of the TAR/TAR\* KL complex, replacing residues of TAR\* such that they match the TAR-S1 sequence (Fig. 4.7d) and modeling the four potential WC base pairs between the TAR and TAR-S1 loops.

**RNA system heating and equilibration.** Each RNA system was then copied several times, and each copy was energy minimized and equilibrated. The minimization was done in two successive steps, first by running 5,000 iterations of the steepest descent algorithm and then 5,000 iterations of the conjugate gradient algorithm, both with tight positional restraints (25 kcal mol<sup>-1</sup> Å<sup>-2</sup>) on the heavy atoms of the RNA. It was followed by three 200-ps equilibration steps: (1) using a standard MD simulation, the temperature was first raised to 150 K with the RNA heavy atoms under tight positional restraints (25 kcal mol<sup>-1</sup> Å<sup>-2</sup>); (2) the system was then heated to the target temperature (see below) with relaxed positional restraints (5 kcal mol<sup>-1</sup> Å<sup>-2</sup>); and (3) the target pressure of 1 bar was then attained by switching on the Berendsen barostat. Finally, the RNA system was simulated for 1 ns using standard MD without any restraint.

**Production T-REMD simulations.** The temperature replica exchange MD (T-REMD) simulations were run in several replicas at evenly distributed temperatures from 300 K to 375 K (inclusive bounds), to allow enhanced sampling of the conformational space. The number and temperature of the replicas were automatically generated for each KL complex to ensure an exchange rate of approximately 20% between adjacent replicas within the temperature scale.[15] This resulted in 58, 60 and 60 replicas for the SLI/SLV, L88/L22 and TAR/TAR\* KL complexes, respectively. The T-REMD simulations were run for 50 ns with 2-fs time steps, attempting exchange between replicas at each 500 steps and recording a frame for each replica every 200 fs. The resulting trajectories were collected and treated with the GROMACS trjconv tool to remove jumps in atomic coordinates introduced by the periodic boundaries conditions (PBC).

**Convergence assessment on demultiplexed trajectories.** To assess the convergence of the MD simulations,[16] the T-REMD trajectories were first demultiplexed to reconstruct continuous trajectories, each one following a single starting structure through the temperature space. The demultiplexing was done in two steps: the GROMACS demux.pl script was used to analyze the T-REMD trajectories and locate pieces of trajectory corresponding to each starting structure. The trjconv GROMACS tool was run multiple times with the output of demux.pl as reference to cut and paste together the relevant portions of trajectories. An RMSD profile was then built for each demultiplexed trajectory by calculating the pairwise heavy-atom RMSD of each frame with respect to the starting structure.



**Computation of Euler angles between the two stems of the KL complexes.** Each frame of the 300-K T-REMD trajectories was analyzed to derive the three Euler angles  $\alpha$ ,  $\beta$  and  $\gamma$  between the two stems of the KL complexes. The Euler angles were computed by adapting a Perl script previously developed for RNA interhelical junctions.[17] The interhelical angle  $\theta$  was derived by taking the supplementary angle of the  $\beta$  Euler angle. For comparative analysis, the  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\theta$  angles were also calculated for the VS ribozyme crystal structure (PDB ID code 4R4P[18]).

**Analysis of the T-REMD trajectories in the context of the crystal structure of the VS ribozyme.** To investigate the 300-K T-REMD trajectory of a given KL complex in the context of the crystal structure of the VS ribozyme, composite models were built that merged the KL interaction of an individual T-REMD frame with a substrate/ribozyme complex, termed  $S_X/R_X$ , taken from the crystal structure of the VS ribozyme A<sub>756</sub>G variant [PDB ID code 4R4P[18]], in which  $S_X$  is the SLI substrate of one protomer and  $R_X$  the helical domains II-VI of the other protomer. First, each frame of the T-REMD trajectories was aligned on the crystal structure by superposing the four core base pairs of stem V (residues 685-688 and 698-701 in the crystal structure; residues 5-8 and 18-21 in the SLI/SLV complex; residues 5-8 and 19-22 in the L88/L22 complex and residues 5-8 and 19-22 in the TAR/TAR\* complex). Subsequently, the internal loop and stem Ia of SLI from the crystal structure was added to the model, superposing stem Ib from the crystal structure (residues 623-625 and 635-637) on the corresponding residues of the SLI/SLV, L88/L22 and TAR/TAR\* KL complexes (residues 4-6 and 18-20, 4-6 and 17-19 and 4-6 and 17-19, respectively). From these composite models,  $d_{PP}$  values were computed by measuring the distance between phosphorus atoms of the scissile phosphate from the crystal structure and the model derived from a given frame of the T-REMD trajectory. From the same models, heavy atom RMSD values were computed between the G<sub>638</sub> loop and closing base pairs (residues 619-623 and 637-640) of the crystal structure and the frame-derived model.

**Principal component analysis of the T-REMD trajectories to characterize the molecular mechanics of KL complexes.** A principal component (PC) analysis was performed on the 300-K T-REMD trajectories to extract the PCs of motions and evaluate their contributions

to the overall dynamic of the KL complexes. The covar GROMACS tool was used to obtain and diagonalize the covariance matrix of the atomic coordinates from the T-REMD trajectories. The ana eig tool was then used to characterize the resulting pairs of eigenvector and eigenvalues by calculating for each eigenvector the two extreme projections along the trajectory on the average structure and interpolating frames to obtain a pseudo-trajectory along a single PC. The interpolated frames of the pseudo-trajectory were represented in the context of the VS ribozyme crystal structure by building composite models as described above for frames of the T-REMD trajectories.

**Root mean square fluctuation analysis of the T-REMD trajectories and identification of the residues responsible for the dynamics of the KL complexes.** The Root Mean Square Fluctuation (RMSF) is the measure of the mean displacement of atoms from their reference positions over time. To capture the dynamics of individual nucleotide with respect to its neighbors in the sequence, the heavy-atom RMSF was computed between the T-REMD starting structure and each frame of the trajectory for each pair of consecutive nucleotides using all frames of a given T-REMD trajectory.. Of the two RMSF values computed for each non-terminal nucleotide, the largest was retained. Heavy-atom RMSF values were calculated with the rmsf tool of GROMACS using the 300-K trajectory as input and the average atomic positions of the heavy atoms as the reference.

***In vitro* selection procedure.** All DNA sequences were obtained from IDT (Integrated DNA Technologies, IA) and gel purified, including the starting single-stranded DNA template library (5'-CCA AGA TGC CTG CAC TGT CGC AAG CGG GCT TGT GCT GTG CTG CAA TAC TGA CCG AAG CCA ATA CCG TCA CTT AGC CTC TAC CAA TCA CAT AAC A(10N)T GCT ACG ACC GTC CAA GGA CGG GAG GTG AGT TCC CTG CTT GCT TTC AGA TCG AGC TCC CAG GCT CTT CTC TGC ATG TCA TAT GTG CTG TGC C-3'), 5'-primer (5'-TAA TAC GAC TCA CTA TAG GCA CAG CAC ATA TGA CAT GCA GAG AAG AGC CTG GGA GCT CGA TCT GAA AGC AAG-3'), 3'-primer (5'-CCA AGA TGC CTG CAC TGT CGC AAG-3'), RT-primer (5'-CCA AGA TGC CTG CAC TGT-3'). Double-stranded DNA libraries were generated by PCR using 5'- and 3'-primers and the Pfu DNA polymerase (95°C for 5 min, followed by 8 cycles of 95°C for 30 s, 62°C for 1 min and 72°C for 3 min), extracted with phenol/chloroform and precipitated with 1/10 volume NaAc 3 M

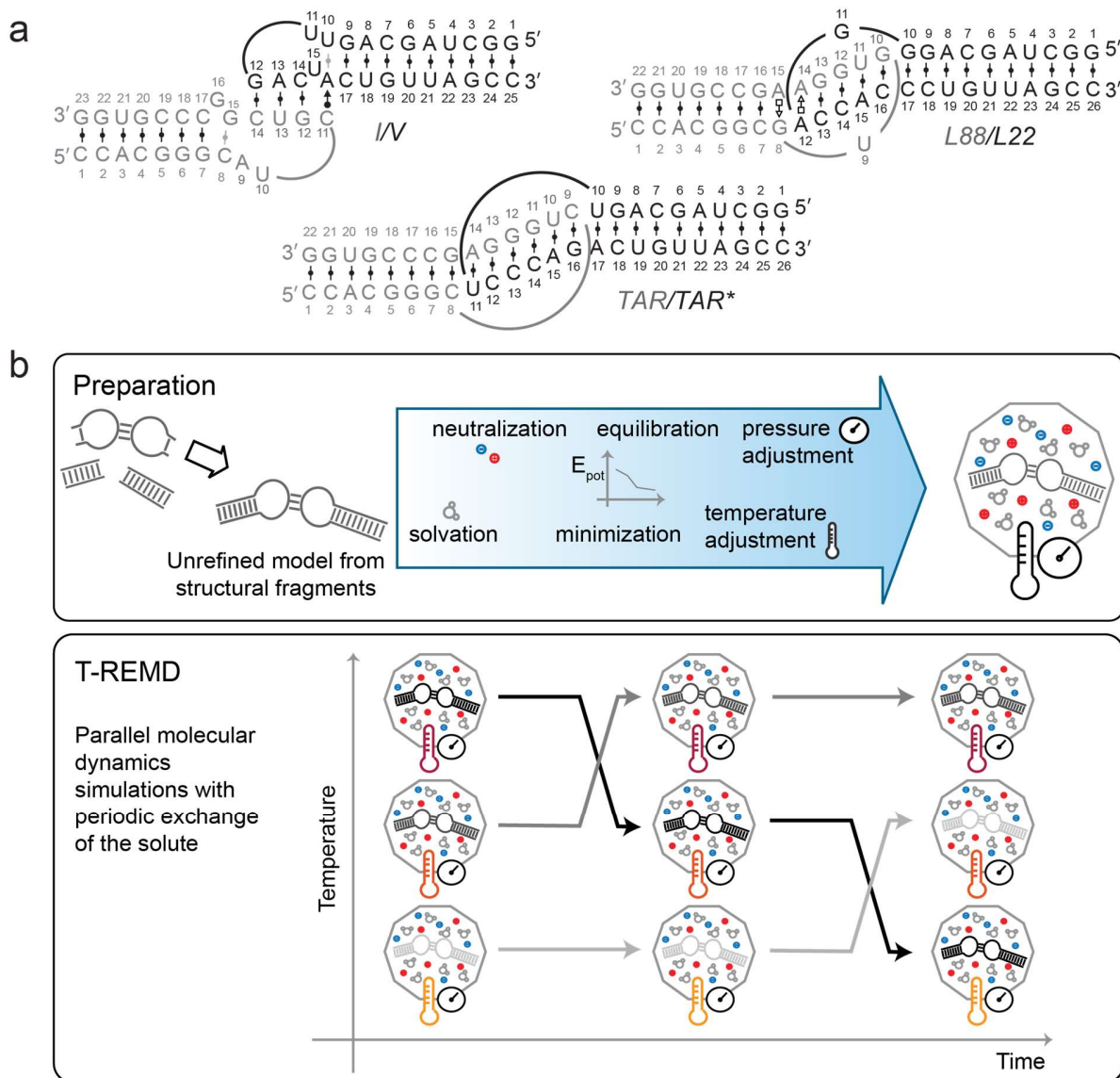
pH 5.3 and 3 volumes ethanol 95%. RNA libraries were prepared by *in vitro* transcription using T7 RNA polymerase in 1X Transcription buffer (40 mM Tris pH 7.6, 50 mM DTT, 1 mM spermidine, 0.1% triton, 7.5 mM MgCl<sub>2</sub> and 4 mM of each ATP, CTP, UTP and GTP) and incubation for 90 min at 37°C. To improve ribozyme cleavage of the active population, the reaction was incubated for an additional 1-20 min (20, 10, 10, 5, 2 and 1 min for cycles 1, 2, 3, 4, 5 and 6, respectively) after addition of 17.5 mM MgCl<sub>2</sub>. The large cleavage product was purified by denaturing gel electrophoresis (10% acrylamide:bisacrylamide (19:1) and 7 M urea). The RNA product was then reverse transcribed using Superscript III reverse transcriptase with the RT-primer, according to the manufacturer's protocol (Invitrogen, CA). The next rounds of selection were performed by following the same protocol, starting by regenerating the double-stranded DNA library, as described above.

**Next-generation sequencing and sequence analysis of selected ribozymes.** PCR products from the 6<sup>th</sup> round of selection were sequenced by Illumina MiSeq Spike (*McGill University and Génome Québec Innovation Centre, Montréal, Canada*) and analyzed using tools from the Galaxy Project.[19] Briefly, the raw fastq output was first filtered using a Phred quality score of 30, such that 141 511 reads were kept. The dataset was first analyzed to verify that there were no mutation within the ribozyme outside the randomized region. Sequences were then truncated, keeping 8 bases flanking both sides of the randomized region, and the data was processed to obtain each unique sequence and its number of reads. Sequences with 10+ reads were kept and analyzed in terms of their potential for contiguous Watson-Crick (WC) base pairing with loop I. Then, the 96.4% of sequences that could formed at least 4 base pairs with loop I were analyzed to identify the positions of loop V residues that could form the 4 base pairs, and from this analysis two main classes were identified (Fig. 7). For each class, position weight matrices were generated by counting the frequency of each base at every position within the randomized region and sequence logo representations were generated using a modified version of SeqLogo.[20]

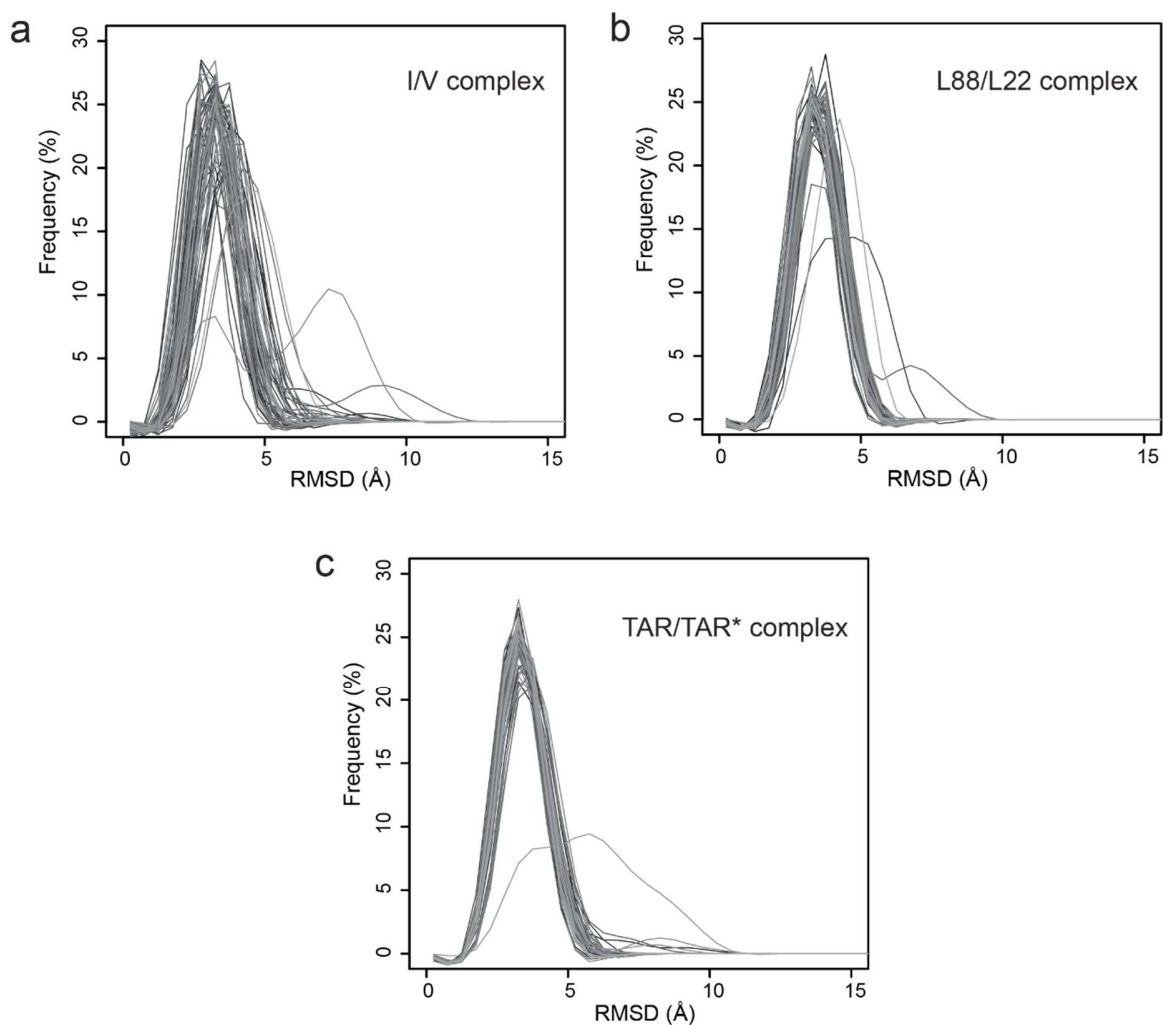
**Single-turnover kinetic studies of the  $S_{TAR}/R_{TAR-SI}$  complex.** The  $S_{TAR}/R_{TAR-SI}$  complex is a derivative of the  $S_0/R_0$  complex previously characterized kinetically,[21,22] modified in stem Ia and stem-loop Ib to match the HIV-1 TAR RNA sequence ( $S_{TAR}$ : 5'-GCA GAG AAG AGC CUG GGA GCU CGA UCU G-3') and in stem-loop V to match the selected  $R_{TAR-SI}$

sequence (Fig. 4.7d). The 5'-end- $^{32}\text{P}$ -labeled  $S_{\text{TAR}}$  and unlabeled *trans* ribozymes ( $R_0$ ,  $R_{\text{TAR}^*-0}$  and  $R_{\text{TAR-SI}}$ ) were prepared using established procedures.[22] Single-turnover kinetic studies with the  $S_{\text{TAR}}/R_{\text{TAR-SI}}$ ,  $S_{\text{TAR}}/R_0$  and  $S_{\text{TAR}}/R_{\text{TAR}^*-0}$  complexes were carried out at 37°C in 1X Kinetic Reaction buffer (50 mM Tris-HCl pH 7.6, 25 mM KCl, 2 mM spermidine and 25 mM  $\text{MgCl}_2$ ) to determine the values of the observed rate constant ( $k_{\text{obs}}$ ) using a substrate concentration ( $[\text{S}]$ ) of 250 pM and a ribozyme concentration ( $[\text{R}]$ ) of 0.5  $\mu\text{M}$ , as previously described.[22]

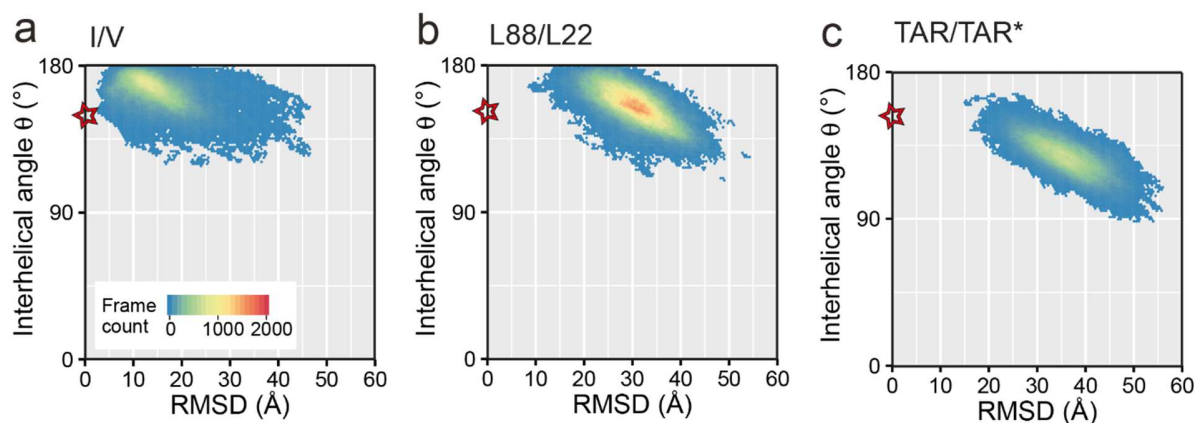
## Supporting figures



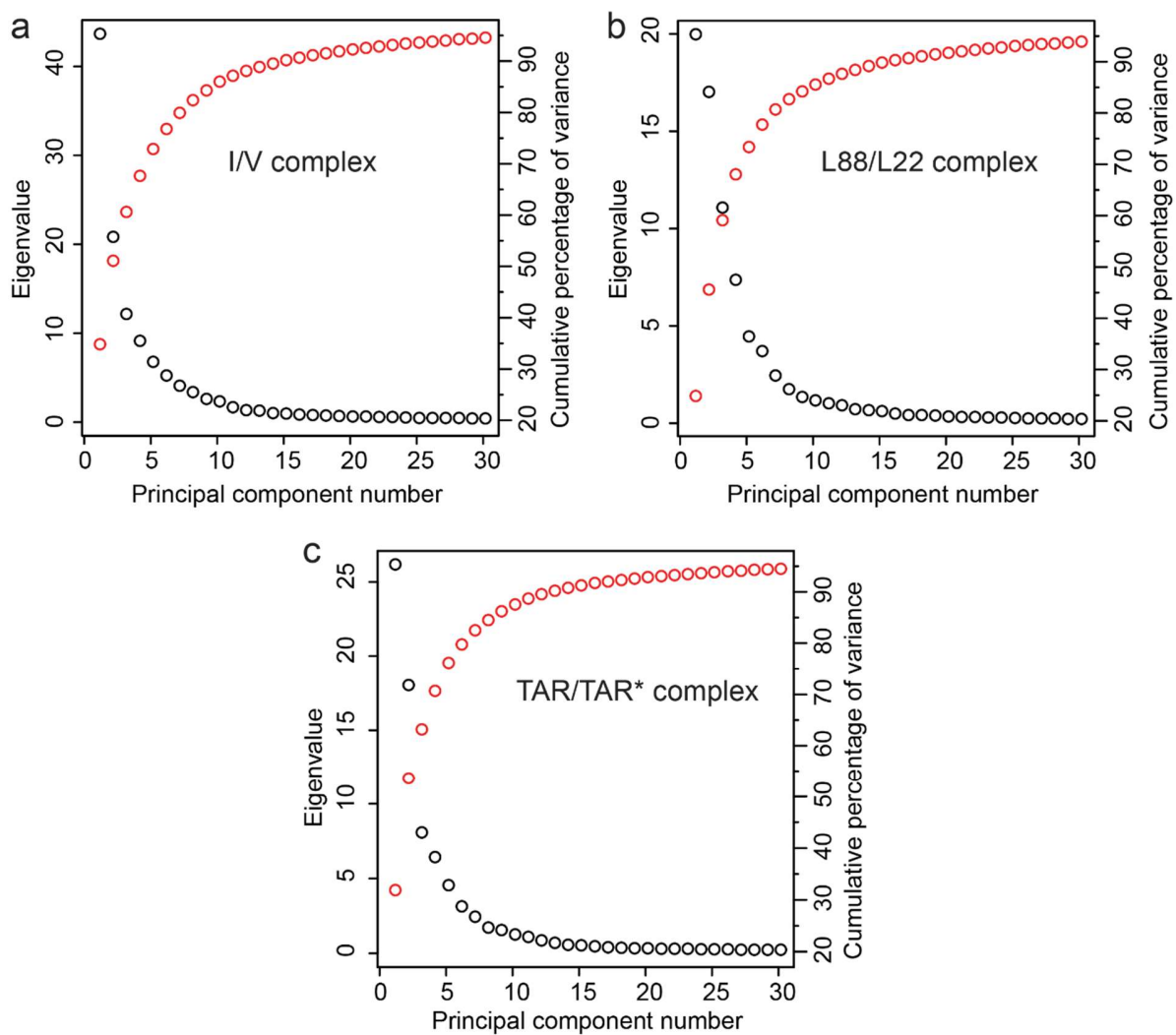
**Figure S4.1. Summary of the T-REMD protocol.** (a) Sequence and secondary structures of the KL complexes investigated by T-REMD simulations. (b) Schematics of the preparation step and T-REMD runs. In the preparation step, unrefined models of the KL complexes shown in (a) were first constructed by combining the deposited structures of each kissing-loop junction with standard stems. These structures were then solvated with explicit solvents, neutralized with  $K^+$  and  $Cl^-$  ions, supplemented with 100 mM KCl, energy minimized, and then subjected to short MD simulation runs during which the system was equilibrated and the temperature and pressure were adjusted. For the T-REMD simulations, 58-60 replicates were created at evenly-distributed temperatures across the temperature range (300-375 K, inclusively). For these replicates, parallel MD simulations were run for 50 ns with 2 fs time steps. Each 500 steps, there were attempts of exchanges of the solute (RNA complex) between neighbor replicas within the temperature ladder, and these attempts were successful when the energy of these replicas were sufficient close.



**Figure S4.2. Convergence analysis of the T-REMD simulations.** Frequency plots of the RMSD between the T-REMD starting structure and each frame of the demultiplexed trajectories for the (a) I/V (58 replicas), (b) L88/L22 (60 replicas) and (c) TAR/TAR\* (60 replicas) KL complexes. Each frequency plot contains data from all demultiplexed trajectories, which are trajectories that follow a given starting structure through the temperature space. During the 50-ns simulations, most trajectories appear to have sampled similar parts of the conformational space, suggesting that the simulations converged.

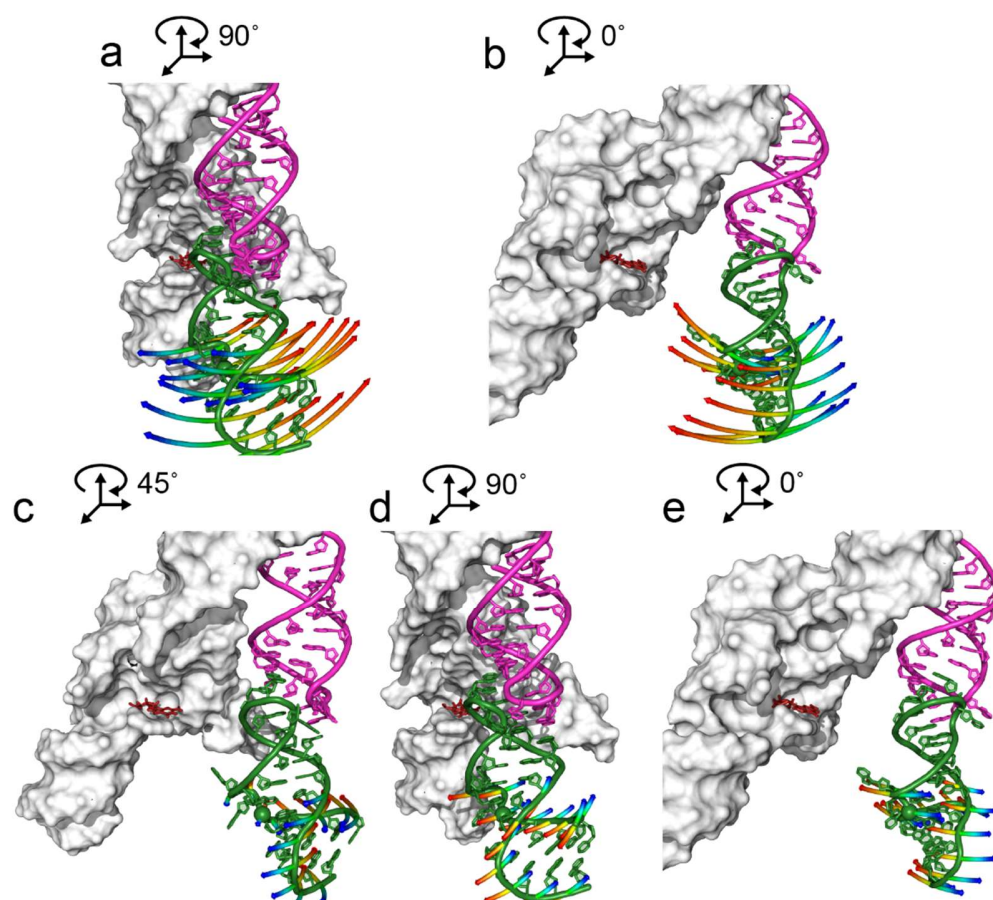


**Figure S4.3. Correlation between the interhelical angle  $\theta$  and the heavy-atom RMSD of the SLI internal loop for the I/V KL complex and surrogate KL complexes.** 2D heat maps of the interhelical angle  $\theta$  versus the heavy atom RMSD representing all frames of the 300-K T-REMD trajectories are shown for the (a) I/V, (b) L88/L22 and (c) TAR/TAR\* KL complexes. In the 2D maps, the star points to values measured for the I/V KLI within the reference crystal structure (PDB entry 4R4P;  $\theta = 149^\circ$  and RMSD = 0 Å).

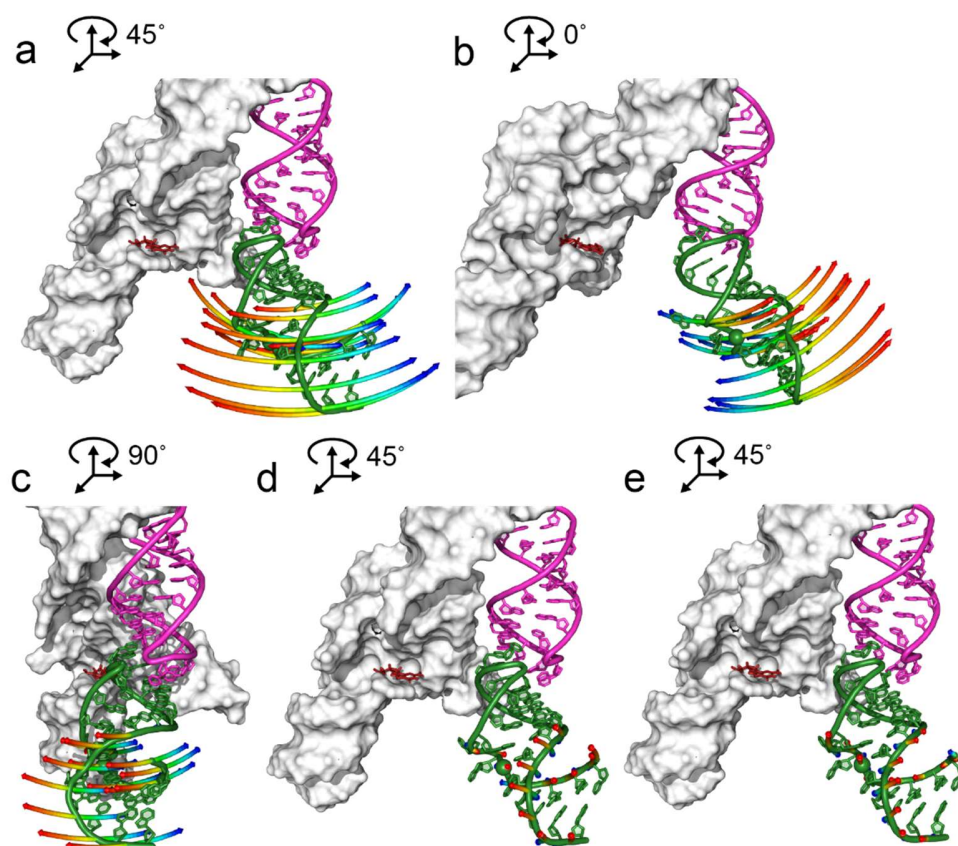


**Figure S.4.4. Scree plots of PC analysis for the T-REMD simulations.** (a) I/V, (b) the L88/L22 and (c) TAR/TAR\* KL complexes. The Scree plots show, for the first 30 PC, the eigenvalues (left axis and black circles) as well as the cumulative percentage of variance (right axis and red circles).

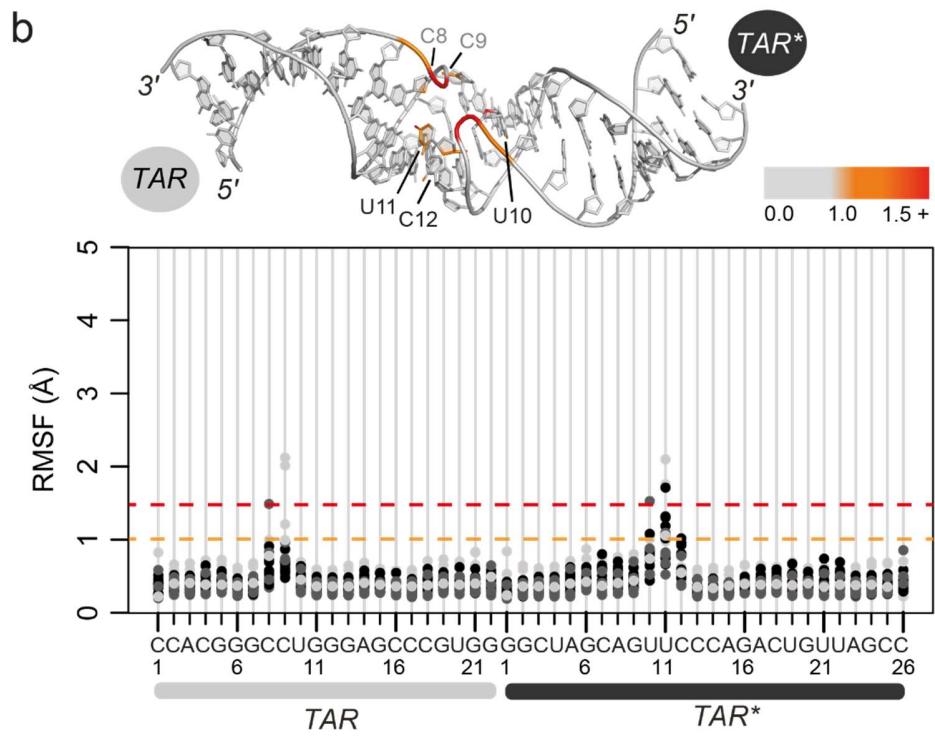
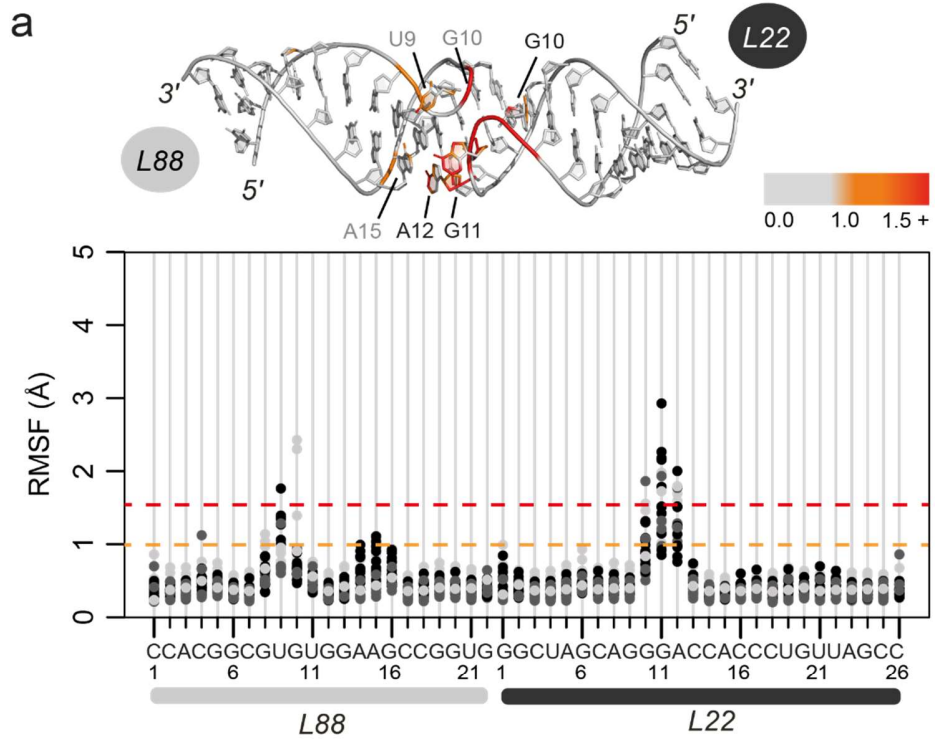




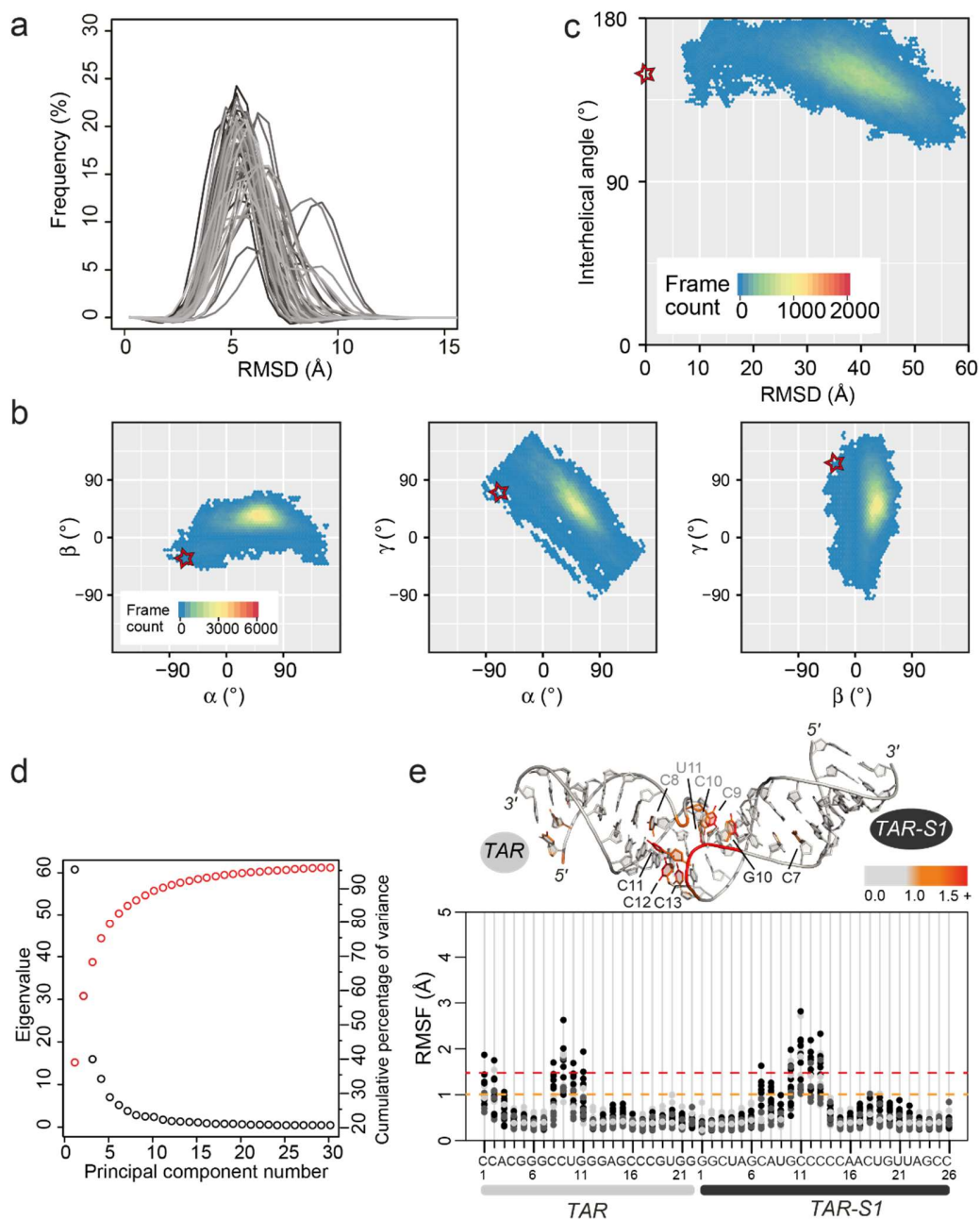
**Figure S4.5. Cartoon representations of the first five PCs of motion for the L88/L22 KL complex.** (a-b) The large bends of the first and second PCs, (c) the small twist of the third PC, (d) the small back-to-front bend of the fourth PC and (e) the small side-to-side bend of the fifth PC. The PCs of motion were extracted from the 300-K T-REMD trajectory for the L88/L22 KL complex and are represented in the context of the VS ribozyme X-ray structure (PDB entry 4R4P; in white surface with A<sub>756</sub> in red stick), by superposing interpolated frames from the pseudo-trajectory (SLV in pink and SLI in green) as described in Experimental Procedures. The movements of the backbone phosphorus atoms of SLI residues 616-623 and 637-643 are materialized as rainbow-colored arrows along with a single interpolated frame taken at the halfway point of the pseudo-trajectory.



**Figure S4.6. Cartoon representations of the first five PCs of motion for the TAR/TAR\* KL complex.** (a) The large bend of the first PC, (b) the large side-to-side bend of the second PC, (c) the small bend of the third PC and (d-e) the two small back-to-front twists of the fourth and fifth PCs. The PCs of motion were extracted from the 300-K T-REMD trajectory for the TAR/TAR\* KL complex and are represented in the context of the VS ribozyme X-ray structure (PDB entry 4R4P; in white surface with A<sub>756</sub> in red stick), by superposing interpolated frames from the pseudo-trajectory (SLV in pink and SLI in green) as described in Experimental Procedures. The movements of the backbone phosphorus atoms of SLI residues 616-623 and 637-643 are materialized as rainbow-colored arrows along with a single interpolated frame taken at the halfway point of the pseudo-trajectory.



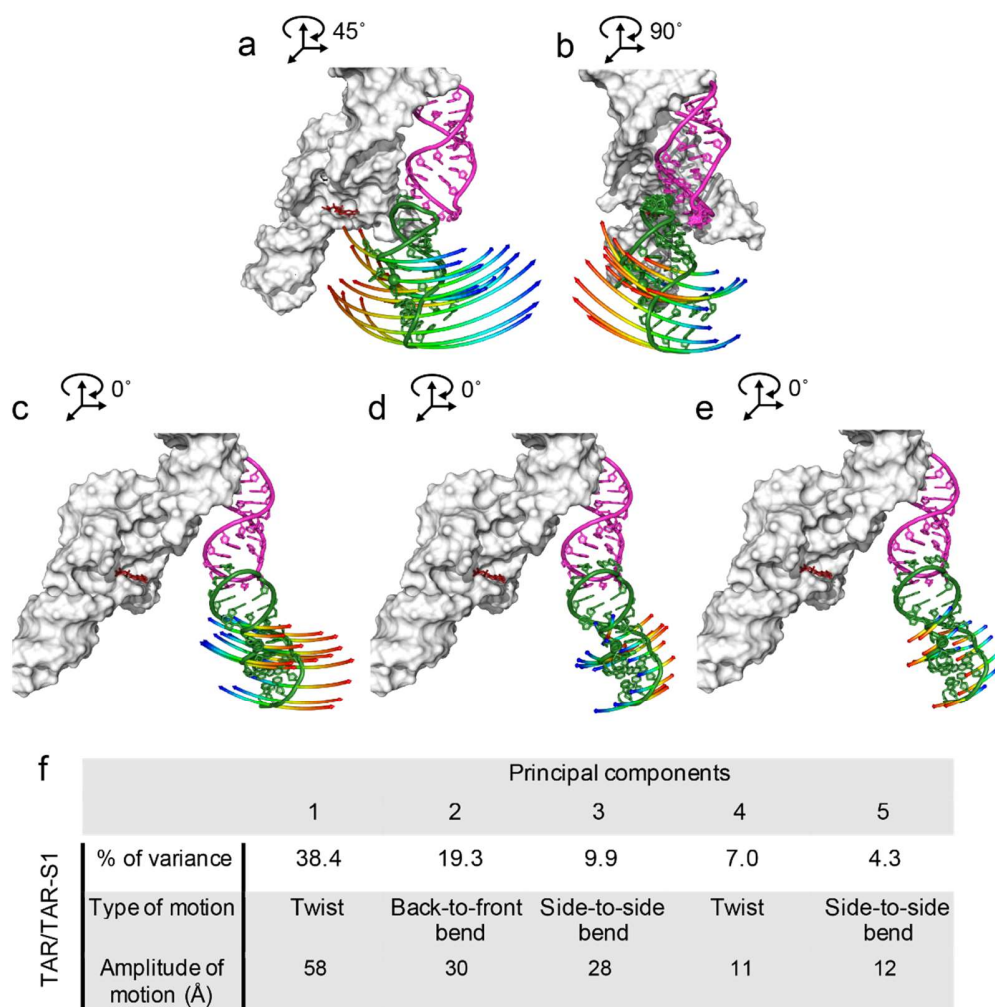
**Figure S4.7. Identification of mobile and rigid residues** in the (a) L88/L22 and (b) TAR/TAR\* KL complexes. RMSF values relative to the starting structure were computed for phosphate atoms (pale grey), ribose ring atoms (dark grey) and base heavy atoms (black) of the 300-K T-REMD trajectories. The RMSF values are mapped according to the displayed color scheme on the three-dimensional structure of the respective KL complex (top panel) that was used to initiate the T-REMD calculation.



**Figure S4.8. Results from T-REMD simulations of the TAR/TAR-S1 KL complex.** (a) Convergence analysis of the T-REMD simulations shown as a frequency plot of the RMSD between the T-REMD starting structure and each frame of the demultiplexed trajectories. (b) 2D heat maps of Euler angle combinations representing all frames of the 300-K T-REMD trajectories. The star points to the Euler angle values measured for the I/V KLI within the reference crystal structure (PDB entry 4R4P;  $\alpha = -69^\circ$ ;  $\beta = -31^\circ$ ; and  $\gamma = 71^\circ$ ). (c) 2D heat

map of the interhelical angle  $\theta$  versus the heavy atom RMSD representing all frames of the 300-K T-REMD trajectories. The star points to values measured for the I/V KLI within the reference crystal structure (PDB entry 4R4P;  $\theta = 149^\circ$  and RMSD = 0 Å). (d) Scree plots of PC analysis showing for the first 30 PC, the eigenvalues (left axis and black circles) as well as the cumulative percentage of variance (right axis and red circles). (e) RMSF values computed for phosphate atoms (pale grey), ribose ring atoms (dark grey) and base heavy atoms (black) of the 300-K T-REMD trajectories. The RMSF values are mapped according to the displayed color scheme on the three-dimensional structure of the KL complex (top panel) that was used to initiate the T-REMD calculation.





**Figure S4.9. Cartoon representations of the first five PCs of motion for the TAR/TAR-S1 KL complex.** (a) The large twist of the first PC, (b) the large back-to-front bend of the second PC, (c) the large side-to-side bend of the third PC, (d) the small twist of the fourth PC and (e) the small side-to-side bend of the fifth PC and (f) the characteristics of the first five PCs. The PCs of motion were extracted from the 300-K T-REMD trajectory for the TAR/TAR-S1 KL complex and are represented in the context of the VS ribozyme X-ray structure (PDB entry 4R4P; in white surface with A<sub>756</sub> in red stick), by superposing interpolated frames from the pseudo-trajectory (SLV in pink and SLI in green) as described in Experimental Procedures. The movements of the backbone phosphorus atoms of SLI residues 616-623 and 637-643 are materialized as rainbow-colored arrows along with a single interpolated frame taken at the

halfway point of the pseudo-trajectory. In (f), the amplitude of motion was calculated as the distance travelled by the scissile phosphate in the pseudo-trajectory.



## References

1. Pronk, S., et al., *GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit*. *Bioinformatics.*, 2013. **29**(7): p. 845-54.
2. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers*. *SoftwareX.*, 2015. **1-2**: p. 19-25.
3. Hornak, V., et al., *Comparison of multiple Amber force fields and development of improved protein backbone parameters*. *Proteins.*, 2006. **65**(3): p. 712-25.
4. Perez, A., et al., *Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers*. *Biophys. J.*, 2007. **92**(11): p. 3817-29.
5. Guy, A.T., T.J. Piggot, and S. Khalid, *Single-stranded DNA within nanopores: conformational dynamics and implications for sequencing; a molecular dynamics simulation study*. *Biophys. J.*, 2012. **103**(5): p. 1028-36.
6. Zgarbova, M., et al., *Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles*. *J Chem. Theory Comput.*, 2011. **7**(9): p. 2886-2902.
7. Darden, T., et al., *New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations*. *Structure (London, England : 1993)*, 1999. **7**(3): p. R55-60.
8. Essmann, U., et al., *A smooth particle mesh Ewald method*. *J.Chem. Phys.*, 1995. **103**(19): p. 8577-8593.
9. Major, F., et al., *The combination of symbolic and numerical computation for three-dimensional modeling of RNA*. *Science.*, 1991. **253**(5025): p. 1255-60.
10. Magee, T.V., et al., *Novel 3-O-carbamoyl erythromycin A derivatives (carbamolides) with activity against resistant staphylococcal and streptococcal isolates*. *Bioorg. Med. Chem. Lett.*, 2013. **23**(6): p. 1727-31.

11. 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. **53**(1): p. 258-69.
12. 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.
13. Jorgensen, W.L., et al., *Comparison of Simple Potential Functions for Simulating Liquid Water*. *J. Chem. Phys.*, 1983. **79**(2): p. 926-935.
14. Auffinger, P., *Ions in molecular dynamics simulations of RNA systems*, in *RNA 3D Structure Analysis and Prediction*, V. Springer, Editor. 2012. p. 299-318.
15. Patriksson, A. and D. van der Spoel, *A temperature predictor for parallel tempering simulations*. *Phys. Chem. Chem. Phys.*, 2008. **10**(15): p. 2073-2077.
16. Henriksen, N.M., D.R. Roe, and T.E. Cheatham, *Reliable oligonucleotide conformational ensemble generation in explicit solvent for force field assessment using reservoir replica exchange molecular dynamics simulations*. *J. Phys. Chem. B.*, 2013. **117**(15): p. 4014-4027.
17. Bailor, M.H., et al., *3D maps of RNA interhelical junctions*. *Nat. Protoc.*, 2011. **6**(10): p. 1536-45.
18. Suslov, N.B., et al., *Crystal structure of the Varkud satellite ribozyme*. *Nat. Chem. Biol.*, 2015. **11**(11): p. 840-6.
19. Afgan, E., et al., *The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update*. *Nucleic Acids Res.*, 2016. **44**(W1): p. W3-w10.
20. Bembon, O. *seqLogo: Sequence logos for DNA sequence alignments*. R package version 1.46.0., 2018.
21. Lacroix-Labonté, J., et al., *Helix-length compensation studies reveal the adaptability of the VS ribozyme architecture*. *Nucleic Acids Res.*, 2011. **40**(5): p. 2284-93.

22. Lacroix-Labonté, J., et al., *Rational engineering of the Neurospora VS ribozyme to allow substrate recognition via different kissing-loop interactions*. *Nucleic Acids Res.*, 2016. **44**(14): p. 6924-34.

## **5. Discussion**

### **5.1. Conclusions majeures des articles présentés**

#### **5.1.1. La conformation active implique un angle entre les hélices I et V proche de la colinéarité**

La capacité du ribozyme VS de cliver des substrats de longueur différente de son substrat naturel (promiscuité) ainsi que la possibilité pour des ribozymes dont les hélices sont allongées ou raccourcies de cliver des substrats dont la longueur les rendaient non clivables par le ribozyme naturel (compensation de longueur d'hélice) mènent toutes deux à la même conclusion, confirmée par l'analyse statistique menée sur des modèles de la structure 3D du complexe I/V (chapitre 2): afin d'être actif, le ribozyme VS doit être pourvu d'une interaction *kissing-loop* I/V qui permet à l'angle entre les axes des hélices V et I de se maintenir ou d'osciller autour d'une valeur proche de la colinéarité. Cette observation a depuis été confirmée à plusieurs reprises, en particulier par les travaux de RMN réalisés sur ce complexe [96] et par les structures obtenues par cristallographie [65, 73]. Dans ces dernières, le ribozyme VS est dans un état fermé dans lequel le site actif est formé et l'angle entre les hélices I et V a été calculé aux alentours de 150°. Les angles interhélicaux dans l'ensemble de RMN vont quant à eux de 127° à 175°, ce qui semble indiquer un complexe très dynamique, capable d'atteindre l'angle interhélicale requis pour la formation du site actif. Enfin, l'article présenté au chapitre 4 démontre que les capacités dynamiques intrinsèques du complexe I/V naturel semblent le restreindre à une gamme d'angles interhélicaux proche de celle observée dans l'ensemble de RMN lorsqu'il est dans une conformation compatible avec la formation du site actif.

#### **5.1.2. Le potentiel d'ingénierie du ribozyme VS et ses conséquences**

Nos travaux ont révélé l'adaptabilité du ribozyme VS en ce qui a trait à sa reconnaissance de substrats différents de son substrat naturel. Que l'on modifie la longueur du substrat, de l'hélice V indispensable à sa reconnaissance (chapitre 2) ou les structures primaires, secondaires et tertiaires de l'interaction boucle-boucle (chapitre 3) elles-mêmes, il est dans une certaine

mesure possible de conserver une activité de clivage, qui peut parfois se comparer favorablement à celle du ribozyme naturel. Ces résultats ont mis en évidence la modularité du ribozyme VS mais également son caractère fortement dynamique. En effet, une vision statique de l'interaction ne peut expliquer ni les résultats de substitution de l'interaction, ni ceux de compensation de longueur d'hélice. D'une part, l'alignement des structures des complexes boucle-boucle sur la structure du ribozyme complet, toutes obtenues par des méthodes biophysiques, offre la vision d'un état ouvert du ribozyme dans lequel les deux parties du site actif sont loin d'être en contact. D'autre part, le simple examen visuel des complexes boucle-boucle de longueurs différentes révèle une évidence : le décalage latéral et l'angle inférieur à 180° entre les axes des hélices I et V (Figure 2.5) implique qu'une modification du nombre de paires de bases de l'hélice V entraîne une combinaison de rotation et de translation de la boucle interne du substrat qu'il n'est pas possible de compenser par une simple modification de la longueur de l'hélice I sans dynamique.

### **5.1.3. La dynamique de l'interaction I/V est distinctive et suffisante pour la formation du site actif**

Comme le montrent les travaux présentés au chapitre 4, la différence majeure entre l'interaction naturelle I/V et les variantes identifiées par la recherche bio-informatique se situe au niveau de leur dynamique intrinsèque. Le complexe boucle-boucle naturel possède plus de résidus non-canoniquement appariés et mobiles dans la boucle terminale du substrat qui par conséquent peut adopter deux conformations, une conformation compacte et une conformation relâchée. Cette transition est à l'origine d'un mouvement particulier de rotation de grande amplitude de l'hélice du substrat sur elle-même qui n'est pas retrouvée chez les variantes. La transition contribue également aux autres modes de mouvement principaux du complexe boucle-boucle naturel, soit des mouvements d'ouverture et de fermeture du complexe avec l'interaction comme charnière qui causent des translations de la boucle interne du substrat vers la boucle A756 de l'hélice VI et le long du corps du ribozyme.

Les travaux de résolution de structure entrepris dans notre laboratoire ont mené à l'élaboration d'un modèle dit « ouvert » du ribozyme VS représentatif d'un ensemble de conformations adoptées en solution dans lesquelles le site actif n'est pas formé, ses deux

composantes, la boucle A756 et la boucle interne du substrat en *trans*, se trouvant séparés par un peu moins de 30 Å. Théoriquement, le passage d'une conformation ouverte à la conformation fermée, active, représentée par la structure cristalline, peut être le fruit de la dynamique de nombreux sous-domaines du ribozyme [75], en particulier les jonctions II-III-VI et III-IV-V et l'interaction I/V. Nos études montrent que le complexe I/V naturel se meut intrinsèquement d'une manière qui permet le rapprochement de la boucle interne du substrat de la boucle A756 de l'hélice VI, grâce à une combinaison complexes de mouvements selon plusieurs axes qui permettent au substrat à la fois de se rapprocher du corps du ribozyme et d'opérer des translations et rotations afin de pouvoir présenter sa boucle interne dans la position la plus appropriée à la formation du site actif.

Ce point a mis en lumière deux principes intéressants : d'une part la très grande différence de capacités dynamiques intrinsèques entre des interactions *kissing-loop* somme toute structurellement très proches. D'autre part, le caractère contre-instinctif du fait de perdre en efficacité fonctionnelle lorsque l'on remplace un sous-domaine très flexible par un sous-domaine similaire qui possède une dynamique intrinsèque plus restreinte.

La différence est telle entre les dynamiques intrinsèques du complexe I/V et ses variantes qu'il est surprenant que ces dernières soient effectivement actives. L'explication réside peut-être dans un phénomène de compensation dynamique par d'autres sous-domaines du ribozyme, par exemple au niveau de la jonction III-IV-V, ou dans des réarrangements des boucles terminales du substrat et/ou de l'hélice V à des échelles de temps plus longues. Des événements rares ont en effet été observés dans les simulations en dynamique moléculaire du complexe TAR/TAR\* durant lesquels des paires de bases de l'interaction se défont et les deux boucles impliquées adoptent des conformations plus relâchées. Des travaux complémentaires, basés sur des simulations plus longues et des simulations intégrant plusieurs sous-domaines du ribozymes VS pourraient permettre d'éclaircir la contribution de ces phénomènes à l'activité des variantes.

#### **5.1.4. L'activité du ribozyme est principalement dépendante de la dynamique de l'interaction I/V**

Les trois interactions boucle-boucle testées ont des valeurs d'affinités proches et ont été spécifiquement sélectionnées pour être structurellement similaires, tant au niveau du nombre de

paires de bases, de la taille des boucles impliquées et de l'angle entre les hélices connectées. Pourtant, l'activité des ribozymes les plus actifs dans lesquels l'interaction I/V a été substituée par une variante est environ 50 fois moindre que celle du ribozyme VS naturel. Ces données couplées à l'analyse des différences de dynamique intrinsèque semble indiquer que le principal facteur limitant est la probabilité que l'interaction explore l'espace conformationnel dans laquelle la formation du site actif est possible. Cela conduit à une réévaluation du protocole de sélection des candidats à la substitution de l'interaction boucle-boucle. Bien qu'initialement il paraissait important d'identifier des candidats dont les caractéristiques structurales se rapprochaient de l'interaction naturelle, il s'avère qu'il est en fait plus important d'identifier les modes de mouvement favorable à la formation du site actif puis de sélectionner des candidats dont les caractéristiques dynamiques intrinsèques sont similaires ou meilleures en termes de probabilité d'exploration des conformations favorables.

Le succès des expériences de sélection *in vitro* focalisées sur la modification de la boucle V ont ajouté une composante additionnelle à la façon d'envisager l'ingénierie. Il semble en effet que conserver les modalités dynamiques de l'ensemble de l'interaction I/V intacte suffise. Que l'essentiel de la dynamique provienne de la boucle I dans le complexe I/V naturel ne semble pas important et cette dynamique peut être reportée ailleurs, dans cet exemple sur la boucle V. Cette observation est encourageante pour d'éventuels futurs travaux d'ingénierie puisque dans l'optique de modifier le ribozyme VS pour lui permettre d'accommoder des substrats d'intérêt biologiques, il n'est bien sûr pas possible de modifier le substrat.

### **5.1.5. Des différences structurales importantes reposent sur des mouvements locaux de faible amplitude**

Les mouvements des complexes I/V seul observés en dynamique moléculaire présentent des déplacements importants de la boucle interne du substrat au cours de sa dynamique (plus de 40 Å dans certains cas). Or, l'étude de l'origine locale de ces déplacements révèle qu'ils proviennent du mouvement coordonné et de faible amplitude d'un nombre réduit de résidus non-canoniquement appariés, situés dans la boucle du substrat. Cette observation est intéressante du point de vue de l'ingénierie. Elle démontre qu'il est possible d'améliorer l'activité d'un ARN

en modifiant son profil dynamique sans recourir à des modifications drastiques de ses structures primaires et secondaires.

### **5.1.6. Le modèle dynamique du complexe I/V est prédictif de l'activité du ribozyme VS**

L'évaluation des capacités dynamiques du complexe I/V ne sont qu'un préalable à la construction d'un modèle dynamique complet du ribozyme VS. Néanmoins, il apparaît que ce modèle dynamique, aussi incomplet qu'il soit pour le moment, permette de prédire l'effet de certaines modifications. La comparaison des capacités dynamiques des différentes interactions testées met en lumière la participation de résidus non-canoniquement appariés dans la boucle terminale du substrat. Par extension, nous avons émis l'hypothèse que, pour les variantes, il devrait être possible d'obtenir de meilleure activité en empêchant quelques paires canoniques dans l'interaction de se faire et ainsi obtenir des boucles terminales avec des résidus libres. Les résultats de la sélection *in vitro* ont confirmé cette hypothèse en révélant que l'essentiel des candidats identifiés à l'issue de la sélection possédait effectivement moins de paires de bases que l'interaction TAR/TAR\* originale. Ce caractère prédictif demanderait à être testé plus en profondeur en substituant l'interaction boucle-boucle par d'autres candidats dont les structures sont connues et en tentant de prédire leur activité.

## **5.2. Corollaires des résultats obtenus**

### **5.2.1. Vers un modèle dynamique complet du ribozyme VS**

L'un des objectifs à long terme de notre laboratoire est la construction d'un modèle à haute résolution du ribozyme VS complété par les informations dynamiques de ses différents sous-domaines structuraux. Le portrait dynamique du complexe I/V fourni par les travaux présentés dans le chapitre 4 est une première étape vers la construction de ce modèle, bien qu'il reste beaucoup à accomplir avant de pouvoir le construire. En particulier, une étude statistique est nécessaire afin de déterminer les populations des différentes conformations métastables. Il serait également nécessaire de produire le même travail sur d'autres sous-domaines du ribozyme VS, notamment les jonctions III-IV-V et II-III-VI afin de connaître leur contribution à la



formation du site actif. Un intérêt additionnel de tels travaux est la récolte de données de DM sur des éléments de structure secondaire et tertiaire peu étudiés sous l'angle dynamique pour le moment.

À terme, le but de ce modèle est de devenir prédictif : en intégrant les informations structurales tridimensionnelles à échelle atomique et les informations de dynamique à différentes échelles de temps et de niveau de structure, il devient théoriquement possible de prédire l'effet d'une ou plusieurs mutations sur l'activité du système modélisé et à l'inverse, de déterminer les mutations requises pour obtenir un paysage énergétique particulier.

### **5.2.2. Utilisation du ribozyme VS pour le clivage de substrats structurés**

Concevoir des ribozymes sur mesure afin de cliver des substrats d'ARN simple brin est aujourd'hui courant et relativement aisé grâce aux techniques de sélection *in vitro*. En revanche, l'utilisation d'ARN catalytiques pour la reconnaissance et le clivage d'ARN possédant des structures secondaires et tertiaires complexes n'est pas un champ de recherche très actif. Cela est à la fois dû au fait de l'augmentation exponentielle de la difficulté à utiliser des techniques d'ingénierie non-rationnelle lorsque le nombre de nucléotides impliqués dans la reconnaissance et le clivage augmente mais aussi parce qu'il n'existe pas beaucoup d'exemples de tels ribozymes sur lesquels baser ces travaux.

Le ribozyme VS est donc particulièrement intéressant dans cette optique. S'il l'on schématise, son fonctionnement de base est simple : la reconnaissance d'un substrat par création d'interactions tertiaires puis rapprochement d'un élément de structure secondaire non-apparié située dans le substrat d'un élément de structure secondaire non-apparié qui se trouve dans le corps du ribozyme.

Ce fonctionnement de base pourrait être adapté au clivage de substrats très différents non seulement en termes de taille mais aussi reconnus par des interactions tertiaires différentes qu'une interaction boucle-boucle et clivés à des sites qui ne sont pas des boucles internes mais des boucles terminales ou des jonctions, par exemple, ce qui permettrait d'élargir l'éventail de motifs d'ARN clivable par des ribozymes.

### 5.2.3. L'interaction boucle-boucle comme bloc de construction

La construction de nanostructures à base d'ARN implique d'être capable de prévoir de quelle façon va s'organiser en trois dimensions une séquence donnée en solution. La modularité structurale de l'ARN permet de simplifier ce problème et il est ainsi possible d'envisager la structure finale en termes de modules tridimensionnels : sachant de quelle façon se replie des petites séquences d'ARN, de plus grandes structures peuvent être construites soit en fusionnant ces séquences de manière à ce que la séquence finale présente un assemblage de tous les sous-modules [161], soit en les produisant individuellement afin qu'elles s'assemblent une fois repliées individuellement [163]. L'interaction boucle-boucle est un module important dans ce contexte car il permet aussi bien de lier de courtes sous-structures entre elles afin d'obtenir des complexes de plus grande taille que de permettre la création d'interactions tertiaire au sein d'une large structure d'un seul tenant. Les recherches actuelles se concentrent principalement sur la mise au point de macrostructures rigides aptes à servir d'échafaudages, ce qui demande d'avoir un catalogue de pièces peu dynamiques et à la conformation très prévisibles. Néanmoins, dans l'optique de construire des nanostructures de grande tailles dotées de capacités catalytiques, analogues au ribosome par exemple, il est important de compléter ce catalogue avec des pièces très dynamiques et à l'espace conformationnel très prévisible. Dans ce contexte, l'interaction boucle-boucle I/V est un élément de construction intéressant car elle se comporte comme une jointure capable de se mouvoir selon plusieurs axes.

Afin d'approfondir nos connaissances de la dynamique du complexe I/V du ribozyme VS, il serait intéressant d'étudier plus précisément les façons de la modifier. En particulier, nous savons que la flexibilité de la boucle terminale du substrat qui permet une rotation particulière de l'hélice I autour de son axe est limitée par les possibilités de déplacement des résidus de la boucle. Concevoir des modifications rationnelles visant à amplifier cette rotation afin de vérifier si le ribozyme devient alors capable de cliver en dehors des conditions de compensation de longueur d'hélice précédemment déterminés pourrait alors être très intéressant.

#### **5.2.4. Intégrer les modes dynamiques spécifiques à l'ARN dans l'évaluation des ARN fonctionnels**

La dynamique rapide à niveau très local et les réarrangements de structure secondaires lents sont des paramètres dont l'importance pour l'activité des enzymes protéiques et des ribozymes est bien connue. Nos travaux montrent cependant l'importance des mouvements amples d'éléments de structure secondaire, sans réarrangement conformationnel majeur, occasionnés par la dynamique de quelques résidus situés à des « points chauds ». Ce type de mouvements, représentés dans l'étude présentée au chapitre 4 par la dynamique du complexe I/V, conduit à envisager l'activité des ARN catalytiques comme dépendante d'une probabilité de rencontre des différents éléments du site actif en plus des facteurs usuellement associés au taux d'activité d'une enzyme, comme son affinité pour son substrat et la stabilité de cette liaison. Bien que cette façon de percevoir ne soit pas nouvelle, il semble important de la prendre en compte d'une manière systématique lors de l'évaluation du mécanisme catalytique d'un ribozyme aussi bien pour l'évaluation des stratégies d'ingénierie.

Cette vision de l'activité implique de moins mettre l'accent sur les paramètres statiques des comparaisons de structures, comme la *root mean square deviation* (RMSD). Ce paramètre pointe un des défauts majeurs de la pensée statique : en effet, il y a une très faible déviation entre un variant actif et un variant inactif si l'on s'intéresse aux expériences de compensation de longueur d'hélice. À l'opposé, il est impossible de dire si un ARN est actif en utilisant la RMSD si la structure statique étudiée a été capturée lors d'un mouvement rapide de grande amplitude.

#### **5.2.5. L'intégration de la dynamique dans la modélisation et l'ingénierie**

L'exemple du rôle central des mouvements rapides du complexe I/V du ribozyme VS dans sa fonction de clivage implique qu'un modèle statique est insuffisant pour prédire son activité. Cette vision dynamique de la catalyse par l'ARN implique dès lors de s'intéresser à l'espace conformationnel exploré par le ribozyme étudié. Produire une variante active et efficace implique que l'espace conformationnel exploré par ce dernier permette la formation du site catalytique. Ces informations nécessitent un modèle dynamique, c'est-à-dire un modèle tridimensionnel auquel sont rattachées les données de mobilité de chaque sous-domaine de la

molécule étudiée. Afin d'éviter de perpétuellement réinventer la roue, ce type de modélisation nécessite des bases de données de structure 3D de modules d'ARN tels que des jonctions multivoies et des interactions boucle-boucle associées aux informations dynamiques qui leur sont propres comme leurs(s) degré(s) de liberté, l'amplitude de leur mouvements intrinsèques et les résidus à l'origine de ces mouvements. Il pourrait être également très informatif de classer les unités structurales de l'ARN selon leur potentiel dynamique. La modularité de l'ARN est un trait qui rend cette démarche d'autant plus pertinente. Reliés par des structures peu dynamiques comme des hélices, ces modules peuvent continuer de se comporter au sein d'une structure complexe de la même manière qu'isolés. Il est alors relativement facile de les remplacer par d'autres modules de dynamique connue.

À défaut de bases de données, il est important de découper les molécules d'ARN fonctionnelles d'intérêt en modules indépendant et de procéder à leur étude dynamique, que ce soit en dynamique moléculaire ou par des méthodes biochimiques aptes à brosser le portrait de leur ensemble conformationnel en solution. Dans la même optique, il pourrait être intéressant d'étudier des modifications chimiques des nucléotides aptes à biaiser la dynamique locale d'ARN d'intérêt vers des portions particulières de leur espace conformationnel.

Incorporer la notion de dynamique dans l'ingénierie d'ARN permet de compléter les techniques existantes. Les méthodes non-rationnelles employées pour modifier des ARN catalytiques connus, comme le SELEX deviennent complexes à mettre en œuvre dès lors que la taille de la molécule considérée devient importante, peuvent être rendues semi-rationnelles en ciblant les sous-domaines responsables des mouvements d'intérêt.

### **5.3. Conclusion : passer de la structure native à l'ensemble natif**

La possibilité d'obtenir la structure native d'une molécule d'ARN par prédiction est encore aujourd'hui un but désirable en bio-informatique. Des initiatives telles que RNA-Puzzles [119, 164], très utiles pour évaluer les progrès de la prédiction de structure 3D d'ARN, mettent l'accent sur l'importance d'obtenir une réponse en accord avec les structures obtenues par des méthodes principalement biophysiques et évaluées par des métriques comme la RMSD des positions atomiques et le *mean of circular quantities score* (MCQ). De plus, de nombreux travaux portant sur les ARN non-codants ignorent encore totalement la structure

tridimensionnelle et l'aspect dynamique de ces derniers. Nos travaux ainsi que de nombreux autres détaillés dans cette thèse semblent pourtant souligner à la fois l'importance pour la fonction de l'ARN de son aspect hautement dynamique et les limitations des études de résolution de structures quand il s'agit de capturer une unique conformation capable d'expliquer la fonction d'un ARN.

Pour ces raisons, il semble important de prendre en compte les spécificités structurales et dynamiques de l'ARN et de se dissocier de la vision à deux structures (liée/non-liée) héritée des protéines pour définitivement considérer l'ARN comme un ensemble de structures à l'équilibre en tout temps.

## 6. Bibliographie

1. Weinberg, R.A. and S. Penman, *Small molecular weight monodisperse nuclear RNA*. J Mol Biol, 1968. **38**(3): p. 289-304.
2. Hodnett, J.L. and H. Busch, *Isolation and characterization of uridylic acid-rich 7 S ribonucleic acid of rat liver nuclei*. J Biol Chem, 1968. **243**(24): p. 6334-42.
3. Lerner, M.R., et al., *Are snRNPs involved in splicing?* Nature, 1980. **283**(5743): p. 220-4.
4. Grabowski, P.J., S.R. Seiler, and P.A. Sharp, *A multicomponent complex is involved in the splicing of messenger RNA precursors*. Cell, 1985. **42**(1): p. 345-53.
5. Brody, E. and J. Abelson, *The "spliceosome": yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction*. Science, 1985. **228**(4702): p. 963-7.
6. Prestayko, A.W., M. Tonato, and H. Busch, *Low molecular weight RNA associated with 28 s nucleolar RNA*. J Mol Biol, 1970. **47**(3): p. 505-15.
7. Maxwell, E.S. and M.J. Fournier, *The small nucleolar RNAs*. Annu Rev Biochem, 1995. **64**: p. 897-934.
8. 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 U S A, 1984. **81**(7): p. 1966-70.
9. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans**. Nature, 1998. **391**(6669): p. 806-11.
10. 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.
11. Cech, T.R., A.J. Zaug, and P.J. Grabowski, *Invitro splicing of the ribosomal-RNA precursor of *Tetrahymena* - involvement of a guanosine nucleotide in the excision of the intervening sequence*. Cell, 1981. **27**(3): p. 487-496.
12. Michel, F., A. Jacquier, and B. Dujon, *Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure*. Biochimie, 1982. **64**(10): p. 867-81.
13. 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.

14. Palade, G.E., *A small particulate component of the cytoplasm*. J Biophys Biochem Cytol, 1955. **1**(1): p. 59-68.
15. Nissen, P., et al., *The structural basis of ribosome activity in peptide bond synthesis*. Science, 2000. **289**(5481): p. 920-30.
16. Palazzo, A.F. and E.S. Lee, *Non-coding RNA: what is functional and what is junk?* Front Genet, 2015. **6**: p. 2.
17. Gilbert, W., *Origin of life: The RNA world*. Nature, 1986. **319**: p. 618.
18. Lewin, R., *RNA Catalysis Gives Fresh Perspective on the Origin of Life: The old chicken-and-egg problem of the origin of life is illuminated in unexpected ways by recent results on the splicing of RNA precursors*. Science, 1986. **231**(4738): p. 545-6.
19. Pace, N.R. and T.L. Marsh, *RNA catalysis and the origin of life*. Orig Life Evol Biosph, 1985. **16**(2): p. 97-116.
20. Julien, K.R., et al., *Conformationally restricted nucleotides as a probe of structure-function relationships in RNA*. RNA, 2008. **14**(8): p. 1632-43.
21. Zirbel, C.L., et al., *Classification and energetics of the base-phosphate interactions in RNA*. Nucleic Acids Res, 2009. **37**(15): p. 4898-918.
22. Treger, M. and E. Westhof, *Statistical analysis of atomic contacts at RNA-protein interfaces*. J Mol Recognit, 2001. **14**(4): p. 199-214.
23. Liu, J. and D.M. Lilley, *The role of specific 2'-hydroxyl groups in the stabilization of the folded conformation of kink-turn RNA*. RNA, 2007. **13**(2): p. 200-10.
24. Paladino, A. and R. Zangi, *Ribose 2'-hydroxyl groups stabilize RNA hairpin structures containing gcuua pentaloop*. J Chem Theory Comput, 2013. **9**(2): p. 1214-21.
25. Abu Almakarem, A.S., et al., *Comprehensive survey and geometric classification of base triples in RNA structures*. Nucleic Acids Res, 2012. **40**(4): p. 1407-23.
26. Leontis, N.B. and E. Westhof, *Geometric nomenclature and classification of RNA base pairs*. RNA, 2001. **7**(4): p. 499-512.
27. Hunter, C.A. and J.K.M. Sanders, *The nature of .pi.-.pi. interactions*. J Am Chem Soc, 1990. **112**(14): p. 5525-5534.
28. Bailor, M.H., X. Sun, and H.M. Al-Hashimi, *Topology links RNA secondary structure with global conformation, dynamics, and adaptation*. Science, 2010. **327**(5962): p. 202-6.

29. Darty, K., A. Denise, and Y. Ponty, *VARNA: Interactive drawing and editing of the RNA secondary structure*. Bioinformatics, 2009. **25**(15): p. 1974-5.
30. Kim, S.H. and J.L. Sussman, *pi turn is a conformational pattern in RNA loops and bends*. Nature, 1976. **260**(5552): p. 645-6.
31. Quigley, G.J. and A. Rich, *Structural domains of transfer RNA molecules*. Science, 1976. **194**(4267): p. 796-806.
32. Heus, H.A. and A. Pardi, *Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops*. Science, 1991. **253**(5016): p. 191-4.
33. Baumruk, V., et al., *Comparison between CUUG and UUCG tetraloops: thermodynamic stability and structural features analyzed by UV absorption and vibrational spectroscopy*. Nucleic Acids Res, 2001. **29**(19): p. 4089-96.
34. Wimberly, B., G. Varani, and I. Tinoco, *The conformation of loop-e of eukaryotic 5s-ribosomal RNA*. Biochemistry, 1993. **32**(4): p. 1078-1087.
35. Vidovic, I., et al., *Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment*. Molecular Cell, 2000. **6**(6): p. 1331-1342.
36. Lescoute, A. and E. Westhof, *Topology of three-way junctions in folded RNAs*. RNA, 2006. **12**(1): p. 83-93.
37. Laing, C., et al., *Tertiary motifs revealed in analyses of higher-order RNA junctions*. J Mol Biol, 2009. **393**(1): p. 67-82.
38. Laing, C. and T. Schlick, *Analysis of four-way junctions in RNA structures*. J Mol Biol, 2009. **390**(3): p. 547-59.
39. Kim, S.H., et al., *The general structure of transfer RNA molecules*. Proc Natl Acad Sci U S A, 1974. **71**(12): p. 4970-4.
40. Cate, J.H., et al., *Crystal structure of a group I ribozyme domain: principles of RNA packing*. Science, 1996. **273**(5282): p. 1678-85.
41. Chang, K.Y. and I. Tinoco, Jr., *Characterization of a "kissing" hairpin complex derived from the human immunodeficiency virus genome*. Proc Natl Acad Sci U S A, 1994. **91**(18): p. 8705-9.
42. Shen, L.X. and I. Tinoco, Jr., *The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus*. J Mol Biol, 1995. **247**(5): p. 963-78.



43. 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.
44. Walter, N.G. and J.M. Burke, *The hairpin ribozyme: structure, assembly and catalysis*. Curr Opin Chem Biol, 1998. **2**(1): p. 24-30.
45. Eiler, D., J. Wang, and T.A. Steitz, *Structural basis for the fast self-cleavage reaction catalyzed by the twister ribozyme*. Proc Natl Acad Sci U S A, 2014. **111**(36): p. 13028-33.
46. Lau, M.W., K.E. Cadieux, and P.J. Unrau, *Isolation of fast purine nucleotide synthase ribozymes*. J Am Chem Soc, 2004. **126**(48): p. 15686-93.
47. Horning, D.P. and G.F. Joyce, *Amplification of RNA by an RNA polymerase ribozyme*. Proc Natl Acad Sci U S A, 2016. **113**(35): p. 9786-91.
48. Samanta, B. and G.F. Joyce, *A reverse transcriptase ribozyme*. Elife, 2017. **6**.
49. Moretti, J.E. and U.F. Muller, *A ribozyme that triphosphorylates RNA 5'-hydroxyl groups*. Nucleic Acids Res, 2014. **42**(7): p. 4767-4778.
50. Winkler, W.C., et al., *Control of gene expression by a natural metabolite-responsive ribozyme*. Nature, 2004. **428**(6980): p. 281-6.
51. Barrick, J.E., et al., *New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control*. Proc Natl Acad Sci U S A, 2004. **101**(17): p. 6421-6.
52. Weinberg, Z., et al., *New classes of self-cleaving ribozymes revealed by comparative genomics analysis*. Nat Chem Biol, 2015. **11**(8): p. 606-10.
53. Roth, A., et al., *A widespread self-cleaving ribozyme class is revealed by bioinformatics*. Nat Chem Biol, 2014. **10**(1): p. 56-60.
54. Prody, G.A., et al., *Autolytic processing of dimeric plant virus satellite RNA*. Science, 1986. **231**(4745): p. 1577-80.
55. Kuo, M.Y., et al., *Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus*. J Virol, 1988. **62**(12): p. 4439-44.
56. 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.
57. 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.

58. Chen, J.H., et al., *A 1.9 Å crystal structure of the HDV ribozyme precleavage suggests both Lewis acid and general acid mechanisms contribute to phosphodiester cleavage*. *Biochemistry*, 2010. **49**(31): p. 6508-18.
59. Mir, A., et al., *Two Divalent Metal Ions and Conformational Changes Play Roles in the Hammerhead Ribozyme Cleavage Reaction*. *Biochemistry*, 2015. **54**(41): p. 6369-81.
60. Jimenez, R.M., J.A. Polanco, and A. Luptak, *Chemistry and biology of self-cleaving ribozymes*. *Trends Biochem. Sci.*, 2015. **40**(11): p. 648-661.
61. 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.
62. Saville, B.J. and R.A. Collins, *RNA-mediated ligation of self-cleavage products of a Neurospora mitochondrial plasmid transcript*. *Proc Natl Acad Sci U S A*, 1991. **88**(19): p. 8826-30.
63. Beattie, T.L., J.E. Olive, and R.A. Collins, *A secondary-structure model for the self-cleaving region of Neurospora VS RNA*. *Proc Natl Acad Sci U S A*, 1995. **92**(10): p. 4686-90.
64. Guo, H.C., et al., *Nucleotide sequence requirements for self-cleavage of Neurospora VS RNA*. *J Mol Biol*, 1993. **232**(2): p. 351-61.
65. Suslov, N.B., et al., *Crystal structure of the Varkud satellite ribozyme*. *Nat Chem Biol*, 2015. **11**(11): p. 840-6.
66. Bouchard, P., et al., *Role of SLV in SLI substrate recognition by the Neurospora VS ribozyme*. *RNA*, 2008. **14**(4): p. 736-48.
67. 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.
68. 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.
69. 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.
70. Pereira, M.J., et al., *Single VS ribozyme molecules reveal dynamic and hierarchical folding toward catalysis*. *J Mol Biol*, 2008. **382**(2): p. 496-509.

71. Lipfert, J., et al., *The complete VS ribozyme in solution studied by small-angle X-ray scattering*. Structure, 2008. **16**(9): p. 1357-67.
72. Bonneau, E., et al., *The NMR structure of the II-III-VI three-way junction from the Neurospora VS ribozyme reveals a critical tertiary interaction and provides new insights into the global ribozyme structure*. RNA, 2015. **21**(9): p. 1621-32.
73. DasGupta, S., N.B. Suslov, and J.A. Piccirilli, *Structural Basis for Substrate Helix Remodeling and Cleavage Loop Activation in the Varkud Satellite Ribozyme*. J Am Chem Soc, 2017.
74. 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.
75. Dagenais, P., et al., *Insights into RNA structure and dynamics from recent NMR and X-ray studies of the Neurospora Varkud satellite ribozyme*. Wiley Interdiscip Rev RNA, 2017. **8**(5).
76. Collins, R.A., *The Neurospora Varkud satellite ribozyme*. Biochem Soc Trans, 2002. **30**(Pt 6): p. 1122-6.
77. Lilley, D.M., *The Varkud satellite ribozyme*. RNA, 2004. **10**(2): p. 151-8.
78. Lilley, D.M., *Catalysis by the nucleolytic ribozymes*. Biochem Soc Trans, 2011. **39**(2): p. 641-6.
79. 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.
80. Lilley, D.M., *Mechanisms of RNA catalysis*. Philos Trans R Soc Lond B Biol Sci, 2011. **366**(1580): p. 2910-7.
81. Wilson, T.J., et al., *Nucleobase-mediated general acid-base catalysis in the Varkud satellite ribozyme*. Proc Natl Acad Sci U S A, 2010. **107**(26): p. 11751-6.
82. 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 U S A, 2007. **104**(14): p. 5818-23.
83. 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.

84. Smith, M.D., et al., *The ionic environment determines ribozyme cleavage rate by modulation of nucleobase pK<sub>a</sub>*. RNA, 2008. **14**(9): p. 1942-9.
85. Cochrane, J.C. and S.A. Strobel, *Catalytic strategies of self-cleaving ribozymes*. Acc Chem Res, 2008. **41**(8): p. 1027-35.
86. 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.
87. 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.
88. Andersen, A.A. and R.A. Collins, *Intramolecular secondary structure rearrangement by the kissing interaction of the Neurospora VS ribozyme*. Proc Natl Acad Sci U S A, 2001. **98**(14): p. 7730-5.
89. 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.
90. Michiels, P.J.A., et al., *Structure of the ribozyme substrate hairpin of Neurospora VS RNA: A close look at the cleavage site*. Rna-a Publication of the Rna Society, 2000. **6**(12): p. 1821-1832.
91. 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.
92. Hoffmann, B., et al., *NMR structure of the active conformation of the Varkud satellite ribozyme cleavage site*. Proc Natl Acad Sci U S A, 2003. **100**(12): p. 7003-8.
93. 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.
94. 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.
95. Bergonzo, C., K.B. Hall, and T.E. Cheatham, 3rd, *Stem-loop V of Varkud Satellite RNA exhibits characteristics of the Mg(2+) bound structure in the presence of monovalent ions*. J Phys Chem B, 2015. **119**(38): p. 12355-64.

96. 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.
97. Bergonzo, C. and T.E. Cheatham, 3rd, *Mg(2+) Binding Promotes SLV as a Scaffold in Varkud Satellite Ribozyme SLI-SLV Kissing Loop Junction*. *Biophys J*, 2017. **113**(2): p. 313-320.
98. 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.
99. 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.
100. 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*, 2014. **53**(3): p. 579-90.
101. Bonneau, E. and P. Legault, *Nuclear magnetic resonance structure of the III-IV-V three-way junction from the Varkud satellite ribozyme and identification of magnesium-binding sites using paramagnetic relaxation enhancement*. *Biochemistry*, 2014. **53**(39): p. 6264-75.
102. Bouchard, P. and P. Legault, *A remarkably stable kissing-loop interaction defines substrate recognition by the Neurospora Varkud Satellite ribozyme*. *RNA*, 2014. **20**(9): p. 1451-64.
103. Frauenfelder, H., S.G. Sligar, and P.G. Wolynes, *The energy landscapes and motions of proteins*. *Science*, 1991. **254**(5038): p. 1598-603.
104. Cruz, J.A. and E. Westhof, *The Dynamic Landscapes of RNA Architecture*. *Cell*, 2009. **136**(4): p. 604-609.
105. Mustoe, A.M., C.L. Brooks, and H.M. Al-Hashimi, *Hierarchy of RNA functional dynamics*. *Annu Rev Biochem*, 2014. **83**: p. 441-66.
106. Hennelly, S.P. and K.Y. Sanbonmatsu, *Tertiary contacts control switching of the SAM-I riboswitch*. *Nucleic Acids Res*, 2011. **39**(6): p. 2416-31.

107. Sharma, M., G. Bulusu, and A. Mitra, *MD simulations of ligand-bound and ligand-free aptamer: molecular level insights into the binding and switching mechanism of the add A-riboswitch*. RNA, 2009. **15**(9): p. 1673-92.
108. Bothe, J.R., et al., *Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy*. Nat Methods, 2011. **8**(11): p. 919-31.
109. Chen, Y. and L. Pollack, *SAXS studies of RNA: structures, dynamics, and interactions with partners*. Wiley Interdiscip Rev RNA, 2016. **7**(4): p. 512-26.
110. Merino, E.J., et al., *RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE)*. J Am Chem Soc, 2005. **127**(12): p. 4223-31.
111. Zhuang, X., et al., *Fluorescence quenching: A tool for single-molecule protein-folding study*. Proc Natl Acad Sci U S A, 2000. **97**(26): p. 14241-4.
112. Zhuang, X., et al., *A single-molecule study of RNA catalysis and folding*. Science, 2000. **288**(5473): p. 2048-51.
113. Murata, K. and M. Wolf, *Cryo-electron microscopy for structural analysis of dynamic biological macromolecules*. Biochim Biophys Acta, 2018. **1862**(2): p. 324-334.
114. Korostelev, A., D.N. Ermolenko, and H.F. Noller, *Structural dynamics of the ribosome*. Current Opinion in Chemical Biology, 2008. **12**(6): p. 674-683.
115. Agirrezabala, X. and M. Valle, *Structural Insights into tRNA Dynamics on the Ribosome*. International Journal of Molecular Sciences, 2015. **16**(5): p. 9866-9895.
116. White, N.A., et al., *Coupling between conformational dynamics and catalytic function at the active site of the lead-dependent ribozyme*. RNA, 2018.
117. Ochieng, P.O., et al., *Intrinsic base-pair rearrangement in the hairpin ribozyme directs RNA conformational sampling and tertiary interface formation*. J Phys Chem B, 2016.
118. Vusurovic, N., et al., *Pseudoknot formation seeds the twister ribozyme cleavage reaction coordinate*. J Am Chem Soc, 2017. **139**(24): p. 8186-8193.
119. Miao, Z., et al., *RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme*. RNA, 2017. **23**(5): p. 655-672.
120. Berman, H.M., et al., *The Protein Data Bank*. Nucleic Acids Res, 2000. **28**(1): p. 235-42.

121. Danaee, P., et al., *bpRNA: large-scale automated annotation and analysis of RNA secondary structure*. Nucleic Acids Res, 2018. **46**(11): p. 5381-5394.
122. Zok, T., et al., *RNApdbee 2.0: multifunctional tool for RNA structure annotation*. Nucleic Acids Res, 2018.
123. Adams, H., et al., *A supramolecular system for quantifying aromatic stacking interactions*. Chemistry, 2001. **7**(22): p. 4863-77.
124. Popena, M., et al., *Automated 3D structure composition for large RNAs*. Nucleic Acids Res, 2012. **40**(14): p. e112.
125. Jossinet, F., T.E. Ludwig, and E. Westhof, *Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels*. Bioinformatics, 2010. **26**(16): p. 2057-9.
126. Xia, Z., et al., *RNA 3D structure prediction by using a coarse-grained model and experimental data*. J Phys Chem B, 2013. **117**(11): p. 3135-44.
127. Cao, S. and S.J. Chen, *Physics-based de novo prediction of RNA 3D structures*. J Phys Chem B, 2011. **115**(14): p. 4216-26.
128. Tan, D., et al., *RNA force field with accuracy comparable to state-of-the-art protein force fields*. Proc Natl Acad Sci U S A, 2018. **115**(7): p. E1346-E1355.
129. Fiorin, G., M.L. Klein, and J. Henin, *Using collective variables to drive molecular dynamics simulations*. Molecular Physics, 2013. **111**(22-23): p. 3345-3362.
130. Torrie, G.M. and J.P. Valleau, *Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling*. J Comput Phys, 1977. **23**(2): p. 187-199.
131. Laio, A. and M. Parrinello, *Escaping free-energy minima*. Proc Natl Acad Sci U S A, 2002. **99**(20): p. 12562-6.
132. Sugita, Y. and Y. Okamoto, *Replica-exchange molecular dynamics method for protein folding*. Chem Phys Lett, 1999. **314**(1-2): p. 141-151.
133. Marinari, E. and G. Parisi, *Simulated tempering - a new monte-carlo scheme*. EPL, 1992. **19**(6): p. 451-458.
134. Haldar, S., et al., *Insights into stability and folding of GNRA and UNGC tetraloops revealed by microsecond molecular dynamics and well-tempered metadynamics*. J Chem Theory Comput, 2015. **11**(8): p. 3866-77.

135. Zhuang, Z., L. Jaeger, and J.E. Shea, *Probing the structural hierarchy and energy landscape of an RNA T-loop hairpin*. Nucleic Acids Res, 2007. **35**(20): p. 6995-7002.
136. 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 U S A, 2012. **109**(24): p. E1530-9.
137. Sripathi, K.N., et al., *Wobble pairs of the HDV ribozyme play specific roles in stabilization of active site dynamics*. Phys Chem Chem Phys, 2015. **17**(8): p. 5887-5900.
138. Besseova, I., et al., *Molecular dynamics simulations suggest that RNA three-way junctions can act as flexible RNA structural elements in the ribosome*. Nucleic Acids Res, 2010. **38**(18): p. 6247-64.
139. Iwakiri, J., et al., *Improved Accuracy in RNA-Protein Rigid Body Docking by Incorporating Force Field for Molecular Dynamics Simulation into the Scoring Function*. J Chem Theory Comput, 2016. **12**(9): p. 4688-97.
140. Besseova, I., et al., *Simulations of A-RNA duplexes. The effect of sequence, solute force field, water model, and salt concentration*. J Phys Chem B, 2012. **116**(33): p. 9899-916.
141. Cunha, R.A. and G. Bussi, *Unraveling Mg(2+)-RNA binding with atomistic molecular dynamics*. RNA, 2017. **23**(5): p. 628-638.
142. Henriksen, N.M., D.R. Davis, and T.E. Cheatham, 3rd, *Molecular dynamics re-refinement of two different small RNA loop structures using the original NMR data suggest a common structure*. J Biomol NMR, 2012. **53**(4): p. 321-39.
143. Cheng, P., J. Peng, and Z. Zhang, *SAXS-Oriented Ensemble Refinement of Flexible Biomolecules*. Biophys J, 2017. **112**(7): p. 1295-1301.
144. Trabuco, L.G., et al., *Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics*. Structure, 2008. **16**(5): p. 673-83.
145. Brooks, B. and M. Karplus, *Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor*. Proc Natl Acad Sci U S A, 1983. **80**(21): p. 6571-5.
146. Go, N., T. Noguti, and T. Nishikawa, *Dynamics of a small globular protein in terms of low-frequency vibrational modes*. Proc Natl Acad Sci U S A, 1983. **80**(12): p. 3696-700.
147. Amadei, A., A.B. Linssen, and H.J. Berendsen, *Essential dynamics of proteins*. Proteins, 1993. **17**(4): p. 412-25.



148. Tuerk, C. and L. Gold, *Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase*. Science (New York, N Y ), 1990. **249**(4968): p. 505-10.
149. Ellington, A.D. and J.W. Szostak, *In vitro selection of RNA molecules that bind specific ligands*. Nature, 1990. **346**(6287): p. 818-22.
150. Shu, D., et al., *Programmable folding of fusion RNA in vivo and in vitro driven by pRNA 3WJ motif of phi29 DNA packaging motor*. Nucleic Acids Res, 2014. **42**(2): p. e10.
151. Perez-Ruiz, M., A. Barroso-DelJesus, and A. Berzal-Herranz, *Specificity of the hairpin ribozyme. Sequence requirements surrounding the cleavage site*. J Biol Chem, 1999. **274**(41): p. 29376-80.
152. Jellinek, D., et al., *High-affinity RNA ligands to basic fibroblast growth factor inhibitor receptor binding*. Proc Natl Acad Sci U S A, 1993. **90**(23): p. 11227-31.
153. Werstuck, G. and M.R. Green, *Controlling gene expression in living cells through small molecule-RNA interactions*. Science, 1998. **282**(5387): p. 296-8.
154. Suess, B., et al., *Conditional gene expression by controlling translation with tetracycline-binding aptamers*. Nucleic Acids Res, 2003. **31**(7): p. 1853-8.
155. Tang, J. and R.R. Breaker, *Rational design of allosteric ribozymes*. Chem Biol, 1997. **4**(6): p. 453-9.
156. Lau, M.W. and A.R. Ferre-D'Amare, *In vitro evolution of coenzyme-independent variants from the glmS ribozyme structural scaffold*. Methods, 2016. **106**: p. 76-81.
157. Dolan, G.F., A. Akoopie, and U.F. Muller, *A Faster Triphosphorylation Ribozyme*. PLoS One, 2015. **10**(11).
158. Diegelman, A.M. and E.T. Kool, *Generation of circular RNAs and trans-cleaving catalytic RNAs by rolling transcription of circular DNA oligonucleotides encoding hairpin ribozymes*. Nucleic Acids Res, 1998. **26**(13): p. 3235-41.
159. Berzal-Herranz, A., et al., *Essential nucleotide sequences and secondary structure elements of the hairpin ribozyme*. EMBO J, 1993. **12**(6): p. 2567-73.
160. 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.

161. Geary, C., P.W. Rothmund, and E.S. Andersen, *RNA nanostructures. A single-stranded architecture for cotranscriptional folding of RNA nanostructures*. *Science*, 2014. **345**(6198): p. 799-804.
162. Geary, C., et al., *Composing RNA nanostructures from a syntax of RNA structural modules*. *Nano Lett*, 2017. **17**(11): p. 7095-7101.
163. Horiya, S., et al., *RNA LEGO: magnesium-dependent formation of specific RNA assemblies through kissing interactions*. *Chem Biol*, 2003. **10**(7): p. 645-54.
164. Cruz, J.A., et al., *RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction*. *RNA*, 2012. **18**(4): p. 610-25.