CATION EFFECTS ON THE FOLDING OF RIBOSOMAL RNA

A Dissertation
Presented to
The Academic Faculty

by

Timothy Kenneth Lenz

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry & Biochemistry

Georgia Institute of Technology
December, 2015

CATION EFFECTS ON THE FOLDING OF RIBOSOMAL RNA

Approved by:

Dr. Nicholas V. Hud, Advisor
School of Chemistry & Biochemistry
Georgia Institute of Technology

Dr. Loren Dean Williams, Co-Advisor
School of Chemistry & Biochemistry
Georgia Institute of Technology

Dr. Roger M. Wartell
School of Biology
Georgia Institute of Technology

Dr. Stephen C. Harvey
School of Biology
Georgia Institute of Technology

Dr. Adegboyega Oyelere
School of Chemistry & Biochemistry
Georgia Institute of Technology

Date Approved: Nov 3rd, 2015

To my high school science teachers: particularly Peter Giordano, Carl Vickers, and Susan Picanco. Thank you for igniting in me a lifelong appreciation for science and a drive to understand how our world works.

ACKNOWLEDGEMENTS

First and foremost, I wish to thank my co-advisors, Dr. Loren Williams and Dr. Nicholas Hud. Their research interests and the enthusiasm with which they speak on them are infectious and I could not help but become vastly fascinated in questions related to the origins of life. While working with my co-advisors I have learned many useful lessons, both of scientific and practical natures. I have gained countless worthwhile skills and insights during my studies, due in large part to their advice and support. They have encouraged me to work independently and make my own judgements, and helped me make difficult decisions regarding the pursuit of elusive results. Thanks to my co-advisors, I am leaving graduate school with a much better understanding of my career path.

I am grateful to the other members of my thesis committee--Dr. Steve Harvey, Dr. Yomi Oyelere, and Dr. Roger Wartell--for their time and advice throughout my studies. They have been available to me not just in formal capacities, but also in countless informal and extremely educational meetings. My committee members have provided me with encouragement and excellent suggestions to improve the interpretation and presentation of my results.

I would also like to acknowledge the members of the Williams and Hud groups, past and present, for their aid in instrumentation, discussion, and support. Members of both research groups have been a pleasure to work with. In particular, I would like to thank Jessica Bowman for limitless experimental, moral, and writing support, Eric O'Neill for supplying extensive protocol knowledge and pristine RNA, Chiaolong Hsiao

for his mentorship and encouragement, and Sue Winters for her generous administrative support. Shreyas Athavale, Chad Bernier, Aaron Engelhart, Jared Gossett, and Anton Petrov also provided valuable tools, information, and discussion crucial to the execution and interpretation of experiments described in this document.

I am very grateful to several hard-working undergraduate researchers that have assisted me in my experiments. Ava Afshar, Amy Boudreau, Joey Hahm, and Ashlyn Norris all provided me with rewarding mentoring experiences, and often taught me just as much if not more than I strived to teach them.

I also thank my Georgia Tech friends, who have provided welcome distraction from the lab, often at times when it was sorely needed. Joel, Marty, and Mike have been part of many entertaining lunch breaks, evenings, and weekends.

Last but certainly not least, I thank my family for their continuous, unflinching love and faith in me and my abilities, even at times when my self-confidence wavered or collapsed entirely. I would not have made it through this endeavor with my sanity (mostly) intact without their persistent reassurance. My wife and best friend, Alicia, has been at my side to celebrate my victories or support me through my frustrations, and has always encouraged me to persevere. My mother, father, and sister, though often far away, consistently help me keep things in perspective. And my dog, Gordon, has provided me with copious grunt-filled solace.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ΔMg^{2+} or ΔFe^{2+} site	position that exhibits significant response to indicated cation
ΔMg^{2+} or ΔFe^{2+} value	effect of indicated cation on SHAPE reactivity
2'-OH	2'-hydroxyl group
a-PTC	ancestral peptidyl transferase center
a-rPeptide	ancestral ribosomal peptide
a-rRNA-*x*	ancestral ribosomal RNA version *x*
AOCN	average observed coordination number
bp	base pair
cDNA	complementary DNA
CE	capillary electrophoresis
cWW	cis Watson-Crick/Watson-Crick base pairing
DCTA	1,2-diaminocyclohexanetetraacetic acid
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
ddNTP	dideoxynucleotide triphosphate

dNTP	deoxynucleotide triphosphate
GOE	great oxidation event
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
I	intermediate RNA folding state
LSU	ribosomal large subunit
LUCA	last universal common ancestor
Mg^{2+}-μc	magnesium microcluster
mRNA	messenger RNA
N	native RNA folding state
NMIA	N-methylisatoic anhydride
NMR	nuclear magnetic resonance
nt	nucleotide
P4-P6	subdomain of the *Tetrahymena thermophila* Group I intron
PAGE	polyacrylamide gel electrophoresis
PDB	protein data bank
PO	non-bridging phosphate oxygen
PTC	peptidyl transferase center

R-PCR	recursive polymerase chain reaction
ribozyme	RNA enzyme
RNA	ribonucleic acid
rPeptide	ribosomal peptide
rProtein	ribosomal protein
rRNA	ribosomal RNA
rRNA-Mg^{2+} state	conformational state of the LSU rRNA in presence of Mg^{2+}
RT	reverse transcription
SHAPE	selective 2'-hydroxyl acylation analyzed by primer extension
SSU	ribosomal small subunit
tRNA	transfer RNA
*Tt*23S	*Thermus thermophilus* 23S DNA plasmid construct
U	unfolded RNA state
UV-Vis	ultraviolet-visible spectroscopy
WT	wild type

SUMMARY

Folding of ribonucleic acids (RNA) invariably involves positively charged metal ions (cations) which neutralize the negatively-charged phosphate groups, allowing backbone atoms to achieve close proximity. This dissertation details multiple studies into the role of divalent cations in the folding of large ribosomal RNAs. We explore the intricate relationship between divalent cations and ribosomal structure, origins, and evolution. A powerful RNA footprinting technique called SHAPE is the primary tool utilized to detect structural transitions of ribosomal RNA induced by cations.

We develop a model of an ancestral ribosomal RNA that is 80% smaller than its modern counterpart. Despite extensive deletions, this ancestral ribosome retains the ability to fold into its predicted secondary structure and associate specifically with Mg^{2+} ions and a ribosomal protein ancestor. These results imply that the functional core of the ribosome is an ancient assembly that has remained stable and largely static over billions of years of evolution.

We also find that protein-free ribosomal RNA of the large ribosomal subunit exhibits widely-dispersed conformational changes upon association with Mg^{2+}, consistent with global collapse to a near-native conformation. Many native tertiary interactions are inferred in this state, even in the absence of all other ribosomal components. Evidence suggests that many long range inter-domain interactions are induced by Mg^{2+}, forming the core architecture of the ribosomal large subunit. By inference, the structural effects of rProteins are largely assumed to be local and nominal.

Predicated on earlier structural and functional experiments, we perform experiments designed to examine the structure of ribosomal RNA under plausible early earth conditions. In particular we observe effects of Fe^{2+} on ribosomal RNA structure, with the goal of understanding the emergence of the translation apparatus under relevant geochemical conditions. Fe^{2+} would have been abundant under the reducing atmosphere of ancient earth, whereas today it is mostly oxidized to insoluble Fe^{3+}. We observe atomic-level mimicry of Fe^{2+} for Mg^{2+} in extant and ancient protein-free ribosomal RNA structures. This substitution is robust; our results suggest that it occurs through multiple cation binding modes. Ribosomal RNA is seen to fold more readily in presence of Fe^{2+} than Mg^{2+}, suggesting increased folding competency under ancient environmental conditions.

Certain trends were observed in the SHAPE data obtained for these large ribosomal RNA systems. Repeating SHAPE patterns are reported for a common class of stem-loop and the well-known K-turn RNA motif. The potential utility of patterns such as these in improving RNA structure prediction and validation is discussed.

CHAPTER 1

INTRODUCTION

An RNA World that predated the modern biological dichotomy of protein and nucleic acid is a widely-accepted model for the emergence of life from simple chemical networks.[1-9] The RNA World Hypothesis is actually a group of related models, with a variety of assumptions and definitions. In all variations of the RNA World Hypothesis, RNA enzymes (ribozymes) predate protein enzymes. The origin and evolution of the ribosome, which is responsible for synthesis of all coded proteins in modern biology, marks the boundary between the two phases.

One central element of any RNA World Hypothesis is a feasible pathway out of the RNA World, into the extant biological triumvirate of DNA/RNA/protein. That is, biology presumably made a polymer transition from the RNA World to the current state of biopolymer co-dependence in which ribozymes, particularly ribosomes, synthesize protein, and the vast majority of chemical transformations are catalyzed and regulated by proteins. This transition must have followed the continuity principle, accomplished by numerous, manageable steps, each maintaining fitness of crucial components. The discovery that the ribosome is a ribozyme[10,11] has been taken as support for the RNA World Hypothesis.

The translation system, of which the ribosome is the central component, has been used for almost half a century to probe some of deepest questions in biology.[12] Using the ribosome as a molecular ‘time machine’, Woese and Fox discovered that life on earth

arose from three primary lineages.[13] They were able to correct the long-standing tree of life that had been used by biologists for decades because the translation system has recorded and retained interpretable information on the ancient past. Numerous studies have now confirmed that molecular structures and chemical processes that directed the early evolution of life on earth are contained in or imprinted on the translation system. A more thorough discussion of these studies is provided in later sections of this work.

In this work, we draw inspiration from the studies of Woese and others by asking questions regarding the structures assumed by ribosomal RNA in the absence of its protein components. Specifically, we seek to increase understanding of the role of cations in the structure of the nucleic acid components of the translation machinery by posing the following general queries:

1) What are the roles of divalent cations in the folding and stability of ancient and modern ribosomal RNA?

2) Were the same divalent cations involved during the entire evolution of the translation machinery, or did large-scale environmental transitions induce a switch in the identity of cations which associate with ribosomal RNA?

1.1 – RNA folding

Portions of this section have been adapted from previously published work: "Bowman, Jessica C.; Lenz, Timothy K.; Hud, Nicholas V.; Williams, Loren Dean. Cations in Charge: Magnesium Ions in RNA Folding and Catalysis. Current Opinion in Structural Biology, 2012, 22, 262-272."[14] The primary contribution of the author of this

document to the published work was a review of recently-developed methods in nucleic acid-cation studies.

When large RNAs fold into compact structures, negatively charged phosphate groups are brought into close proximity. RNA compaction requires metal cations and polyamines that accumulate in and around the folded RNA. RNA folding is generally understood to be a hierarchical process in which RNA proceeds from an unfolded state (U) to one or more intermediate states (I) comprised of monovalent cation-mediated secondary structure but no stable tertiary structure, and finally to its native state (N) which includes stable tertiary interactions, often mediated by multivalent cations, proteins/peptides, or other co-factors.[15] For large RNAs, the importance and complexity of the transition from I to N are compounded, as many more negatively-charged phosphates must be brought into close proximity in order to achieve the globular native structure. Diffuse cations remain hydrated and make primary contributions to global stability by mitigating electrostatic repulsion of the negatively charged backbone. Chelated ions are less abundant, but are often essential for achieving fully-native RNA tertiary structure. Multivalent cations are particularly useful in these chelated ion interactions due to their multiple positive charges, which can neutralize an equivalent number of phosphate negative charges and bring together regions of RNA that are distant in primary sequence and secondary structure. The interactions of cations with RNA are subtle and polymorphic, and depend on multiple factors, including RNA sequence and structure, cation type, and presence of other cations in the system.

Common conceptual practice for RNA-cation interactions is to partition RNA-interacting metals into two limiting modes: ‘diffuse’ and ‘site-bound’.[16] Diffuse cations

are commonly hydrated, abundant, and retain near bulk-solution like mobility. They interact with RNA via weak but numerous long-range electrostatic interactions. Diffuse cations, due to their overwhelmingly larger populations, make the primary contributions to stability of folded RNAs. Site-bound ions interact strongly with the RNA at short distances, which vary depending on the type of cation but often involve first shell interactions. Mobilities of site-bound ions are low, and often determined by those of the RNA itself. Highly chelated ions, with two or more first shell ligands contributed by RNA, are the least abundant but in some cases make important contributions to specific local or even global conformations. Site-bound ions are sometimes required to access the native state. Site-bound cations are sometimes elevated to artificial significance because they are readily observable by physical techniques.

1.1.1 The special role of Mg^{2+} in RNA folding

Mg^{2+} is uniquely suited as a structural cofactor for RNA. Mg^{2+} was seen early on to be play an important role in tRNA structure.[17-19] It is now recognized that Mg^{2+} plays important roles in folding of essentially all large RNAs.[15,20,21] Hexahydrated Mg^{2+} interacts with RNA in both restricted and diffuse modes. Chelated Mg^{2+} ions, while fewer in number, are essential for achieving native, collapsed RNA states[22,23]. Some ribozymes appear to utilize Mg^{2+} ions to assist directly in catalysis by contributing to transition state stabilization.[24,25] Magnesium is the most common divalent cation in biological systems, and is widely available on earth's surface (2% of the earth's crust). Mg^{2+} is highly soluble near neutral pH [K_{sp} of $Mg(OH)_2 = 10^{-12}$] and is insensitive to O_2.

Mg^{2+} orients and polarizes first shell water molecules, activating them to molecular recognition and enzymatic mechanism. Mg^{2+} has a small ionic radius, but high charge density [Mg^{2+}_r (ionic radius) = 0.65 Å, Ca^{2+}_r = 0.99 Å, Na^{+}_r = 0.95 Å, K^{+}_r = 1.52 Å].[26-29] In aqueous solution, the first coordination shell of Mg^{2+} contains six tightly-packed water molecules with octahedral geometry. These water molecules exhibit increased acidity compared to bulk water, and also have elevated hydrogen-bond donating potential (pK_a of $Mg(H_2O)_6^{2+}$ = 11.4, pK_a of $Na(H_2O)_{6\text{-}8}^{+}$ = 14.4, pK_a of H_2O_{bulk} = 15.7).[30] Mg^{2+}-coordinated waters are also compressed and electro-restricted compared to hydrates of other biological cations, giving a large negative partial molal volume to Mg^{2+} in water (Mg^{2+}_V = -30 mL/mol; Na^{+}_V = -5.7 mL/mol).[31] Oxygen atoms of the waters are directed inwards toward the metal center and the protons are directed outwards, available for abstraction or donation. The dynamics of these water molecules are suppressed; exchange of water from the first shell of Mg^{2+} is almost four orders of magnitude slower than from the first shell of Na^{+}.[32] The enthalpy of hydration of Mg^{2+} is very large in magnitude (-450 kcal/mol) compared to other biologically-relevant cations (Na^{+}, -100 kcal/mol).[27]

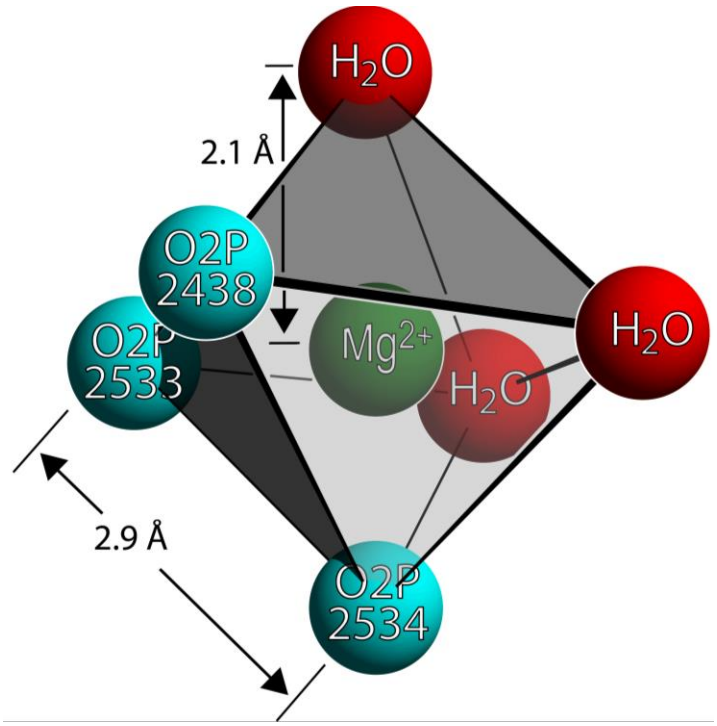


Figure 1.1. A Mg^{2+} ion chelated by RNA (Mg^{2+} 8001 from 23S rRNA of the *H. marismortui* LSU rRNA; PDB entry 1JJ2). This Mg^{2+} ion (green sphere) is octahedral, with three first shell phosphate oxygens of the rRNA (cyan) and three first shell water oxygens (red). Mg^{2+}-oxygen distances are around 2.1 Å. Mg^{2+} coordination tightly packs oxygen atoms, imposing oxygen-oxygen distances of around 2.9-3.2 Å. For clarity the radii of the spheres are reduced from the van der Waals radii of the atoms, and have no physical significance. Reprinted with permission.[14]

RNA and Mg^{2+} participate in well-defined coordination regimes. Chelated Mg^{2+} increases local RNA rigidity by tightly packing functional groups in its first coordination shell (Figure 1.1). Phosphate groups are the preferred RNA ligands to Mg^{2+}.[22] When a phosphate oxygen (PO) of RNA interacts with the first shell of Mg^{2+}, the attached phosphorus atom is activated to nucleophilic attack due to shifting of electron density towards the Mg^{2+}. As a result of this phenomenon, Mg^{2+} increases RNA hydrolysis rates. The enthalpy of exchange of a first shell water molecule for a phosphate oxygen is very close to zero, even though the dehydration enthalpy is highly unfavorable. The ratio of Mg^{2+} hydrate volume (V_h = the volume of $Mg^{2+}(H_2O)_6$) to ionic volume (V_i = the volume of Mg^{2+} alone) is especially large (V_h/V_i = 400) compared to that of Na^+ (V'_h/V'_i = 25)

and other relevant cations.[28] Therefore, the effects of Mg^{2+} dehydration on RNA structure are particularly pronounced. Mg^{2+} prefers oxygen ligands, although nitrogen ligands are observed in some systems (e.g, hemes or nucleobase heteroatoms).

It is sometimes assumed that cations participate in RNA folding simply by neutralization of negative backbone charge during RNA compaction. For weakly-interacting cations such as Na^{+}, K^{+}, polyamines, or hexahydrated Mg^{2+}, this narrow electrostatic model can provide suitably accurate approximations of reality. This approximation fails for site-bound Mg^{2+} ions, which are distinct from complexes with other biologically available cations. For Mg^{2+}, specific coordination chemistry and physicochemical phenomenon are important factors in structure and stability. Site-bound Mg^{2+} ions compact, electro-restrict and polarize their first shell ligands, which interact with Mg^{2+} not only by electrostatic interactions but also by 'non-electrostatic' interactions including charge transfer, polarization and exchange correlation.

The local properties of RNA influence its Mg^{2+} interactions. RNA chain flexibility, positioning of phosphate groups, and charge density influence chelated RNA-Mg^{2+} interactions. Mg^{2+} forms site-bound complexes more readily with single-stranded RNA[33,34] and compacted RNA than with double-stranded RNA. The ribosome provides a useful case study of the frequency of site-bound Mg^{2+} ions in folded RNAs. Around 20% of RNA PO atoms within 20 Å of the Peptidyl Transferase Center (PTC) of the ribosomal Large Subunit (LSU) form first shell interactions with Mg^{2+}.[23,35] By contrast, the RNA near the surface of the LSU shows very few first shell interactions with Mg^{2+}. The frequency of site-bound Mg^{2+} in compact RNAs can vary widely. RNA in the vicinity of

catalytic sites tends to be Mg^{2+}-rich, in part due to Mg^{2+} involvement in RNA catalytic mechanisms.

Energy decomposition analysis demonstrates that first shell RNA-Mg^{2+} interactions have significant 'non-electrostatic' components, which are important determinants of structure and stability.[36] Non-electrostatic components of the energy include polarization, charge transfer and exchange correlation (defined by Natural Energy Decomposition Analysis [37,38]). These contributions can be significant, and are related to the type of cation, the type of RNA ligand, and specific geometry of the coordination complex.[36,39-41] The net binding energy of a site-bound Mg^{2+} is composed of favorable electrostatic and 'non-electrostatic' components between cation and ligands balanced by unfavorable dehydration and ligand-ligand contributions (Figure 1.2).

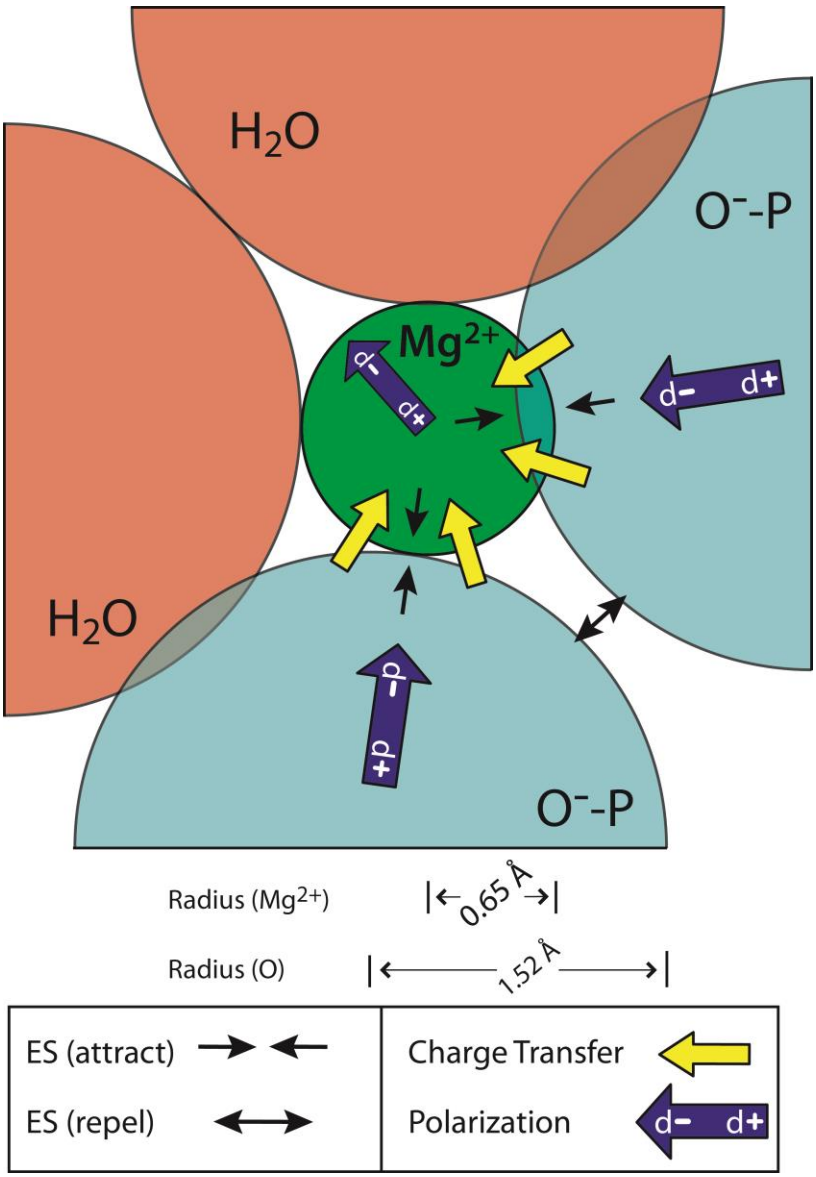


Figure 1.2. Interactions of a magnesium ion with two anionic phosphate oxygen atoms of RNA and four water molecules (the axial water molecules are omitted for clarity). Arrows represent electrostatic, polarization, and charge transfer components of the interaction energy. Only the major components of the interaction energy are shown. The exchange term, which is favorable but significantly weaker than the charge transfer and polarization terms, is omitted from the schematic diagram for clarity. The atoms are colored as in Figure 1.1. Interaction arrows are not to scale. Adapted with permission.[14]

The non-electrostatic components of site-bound interaction energies (polarization and charge transfer) make larger contributions for Mg^{2+} than for other relevant cations.[36] Non-electrostatic components are negligible for Na^{+}, K^{+}, Ca^{2+} and polyamines due to long cation-ligand distances and low charge densities.

Intermediate, near-native folding states of large RNAs are accessible in the presence of high concentrations of monovalent cations alone, without Mg^{2+}. These quasi-folded RNAs contain native-like tertiary interactions (i.e., native long-range RNA-RNA interactions) but are not true native states. These I states lack sites for chelated Mg^{2+} ions,[42] which assemble only in the presence of Mg^{2+}. RNA conformation and site-specific Mg^{2+} binding are interrelated. In absence of Mg^{2+}, RNA cannot access certain conformation spaces, including states in which multidentate chelation of Mg^{2+} is possible.

The Mg^{2+} clamp[36,43] is an excellent example of how magnesium ions induce binding site formation. A Mg^{2+} clamp is formed by two phosphates from adjacent nucleotides. Non-bridging phosphate oxygens penetrate the first shell of a common Mg^{2+} ion (Figure 1.3). The Mg^{2+} clamp is the most frequent mode of bidentate chelation of Mg^{2+} found in large RNAs.[36,43,44] Twenty-five Mg^{2+} clamps are found in the *Haloarcula marismortui* LSU (PDB ID 1JJ2),[11] two in the P4-P6 domain of the Tetrahymena Group I intron,[45,46] one in a self-splicing Group II intron from *Oceanobacillus iheyensis*,[47] one in the *in vitro* evolved L1 ligase,[48] and one in the synthetic M6C″ riboswitch.[49]

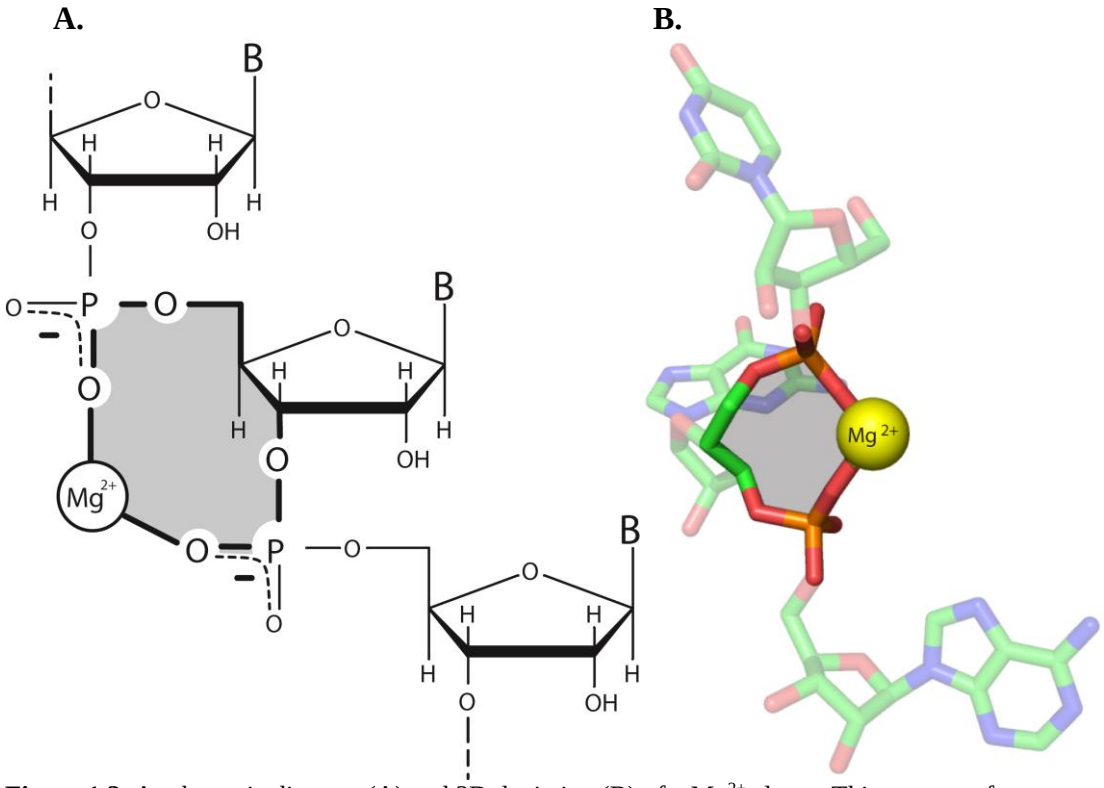


Figure 1.3. A schematic diagram (A) and 3D depiction (B) of a Mg^{2+} clamp. This structure features bidentate RNA chelation of Mg^{2+}, and is formed when adjacent phosphate groups enter the first coordination shell of a common magnesium ion. A 10-membered ring (shaded) characterizes the Mg^{2+} clamps. Reprinted with permission.[14,36]

1.1.2 Fe^{2+}: An ancient cofactor for RNA

Though Mg^{2+} occupies a reserved role in RNA folding in modern biology, mounting evidence suggests that Fe^{2+} was charged with this task during early evolution, acting instead of or alongside Mg^{2+}. Understanding the origins of RNA folding requires consideration of the conditions under which RNA originated and existed for hundreds of millions of years. To this end, recent studies have explored the relationship between geochemical conditions of the ancient earth and the folding and function of RNA.[50-52] The anoxic atmosphere of ancient earth, prior to the Great Oxidation Event (GOE), facilitated solubility of iron by preventing oxidation of soluble Fe^{2+} to insoluble Fe^{3+}. The geologic record indicates that the early oceans were devoid of O_2 and contained vast

quantities of Fe^{2+} (high μM concentrations compared to pM in modern oceans).[53] Modeling experiments and comparison of physicochemical properties of Fe^{2+} and Mg^{2+} suggest that Fe^{2+} may have interacted more readily than Mg^{2+} with RNA under pre-GOE environmental conditions. RNA catalysis and structural probing experiments support the plausibility of this substitution.[50,51] In fact, rates of ribozyme catalysis appear greater with Fe^{2+} as a cofactor than with Mg^{2+}. A model has emerged from these results under which Fe^{2+} was an essential cofactor for nucleic acids on the ancient earth, with important roles in RNA folding and catalysis, acting instead of or alongside Mg^{2+} in RNA-cation interactions. Further background regarding the substitution of Fe^{2+} for Mg^{2+} in RNA structures is discussed extensively in chapter 5.

1.1.3 Broadening the classes of RNA-cation interactions

A fine-grained conceptual framework is the most accurate way to manage classification of cations bound to RNA. The partitioning of cations into two modes (diffuse and site-bound) is useful for many applications, but is limiting in the sense that many cations fall between these two classes. We propose a classification scheme in which ions in association with RNA are more accurately divided into four classes: free, condensed, glassy and chelated. The continuum nature of the phenomena and interdependence of parameters characterizing the four classes are illustrated schematically in Figure 1.4, while 3D structures are illustrated in Figure 1.5. Each cation class is described in terms of relative population, extent of coordination, rate and dimensionality of diffusion, thermodynamic contribution to stability, and influence on specific structural states (Figure 1.4). There are many more condensed ions than glassy or

chelated ions (Figure 1.4A). For monovalent cations, the number of first shell ligands contributed by a nucleic acid can vary from zero (condensed) to eight (chelated), while for Mg^{2+}, the number of first shell ligands contributed by the RNA can vary from zero to four (Figure 1.4B). Envelopes containing condensed cations extend well beyond the van der Waals surface of the collapsed nucleic acid. Regions of RNA structure occupied by condensed cations are illustrated in Figure 1.5B by iso-surfaces for the densities of mobile charges, which were calculated using the Poisson-Boltzmann equation as implemented in APBS[54] for a solution of 100 mM KCl, 20 mM $MgCl_2$, with a 1.4 Å solvent probe at 298 ºC.

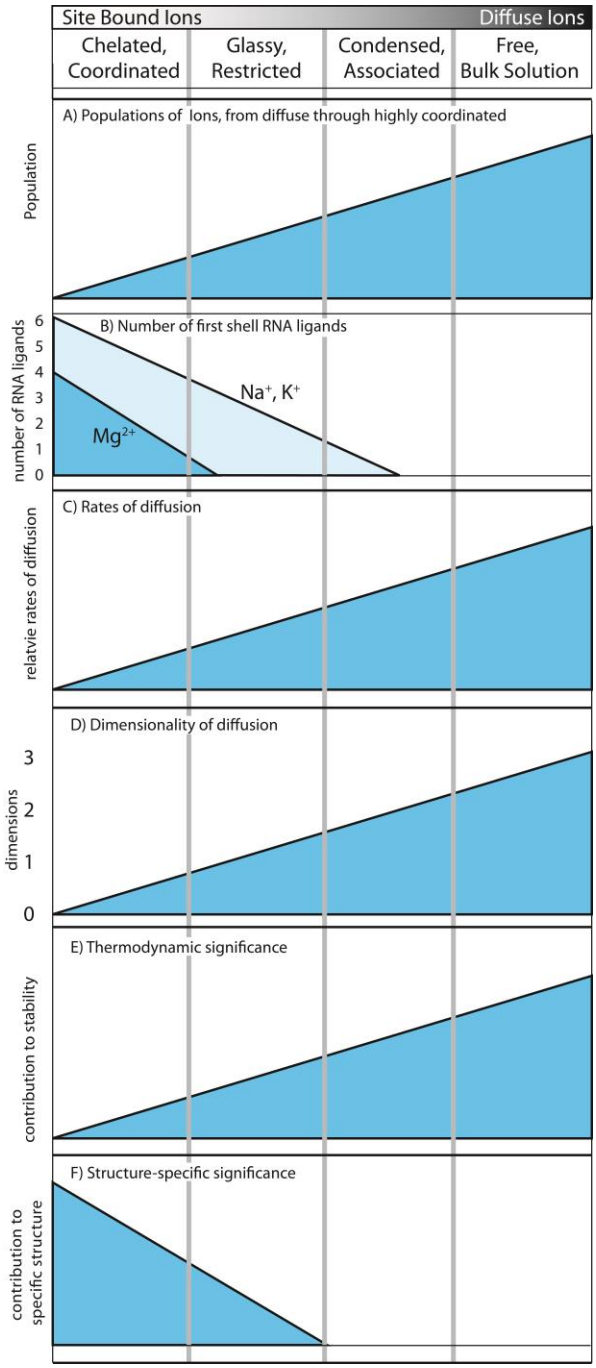


Figure 1.4. Schematic illustration of parameters describing RNA-cation interactions. A) The population of diffuse cations is much greater than the population of site-bound cations. B) Diffuse ions are not directly coordinated by RNA. The number of first shell ligands contributed by RNA to Na^+ or K^+ can generally vary from zero to six. In G-quadruplexes, monovalent cations are coordinated by up to eight first shell ligands from DNA or RNA. The number of first shell ligands contributed by RNA to Mg^{2+} can vary from zero to four. C) As number of first shell ligands contributed by RNA increases, the rate of diffusion of the cation decreases. D) As number of first shell ligands contributed by RNA increases, dimensionality of diffusion of the cation decreases. Cations in the grooves of RNA are not free to diffuse in three dimensions. E) As the number of first shell ligands contributed by RNA increases, the thermodynamic significance of cation association decreases, primarily because the number of cations with first shell RNA ligands is small. F) The specific structural significance of a cation increases with the number of first shell RNA ligands. Reprinted with permission.[14]

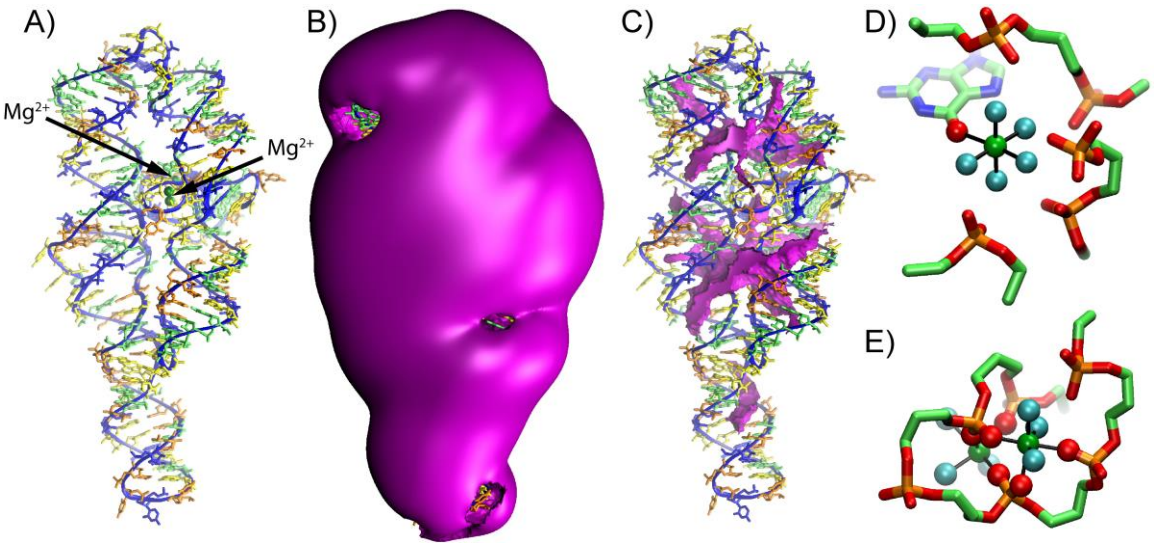


Figure 1.5. Diffuse, glassy, and coordinated cations in association with the P4-P6 RNA. A) The structure of the P4-P6 domain of the *Tetrahymena* group 1 intron RNA [PDB ID 1HR2]. Several highly coordinated Mg^{2+} ions are indicated by green spheres. The coordination of these ions is shown in detail in panels D and E. B) The envelope containing diffuse cations surrounding the P4-P6 RNA. This envelope was calculated with a mobile charge density of +0.5 M*e*. Ions within this envelope are well-hydrated, with near bulk-like diffusion properties. C) Regions of glassy cation localization within the grooves, calculated with a mobile charge density of +5.0 M*e*. D) Coordination of a glassy Mg^{2+} ion. This ion is coordinated by RNA and five water molecules. The rate and dimensionality of diffusion of this cation are restricted. A guanine base and selected backbone atoms of RNA are shown to illustrate positions and orientations of the second coordination shell of the Mg^{2+} ion (phosphorus, orange; oxygen, red; carbon, green; nitrogen, blue). Oxygen atoms of first-shell water molecules are cyan [Mg^{2+} 6766 of 1HR2]. E) Highly coordinated Mg^{2+} ions induce specific conformational states of RNA. Two trichelate Mg^{2+} ions contain tightly packed RNA PO atoms in their first coordination shells [Mg^{2+} ions 6756 and 6758 of 1HR2]. In panels D and E, the radii of the atom spheres are reduced from their van der Waals/ionic radii for clarity. Reprinted with permission.[14]

Glassy ions are closely associated with RNA (Figure 1.5C). A Mg^{2+} ion with one first shell nucleic acid ligand (Figure 1.5D) is in a glassy state. Chelated Mg^{2+} ions, with two or more first shell RNA ligands, are shown in Figure 1.5E. The RNA conformation is specifically dependent on the positions and coordination of these Mg^{2+} ions. The greater the number of first shell nucleic acid ligands, the slower the rate of diffusion (Figure 1.4C). Therefore, there are more cations in the condensed envelope with high rates of diffusion than with low rates of diffusion. Dimensionality of diffusion tracks the rate of diffusion (Figure 1.4D) because cations in bulk solution diffuse freely in three dimensions while movement of cations within helical grooves, for example, is more

restrained; cations within the grooves are glassy, with limited rates and dimensionality of diffusion. Increasing the number of first shell RNA ligands decreases both the rate and dimensionality of diffusion. Thermodynamic significance to folding of the native structure is illustrated in Figure 1.4E. The number of cations with few or no first shell RNA ligands greatly exceeds the population of cations with many first shell ligands due to considerable electrostatic screening requirements for RNA folding, and therefore the net thermodynamic contribution to folding decreases with decreasing number of first shell RNA ligands. Small numbers of ions are highly chelated by nucleic acids (Figures 1.4F and 1.5E), but these ions are of utmost importance in stabilization of specific three-dimensional structure, and cannot be substituted by other ion-types.

1.1.4 Recent technical developments in RNA-cation studies

RNA footprinting provides direct and indirect information about changes to RNA conformation induced by cations. By far the most influential RNA footprinting method developed in recent years is Selective 2'-Hydroxyl Acylation analyzed by Primer Extension (abbreviated SHAPE).[55] The final section of this chapter will be dedicated to a detailed description of the SHAPE method and its application in RNA-cation investigations. Several groups are pursuing time-resolved chemical footprinting at nucleotide resolution by hydroxyl radical cleavage on increasingly large and complex RNA assemblies.[56,57] This approach can detect time-dependent tertiary structure formation and protein interactions during folding and assembly. Local measures of folding can be combined with more global measures (SAXS, etc., see below) to provide a comprehensive depiction of folding pathways.

Global monitoring techniques provide information regarding large-scale conformational transitions of nucleic acids. Small angle X-ray scattering (SAXS) and anomalous SAXS (ASAXS) can characterize conformations and ion distributions of nucleic acids at resolutions of ~10 Å.[58] SAXS yields information about size, shape, compactness, and molecular weight of RNAs. ASAXS reports on diffuse cations within the envelope around nucleic acids, and has been used to detect differential monovalent cation distributions between B-form DNA and A-form RNA.[59] Single molecule Förster resonance energy transfer (smFRET) measures distances within or between RNA or DNA. The Pollack group has studied the Mg^{2+}-dependence of unstructured ssDNA and ssRNA with SAXS and smFRET,[60] and their results confirm that ssDNA and ssRNA have different conformations in solution, as expected from distinct sugar pucker and stacking preferences. Of more relevance here, they observe that for both ssRNA and ssDNA, charge screening by Mg^{2+} is anomalously efficient.

Woodson and coworkers[61] monitored folding of the *Azoarcus* and *Tetrahymena* Group I ribozymes in various ionic environments, monitored by SAXS. Decreases in the radius of gyration (R_g) are observed upon addition of cations, corresponding to collapse. The results suggest that total charge, not valence or charge density of the counter ions, is the most important characteristic of the cations for initial collapse from the U to I state. Polyamines induce the collapse of the *Azoarcus* ribozyme at mid-micromolar concentrations and Mg^{2+} induces collapse at high-micromolar concentrations, while monovalent cations induce require considerably higher mid-millimolar concentrations. Subtle differences in R_g for various ions demonstrate that even for a low resolution assay like SAXS, specific effects of Mg^{2+} on the collapsed state are observable. The collapsed

state is slightly more compact with Mg^{2+} than with monovalent cations or polyamines, consistent with folding to its N state. Although the ability of SAXS to reliably detect Mg^{2+}-specific effects on RNA folding remains an open question, these results support the model of collapse described above,[42] in which RNA can collapse to a near-native state in the presence of Na^{+}, K^{+} or polyamines at sufficiently high concentrations. These compact RNAs can contain many native RNA-RNA tertiary interactions but are not conformationally competent for site-binding of Mg^{2+} ions.

Temperature-controlled smFRET has been used to explore the Mg^{2+}-dependent thermodynamics and kinetics of RNA folding/unfolding in a model system.[62] Increasing [Mg^{2+}] is observed to promote tetraloop–receptor interaction by reducing both the entropic activation barrier and the net entropy of the transition with minimal effects on activation enthalpy and net enthalpy. These results appear to be consistent with a previous proposal[44] that during RNA folding, Mg^{2+} can form chelation complexes preferentially with flexible regions of RNA, locking out conformational heterogeneity.

Methods for characterizing site-bound cations are an integral part of RNA-cation studies. Fierke and coworkers[63] report that a combination of extended X-ray absorption fine structure (EXAFS) and paramagnetic line-broadening experiments by nuclear magnetic resonance (NMR) reveals a hexacoordinated Zn^{2+} interacting with a mimic of the conserved P4 helix of ribonuclease P, with inner-sphere coordination at two specific residues (average Zn-O/N distance of 2.08 Å). Christian, Harris and coworkers report attenuation of the Raman signal of symmetric vibrations of RNA PO atoms by electrostatic, hydrogen bond and inner-sphere interactions with metals.[64] They also report

cation-specific shifts (based on hardness and electronegativity) to higher wavenumbers with inner-sphere metal coordination.

Woodson and coworkers used quasielastic neutron scattering spectroscopy to reach the counter-intuitive conclusion that Mg^{2+} increases transfer RNA (tRNA) dynamics on the picosecond to nanosecond timescale while stabilizing the folded state.[65] In an unusual, minimally-hydrated state, it seems that tRNA compaction can accompany increases in local molecular dynamics. These results suggest that water lubricates conformational motions of the macromolecules, but differences in the temperature dependencies of the mobilities of folded and unfolded tRNA were interpreted to suggest that dynamics are not controlled solely by hydrating water but are significantly affected by the electrostatic nature of the RNA surface. Specifically, charge screening by counterions increases the local motion of both tRNA and a synthetic charged polyelectrolyte that does not fold into a specific structure.

Computational and theoretical tools allow the conceptualization of RNA-cation interactions at levels of theory that are often impractical for experimental techniques. Herschlag and coworkers measured the unfolding of a DNA hairpin using a simple experimental system designed to obtain interpretable data with the potential to validate or falsify various theories.[66] Measurements were made on single molecules with constrained conformations. Their results show that Poisson-Boltzmann theory, long used in modeling of nucleic acid cation interactions, can successfully account for Na^{+}-dependence of stability of a simple folded DNA. This observation is consistent with many other experiments.[36,39,40,67,68] However, in the presence of Mg^{2+}, Poisson-Boltzmann Theory, which describes ions as non-interacting point charges, fails to correctly predict energetics

of DNA hairpin formation due to an inability to address ion-ion correlations.[66,69,70] To treat correlations, Chen has partitioned cations into bound and diffuse classes, and assigned the space occupied by the two classes of ions as bound regions and diffuse regions.[71,72] This “Tightly Bound Ion" (TBI) model successfully predicts that Mg^{2+} is more efficient at charge screening than Na^{+}, beyond considerations of ionic strength alone. The high efficiency of Mg^{2+} screening is most pronounced for compact folded structures. The TBI method yields good agreement with experimentally observed salt dependence of stabilities for several model systems. Accurate theoretical models of RNA-Mg^{2+} interactions are elusive, and are the focus of many ongoing studies.

1.2 – The ribosome

1.2.1 Structure of the modern ribosome

The ribosome is a large RNA-protein complex that synthesizes all coded protein.[73,74] It is responsible for the translation of the genetic code, carried by messenger RNA (mRNA), into functional proteins through tRNA adapters. The ribosome consists of a large subunit (LSU) that binds amino-acylated tRNAs and catalyzes formation of peptide bonds, and a small subunit (SSU) that decodes mRNA. In bacteria, the LSU is comprised of two ribosomal RNAs (rRNAs)—23S and 5S—and ribosomal proteins (rProteins). The ~2,900 nucleotide (nt) 23S rRNA is the most important biopolymer in the ribosome, as its residues form the active site (PTC) where new peptide bonds are formed.[75] In this document, the 23S rRNA is often referred to as the LSU rRNA, as it is

the major RNA component of the LSU. rRNAs are the largest folded RNAs in modern biology.

The past 15 years have seen an advent in high-resolution x-ray crystal structures of the ribosome. Steitz and coworkers solved the structure of an archaeal LSU,[11] and shortly thereafter Yonath released a structure for the LSU of a bacterium.[76] The Ramakrishnan group solved the structure of a full bacterial ribosome from *Thermus thermophilus* complexed with mRNA and tRNA.[75] Most recently, structures have been released at increasingly higher resolutions for the complete ribosomes of simple eukaryotes by the Yusopov group.[77,78] These high-resolution structures have provided a wealth of information regarding the assembly, evolution, and function of the ribosome, including the exact coordination patterns of dozens of associated Mg^{2+} ions. Ribosomes are stabilized in large part by a network of RNA-Mg^{2+} interactions.[43,75] Mg^{2+} is crucial for folding and peptidyl transferase activity of the LSU.[23,74,79,80] However, no metal ions are observed in the LSU active site, setting its catalytic mechanism apart from that of most known ribozymes. Further background regarding the interactions between Mg^{2+} and rRNA is provided in chapters 3 and 4.

One of the most important discoveries made possible by ribosomal crystal structures was the confirmation that the ribosome is a ribozyme; ribosomal proteins are not found within 18 Å of the active site.[81] Ribosomes are universally conserved across all domains of extant life, demonstrating that they are evolutionarily very old.[13] The ribosome is, at its essence, an ancient ribozyme; the LSU active site is comprised entirely of RNA.[82] The LSU acts primarily as an 'entropy cage', positioning tRNA substrates in a conformation in which the peptidyl transferase reaction is kinetically more probable,

thereby greatly accelerating the reaction rate by reduction of entropy.[82] There is debate regarding which residues of the LSU rRNA, if any, are involved directly in the peptidyl transfer reaction through covalent bonding with substrates during any part of the reaction mechanism.[83,84] The growing polypeptide chain is guided through a pore in the ribosome called the 'exit tunnel' to the exterior of the LSU.[85-87]

1.2.2 The ribosome and the RNA world hypothesis

The RNA world hypothesis is a widely accepted general model for the origin and early evolution of life on earth.[1-7,88] It posits that RNA was life's original biopolymer due to its dual informational and functional capabilities; RNA is able to store and convey genetic information, and also to perform enzymatic catalysis. Other biopolymers evolved later that largely subsumed these roles; DNA became the primary system of information storage, while protein assumed many functional roles. From these transitions emerged the modern paradigm of the Central Dogma of Molecular Biology, in which RNA remains the central biopolymer.[89] However, many relics of the RNA world remain deeply entrenched in modern biology.[90] The discovery of RNA enzymes, termed 'ribozymes', rekindled interest in the RNA world hypothesis.[91,92] Most ribozymes catalyze phosphoryl-transfer reactions which ligate, cleave, or splice RNA.[93-95] Some ribozymes have been produced synthetically that widen the functional repertoire of RNA.[96,97]

Because the ribosome synthesizes all coded protein in modern biology, its emergence is one of the fundamental questions of the RNA world model.[98] The origin of translation machinery represents the transition from a biology dominated by RNA to an RNA/protein world in which proteins appropriate many functional aspects. Many lines of

evidence support the ancient nature of the ribosome, including a nearly universal genetic code[99,100] and universally conserved molecular structures,[101-103] assemblies,[35] biopolymer sequences,[104,105] and chemical processes.[83,106,107] In particular, the residues that form the PTC are universally structurally conserved across all branches of the tree of life, suggesting that these structures originated prior to the last universal common ancestor (LUCA).[108] The PTC is thought to predate coded protein.[35,98,109-111] This would place the PTC as a direct biochemical link to the distant evolutionary past. Reconstructed models of the biochemistry present in LUCA suggest that the common descendent already possessed sophisticated translation machinery.[112,113] The core of the LSU, including the PTC, is particularly rich in Mg^{2+} ions.[23,35,43] A more detailed discussion of ribosomal origins and evolution is provided in chapter 3.

1.3 – SHAPE

1.3.1 RNA footprinting with SHAPE

SHAPE (Selective 2'-Hydroxyl Acylation analyzed by Primer Extension) is a single nucleotide-resolution RNA footprinting method that provides quantitative information about base pairing and secondary structure.[55,114] Developed and championed by Weeks and coworkers,[114] SHAPE exploits the variable susceptibility of the 2'-hydroxyl (2'-OH) group of RNA to reaction with a small electrophilic molecule, which acylates exclusively at the 2'-OH position. Reactivity of the 2'-OH is modulated by its proximity to PO atoms of the RNA backbone, which alters nucleophilicity of the hydroxyl. When a nucleotide is constrained by canonical Watson-Crick base-pairing interactions, the

interatomic distance between its 2'-OH and PO is small, and consequently that nucleotide is largely unreactive to the SHAPE modifying agent due to suppressed nucleophilicity. For an unpaired, flexible nucleotide, this interatomic distance fluctuates considerably, and the nucleotide is able to access conformations in which the 2'-OH is more nucleophilic and therefore more susceptible to acylation. The reactivity of nucleotides that are constrained by unusual base pairing or other tertiary interactions depends largely on the 2'-OH/PO interatomic distance induced by the constraining interactions, so these positions exhibit variable reactivity. Different SHAPE reagents have distinct advantages. N-methylisatoic anhydride (NMIA) was the original acylating agent used in the earliest applications of SHAPE. NMIA reacts relatively slowly (tens of minutes for complete reaction) and exhibits a broad, general utility.[55] Other reagents, such as 1-methyl-7-nitroisatoic anhydride and benzoyl cyanide, react to completion in minutes or seconds, and are often used in time-resolved SHAPE experiments, allowing snapshots of short-lived folding intermediates.[115,116] NMIA is used exclusively in the experiments described in later chapters of this document. Reaction with the 2'-OH occurs at the benzylic carbonyl of NMIA, and is in competition with a slower hydrolysis reaction in aqueous solution. Both acylation and hydrolysis reactions are irreversible, as they result in loss of a molecule of CO_2.

SHAPE is ultimately an averaging experiment; flexible nucleotides are modified more frequently overall than those that are constrained, though base-paired positions may still exhibit minor reactivity. If the population of RNA molecules in a sample is conformationally heterogeneous, the resulting SHAPE reactivities will represent an average of the reactivities of nucleotides in each conformational state. Therefore, it is

important to anneal the RNA prior to modification under conditions that are expected to induce a homogeneously-folded population of RNA polymers. SHAPE reagent concentrations are optimized such that each RNA molecule in a sample is modified only a few times (single-hit conditions are ideal).

SHAPE modifications to RNA are detected by primer extension, similar to earlier footprinting techniques.[55] A DNA primer, often fluorescently-labeled, is annealed to the RNA. For short RNAs, a single primer is sufficient, but RNAs longer than a few hundred nucleotides require multiple primers spaced throughout the sequence due to decay caused by early termination (enzyme fall-off or over-modification). Reverse transcriptase extends the DNA primer 5' to 3', 'reading' the RNA towards its 5' terminus until a 2'-adduct (SHAPE modification) is encountered. At this point, the reverse transcriptase is unable to pair a nucleotide across from the acylated position, and reverse transcription (RT) is terminated. The resulting complementary DNA (cDNA) product represents the n+1 position of the acylated RNA nucleotide. The cDNA library is analyzed by capillary or gel electrophoresis.

Processing SHAPE data is a multi-step process that includes alignment and assignment of SHAPE peaks to sequencing data, peak integration, background subtraction, decay and scaling corrections, and normalization. Multiple software suites have been generated to process SHAPE data, and different RNA systems often present unique data processing challenges.[117-119]

Because SHAPE reagents react with the RNA backbone rather than the nucleobase, SHAPE provides information about the conformation of every nucleotide. This affords SHAPE with an advantage over most RNA footprinting techniques, which

often only interrogate certain nucleotides or limited regions of the folded structure, yielding piecemeal data sets. Hydroxyl radicals, due to their high reactivity, probe only the nucleotides on the exterior of a folded RNA. DMS, CMCT, and kethoxal footprinting only modify certain unpaired bases; DMS methylates only at positions where the N1 of adenosine or N3 of cytosine are accessible to the modifying agent, CMCT primarily acylates exposed uridine residues, and kethoxal reacts solely with exposed guanine residues. Because SHAPE reagents react with the backbone of RNA, SHAPE reactivity is largely insensitive to base identity.[120] SHAPE data are broadly used to validate RNA secondary structure and monitor RNA structural transitions. SHAPE has been used to probe a wide variety of RNAs, including ribozymes,[50,121] RNA specificity domains,[122] tRNA,[55,123] riboswitches,[124] and large viral RNAs.[125-127] Recently, development of soluble, slow-reacting SHAPE reagents has allowed for in-cell SHAPE, with which structure interrogation can be carried out *in vivo*.[128] Data can also be used in prediction of secondary structure for RNA systems of unknown conformation,[129,130] though critics of this utility advise caution regarding SHAPE's predictive capabilities.[131] Das has been the most prevalent critic of the SHAPE technique, particularly regarding the use of SHAPE data in structure prediction algorithms. The Das group has provided evidence that SHAPE-weighted RNA structure prediction may yield incorrect structures for a variety of small and large RNAs of known structure.[131] They propose an alternative structure-probing strategy, the 'mutate-and-map' method, which involves systematic mutation of each nucleotide position of the RNA, and subsequent modification of all mutant RNAs by a chemical probe (ie. DMS or SHAPE reagents).[132-134] Comparison of data from mutant and parental RNA constructs can elucidate which nucleotide participates in long-range

base-pairing interactions with the mutated base. While this technique has great potential, generation of a comprehensive dataset requires production of a large library of mutant RNA sequences and massively parallel experimentation, which is impractical for very large RNA systems such as rRNAs. Criticism of SHAPE has generally been limited to its application in RNA structure prediction; no known objection has been expressed regarding utilization of SHAPE data in confirmation of independently-developed model RNA structures or in monitoring of structural transitions.

rRNA has previously been probed using footprinting techniques. Early primer extension experiments elucidated 23S residues involved in interactions with tRNA, leading to identification of the active site before crystal structures were available.[135] SHAPE has been performed on assembled yeast ribosomes,[136] deproteinized *E. coli*, *C. difficile*, and *H. volcanii* rRNA,[129,137] and also domain III (DIII) of the *T. thermophilus* LSU rRNA.[118]

1.3.2 Utility of SHAPE to monitor cation-dependent effects on RNA folding

SHAPE is occasionally used to monitor RNA structural transitions induced by cations, small molecules, or proteins.[115,138,139] This requires comparison of multiple SHAPE data sets for the same RNA system collected under two or more conditions, which we refer to as 'comparative SHAPE'. By varying the concentration of RNA folding agents in SHAPE experiments, induced effects on RNA structure can be monitored. These effects may be direct, occurring at or near positions which interact directly with the folding agent, or indirect, in which case interaction of the folding agent with one part of the RNA may induce conformational changes that bring structurally

distant RNA regions into close proximity (ie. a loop-loop or loop-receptor interaction). Comparative SHAPE monitors induced folding from RNA folding states comprising only secondary structure to native or near-native states, or between multiple biologically-relevant conformations. Here, we review several studies that utilize SHAPE to monitor cation-induced conformational changes.

Shortly after developing the SHAPE method, Weeks and coworkers described Mg^{2+}-dependent folding of two small RNA systems. Comparative SHAPE results for the 154 nt *B. subtilis* RNase P specificity domain revealed significant conformational changes upon addition of 6 mM Mg^{2+}.[115] The induced changes, primarily decreases in reactivity, are mostly located at nucleotides involved in tertiary interactions, specifically complex interhelical junctions. The Weeks group also examined cation effects on yeast $tRNA^{Asp}$, collecting SHAPE data at a variety of Mg^{2+} concentrations between 0-20 mM.[123] Armed with a large number of [Mg^{2+}] data points, they were able to monitor the Mg^{2+} response of individual nucleotides and estimate midpoint Mg^{2+} values for certain regions of $tRNA^{Asp}$, including the D-loop and anticodon stem. A Mg^{2+} midpoint of 0.80 mM was identified for a complex rearrangement between a partially-folded intermediate structure and the canonical, native tRNA structure. Detailed unfolding of $tRNA^{Asp}$ induced by the cationic aminoglycoside tobramycin (total charge: +5) was also reported.

Riboswitches have been a common target of RNA-cation SHAPE studies. Winkler and coworkers examined a largely-uncharacterized 265 nt riboswitch that they term the "M-box", which is interrelated with cations on multiple levels.[140] This bacterial riboswitch is involved in Mg^{2+} homeostasis, controlling regulation of Mg^{2+} transport genes. SHAPE experiments on M-box RNA annealed in 0-10 mM Mg^{2+} contribute to the

determination that Mg^{2+} association induces a drastic conformational change in the riboswitch. The structure seems to switch between two different helical arrangements, involving significant base-pairing rearrangement. This conclusion is supported by a wide array of other *in vitro* and *in vivo* experiments, which together suggest that Mg^{2+} is the effective ligand of this riboswitch.

Lafontaine and coworkers were able to detect formation of two long-range tertiary interactions in a lysine riboswitch of *B. subtilis*.[138] These loop-loop and loop-receptor interactions have Mg^{2+} midpoints around 1 mM, and are fully-formed at 5 mM Mg^{2+}. In low Mg^{2+}, lysine, the ligand of this ribozyme, is seen to induce similar structural changes. The Sanbonmatsu group utilize comparative SHAPE data as their primary evidence that the S-adenosyl methionine (SAM) riboswitch requires both Mg^{2+} and SAM in order to form the native, bound state.[141] Recently, Weeks and coworkers performed in-cell probing of the adenine riboswitch with the SHAPE reagent 1M7,[142] and compared its observed *in vivo* conformation with that of multiple *in vitro* samples prepared at several Mg^{2+} concentrations. They conclude that the cellular environment stabilizes the tertiary structure of the adenine riboswitch, surpassing the level of stabilization observed at high [Mg^{2+}]. This conclusion is supported by experiments performed by the Bevilacqua group, who implemented SHAPE experiments on $tRNA^{Phe}$ under *in vitro* conditions designed to mimic the cellular environment.[143] They use SHAPE data collected in low [Mg^{2+}] at increasing temperatures to complement thermal denaturation studies which exhibit higher cooperativity of tRNA folding in solutions containing various molecular crowding agents and cosolutes.

Viral RNAs have also been the subject of several cation-dependent SHAPE studies in an effort to identify target regions for therapeutics. A study on a 63 nt fragment of the RNA genome of influenza A virus was undertaken by Turner and coworkers.[144] They apply SHAPE, along with multiple other footprinting techniques, to discover that this short viral RNA exists in equilibrium between two distinct conformations. The dominant conformation is dependent on presence of multivalent cations (5 mM cobalt hexamine or 10 mM Mg^{2+}). Recently, the Martinez-Salas group utilized SHAPE to probe the structure of a picornavirus internal ribosome entry site (IRES) element from foot-and-mouth disease virus.[145] IRES elements interact with initiation factor proteins to target regions of the viral genome for translation by the ribosome. Results from Martinez-Salas and coworkers demonstrate Mg^{2+}-induced tertiary folding of the picornavirus IRES RNA in protein-free samples to a structure similar to the protein-bound conformation.

The cation-dependence of SHAPE reactivities appears to be quite general and informative, and has been demonstrated for a wide variety of small RNA systems. Though commonly used in tandem with other structural biology techniques, SHAPE data alone provide a wealth of information in terms of structural validation and transitions.

1.4 – Overview of thesis

The overall goal of this work is the exploration of interactions between rRNA and divalent cations, with the objective of increasing knowledge regarding the roles of divalent cations in the origins, evolution, and assembly of the ribosome. The powerful RNA footprinting method SHAPE is the primary tool applied to observe cation-induced

RNA conformational changes. Chapter 2 details the manipulation and generation of several RNA genes and constructs for use in RNA-cation studies, including rRNA and ribozyme constructs. We focus on ancient rRNA in chapter 3, describing the design and characterization of a consensus model of the ancestral ribosome, based on the modern LSU rRNA of bacteria. Despite excision of roughly 80% of extant LSU rRNA, the remaining ancestral fragments form predicted structures, including association with Mg^{2+}. In chapter 4, the scope is expanded to include the entire bacterial LSU rRNA. Mg^{2+} effects on the structure of the LSU rRNA are elucidated with nucleotide resolution. Mapping these responsive sites with respect to tertiary interactions observed in the fully-assembled LSU provides an understanding of the degree to which Mg^{2+} alone determines LSU conformation, and structural contributions of rProteins are inferred. We then compare the effects of Mg^{2+} and Fe^{2+} on ancient and extant rRNA in chapter 5, in order to examine their assumed conformations under anoxic ancient earth conditions, linking the origins of modern biology to the geological record. The substitution of Fe^{2+} for Mg^{2+} in rRNA systems is discussed in detail. Chapter 6 outlines the potential of SHAPE data to identify recognized RNA motifs based on observations and analyses of the library of rRNA SHAPE data generated in the preceding experiments.

CHAPTER 2

DESIGN, MANIPULATION, AND ITERATION OF DNA GENES AND RNA CONSTRUCTS

2.1 – Introduction

2.1.1 *Thermus thermophilus* 23S rRNA

Thermus thermophilus (*T. thermophilus*) is a highly thermophilic, gram-negative bacterium of the class *Deinococci*.[146] *T. thermophilus* was first isolated from a Japanese hot spring, and its optimal growth temperature is between 65-72 °C, classifying it as an 'extremophile'. Multiple strains of *T. thermophilus* have been isolated. Throughout this work, all *T. thermophilus* ribosomal components described and utilized are those of the HB8 strain, for which complete genomic sequence information is available.[147] The GC content of *T. thermophilus* DNA is 69%, consistent with the organism's high growth temperature. Proteins of *T. thermophilus* are exceptionally thermally stable, exhibiting extraordinary resistance to heat-induced denaturation. *T. thermophilus* is a commonly used bacterial model system for many biological and biochemical applications.

The *T. thermophilus* 70S particle was the first complete ribosome structure determined at high resolution (<3 Å).[75] Structures of the ribosome available prior to the publication of the *T. thermophilus* 70S structure were either low resolution (>3 Å) or consisted of only one subunit (LSU or SSU). The thermophilic nature of *T. thermophilus* and related stability of its biomolecular components aided in the acquisition of a highly-ordered crystal, since cellular components are able to withstand high optimal growth

temperatures. The 23S rRNA of *T. thermophilus* HB8 is ~2900 nt long with a GC content of over 63%. The full sequence is provided in Appendix A.

The 23S rRNA comprises nearly the entire catalytic core of the LSU.[35] The active site of the LSU, where the peptide bond is formed, is comprised entirely of 23S rRNA residues. *in vivo* the 23S rRNA associates with over 30 rProteins, many Mg^{2+} ions bound in various modes, and the 5S rRNA to form the assembled LSU. In the 3D structure of the *T. thermophilus* ribosome, dozens of Mg^{2+} ions associate with the 23S rRNA through first-shell interactions, and many hydrated Mg^{2+} ions are also observed to bind specifically.[75]

Here, we describe the manipulation of an *in vitro* construct of *T. thermophilus* 23S rRNA and model ancestral rRNA systems composed primarily of 23S rRNA. Site-directed mutagenesis is performed to correct non-native mutations to the 23S gene and introduce designed iterations to the ancestral rRNA system, and these constructs are overexpressed in and purified from bacterial cells. RNA coded by these genes is transcribed using run-off *in vitro* transcription, then purified and characterized by denaturing PAGE. These rRNAs are used in experiments described later in this document, and also in other publications.[51,118]

2.1.2 P4-P6 RNA

The 160 nt independently-folding RNA commonly referred to as ‘P4-P6’ is a subdomain of the *Tetrahymena thermophila* (*T. thermophila*) Group I intron, which was the first self-splicing RNA to be discovered.[148] This intron is an intervening sequence in

the 26S rRNA-coding region of a *T. thermophila* rRNA gene. P4-P6 is one of the most extensively characterized fragments of RNA, and is commonly used as a model system in RNA-folding and RNA-cation studies.[149-152] An x-ray crystal structure of P4-P6, one of the earliest high-resolution RNA structures, demonstrates specific binding of Mg^{2+} by an A-rich loop in the folded RNA.[153] The A-rich loop interacts with a hairpin-receptor to connect the two halves of the P4-P6 domain.

We demonstrate manipulation of a plasmid bearing a mutant P4-P6 gene. Two variations of P4-P6 are generated using site-directed mutagenesis: a wild type (WT) P4-P6 version, and another iteration containing a tertiary fold-destabilizing mutation to the A-rich loop.[154] Plasmids containing P4-P6 constructs have been overexpressed, purified, and used to *in vitro*-transcribe P4-P6 RNA. RNA generated from these constructs has been studied by members of the Williams group, often as control systems in binding and interaction studies.

2.1.3 Site-directed mutagenesis

The introduction of specific mutations to circular DNA constructs using designed oligonucleotides is commonly referred to as 'site-directed mutagenesis'.[155] Michael Smith received the Nobel Prize in Chemistry for the development of this technique, along with Kary Mullis (for the development of PCR). Site-directed mutagenesis is one of the most powerful tools in synthetic biology, allowing for the alteration, insertion, or deletion of short sequence regions in a circular DNA construct. Oligonucleotide primers are designed which contain desired mutations, insertions, or deletions that are not present in

the template sequence. Primers are extended *in vitro* by DNA polymerase to produce daughter plasmids that contain the target mutations. The parental DNA is selectively digested, leaving only the mutated daughter plasmid, which is then transformed and overexpressed in bacterial cells. Here, we describe the use of site-directed mutagenesis to manipulate the sequence of RNA-coding genes ranging in size from 160 to almost 3,000 nucleotides.

2.2 – Methods

2.2.1 Manipulation of *T. thermophilus* 23S rRNA

2.2.1.1 Transformation and purification

The gene coding for the 23S rRNA of *T. thermophilus* was provided in pUC19 plasmid (construct referred to here as *Tt*23S). The gene was originally amplified from genomic *T. thermophilus* DNA, and cloned into pUC19 by J.C. Bowman. *Tt*23S construct (0.2 μL of ligation reaction product) was transformed into ~70 μL of competent cells (Z-comp cells) by addition of thawed cells to DNA, followed by incubation on ice for 45-60 min. Transformed cells were plated on LB-Amp and incubated overnight at 37 °C. Single colonies from the plates were inoculated into 2 mL LB-Amp broth to create starter cultures, which were incubated overnight with shaking at 37 °C. Starter cultures were then inoculated into 100 mL cultures to be used in large scale plasmid purification. These cultures were grown again at 37 °C overnight with shaking. Sterile techniques were used consistently in handling of cells and cell cultures here, and throughout the research presented this document. *Tt*23S plasmid was purified from 100 mL cultures using the

Qiagen Endo Free Plasmid Maxi Kit, according to the provided protocol "Plasmid or Cosmid DNA Purification" with the following changes: i) the centrifugation steps (steps 14 & 15) were performed at the optional conditions (15,000 rpm for 60 min @ 4 °C), ii) in step 16, the pellet was air-dried for 15-20 min @ 37 °C, and iii) purified plasmid pellets were resuspended in 500 µL of nuclease free dH_2O. *Tt*23S plasmid concentrations and purity were verified by collection of UV absorbance spectra on a NanoDrop 2000 instrument (Thermo).

2.2.1.2 Gel purification of *Tt*23S construct

Purified DNA was run on a 2% low-melting agarose gel in 1x TAE in order to separate *Tt*23S plasmid from contaminating DNA fragments. Target bands were extracted from the gel and purified using the QIAquick Gel Extraction Kit using the included microcentrifuge protocol, with the following changes: i) gel slices were melted at 55 °C, ii) the optional wash step with Buffer QG was performed in step 9, and iii) DNA was eluted into 50 µL of nuclease-free dH_2O.

2.2.2 Site-directed mutagenesis

2.2.2.1 *T. thermophilus* 23S rRNA gene corrections

Site-directed mutagenesis primers for the *Tt*23S construct were designed by J.C. Bowman, with the following sequences: 5'-GGGAAGTGCGAATGCCGGCATGAGTA ACGATAAAGAGGGTGAG-3' (sense primer), and 5'-CTCACCCTCTTTATCGTTAC TCATGCCGGCATTCGCACTTCCC-3' (anti-sense primer). Site-directed mutagenesis

of the *Tt*23S construct was performed using the QuikChange Lightning Site-Directed Mutagenesis Kit according to the enclosed protocol, with the following modifications: i) 50 ng of *Tt*23S plasmid template was used in reactions, ii) extension time for PCR cycles was 3 min (6 kb plasmid, 30 s/kb of plasmid length), and iii) *Dpn* I-treated DNA was transformed into lab stock dH5α cells instead of XL10-Gold ultracompetent cells. For transformation, 6 μL of *Dpn* I-treated DNA was combined with 50 μL of thawed dH5α cells and incubated on ice for 1 h, followed by plating on LB Amp and overnight incubation at 37 °C. Single colonies were inoculated into 2 mL LB-Amp broth, and cultures were grown overnight at 37 °C with shaking. DNA was purified from cultures using the QIAprep Miniprep kit (Qiagen) according to the enclosed protocol, and eluted into 50 μL of nuclease-free dH_2O.

Primers for secondary PCR of DNA purified from site-directed mutagenesis targeted the 5' and 3' ends of the 23S gene. The FailSafe enzyme/buffer system were used for secondary PCR (FailSafe buffer J, EpiBio), with the following temperature cycle: Initial denaturation; 2 min @ 95 °C, 30 cycles of [1 min denaturation @ 95 °C, 1 min annealing @ 58 °C, 3 min extension @ 72 °C], and a final extension of 5 min @ 72 °C. Secondary PCR products were run on a 1.2% agarose gel in 1x TAE. Purified *Tt*23S DNA was sequenced by Operon.

2.2.2.2 a-rRNA gene iterations

Full description of the origins and design of ancestral rRNA (a-rRNA) versions can be found in Chapter 3. Template construct for site-directed mutagenesis was a-rRNA-β1 gene in pUC19, provided by J.C. Bowman. RNA sequence is provided in Appendix A.

Sequences of primers used to generate desired mutations are listed in Table 2.1. Mutagenesis was carried out with the QuikChange Lightning Site-Directed Mutagenesis kit as described above for the *Tt*23S construct, with the following exceptions: i) an extension time of 2 min was used in the amplification step, and ii) mutated DNA was transformed into XL10-Gold ultracompetent cells. Mutations were made subsequently; A246TG was carried out first, followed by G257CT (a-rRNA-β1 numbering). Mutated plasmid was cultured and purified via miniprep as described above, and purified plasmid was submitted to Operon for sequencing. Mutated a-rRNA construct was transformed, cultured, and purified via maxiprep as described above for the *Tt*23S construct.

Table 2.1. Primers used in site-directed mutagenesis of a-rRNA-β1 gene

Mutation	Sense primer sequence (5'→3')	Anti-sense primer sequence (5'→3')
A246TG	GCCGGCATGAGTAACGTGGCCGTA AGGCGCGGG	CCCGCGCCTTACGGCCACGTTACT CATGCCGGC
G257CT	CGTGGCCGTAAGGCCTCGGGAGAA CCCTCGCC	GGCGAGGGTTCTCCCGAGGCCTTA CGGCCACG

2.2.2.3 P4-P6 gene variations

Template construct was P4-P6 gene in pUC19 plasmid. Primer sequences used to generate desired P4-P6 point mutations are listed in Table 2.2. G109C and C212G mutations to generate the WT P4-P6 gene were made simultaneously using the QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Stratagene) according to the enclosed protocol, with the following modifications: i) 50 ng of plasmid template was used, ii) extension time for PCR amplification was 90s (~3 kb template, 30 s/kb). Culture preparation and DNA purification via miniprep was performed as described above for the *Tt*23S construct. WT P4-P6 construct was transformed into dH5α cells and a maxiprep-

scale purification was carried out as described above for *Tt*23S construct. The A187U mutation was performed subsequently as described above, using the WT P4-P6 gene in pUC19 as a template. All constructs were sequenced by Operon. Both WT P4-P6 and A187U constructs were transformed into dH5α cells and maxiprep plasmid purifications were performed as described above for *Tt*23S construct. Sequences of all relevant P4-P6 variants are provided in Appendix A.

Table 2.2. Primers used in site-directed mutagenesis of P4-P6 genes

Mutation	Sense primer sequence (5'→3')	Anti-sense primer sequence (5'→3')
G109C	GACTCACTATAGGGGAATTGCGG GAAAGGGGTCAACAGCCG	CGGCTGTTGACCCCTTTCCCGCAA TTCCCCTATAGTGAGTC
C212G	GACGGACATGGTCCTAACCCGCA GCCAAGTCCTAAGTCAACAG	CTGTTGACTTAGGACTTGGCTGCG GGTTAGGACCATGTCCGTC
A187U	CTTGCAAAGGGTATGGTAATTAGC TGACGGACATGGTCCTAACCCGC	GCGGGTTAGGACCATGTCCGTCAG CTAATTACCATACCCTTTGCAAG

2.2.3 *in vitro* transcription/purification of RNA

Plasmid constructs were digested at the 3' end of the RNA-coding gene with HindIII (NEB) in preparation for run-off *in vitro* transcription; 1 µg of plasmid was digested in each 50 µL reaction, incubated overnight at 37 °C. Cut plasmid was purified from digestions using the 'DNA Clean + Concentrator – 5' kit (Zymo), and eluted in 40-50 µL nuclease-free dH_2O. Digested plasmid (500 ng) was transcribed *in vitro* using the T7 MEGAscript kit (Applied Biosystems) according to the enclosed protocol (minimum 2 h reaction time). The optional TURBO DNase step was performed to digest DNA. In most cases, transcribed RNA was purified from transcription reactions using the RNeasy Mini kit (Qiagen) according to the enclosed "RNA Cleanup" protocol. In one instance, the MegaClear kit (AB) was used to purify transcribed RNA according to the enclosed

protocol, with the following changes: i) in steps 4c, 5a, and 5b, a 1 min centrifugation was performed at 14,000 g, and ii) in step 5c, a 30 s centrifugation was performed. Transcribed RNA (500 ng) was run on 5% denaturing PAGE (SequaGel UreaGel System, National Diagnostics) in 1x TBE with a 100 bp ladder (NEB).

2.3 – Results

2.3.1 *T. thermophilus* 23S gene

The *T. thermophilus* 23S gene was successfully obtained in high yield, and a single point mutation was corrected using site-directed mutagenesis. Yields from maxiprep of the *Tt*23S construct ranged from 76-160 µg of high purity plasmid. Secondary PCR was carried out using primers which flanked the 23S gene insert, and bands of expected size were observed. Gel purification was carried out in order to separate the amplified bands of expected size for the 23S gene (~3 kilo base pairs, kbp) from template DNA. Yield from gel purification was ~1 µg of plasmid, a portion of which was submitted for sequencing. Upon inspection of sequencing results, the *Tt*23S construct described above was found to have a single mutation compared to the known sequence, C1339T. Site-directed mutagenesis was used to correct this mutation (target mutation: T1339C). After purification of the mutated plasmid, 3.5 µg of plasmid was recovered. Secondary PCR was performed on the mutated plasmid DNA, which confirmed presence of an insert of the approximate size of the 23S rRNA gene (Figure 2.1). This plasmid was submitted for sequencing, and results confirmed successful

mutation of the target site back to the WT sequence (C1339). Characterization of 23S rRNA generated from this gene is described in chapters 4-6 of this document.

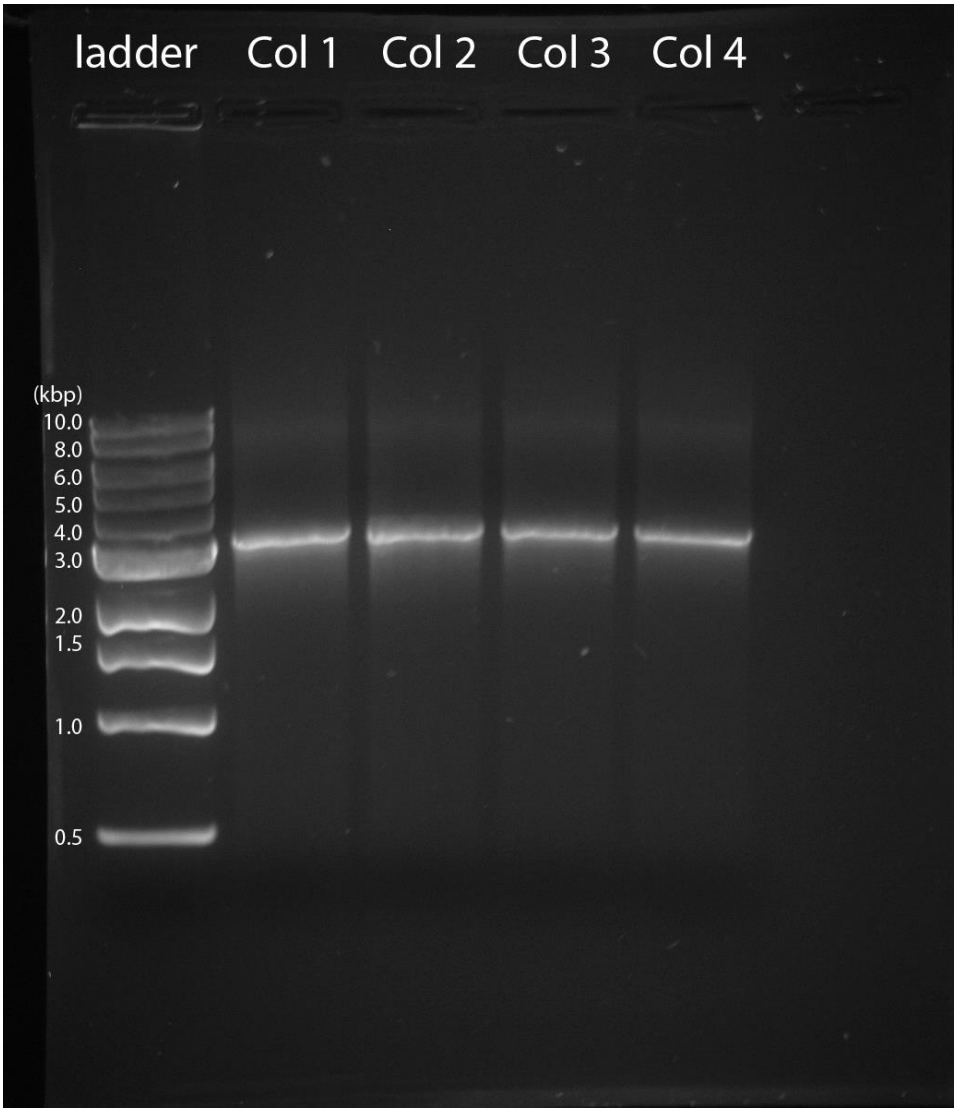


Figure 2.1. Agarose gel (1.2%) of *Tt*23S secondary PCR products following site-directed mutagenesis. Gel was run in 1x TAE with 1 kb ladder (NEB). "Col 1-4" represent plasmid purified from four individual colonies.

2.3.2 Iterations of a-rRNA

Experimentally-informed iteration of the model ancestral rRNA (a-rRNA) was carried out successfully using site-directed mutagenesis. Design and synthesis of a-rRNA

is described in chapter 3. RNase H footprinting data performed on the first *in vitro* version of a-rRNA, a-rRNA-β1, suggested that one of the gccGUAAggc tetraloops used to stitch together disparate elements of 23S-derived rRNA was not folding correctly (data obtained by L. Lie). Two mutations were introduced to a-rRNA-β1 in an effort to increase favorability of formation for this tetraloop: A246TG, and G257CT (Figure 2.2). In the transcribed RNA, these mutations add one G-C pair to the stem of the tetraloop, increasing thermodynamic favorability of tetraloop formation, and also add an additional U residue to both the 5' and 3' ends of the stem for increased flexibility, in an effort to relieve any stress or strain caused by folding of adjacent RNA regions. Yield from a maxiprep of the mutated plasmid was 96 µg. Sequencing results from this plasmid confirmed presence of all desired mutations. This new iteration of a-rRNA was termed 'a-rRNA-β2', and its characterization is described in chapter 3.

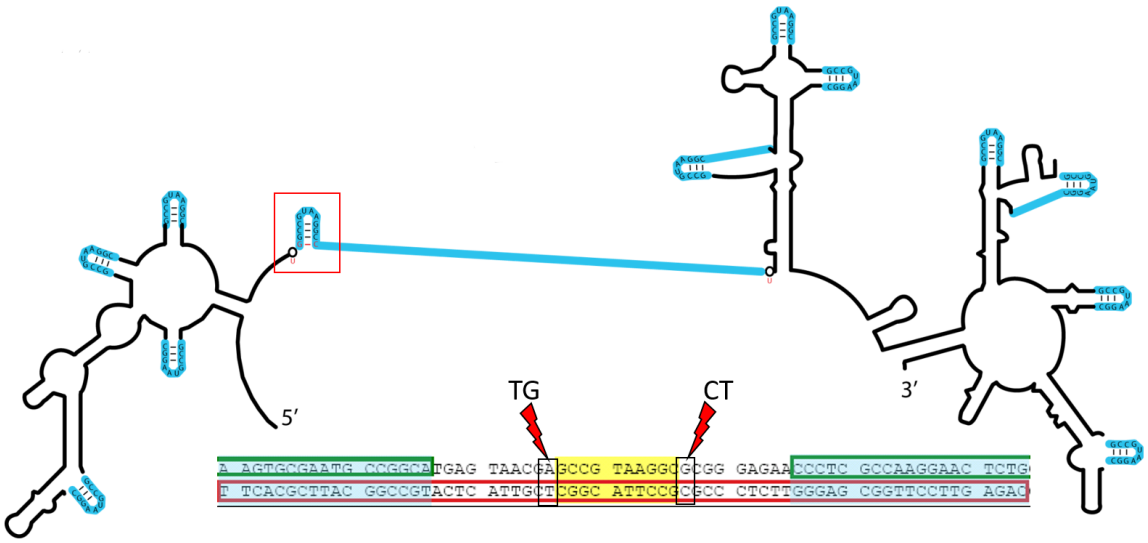


Figure 2.2. Mutations involved in iteration of a-rRNA-β2. Black line: 23S-derived rRNA retained in a-rRNA, displayed as per the canonical secondary structure. Cyan line: Stitching gccGUAAggc tetraloops. Red box: Target tetraloop for mutation. Red letters indicate mutations present in a-rRNA-β2. Inset: Region of sequence where mutations were made to generate a-rRNA-β2.

2.3.3 Generation of P4-P6 variants

Two versions of the P4-P6 gene were successfully created by site-directed mutagenesis. The P4-P6 construct available in the Williams lab contained two point mutations (C109G and G212C) which increase the propensity of the corresponding RNA to be crystallized. Positions 109 and 212 form a base pair in the folded P4-P6 RNA.[153] In order to create a WT version of the gene, these mutations were reversed to their native sequence via G109C and C212G mutations. After mutagenesis, the range of miniprep-purified plasmid was 8.3-10.8 µg. Sequencing results confirmed presence of the desired mutations. A large scale maxiprep purification was performed on this construct, termed 'WT P4-P6', from which ~306 µg of plasmid was obtained. Using the WT P4-P6 plasmid as a template, another mutation was introduced through a second round of site-directed mutagenesis; A187U is a single point mutation known to destabilize the A-rich loop of the P5abc domain, thereby destabilizing the tertiary fold of the entire domain.[154] The construct bearing the A187U mutation was termed 'P4-P6-UNFLD'. A large scale purification of the P4-P6-UNFLD construct yielded 42 µg of plasmid. WT and P4-P6-UNFLD RNAs have been utilized by others in the Williams lab in both structural and functional studies.

2.3.4 *in vitro* transcription/RNA purification

RNA was successfully generated by run-off *in vitro* transcription of four plasmid constructs: *T. thermophilus* 23S rRNA, a-rRNA-β1, a-rRNA-β2, and WT P4-P6. Recovery from purification of HindIII-digested plasmids used as templates was consistently >75%. Transcribed 23S rRNA and a-rRNA versions were purified using the

RNeasy Mini kit, whereas the MegaClear kit was used to purify WT P4-P6 RNA. Yields of purified *in vitro*-transcribed RNA ranged from 65-75 µg. Denaturing PAGE of purified RNA demonstrates overall purity of transcribed RNA; each RNA sample displays a single dominant band, with minimal smearing and secondary bands. The ladder is comprised of double-stranded DNA, and is used solely for reference and comparison purposes with similar constructs previously analyzed by PAGE. For a-rRNA-β1 and a-rRNA-β2, the most prominent band is larger than the highest ladder band. WT P4-P6 RNA exhibits a single diffuse band between 300-400 bp. Band migration and lane profile of a-rRNA-β1, a-rRNA-β2, and WT P4-P6 are consistent with previous analyses. *T. thermophilus* 23S rRNA does not enter the gel matrix in this case due to its size (~3 kb), but general appearance of this lane is also consistent with previous PAGE analyses. These results are consistent with successful production of the four described RNA constructs.

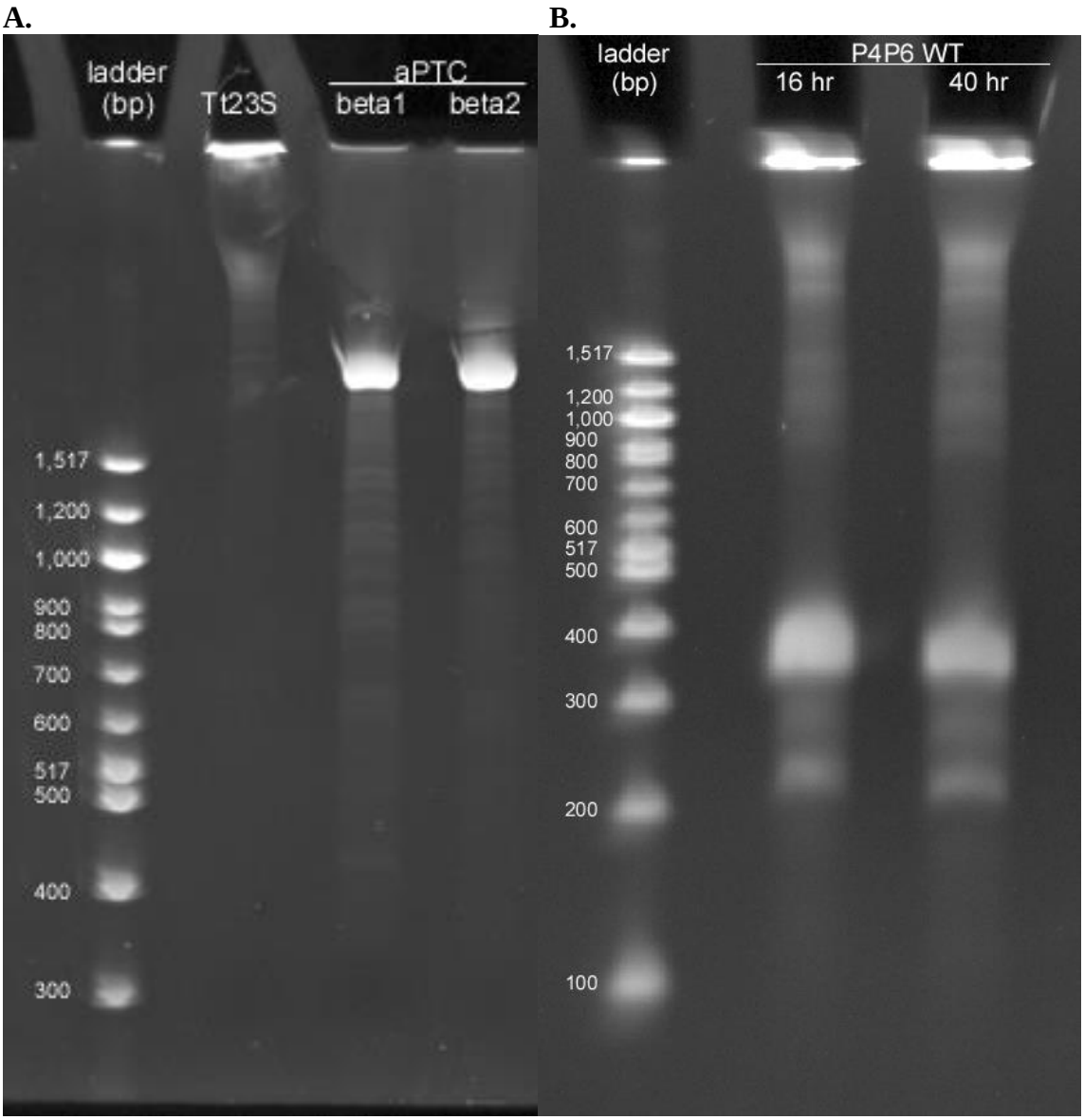


Figure 2.3. Denaturing PAGE gels of *in vitro*-transcribed RNA. RNA was generated by run-off *in vitro* transcription. 500 ng of RNA loaded in each lane. A) Tt23S: *T. thermophilus* 23S rRNA. aPTC: a-rRNA, versions β1 (beta1) or β2 (beta2). B) WT P4-P6 RNA, transcribed *in vitro* for either 16 or 40 h. 5% denaturing PAGE gel, run with 100 bp DNA ladder.

2.4 – Discussion

2.4.1 Further uses of described genes/RNA constructs

RNA generated from the DNA constructs described above has been utilized in several studies. The *Tt*23S construct was used to generate *T. thermophilus* 23S rRNA used in experiments described in chapters 4-6 of this document, studies involving interactions of LSU DIII with the remainder of the LSU rRNA,[118] and catalysis of electron transfer by RNA with Fe^{2+} as a cofactor.[51] a-rRNA-β1 and a-rRNA-β2 RNA were used in thermal stability and SHAPE footprinting studies described in chapter 3. The SHAPE results in particular informed specific revisions to future iterations of a-rRNA. P4-P6 UNFLD RNA was used in experiments related to high-affinity Mg^{2+} binding sites in RNA.[156]

2.5 – Conclusions

The work documented here describes the successful manipulation of several RNA-coding DNA plasmids. Utilized techniques include site-directed mutagenesis, DNA cloning, gel electrophoresis (agarose and PAGE), secondary amplification PCR, DNA/RNA purification, and run-off *in vitro* transcription. RNA produced from these constructs has been used extensively in published research, and the rRNA constructs in particular are characterized and discussed extensively in the following chapters.

CHAPTER 3

MOLECULAR PALEONTOLOGY: EXPLORING THE STRUCTURE OF ANCESTRAL LSU RIBOSOMAL RNA

3.1 - Introduction

Portions of this chapter are adapted from previously published work: "Hsiao, Chiaolong; Lenz, Timothy K.; Peters, Jessica K.; Fang, Po-Yu; Schneider, Dana M.; Anderson, Eric J.; Preeprem, Thanawadee; Bowman, Jessica C.; O'Neill, Eric B.; Lie, Lively; Athavale, Shreyas S.; Gossett, J. Jared; Trippe, Catherine; Murray, Jason; Petrov, Anton S.; Wartell, Roger M.; Harvey, Stephen C.; Hud, Nicholas V.; Williams, Loren Dean. Molecular Paleontology: A Biochemical Model of the Ancestral Ribosome. Nucleic Acids Research, 2013, 41, 3373-3385."[157] The author of this document contributed to this work primarily through SHAPE footprinting experiments, data analysis, figure creation/design, and overall organization of the manuscript.

3.1.1 Consensus of ancient LSU rRNA

Some components of the ribosome are highly conserved throughout extant life[108] and are considered to be among the oldest structures in biology.[104,105,158-161] The PTC, for example, is thought to predate coded protein.[35,98,109-111] If so, the PTC emerged from an ancient biological world, possibly an 'RNA World',[1-4,162] before life adopted all the processes of Crick's "central dogma".[89] In this scenario the PTC was an active participant in the origins of current biology and is one of our most direct biochemical links to the

distant evolutionary past. Several groups, including Fox,[111] Noller,[98] Steinberg,[109] Williams,[35] and Gutell and Harvey,[108] have proposed molecular-level events in early ribosomal evolution or have determined universally conserved ribosomal components (Figure 3.1). These proposed evolutionary pathways can be used to predict specific sequences and structures of ancestral rRNAs and polypeptides.

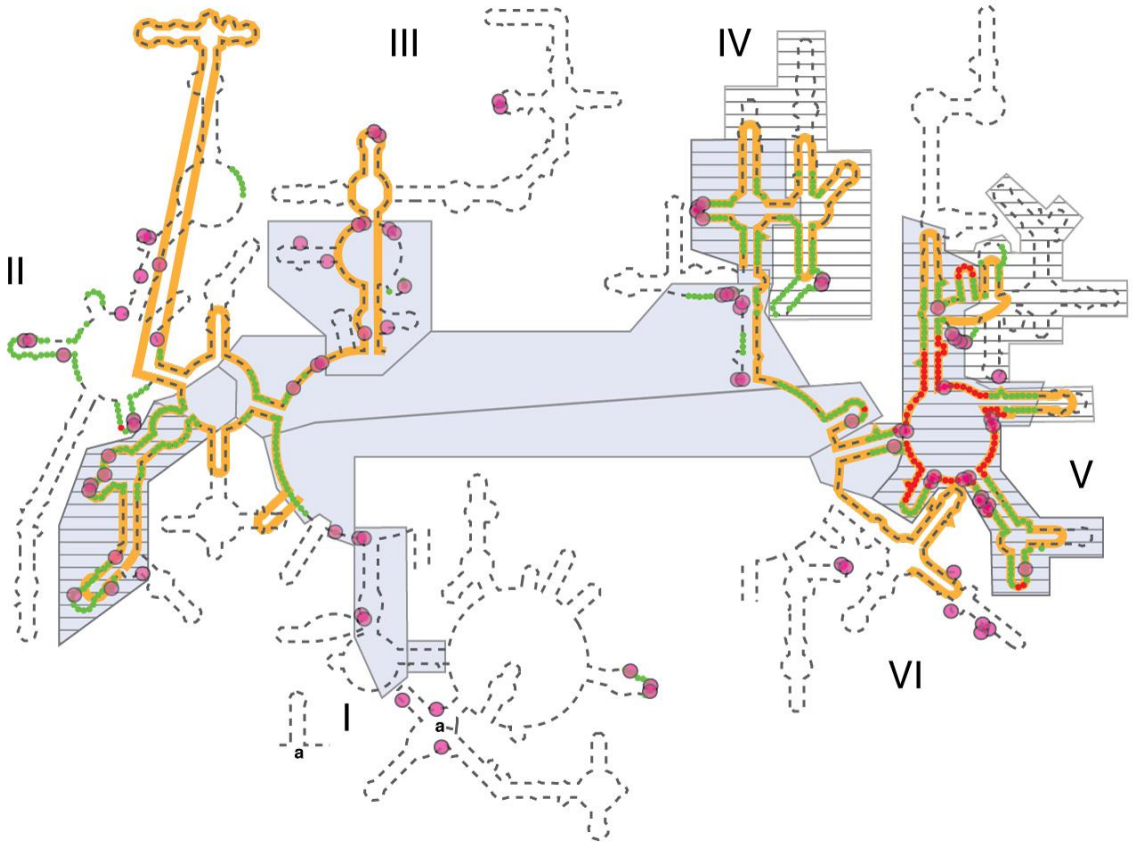


Figure 3.1. Various models of 23S rRNA evolution. The dashed line illustrates the canonical secondary structure of the *T. thermophilus* 23S rRNA. Secondary structural domains are indicated by Roman numerals. Red and green circles show the two inner shells of the ribosomal onion of Hsiao and Williams, marking the rRNA that is in closest proximity, in three dimensions, to the site of peptidyl transfer. Gray boxes are ancient according to the 'A-minor' method of Steinberg. Hashed boxes (with black horizontal lines) are ancient according to the networking analysis of Fox. Multidentate Mg^{2+}-phosphate interactions, also proposed as an indicator of ancient rRNA, are indicated by magenta circles. The orange line shows the universally conserved portions of the 23S rRNA in bacteria, archaea, eukarya, and in mitochondria, as determined by Gutell and Harvey.

3.1.2 Design of a model ancestral ribosome

Here we use a consensus of proposed evolutionary pathways and of conserved ribosomal components (Figure 3.1) to design and construct a molecular-level model of an ancestral PTC (a-PTC, Figure 3.2). This model is intended as a starting platform for an iterative hypothesis-testing approach for understanding the origins of translation. Our design process relies substantially on three-dimensional structures, which are more conserved than sequence over long evolutionary time frames.[35,163,164] The a-PTC incorporates fragments of the 23S rRNA, fragments of ribosomal proteins and divalent cations.

The ancestral rRNA fragments inferred here from consensus (Figure 3.1) are joined together to form a single RNA polymer (a-rRNA, Figure 3.2). The a-rRNA contains rRNA that forms and surrounds the PTC, which is composed of fragments from DII, DIV and DV. Using the 3D structure of the *T. thermophilus* LSU,[75] the 23S rRNA was 'shaved' to a rough sphere of around 30 Å in radius, centered at the site of peptidyl transfer. This shaving process created 13 rRNA fragments. To facilitate re-connection of these fragments to form a polymer, termini were selected preferentially in A-form helical regions of the rRNA 3D structure and were capped with stem-loops. The result is a single RNA polymer containing the most ancient 20% of the *T. thermophilus* 23S rRNA. The fragments are stitched together by stem-loops in such a way that the 3D structure of the PTC will be maintained (Figure 3.2). The a-rRNA contains rRNA that is (i) universally conserved in secondary and three-dimensional structure in extant organisms and organelles,[35,108] (ii) tightly networked by molecular interactions,[111] (iii) densely coordinated by Mg^{2+} ions,[43] and (iv) found in the central shells of the ribosomal 'onion'.

The ribosomal onion is a conceptual model in which the age of elements of the ribosome are considered to correlate inversely with radial distance from the site of peptidyl transfer.[35]

Ribosomal proteins do not engage directly in catalytic processes in the ribosome.[11,80] However, the non-globular 'tails' of some ribosomal proteins do penetrate the LSU and interact extensively with the rRNA that forms the catalytic PTC. The tails of ribosomal proteins uL2, uL3, uL4, uL15, and uL22 penetrate the 30 Å sphere that defines the a-rRNA and are highly conserved in conformation throughout the three domains of life.[77,165,166] We shaved those proteins at the 30 Å boundary used for the 23S rRNA, and included the resulting peptides in the a-PTC. The shaving process did not disrupt protein secondary structure; none of these protein tails are globular. The ancestral peptide components of the a-PTC are called a-rPeptide uL2, a-rPeptide uL3, a-rPeptide uL4, a-rPeptide uL15, and a-rPeptide uL22. It was previously proposed that these tails are more ancient than any protein with globular structure.[35]

Using *in vitro* experiments, we test the hypotheses that (i) a-rRNA adopts LSU-like secondary structure, (ii) a-rRNA in association with Mg^{2+} adopts LSU-like tertiary interactions, and (iii) a-rRNA forms specific LSU-like complexes with a-rPeptides. Components of the a-PTC were assembled *in vitro* and these hypotheses tested by various footprinting and binding assays.

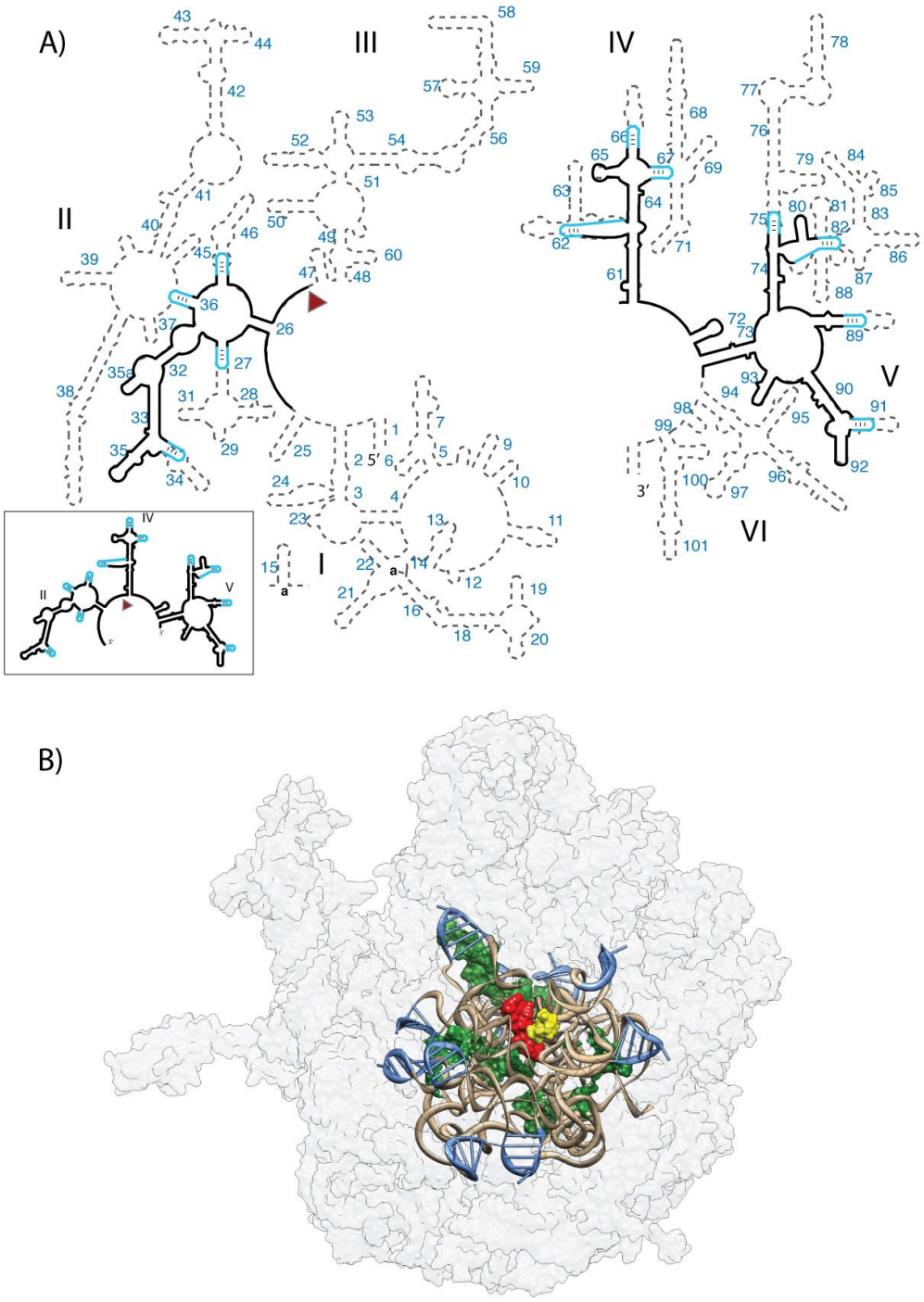


Figure 3.2. Model structures of the a-PTC. (A) Predicted secondary structure of the ancestral rRNA (a-rRNA-γ). Ancestral fragments of rRNA, indicated by black lines in the secondary structure, are derived from a consensus of rRNA evolution models. The ancestral rRNA elements are stitched together by stem-loops (cyan). The RNA sequences are from the *T. thermophilus* 23S rRNA. Helix numbers are indicated. The predicted secondary structure of a-rRNA-γ alone is highlighted in the outbox. (B) Three-dimensional model of the a-PTC. This 3D model contains a-rRNA-γ plus five a-rPeptides (ancestral fragments of ribosomal proteins uL2, uL3, uL4, uL15, and uL22). a-rRNA-γ is shown in ribbon (brown), the stem-loops are blue, and the peptides are surface representation (green). For reference, A-site (yellow) and P-site (red) substrates are shown in the figure, but are not components of the a-PTC. The modern LSU surface is shown for comparison (light gray, transparent).

3.1.3 Recursive PCR

Portions of this section are adapted from previously published work: "Bowman, Jessica C.; Azizi, Bahareh; Lenz, Timothy K.; Roy, Poorna; Williams, Loren Dean. Preparation of Long Templates for RNA In Vitro transcription by Recursive PCR. Methods in Molecular Biology, 2012, 941, 19-41."[167] The author of this document contributed to the published work primarily through generation of the 'Notes' section describing alternative experimental approaches and strategies.

The gene used to produce a-rRNA was constructed using an advanced PCR technique known as recursive PCR (R-PCR), which can be used to generate long, double-stranded DNA from appropriate collections of long DNA oligonucleotides (oligos). This method is named for its similarity to recursion in mathematics and computer science applications. R-PCR is also referred to as assembly PCR,[168] parallel overlap assembly,[169] or polymerase cycling assembly.[170] R-PCR has the advantages of (i) one-pot synthesis, (ii) a single dominant product, (iii) speed (often requires less than a day), (iv) material efficiency (chemical synthesis required for only half of the duplex DNA product), (v) modularity (facilitates nesting, modular synthesis, retrosynthesis, and synthesis of targeted libraries), and (vi) robustness (no requirement for ligation).

PCR-based methods for splicing double-stranded DNA by overlap extension,[171] anticipated R-PCR. The first published applications of R-PCR were the synthesis of a 522 bp double-stranded DNA gene from ten oligos,[172] and a 220 bp gene.[173] Stemmer has used R-PCR for the single-step synthesis of DNA sequences up to 3 kb[168] and the approach now enjoys wide application.[169,174,175]

In R-PCR, synthetic DNA oligos, with sequences identical to the sense and anti-sense strands of the target gene, are designed to form a tile. The tile contains alternating double-stranded and single-stranded regions. The 50-70-mer oligos used to construct the tile are complementary on their ends yielding duplex regions interleaved by single-stranded regions. During an R-PCR, each 3' terminus is extended, initially giving a mixture of products that coalesce to a single product as the number of PCR cycles increases. Modifying an R-PCR product can be as simple as swapping one or more oligos for one or more replacements, and repeating the R-PCR. Once a desired RNA-coding gene is obtained, it can be used for *in vitro* transcription without further manipulation.

3.2 - Methods

3.2.1 *in silico* design of a-rRNA

The *T. thermophilus* 23S rRNA was shaved to yield the rRNA fragments indicated by the black lines in Figure 3.2A. The shaving process left 13 rRNA fragments. Most of the fragment termini are within helical regions, allowing facile reconnection with stem-loops to create a continuous RNA chain. For 11 of the connections, the stem-loop is 5'-gccGUAAggc-3', which is a GNRA tetraloop (capital letters) on a three base pair stem (lower case letters). The stem-loops are indicated by blue lines in Figure 3.2A. In a-rRNA-γ, two of the rRNA fragments were connected directly, without a stem-loop.

3.2.2 Synthesis of a-rRNA

The DNA genes encoding a-rRNA versions a-rRNA-β1 and a-rRNA-γ were constructed using R-PCR.[167] The 13 fragments of *T. thermophilus* DNA (sequence NCBI 3169129) that comprise the rRNA component of the a-rRNA gene were connected with 5'-gccGTAAGggc-3' stem-loops to form a single, contiguous DNA gene (Figure 3.3). A stability sequence (5'-GTGG-3'), EcoRI recognition sequence (5'-GAATTC-3'), and T7 promoter (5'-TAATACGACTCACTATAGGG-3') were appended upstream of the a-rRNA gene. A HindIII recognition sequence (5'-AAGCTT-3') and stability sequence (5'-GGTG-3') were appended downstream of the a-rRNA gene. The resulting sequences were conceptually divided into sixteen partially complementary DNA oligos by J.C. Bowman as described,[167] forming a tile. High purity salt-free DNA oligos and a standard amplification primer pair (5'-GTGGGAATTCTAATACGACTC-3' and 5'-CACCAAGCTTGCCCG-3') were obtained from MWG Operon.

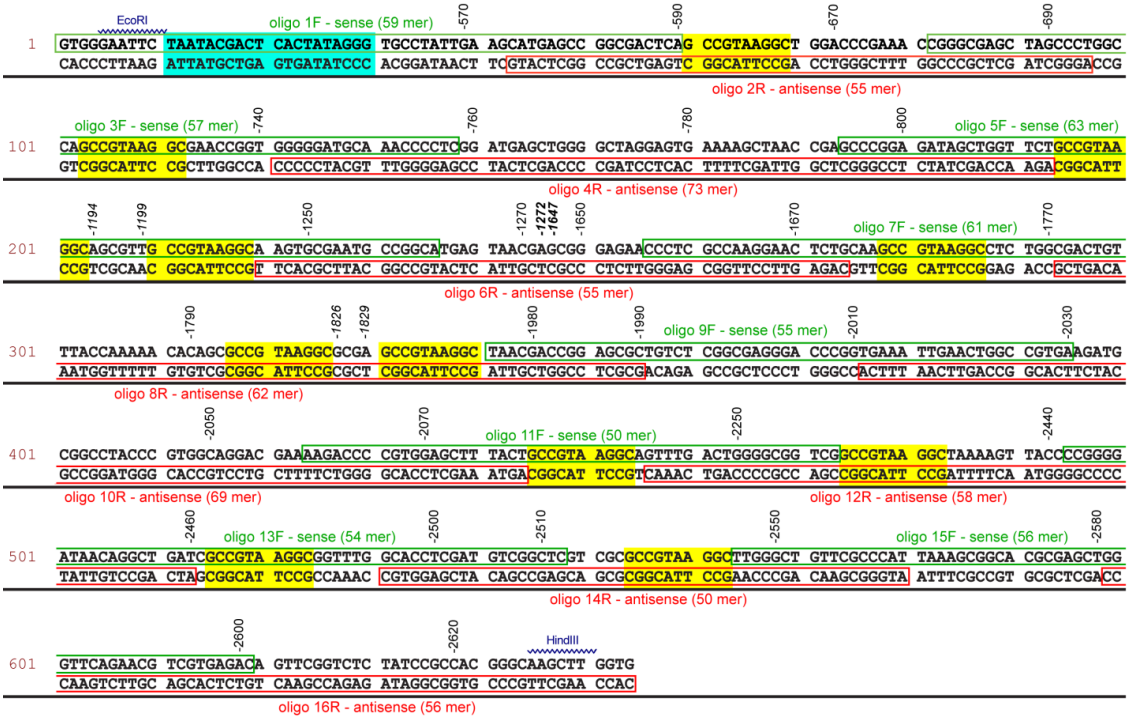


Figure 3.3. Oligonucleotide design for synthesis of the a-rRNA-γ gene by R-PCR. Oligos are outlined in green (sense) and red (antisense) boxes. Restriction endonuclease recognition sequences included to facilitate vector ligation are shown in dark blue font. The T7 promoter is highlighted cyan. 5'-gccGTAAggc-3' stem-loops connecting fragments of the *T. thermophilus* 23S gene are highlighted yellow. Sense strand residues are labeled according to the *E. coli* numbering scheme for 23S rRNA. Residues numbered by inference from *E. coli* numbering are italicized. The sequence within each fragment separated by a stem-loop represents a contiguous *T. thermophilus* 23S rDNA sequence except as shown in bold (A1272-G1647), where 23S rRNA segments were joined directly (without a stem-loop).

Oligos and primers were dissolved in nuclease-free water to working stock solutions of 5 µM (oligo) or 20 µM (primer). An oligo master mix was prepared by combining equal volumes of each oligo working stock solution. Final PCR conditions were 1x Cloned Pfu DNA Polymerase Buffer (Agilent Technologies), 250 µM each dNTP (New England Biolabs), 500 nM of each primer, 50 nM of each DNA oligo, and approximately 3 U Pfu DNA polymerase. In an Eppendorf Mastercycler thermocycler, reaction mixtures were heated to 95 °C for 2 min and cycled 25 times through denaturing

(95 °C for 1 min), annealing (55.2 °C for 1 min), and elongation (72 °C for 1 min). The final step in the reaction was an extended elongation (72 °C for 5 min).

Approximately 1 µg of purified amplification product and pUC19 vector were digested independently and sequentially with EcoRI and HindIII. The 5' phosphate groups were removed from the vector with Antarctic Phosphatase, following the enzyme manufacturer's instructions (New England Biolabs). Enzymes were heat inactivated at 65 °C (20 min), and the product and vector purified by DNA Clean & Concentrator Kit (Zymo Research). Product and vector were ligated with the New England Biolabs Quick Ligation Kit and used to transform competent *E. coli* cells (5 µL ligation mix + 100 µL competent DH5α *E. coli* cells incubated 45 min at 4 °C prior to plating to LB agar containing 50 µg/mL ampicillin). Colonies present after 20 h incubation at 37 °C were screened for inserts by colony PCR with OneTaq Quick-Load Master Mix (New England Biolabs). Colonies positive for insert were used to inoculate 2 mL of LB containing 50 µg/mL ampicillin. After 20 hours of incubation at 37 °C, plasmids were purified by QIAPrep Spin Miniprep Kit (Qiagen) and sequenced bi-directionally by Operon MWG.

pUC19 containing the a-rRNA genes were linearized at the 3' end of the a-rRNA gene with HindIII. The enzyme was heat inactivated and the DNA was purified with the DNA Clean & Concentrator Kit. Approximately 500 ng of DNA per 20 µL reaction volume was transcribed for 2 to 4 h at 37 °C with the MEGAscript High Yield Transcription Kit (Applied Biosystems), followed by incubation with TURBO DNase for 15 min to 1 h at 37 °C. RNA was purified by ammonium acetate precipitation, incubated in 2.5 M ammonium acetate at 4°C for 15 min, centrifuged at 16,000xg at 4 °C for 15 min, and the pellet washed with 80% ethanol before drying via Speedvac or

lyophilization. Pellets were resuspended in nuclease-free water and yields quantified by UV absorbance (NanoDrop 2000, ThermoScientific). Integrity of RNA products from each reaction was assessed by loading 0.75 µg – 1 µg of heat-denatured product in Ambion Gel Loading Buffer II (48% formamide) on a 5% acrylamide, 8 M Urea gel with TBE buffer (89 mM Tris base, 89 mM Boric Acid, 2 mM EDTA) for 1 h at 120 V and stained with ethidium bromide (Fisher).

3.2.3 SHAPE reactions

SHAPE methods were adapted from published protocols.[176] *in vitro*-transcribed a-rRNA was prepared in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) at 100 ng/µL a-rRNA. Thirty-two µL aliquots of the RNA solution were added to 4 µL of 10x folding buffer (500 mM NaHEPES pH 8.0, 2 M NaOAc, varying $MgCl_2$) and incubated at 37 °C for 20 min. A 10x NMIA (Tokyo Chemical Industry Co., Ltd.) solution in dimethyl sulfoxide (DMSO, 2 µL) was added to solutions of 18 µL annealed a-rRNA. Control reactions contained DMSO only. Reactions were carried out for 1 h at 37 °C. For NMIA modification under denaturing conditions, reactions were run at 90 °C for 4 min. NMIA-modified RNA was purified from reaction mixtures with the Zymo RNA Clean + Concentrator-25 kit (Zymo Research). a-rRNA was eluted with 25 µL TE buffer. Recovery after purification was >75%.

3.2.4 Reverse transcription of NMIA-modified a-rRNA

Four different 5'-[6-FAM]-labeled DNA oligonucleotides (Eurofins MWG Operon) were used as primers (each helix in parentheses either contains or succeeds the primer binding site): 5'-TGCCCGTGGCGGATAGAGAC-3' (helix 73), 5'-ACATCGAGGTGCCAAACCGCC-3' (helix 89), 5'-GTTCAATTTCACCGGGTCCCTCG-3' (helix 61), and 5'-CGTTACTCATGCCGGCATTCGC-3' (helix 26). The same reverse transcription primers were used for all iterations of a-rRNA. Modified RNA (20 µL) was added to 8 pmol of each primer in 10 µL of TE buffer. To anneal primers, samples were heated to 95 °C for 1 min, held at 65 °C for 3 min, and then placed on ice. SuperScript III Reverse Transcriptase (Invitrogen) was used in RT reactions. RT buffer (19 µL) was added at 30 °C to yield 50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 2 mM DTT, and 250 µM of each dNTP (final concentrations in 50 µL). RT mixtures were heated to 55 °C for 1 min before addition of 1 µL Superscript III Reverse Transcriptase enzyme mix (200 U). Reactions were incubated at 55 °C for 2 h and terminated by heating to 70 °C for 15 min. Sequencing control reaction of unmodified *in vitro*-transcribed RNA were prepared in TE buffer at 31 ng/µL. Aliquots of RNA solution (20 µL) were annealed to the DNA primers described above. RNA was sequenced by reverse transcription/chain termination using all four dideoxynucleotide triphosphates (ddNTP's), at a ratio of 8:1 ddNTP to dNTP (deoxynucleotide triphosphate), and a control reaction without ddNTPs was also prepared.

3.2.5 Capillary electrophoresis of RT reaction products

RT reaction mixture (1 µL) was mixed with 0.3 µL of ROX-labeled DNA sizing ladder (for alignment of disparate traces) and 8.7 µL Hi-Di Formamide (Applied Biosystems) in a 96-well plate. Plates were heated to 95 °C for 5 min and RT products were resolved by capillary electrophoresis (CE) using a 3130 Genetic Analyzer (Applied Biosystems) at 65 °C with a custom fluorescence spectral calibration. The capillary array was loaded with Performance Optimized Polymer-4 (Applied Biosystems).

3.2.6 SHAPE data processing

SHAPE data were processed as described previously.[118] Several nucleotides were excluded from SHAPE analysis of a-rRNA-γ because the reverse transcriptase gave high background termination in the (-) NMIA reactions. These nucleotides were determined statistically (termination value ≥90% of the (+) NMIA value, where both values exhibit ≥0.25 normalized intensity).

3.2.7 Thermal denaturation studies of a-rRNA iterations

Extinction coefficients were calculated using sequence (nearest neighbor method) with the online Oligonucleotide MW and Extinction Coefficient Calculator (Ambion). Estimated extinction coefficients were 2.8×10^7 $cm^{-1}M^{-1}$ for *T. thermophilus* 23S rRNA, and 6.1×10^6 $cm^{-1}M^{-1}$ for a-rRNA versions. Preliminary thermal denaturation with of a-rRNA-β2 used ~0.5 OD of *in vitro*-transcribed RNA in 650 µL samples, measured in quartz cuvettes. No buffer was used, and readings were taken with or without 200 µM

$MgCl_2$. RNA was annealed by heating in a 1.4 L 90 °C water bath for 90 s. Water bath and samples were then removed from heat and stirred until water bath temperature reached 30 °C. Samples were heated from 5 °C to 95 °C in a Peltier-controlled UV-Vis spectrophotometer (Agilent 8453), followed by cooling to 5 °C. This heating/cooling cycle was repeated, and spectra were collected every degree. Data were processed using IgorPro. Baseline correction was applied using the region between 370-400 nm.

Subsequent rounds of thermal denaturation studies were performed in 1 mm cuvettes with 350 μL samples containing ~0.25 OD of RNA (~400 nM for a-rRNA versions, ~88 nM for the *T. thermophilus* 23S rRNA). Folding conditions were 200 mM NaOAc, 50 mM NaHEPES pH 8, with either 10 mM $MgCl_2$ or no $MgCl_2$. To anneal RNA, samples were heated to 90 °C in a heating block for 90 s, then removed and allowed to cool to room temperature. UV measurements were obtained in a Peltier-controlled Cary 50 UV-Vis spectrometer (Varian) with a 6-cell changer. The temperature cycle used was as follows: 0 °C→95 °C→0 °C→95 °C→0 °C.

In thermal denaturation studies with a-rPeptide L4, 400 nM peptide was added before folding (1:1 ratio with a-rRNA). A mineral oil overlay was used in this case to prevent evaporation. Certain samples contained 50 mM NaHEPES, pH 8 with no other monovalent present (as indicated).

Absorbance data at 260 nm was fit to a two-state model using OligoMelt (developed by C. Bernier). Melting temperatures were extracted from these fits.

3.3 – Results

3.3.1 SHAPE experiments on early iterations of a-rRNA

a-rRNA-β1 was the first physical version of a-rRNA, followed by a-rRNA-β2. See chapter 2 for a discussion on the reasoning and methodology behind this iteration. SHAPE was performed on both of these early versions to test for induction of the model secondary structure (Figure 3.2). SHAPE data described in this chapter was processed by J.G. Gossett. Figure 3.4 depicts SHAPE data compared for a-rRNA-β1 and a-rRNA-β2. These data sets are not directly comparable, as they were modified in solutions containing different concentrations of NMIA due to ongoing optimization of our SHAPE protocols; a-rRNA-β1 was modified with 13 mM NMIA, while a-rRNA-β2 was modified with 3 mM NMIA. However, effects of the mutations made between these two a-rRNA versions are apparent (Figure 3.4, cyan box). In both versions of a-rRNA, the nucleotides of the loop exhibit moderate reactivity (0.5-1 SHAPE units), which is expected since they are designed to be unpaired. However, in a-rRNA-β1, the neighboring nucleotides of the tetraloop stem are somewhat SHAPE reactive, suggesting that the stem is not fully-formed as designed. In a-rRNA-β2, the nucleotides of the stem exhibit near-zero SHAPE reactivity, indicating stable formation of the tetraloop stem as designed (Figure 3.4 and 3.5).

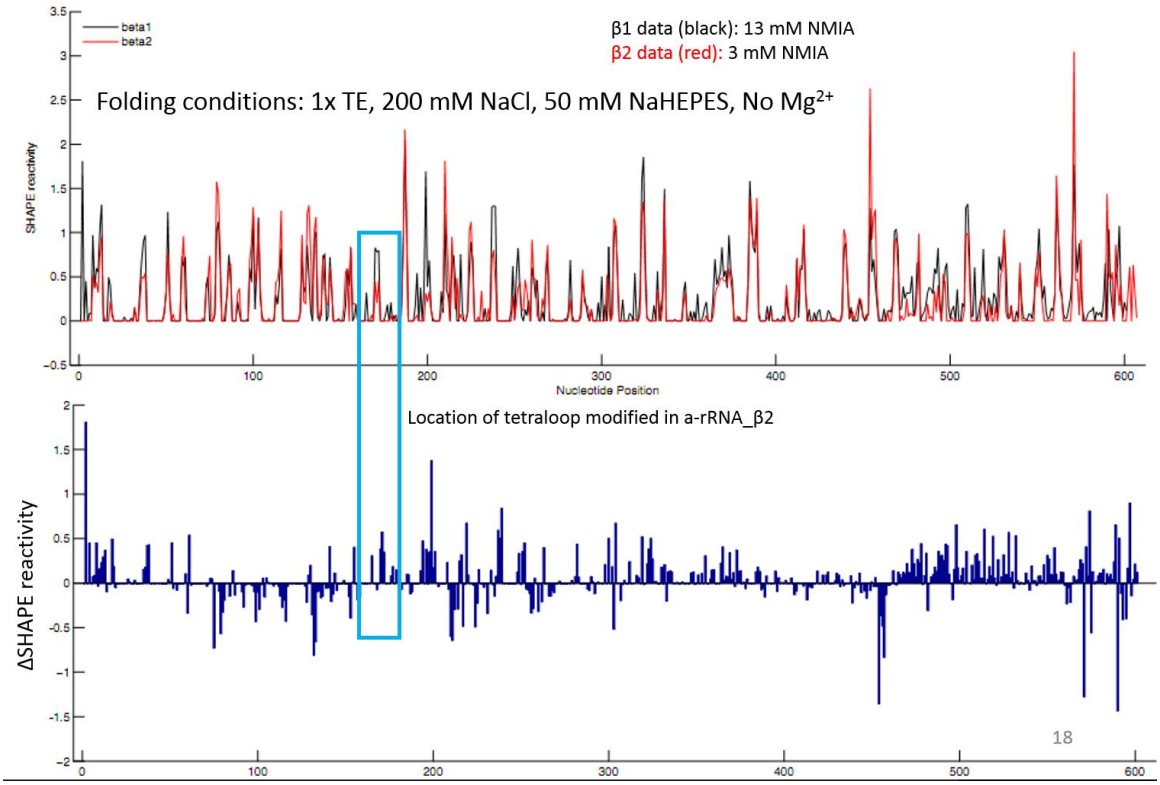


Figure 3.4. Processed SHAPE data for early iterations of a-rRNA. Top panel: SHAPE reactivities for β1 (black) and β2 (red) versions of a-rRNA. RNA was folded and modified in 1x TE, 200 mM NaCl, 50 mM NaHEPES pH 8. Bottom panel: Comparison of reactivities in a-rRNA-β1 vs. a-rRNA-β2. Cyan box indicates the tetraloop to which changes were made between the β1 and β2 versions (see section 2.3.2).

SHAPE data for a-rRNA-β2 exhibit excellent agreement with the model secondary structure. Figure 3.5 depicts SHAPE data obtained for a-rRNA-β2 in 250 mM Na^+ (no Mg^{2+}) mapped onto the predicted secondary structure of a-rRNA-β2. Overall, low reactivity is observed in regions of a-rRNA-β2 expected to be helical, and virtually all reactive sites are restricted to loops, inter-helical regions, or otherwise unpaired nucleotides. These results suggest that a-rRNA-β2 forms the predicted secondary structure in presence of moderate [Na^+].

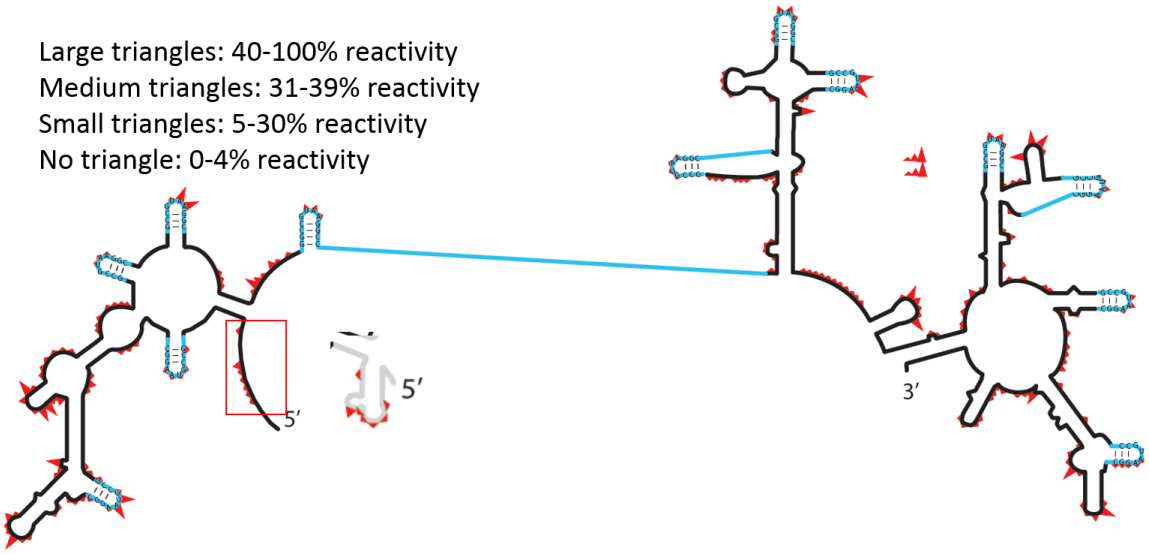


Figure 3.5. SHAPE reactivities of a-rRNA-β2 in 250 mM Na^+ mapped onto the expected secondary structure. Red triangles indicate extent of SHAPE reactivity at each position, normalized to the most reactive nucleotide. Black lines indicate 23S-derived rRNA, while cyan lines indicate non-native gccGUAAggc tetraloops used to stitch together a-rRNA. The red box highlights the 5' region. The grey lines illustrates a version of the secondary structure for the 5' region that is more consistent with the 3D structure. SHAPE reactions were performed in 50 mM NaHEPES pH 8.0, 200 mM NaOAc, 3 mM NMIA.

SHAPE data for the 5' end of a-rRNA hint at flaws in the canonical secondary structure of the 23S rRNA. The region of a-rRNA-β2 indicated in Figure 3.5 (red box) exhibits near-zero SHAPE reactivity at many nucleotides that appear to be unpaired in the predicted secondary structure, which is based on the canonical 23S secondary structure.[177] Upon inspection of the 3D structure of the *T. thermophilus* ribosome, several base pairs are observed between positions analogous to the unreactive nucleotides of the 5' region of a-rRNA. A more accurate, 3D-based secondary structure for this region is shown in grey in Figure 3.5. SHAPE data agree much more closely with this updated secondary structure model.

Molecular dynamics simulations performed by T. Preeprem suggested that tetraloop 5 (fifth tetraloop from the 5' end of a-rRNA-β2) would remain unstable despite the mutations introduced to stabilize its formation, although the stem nucleotides of this

tetraloop yield low SHAPE reactivities indicating they are base-paired or otherwise constrained (Figures 3.4 and 3.5). Based on the modeling data, a new version of a-rRNA was designed in which the 23S-derived rRNA on either side of tetraloop 5 is sealed directly, rather than with a gccGUAAggc tetraloop. In the spirit of the iterative naming convention already established for a-rRNA, this new version was termed 'a-rRNA-γ'.

3.3.2 Design and generation of a-rRNA-γ

The completed a-rRNA-γ (Figure 3.2) is composed of 505 nucleotides derived from the 23S rRNA plus 110 nucleotides from 11 stem-loops. The complete a-PTC contains a-rRNA-γ (615 nucleotides), Mg^{2+} ions, and 5 a-rPeptides. a-rRNA-γ was generated by *in vitro* transcription. The RNA product of the transcription reaction runs as a tight band on a denaturing gel (Figure 3.6).

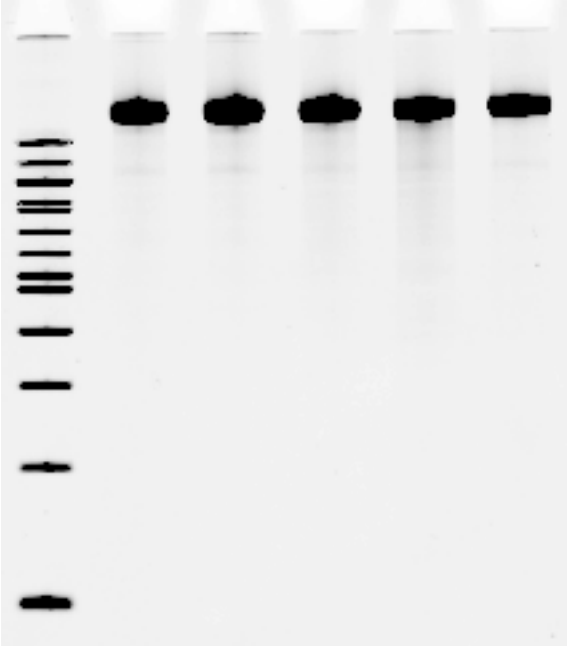


Figure 3.6. a-rRNA-γ *in vitro* transcription reaction replicates and 100 bp dsDNA ladder. The integrity of each a-rRNA transcription reaction (lane 2 – lane 6) was assessed by running 0.75 µg – 1 µg of product denatured in Ambion Gel Loading Buffer II (48% formamide) on a 5% acrylamide, 8 M Urea gel with 1x TBE buffer (89 mM Tris base, 89 mM Boric Acid, 2 mM EDTA) for ~1 h at 120 V and staining with ethidium bromide (Fisher). The 100 bp dsDNA ladder (lane 1) serves as an RNA mobility and gel quality standard among gels. Gel electrophoresis performed by E.B. O'Neill.

In the 3D model, the components that are common between the a-PTC and the *T. thermophilus* LSU crystal structure have conserved conformation and interactions. The stem-loops are constrained to canonical GNRA tetraloop conformation[178] and are located on the surface of the a-PTC. The stem-loops do not engage in unfavorable steric contacts with other parts of the a-PTC. The model is stereochemically reasonable.

The predicted secondary structure of the a-PTC shown in Figure 3.2 is based on the canonical secondary structure of the 23S rRNA. There are some small differences between the canonical secondary structure and the actual secondary structure found in the crystal structure.

3.3.3 SHAPE characterization of a-rRNA-γ secondary structure

The SHAPE reactivity of a-rRNA-γ was determined in the presence and absence of Mg^{2+}. SHAPE utilizes an electrophile, in this case NMIA (13 mM in modification reactions), that reacts with 2'-OH groups of RNA.[55] SHAPE reactivity is modulated by RNA flexibility. Base-paired nucleotides exhibit low reactivity since their flexibility is constrained. Single-stranded nucleotides are flexible and reactive to the electrophile. SHAPE has been shown to be accurate for mapping secondary structure of many RNAs, including *E. coli* rRNA.[129] SHAPE data were processed by J.G. Gossett as previously described,[50,118] and graphs of processed SHAPE data are provided in Figure 3.7. Full numerical SHAPE reactivities are provided in a multimedia file (Dataset 1).

The SHAPE data show excellent correspondence with the predicted secondary structure of a-rRNA-γ (Figure 3.8A). These data were obtained by probing a-rRNA-γ in the presence of monovalent cations (250 mM Na^{+}) and the absence of Mg^{2+}. These conditions are known to support the formation of RNA secondary structure, but not tertiary structure.[14,15,179] SHAPE data were obtained for 459 of the 505 nucleotides (91%) of a-rRNA-γ derived from the 23S (stem-loop nucleotides are analyzed separately below). The values and statistics provided in the remainder of this section refer only to these nucleotides derived from the 23S rRNA. Data were not obtainable for RNA near the termini, or for a few nucleotides with high background (i.e., high reverse transcriptase termination in samples not treated with NMIA). These regions are shown in gray on the a-rRNA secondary structure in Figure 3.8A.

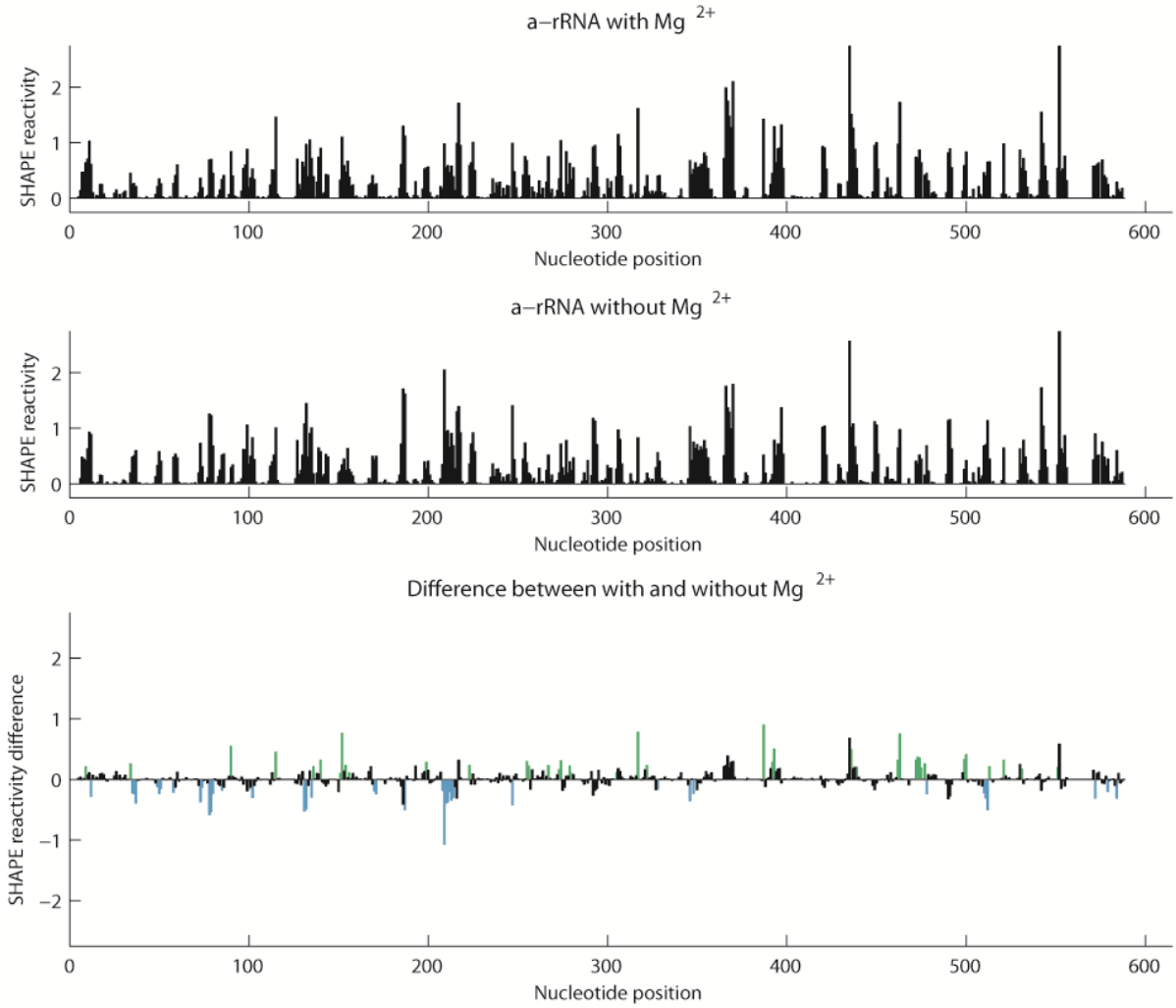


Figure 3.7. SHAPE reactivity of a-rRNA-γ in the presence (A) and absence (B) of 10 mM Mg^{2+}. (C) Changes in SHAPE reactivity (nucleotide flexibility) of a-rRNA-γ upon addition of 10 mM Mg^{2+}. Green indicates nucleotides at which the SHAPE reactivity increases by at least 40%, and blue, decreases of at least 40%, where at least one SHAPE value (with or without 10 mM Mg^{2+}) is greater than 0.3. Colored positions are the same as in Figure 3.8.

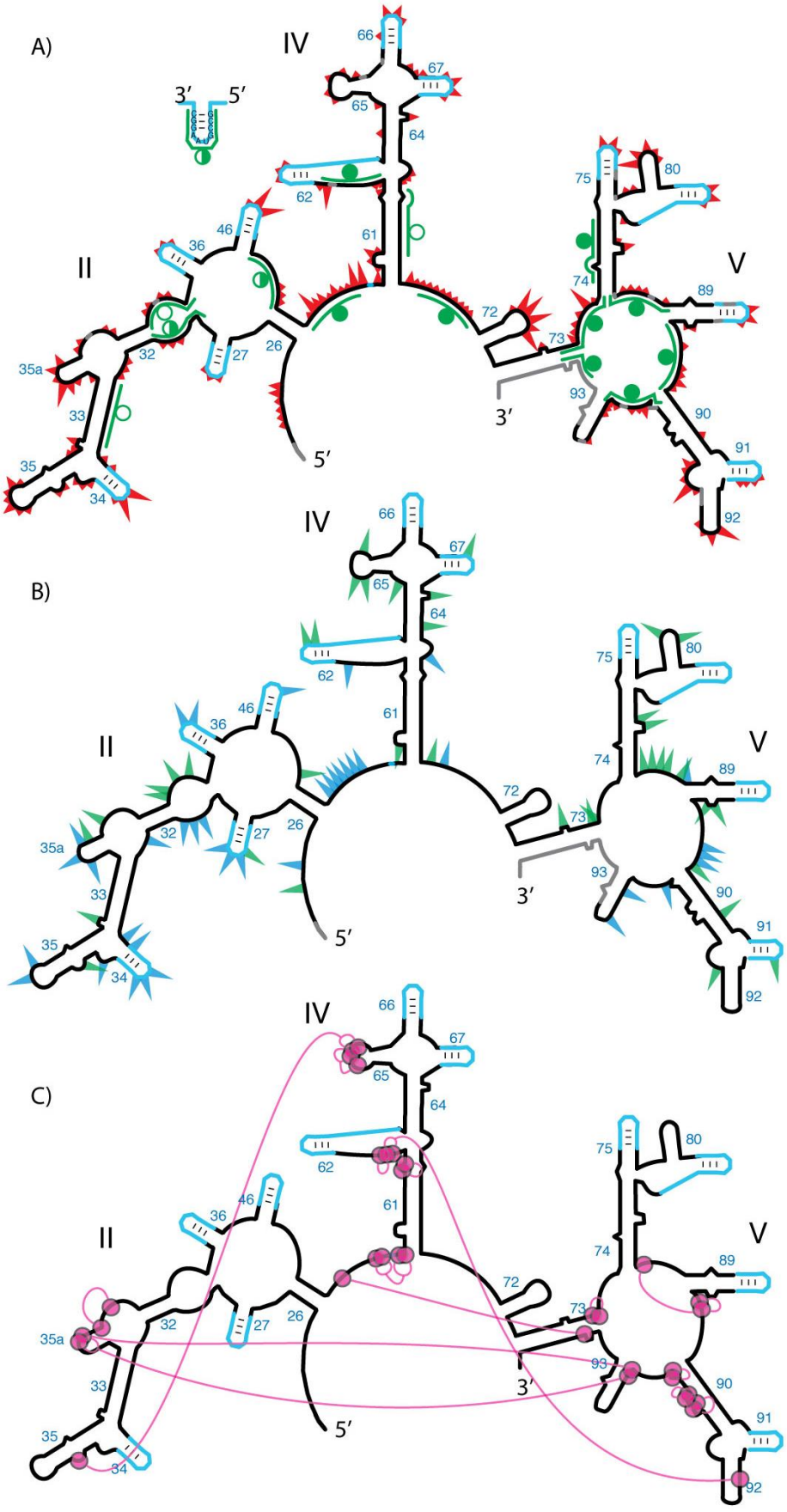

A)
IV
II
V
3′
5′
66
67
65
64
62
61
46
36
32
27
26
35a
33
35
34
72
73
75
80
74
89
93
90
91
92
B)
C)

Figure 3.8. Probing the secondary and tertiary structure of a-rRNA-γ. A) SHAPE and RNase H mapping. Red triangles mark SHAPE reactivities in 250 mM Na^+, mapped onto the predicted secondary structure of a-rRNA-γ. Larger triangles indicate greater SHAPE reactivity. RNase H DNA probes are indicated by green lines. Circles indicate extent of RNA digestion by RNase H: filled circles (more than 75%), half-filled circles (between 25% and 75%), and empty circles (less than 25%). RNase H data collected by L. Lie. B) Effects of 10 mM Mg^{2+} on SHAPE reactivity suggest formation of tertiary structure. Green triangles show the greatest increases in SHAPE reactivity upon addition of Mg^{2+}. Blue triangles show the greatest decreases in reactivity. C) Multidentate Mg^{2+}-phosphate interactions observed in the *T. thermophilus* LSU (PDB 2J01) are mapped onto the predicted secondary structure of a-rRNA. Magenta circles indicate first-shell Mg^{2+}-PO interactions. Magenta lines indicate PO-Mg^{2+}-PO linkages. Gray shading in panels A and B indicates rRNA where SHAPE data were not obtainable. SHAPE reactions were performed in 50 mM NaHEPES pH 8.0, 200 mM NaOAc, 0 or 10 mM $MgCl_2$.

Nucleotides were binned into four groups according to their normalized SHAPE reactivity. U2574 (NCBI *E. coli* numbering) is the most reactive nucleotide, with an absolute reactivity of 3.24 (arbitrary units), and is considered 100% reactive. Bin assignments are indicated by the size of red triangles in Figure 3.8A. The largest triangles denote the 13 nucleotides with normalized SHAPE reactivities ranging from 38-100%. Triangles of intermediate size indicate the 24 nucleotides with reactivities from 25-37%. The smallest triangles indicate the 109 nucleotides with reactivities from 8-24%. Nucleotides with reactivities of ≤ 7% are considered to be unreactive.

Of the 146 nucleotides that are reactive to the SHAPE reagent, 133 (91%) fall in regions expected to be single-stranded, in that they do not form base pairs in the predicted secondary structure. These nucleotides are in loops, bulges, or other single-stranded regions. The resolution of SHAPE is sufficiently high to allow identification of small bulges such as those in Helices 35, 61, 64, 73, 74, and 89 (Figure 3.8A). Thirteen reactive nucleotides are base-paired in the predicted secondary structure. These nucleotides represent less than 6% of the 232 nucleotides that are expected to be double-stranded, the remaining 94% of which exhibit low SHAPE reactivity as predicted. The 133 nucleotides

that SHAPE identifies as flexible compose 59% of the 227 single-stranded nucleotides in the predicted secondary structure.

SHAPE data were obtained for 106 (96%) of the 110 nucleotides within the 11 stem-loops of a-rRNA-γ. The stem regions are mostly unreactive. Only 5 of the observable 62 nucleotides in the stem regions (8%) exhibit reactivity, and all fall into the 'low reactivity' bin. The loop regions yield a consistent pattern of reactivity that is independent of the surrounding sequence. In 10 of the 11 stem-loops, the UAA nucleotides are more reactive than the preceding G. This pattern of reactivity is consistent with known patterns of flexibility of GNRA tetraloops,[150,178] and will be expounded upon in greater detail in chapter 6. Nucleotides NRA of a GNRA tetraloop are more polymorphic and flexible than the G. The N nucleotide is the most polymorphic of all.

The SHAPE results suggest that the stem-loop replacing helix 91 (Figure 3.8A) might not form the predicted secondary structure. The stem-loop at this position exhibits a distinctive pattern of reactivity. Two nucleotides in the stem appear to be anomalously flexible. This SHAPE data could be used to inform design changes in future iterations of a-rRNA.

3.3.4 SHAPE characterization of a-rRNA-γ tertiary interactions

RNA requires Mg^{2+} ions to form compact structures.[14,15,179] The core of the assembled LSU is particularly rich in Mg^{2+} ions.[23,35,43] In our 3D model of the a-PTC, the conformation of the a-rRNA is stabilized in part by Mg^{2+} ions (Figure 3.8C) and also by tertiary interactions between RNA elements that are remote in the primary sequence.

Here we ask if a-rRNA-γ *in vitro*, upon the addition of Mg^{2+}, forms a compact structure consistent with the *in silico* a-PTC model (Figure 3.2B).

A network of phosphate-Mg^{2+}-phosphate interactions is anticipated in the a-PTC, based on the network in the LSU crystal structure (PDB 2J01).[43] This phosphate-Mg^{2+}-phosphate network is indicated in Figure 3.8C, where magenta circles represent nucleotides with phosphate oxygens that make direct contacts with Mg^{2+} ions (<2.4 Å cutoff distance). Long-range phosphate-Mg^{2+}-phosphate linkages are indicated by magenta lines. The proposed network includes all Mg^{2+} ions that interact with phosphates of two or more of the nucleotides. Formation of the phosphate-Mg^{2+}-phosphate network is coupled with formation of base-base tertiary interactions. Therefore Mg^{2+} is expected to influence the rRNA flexibility and SHAPE reactivity of nucleotides that contact Mg^{2+} or are involved in long-range tertiary interactions. This pattern of Mg^{2+}-dependent SHAPE reactivity has previously been observed for tRNA, RNase P, the P4-P6 domain of the *Tetrahymena* Group I intron and Domain III of the 23S rRNA.[50,55,116,118]

The influence of Mg^{2+} on a-rRNA-γ SHAPE reactivity (Figures 3.7 and 3.8B), suggests that Mg^{2+} does indeed induce global folding. Changes in SHAPE reactivity upon addition of Mg^{2+} (10 mM) were determined. The greatest changes in SHAPE reactivity (>40% change) are mapped onto the predicted secondary structure of a-rRNA-γ (increases: green; decreases: blue, Figure 3.8B). Some Mg^{2+}-induced changes in SHAPE reactivity of a-rRNA-γ are clustered around hypothesized regions of direct RNA-Mg^{2+} contact similar to those in the folded LSU. Others are observed in regions of hypothesized long-range tertiary interactions. Nucleotides with large changes in reactivity are dispersed throughout the sequence (Figure 3.7) and the secondary structure.

Nucleotides for which the absolute SHAPE reactivity is small (≤ 0.3) were omitted from the analysis to avoid attributing artificial significance to small changes in absolute value.

Of the 65 rRNA nucleotides (excluding stem-loops) with altered SHAPE reactivities upon addition of Mg^{2+} (Figure 3.8B), 25 (38%) are within 3 residues of a nucleotide that contacts a Mg^{2+} ion in the predicted interaction network depicted in Figure 3C. Most of the other Mg^{2+} effects are at or near bulges and loops, assumed to be involved in tertiary interactions. The data are consistent with induction by Mg^{2+} of local and long distance interactions that alter nucleotide flexibility. The results suggest a transition of the a-rRNA-γ from an extended secondary structure to a collapsed structure with tertiary interactions.

The SHAPE reactivities argue against significant alteration of the secondary structure upon addition of Mg^{2+}. Only a small number of helical nucleotides show Mg^{2+}-dependent changes. Of the 65 rRNA nucleotides that show significant Mg^{2+} effects, only a few (9 nucleotides) are found in regions that are paired in the predicted secondary structure, and several of these are adjacent to predicted bulges or otherwise single-stranded regions.

3.3.5 Thermal denaturation studies of a-rRNA

The thermal stability of a-rRNA iterations, as well as full-length *T. thermophilus* 23S rRNA, was studied in order to compare stabilities of successive iterations to that of the parental LSU rRNA. In general, transitions were very broad, which is to be expected considering the size and complex structure of the studied rRNAs.

Melting temperatures of successive iterations of a-rRNA exhibit increasing thermal stability. Thermal denaturation data were collected after annealing in 50 mM NaHEPES, pH 8. When fit to a two-state model, the estimated T_m of a-rRNA-γ is 60.2 °C, while the estimated T_m of both a-rRNA-β1 and a-rRNA-β1 is 58.8 °C. This suggests a slight increase in overall thermal stability of a-rRNA-γ. By comparison, the estimated T_m of intact 23S rRNA was estimated at >85 °C.

a-rRNA-γ exhibits higher thermal stability in presence of a-rPeptide uL4. Based on the *in silico* model of the a-PTC (Figure 3.2B), a-rPeptide uL4 should bind to a-rRNA-γ. The sequence of a-rPeptide uL4 is: KTRGEVAYSGRKIWPQKHTGRARHGDI GAPIFVGGGVVFGP (41 amino acids). a-rRNA-γ was thermally denatured in the presence and absence of a-rPeptide L4 (50 mM NaHEPES, pH 8), and monitored by absorbance at 260 nm. When fit to a two-state model, a-rRNA-γ exhibits a reversible transition with a T_m of 54.4 °C in the absence of a-rPeptide uL4, and 57.3 °C in its presence, suggesting modest stabilization. This result is consistent with continuous variation experiments performed by C. Hsiao and J.K. Peters that exhibit specific 1:1 binding between a-rRNA-γ and a-rPeptide uL4.[157]

3.4 – Discussion

Translation and the ribosome are some of the most illuminating biochemical links between current and ancient biological systems.[5] Translation provides a powerful experimental and theoretical system for exploring life's oldest biological processes and molecules.[104,105,158-161] Some parts of the ribosome are thought to predate our current biology of DNA, RNA, and coded protein. Following previous work,[10,110,180-182] we have

undertaken an effort to exploit the ribosome as a tool for the study of ancient biology. Ancient components of the LSU, as inferred from consensus among models of early ribosomal evolution, were constructed and characterized. The conclusion here is that the core of the LSU has retained an ancient ability to fold and assemble over vast evolutionary time frames of billions of years. Relatively small ancestral components of the LSU are folding-competent and assembly-competent.

Advances in technology allow us to address fundamental questions of paleobiochemistry, a discipline first conceived by Zuckerkandl and Pauling.[183] Here we have constructed an *in vitro* model of an ancestral Peptidyl Transferase Center. The a-PTC is designed from a consensus among previous proposals of ribosomal evolution by the laboratories of Fox,[111] Steinberg,[109] Noller,[98] Williams,[35] and others. The a-PTC is an experimental model of one of life's most ancient assemblies, composed of fragments of rRNA and protein from the oldest part of the extant ribosome (Figure 3.1). These rRNA fragments are joined together here by stem-loops to form a single RNA polymer called the a-rRNA (Figure 3.2A).

3.4.1 SHAPE results for a-rRNA-γ support formation of predicted structure

a-rRNA-γ folds as expected *in vitro*. This conclusion is based in part on i) SHAPE reactivity, which supports formation of the predicted secondary structure of a-rRNA-γ, ii) the Mg^{2+}-dependence of the SHAPE profile, which indicate LSU-like Mg^{2+} interactions within a-rRNA and Mg^{2+}-induced formation of tertiary interactions, and (iii) specific binding of a-rPeptide uL4, supported by thermal denaturation studies. Data also confirm the folding and increasing thermal stability of earlier versions of a-rRNA.

Yonath and coworkers previously described a model for the proto-ribosome[110,181] that is significantly smaller than our model a-rRNA. In fact, Yonath's proto-ribosome is fully contained within our a-rRNA. In contrast to the a-PTC, there are no peptide components in the proto-ribosome. Yonath's proto-ribosome appears to represent a model of a smaller, even more primitive ancestor of the a-PTC.

3.4.2 The cooption model of early ribosomal evolution

The fundamental hypothesis that underlies our approach is called the "*cooption model*".[98,161] The cooption model suggests the following general order of events in ribosomal evolution: (i) ancestors of the SSU and LSU originated and evolved independently of each other, with autonomous functionalities, (ii) an ancestor of the LSU, independent of the SSU, contained the PTC, which catalyzed production of short, non-coded peptide or combined peptide/ester[184] oligomers, (iii) some of the non-coded oligomer products of the LSU ancestor associated with the LSU, conferring advantage, (iv) an ancestor of the SSU had a function that is more tentative but, as proposed by Noller, may have involved RNA binding or polymerization, (v) ancestral LSUs and SSUs joined, in a cooption process, enabling coded synthesis, and (vi) the peptide/ester oligomers associated with the ancestral LSU eventually became the tails of ribosomal proteins that penetrate deep within the extant LSU.

In the cooption model, and other models of ribosomal evolution, changes over evolution are restricted to those that maintain the structure and functionality of the PTC and the decoding center. The catalytic core of the LSU, and the decoding center of the SSU, are assumed to predate the cooperative relationship between the LSU and SSU.

This cooption model predicts that an ancestor of the 23S rRNA, lacking more recent components required for association with the SSU, will retain folding, recognition and assembly capabilities. Here we test that aspect of the cooption model by determining if the a-rRNA, a model of a 23S rRNA ancestor, can fold and assemble to form the a-PTC.

3.5 – Conclusions

The results here support elements of the cooption model of ribosomal evolution. The proposed ancestral elements of the LSU are highly robust in folding and assembly. We have shaved around 2400 nucleotides from the 23S rRNA. The remaining a-rRNA retains the ability to fold and assemble with Mg^{2+} and certain a-rPeptides. This robustness is consistent with the premise that the ribosome is an ancient assembly that evolved in a primitive biological environment and has survived billions of years of evolution, growing in size and complexity, without major changes in core structure or function. Although the modern LSU has the appearance of a massive monolithic assembly,[11] our results here, and elsewhere,[118] indicate that it can be understood as much smaller elements of RNA and peptide that retain ancient abilities to fold and independently assemble.

CHAPTER 4

THE LSU RIBOSOMAL RNA FOLDS TO A NEAR-NATIVE STATE IN THE PRESENCE OF MAGNESIUM

4.1 – Introduction

4.1.1 rRNA-Mg^{2+} interactions

Ribosomes are stabilized by intricate networks of RNA-RNA, RNA-protein, and RNA-cation interactions.[43,75] Mg^{2+} is crucial for folding and peptidyl transferase activity of the LSU.[23,74,79,80]

Cations are known to facilitate the folding of large RNAs into compact native structures by accumulating in and around the RNA and allowing negatively-charged phosphates to achieve close proximity.[14,15] Monovalent cations facilitate formation of RNA secondary structure, and Mg^{2+} promotes native 3D structure, including long-range RNA-RNA tertiary interactions. Mg^{2+} can increase local RNA density and rigidity by tightly packing RNA functional groups in its first coordination shell. Mg^{2+} is particularly potent in mediating RNA-RNA interactions due to its high charge density, which can bring distant RNA elements into close proximity with well-defined geometry. In RNA-Mg^{2+} interactions, the most common inner shell ligands of Mg^{2+} are PO atoms, while nucleobase heteroatoms are common outer shell ligands.[22] The same coordination trends are observed in LSU rRNA.[23,44]

Here we investigate structural collapse of LSU rRNA (Figure 4.1) from an extended secondary structure in the presence of Na^+ into a quasi-folded state upon addition of Mg^{2+}. We characterize the folding of rRNA in the absence of ribosomal proteins (rProteins) and other native interaction partners, assaying for specific local and long-range interactions. The objective is to determine the degree to which rRNA sequence alone determines its native structure, and infer the roles of binding partners such as Mg^{2+} and rProteins in induction of fully native ribosomal structure.

4.1.2 Updated rRNA secondary structures

The LSU rRNA secondary structure depicted in Figure 4.1 is updated compared to those used predominantly in chapters 2-3. This updated secondary structure is based on 3D structural data, which was unavailable when the original LSU secondary structure was produced by covariation analysis.[177] The revised LSU secondary structure includes a central core domain, Domain 0 (D0), from which all other secondary domains radiate.[185] Two new non-canonical helices are introduced in D0: Helices 25a and 26a. While H25a and H26a contain many non-canonical base pairs, their 3D structures are clearly helical. Inspection of the LSU 3D structure, which was motivated by SHAPE data for a-rRNA, revealed early evidence for H25a (see section 3.3.1). This revised secondary structure is used consistently in this document from this point forward.

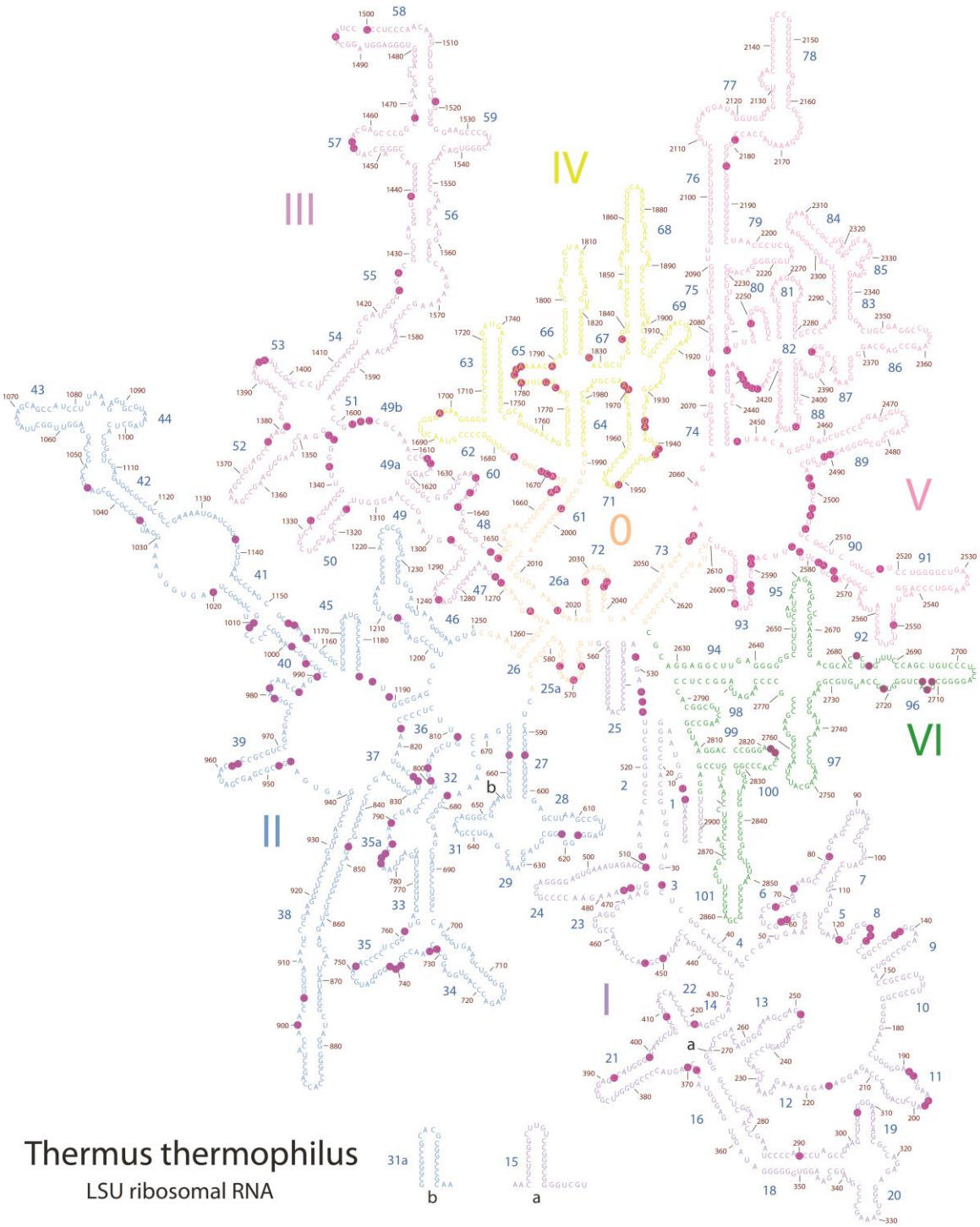


Figure 4.1. 3D-based secondary structure of LSU rRNA from *T. thermophilus*. Nucleotides are colored by domain. Nucleotide numbers are indicated every ~10 nt (*E. coli* numbering). Helix numbers (blue), domain numbers, and insertion points are shown. Underlaid magenta circles denote nucleotides observed to interact directly with Mg^{2+} atoms (first-shell interactions, 2.4 Å cut-off) in the crystal structure of the *T. thermophilus* ribosome (PDB IDs: 2J00 and 2J01). Figure generated with RiboVision.

4.1.3 The rRNA-Mg^{2+} state

We propose a model of Mg^{2+}-induced folding of the LSU rRNA to an 'rRNA-Mg^{2+}' state, comprised of many but not all features of the native LSU state as observed in three-dimensional structures of assembled ribosomes. In the rRNA-Mg^{2+} state all base pairs of the canonical secondary structure are intact. The rRNA-Mg^{2+} state includes tertiary interactions dispersed throughout the rRNA, and inter-domain interactions involving all domains of the LSU rRNA. Some long range RNA-RNA interactions are not supported in the rRNA-Mg^{2+} state, highlighting the importance of proteins in ribosomal function. The rRNA-Mg^{2+} state represents the core architecture of the LSU which, while not catalytically active,[186] positions the residues of the LSU rRNA in such a way as to promote native interactions with rProteins to ultimately form a functional LSU.

4.1.4 SHAPE and rRNA

We assayed rRNA structure by comparing footprinting data obtained in the presence and absence of Mg^{2+}. SHAPE provides quantitative, nucleotide-resolution information about local nucleotide conformation and base pairing.[115] SHAPE is sometimes used to monitor RNA structural transitions induced by cations, small molecules, or proteins.[115,138,139] Intramolecular (tertiary) and intermolecular (crystal packing) interactions have been detected using SHAPE in the well-studied RNA P4-P6.[121] SHAPE has also been performed on deproteinized *E. coli*, *C. difficile*, and *H. volcanii* rRNA,[137] but data were not intended to infer structural transitions. Comparison of SHAPE data collected under varied conditions, which we refer to as 'comparative

SHAPE', monitors induced folding from secondary structure to native or near-native states via detection of tertiary interactions. We have previously utilized comparative SHAPE to monitor Mg^{2+}-induced structural transitions in two isolated subsets of LSU rRNA, including formation of intra- and inter-domain tertiary interactions.[118,157]

4.2 – Methods

4.2.1 RNA *in vitro* transcription/purification

T. thermophilus LSU rRNA was *in vitro*-transcribed using previously described constructs and protocols.[118] RNA was purified stringently to limit cation contamination, according to previously published protocols.[156] SHAPE was performed here with and without Mg^{2+} on the ~2,900 nt *T. thermophilus* 23S rRNA (LSU rRNA, Figure 4.1). Cation conditions were 250 mM Na^+ (abbreviated Na^+) or 250 mM Na^+/10 mM Mg^{2+} (abbreviated Na^+/Mg^{2+}). LSU rRNA lacked rProteins and all other components of the assembled ribosome, since it was generated by *in vitro* transcription.

4.2.2 NMIA modification of *T. thermophilus* LSU rRNA

The SHAPE method exploits the variable susceptibly of the 2'-hydroxyl group to attack by electrophiles such as NMIA. Reactivity to NMIA is modulated by the interatomic distance between the 2'-OH and PO, and is often regarded as commensurate with local RNA flexibility. Generally, nucleotides that are constrained by canonical cis Watson-Crick/Watson-Crick (cWW) base pairing exhibit low reactivity toward SHAPE reagents, while nucleotides that participate in non-canonical or unusual RNA-RNA

interactions exhibit variable reactivity, and flexible unpaired nucleotides are highly SHAPE reactive. Final conditions in SHAPE modification reactions were 400.8 nM *in vitro*-transcribed *T. thermophilus* LSU rRNA, 6.5 mM NMIA, 200 mM NaOAc, 50 mM NaHEPES, pH 8 and either 1 mM 1,2-diaminocyclohexanetetraacetic acid (DCTA) or 10 mM $MgCl_2$ in a total volume of 440 µL. Lyophilized RNA was rehydrated with nuclease-free water in a vinyl anoxic chamber (Coy Laboratory Products). Anoxic conditions were required for parallel experiments reported in chapter 5. Buffer and either DCTA (Na^+ samples) or $MgCl_2$ (Na^+/Mg^{2+} samples) were added, and RNA was refolded by heating to 60 °C for 30 s followed by equilibration at room temperature. In Na^+, the LSU rRNA is expected to form secondary structure only, and changes observed upon addition of Mg^{2+} can be attributed to tertiary interactions induced by direct or indirect effects of RNA-Mg^{2+} interactions present in the near-native folded state.[14] DCTA, a high-affinity chelator of divalent cations, was included in Na^+ samples to ensure absence of contaminating divalent cations in solution, providing a 'divalent-free RNA' baseline reading. This strategy has been employed previously to study low-concentration cation interactions in other RNA systems.[156] Folded RNA samples were divided equally into SHAPE modification and background control reactions. NMIA prepared in anhydrous DMSO was added to SHAPE reactions, and anhydrous DMSO alone was added to background control reactions. Modification reactions were carried out at 37 °C for 1 h. Divalent cations were removed from samples by mixing with Chelex 100 resin (Bio-Rad) followed by centrifugation through 0.22 µm filters. RNA was purified from reaction mixtures outside the anaerobic chamber via sodium acetate precipitation, and pellets were

resuspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). RNA recovery was >80%.

4.2.3 Analysis of NMIA-modified *T. thermophilus* LSU rRNA by primer extension

NMIA-modified RNA solutions were divided equally into reactions with 11 different 5'-[6-FAM]-labeled DNA reverse transcription (RT) primers (primer sequences can be found in Table 4.1). Primers were designed such that their 5' ends targeted non-helical regions of the LSU rRNA, based on the known secondary structure. Modified RNA (20 µL) was added to 12 pmol of each primer in 15 µL of TE buffer. Primers were annealed to RNA by heating to 95 °C for 30 s, incubation at 65 °C for 3 min, and then cooling to 4 °C. RT reactions (50 µL total) were assembled with 10 µL 5x Maxima RT Buffer (Thermo Scientific), a final dNTP concentration of 2.5 mM, and 200 U Maxima Reverse Transcriptase (Thermo Scientific) added last. Reactions were incubated at 55 °C for 2 h and terminated by heat inactivation at 85 °C for 5 min. Sequencing reactions for use in data processing were run with 3 pmol of unmodified RNA, prepared and run as previously described for RT reactions. RNA was sequenced by RT/chain termination with all four ddNTPs at a ratio of 1:1 ddNTP to dNTP. A control reaction without ddNTPs was also prepared.

RT reaction mixture (1 µL) was mixed with 0.3 µL of ROX-labeled DNA sizing ladder (for alignment of disparate traces) and 8.7 µL Hi-Di Formamide (Applied Biosystems) in a 96-well plate. SHAPE and sequencing samples generated with the same RT primers were loaded on the same plate to minimize variation in electrophoretic migration. This allows for easier alignment/assignment of data peaks to corresponding

sequencing peaks during data processing. Plates were heated at 95 °C for 5 min and resolved by CE using a 3130 Genetic Analyzer (Applied Biosystems) at 65 °C with a custom fluorescence spectral calibration. The capillary array was loaded with Performance Optimized Polymer-4 (Applied Biosystems).

Table 4.1. Reverse transcription primers used in SHAPE analysis.

Primer Sequence (5'→3')	Target region
AAGCTTGGAGGGGTCAAGACCTC	2891-2913[a]
GCCCGTGGCGGATAGAGAC	2608-2626
CCGACATCGAGGTGCCAAACCTCC	2486-2509
CCCGCCAAGCGCTGTC	2224-2239
GTTCAATTTCACCGGGTCCC	2002-2021
CGCGCCTGAGTGCTCTTG	1631-1647
CGTTACTCATGCCGGCATTCGC	1250-1271
CCTTAGCTGGCGGTCTC	987-1003
GCTATCTCCGGGCTCGG	790-806
CTTTTCACCTTTCCCTCACGG	461-481
CAGGGAGTCGATTTCTCTTTCC	219-240

[a]Native *T. thermophilus* LSU rRNA sequence ends at nucleotide 2902. Nucleotides 2903-2913 are remnants of cloning/*in vitro* transcription, and are not expected to interact with LSU rRNA.

4.2.4 SHAPE data processing and analysis

Capillary electropherograms were converted into SHAPE reactivities using previously-described in-house MATLAB scripts.[118,126] In summary, data processing involves alignment of disparate traces, baseline correction, peak detection and annotation, peak integration, decay correction and scaling, background subtraction, normalization, and averaging of data from multiple primers in regions with overlapping data. Reverse transcription primers were able to read between 198 and 590 nt. In certain regions, two or three distinct values for the same nucleotide from different primer reads were averaged to yield a final SHAPE value, providing a degree of internal control. Generally, overlapping SHAPE data from independent primer reads were in excellent agreement. Processed

SHAPE data were inspected for disagreement between overlapping regions from independent primer reads, and a small proportion of data was discarded based on these disagreements.

4.3 – Results

4.3.1 Production of LSU rRNA

LSU rRNA was generated and purified by E.B. O'Neill. Purified RNA runs as a single tight band on a denaturing PAGE gel (SequaGel UreaGel System, National Diagnostics, Figure 4.2).

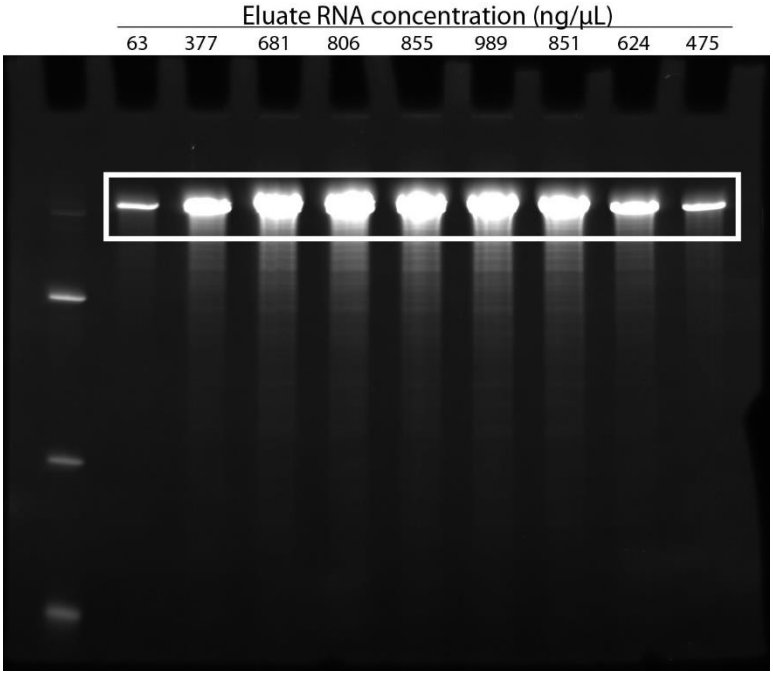


Figure 4.2. Denaturing PAGE gel of purified, *in vitro*-transcribed LSU rRNA. Equal volumes of eluate were loaded in each lane. RNA concentrations of each sequential eluate fraction are indicated, as determined by absorbance at 260 nm. White box denotes bands representing LSU rRNA.

4.3.2 SHAPE reactivity for LSU rRNA is consistent with secondary structure

SHAPE data for the LSU rRNA in Na^+ and Na^+/Mg^{2+} exhibit excellent agreement with predictions of the canonical secondary structure.[177,187] SHAPE data were collected for 2890 of the 2911 LSU rRNA nucleotides (>99%). Figure 4.3 depicts SHAPE reactivities mapped onto the LSU rRNA secondary structure in Na^+ or Na^+/Mg^{2+}, and full numeric SHAPE data are available in Dataset 1. The Na^+ condition used here is known to promote formation of secondary but not tertiary structure.[14,15] Unless otherwise noted, results described in this section are consistent between Na^+ and Na^+/Mg^{2+} conditions. Virtually all LSU rRNA predicted to be double-stranded by the canonical secondary structure appears double-stranded here in Na^+, as indicated by low SHAPE reactivity (<0.6). Less than 1% of internal helical nucleotides are SHAPE reactive. Regions of the canonical secondary structure that do not participate in local cWW base pairs (non-canonical base pairs, bulges, loops, and other unpaired nucleotides) exhibit variable reactivity as expected, with many nucleotides exhibiting moderate to high SHAPE values (Table 4.2). Virtually all LSU rRNA nucleotides that exhibit moderate to high SHAPE reactivity (>0.6) are in loops, bulges, mismatches, or otherwise single-stranded regions of the canonical secondary structure. Examples of long canonical helices that display consistently low SHAPE reactivity can be found in Table 4.2. Many rRNA helices bear small bulges (1-3 nt), mismatches, and internal loops. These features exhibit moderate to high SHAPE reactivity (Table 4.2, Figure 4.3) but the flanking canonical base pairs still exhibit low SHAPE values. More often, helices bearing small non-canonical structures exhibit consistently low reactivity, even at the non-canonical positions (Table 4.2, Figure 4.3). Although these nucleotides may not be involved in local

canonical base pairs, they are still constrained (by non-local or non-canonical interactions) in such a way as to be unreactive to SHAPE. Less than 10% of helices contain one or more expected helical nucleotides that exhibit moderate to high SHAPE reactivity (>0.6, Table 4.2). Greater than half of reactive helical nucleotides are adjacent to non-canonical structures (bulges, mismatches, loops), which could induce increased mobility/flexibility of neighboring nucleotides. Prime examples of SHAPE reactive non-canonical regions are listed in Table 4.2, including helix-capping loops, internal loops/bulges, and intervening non-canonical regions between helices. Low SHAPE reactivities observed for helical nucleotides and restriction of reactive sites to non-helical regions support formation of the canonical LSU rRNA secondary structure in the absence of all other LSU components.

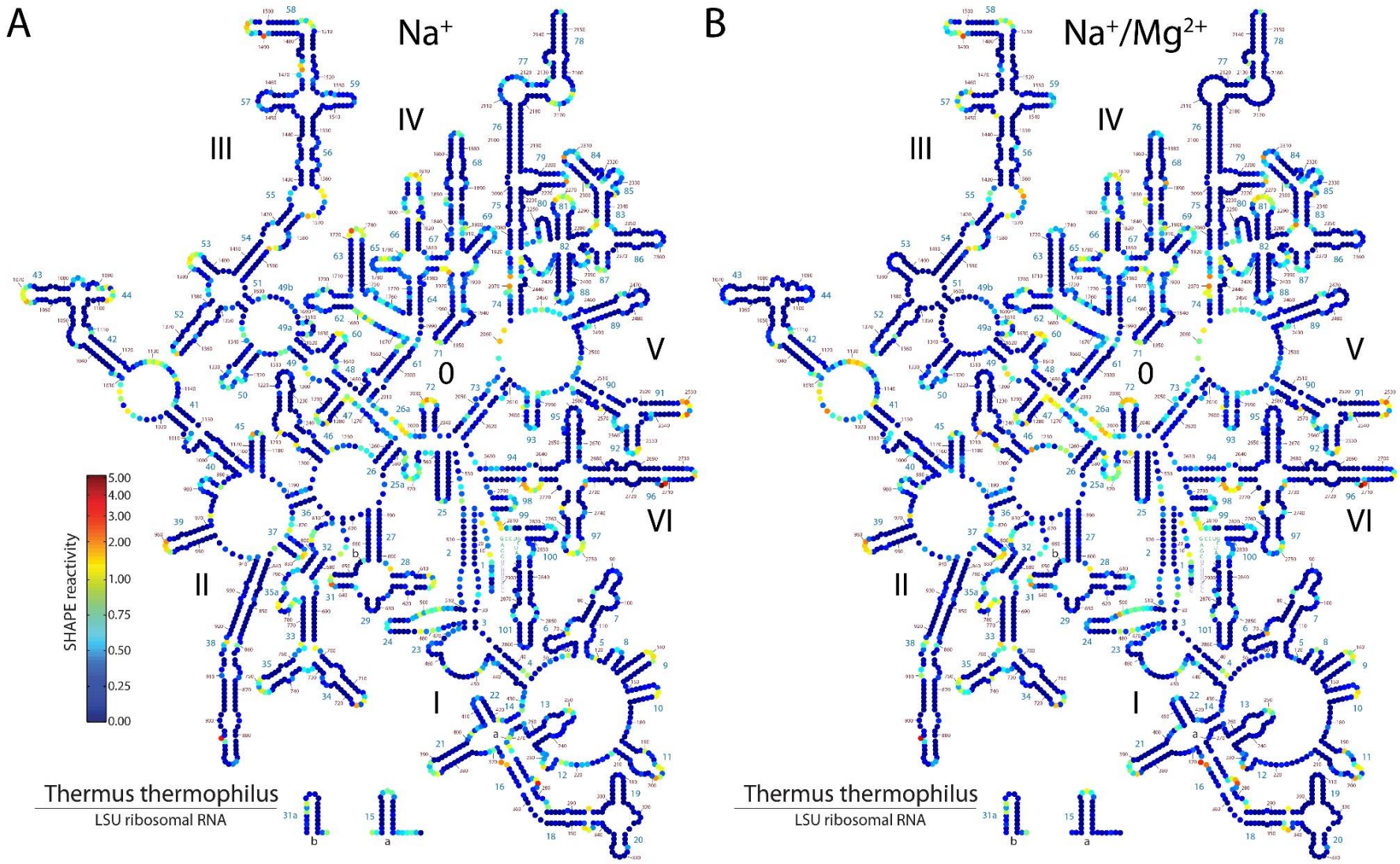

A
Na+
B
Na+/Mg2+
III
IV
V
VI
II
I
0
SHAPE reactivity
5.00
4.00
3.00
2.00
1.00
0.75
0.50
0.25
0.00
Thermus thermophilus
LSU ribosomal RNA
Thermus thermophilus
LSU ribosomal RNA

Figure 4.3. SHAPE reactivities for the *T. thermophilus* LSU rRNA. SHAPE reactivities are mapped onto LSU rRNA secondary structure in presence of A) Na^+ or B) Na^+/Mg^{2+}. All samples contained 200 mM NaOAc, 50 mM NaHEPES, pH 8. Helix numbers (blue), domain numbers, and insertion points are indicated. Nucleotide numbers are denoted every ~10 nt (*E. coli* numbering). Regions where SHAPE data is not available (5' and 3' ends) are displayed as sequence only. Larger versions of these figures are provided in Appendix C (Figures C.1 and C.2). Figures generated with RiboVision.

Table 4.2. LSU rRNA helices exhibiting low vs. high SHAPE reactivity in both Na^+ & Na^+/Mg^{2+}

Low SHAPE reactivity in Na^+ & Na^+/Mg^{2+}	
Long unreactive LSU rRNA helices	Helices bearing unreactive bulges, mismatches, or internal loops
2	2
21	7
25	13
27	28
33	34
42	35
54	39
63	41
76	50
	52
	77
	78
	86
	88
	95
	101

Helices bearing indicated features with moderate to high SHAPE reactivity in Na^+ & Na^+/Mg^{2+}				
Helix-capping loops	Internal loops/bulges	Inter-helical regions	Bulges, mismatches, or internal loops	Reactive helical nucleotides
9	16	41/42	18	1
31	18	55/56	38	4
34	38	61/62	62	31a
39	58	73/74	66	35a
66	96		68	47
72	98		89	65
81			90	67
91			91	73
			93	74
			94	
			96	

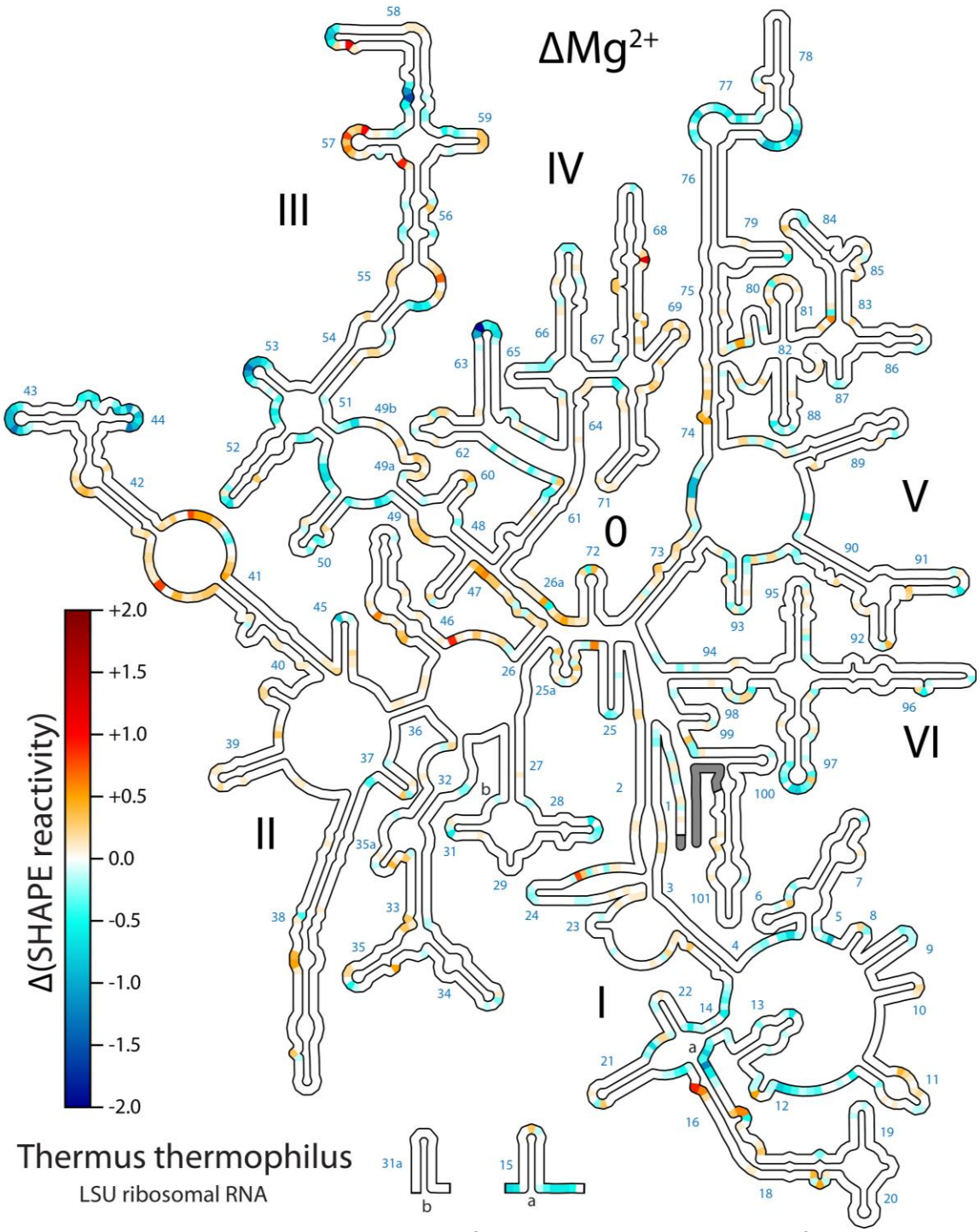


Figure 4.4. *T. thermophilus* LSU rRNA exhibits Mg^{2+}-dependent structural changes. Mg^{2+}-induced changes in SHAPE reactivity (ΔMg^{2+} values) mapped onto LSU rRNA secondary structure. Positive values; increased SHAPE reactivity, negative values; decreased SHAPE reactivity in presence of 10 mM Mg^{2+}. Grey; data not available. Domain numbers, helix numbers, and insertion sites indicated. Figure generated using RiboVision. Version with more detail provided in Figure 4.5.

4.3.3 The LSU rRNA exhibits Mg^{2+}-induced structural changes

Comparative SHAPE results for the LSU rRNA indicate significant changes in reactivity upon addition of Mg^{2+}. For each nucleotide, Na^+ SHAPE reactivity was subtracted from Na^+/Mg^{2+} reactivity (ΔMg^{2+}, Dataset 1). In analysis of Mg^{2+}-dependent SHAPE changes, values >0.3 SHAPE units (positive or negative) are termed 'ΔMg^{2+} sites', ie. sites which experience a significant Mg^{2+}-dependent alteration in SHAPE reactivity; ~7.5% of LSU rRNA nucleotides exceed this threshold. In presence of 10 mM Mg^{2+}, a larger proportion of the LSU rRNA (128 nt) becomes less reactive to SHAPE than exhibits increased reactivity (92 nt). Overall, Mg^{2+} dependency of SHAPE reactivities are considered to be related to Mg^{2+}-induced changes in nucleotide conformation or interactions. The extent and distribution of changes in SHAPE reactivity suggest a global structure transition from a state comprising secondary structure and minimal local tertiary interactions to the rRNA-Mg^{2+} state containing additional local and long-range tertiary interactions.

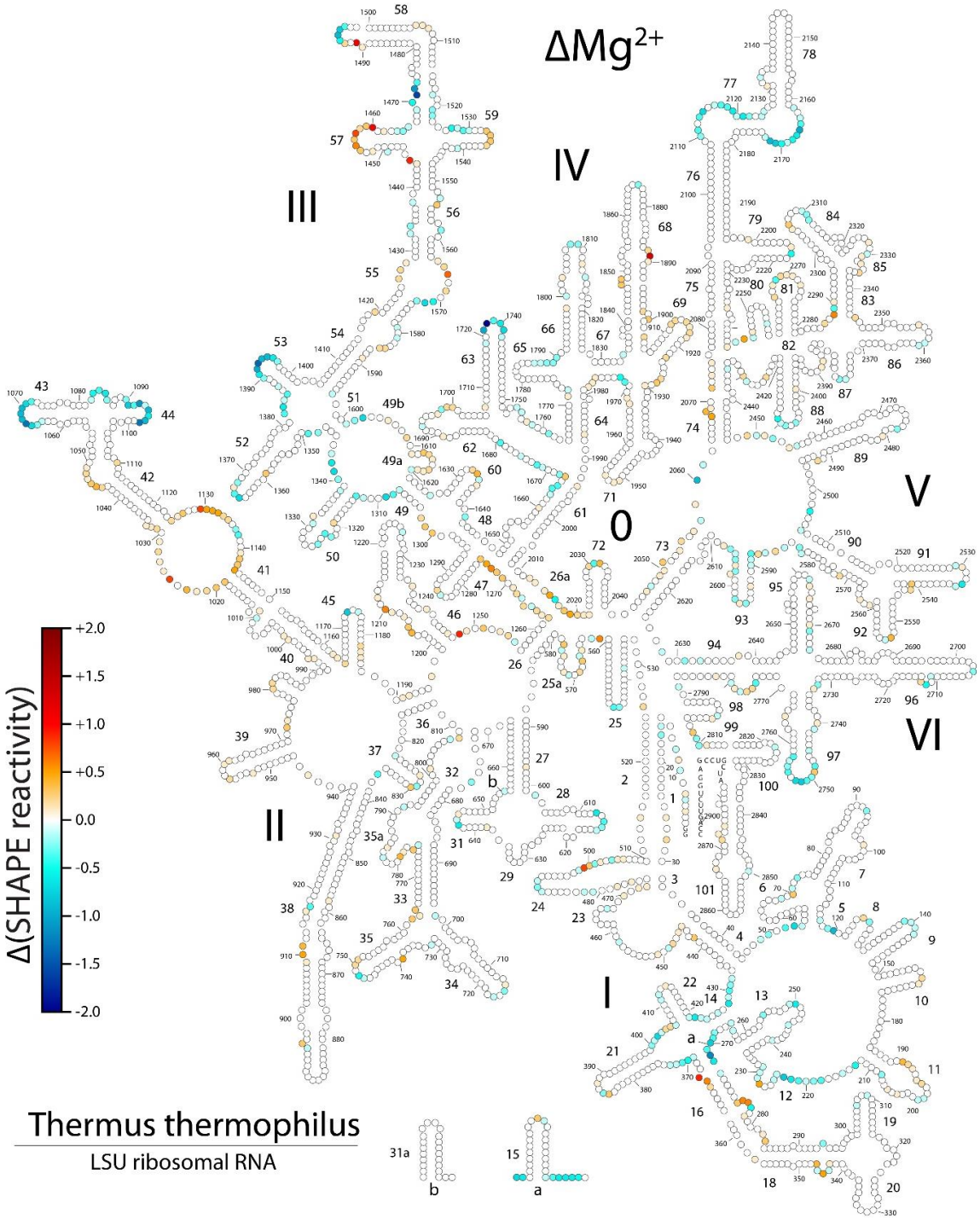


Figure 4.5. Mg^{2+}-dependent structural changes in *T. thermophilus* LSU rRNA. Mg^{2+}-induced changes in SHAPE reactivity are mapped onto LSU rRNA secondary structure, compared against data obtained in presence of Na^{+} only. ΔMg^{2+} value for each nucleotide is indicated by colored circles. Positive values indicate nucleotides with increased SHAPE reactivity in presence of 10 mM Mg^{2+} and negative values denote decreased reactivity. Regions where SHAPE data is not available in at least one data set (5' and 3' ends) are displayed as sequence only. Nucleotide numbers are indicated every ~10 nt (*E. coli* numbering). Helix numbers (blue), domain numbers, and insertion points are indicated. Figure generated with RiboVision.

4.3.4 LSU rRNA helices are not responsive to Mg^{2+}

Canonical cWW helices are invariant to Mg^{2+} (Figures 4.4 and 4.5), consistent with little or no change in their base pairing. Of the >100 canonical helices found in the LSU rRNA, only three—H16, H77, and H83—contain internal ΔMg^{2+} sites not immediately adjacent to a non-canonical region. These sites all exhibit decreases in SHAPE reactivity consistent with formation/stabilization of a canonical base pairing. Three ΔMg^{2+} sites are found at guanine nucleotides of H16, in a short segment at the 5' end of the helix (positions 272B-272D). These nucleotides decrease significantly in SHAPE reactivity in presence of Mg^{2+}, indicating reduced flexibility/increased stability of the 5' end of H16. Single guanine residues near the 5' ends of H77 (G2121) and H83 (G2290) exhibit similarly decreased SHAPE reactivity in presence of Mg^{2+}, consistent with helix ends that are frayed in the absence of Mg^{2+} and stabilized in its presence. The overall absence of helical ΔMg^{2+} sites confirms that, in presence of moderate [Na^{+}], Mg^{2+} does not cause changes to RNA secondary structure.

4.3.5 ΔMg^{2+} sites are restricted to non-helical regions of the LSU rRNA

Mg^{2+} induces increases and decreases in SHAPE reactivity, which are dispersed throughout the LSU rRNA in primary and secondary structure. Overall, only those nucleotides found in loops, bulges and otherwise unpaired secondary structural regions are responsive to Mg^{2+} (Figures 4.4 and 4.5) indicating that they experience structural changes consistent with formation of tertiary structure. ΔMg^{2+} sites are frequently found in helix-capping RNA loops. Many internal loops and bulges bear one or more ΔMg^{2+} sites. Substantial Mg^{2+}-dependent changes in SHAPE reactivity are also observed in

inter-helical regions. A summary of LSU rRNA regions that bear ΔMg^{2+} sites is provided in Table 4.3. No correlation is observed between ΔMg^{2+} sites in the LSU rRNA and RNA-Mg^{2+} first-shell interaction sites observed in the assembled ribosome. The extent of observed ΔMg^{2+} sites, in combination with their restriction to non-helical regions, is consistent with a model of Mg^{2+}-induced tertiary collapse of the LSU rRNA to the rRNA-Mg^{2+} state through formation of tertiary interactions.

Table 4.3. LSU helices that exhibit ΔMg^{2+} sites

Classification of ΔMg^{2+} site-bearing helices		
Helix-capping loops	Internal loops/bulges	Intervening non-canonical regions between helices
25	16	2/24
28	18	4/5
31	35	4/14
35	38	5/8
37	46	11/12
43	50	14/15
44	56	15/16
45	58	16/21
52	61	21/22
53	68	25/25a
57	74	26/46
58	96	41/42
59		49/50
60		49a/51
63		50/51
69		52/53
72		55/56
79		73/74
81		76/77
84		77/78
88		
91		
92		
97		

4.4 – Discussion

4.4.1 Mg^{2+}-induced SHAPE changes are consistent with tertiary interactions

We propose a model in which Mg^{2+}-induced changes in SHAPE reactivity are related to induction of native tertiary interactions in the LSU rRNA by Mg^{2+}. In this model, the interactions suggested by ΔMg^{2+} sites are similar to those observed in the assembled ribosome. Five lines of evidence support this hypothesis: i) ΔMg^{2+} sites are highly focused in rRNA loops, bulges, or otherwise non-helical regions, which are expected to form tertiary interactions, ii) a robust correlation is observed between ΔMg^{2+} sites in the LSU rRNA and nucleotides involved in tertiary interactions in the assembled LSU, iii) ΔMg^{2+} sites are absent from regions which are not expected to form tertiary interactions, such as helices and non-interacting loops, iv) correlations are observed between ΔMg^{2+} sites and rRNA regions involved in inter-domain interactions in the assembled LSU, and v) it has been demonstrated previously that Mg^{2+} effects on SHAPE reactivity are related to formation of RNA tertiary interactions.[115,138,139]

4.4.2 ΔMg^{2+} sites correlate with tertiary interactions

Using the RiboVision visualization suite,[188] we have compared RNA-RNA tertiary interactions[189] observed in the assembled *T. thermophilus* ribosome[75] to ΔMg^{2+} sites. To simplify our analysis, 28 ΔMg^{2+} regions of rRNA were selected which represent the highest local concentrations of ΔMg^{2+} sites (Figures 4.6 and 4.7). There is a clear correlation between these ΔMg^{2+} regions and nucleotides involved in RNA-RNA tertiary interactions; nearly all ΔMg^{2+} sites occur at or directly adjacent to one or more

nucleotides that participate in tertiary interactions in the assembled LSU (Figures 4.6 and 4.7). Conversely, loops that do not participate in RNA-RNA tertiary interactions within the assembled LSU exhibit ΔMg^{2+} values close to zero; counter-examples that meet this criteria include the loops that cap H15, H34, H45, H68, and H69 (Figures 4.4 and 4.5). Most nucleotides in these loops experience little to no Mg^{2+}-dependent change in SHAPE reactivity, and those few nucleotides that do yield significant ΔMg^{2+} values exhibit decreases, suggesting local Mg^{2+}-induced stabilization of loop structures. H34, H68, and H69 are involved in intersubunit bridging interactions with the SSU *in vivo*.[190] In the experiments presented here only the LSU rRNA was present, and therefore these loops are not expected to form the intersubunit interactions found in the assembled ribosome. In general, we consider the Mg^{2+}-dependent formation of a specific native tertiary interaction to be supported by comparative SHAPE if a ΔMg^{2+} site is observed at or directly adjacent to one or more nucleotides involved in the interaction (certain RNA-RNA interaction types may not induce SHAPE-detected structural changes in both regions involved). To illustrate this relationship between tertiary interactions and ΔMg^{2+} sites, structures including native tertiary interactions are presented for two ΔMg^{2+} regions in Figure 4.6.

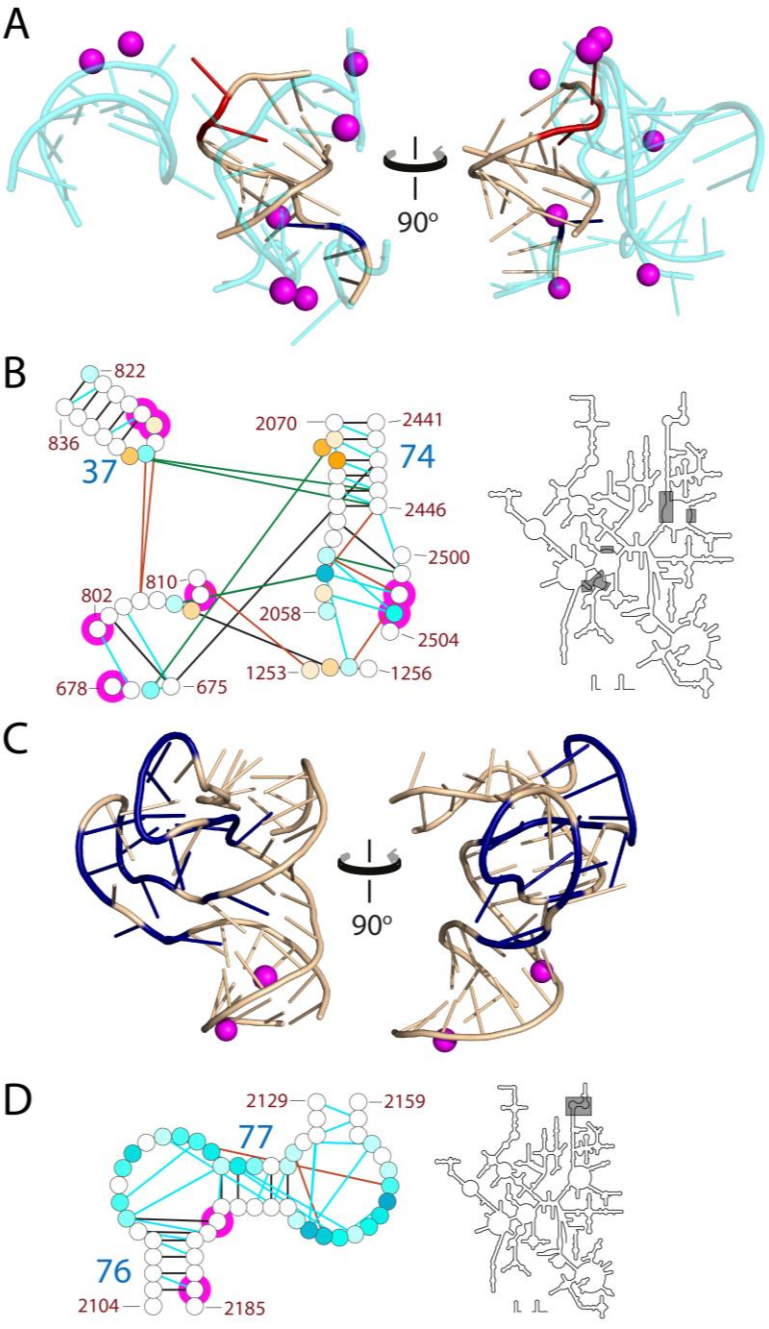
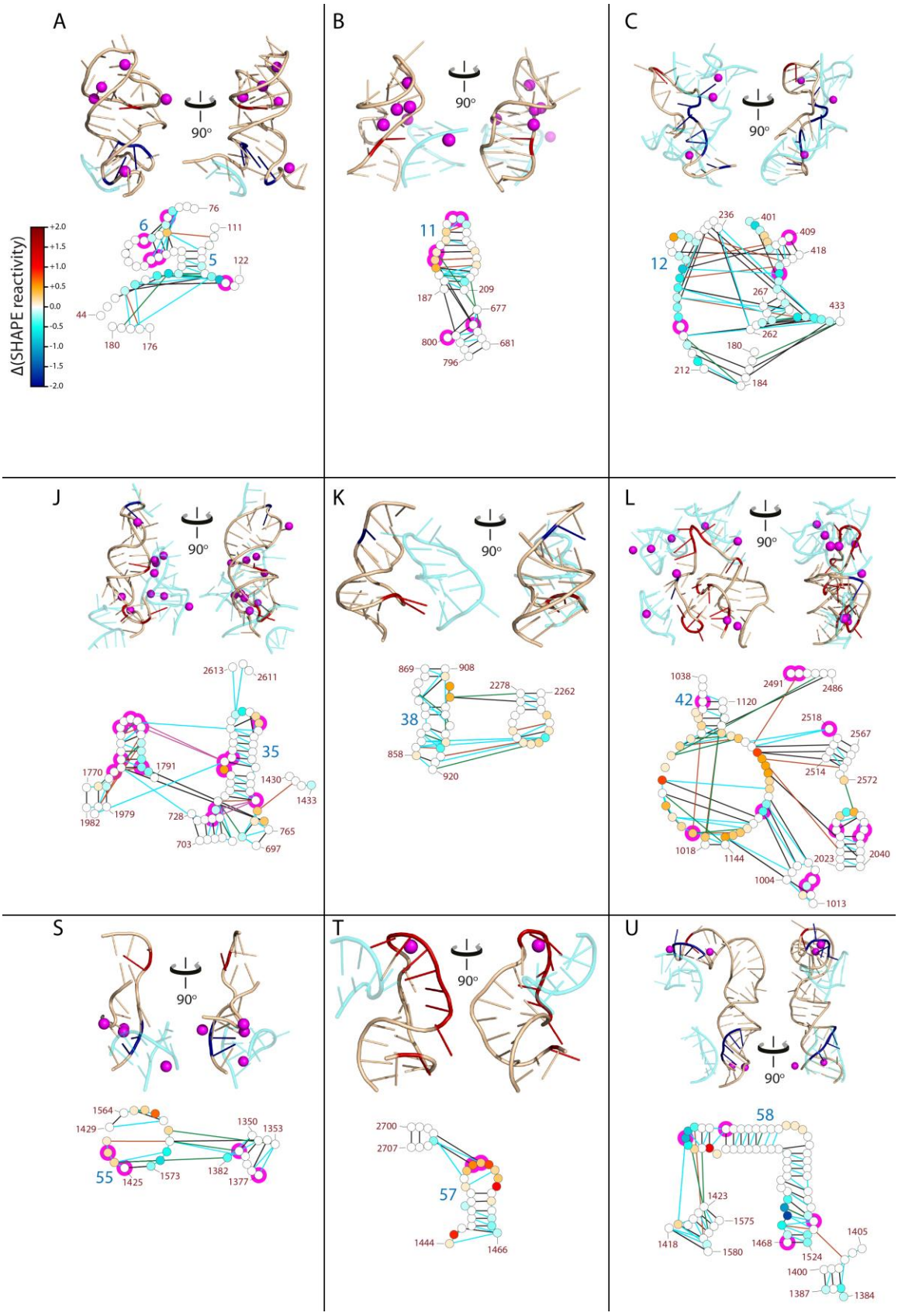
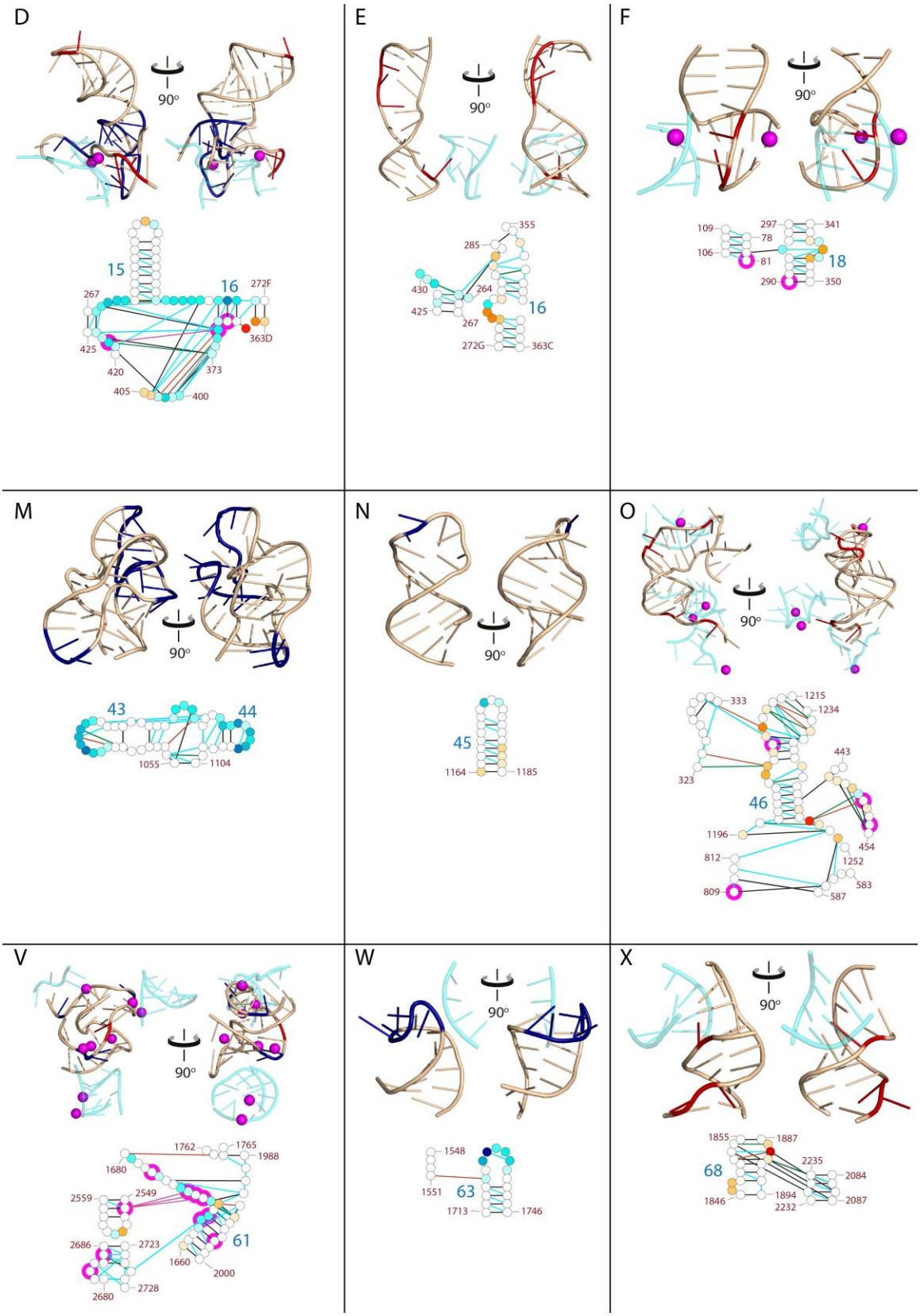

A
90°
B
822
836
37
2070
2441
74
2446
2500
802
810
2058
2504
678
675
1253
1256
C
90°
D
2129
2159
77
76
2104
2185
E
D VI V IV III II I
0 ✓ ✓ ✓ ✓
I ✓ ✓
II ✓ ✓ ✓ ✓
III
IV ✓
V ✓
Interaction
Native
rRNA-Mg2+
None

Figure 4.6. Mg^{2+}-induced structural changes in LSU rRNA are consistent with tertiary interactions. A & C) 3D structure of interacting LSU rRNA regions from the *T. thermophilus* ribosome crystal structure (PDB IDs: 2J00 and 2J01). H74 (A) and the uL1 protuberance (C) are colored by ΔMg^{2+} values (red, increased reactivity; dark blue, decreased reactivity; wheat, little/no reactivity change). Translucent cyan; LSU rRNA segments that interact near ΔMg^{2+} sites. Magenta spheres; first shell-interacting Mg^{2+} cations (2.4 Å cut-off). Figures generated with PyMol. B & D) Mg^{2+}-induced changes in SHAPE reactivity mapped onto secondary structure for H74 (B) and the uL1 protuberance (D). Nucleotides represented as circles, colored as in Figure 4.4. Lines; RNA-RNA interactions observed in 3D (black, base-base; orange, base-phosphate; blue, base-stacking; and green, base-sugar). Underlaid magenta circles; nucleotides in first-shell interaction with Mg^{2+} ions. Gray boxes on full secondary structure highlight displayed regions. E) Inter-domain interactions of the LSU rRNA. Green; domain pairs which interact through long-range RNA-RNA interactions in assembled LSU. Checkmarks; data supports presence of partial or complete inter-domain interactions in the rRNA-Mg^{2+} state.


A
B
C
J
K
L
S
T
U
90°
Δ(SHAPE reactivity)
+2.0
+1.5
+1.0
+0.5
0.0
-0.5
-1.0
-1.5
-2.0
6
5
76
111
122
44
180
176
11
187
209
677
800
681
796
12
236
401
409
418
267
433
262
180
212
184
35
2613
2611
1770
1791
1430
1433
1982
1979
728
765
703
697
38
869
908
2278
2262
858
920
42
1038
1120
2491
2486
2518
2567
2514
2572
1018
1144
2023
2040
1004
1013
55
1564
1429
1350
1353
1425
1573
1382
1377
57
2700
2707
1444
1466
58
1423
1575
1418
1580
1468
1405
1524
1400
1387
1384


D
90°
15
16
272F
267
363D
425
420
373
405
400
E
90°
355
285
430
264
425
16
267
272G
363C
F
90°
109
297
341
78
106
81
18
290
350
M
90°
43
44
1055
1104
N
90°
45
1164
1185
O
90°
333
1215
1234
443
323
46
1196
454
812
1252
583
809
587
V
90°
1762
1765
1988
1680
2549
2559
2686
2723
61
1660
2000
2728
2680
W
90°
1548
1551
63
1713
1746
X
90°
1855
1887
2235
68
2084
1846
1894
2087
2232

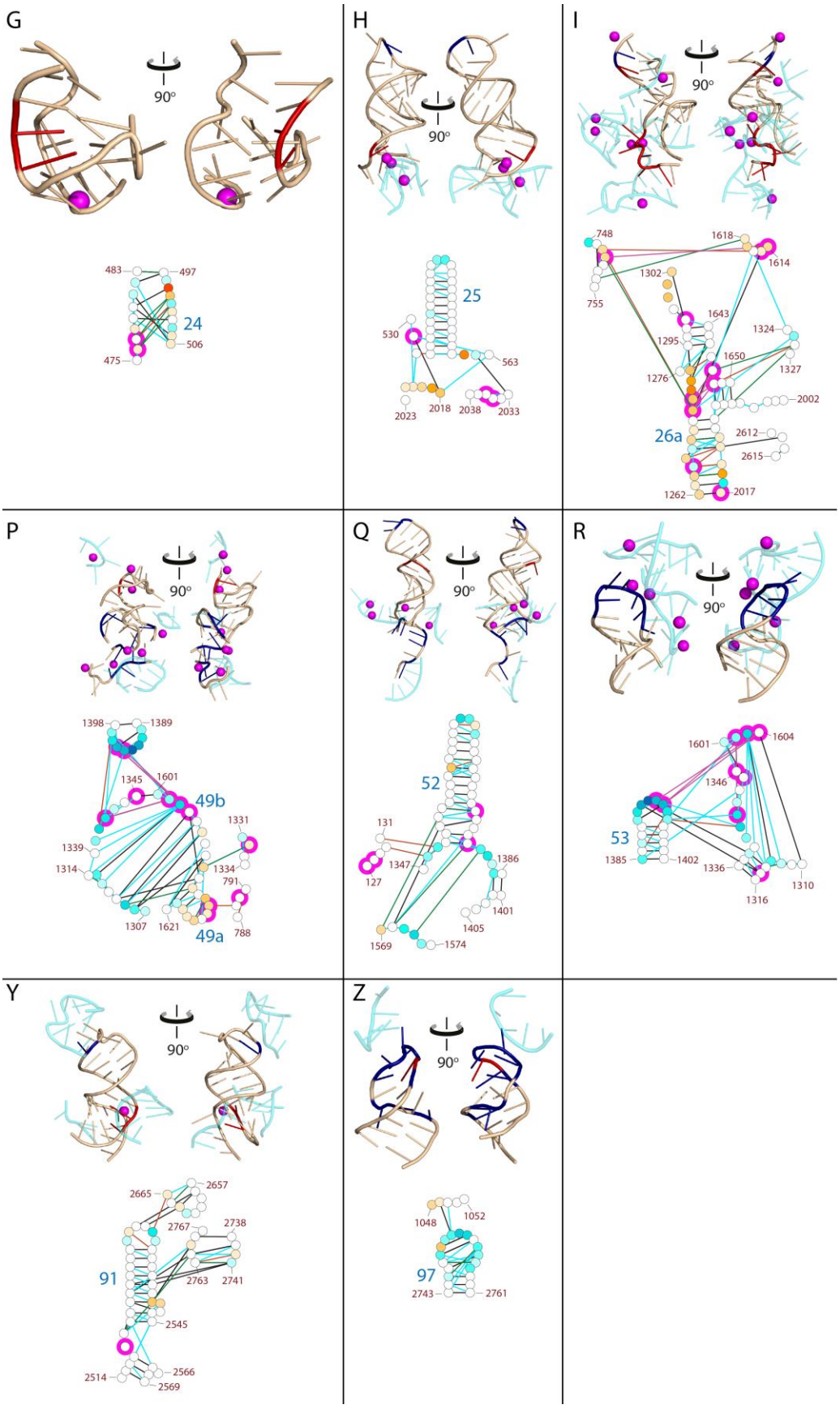

G
90°
483
497
24
506
475
H
90°
25
530
563
2023
2018
2038
2033
I
90°
748
1618
1614
1302
755
1643
1324
1295
1650
1327
1276
2002
26a
2612
2615
1262
2017
P
90°
1398
1389
1345
1601
49b
1331
1339
1314
1334
791
1307
1621
49a
788
Q
90°
52
131
1347
1386
127
1401
1405
1569
1574
R
90°
1601
1604
1346
53
1385
1402
1336
1310
1316
Y
90°
2665
2657
2767
2738
91
2763
2741
2545
2514
2566
2569
Z
90°
1048
1052
97
2743
2761

Figure 4.7. ΔMg^{2+} regions of the LSU rRNA. Selected regions of LSU rRNA with high densities of ΔMg^{2+} sites are shown with regions of rRNA that interact at or near the ΔMg^{2+} sites. Top Panels: 3D structure of selected regions of LSU rRNA from the *T. thermophilus* ribosome crystal structure (PDB IDs: 2J00 and 2J01). Nucleotides are colored by ΔMg^{2+} values (red, increased reactivity; dark blue, decreased reactivity; wheat, little to no change in reactivity). LSU rRNA segments that interact at or near ΔMg^{2+} sites of main helix are colored translucent cyan. Mg^{2+} cations in first shell contact with the displayed RNA (2.4 Å cut-off) are represented by magenta spheres. RNA is displayed with PyMol in cartoon representation. Bottom Panels: Mg^{2+}-induced changes in SHAPE reactivity mapped onto LSU rRNA secondary structure. Nucleotides are represented as circles, colored as in Figure 4.4. Lines represent RNA-RNA interactions observed in the *T. thermophilus* ribosome crystal structure (PDB 2J01), as determined by FR3D (black, base-base; orange, base-phosphate; blue, base-stacking; and green, base-sugar). Underlaid magenta circles indicate nucleotides observed to interact directly with Mg^{2+} ions. Same nucleotides are displayed in corresponding 3D and 2D representations. All samples contained 200 mM NaOAc, 50 mM NaHEPES, pH 8. Helix and nucleotide numbers are indicated.

H74, which lies at the functional and structural core of the LSU, exhibits Mg^{2+}-dependent SHAPE reactivity consistent with formation of tertiary interactions. In the assembled LSU, H74 interacts with RNA from LSU D0, DII, and DV, in addition to Mg^{2+} ions (Figure 4.6, panels A and B) to form a portion of the LSU functional core, the peptidyl transferase center. Canonical helix regions in H74 and its interaction partner H37 are marked by stretches of cWW base pairs (black lines, Fig 4.6B) and cross-strand base-stacking interactions (cyan lines). H74 exhibits three ΔMg^{2+} sites; A2060 decreases in SHAPE reactivity, while G2067 and U2068 increase in SHAPE reactivity. In the assembled LSU, A2060 is involved in several long-distance tertiary interactions: a base-sugar interaction between its primary amine and the 2'-OH of U807, and base-stacking interactions with U1255 and G2502. The ΔMg^{2+} data imply that one or more tertiary interactions are formed, restraining A2060 and making it significantly less SHAPE reactive. Increased reactivity at G2067 and U2068 can also be attributed to nearby tertiary interactions. U2068 is a single nucleotide bulge, and G2067 neighbors the bulge. Neither U2068 nor G2067 are directly involved in long-range RNA-RNA tertiary interactions as classified by FR3D, but neighboring nucleotide G2609 participates in a

base-sugar interaction with A676 in which the amine of A676 contacts the 2'-OH of G2609. U827 and G805 contain atoms within 5 Å of U2608, but not close enough to be classified as a direct interaction by the geometric pattern recognition used by FR3D. Indirect Mg^{2+}-induced formation of one or more of these tertiary interactions may cause U2068 to 'flip-out' of H74, into a conformation with a greater 2'-OH/PO interatomic distance, resulting in increased SHAPE reactivity. Higher reactivity at G2067 may be related to increased instability and strain caused by formation of tertiary interactions at positions U2068 or G2609. Nearby base-sugar tertiary interactions between the loop of H37 and G2445/G2446 may also influence SHAPE reactivity of G2067 and U2068. ΔMg^{2+} sites in H74 correlate with nucleotides involved in tertiary interactions that connect DV to distant regions of D0 and DII, consistent with formation of part of the PTC. These few interactions alone would be enough to cause a significant decrease in the radius of the LSU rRNA.

Mg^{2+}-dependent SHAPE changes in the uL1 protuberance suggest formation of a network of native base-stacking and base-phosphate interactions. The uL1 protuberance (Figure 4.6C-D) is an independently-folding LSU rRNA region which protrudes from the exterior of the LSU and interacts with rProtein uL1.[191] In the assembled LSU, the uL1 protuberance forms tertiary loop-loop interactions between nucleotides 2109-2119 and 2161-2173 (Figure 4.6C-D). Within the uL1 protuberance, helical regions of H76 and H77 are marked by stretches of cWW base pairs and cross-strand base-stacking interactions (Figure 4.6D). The inter-helical loops decrease in SHAPE reactivity upon introduction of Mg^{2+}, consistent with structural stabilization (Figure 4.3). Thirteen ΔMg^{2+} sites are found in the uL1 protuberance, all negative, indicating increased

stability/decreased flexibility in response to Mg^{2+} (Table 4.4). These changes in SHAPE reactivity are consistent with formation of local tertiary interactions at or near the ΔMg^{2+} sites. Eight base-stacking tertiary interactions and two base-phosphate interactions are observed within the uL1 protuberance in the assembled LSU (Figure 4.6D, Table 4.4). All ΔMg^{2+} sites are at or near nucleotides involved in these local tertiary interactions, and are in agreement with formation of that interaction network. Two base-phosphate tertiary interactions occur within the uL1 protuberance; the primary amine group of G2165 contacts a PO of G2116, and the carbonyl group of G2125 contacts a PO of A2171. The 5' strand of the 5 base-pair H77 exhibits decreased SHAPE reactivity, supporting Mg^{2+}-induced stabilization (Figure 4.6D). A robust correlation is observed between ΔMg^{2+} sites observed in the uL1 protuberance and nucleotides involved in loop-loop interactions (Figure 4.6D). As in H74 and the uL1 protuberance, virtually all ΔMg^{2+} sites observed in the LSU rRNA occur at or directly adjacent to nucleotides involved in local or long-range tertiary interactions in the assembled LSU (Figure 4.7). In our model, these Mg^{2+}-dependent structural changes suggest formation of an intricate tertiary interaction network, consistent with global collapse of the LSU rRNA into the rRNA-Mg^{2+} state.

Table 4.4. ΔMg^{2+} sites and base-stacking interactions in the uL1 protuberance

Sites of Mg^{2+}-induced SHAPE reactivity decrease	Nucleotide pairs involved in base-stacking interactions[a]
U2113	G2110/G2120
A2114	C2111/U2118
A2117	G2116/A2171
U2118	A2119/A2170
A2119	A2119/A2169
G2121	C2128/U2172
G2165	G2162/U2172
G2166	G2166/A2170
U2167	
G2168	
A2170	
A2171	
U2172	

[a] Interactions defined by FR3D

Nucleotides that directly contact Mg^{2+} are dispersed throughout the secondary structure, in all seven domains of the LSU rRNA (Figure 4.1). In 3D, proportional incidence of Mg^{2+} is highest in the functional core of the LSU and decreases drastically with increasing radial distance from the PTC.[23,35] Some ΔMg^{2+} sites are at or near direct Mg^{2+}-coordinating nucleotides, while a roughly equal proportion are distant in primary, secondary, and 3D structure from nucleotides expected to interact directly with Mg^{2+} (Figures 4.6 and 4.7). Conversely, a considerable proportion of nucleotides that bind Mg^{2+} in the assembled LSU are distant from ΔMg^{2+} sites, suggesting that Mg^{2+}-binding does not consistently translate to altered local SHAPE reactivity. Alternatively, some expected Mg^{2+}-binding sites may not be occupied under the studied conditions. ΔMg^{2+} sites that are distant from direct Mg^{2+}-coordination sites may be influenced by interactions with hydrated Mg^{2+}, loosely bound Mg^{2+} not observed in the 3D structure, or structurally-distant RNA-Mg^{2+} binding events that allow distal RNA regions to come into close proximity in 3D to form RNA-RNA tertiary interactions. In H74, a Mg^{2+} cation directly contacts a phosphate oxygen of G2502 in close proximity to several ΔMg^{2+} sites,

but no Mg^{2+} cations coordinate directly with RNA at or near ΔMg^{2+} sites G2067 or U2068 (Figure 4.6A-B). Two Mg^{2+} cations are coordinated directly by the uL1 protuberance, but none at or directly adjacent to ΔMg^{2+} sites (Figure 4.6C-D). Several hexahydrated Mg^{2+} cations interact in a specific manner with the uL1 protuberance, some at nucleotides involved in the loop-loop interaction. The Mg^{2+} effects observed in the uL1 protuberance are consistent with hydrated Mg^{2+} binding, structurally-distant RNA-Mg^{2+} binding events, or non-specific stabilization, not by direct Mg^{2+} coordination. Ultimately, tertiary structure formation and RNA-Mg^{2+} interactions are inextricably entangled, and it is virtually impossible to decouple the two, especially for large RNA structures such as the LSU rRNA.

4.4.3 Defining the rRNA-Mg^{2+} state

If ΔMg^{2+} sites are indeed related to tertiary interaction formation, our data supports formation of the core inter-domain architecture of the assembled LSU in the rRNA-Mg^{2+} state. According to our model, most domain pairs observed to interact in the assembled LSU are inferred to interact in a complete or partial manner in the rRNA-Mg^{2+} state (Figure 4.6E). In the assembled ribosome, long-range inter-domain tertiary interactions interconnect disparate regions of LSU rRNA.[185] We have observed limited inter-domain tertiary interactions with DIII rRNA.[118] Inter-domain interactions are not likely to be present in Na^{+} alone, because moderate monovalent cation concentrations induce secondary structure only.[192] Mg^{2+} is implicated in facilitation of inter-domain contacts.[23] While local intra-domain tertiary interactions may form in absence of Mg^{2+}, inter-domain contacts are dependent on collapse into a globular structure, which requires

divalent cations. The largest number of long-range inter-domain interactions in the assembled LSU is between DII and DV. We observe evidence consistent with Mg^{2+}-induced formation of 3 distinct sets of RNA-RNA interactions between DII/DV (Figures 4.7B, 4.7O, and Table 4.5). In the rRNA-Mg^{2+} state, all long-range inter-domain interactions observed in the assembled LSU are inferred between D0/DIII (Figure 4.7I), D0/DV (Figure 4.7V), DII/DVI (Figures 4.7T, 4.7Z), and DV/DVI (Figure 4.7Y). Mg^{2+}-induced formation of partial inter-domain interactions are inferred for 7 additional LSU domain pairs (Table 4.5). We observe comparative SHAPE evidence in support of partial or comprehensive formation of 12 out of 14 native domain pair interactions (Figure 4.6E, Table 4.5). The network of inter-domain interactions inferred in the rRNA-Mg^{2+} state supports formation of the core domain architecture of the LSU, in the absence of other native interaction partners.

Only two domain pairs that interact in the assembled LSU are not inferred to interact in the rRNA/Mg^{2+} state: DI/DV and DIV/DVI. In both cases, specific rProteins contact the disparate regions that interact natively; bL28 mediates the interaction of DI/DV rRNA, and bL19 mediates interaction of DIV/DVI. These may represent late or otherwise less crucial steps in the overall assembly of the LSU.

4.4.4 Inferring the role of rProteins in LSU assembly

Inter-domain interactions that are not implied in the rRNA-Mg^{2+} state may be pre-organized for protein-mediated interaction by inferred interactions. In the rRNA-Mg^{2+} state, tertiary interactions not supported by correlation of ΔMg^{2+} sites to assembled LSU-interacting regions would fall into one of three classes; i) already present in Na^{+}, ii)

present only in Na^{+}/Mg^{2+}, but undetectable by comparative SHAPE, or iii) absent or unstable in Na^{+}/Mg^{2+}, in which case their formation is dependent on agents other than or in addition to Mg^{2+}. Given the extent and wide distribution of observed ΔMg^{2+} sites, which suggests global collapse of the LSU rRNA, it is unlikely that long-range inter-domain interactions would be formed in absence of Mg^{2+}. Based on this reasoning, long-range interactions for which we have no supporting evidence must fall into either class 'ii' or 'iii'. In the interests of caution, we do not assume formation of any interactions not supported by interpretation of ΔMg^{2+} sites as being induced by tertiary interaction formation. Tertiary interactions that are not present in the rRNA-Mg^{2+} state are likely to require mediation by one of the other interaction partners of the LSU rRNA through electrostatic effects (ie. charge screening), kinetic effects (ie. decreases in local dynamics), or induction of minor conformational changes. Assuming the interactions not directly supported by our data are absent or unstable in the rRNA-Mg^{2+} state, the extent of supported Mg^{2+}-induced tertiary interactions (Figures 4.6 and 4.7) suggests that most nucleotides involved in the absent interactions are pre-organized for association, contingent upon addition of a mediating rProtein (Table 4.6). Several sets of tertiary interactions between DII and DIV are found in the assembled ribosome for which evidence is not found in the rRNA-Mg^{2+} state by comparative SHAPE. The unsupported DII/DIV interactions involve nucleotides in or near H65 and several helices near H35 in secondary structure (Table 4.6). Our data supports Mg^{2+}-induced formation of a long range interaction between DII/DIV in the rRNA-Mg^{2+} state, involving H35 and H65 (Figure 4.7J), the formation of which would undoubtedly bring nearby regions involved in the remaining DII/DIV interactions into much closer proximity in 3D. Using

RiboVision to map locations where each rProtein contacts rRNA in the assembled LSU,[188] it is observed that rProtein uL2 contacts rRNA at or near nucleotides involved in all of the DII/DIV interactions that are not implied in the rRNA-Mg^{2+} state. The inferred inter-domain interaction between H35/H65 may pre-organize DII and DIV such that, if uL2 were present, it could mediate and/or stabilize nearby RNA/RNA interactions. Similar cases can be made for a majority of native tertiary interactions that are not inferred in the rRNA-Mg^{2+} state (Table 4.6); interactions between DII/DV are pre-organized by data-supported interactions between H38/H81 (Figure 4.7K), H42/H89, or H42/H90 (Figure 4.7L) for protein-mediated interaction by uL15 or uL16, interactions between DI/DIII are pre-organized by the data-supported interaction between H8/H52 (Figure 4.7Q) for protein-mediated interaction by uL23, etc. In this model, formation of certain native tertiary interactions is much more favorable in the rRNA-Mg^{2+} state than in Na^{+} alone.

Table 4.5. Long-range inter-domain interactions observed in the assembled LSU, colored by their inferred status in the rRNA-Mg^{2+} state[a]

D	VI	V	IV	III	II	I
0	H61/H96	H61/H92		H26a/H49a H26a/H50	H26a/H35 H35/H73 H42/H72	
	H61/H100 H73/H100				H25a/H39 H40/H72 H35/H73	
I				H8/H52	H4/H46 H11/H32 H20/H46	
				H9/H51	H2/H26 H3/H46 H4/H46 H13/H28 H23/H32 H23/H35a	
		H11/H74 H21/H75			H23/H31 H25/H46 H25/H40	
II	H57/H96 H41/H97	H38/H81 H42/H89 H42/H90	H35/H65	H35/H56		
		H37/H74 H39/H80 H39/H81 H39/H88 H39/H89 H42/H89	H33/H65 H34/H65 H35a/H65			
		H29/H88 H39/H74 H42/H74		H33/H52		
III						
IV		H68/H75				
	H63/H101	H71/H92				
V	H91/H95 H91/H97					

[a]Green shading indicates long-range inter-domain interactions inferred in the rRNA-Mg^{2+} state. Yellow shading indicates native interactions that are not seen to be present in the rRNA-Mg^{2+} state, but may be pre-organized for protein-mediated interaction (see Table 4.6 below). Red shading indicates native interactions not consistent with the interactions inferred in the rRNA-Mg^{2+} state, with no implied pre-organizing interactions. Helix numbers are used for simplicity, though in some cases the interacting regions may be located in nearby non-helical regions.

Table 4.6: LSU long-range inter-domain interactions not inferred in the rRNA-Mg^{2+} state and nearby inferred inter-domain interactions that could pre-organize natively interacting regions for mediation by rProteins

	Interaction(s) not supported in the rRNA-Mg^{2+}-state[a]	Nearby inferred interaction[a]	Figure	Mediating/stabilizing agent
D0-DII	H25a/H39 H40/H72	H42/H72	S4L	bL20 bL20
	H35/H73	H35/H73	S4J	uL3
D0-DVI	H61/H100	H61/H96	S4V	uL3/uL4/bL17
	H73/H100	H91/H97	S4Y	uL3
DI-DII	H2/H26 H3/H46 H4/H46	H4/H46	S4O	bL20 bL20 uL4/bL20
	H13/H28 H23/H32 H23/H35a	H11/H32	S4B	uL15 uL4/bL34 bL34
DI-DIII	H9/H51	H8/H52	S4Q	uL23
DII-DIV	H33/H65 H34/H65 H35a/H65	H35/65	S4J	uL2 uL2 uL2
DII-DV	H37/H74 H39/H80 H39/H81 H39/H88 H39/H89	H38/H81	S4K	uL15 uL16 uL16 uL15 uL16
	H42/H89	H42/H89&H90	S4L	uL16

[a]Helix numbers are used, though in some cases the interacting regions may be located in nearby non-helical regions.

rProteins known to be crucial for ribosome function are implied in the mediation of interactions that are not supported by our model of the rRNA-Mg^{2+} state. Individual omission of rProteins uL2, uL3, uL4, uL15, uL16, and bL20 from reconstituted minimal LSU particles cripples peptidyl transferase activity.[193] These rProteins also exhibit strong interconnectivity and interdependence in LSU assembly maps, and dominate the list of rProteins expected to mediate inter-domain interactions that are not supported by our model of the rRNA-Mg^{2+} state (Table 4.6). Four of these functionally-required rProteins—uL2, uL3, uL4, and uL15—penetrate deeply into the ribosome and include the protein components nearest to the PTC (within 30 Å); these regions constitute the a-rPeptides discussed in chapter 3, which are expected to interact with ancestral models of

LSU rRNA.[157] The correlation between functional and structural importance of this set of rProteins suggests that the interactions they mediate bring the LSU closer to its native, active form.

4.5 – Conclusions

LSU rRNA adopts a well-defined near-native tertiary structure (the rRNA-Mg^{2+} state) in complex with Mg^{2+} alone, absent all other native interaction partners. *in vivo*, the LSU rRNA interacts with the SSU, 5S rRNA and over 30 LSU proteins.[75] The experiments described here involve *in vitro*-transcribed LSU rRNA in solutions containing only buffer and ions (within reasonable physiological ranges), with no additional structural elements. We have used comparative SHAPE to provide single-nucleotide information regarding Mg^{2+}-induced structural changes of the entire LSU rRNA. The LSU rRNA retains its native secondary structure and forms many native tertiary interactions in association with Mg^{2+}. We propose a model of this near-native state based on correlations of Mg^{2+}-induced structural changes with locations known to be involved in tertiary interactions in the assembled LSU. This model, termed the rRNA-Mg^{2+} state, is collapsed with a significantly reduced radius, and includes the core inter-domain architecture of the native LSU, but it is not expected to be catalytically competent. Peptidyl transferase competency is not attributable to a single ribosomal component.[193] Protein-free LSU rRNA is not capable of peptidyl transferase activity, even in presence of Mg^{2+}.[186,194] Certain rProtein-rRNA interactions are expected to be cation-mediated, underscoring the interdependence of rProteins and cations to the

structure of the assembled LSU.[23] rProteins are crucial to ribosomal function, but our model suggests that many of their structural effects are ancillary.

Our results for the LSU rRNA describe the use of comparative SHAPE to support formation of specific Mg^{2+}-induced tertiary interactions in a system >10x times larger than prior studies.[115,138,139] Previously we performed similar Mg^{2+}-dependent SHAPE experiments with two isolated subsets of *T. thermophilus* LSU rRNA: a 615 nt model ancestral LSU rRNA[157] and DIII rRNA,[118] both prepared by *in vitro*-transcription. The observations presented here are consistent with those previous results, in which smaller rRNA systems adopt monovalent cation-induced secondary structures in agreement with established structure, and exhibit Mg^{2+}-dependent structural effects.[118,157]

The results presented here are consistent with previous experiments, in which removal of Mg^{2+} results in global unfolding of the LSU rRNA, as detected by various biophysical techniques.[79] Virtually all aspects of ribosome structure and function involve Mg^{2+} to some degree.[74] Ribosome function can survive significant deproteinization, but is completely abolished upon exposure to divalent cation-chelating compounds, highlighting the necessity of Mg^{2+}-induced interactions for peptidyl transferase activity.[80] We see that Mg^{2+} is crucial to the formation of inter-domain contacts, and the resulting formation of the LSU core architecture. When the assembled LSU is depleted of Mg^{2+}, this core architecture is lost, resulting in global unfolding of the domains and loss of ribosomal function. Our results support a model in which the LSU rRNA requires only Mg^{2+} to fold to a well-defined, collapsed near-native state, and elements such as rProteins induce nominal, local structural changes that foster formation of the remaining interactions to form a fully catalytic LSU.

CHAPTER 5

IRON(II) AND MAGNESIUM(II) BINDING TO LSU RIBOSOMAL RNA AND AN ANCESTRAL CORE UNDER EARLY EARTH CONDITIONS

5.1 – Introduction

5.1.1 Fe^{2+}: RNA's 'first wife'?

When life originated and first proliferated, iron (Fe) was abundant, soluble and benign. The GOE approximately 2.4 billion years ago introduced the modern conditions of iron scarcity and iron-mediated oxidative damage to biological systems. We have developed a model for the ancient earth in which Fe^{2+} was an essential cofactor for nucleic acids with important roles in RNA folding and catalysis, acting instead of or alongside Mg^{2+} in RNA-cation interactions. While Mg^{2+} is the primary divalent cation interaction partner of RNA in modern biology (RNA's extant metaphorical 'significant other'), Fe^{2+} may be thought of as RNA's 'first wife', in that Fe^{2+} may have interacted more readily with RNA under pre-GOE environmental conditions. In this model, modern biochemical systems retain the latent ability to revert to primordial Fe^{2+}-associated states when exposed to pre-GOE conditions. Our previous quantum mechanical calculations, along with folding and ribozyme experiments,[50,51] suggest that Fe^{2+} can substitute for Mg^{2+} in a variety of RNA systems, and that rates of ribozyme catalysis appear greater with Fe^{2+} as a cofactor than with Mg^{2+}.

A summary of important physical and chemical characteristics of Mg^{2+} and Fe^{2+} can be found in Table 5.1. Fe^{2+} has a slightly larger ionic radius than Mg^{2+}, but their usual coordination numbers are identical. While the enthalpy of formation for Fe^{2+} and Mg^{2+} hydrates are very similar, the pK^a of hydrated Fe^{2+} is two units lower than hydrated Mg^{2+}, suggesting that hydrated Fe^{2+} could more readily act as an acid than hydrated Mg^{2+}. In modeling experiments with RNA clamp structures (Figure 1.3),[36,43] in which divalent cations are pinched between PO atoms, Fe^{2+} is estimated to be 1.3 kcal/mol more favorable than Mg^{2+}.[50]

Table 5.1. Comparison of physical and chemical characteristics of Mg^{2+} and Fe^{2+}

	r (Å)[(a)]	AOCN[(b)]	$-\Delta H_{hyd}$[(c)]	pK_a[(d)]	ΔH[(h)]
Mg^{2+}	0.65	6	458[(e)]	11.4	-36.6
$Fe^{2+(f)}$	0.74	6	464[(g)]	9.5	-37.9

(a) Ionic radius;[195] (b) Average Observed Coordination Number;[195] (c) Hydration enthalpy (kcal mol^{-1}); (d) pK_a of $M^{2+}(H_2O)_6$ where $M^{2+} = Fe^{2+}$ or Mg^{2+};[196] (e) From Rashin & Honig;[27] (f) High spin; (g) From Uudsemaa & Tamm;[197] (h) Interaction enthalpy (kcal mol^{-1}) for RNA clamp formation.[50]

Iron, an essential element for life, tends to be octahedrally coordinated[195] and can be found in mixed oxidation state complexes with oxygen and sulfur.[198,199] However, Fe^{2+} inflicts oxidative damage[200] and thus cellular Fe^{2+} concentrations are carefully regulated.[201] Ferrous iron (Fe^{2+}) is highly soluble in aqueous media while Fe^{3+} is highly insoluble.[202] It is commonly believed that iron became deeply enmeshed in protein biochemistry during the first half of Earth's history.[203-205] The widespread use of iron in biological reactions,[201] despite its low concentrations on the oxic surface of the modern earth, highlights this element's deep evolutionary history. We posit that Fe^{2+} was a primeval RNA cofactor that has been largely supplanted by Mg^{2+} in extant biology.

5.1.2 Availability of Fe^{2+} on ancient earth

The transition from chemistry to early biology, and thereafter to modern biology, can be understood only in the context of relevant ambient geochemical conditions. In this case, we undertake the investigation of rRNA, which facilitates the transduction of genetic information into functional units, under relevant early earth conditions.

Ferrous iron (Fe^{2+}) was abundant and benign when life originated and first proliferated. Iron's important role in biology appears to date back to life's origin in the anoxic era during the first half of Earth's history. The geologic record indicates that the early oceans were devoid of O_2 and contained vast quantities of Fe^{2+} (high μM concentrations compared to pM now).[53] These data suggest that Fe^{2+} was broadly available over the Earth's surface. Thus, at the time of life's origin, Fe was abundant, soluble and benign.[53,206-209] For ~2 billion years, the reducing conditions of the ancient Earth sustained soluble Fe^{2+} rather than insoluble Fe^{3+}, preventing destructive Fe^{2+}-mediated oxidative processes such as Fenton chemistry.[210,211]

Approximately 2.4 billion years ago, for reasons under debate, molecular O_2 produced by cyanobacteria began accumulating in the atmosphere, precipitating radical shifts in biochemistry and metabolisms. Surface oxidation introduced the modern condition of iron scarcity and iron-mediated oxidative damage to biological systems, and drove substitution of other metals for iron in protein enzymes.[203-205,212-218] We suggest that as the GOE drove Fe^{2+} replacement by Cu^{2+}, Zn^{2+}, Mn^{2+} and other metals in protein enzymes,[203,219] it also drove Fe^{2+} replacement by Mg^{2+} as the primary divalent cation cofactor for RNA.

5.1.3 Fe^{2+} mimics Mg^{2+} in RNA structure

Fe^{2+} can substitute for Mg^{2+} in RNA structure. Our previous results with the P4-P6 domain of the Protozoan *T. thermophila* Group I intron suggest that RNA can access native structures in pre-GOE conditions.[50] We suggest that differences in electronic structure lead to increased folding stability of Fe^{2+}-RNA vs. Mg^{2+}-RNA complexes. The phosphorus atom is a better electrophile in first shell RNA complexes with Fe^{2+} than in analogous complexes with Mg^{2+}. Our calculations and experiments show that Fe^{2+} can readily substitute for Mg^{2+} in Mg^{2+}-clamp structures (Figure 1.3) that are common in large RNAs such as LSU rRNAs[36] and self-splicing Group II introns.[50] Table 5.1 suggests that the substitution of Mg^{2+} by Fe^{2+} is plausible for diverse RNAs based on the physical parameters of the two metals. Previously, Hillen used RNA cleavage by Fenton chemistry to show that Fe^{2+} interacts at the same sites as Mg^{2+}.[220] Although Fe^{2+}-RNA complexes degrade quickly in the presence of atmospheric O_2, they are relatively long-lived in anoxia.

5.1.4 RNA/Fe^{2+} catalysis

Previous catalysis experiments confirm predictions of quantum mechanical calculations, namely that Fe^{2+} can substitute for Mg^{2+} in a variety of functional RNAs. RNA can access native structures in pre-GOE conditions to form stable, catalytically active assemblies. These observations were obtained from a variety of RNAs, including an *in vitro*-selected L1 ligase, the minimal HHα1 hammerhead ribozyme, and several rRNA constructs.[50,51] We have observed that certain ribozymes are more active with Fe^{2+} than with Mg^{2+}. Very recently, Popović and Ditzler performed what was, to our

knowledge, the first ever *in vitro* RNA selection under plausible pre-GOE conditions.[221] They demonstrate that diverse ribozymes obtained by selection with Fe^{2+} are active upon substitution of Mg^{2+} for Fe^{2+}. Sen demonstrated that RNA can enhance the redox-activity of iron-protoporphyrin IX.[222]

5.1.5 Hypotheses regarding Fe^{2+}-rRNA interactions

In this study, we compare folding efficiency of LSU rRNA in the presence of Mg^{2+} or Fe^{2+}. Divalent cations are essential for ribosomal structure and function.[23,43] Crystal structures have been obtained from ribosomes purified from the thermophilic bacterium *T. thermophilus*,[75,223] the halophilic archaeon *H. marismortui*,[23] and the mesophilic bacterium *E. coli*.[224] These structures, and all other ribosome structures to date, indicate at least eight deeply-buried tightly-bound Mg^{2+} ions within Mg^{2+} micro-clusters[43] along with several hundred less-localized Mg^{2+} ions.[23,43] However, all existing ribosomal structures originate from cells grown, and ribosomes crystallized, under oxic conditions characteristic of our modern atmosphere, which is markedly different from the anoxic, pre-GOE environment. We have observed previously that the LSU rRNA (chapter 4) and a-rRNA[157] (chapter 3) experience Mg^{2+}-dependent structural changes to form LSU-like tertiary interactions. Here, we revisit ancient Earth biochemistry by recreating anoxic, micromolar to millimolar Fe^{2+} conditions (referred to here as "pre-GOE") in the laboratory. We test the ability of the ribosome to use Fe^{2+} in place of Mg^{2+}, exploring Fe^{2+} involvement in RNA-catalyzed protein synthesis and mediation of information transfer between RNA and protein. Previous evidence suggests Fe^{2+} incorporation into ribosomes in certain diseased eukaryotic cells.[225]

Particularly, we examine the effects of Mg^{2+} and Fe^{2+} on the structure of *T. thermophilus* LSU rRNA and its ancestral predecessor, a-rRNA (Figures 5.1 and 5.2). Mg^{2+} and Fe^{2+} have nearly-identical geometric and electronic properties, suiting them both as ideal interaction partners for RNA phosphate oxygens.[14,50] The origin of the ribosome pre-dates LUCA, placing the emergence of an ancestral ribosome at least a billion years prior to the GOE.[53,207] During this early stage of evolution, significant amounts of soluble Fe^{2+} existed in aqueous environments in concentrations estimated up to ~10 mM. In fact, life thrived for hundreds of millions of years in the absence of atmospheric or dissolved oxygen, with abundantly available aqueous Fe^{2+}.

Our primary motivation in this work is to discern if Fe^{2+} can mimic Mg^{2+} in interactions with rRNA. Research involving the interactions of ribosomal RNA with divalent cations, particularly Fe^{2+} and Mg^{2+}, will help us understand how early evolution achieved the transition from a predominantly-RNA World with small chemical or peptide co-factors,[1-4,162] to the "Central Dogma"-driven biology that pervades modern organisms.[89] We have chosen the ribosome as our specific research target because of its biological centrality[159] and unsurpassed architectural conservation.[177] The ribosome is crucial to all life on earth, and understanding its interactions with key structural partners such as divalent cations will provide insight into its origins and early development. a-rRNA represents a very early ribosomal ancestor, while the bacterial LSU rRNA represents a later, though still pre-GOE, ancestor of all modern ribosomes.[226] The bacterial LSU rRNA represents a common core that is found in the LSU rRNAs of all other species.[226] Since the earliest branching events of the tree of life took place well before the GOE, the evolution leading to this common core would have occurred in

presence of soluble Fe^{2+}. Therefore, ribosomes in modern microbes can be interrogated as fossil remnants from the first half of Earth's history[35] when ferrous iron was abundant and soluble prior to the GOE.[206]

In addition to information regarding the interactions of Fe^{2+} with rRNA, comparison of a-rRNA and LSU structural data provides us with valuable information about important RNA-RNA interactions in the LSU. We also utilize the wealth of footprinting data generated from the described experiments to interrogate the SHAPE patterns exhibited by non-canonical RNA helices, such as those found in D0.

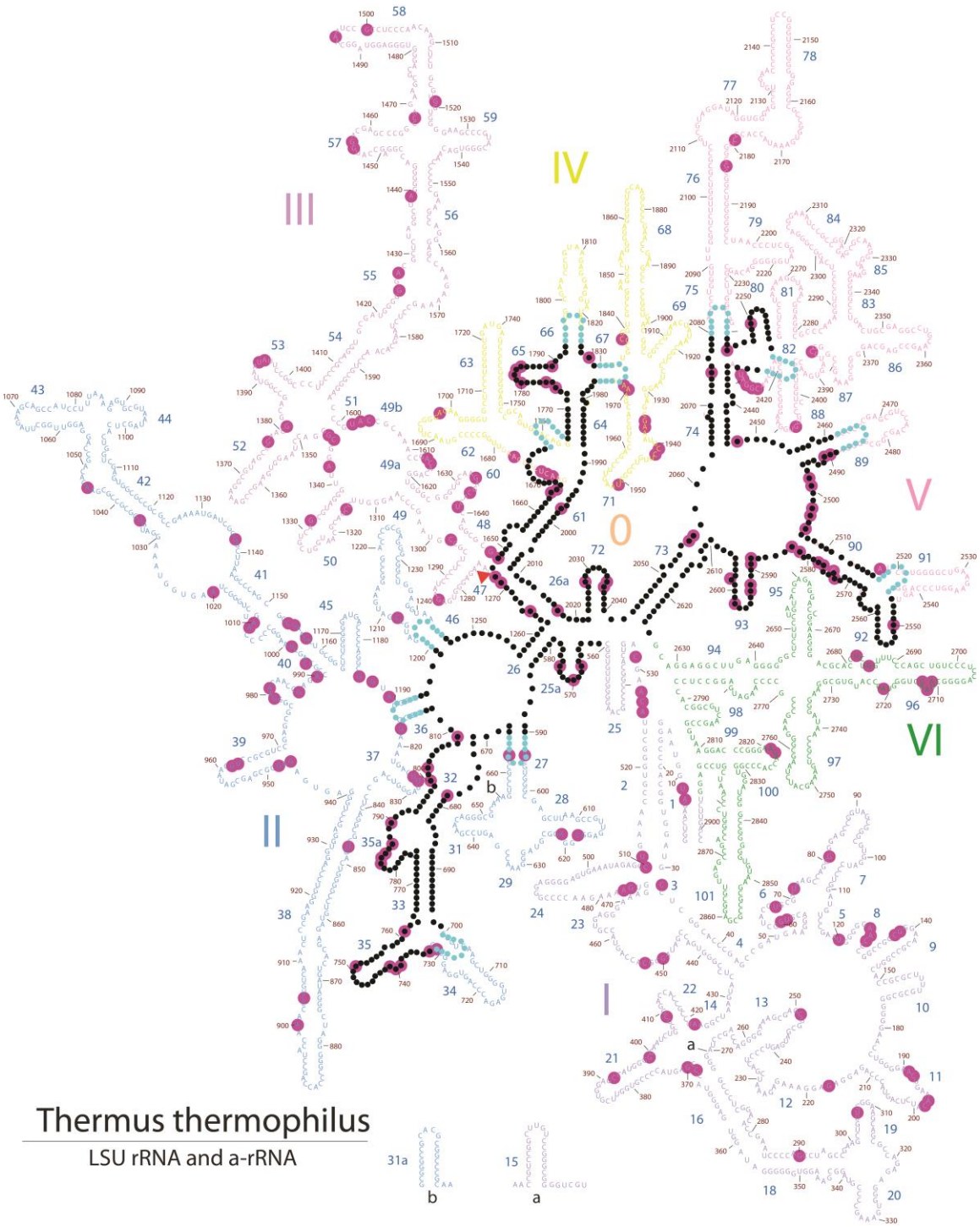


Figure 5.1. Comparison of a-rRNA and *T. thermophilus* 23S rRNA secondary structures. Ancient 23S sequences (black circles) and stitching tetraloops (cyan circles) comprise a-rRNA, a single 615 nt RNA polymer. A direct sealing of 23S sequence is indicated by a red arrow. 23S rRNA secondary structure is underlaid for comparison (letters colored by domain), with 23S helix numbers in blue, and nucleotide numbers indicated approximately every 10 nt (*E. coli* numbering). Magenta circles denote nucleotides observed to interact directly with Mg^{2+} atoms (first-shell interactions, 2.4 Å cut-off) in the *T. thermophilus* LSU crystal structure (PDB IDs: 2J00 and 2J01). Figure generated using RiboVision.

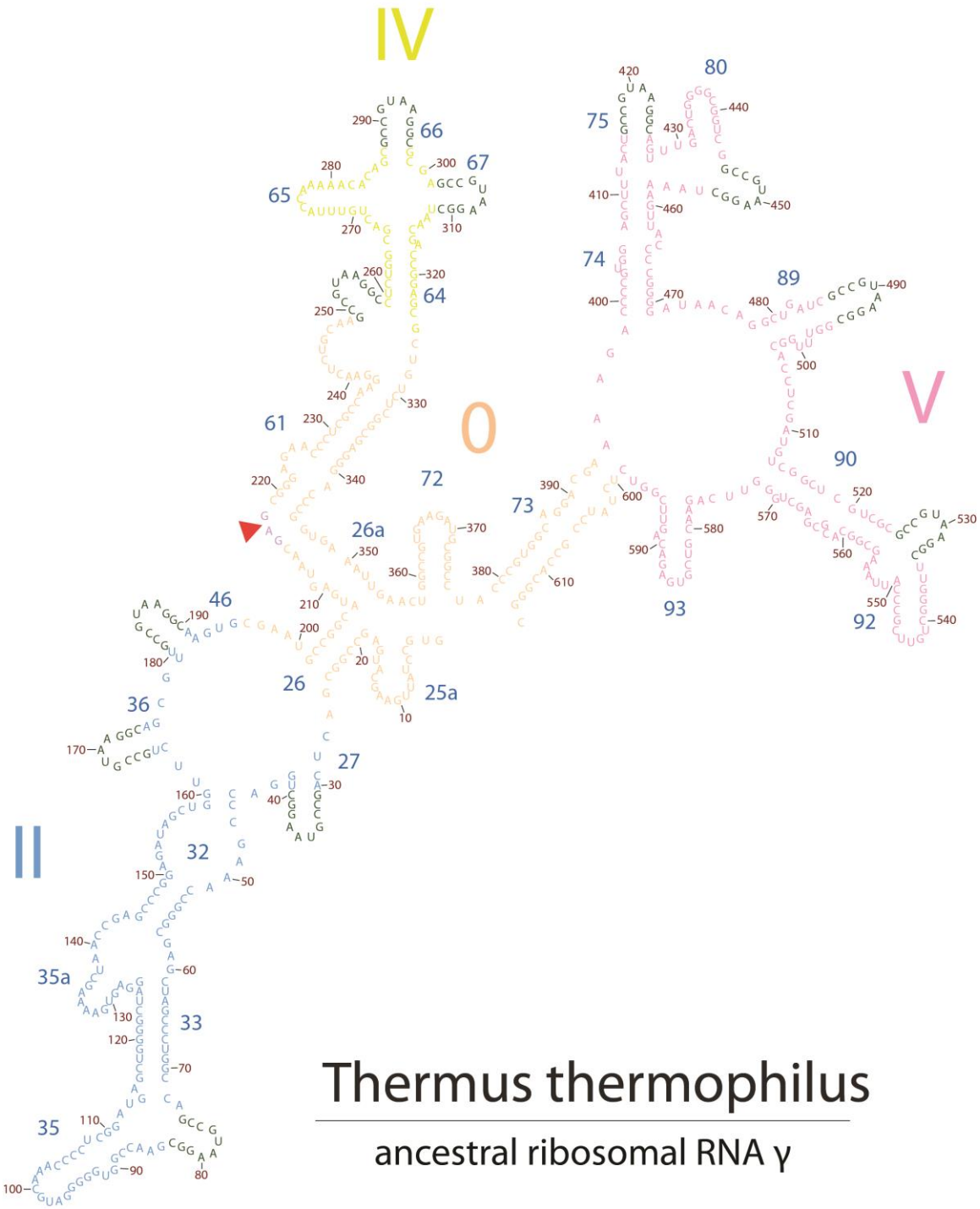


Figure 5.2. Detailed secondary structure of a-rRNA. This secondary structure is based on an updated 3D-based LSU rRNA secondary structure that introduces a new central domain, D0.[185] 23S-derived nucleotides are colored by domain, and stitching gccGUAAggc tetraloops are dark green. Nucleotide, domain, and helix numbers are indicated.

5.2 – Methods

T. thermophilus 23S rRNA and a-rRNA were *in vitro*-transcribed and purified using previously described constructs and protocols[118,156,157]. SHAPE experiments for a-rRNA were identical to those described in section 4.2 for LSU rRNA, including RNA concentration, with the sole exception of reverse transcription primers specific to a-rRNA (described in section 3.2.4). Additional samples were prepared with 2.5 mM $FeCl_2$ (final concentration in modification reaction) or 2.5 mM $MgCl_2$ (a-rRNA only), added before RNA annealing.

5.2.1 Anoxic techniques

Anoxic SHAPE methods were adapted from previously published protocols.[157] Generally, experiments were executed exactly as described in chapter 4. Adaptations were made to facilitate folding and NMIA modification of RNA in a vinyl anaerobic chamber (Coy Laboratory Products). Gas mix in the chamber was 95% argon/5% hydrogen, and the chamber maintained an O_2 level of 7-15 ppm. In order to keep RNA samples oxygen-free, samples containing RNA/buffer were lyophilized, transferred into anoxic chamber, and rehydrated with nuclease-free, degassed water inside the chamber. After RNA modification, divalent cations were removed from samples as described in section 4.2.2, after which the samples could safely be removed from the chamber without concern of iron-induced RNA degradation.

5.3 – Results

All SHAPE experiments described here were performed under pre-GOE conditions, with virtually no atmospheric O_2. Solutions were degassed with the atmospheric gas mix of the anoxic chamber to ensure absence of dissolved O_2. SHAPE data were collected for *T. thermophilus* LSU rRNA under three conditions: Na^+ (1 mM DCTA), Na^+/Mg^{2+} (10 mM $MgCl_2$), and Na^+/Fe^{2+} (2.5 mM $FeCl_2$). Lower [Fe^{2+}] was used compared to [Mg^{2+}] to limit iron-induced degradation of RNA due to trace Fe^{3+} present in $FeCl_2$ solutions and trace O_2 in the atmosphere of the chamber. Control experiments support that RNA is stable in Na^+/Fe^{2+} over the course of the experiment. When experiments were performed on a-rRNA in 2.5 mM $MgCl_2$ ($Na^+/low\ Mg^{2+}$) in order to generate analogous data for Fe^{2+} and Mg^{2+}, the induced changes were difficult to discern from background variations in SHAPE reactivity. Therefore additional data were collected at 10 mM Mg^{2+}, at which concentration noticeable changes in SHAPE reactivity are induced, as previously observed (chapter 3). All samples contained 200 mM NaOAc and 50 mM NaHEPES pH 8 (250 mM Na^+ total); relatively high Na^+ concentrations were used so that observed changes could be confidently attributed to the addition of divalent cations. In 250 mM Na^+, most if not all secondary structure should be present due to local electrostatic screening effects.[15,179] No a-rPeptides, rProteins, or other ribosomal components were present in any of the experiments expounded here. Where a-rRNA results are presented, the term is used exclusively in reference to the a-rRNA-γ version, which is the iteration described most extensively in chapter 3. The structure of a-rRNA portrayed in this chapter (Figures 5.1 and Figure 5.2) is based on the 3D-based

LSU rRNA secondary structure introduced in section 4.1.2, rather than the canonical secondary structure.

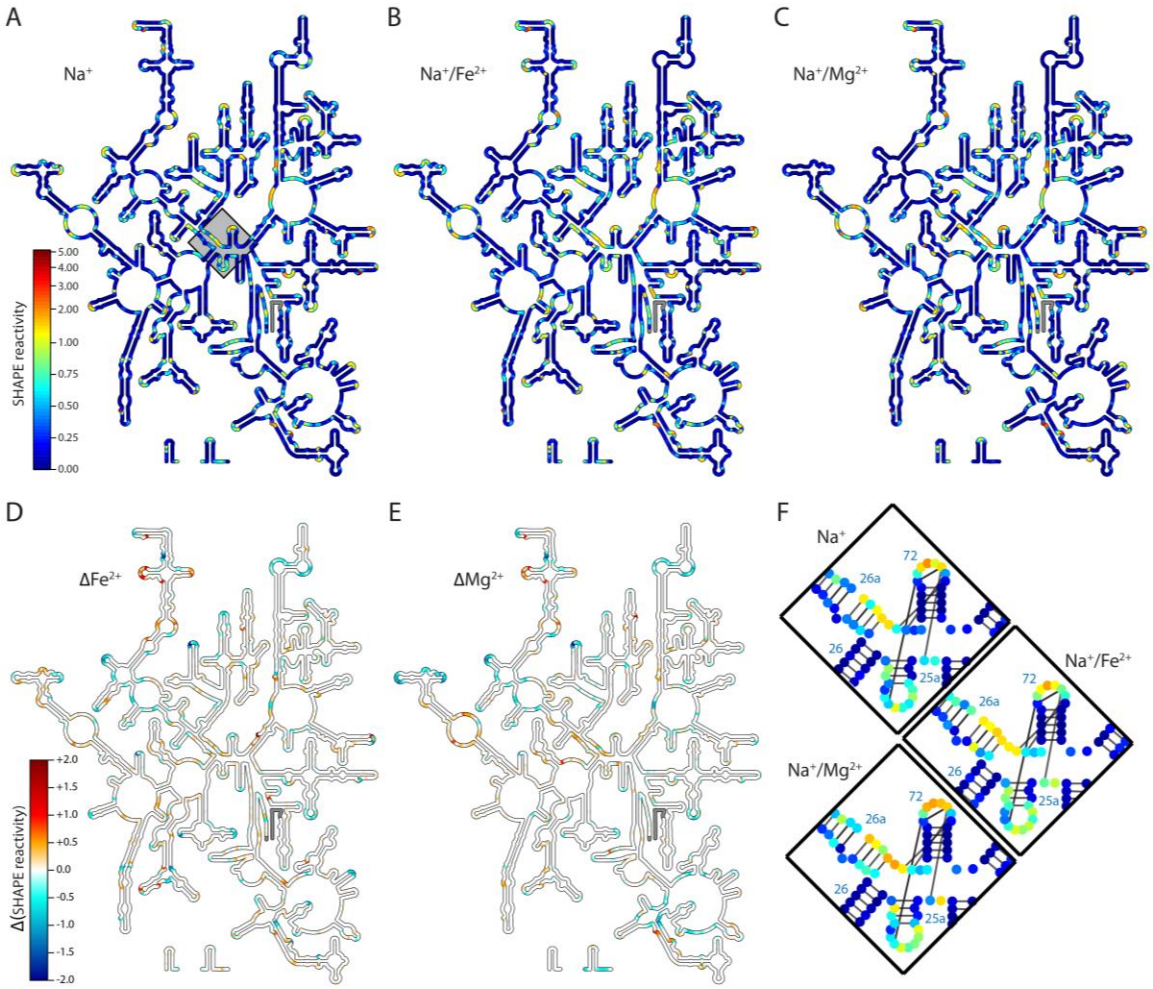


Figure 5.3. SHAPE reactivities and divalent-induced changes in SHAPE reactivity for the LSU rRNA. A-C) SHAPE reactivities mapped onto LSU rRNA secondary structure in A) Na^+, B) Na^+/Fe^{2+}, or C) Na^+/Mg^{2+}. All samples contained 200 mM NaOAc, 50 mM NaHEPES, pH 8, and data was collected under anoxic conditions. Legend in panel A applies also to panels B, C, and F. Box in panel A indicates region highlighted in panel F. D and E) Divalent cation-induced changes in SHAPE reactivity mapped onto LSU rRNA secondary structure, compared against SHAPE reactivities in Na^+. Positive values indicate nucleotides with increased SHAPE reactivity in presence of indicated divalent, while negative values denote decreased reactivity. Legend in panel D also applies to panel E. F) SHAPE reactivities for D0, including recently discovered non-canonical helices 25a and 26a, under specified conditions. Lines indicate nucleotide-nucleotide interactions observed between displayed nucleotides in the *T. thermophilus* LSU crystal structure (PDB 2J01). For insertion sites and nucleotide/helix numbering, see Figure 5.1. Regions where data are not available (5' and 3' ends) are grey. Line segment length in A-E is not proportionate to number of nucleotides due to inconsistent nucleotide spacing in the secondary structure. Larger, more detailed versions of panels A-E may be found in chapter 4 (Figures 4.3, 4.4, and 4.5) or Appendix C (Figures C.1, C.2, C.3, and C.4). Figures generated using RiboVision.

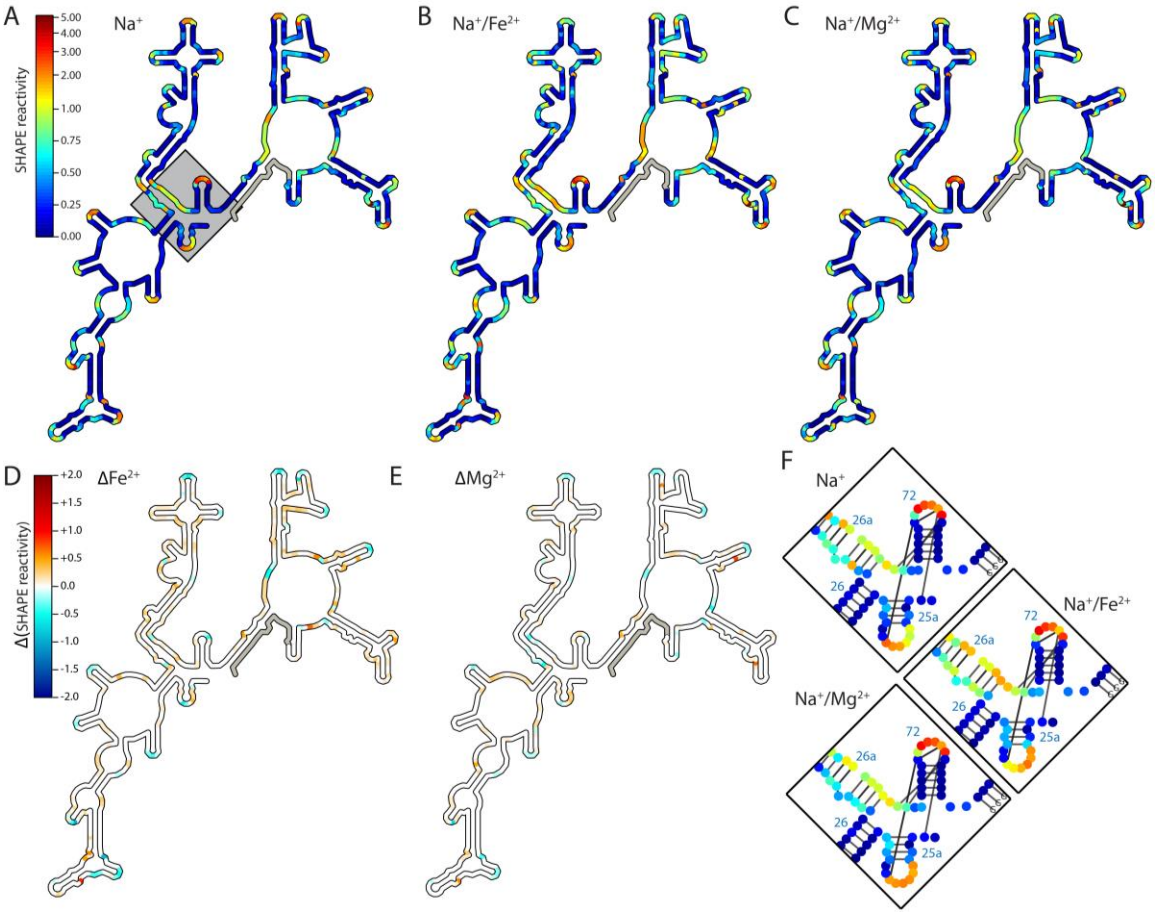


Figure 5.4. SHAPE reactivities and divalent-induced changes in SHAPE reactivity for a-rRNA. A-C) SHAPE reactivities mapped onto predicted a-rRNA secondary structure in A) Na^+, B) Na^+/Fe^{2+}, or C) Na^+/Mg^{2+}. All samples contained 200 mM NaOAc, 50 mM NaHEPES, pH 8, and data was collected under anoxic conditions. Legend in panel A applies also to panels B, C, and F. Box in panel A indicates region highlighted in panel F. D and E) Divalent-induced changes in SHAPE reactivity mapped onto predicted a-rRNA secondary structure, compared against samples with no divalent cations. Positive values indicate nucleotides with increased SHAPE reactivity in presence of indicated divalent, while negative values denote decreased reactivity. Legend in panel D also applies to panel E. F) SHAPE reactivities for D0, including recently discovered non-canonical helices 25a and 26a, under specified conditions. Lines indicate nucleotide-nucleotide interactions observed between analogous nucleotides in the *T. thermophilus* LSU crystal structure (PDB 2J01). Regions where data are not available (5' and 3' ends) are either grey (A-E) or shown as sequence only (F). For domain/nucleotide/helix numbering, see Figure 5.2. Line segment length in A-E is not proportionate to number of nucleotides due to inconsistent nucleotide spacing in the secondary structure. Figures generated using RiboVision.

5.3.1 LSU rRNA and a-rRNA maintain expected secondary structure with Fe^{2+}

Formation of rRNA helices is dependent only on monovalent cations, and is not influenced by divalent cations or rProteins. SHAPE data collected under pre-GOE conditions demonstrate excellent agreement with the expected secondary structures for the LSU rRNA and a-rRNA under all cation conditions studied. Figure 5.2A-C depicts SHAPE data for the LSU rRNA mapped onto the canonical secondary structure, and full numerical SHAPE data are available in Dataset 1. Broad agreement between the LSU rRNA secondary structure and SHAPE data obtained in Na^+ and Na^+/Mg^{2+} is described in-depth in section 4.3.2. This agreement is also observed in Na^+/Fe^{2+} (Figure 5B), in which virtually all nucleotides involved in local cWW base pairing exhibit low SHAPE reactivity under pre-GOE conditions; the same unreactive helices listed in Table 4.2 are also unreactive in Na^+/Fe^{2+}. Some bulges, mismatches, and loops are also unreactive to SHAPE, consistent with observations in Na^+ and Na^+/Mg^{2+}. Under all conditions tested, formation of virtually all canonical LSU rRNA helices is supported by SHAPE data. Recently-identified helices 25a and 26a (Figure 5.3F) do not yield uniformly low reactivity to SHAPE, despite the fact that they clearly form helices in 3D;[185] this phenomenon is further discussed below. SHAPE reactive nucleotides in Na^+/Fe^{2+} are generally restricted to nucleotides that are unpaired in the secondary structure (loops, bulges, mismatches, and inter-helical regions). Fe^{2+}, like Mg^{2+}, does not induce secondary structure rearrangement in the LSU rRNA, and therefore all Fe^{2+}-induced changes are expected to be related to tertiary interactions.

SHAPE data obtained for a-rRNA under pre-GOE conditions exhibits similar trends as LSU rRNA when compared with the model secondary structure, regardless of

cation environment. a-rRNA has been previously characterized by SHAPE with Na^+ in the presence and absence of Mg^{2+} (see chapter 3),[157] but these experiments were not conducted under pre-GOE conditions. Mapped SHAPE data for a-rRNA with Na^+, Na^+/Fe^{2+}, or Na^+/Mg^{2+} are displayed in Figure 5.3A-C. Virtually all nucleotides expected to participate in cWW helices exhibit low SHAPE reactivity in Na^+, consistent with constraint of these nucleotides by base pairing. These helices are maintained in Na^+/Fe^{2+} and Na^+/Mg^{2+}, consistent with persistent secondary structure stability in presence of divalent cations. As is observed in the LSU rRNA, helices 25a and 26a (Figure 5.4F) do not exhibit uniformly low reactivity to SHAPE, despite the fact that the analogous LSU rRNA forms helices in 3D;[185] this phenomenon is further discussed below. As is observed for the LSU rRNA, SHAPE reactive sites in a-rRNA are generally constrained to regions of the predicted secondary structure that do not participate in local cWW base pairing. Under all cation conditions studied, reactive sites are found in loops, bulges, and otherwise single-stranded or inter-helical regions of the predicted a-rRNA secondary structure. Data for a-rRNA in Na^+ and Na^+/Mg^{2+} are generally consistent with data collected separately under modern atmospheric conditions (see chapter 3). SHAPE data for a-rRNA and LSU rRNA are consistent with a model in which only Na^+ is necessary for local helix formation, and when Na^+ is present, divalent cations do not induce changes to rRNA secondary structure.

5.3.2 Fe^{2+} and Mg^{2+} induce qualitatively similar structural changes in LSU rRNA

Addition of Fe^{2+} or Mg^{2+} to LSU rRNA induces widespread structural changes. Similar to the process for Mg^{2+} reported in chapter 4, Fe^{2+}-induced SHAPE changes were

calculated by subtracting Na^+ data from Na^+/Fe^{2+} data for each nucleotide; the resulting data represent changes in SHAPE reactivity upon addition of Fe^{2+}, and are referred to as ΔFe^{2+} values (Dataset 1). To compare the effects of Fe^{2+} and Mg^{2+} on LSU rRNA structure, we generated LSU rRNA 'heat maps' (Figure 5.3D-E), in which ΔFe^{2+} or ΔMg^{2+} values are mapped onto the canonical LSU rRNA secondary structure. Important features of the ΔFe^{2+} and ΔMg^{2+} heat maps include:

i) Helices (white) do not change their base pairing states in response to addition of either Fe^{2+} or Mg^{2+}, confirming the expectation that in Na^+, divalent cations do not cause changes in rRNA secondary structure.[15,179] Helical regions of the LSU rRNA are invariant to addition of Fe^{2+} or Mg^{2+}.

ii) Loops and other unpaired nucleotides frequently change their extent of base pairing (their SHAPE reactivity) in response to addition of either Fe^{2+} or Mg^{2+}, confirming the expectation that divalent cations induce formation of tertiary structure between non-helical regions.

iii) The ΔFe^{2+} and ΔMg^{2+} heat maps are broadly similar, suggesting that Fe^{2+} and Mg^{2+} have similar effects on rRNA tertiary folding.

Overall, the heat maps indicate that Fe^{2+} and Mg^{2+} induce very similar effects on the structure of the LSU rRNA.

LSU rRNA forms a structure related in many ways to the rRNA-Mg^{2+} state (described at length in chapter 4) when folded under pre-GOE conditions (anoxic atmosphere, Fe^{2+}). In a majority of cases, this similarity extends to specific nucleotides which exhibit parallel changes in SHAPE reactivity upon addition of Fe^{2+} or Mg^{2+}. Most

regions exhibit similar divalent cation effects whether annealed with Fe^{2+} or Mg^{2+} (similar ΔFe^{2+} or ΔMg^{2+} patterns), in that the analogous positions/regions generally exhibit the same type of change in SHAPE reactivity (positive or negative), though these changes may vary in intensity. The magnitude of the SHAPE change often varies significantly, which is attributable to either the differing concentrations used for Fe^{2+} (2.5 mM) and Mg^{2+} (10 mM) or a propensity of certain divalent binding sites to preferentially bind either Mg^{2+} or Fe^{2+} based on slight structural perturbations. To capture the similarity between Fe^{2+}- and Mg^{2+}-responsiveness of specific nucleotides, we have binned nucleotides into three categories based on their general response to each divalent: i) increased SHAPE reactivity, ii) decreased SHAPE reactivity, or iii) little to no change in SHAPE reactivity. This method allows us to interrogate structural changes induced by Mg^{2+} or Fe^{2+}, while largely ignoring the magnitude of the change. In this analysis, we only consider nucleotides as exhibiting a significant change in SHAPE reactivity if either the original reactivity (in Na^{+}) or final reactivity (in Na^{+}/Fe^{2+} or Na^{+}/Fe^{2+}) is above 0.3 SHAPE units, so as not to artificially attribute significance to small overall changes in SHAPE reactivity. Based on this analysis, >86% of LSU rRNA nucleotides exhibit the same general response to Mg^{2+} as to Fe^{2+} (all values and calculations can be found in Dataset 1). If we focus instead on the sites that experience the largest overall changes in Mg^{2+}, 43 of the 50 nucleotides that exhibit the largest Mg^{2+} response (in terms of % change) exhibit the same general type of response to Fe^{2+} (86%). When that threshold is lowered to include all LSU rRNA nucleotides that experience a Mg^{2+}-induced SHAPE change of >33% (441 nucleotides total), >75% of these sites demonstrate the same type of response to Fe^{2+}. These statistics suggest a global similarity between the rRNA-Mg^{2+}

state postulated in chapter 4 and the rRNA-Fe^{2+} state. Possible discrepancies between the rRNA-Mg^{2+} and rRNA-Fe^{2+} states and the magnitude of response between Fe^{2+} and Mg^{2+} are discussed further below.

Certain LSU rRNA domains exhibit ΔFe^{2+} and ΔMg^{2+} patterns that are qualitatively extremely similar (Figure 5.3D-E). DIII in particular exhibits extraordinary overall agreement in its response to Fe^{2+} and Mg^{2+}. DIV also demonstrates strikingly similar ΔFe^{2+} and ΔMg^{2+} locations, with the exception of minor deviations in H68. DVI is broadly similar in its response to Fe^{2+} and Mg^{2+}, although certain helices bear nucleotides which exhibit greater response to Fe^{2+} (H94, H96 and H99). D0 displays similar responsivity to Fe^{2+} or Mg^{2+}, though a bulge in H73 exhibits the largest Fe^{2+}-induced increase in SHAPE reactivity of the entire LSU rRNA while Mg^{2+} has almost no effect at this position. Virtually all ΔMg^{2+} sites observed in DV are also responsive to Fe^{2+}, though this domain bears significantly more ΔFe^{2+} sites overall, including regions in H91 and H78 that are responsive only to Fe^{2+}. A similar trend is observed in DII, in which many positions in H28-39 are more responsive to Fe^{2+} than Mg^{2+}. H43-44 in DII deviate significantly in their Fe^{2+} and Mg^{2+} responses. DI appears to be more responsive overall to Mg^{2+} than to Fe^{2+}, particularly in the inter-helical regions near H4, H12, H14, and H15, all of which exhibit decreased in flexibility upon addition of Mg^{2+} but little to no change in response to Fe^{2+}. However, other regions of DI (ie. H16-18) exhibit similar ΔFe^{2+} and ΔMg^{2+} patterns. A quantitative comparison of the overall Fe^{2+} and Mg^{2+} response of each LSU rRNA domain is provided in Table 5.2. Overall, all seven LSU rRNA domains exhibit similar ΔFe^{2+} and ΔMg^{2+} patterns to some extent, though Fe^{2+} is seen to induce structural changes at additional positions, and also induces changes of

greater magnitude in numerous cases. This observation is particularly striking given that [Fe^{2+}] was 4-fold lower than [Mg^{2+}] in the experiments being compared.

5.3.3 Fe^{2+} and Mg^{2+} induce qualitatively similar structural changes in a-rRNA

Addition of Fe^{2+} or Mg^{2+} to the LSU rRNA's model predecessor, a-rRNA, also results in structural changes dispersed throughout primary and secondary structure. ΔFe^{2+} and ΔMg^{2+} values were calculated (Dataset 1), and heat maps for a-rRNA were generated based on the predicted secondary structure (Figure 5.4D-E). The range of ΔFe^{2+} and ΔMg^{2+} values observed in a-rRNA is roughly half that of the LSU rRNA, but extensive changes are still detected. The same general features are observed in a-rRNA as in the LSU rRNA: i) cWW helices are invariant to addition of Fe^{2+} or Mg^{2+}, ii) unpaired nucleotides in loops, bulges, and inter-helical regions frequently experience structural changes in response to Fe^{2+} or Mg^{2+}, indicating formation of tertiary structures involving non-helical regions, and iii) ΔFe^{2+} and ΔMg^{2+} heat maps are in general agreement, suggesting that Fe^{2+} and Mg^{2+} induce similar tertiary interactions. The loop nucleotides of the gccGUAAggc stem-loops used to stitch together a-rRNA frequently exhibit reduced reactivity in response to Fe^{2+} and Mg^{2+}, consistent with added stabilization of the loop nucleotides. These observations are consistent with a largely stable a-rRNA secondary structure that is insensitive to divalent cations, and global collapse into a native-like state upon addition of Fe^{2+} or Mg^{2+} through induction of RNA-RNA tertiary interactions.

Qualitatively, Fe^{2+} induces changes at virtually all Mg^{2+}-responsive positions. The same binning scheme applied to LSU rRNA data was used to classify a-rRNA nucleotides as exhibiting significant SHAPE reactivity increases, decreases, or little to no

response upon addition of Fe^{2+} or Mg^{2+}. Over 84% of all measured a-rRNA nucleotides were found to experience the same overall type of response to Fe^{2+} and Mg^{2+}. Of the 50 nucleotides exhibiting the largest degree of Mg^{2+}-dependent change, 44 (88%) exhibit the same type of change (increase or decrease in SHAPE reactivity) upon addition of Fe^{2+}. When all nucleotides that exhibit a 10% or greater SHAPE response to Mg^{2+} (161 nucleotides) are considered, 84.5% of these nucleotides are similarly responsive to Fe^{2+}. The cut-off value used for significant divalent-induced changes is lower here than for the LSU rRNA (33%) due to the reduced range of ΔFe^{2+} and ΔMg^{2+} values observed for a-rRNA.

The overall similarity of a-rRNA's response to Mg^{2+} and Fe^{2+} is underscored by comparison of individual domains. a-rRNA nucleotides are partitioned into domains according to the domain to which the analogous LSU rRNA nucleotide belongs. Stitching tetraloops are considered to be part of the same a-rRNA domain as the neighboring 23S-derived RNA. The ΔFe^{2+} and ΔMg^{2+} patterns of DII are almost identical, though intensities vary; several nucleotides of H35 and H35a are more responsive to Fe^{2+}, while a single nucleotide in H32 exhibits greater response to Mg^{2+}. The Mg^{2+} and Fe^{2+} responses of both D0 and DIV are almost identical. In DV, which contains the proposed functional center of a-rRNA, the PTC, most nucleotides that exhibit ΔMg^{2+} values also exhibit similar values in Fe^{2+}, though the overall response to Fe^{2+} is considerably stronger, suggesting increased folding competency of the PTC in Fe^{2+}. a-rRNA does not contain rRNA derived from DI, DIII, or DVI. Overall average SHAPE change for each a-rRNA domain are calculated in Table 5.2.

Table 5.2. Mean overall SHAPE change of LSU and a-rRNA induced by Fe^{2+} vs. Mg^{2+}

	Mean change$_{LSU}$[a]		Mean change$_{a\text{-}rRNA}$[a,b]	
Domain	2.5 mM Fe^{2+}	10 mM Mg^{2+}	2.5 mM Fe^{2+}	10 mM Mg^{2+}
0	0.123	0.100	0.105	0.076
I	0.094	0.096	--	--
II	0.107	0.084	0.107	0.074
III	0.152	0.139	--	--
IV	0.105	0.085	0.127	0.075
V	0.096	0.080	0.130	0.063
VI	0.080	0.065	--	--
Overall	0.107	0.092	0.116	0.071

[a]Average of $|\Delta Fe^{2+}|$ or $|\Delta Mg^{2+}|$ values of all nucleotides in each domain or overall. All values are in normalized SHAPE reactivity units.
[b]a-rRNA does not contain RNA from LSU DI, DIII, or DVI. Nucleotides of gccGUAAggc stitching loops were assigned to the domain of adjacent rRNA.

5.3.4 Fe^{2+} induces larger, more numerous changes to rRNA structure than Mg^{2+}

As alluded to in the two previous sections, the LSU rRNA and a-rRNA are more responsive overall to Fe^{2+} than to Mg^{2+}, both in terms of number of responsive sites and overall average induced structural change. Multiple metrics were used to measure responsiveness to each cation. Table 5.2 presents mean overall response and response of individual domains to Fe^{2+} and Mg^{2+}, calculated using absolute values of ΔFe^{2+} and ΔMg^{2+} values. Overall, a-rRNA exhibits a 63% greater change in overall SHAPE reactivity upon addition of Fe^{2+} vs. Mg^{2+}. Every individual domain of a-rRNA exhibits increased responsiveness to Fe^{2+} over Mg^{2+}, ranging from 38% (D0) to 106% (DV). These observations are consistent with the qualitative results reported in section 5.3.3. The LSU rRNA also exhibits higher mean structural change with Fe^{2+} than with Mg^{2+}, though to a somewhat lesser extent than observed for a-rRNA; global Fe^{2+} response is 16% higher than the response to Mg^{2+}. Six of seven LSU domains exhibit greater Fe^{2+}

response, ranging from 8% (DIII) to 27% (DII) when compared to the overall effect of Mg^{2+} on each domain. DI exhibits a similar overall response to Fe^{2+} and Mg^{2+}, though there is some variation in the exact locations effected by each cation (see section 5.3.2).

Table 5.3. Pairwise comparison of SHAPE data by S-factor

rRNA system		$SHAPE_{init}$	$SHAPE_{final}$	S-factor
LSU rRNA		Na^+	Na^+/Fe^{2+}	0.412
		Na^+	Na^+/Mg^{2+}	0.356
		Na^+/Fe^{2+}	Na^+/Mg^{2+}	0.356
	Null values	Na^+	Na^+_{random}	1.285[a]
				1.285[a]
		Na^+/Fe^{2+}	Na^+/Fe^{2+}_{random}	1.221[a]
				1.210[a]
		Na^+/Mg^{2+}	Na^+/Mg^{2+}_{random}	1.325[a]
				1.291[a]
a-rRNA		Na^+	Na^+/Fe^{2+}	0.371
		Na^+	$Na^+/low\ Mg^{2+}$	0.188
		Na^+	Na^+/Mg^{2+}	0.238
		Na^+/Fe^{2+}	Na^+/Mg^{2+}	0.243
	Null values	Na^+	Na^+_{random}	1.218
		Na^+/Fe^{2+}	Na^+/Fe^{2+}_{random}	1.058
		$Na^+/low\ Mg^{2+}$	$Na^+/low\ Mg^{2+}_{random}$	1.098
		Na^+/Mg^{2+}	Na^+/Mg^{2+}_{random}	1.120
LSU rRNA vs. a-rRNA[b]		$Na^+_{LSU\ rRNA}$	$Na^+_{a\text{-}rRNA}$	0.640
		$Na^+/Fe^{2+}_{LSU\ rRNA}$	$Na^+/Fe^{2+}_{a\text{-}rRNA}$	0.568
		$Na^+/Mg^{2+}_{LSU\ rRNA}$	$Na^+/Mg^{2+}_{a\text{-}rRNA}$	0.596

[a]Data were compared against a randomized version of the $SHAPE_{init}$ data set, shuffled in two different configurations.
[b]Analogous positions of the LSU rRNA and a-rRNA were compared. Only LSU-derived nucleotides of a-rRNA were used in this analysis (not stitching tetraloops).

Pairwise comparison of SHAPE values also supports enhanced structural changes under pre-GOE conditions (ie. Fe^{2+}). We have developed a discrepancy index metric, which we term "S-factor", to compare the overall change based on induced SHAPE reactivity changes at individual positions, similar to the method used to compare goodness of fit between model structure and electron density in solution of x-ray crystal structures:

$$S = \frac{\sum |SHAPE_{init} - SHAPE_{final}|}{\sum |SHAPE_{init}|}$$

where $SHAPE_{init}$ is the initial SHAPE value used for comparison, and $SHAPE_{final}$ is the SHAPE value of the same or analogous position in the compared data set. Calculated S-factors are provided in Table 5.3. S-factors calculated for a-rRNA are consistent with greater structural change in response to Fe^{2+}; the value obtained when Na^{+}/Fe^{2+} and Na^{+} structures are compared is significantly higher than S-factors generated by comparison with either of the Mg^{2+} data sets (Table 5.3). a-rRNA is significantly more responsive to 2.5 mM Fe^{2+} than 2.5 mM Mg^{2+}. Comparison of the S-factors under Na^{+}/Fe^{2+} and $Na^{+}/low\ Mg^{2+}$, in equal amounts of the respective divalent cations (2.5 mM), suggest that a-rRNA experiences much lower structural change in response to Mg^{2+} than Fe^{2+} at comparable divalent cation concentrations. This observation is supported by qualitative observations of considerably lower magnitude and number of SHAPE changes induced by 2.5 mM Mg^{2+} (Appendix C, Figure C.5). The S-factor obtained for Na^{+}/Mg^{2+} is higher than that reported for $Na^{+}/low\ Mg^{2+}$, consistent with expectations of greater overall increases in tertiary structure formation with increasing [Mg^{2+}]. LSU rRNA exhibits larger overall S-factors compared to a-rRNA, though trends are similar; Fe^{2+} exhibits a greater degree of structural change than Mg^{2+}. The higher level of Fe^{2+} responsivity at specific positions observed for both rRNA systems by S-factor comparison is even more striking considering that 2.5 mM Fe^{2+} is inducing larger structural changes than 4 times more Mg^{2+} (10 mM). Comparison of the Na^{+}/Fe^{2+} and Na^{+}/Mg^{2+} structures of both a-rRNA and the LSU rRNA using S-factors indicates that the structures are broadly similar, but do experience some variation, consistent with observations of varying magnitude of response to Fe^{2+} or Mg^{2+} at specific positions. Null values for S-factors were generated to

better understand the significance of the values obtained when SHAPE data for each rRNA system were compared under varied conditions. Data were compared with randomized sets of the same data to produce null S-factor values for a-rRNA (1.058-1.218) and LSU rRNA (1.210-1.325). These null values represent the level at which there is no statistical pairwise similarity between the compared data sets. All values obtained for a-rRNA and LSU rRNA are significantly lower than the corresponding null values, suggesting that the compared structures are broadly similar, with some local variation.

Increased sensitivity of a-rRNA and LSU rRNA to Fe^{2+} manifests both in higher incidence of Fe^{2+}-responsive nucleotides and in greater magnitude of Fe^{2+}-induced structural interactions at individual positions. In a-rRNA, 43 nucleotides exhibit $|\Delta Fe^{2+}|>0.3$, while only 17 nucleotides demonstrate a similar extent of change in response to 10 mM Mg^{2+}. This trend is also observed in LSU rRNA, though to a lesser extent; 163 nucleotides exhibit a $|\Delta Fe^{2+}|>0.4$, while only 126 sites experience the same degree or greater SHAPE change upon addition of Mg^{2+}. A slightly higher cut-off value was used for the LSU rRNA due to the larger overall range of ΔFe^{2+} and ΔMg^{2+} values. Regardless of the applied minimum change value used, Fe^{2+} consistently induces structural changes at more positions than Mg^{2+}, despite 4-fold lower [Fe^{2+}]. Qualitatively and quantitatively, the LSU rRNA and a-rRNA are more responsive to Fe^{2+} than to Mg^{2+}.

5.3.5 Non-canonical RNA helices exhibit anomalous SHAPE patterns

Helices composed of non-canonical base pairs yield unexpected SHAPE reactivities. Recently-identified helices of D0, H25a and H26a, are composed of 50% or

more non-canonical base pairs (ie. non-cWW pairs), though they appear generally helical when inspected in 3D.[185] When SHAPE data obtained in each of the studied cation environments is mapped onto the secondary structure of these helices (Figures 5.3F and 5.4F), reactivity patterns are not consistent with a model of SHAPE data interpretation in which base-paired nucleotides exhibit low reactivity. In both a-rRNA and the LSU rRNA, the 3' strand of H26a yields moderate to high reactivities. In both rRNA systems, the 5' strand of helices 25a and 26a exhibit lower reactivity overall; this discrepancy is more pronounced in H26a. These data suggest that the helical nucleotides of the 5' strand are less flexible, or otherwise constrained in a manner that makes them largely unreactive to the SHAPE modifying agent, possibly by stacking interactions. Normally, low reactivity would be interpreted as base pairing of these nucleotides, but the higher reactivity of the 3' strands of the same helices introduces an added layer of complexity. The larger overall difference in reactivity observed for 26a may be related to the occurrence of 5 consecutive non-cWW pairs; non-canonical base pairs in H25a are flanked by cWW pairs, which could contribute to lower overall reactivity of this stem. Helical nucleotides of H25a are largely insensitive to Fe^{2+} or Mg^{2+} in the context of a-rRNA (Figure 5.4D-E), while certain positions in H26a exhibit low to moderate divalent dependencies. Divalent cation dependencies of H25a and H26a are marginally larger in the LSU rRNA (Figure 5.3D-E) than in a-rRNA (5.4D-E). Overall, helical regions of D0 do not exhibit significant structural transitions in response to Fe^{2+} or Mg^{2+}.

5.4 – Discussion

5.4.1 Fe^{2+} mimics for Mg^{2+} in LSU rRNA and a-rRNA interactions

RNA folds and catalyzes chemical transformations in the presence of both Fe^{2+} and Mg^{2+} due to profound similarities in their coordination chemistries and interactions with water, phosphate and other moieties (Table 5.1).[50,51] Qualitative and quantitative evidence support Fe^{2+} mimicry in rRNA interactions that involve Mg^{2+} in the modern ribosome, including broad similarity of ΔFe^{2+} and ΔMg^{2+} heat maps and high correlations between observed ΔMg^{2+} and ΔFe^{2+} sites observed in binning analyses. Similarity of Fe^{2+} and Mg^{2+} effects suggest induction of a generally comparable network of tertiary interactions by both cations, and collapse of both the LSU rRNA and a-rRNA to near-native states in absence of any other ribosomal components, similar to the rRNA-Mg^{2+} state described in chapter 4. As in the rRNA-Mg^{2+} state, many ΔFe^{2+} sites can be confidently attributed to structural changes involving specific tertiary interactions. Fe^{2+} substitution for Mg^{2+} in RNA-cation interactions has been observed previously by SHAPE in the small P4-P6 RNA.[50] LSU rRNA and a-rRNA have both been shown previously to interact with Fe^{2+} in electron transfer catalysis experiments.[51] Mimicry of Fe^{2+} for Mg^{2+} is plausible based on similarities between physical and chemical properties of the two cations (Table 5.1), and comparable energetics observed in modeling experiments performed at high levels of theory.[50]

Divalent cation-induced SHAPE changes observed in the experiments presented here can be attributed specifically to addition of Fe^{2+} or Mg^{2+}. If the observed effects were attributable to non-specific electrostatic screening, the structures that cause changes

in reactivity would be induced by the relatively high [Na^+],[179] and addition of divalent cations would not induce further changes.

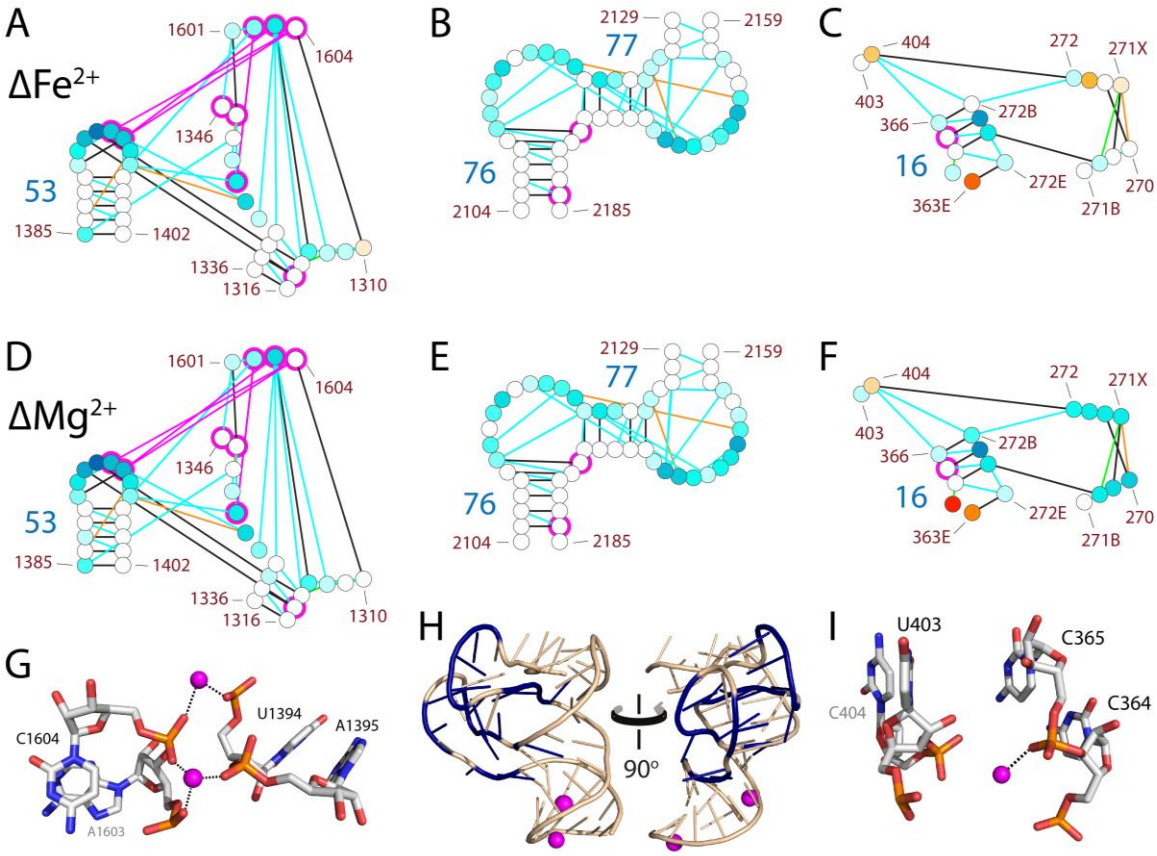


Figure 5.5. Fe^{2+} and Mg^{2+} effects on specific LSU rRNA structures. A-F) 2D structures of selected regions of the LSU rRNA and regions of RNA with which they interact directly: H53 (A & D), the uL1 protuberance (B & E), and H16 (C & F). Nucleotides are colored by ΔFe^{2+} (A-C) or ΔMg^{2+} (D-F) values as in Figure 5.3. Helix and nucleotide numbers are indicated. Lines represent direct and Mg^{2+}-mediated RNA-RNA interactions observed in the *T. thermophilus* ribosome crystal structure, as determined by FR3D (magenta, RNA-Mg^{2+}-RNA; black, base-base; orange, base-phosphate; blue, base-stacking; and green, base-sugar). Underlaid magenta circles indicate nucleotides observed to participate in first-shell interactions with Mg^{2+} ions (2.4 Å cut-off). G) Direct coordination of Mg^{2+} by multiple phosphate oxygens from U1394, A1395, A1603, and C1604 in H53. Atoms are colored as follows: Mg^{2+}, magenta; C, grey; N, blue; O, red; and P, orange. Black dashed lines indicate first-shell interactions between Mg^{2+} and RNA ligands. H) Cartoon representation of the same uL1 protuberance nucleotides depicted in panels B and E. Mg^{2+} in first-shell interaction with RNA (nucleotides C2179 and G2184) is shown as magenta spheres. Nucleotides are colored by $\Delta Fe^{2+}/\Delta Mg^{2+}$ values (dark blue, decreased reactivity; wheat, little to no change in reactivity). I) Direct and water-mediated binding of Mg^{2+} to nucleotides C364, C365, U403, and C404. Atoms and interactions illustrated as in panel G. Water-mediated interactions between rRNA and Mg^{2+} are not explicitly depicted. 2D structures are generated with RiboVision. 3D structures adapted from the *T. thermophilus* ribosome crystal structure (PDB IDs: 2J00 and 2J01), generated using PyMol.

Fe^{2+} mimics for Mg^{2+} in LSU rRNA via multiple binding modes. The binding states of cations associated with nucleic acids are most appropriately considered as a continuum which takes into account overall number of first-shell RNA or DNA ligands.[14] According to this continuum, cations are considered to be tightly bound (multiple first-shell RNA ligands), loosely bound (single first-shell ligand), glassy (interactions with hydrated cations), or diffuse (no specific nucleic acid association). Figure 5.5 depicts secondary and 3D structures for selected regions of LSU rRNA where Fe^{2+} mimicry is readily apparent in several distinct cation binding modes. As is described in chapter 4 for Mg^{2+}, Fe^{2+}-induced changes in SHAPE reactivity are attributable to known tertiary interactions (Figure 5.5A-C). The loop region of H53 is observed to form several intra-domain interactions with other parts of DIII. ΔFe^{2+} and ΔMg^{2+} patterns of H53 and nearby RNA are identical (Figures 5.5A and 5.5D), supporting Fe^{2+} mimicry in this particular tertiary interaction network. Two Mg^{2+} cations directly coordinate PO atoms of C1605, A1603, U1394, and A1395 in the assembled LSU (Figure 5.5G). These tightly-bound Mg^{2+} ions participate in a Mg^{2+} microcluster (Mg^{2+}-μc) structure and serve as bridges between two regions of DIII.[36,43] RNA-Mg^{2+}-RNA interactions of these Mg^{2+}-μc's are depicted on the secondary structure in Figures 5.5A and 5.5D (magenta lines), and occur at or near the sites bearing the largest magnitude ΔFe^{2+} and ΔMg^{2+} values observed in the depicted regions. Fe^{2+} substitutes directly for these two tightly-bound Mg^{2+} ions, inducing identical structural changes. Previous modeling experiments in which Fe^{2+} is substituted for Mg^{2+} in Mg^{2+} clamp structures (which are found in Mg^{2+}-μc's) support plausibility of this replacement.[50] The network of divalent cation-induced tertiary interactions depicted in Figures 5.5A and 5.5D is largely responsible for the

structural collapse of DIII into its native state. Mg^{2+}-μc's are considered to be an ancient type of RNA-cation interaction due to their high incidence in the oldest regions of LSU rRNA.[43] Substitution of a plausible cofactor of ancient RNA such as Fe^{2+} in these sites is consistent with Mg^{2+}-μc's as a molecular fossil of earlier Fe^{2+}-μc structures.

Fe^{2+} also mimics for Mg^{2+} in LSU rRNA interactions involving glassy cations. The uL1 protuberance folds independently, and protrudes from the exterior of the assembled LSU.[75,191] A detailed account of the effect of Mg^{2+} on the uL1 protuberance and the tertiary interaction network observed in this region of the assembled LSU can be found in section 4.4.2, Figure 4.7, and Table 4.4. The Fe^{2+} response of the uL1 protuberance is indistinguishable from that observed with Mg^{2+} (Figures 5.5B and 5.5E); all nucleotides that are reactive to SHAPE in the absence of divalent cations become completely unreactive in Na^{+}/Fe^{2+} (Figure 5.3). While two Mg^{2+} cations are observed to interact through first shell-interactions with H76 in the assembled LSU, these Mg^{2+}-interacting nucleotides are structurally distant from the nucleotides that exhibit divalent cation-induced structural changes. Two Mg^{2+} cations are observed near the loops that interact in the folded uL1 protuberance; these Mg^{2+} ions are at a distance consistent with water-mediated contact with loop nucleotides. The network of tertiary interactions observed between these loops is expected to experience significant stabilization by these hydrated Mg^{2+} ions. Substitution of Fe^{2+} at these positions suggests that rRNA structures are similarly responsive to binding of hydrated Fe^{2+}. Mg^{2+} and Fe^{2+} both promote formation of a native-like uL1 protuberance structure, similar to the rRNA structure observed in the assembled LSU, even in the absence of rProtein uL1.

Fe^{2+} substitutes for Mg^{2+} in loosely bound cation interactions with rRNA. H16 is found in DI of the LSU rRNA. Mg^{2+}-induced stabilization of the 5' end of H16 is described in section 4.3.4. The ΔFe^{2+} pattern observed in the 5' region of H16 is very similar to the ΔMg^{2+} pattern (Figures 5.5C and 5.5F); nucleotides of the 5' strand (272B-272E) exhibit significant decreases in SHAPE reactivity in response to both cations. A loosely bound Mg^{2+} cation is observed in first-shell contact with the phosphate group of C365 (Figure 5.5I), which is paired with G272C, and this same Mg^{2+} also participates in water-mediated interactions with C404. Stabilization of this short helical region is attributed here to association of either loosely-bound Fe^{2+} or Mg^{2+} at this position.

Substitution of Fe^{2+} for Mg^{2+} in LSU rRNA-cation interactions is observed in interactions involving H53, the uL1 protuberance, and H16, via multiple classes of cation-binding (Figure 5.5). The broad similarity of ΔFe^{2+} and ΔMg^{2+} heat maps (Figures 5.3 and 5.4) indicates that Fe^{2+} substitutes for Mg^{2+} in a majority of RNA-cation interactions detected by comparative SHAPE. The extent of formation for certain tertiary interactions may vary, as indicated by variations in ΔFe^{2+} and ΔMg^{2+} values. Affinity for the two cations may vary depending on the coordination geometry and chemical environment of specific binding sites.

5.4.2 rRNA folds more readily in response to Fe^{2+}

Small differences in the electronic structure of Fe^{2+} and Mg^{2+} lead to increased reactivity of Fe^{2+}-RNA versus Mg^{2+}-RNA complexes, as demonstrated by previous results suggesting that the catalytic potential of RNA may be greater than generally understood under pre-GOE conditions.[50,51] However, structural experiments with P4-P6

did not demonstrate significantly different response to Fe^{2+} than Mg^{2+}. Modeling experiments show a modest increase in the interaction energy of RNA-Fe^{2+} complexes compared to RNA-Mg^{2+} complexes in Mg^{2+} clamp structures (1.3 kcal/mol, Table 5.1). Multiple Mg^{2+} clamps are found in LSU rRNA, often in the context of Mg^{2+}-μc's, which are responsible for forming the overall architecture of the PTC.[43] SHAPE data obtained here for rRNA systems suggests greater structural response to Fe^{2+} than Mg^{2+}, both in terms of magnitude of SHAPE change at common sites and an overall increase in Fe^{2+}-responsive sites, even at 4-fold lower [Fe^{2+}]. These results suggest that rRNA folding is more robust with Fe^{2+} than Mg^{2+}, even at 4-fold higher [Mg^{2+}]. When a-rRNA is folded in equal concentrations of Fe^{2+} or Mg^{2+}, the enhanced effect of Fe^{2+} on SHAPE reactivity is more pronounced (Table 5.3); 2.5 mM Mg^{2+} induces only minor structural effects, while 2.5 mM Fe^{2+} causes a considerably higher degree of overall change, both at unique and Mg^{2+}-responsive positions. This suggests that, at low [Mg^{2+}], many native-like a-rRNA tertiary interactions are either partially-formed, unstable, or wholly absent. Data for both a-rRNA and the LSU rRNA are consistent with a lower required ion concentration for folding with Fe^{2+} (2.5 mM) than for Mg^{2+} (10 mM). Larger degree of changes observed upon addition of Fe^{2+} vs. Mg^{2+} indicate either a more significant change to the structure, or a larger proportion of the total RNA folded into a structure that alters flexibility/reactivity at that position. Thus, small divalent-induced changes in SHAPE reactivity may represent RNA structures induced by binding events near their K_d in which some fraction of the RNA is in a conformation that alters reactivity at that position. Higher overall incidence of ΔFe^{2+} sites observed in the LSU rRNA confirms formation of additional tertiary interactions not present in the rRNA-Mg^{2+} state.

Substitution of Fe^{2+} for Mg^{2+} yields greater structural effects in ancient LSU rRNA. Increased response of a-rRNA to Fe^{2+} vs. Mg^{2+} is markedly greater than the response gap observed for the LSU rRNA (Table 5.2), despite lower ranges of ΔFe^{2+} and ΔMg^{2+} values reported for a-rRNA. The ratio of ΔFe^{2+} sites to ΔMg^{2+} sites is also higher in a-rRNA. DV in particular is considerably more responsive to Fe^{2+} than Mg^{2+}, consistent with increased folding competence and stability of the PTC in Fe^{2+}. The higher responsivity of a-rRNA to Fe^{2+} may suggest that ancient LSU rRNA originated in an Fe^{2+}-rich environment, and retains the ability to revert to an Fe^{2+}-RNA state. The lower folding threshold of a-rRNA with Fe^{2+} may have played an important role in the overall stability of the ancestral LSU during a period when little or no peptide was present, very early in the transition from an RNA-dominated biology to the modern paradigm of functional proteins. Higher incidence of Mg^{2+} close to the PTC in the assembled LSU may also contribute to the larger discrepancy in response between a-rRNA and LSU rRNA;[35] RNA that reacts more frequently with divalent cations would reasonably yield a greater response to cation substitution than less cation-rich RNA. However, increased Mg^{2+} density observed in the PTC region is thought to indicate its ancient origin, so these factors are interrelated. In sum, these observations are consistent with a model in which ancient Fe^{2+}-RNA complexes exhibit greater folding competence.

The ultimate chemical origins of RNA's hyperactivity with Fe^{2+} vs. Mg^{2+} as the cofactor may be related to increased affinity of Fe^{2+} for phosphate groups. Due to availability of low lying d orbitals, Fe^{2+} interacts with slightly greater affinity than Mg^{2+} for phosphate oxygens.[50] A modestly tighter binding divalent cation could facilitate improved folding and function.

5.4.3 Isolated LSU rRNA regions exhibit divergent responses to Fe^{2+} and Mg^{2+}

The only regions that noticeably diverge from the overall similarity observed in the ΔFe^{2+} and ΔMg^{2+} heat maps of LSU rRNA are H43/44 and, to a lesser extent, several inter-helical regions in DI. H43/44 becomes uniformly inflexible (reduced SHAPE reactivity) upon addition of Mg^{2+}, but in presence of Fe^{2+} this region exhibits an overall increase in reactivity (Figure 5.3). Stabilization of H43/44 by Mg^{2+} has been reported previously.[227] These helices interact with uL11 in the assembled ribosome, although their structure is not solved in the *T. thermophilus* ribosome crystal structure.[75] In the *H. marismortui* LSU structure, two hydrated Mg^{2+} ions are bound in the core of the H43/44 fold.[11] Our SHAPE data suggest that Fe^{2+} is not able to adequately mimic Mg^{2+} in these positions, leading to divergent Fe^{2+} and Mg^{2+} responses. Several Mg^{2+} ions interact via first-shell interactions at or near the inter-helical regions of DI which exhibit significantly different responses to Fe^{2+} or Mg^{2+}. Again, these regions appear more stable in Mg^{2+} (Figure 5.3), contrary to trends observed in the majority of the LSU rRNA. These specific Mg^{2+} binding sites may also represent structures that favor Mg^{2+} over Fe^{2+}. These isolated divergent regions suggest a degree of subtlety involved in Fe^{2+}/Mg^{2+} substitution.

Models of the most ancient LSU rRNA do not include H43/44 or the divergent regions of DI.[35,109,111,181,226] According to recent estimates of the LSU rRNA's evolutionary history, these regions were accreted in the latter half of the evolution of the ribosomal common core (analogous to the bacterial LSU rRNA).[226] The divergent responses of H43/44 and DI to Fe^{2+} and Mg^{2+} suggest that these regions may have originated in a late period of LSU evolution during which the ribosome had begun to incorporate Mg^{2+} as a structural cofactor instead of or alongside Fe^{2+}. Alternatively, if

Fe^{2+} was bound in these regions earlier in ribosomal evolution, the resulting increased flexibility might have conferred a fitness advantage of some kind upon the ancient ribosome or its products. A comparison of Mg^{2+} vs. Fe^{2+} responsiveness of the LSU rRNA analyzed according to estimated evolutionary age may illuminate this phenomenon further, possibly using a larger LSU rRNA system spanning a longer overall evolutionary history.

5.4.4 Non-canonical RNA helices exhibit anomalous SHAPE patterns

Helices 25a and 26a were not identified by comparative sequence analysis due to a high proportion of non-cWW base pairs,[177] though they clearly form generally helical structures in ribosomal crystal structures.[185] SHAPE reactivity is low for cWW base pairs due to constraint in a conformation in which the 2'-OH and PO are in close proximity in 3D.[55] Non-cWW base paired nucleotides exhibit variable SHAPE reactivity; although they may be constrained by other types of secondary or tertiary interactions, their susceptibility to electrophilic attack by SHAPE reagents is dependent on the specific conformation or conformations (for flexible nucleotides) they access. For instance, purine-purine base pairs in H26a seem to yield higher reactivity in nucleotides of the 3' strand. It is possible that this nucleotide is conformationally restrained in a way that increases SHAPE reactivity (ie. 2'-OH distant in 3D from PO). Alternatively, these anomalous SHAPE results may highlight a region of LSU rRNA that is dynamic and mobile in solution in order to facilitate a particular ribosomal function such as translocation; the ribosome would not function if it were uniformly static. Other regions of the LSU rRNA that appear unpaired in the canonical secondary structure actually form

extended helices, often involving non-cWW base pairs. Solution SHAPE data could be combined with static 3D structural information to support further improvements to rRNA secondary structures. In a greater context, establishment of SHAPE patterns for certain types of interactions or base pairs could inform improvements to SHAPE-directed secondary structure predictions by allowing for facile identification of unusual RNA helices in addition to canonical cWW helices. Further discussion regarding SHAPE-assisted identification of RNA structures will be provided in chapter 6.

5.5 – Conclusions

rRNA folds more readily under pre-GOE conditions. We hypothesize that, in anoxic ancient earth environments, RNA and Fe^{2+} collaborated in an array of biomolecular structures with greater stability and activity than RNA and Mg^{2+} in today's oxic world. This hypothesis is supported by previous results of catalysis assays, *in vitro* structure probing, and modeling experiments.[50,51] Divalent cations are crucial to ribosomal function; virtually all aspects of ribosome structure and function involve Mg^{2+} to some degree.[74] Experiments performed by the Glass group at Georgia Tech demonstrate incorporation of Fe^{2+} into bacterial ribosomes from *E. coli* grown under anoxic conditions (publication in preparation). In the results presented here, Fe^{2+} and Mg^{2+} are both seen to induce folding of rRNA into near-native states. Fe^{2+} induces folding at a lower concentration than Mg^{2+}, suggesting that Fe^{2+} interacts more effectively with rRNA. Many tertiary interactions are induced similarly by Fe^{2+} and Mg^{2+}, and these interactions involve divalent cations bound to rRNA via chelated, loose, and glassy binding modes. Certain isolated, less-ancient rRNA structures seem to respond

more readily to Mg^{2+}; we consider this to be evidence of the eventual transition from Fe^{2+} to Mg^{2+} as the primary RNA structural cofactor. To interrogate the effect of Fe^{2+} vs. Mg^{2+} on individual tertiary interactions, comparison of rRNA SHAPE data obtained in several varied [Fe^{2+}] and [Mg^{2+}] is necessary. While laborious to obtain, this detailed data should provide dissociation constants for specific tertiary interactions, and provide definitive evidence for which interactions are more readily-induced by Fe^{2+} or Mg^{2+}.[123]

We have explored relationships between geochemical conditions of the ancient earth and the folding of rRNA. The anoxic atmosphere of ancient earth facilitated solubility of iron by preventing oxidation of Fe^{2+} to Fe^{3+}. When subjected to pre-GOE conditions, substitution of Fe^{2+} for Mg^{2+} appears to confer increased folding competency and stability to modern and ancient rRNA systems. These results are consistent with our hypothesis that ancient RNAs retain the latent ability to revert to Fe^{2+}-RNA structures.

CHAPTER 6

SHAPE PATTERNS OF RNA MOTIFS

6.1 – Introduction

6.1.1 RNA motifs

As RNA structure is still a relatively young field, the absolute definition an RNA motif is still evolving, as are methods for their accurate classification.[228] Leontis and Westhof describe RNA motifs as ordered stacked arrays of non-Watson-Crick base pairs forming distinctive backbone conformations of the interacting strands.[229] RNA motifs are commonly partitioned into distinct families including hairpin loops (GNRA loops, UNCG loops, T loops), asymmetrical loops (A-minor, K-turn, sarcin-ricin, C-motif), symmetrical internal loops (loop E), junction loops (Hook-turn), and pseudoknots. Here we focus primarily on GNRA loop and K-turn motifs, for which example 3D structures are provided in Figures 6.1 and 6.2. rRNA has provided an immense wealth of information regarding RNA motifs,[230] due in part because rRNAs fold into intricate compact structures. rRNAs are also the largest RNAs for which atomic resolution structural data is available, and size of folded RNA is correlated to incidence of tertiary motifs.[230] D0, the structural core of the LSU, includes an example of the sarcin-ricin loop motif, encompassed by H26a. RNA motifs are important for mediation of tertiary interactions involved in compact folding of complex RNAs, and also for protein or ligand recognition. Motifs are often characterized by consensus sequences that fold into essentially identical 3D structures, including matching base-pairing patterns. Certain

classes of RNA motifs are sequence-independent and are identified only by specific 3D geometries.[228] New motifs are continually reported as more RNA 3D structures are made available, though some would argue that motifs cannot be adequately classified based on single examples as is often attempted.[228] Certain motifs nest to form more complex motifs; adenine residues found in GNRA loops are often involved in A-minor motifs or other loop-receptor motifs.

Base-pair classification is often the first step in identifying RNA motifs based on 3D structure. Leontis and Westhof have developed a robust scheme in which base pairs are partitioned into 12 non-overlapping classes based on geometric properties.[231] High resolution structures are available for almost all base pair types.

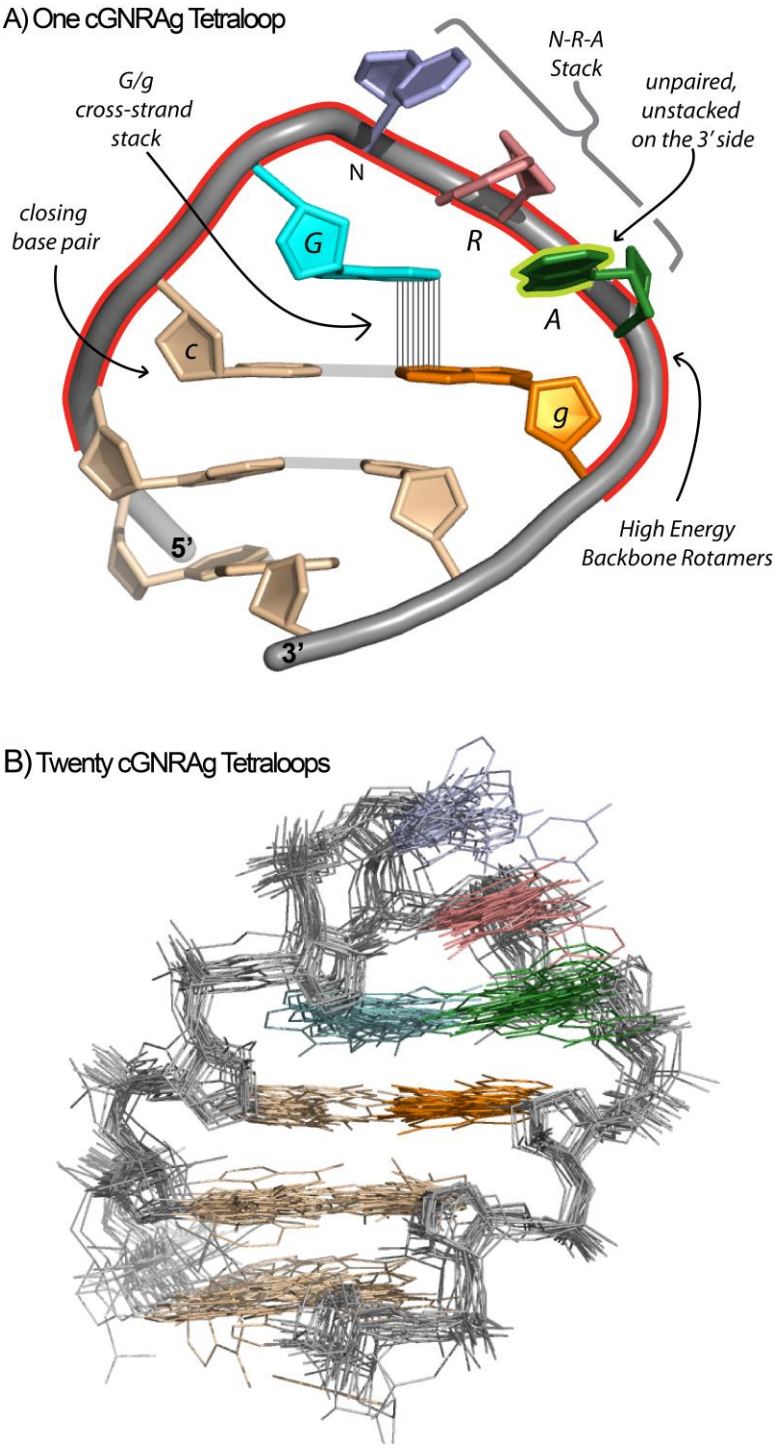


Figure 6.1. A) A representative cGNRAg tetraloop: G, cyan; N, purple; R, pink; and A, green. The tetraloop contains four unpaired loop nucleotides (GNRA) that link the antiparallel helical strands. The closing base-pair helix is the consensus c-g base pair (lowercase letters: c, beige; g, orange). The first residue of the loop (G) is stacked predominantly on the cross-strand g of the closing base-pair and forms a single hydrogen bond to A (the hydrogen bond is not shown). The cross-strand G/g stack is denoted by vertical lines. The rotameric stress of the backbone is indicated by red highlight. The unstacked and unpaired A is highlighted in green. Gray shading indicates base pairing. B) Superimposition of backbone atoms of 20 cGNRAg tetraloops identified in PDB entries 1JJ2 and 2J00. Reprinted with permission.[178]

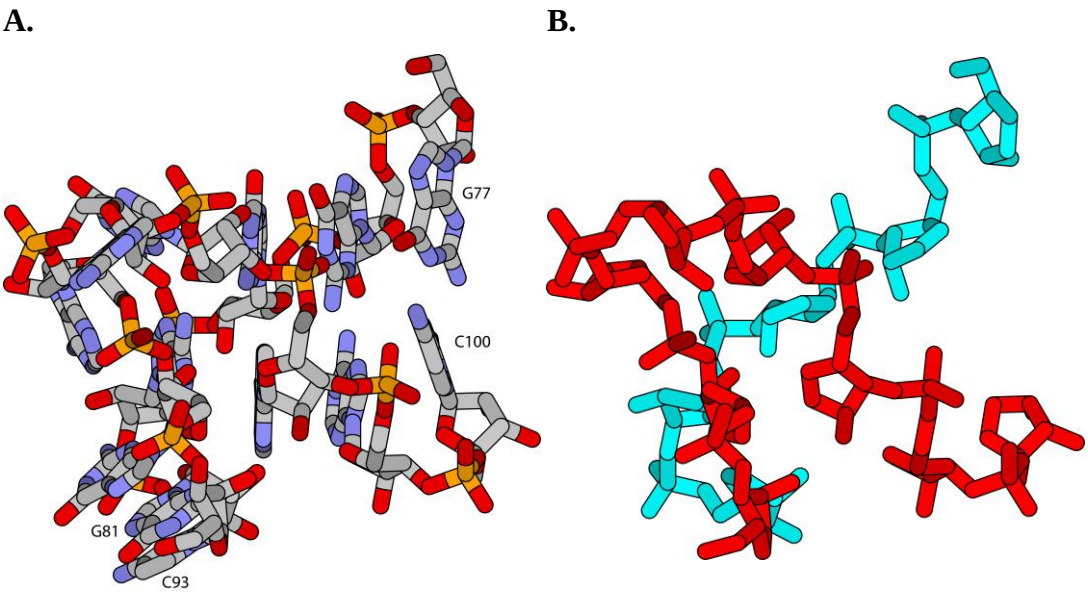


Figure 6.2. 3D images of all-atom (A) and backbone-only (B) representations of a K-turn from the *H. marismortui* LSU rRNA. A) Atoms are colored as follows: carbon, grey; oxygen, red; nitrogen, blue; and phosphorous, orange. B) Nucleotide numbers for the end of each strand of the K-turn are indicated. In panel B, the 5' strand is colored cyan, and the 3' strand of the K-turn is colored red.

6.1.2 Hypotheses regarding SHAPE patterns of RNA motifs and interactions

Here, we explore the potential of SHAPE data to assist in RNA motif prediction and validation. While sequence analyses and folding algorithms are usually the first-line methods for identification of potential RNA motifs, their results are often ambiguous (multiple minimum energy structures and base pairing configurations), especially for large RNA systems where the number of reasonable structural permutations is considerably greater. Given sufficient reference data, a robust, nucleotide-resolution structural method such as SHAPE may hold great utility in motif identification.

SHAPE is sometimes used to predict secondary structure of RNAs for which 3D structural information is not available.[125,130] Utility of SHAPE in secondary structure prediction is generally restricted to weighting parameters which influence whether or not nucleotides participate in cWW interactions in the predicted structure; high reactivity

sites in particular are weighted against inclusion in helices. Automated motif-identification methods (such as FR3D) detect base-pair type and RNA motif candidates based on comparison of a reference set to geometric patterns observed in structural data.[189] Because SHAPE reactivity is modulated by nucleotide geometry and base-pairing interactions, and motifs are commonly defined by geometry and base-pairing classifications, SHAPE-assisted identification of RNA motifs seems theoretically plausible. Chemical probing experiments have been used to identify sarcin/ricin motifs in rRNA previously.[232] The utility of SHAPE in structure prediction may extend to detection of single non-canonical base pairs. Comparison of SHAPE reactivity for nucleotides known to form certain base pair types may reveal unique SHAPE fingerprints.

6.2 – Methods

The preliminary analyses presented here were performed using SHAPE data sets described in chapters 3-5 of this document. Numerical data used in these analyses can be found in Dataset 1.

6.3 – Results

6.3.1 GNRA tetraloops exhibit a consistent SHAPE pattern

The same GNRA tetraloop sequence (5'-gccGUAAggc-3') was used 11 separate times to terminate helices at positions where the LSU rRNA was 'cut' to produce a-rRNA_γ (chapter 3). These stitching tetraloops were designed to protrude from the

exterior of the folded RNA to minimize the probability of tetraloops forming tertiary interactions with the LSU rRNA-derived regions.[157] The a-rRNA system provides a unique opportunity to study 11 incidences of identical tetraloop sequence in different surrounding structural and sequence contexts within the same RNA. GNRA tetraloops are known to be very stable, with a distinct, rigid geometry (Figure 6.1).[178,233] Upon inspection of partially-processed SHAPE data for a-rRNA-β2, collected under varied NMIA concentrations (Figure 6.3), a consistent SHAPE pattern is observed in the stitching tetraloops. The NRA positions (UAA in this case) exhibit a repeating pattern of elevated SHAPE reactivity, while the G position and loop-adjacent stem nucleotides generally yield low reactivity.

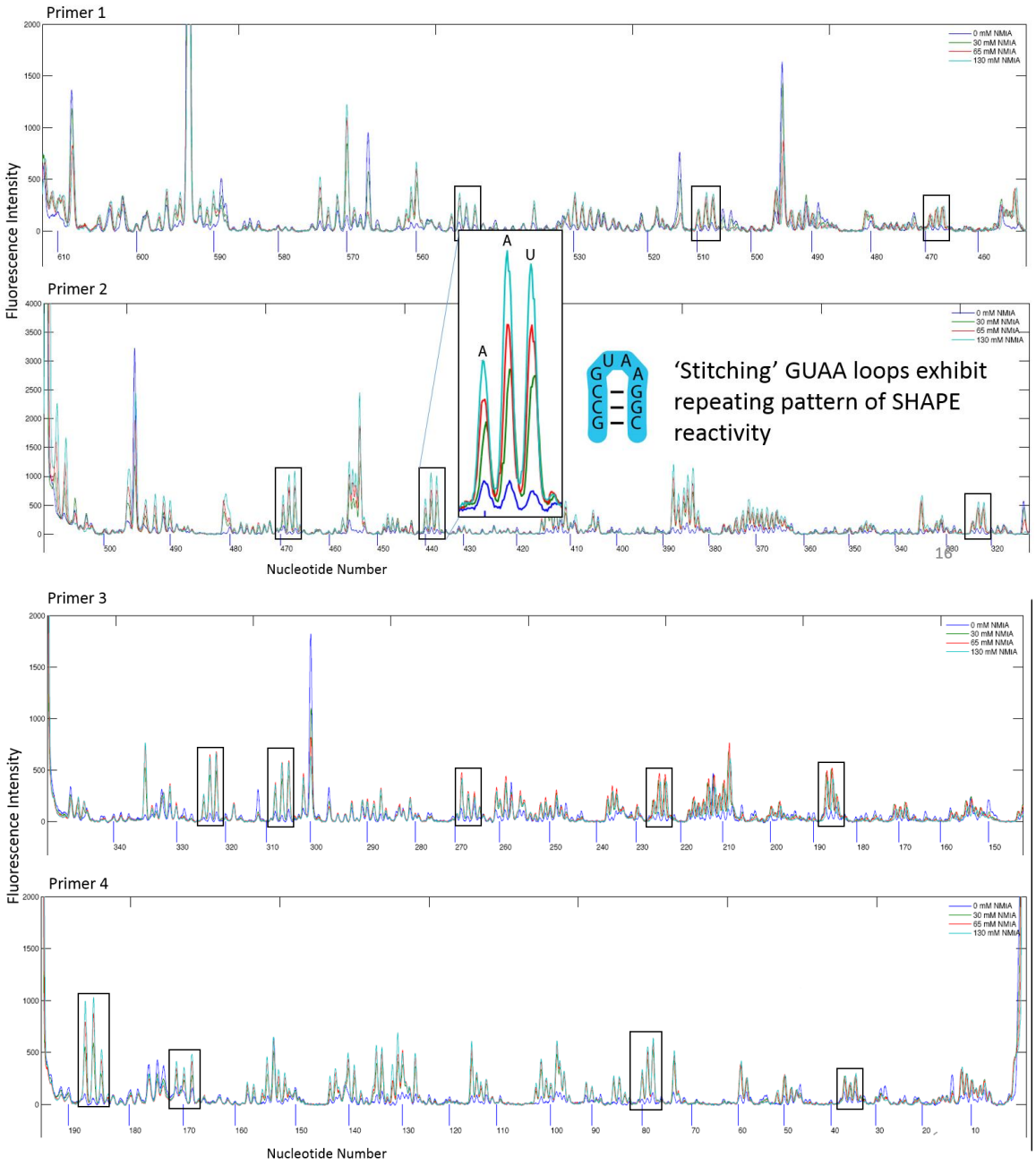


Figure 6.3. GUAA loops exhibit repeating SHAPE patterns. Partially-processed data for a-rRNA-β2 reverse transcribed with the four a-rRNA primers are graphed. Locations of GUAA loops are indicated by black boxes. Inset: close-up view of a typical SHAPE pattern for the UAA positions. SHAPE data from raw electropherograms have been aligned, but no further processing has been performed (ie. background subtraction, peak integration, etc.). NMIA values indicated are 10x stock solutions; [NMIA] in modification reactions was 10-fold lower. Data reads 3' to 5' from left to right. SHAPE reactions were performed in 50 mM NaHEPES pH 8.0, 200 mM NaOAc.

SHAPE data were obtained for all 110 nucleotides within the 11 GUAA tetraloops of a-rRNA-γ (see chapter 5) under Na^+, Na^+/Mg^{2+}, and Na^+/Fe^{2+} conditions. Mean SHAPE reactivities for positions of the GUAA tetraloops of a-rRNA-γ are graphed in Figure 6.4A (numerical values are provided in Dataset 1). The stem regions are mostly unreactive, with few outliers. The loop regions yield a consistent pattern of reactivity that is independent of the surrounding sequence. In 10 of the 11 stem-loops, the UAA nucleotides are more reactive than the preceding G. This pattern of reactivity is consistent with known patterns of GNRA tetraloop structure (Figure 6.1).[150,178] Nucleotides NRA of a GNRA tetraloop are more polymorphic and flexible than the G. The G position is constrained by a cross-strand stacking interaction with the terminal stem base pair. The N nucleotide is the most polymorphic of all, and is the most reactive to SHAPE. Out of the NRA positions, we observe that A is the least reactive, in agreement with the consensus structure; the terminal A position is flipped out of the helix and unstacked, but constrained by a *trans* Hoogsteen/sugar-edge base pairing with the G position. Some variation is observed in the 3' stem nucleotides, due largely to two outlier tetraloops which appear to be partially-formed at best based on SHAPE data. At present, we do not have confirmation that all tetraloops form the consensus structure in 3D. Divalent cations (10 mM Mg^{2+} or 2.5 mM Fe^{2+}) have little effect on tetraloop reactivity, except for a modest decrease in mean SHAPE reactivity at the R position.

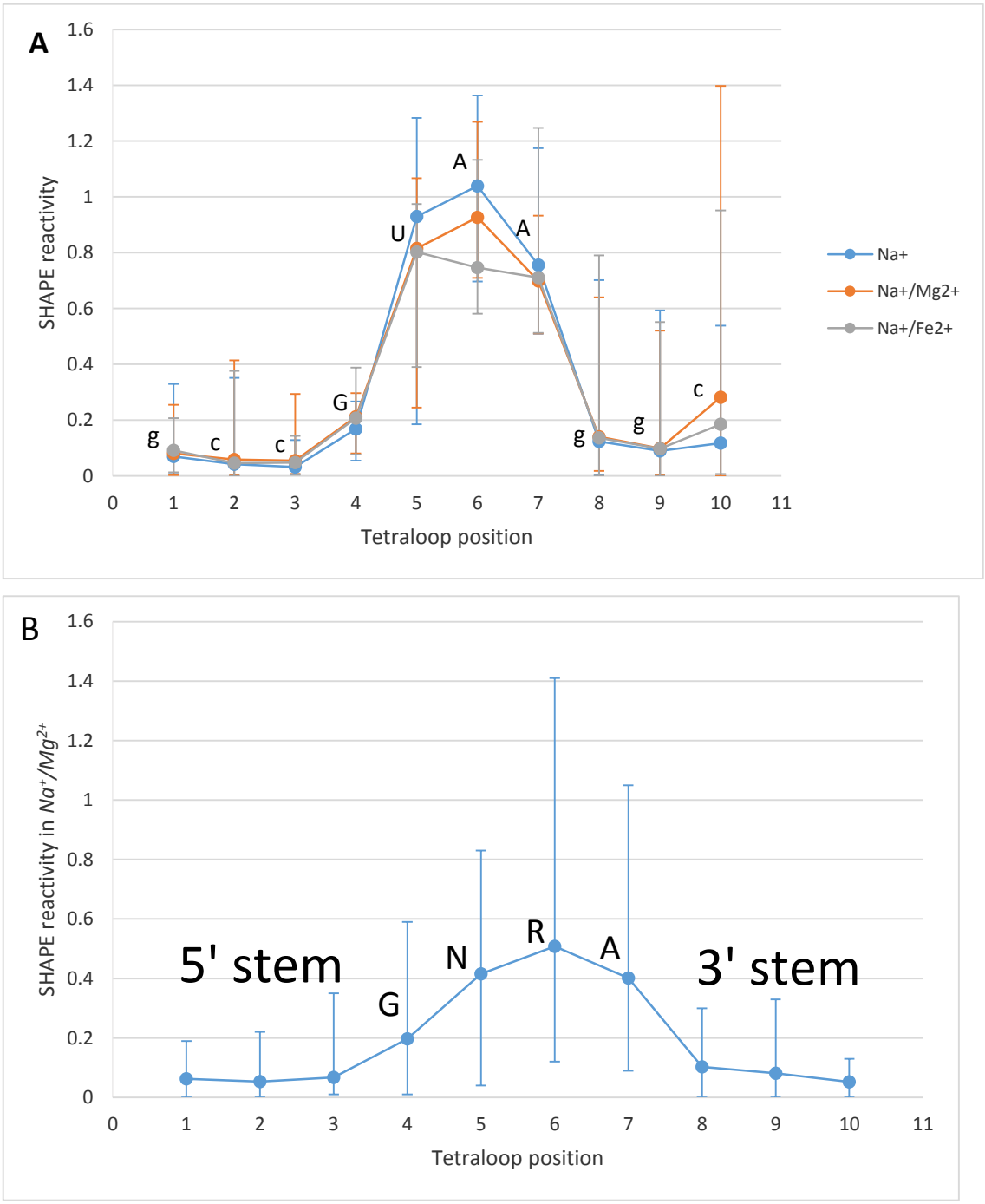


Figure 6.4. Mean SHAPE reactivity of gccGUAAggc tetraloop nucleotides in a-rRNA-γ (A) and standard GNRA tetraloops in the LSU rRNA[35] (B). Error bars represent range at each position. Data was collected under Na^+, Na^+/Mg^{2+}, or Na^+/Fe^{2+} conditions (as indicated), as described in chapters 4 and 5.

GNRA tetraloops of the LSU rRNA exhibit similar SHAPE patterns to those observed in a-rRNA. The LSU rRNA also contains standard GNRA tetraloops in addition to non-standard members of the same tetraloop family.[234] The loops found in the LSU rRNA vary in sequence compared to the gccGUAAggc loops used to stitch together a-rRNA at the NR positions of the loop and stem nucleotides. SHAPE data collected in Na^{+}/Mg^{2+} for eleven standard tetraloops found previously in the *T. thermophilus* LSU rRNA[35] are graphed in Figure 6.4B, and numerical values are provided in Dataset 1. The general trends in tetraloop SHAPE reactivity observed in a-rRNA hold true for LSU rRNA tetraloops; loop nucleotides are more reactive to SHAPE than stem nucleotides, and the NRA positions are more reactive on average than the G position. However, in the LSU rRNA there is a considerably larger variation in SHAPE reactivity for the loop nucleotides. Most loop nucleotides of the LSU rRNA exhibit high reactivity, consistent with analogous values for a-rRNA tetraloops. Several tetraloops yield loop nucleotides with very low maximum SHAPE reactivities (0.18, 0.24, 0.26, 0.29), pulling the average SHAPE reactivity of NRA loop positions down to roughly half of the comparable average values observed in a-rRNA GUAA loops. These low-reactivity LSU rRNA loops would generally be considered largely unreactive to SHAPE, although the loop positions are still consistently more reactive than the stem nucleotides.

6.3.2 A SHAPE pattern observed for K-turns motifs of the LSU rRNA

Four K-turn motifs have been identified in bacterial and archaeal LSU rRNA using automated geometric pattern-recognition methods.[189,235,236] These K-turns are dispersed throughout the secondary structure of the LSU rRNA. Figure 6.5 depicts

SHAPE data of each of these K-turns from the *T. thermophilus* LSU rRNA in Na^{+}/Mg^{2+} (closest to *in vivo* conditions), mapped onto the canonical secondary structure. A pattern is immediately recognizable on the longer side of the asymmetrical loop; the highest SHAPE reactivity of this strand is consistently at position 3 of the K-turn consensus sequence (U99, A1045, A1210, and G2131). In the consensus sequence, this position is occupied by any nucleotide except C, and it is the only nucleotide of the entire motif that is not base-paired (Figure 6.2).[189] The unpaired nature of this position is consistent with observed moderate to high SHAPE reactivity. Reactivity at this position ranges from 0.57-1.92 (normalized SHAPE units). All other nucleotides of the long strand exhibit near-zero SHAPE reactivity, consistent with constraint by base-pairing interactions, in agreement with the consensus structure. The shorter strands of the observed K-turns yield consistently low reactivity at 3 of the 5 positions. Position 3 of the short strand is slightly reactive to SHAPE in half of the K-turn instances of the LSU rRNA (Figure 6.3A-B), and position 4 is moderately reactive in just one K-turn (Figure 6.3B). The common pattern observed here for K-turns may assist in identification of this motif in RNAs for which detailed structural information is unavailable.

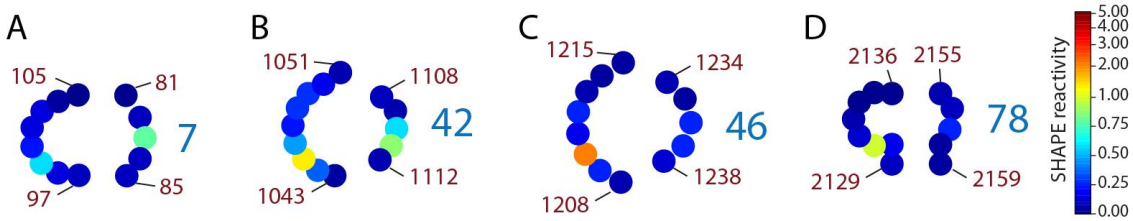


Figure 6.5. SHAPE reactivity for K-turn motifs of the LSU rRNA. Data were collected as described in chapter 4. Nucleotides are colored by SHAPE reactivity in 250 mM Na^{2+}, 10 mM Mg^{2+}, and 50 mM HEPES pH 8. Helix and nucleotide numbers are indicated.

6.4 – Discussion

6.4.1 Utility of SHAPE in identification of GNRA loops

The pattern observed in the GUAA loops of a-rRNA (Figures 6.3 and 6.4A) and GNRA loops of the LSU rRNA suggests a general utility of SHAPE data to identify certain classes of stem-loop motifs. SHAPE reactivity of the NRA positions is consistent with the consensus 3D structure and base-pairing configuration of GNRA loops (Figure 6.1).[234] In RNA secondary structure prediction methods, automated detection of this SHAPE data pattern could be used to weight the predicted local output structure in favor of a GNRA loop.

GNRA SHAPE patterns suggest that the basis of SHAPE reactivity is more nuanced than generally understood. Based on the principles of SHAPE, reactivity to the acylating agent is dictated solely by nucleotide flexibility via the ability of the nucleotide to adopt conformations in which the 2'-OH is distant in space from the adjacent PO, increasing nucleophilicity of the hydroxyl group.[55] In the case presented here, loop nucleotides UAA (NRA positions) are consistently reactive to NMIA under all conditions tested, consistent with previously published data.[157] This is inconsistent with the underlying principles of SHAPE, as NRA nucleotides have been shown to adopt rigid, limited conformations (Figure 6.1). We propose that a more accurate model of SHAPE reactivity involves reaction not only to the 2'-OH of flexible nucleotide positions, but also of nucleotides whose conformation is restricted in such a way that the 2'-OH is distant from the adjoining PO, such as the 'NRA' loop nucleotides of GNRA tetraloops. A detailed comparison of 2'-OH/PO distances in 3D RNA structures and SHAPE reactivity at corresponding conditions is necessary to further explore this hypothesis. A

correlation between SHAPE reactivity and the generalized NMR order parameter has been previously reported,[237] but to our knowledge a large-scale study of SHAPE reactivity and precise 3D nucleotide conformation has not been performed. Analyses of this type will assist in the general understanding of factors that influence SHAPE reactivity, which will ultimately contribute to better-informed application of SHAPE data.

The pattern observed here for the GUAA loops of a-rRNA is expected to hold for all GNRA loops that do not participate in tertiary interactions (A-minor or other tetraloop-receptor motifs, for example). SHAPE reactivity of the 'NR' positions may exhibit interaction-dependent variation, since the nucleobases at these positions are flipped out and available for interaction with other RNA regions. Therefore, GNRA tetraloops involved in tertiary interactions may exhibit a somewhat different overall pattern of SHAPE reactivity. The LSU rRNA contains at least 40 tetraloops that belong to the GNRA family, some of which participate in tertiary interactions.[234] The wide variation observed in SHAPE reactivities of the naturally-occurring standard GNRA loops of the LSU rRNA may present a challenge in SHAPE-assisted identification of all instances of the motif for RNAs of unknown structure. This variation seems to contradict the concept of GNRA loops as rigid structures; if they are truly locked into a limited conformational space, why would the loop nucleotides vary so widely in reactivity to a reagent which is sensitive to local nucleotide conformation? An answer may lie in the surrounding structural contexts of the LSU rRNA tetraloops. Unlike the repeating GUAA loops of a-rRNA, which are all designed to protrude from the exterior of the folded RNA, tetraloops in the intact LSU rRNA are found in a variety of structural contexts, often

participating in tertiary interactions that could conceivably influence SHAPE reactivity. Further analysis of these loops under Na^+ cation conditions, in which fewer tertiary interactions are expected, and sub-classification of the loops into categories by their tertiary interaction status in the assembled LSU, or by their sequence at the variable NR or stem positions, may elucidate the subtleties regarding expected SHAPE reactivities of GNRA loop motifs. Distinction between patterns observed for interacting and non-interacting GNRA loops could also prove useful in structure prediction or validation. Comparison with SHAPE reactivities for non-standard tetraloops with variations such as 3-2 switches, insertions, deletions, and strand clips[234] would also be prudent.

Two of the GUAA tetraloops of a-rRNA exhibit reactivity that diverges from the common pattern. We consider this to be evidence that these particular stem-loops are unstable, or not folding as designed. Because properly-folded GUAA loops yield a well-defined, quantitatively similar SHAPE pattern, it may be possible to estimate the properly-folded proportion of an unstable tetraloop. Because SHAPE reactivity is ultimately an average measure of the reactivity of all RNA molecules within a given sample, comparison of SHAPE reactivity for nucleotides in unstable motifs could theoretically be compared to well-defined consensus SHAPE patterns to yield a quantitative measure of the stability of that particular motif instance. The use of SHAPE to characterize conformational heterogeneity in a single sample has been described previously by Adamiak and coworkers.[127]

6.4.2 Preliminary evidence supports distinct SHAPE patterns of RNA motifs

K-turns exhibit a common pattern of SHAPE reactivity (Figure 6.5). The K-turn consensus sequence consists of many polymorphic positions; any nucleotide is allowed in 6 out of the 12 consensus positions, and only two positions are reserved for a single nucleotide.[236] Polymorphic RNA motifs such as the K-turn could be difficult to confidently identify using sequence information alone. As described for GNRA loops above, SHAPE data collected for RNAs of unknown structure may be analyzed for the K-turn pattern, and candidate K-turns weighted in structure prediction algorithms. Alternatively, candidate K-turns identified by sequence analysis could be confirmed using SHAPE data. Discerning whether the utility of SHAPE reactivity in motif identification extends to other internal loop motifs will require further analysis. Qualitative inspection of LSU rRNA regions identified as forming loop E motifs[232] does not reveal a noticeable SHAPE pattern.

Automated tools are now available with the ability to identify all hairpin and internal loop motifs that have been reported in the literature, including sarcin-ricin, K-turn, C-loop, GNRA loop, T-loop, and UNCG loop motifs.[236] Other automated motif identification methods have been developed by members of the Williams group using geometric pattern recognition, backbone conformation, and/or torsion angles.[234,235,238,239] The RNA 3D Motif Atlas project has applied automated methods to the detection of hairpin and internal loop structures for a *T. thermophilus* ribosome crystal structure (PDB ID: 4QCN).[236] Using one or more of these resources, a motif map of *T. thermophilus* RNA can be generated and compared to SHAPE data obtained for the putative motif regions to identify additional RNA motifs with distinct SHAPE patterns. It is

recommended that the SHAPE data used in this comparison be collected using fully-assembled ribosomes to ensure that all tertiary interactions and RNA motifs observed in the 3D structure are present and stable in the probed rRNA. Due to increased incidence of unpaired or unusually-paired nucleotides, RNA motifs are often less stable than cWW helices. For the same reasons, RNA motifs are frequently involved in protein interactions. Due to these factors, formation of all LSU rRNA motifs cannot be safely assumed in the experiments presented in this document, which were performed in the absence of rProteins, the SSU, or any other ribosomal components.

6.5 – Conclusions

We have described preliminary analyses of the SHAPE reactivity patterns of two common RNA motifs: GNRA loops and K-turns. In both cases, distinctive patterns of SHAPE reactivity are observed. The implications for these patterns include potential improvements to SHAPE-directed RNA secondary structure prediction methods. This is of particular use for RNAs for which detailed structural data is not available.

SHAPE motif patterns may also have utility in verification of motifs that have been observed in static crystal structures. SHAPE provides solution structure data under plausible *in vivo* conditions. If SHAPE data were to contradict crystal structure data, it may suggest that the structures of contradictory regions are not stable in solution. While NMR can provide regarding RNA solution dynamics, the utility of NMR techniques is restricted to small RNAs. As has been demonstrated here, SHAPE data can be collected for very large RNAs.

Though not discussed at length here, we expect some utility may exist in the application of SHAPE reactivities to confirmation or prediction of specific types of base pairs. As is the case for RNA motifs, base pairing classes are defined by their geometry, and SHAPE reactivity is sensitive to local nucleotide geometry.

The SHAPE patterns proposed here could be particularly useful in large RNA systems which may yield multiple candidate secondary structures in prediction algorithms. The application of SHAPE pattern recognition may strongly support one candidate structure over others. Identification of SHAPE reactivity patterns is a promising prospect that could add an additional dimension to the versatility of this robust RNA structural technique.

CHAPTER 7

CONCLUSIONS AND FUTURE DIRECTIONS

This work has detailed multiple SHAPE studies into the role of divalent cations in the folding of rRNA, towards a common goal of increasing understanding of the relationship between divalent cations and ribosomal structure, origins, and evolution. To facilitate these RNA-cation experiments, several RNA constructs were successfully manipulated using common molecular biology techniques (chapter 2). RNA prepared from these constructs was used extensively in experiments described both here and elsewhere.

7.1 – Ancestral rRNA

Design of a-rRNA, a model ancestral rRNA designed by consensus of ribosomal evolution models, was informed and iterated based in part on SHAPE experiments performed on early versions of the model (chapter 3). ~80% of extant LSU rRNA was excised, and the remaining RNA synthesized into a single RNA polymer. Despite these substantial excisions, SHAPE data support that a-rRNA retains the ability to fold into the predicted secondary structure. Comparative SHAPE illuminates the assembly competency of a-rRNA with Mg^{2+} by detection of broad, dispersed Mg^{2+}-induced structural changes. Thermal denaturation studies suggest increased stability of successive a-rRNA iterations and upon association with an ancestral ribosomal peptide. These data are consistent with a model in which the functional core of the ribosome is an ancient

assembly that has remained stable and largely static over billions of years of evolution. This represents the first attempt to resurrect ancestral assemblies of the ribosome *in vitro*. The resurrection process is predicated on trial and error, and other members of the Williams group have since produced a fine-grained model of ribosomal evolution based on accretion of small rRNA segments.[226] Present and future efforts are directed towards the structural and functional characterization of rRNA constructs based on the updated model.

7.2 – The role of Mg^{2+} in LSU structure

Comparative SHAPE experiments performed on extant, protein-free LSU rRNA reveal widely-dispersed structural changes upon association with Mg^{2+}, consistent with global collapse to a near-native conformation that we define as the rRNA-Mg^{2+} state (chapter 4). We consider a model in which Mg^{2+}-responsive sites are induced by formation of specific tertiary interactions observed in the assembled LSU. In our model, Mg^{2+} induces a conformation very close to that observed in the assembled LSU, even in the absence of all other ribosomal components (rProteins, the SSU, etc.). Evidence supports formation of considerable long range inter-domain interactions in the rRNA-Mg^{2+} state, and of those for which evidence is not available, we propose that many are pre-organized for rProtein-mediated interaction based on comparison with ribosome crystal structures. This data is consistent with induction of the core domain architecture of the LSU by Mg^{2+} alone, and by inference the structural effects of rProteins are local and nominal. While single-condition SHAPE experiments have been performed

previously on rRNA systems, this study marks the first use of comparative SHAPE to monitor structural transitions in a large folded RNA.

To improve the understanding of the conformational influence of rProteins, comparison with SHAPE data performed on intact, assembled ribosomes is required. Alternatively, experiments carried out on reconstituted ribosomes associated with one or more rProteins could provide information regarding the structural roles of individual ribosomal components. Application of chemical cross-linking or mutate-and-map techniques such as those developed by Das and coworkers[132] could confirm the assignment of Mg^{2+}-induced SHAPE changes to formation of specific tertiary interactions. The data provided here suggest multiple regions of interest for future *in vitro* studies into specific long-range RNA-RNA interactions involved in LSU assembly.

7.3 – Substitution of Fe^{2+} for Mg^{2+} in rRNA

Modern and ancient protein-free LSU rRNA were probed by SHAPE under anoxic conditions, consistent with those of the environment contemporary with the origin of life (chapter 5). Comparison of data collected in Fe^{2+} vs. Mg^{2+} or in absence of divalent cations suggests the ability of Fe^{2+} to mimic Mg^{2+} in rRNA interactions. This substitution is robust in that the data supports its occurrence in multiple binding modes, ranging from glassy to chelated RNA-cation interactions. Effects of Fe^{2+} on both modern and ancient rRNA are more pronounced than those of Mg^{2+}, inducing greater structural changes to rRNA at lower concentrations. The effect of ancient conditions on ancient rRNA is greater overall.

Certain isolated regions exhibit divergent responses to Fe^{2+} and Mg^{2+}. Further experiments in a wider range of Fe^{2+} and Mg^{2+} concentrations should produce a more fine-grained representation of the response of individual rRNA structures to each cation. Generation of this data would also provide further insight into the extent of tertiary interaction formation at each concentration, and would ultimately yield Mg^{2+} and Fe^{2+} midpoint values for individual structural elements. The next logical goals of this project are to obtain 3D structures of ribosomes crystallized under ancient earth conditions (anoxic, Fe^{2+}), and to assay Fe^{2+}-associated ribosomes for protein synthesis function.

7.4 – SHAPE patterns for common RNA motifs

Certain trends were observed during collection and processing of the >10,000 SHAPE data points obtained for rRNA systems throughout this work (chapter 6). Repeating patterns were detected for the GNRA tetraloops used to seal excision positions in a-rRNA. GNRA tetraloops have a well-defined, rigid structure, and the observed SHAPE pattern is in good agreement with the consensus structure. High reactivity for constrained GNRA loop nucleotides is in disagreement with the prevailing model of SHAPE reactivity solely as a measure of nucleotide flexibility. Four K-turn motifs in the LSU rRNA also yield similar SHAPE patterns, particularly consistently elevated reactivity at position 3 on the long strand of the motif. Utilizing these SHAPE patterns in RNA structure predictions may increase predictive accuracy, or allow for facile discrimination between multiple candidate structures.

A more detailed comparison of SHAPE data for a wider range of known RNA motifs is necessary to elucidate the full extent of this utility. Additional SHAPE data sets

made available by other research groups could be employed to this end. A more complete understanding of the factors that influence SHAPE reactivity should increase efficacy of the application of SHAPE data to structure validation and prediction.

7.5 – New research questions

In addition to the specific experiments proposed above, larger questions have emerged throughout the work presented here. Could certain cations and/or non-protein ribosomal interaction partners, at sufficiently high concentration, induce an LSU rRNA conformation that is functionally competent in terms of peptidyl transferase activity? If so, exactly which native interactions are induced under the conditions that induce function? Are there cation concentrations at which no additional structural changes are observed? Does rRNA structure or function suffer at high cation concentrations? Research aimed at answering these questions could serve to further bridge the gap between the RNA and protein worlds of early evolution. In particular, assignment of LSU function to a specific set of tertiary interactions would be a major development in solving the relationship between structure and function for this intricate biomolecular complex.

Do RNA structures that evolved after the great oxidation event fold in Fe^{2+} or do they prefer Mg^{2+}, the modern divalent substitute? And a related question, how did the ribosome survive the largest atmospheric event in life's history? Evidence of the proposed transition from Fe^{2+} to Mg^{2+} may be imprinted on the ribosome, and could be elucidated based on systematic comparisons of RNA-cation preferences for rRNA systems that contain components that originated both before and after the GOE. For instance, eukaryotic ribosomes essentially contain a bacterial LSU within their expanded

ribosomes; do the ancient core components exhibit similar response to Fe^{2+} as the more recently-evolved eukaryotic rRNA expansion segments?

How far can the utility of SHAPE experiments be extended? How much information is encoded within the SHAPE reactivity values for a single nucleotide or a given RNA region? What factors truly influence SHAPE reactivity? At ten years old, SHAPE is still a relatively 'young' technique. It is clear that SHAPE has broad utility, and SHAPE experiments can provide a wealth of information regarding RNA secondary structure. However, as a wider variety of RNAs are probed by SHAPE, encompassing different interaction modes and structural spaces, trends will undoubtedly emerge that support more powerful applications or developments of the technique.

APPENDIX A

DNA/RNA SEQUENCES

T. thermophilus 23S rRNA DNA gene sequence (sense)

```
1    GTGGGAATTC CGTAATACGA CTCACTATTA GGGTCAAGAT GGTAAGGGCC
51   CACGGTGGAT GCCTCGGCAC CCGAGCCGAT GAAGGACGTG GCTACCTGCG
101  ATAAGCCAGG GGGAGCCGGT AGCGGGCGTG GATCCCTGGA TGTCCGAATG
151  GGGGAACCCG GCCGGCGGGA ACGCCGGTCA CCGCGCTTTT GCGCGGGGGG
201  AACCTGGGGA ACTGAAACAT CTCAGTACCC AGAGGAGAGG AAAGAGAAAT
251  CGACTCCCTG AGTAGCGGCG AGCGAAAGGG GACCAGCCTA AACCGTCCGG
301  CTTGTCCGGG CGGGGTCGTG GGGCCCTCGG ACACCGAATC CCCAGCCTAG
351  CCGAAGCTGT TGGGAAGCAG CGCCAGAGAG GGTGAAAGCC CCGTAGGCGA
401  AAGGTGGGGG GATAGGTGAG GGTACCCGAG TACCCCGTGG TTCGTGGAGC
451  CATGGGGGAA TCTGGGCGGA CCACCGCCTA AGGCTAAGTA CTCCGGGTGA
501  CCGATAGCGC ACCAGTACCG TGAGGGAAAG GTGAAAAGAA CCCCGGGAGG
551  GGAGTGAAAT AGAGCCTGAA ACCGTGGGCT TACAAGCAGT CACGGCCCCG
601  CAAGGGGTTG TGGCGTGCCT ATTGAAGCAT GAGCCGGCGA CTCACGGTCG
651  TGGGCGAGCT TAAGCCGTTG AGGCGGAGGC GTAGGGAAAC CGAGTCCGAA
701  CAGGGCGCAA GCGGGCCGCA CGCGGCCCGC AAAGTCCGCG GCCGTGGACC
751  CGAAACCGGG CGAGCTAGCC CTGGCCAGGG TGAAGCTGGG GTGAGACCCA
801  GTGGAGGCCC GAACCGGTGG GGGATGCAAA CCCCTCGGAT GAGCTGGGGC
851  TAGGAGTGAA AAGCTAACCG AGCCCGGAGA TAGCTGGTTC TCCCCGAAAT
901  GACTTTAGGG TCAGCCTCAG GCGCTGACTG GGGCCTGTAG AGCACTGATA
951  GGGCTAGGGG GCCCACCAGC CTACCAAACC CTGTCAAACT CCGAAGGGTC
```

1001 CCAGGTGGAG CCTGGGAGTG AGGGCGCGAG CGATAACGTC CGCGTCCGAG

1051 CGCGGGAACA ACCGAGACCG CCAGCTAAGG CCCCCAAGTC TGGGCTAAGT

1101 GGTAAAGGAT GTGGCGCCGC GAAGACAGCC AGGAGGTTGG CTTAGAAGCA

1151 GCCATCCTTT AAAGAGTGCG TAATAGCTCA CTGGTCGAGT GGCGCCGCGC

1201 CGAAAATGAT CGGGGCTTAA GCCCAGCGCC GAAGCTGCGG GTCTGGGGGA

1251 TGACCCCAGG CGGTAGGGGA GCGTTCCCGA TGCCGATGAA GGCCGACCCG

1301 CGAGGGCGGC TGGAGGTAAG GGAAGTGCGA ATGCCGGCAT GAGTAACGAT

1351 AAAGAGGGTG AGAATCCCTC TCGCCGTAAG CCCAAGGGTT CCTACGCAAT

1401 GGTCGTCAGC GTAGGGTTAG GCGGGACCTA AGGTGAAGCC GAAAGGCGTA

1451 GCCGAAGGGC AGCCGGTTAA TATTCCGGCC CTTCCCGCAG GTGCGATGGG

1501 GGGACGCTCT AGGCTAGGGG GACCGGAGCC ATGGACGAGC CCGGCCAGAA

1551 GCGCAGGGTG GGAGGTAGGC AAATCCGCCT CCCAACAAGC TCTGCGTGGT

1601 GGGGAAGCCC GTACGGGTGA CAACCCCCCG AAGCCAGGGA GCCAAGAAAA

1651 GCCTCTAAGC ACAACCTGCG GGAACCCGTA CCGCAAACCG ACACAGGTGG

1701 GCGGGTGCAA GAGCACTCAG GCGCGCGGGA GAACCCTCGC CAAGGAACTC

1751 TGCAAGTTGG CCCCGTAACT TCGGGAGAAG GGGTGCTCCC TGGGGTGATG

1801 AGCCCCGGGG AGCCGCAGTG AACAGGCTCT GGCGACTGTT TACCAAAAAC

1851 ACAGCTCTCT GCGAACTCGT AAGAGGAGGT ATAGGGAGCG ACGCTTGCCC

1901 GGTGCCGGAA GGTCAAGGGG AGGGGTGCAA GCCCCGAACC GAAGCCCCGG

1951 TGAACGGCGG CCGTAACTAT AACGGTCCTA AGGTAGCGAA ATTCCTTGTC

2001 GGGTAAGTTC CGACCTGCAC GAAAAGCGTA ACGACCGGAG CGCTGTCTCG

2051 GCGAGGGACC CGGTGAAATT GAACTGGCCG TGAAGATGCG GCCTACCCGT

2101 GGCAGGACGA AAAGACCCCG TGGAGCTTTA CTGCAGCCTG GTGTTGGCTC

2151 TTGGTCGCGC CTGCGTAGGA TAGGTGGGAG CCTGTGAACC CCCGCCTCCG

2201 GGTGGGGGGG AGGCGCCGGT GAAATACCAC CCTGGCGCGG CTGGGGGCCT

2251 AACCCTCGGA TGGGGGGACA GCGCTTGGCG GGCAGTTTGA CTGGGGCGGT

2301 CGCCTCCTAA AAGGTAACGG AGGCGCCCAA AGGTCCCCTC AGGCGGGACG

2351 GAAATCCGCC GGAGAGCGCA AGGGTAGAAG GGGGCCTGAC TGCGAGGCCT

2401 GCAAGCCGAG CAGGGGCGAA AGCCGGGCCT AGTGAACCGG TGGTCCCGTG

2451 TGGAAGGGCC ATCGATCAAC GGATAAAAGT TACCCCGGGG ATAACAGGCT

2501 GATCTCCCCC GAGCGTCCAC AGCGGCGGGG AGGTTTGGCA CCTCGATGTC

2551 GGCTCGTCGC ATCCTGGGGC TGAAGAAGGT CCCAAGGGTT GGGCTGTTCG

2601 CCCATTAAAG CGGCACGCGA GCTGGGTTCA GAACGTCGTG AGACAGTTCG

2651 GTCTCTATCC GCCACGGGCG CAGGAGGCTT GAGGGGGGCT CTTCCTAGTA

2701 CGAGAGGACC GGAAGGGACG CACCTCTGGT TTCCCAGCTG TCCCTCCAGG

2751 GGCATAAGCT GGGTAGCCAT GTGCGGAAGG GATAACCGCT GAAAGCATCT

2801 AAGCGGGAAG CCCGCCCCAA GATGAGGCCT CCCACGGCGT CAAGCCGGTA

2851 AGGACCCGGG AAGACCACCC GGTGGATGGG CCGGGGGTGT AAGCGCCGCG

2901 AGGCGTTGAG CCGACCGGTC CCAATCGTCC GAGGTCTTGA CCCCTCCAAG

2951 CTTGGTG

T. thermophilus 23S rRNA sequence

```
1    GGUCAAGAUG GUAAGGGCCC ACGGUGGAUG CCUCGGCACC CGAGCCGAUG
51   AAGGACGUGG CUACCUGCGA UAAGCCAGGG GGAGCCGGUA GCGGGCGUGG
101  AUCCCUGGAU GUCCGAAUGG GGGAACCCGG CCGGCGGGAA CGCCGGUCAC
151  CGCGCUUUUG CGCGGGGGGA ACCUGGGGAA CUGAAACAUC UCAGUACCCA
201  GAGGAGAGGA AAGAGAAAUC GACUCCCUGA GUAGCGGCGA GCGAAAGGGG
251  ACCAGCCUAA ACCGUCCGGC UUGUCCGGGC GGGGUCGUGG GGCCCUCGGA
301  CACCGAAUCC CCAGCCUAGC CGAAGCUGUU GGGAAGCAGC GCCAGAGAGG
351  GUGAAAGCCC CGUAGGCGAA AGGUGGGGGG AUAGGUGAGG GUACCCGAGU
401  ACCCCGUGGU UCGUGGAGCC AUGGGGGAAU CUGGGCGGAC CACCGCCUAA
451  GGCUAAGUAC UCCGGGUGAC CGAUAGCGCA CCAGUACCGU GAGGGAAAGG
501  UGAAAAGAAC CCCGGGAGGG GAGUGAAAUA GAGCCUGAAA CCGUGGGCUU
551  ACAAGCAGUC ACGGCCCCGC AAGGGGUUGU GGCGUGCCUA UUGAAGCAUG
601  AGCCGGCGAC UCACGGUCGU GGGCGAGCUU AAGCCGUUGA GGCGGAGGCG
651  UAGGGAAACC GAGUCCGAAC AGGGCGCAAG CGGGCCGCAC GCGGCCCGCA
701  AAGUCCGCGG CCGUGGACCC GAAACCGGGC GAGCUAGCCC UGGCCAGGGU
751  GAAGCUGGGG UGAGACCCAG UGGAGGCCCG AACCGGUGGG GGAUGCAAAC
801  CCCUCGGAUG AGCUGGGGCU AGGAGUGAAA AGCUAACCGA GCCCGGAGAU
851  AGCUGGUUCU CCCCGAAAUG ACUUUAGGGU CAGCCUCAGG CGCUGACUGG
901  GGCCUGUAGA GCACUGAUAG GGCUAGGGGG CCCACCAGCC UACCAAACCC
951  UGUCAAACUC CGAAGGGUCC CAGGUGGAGC CUGGGAGUGA GGGCGCGAGC
1001 GAUAACGUCC GCGUCCGAGC GCGGGAACAA CCGAGACCGC CAGCUAAGGC
1051 CCCCAAGUCU GGGCUAAGUG GUAAAGGAUG UGGCGCCGCG AAGACAGCCA
```

1101 GGAGGUUGGC UUAGAAGCAG CCAUCCUUUA AAGAGUGCGU AAUAGCUCAC

1151 UGGUCGAGUG GCGCCGCGCC GAAAAUGAUC GGGGCUUAAG CCCAGCGCCG

1201 AAGCUGCGGG UCUGGGGGAU GACCCCAGGC GGUAGGGGAG CGUUCCCGAU

1251 GCCGAUGAAG GCCGACCCGC GAGGGCGGCU GGAGGUAAGG GAAGUGCGAA

1301 UGCCGGCAUG AGUAACGAUA AAGAGGGUGA GAAUCCCUCU CGCCGUAAGC

1351 CCAAGGGUUC CUACGCAAUG GUCGUCAGCG UAGGGUUAGG CGGGACCUAA

1401 GGUGAAGCCG AAAGGCGUAG CCGAAGGGCA GCCGGUUAAU AUUCCGGCCC

1451 UUCCCGCAGG UGCGAUGGGG GGACGCUCUA GGCUAGGGGG ACCGGAGCCA

1501 UGGACGAGCC CGGCCAGAAG CGCAGGGUGG GAGGUAGGCA AAUCCGCCUC

1551 CCAACAAGCU CUGCGUGGUG GGGAAGCCCG UACGGGUGAC AACCCCCCGA

1601 AGCCAGGGAG CCAAGAAAAG CCUCUAAGCA CAACCUGCGG GAACCCGUAC

1651 CGCAAACCGA CACAGGUGGG CGGGUGCAAG AGCACUCAGG CGCGCGGGAG

1701 AACCCUCGCC AAGGAACUCU GCAAGUUGGC CCCGUAACUU CGGGAGAAGG

1751 GGUGCUCCCU GGGGUGAUGA GCCCCGGGGA GCCGCAGUGA ACAGGCUCUG

1801 GCGACUGUUU ACCAAAAACA CAGCUCUCUG CGAACUCGUA AGAGGAGGUA

1851 UAGGGAGCGA CGCUUGCCCG GUGCCGGAAG GUCAAGGGGA GGGGUGCAAG

1901 CCCCGAACCG AAGCCCCGGU GAACGGCGGC CGUAACUAUA ACGGUCCUAA

1951 GGUAGCGAAA UUCCUUGUCG GGUAAGUUCC GACCUGCACG AAAAGCGUAA

2001 CGACCGGAGC GCUGUCUCGG CGAGGGACCC GGUGAAAUUG AACUGGCCGU

2051 GAAGAUGCGG CCUACCCGUG GCAGGACGAA AAGACCCCGU GGAGCUUUAC

2101 UGCAGCCUGG UGUUGGCUCU UGGUCGCGCC UGCGUAGGAU AGGUGGGAGC

2151 CUGUGAACCC CCGCCUCCGG GUGGGGGGGA GGCGCCGGUG AAAUACCACC

2201 CUGGCGCGGC UGGGGGCCUA ACCCUCGGAU GGGGGGACAG CGCUUGGCGG

2251 GCAGUUUGAC UGGGGCGGUC GCCUCCUAAA AGGUAACGGA GGCGCCCAAA

2301 GGUCCCCUCA GGCGGGACGG AAAUCCGCCG GAGAGCGCAA GGGUAGAAGG

2351 GGGCCUGACU GCGAGGCCUG CAAGCCGAGC AGGGGCGAAA GCCGGGCCUA

2401 GUGAACCGGU GGUCCCGUGU GGAAGGGCCA UCGAUCAACG GAUAAAAGUU

2451 ACCCCGGGGA UAACAGGCUG AUCUCCCCCG AGCGUCCACA GCGGCGGGGA

2501 GGUUUGGCAC CUCGAUGUCG GCUCGUCGCA UCCUGGGGCU GAAGAAGGUC

2551 CCAAGGGUUG GGCUGUUCGC CCAUUAAAGC GGCACGCGAG CUGGGUUCAG

2601 AACGUCGUGA GACAGUUCGG UCUCUAUCCG CCACGGGCGC AGGAGGCUUG

2651 AGGGGGGCUC UUCCUAGUAC GAGAGGACCG GAAGGGACGC ACCUCUGGUU

2701 UCCCAGCUGU CCCUCCAGGG GCAUAAGCUG GGUAGCCAUG UGCGGAAGGG

2751 AUAACCGCUG AAAGCAUCUA AGCGGGAAGC CCGCCCCAAG AUGAGGCCUC

2801 CCACGGCGUC AAGCCGGUAA GGACCCGGGA AGACCACCCG GUGGAUGGGC

2851 CGGGGGUGUA AGCGCCGCGA GGCGUUGAGC CGACCGGUCC CAAUCGUCCG

2901 AGGUCUUGAC CCCUCCAAGC UU

a-rRNA-β1 RNA sequence

```
1    GUGCCUAUUG AAGCAUGAGC CGGCGACUCA GCCGUAAGGC UGGACCCGAA
50   ACCGGGCGAG CUAGCCCUGG CCAGCCGUAA GGCGAACCGG UGGGGGAUGC
100  AAACCCCUCG GAUGAGCUGG GGCUAGGAGU GAAAAGCUAA CCGAGCCCGG
151  AGAUAGCUGG UUCUGCCGUA AGGCAGCGUU GCCGUAAGGC AAGUGCGAAU
201  GCCGGCAUGA GUAACGAGCC GUAAGGCGCG GGAGAACCCU CGCCAAGGAA
251  CUCUGCAAGG CCGUAAGGCG CUCUGGCGAC UGUUUACCAA AAACACAGCG
301  CCGUAAGGCG CGACGGCCGU AAGGCCGUAA CGACCGGAGC GCUGUCUCGG
351  CGAGGGACCC GGUGAAAUUG AACUGGCCGU GAAGAUGCGG CCUACCCGUG
401  GCAGGACGAA AAGACCCCGU GGAGCUUUAC UGCCGUAAGG CAGUUUGACU
451  GGGGCGGUCG GCCGUAAGGC UAAAAGUUAC CCCGGGGAUA ACAGGCUGAU
501  CGCCGUAAGG CGGUUUGGCA CCUCGAUGUC GGCUCGUCGC GCCGUAAGGC
551  UUGGGCUGUU CGCCCAUUAA AGCGGCACGC GAGCUGGGUU CAGAACGUCG
601  UGAGACAGUU CGGUCUCUAU CCGCCACGGG C
```

a-rRNA-β2 RNA sequence

```
1    GUGCCUAUUG AAGCAUGAGC CGGCGACUCA GCCGUAAGGC UGGACCCGAA

51   ACCGGGCGAG CUAGCCCUGG CCAGCCGUAA GGCGAACCGG UGGGGGAUGC

101  AAACCCCUCG GAUGAGCUGG GGCUAGGAGU GAAAAGCUAA CCGAGCCCGG

151  AGAUAGCUGG UUCUGCCGUA AGGCAGCGUU GCCGUAAGGC AAGUGCGAAU

201  GCCGGCAUGA GUAACGUGGC CGUAAGGCCU CGGGAGAACC CUCGCCAAGG

251  AACUCUGCAA GGCCGUAAGG CGCUCUGGCG ACUGUUUACC AAAAACACAG

301  CGCCGUAAGG CGCGACGGCC GUAAGGCCGU AACGACCGGA GCGCUGUCUC

351  GGCGAGGGAC CCGGUGAAAU UGAACUGGCC GUGAAGAUGC GGCCUACCCG

401  UGGCAGGACG AAAAGACCCC GUGGAGCUUU ACUGCCGUAA GGCAGUUUGA

451  CUGGGGCGGU CGGCCGUAAG GCUAAAAGUU ACCCCGGGGA UAACAGGCUG

501  AUCGCCGUAA GGCGGUUUGG CACCUCGAUG UCGGCUCGUC GCGCCGUAAG

551  GCUUGGGCUG UUCGCCCAUU AAAGCGGCAC GCGAGCUGGG UUCAGAACGU

601  CGUGAGACAG UUCGGUCUCU AUCCGCCACG GGC
```

a-rRNA-γ RNA sequence

```
1    GUGCCUAUUG AAGCAUGAGC CGGCGACUCA GCCGUAAGGC UGGACCCGAA

51   ACCGGGCGAG CUAGCCCUGG CCAGCCGUAA GGCGAACCGG UGGGGGAUGC

101  AAACCCCUCG GAUGAGCUGG GGCUAGGAGU GAAAAGCUAA CCGAGCCCGG

151  AGAUAGCUGG UUCUGCCGUA AGGCAGCGUU GCCGUAAGGC AAGUGCGAAU

201  GCCGGCAUGA GUAACGAGCG GGAGAACCCU CGCCAAGGAA CUCUGCAAGC

251  CGUAAGGCCU CUGGCGACUG UUUACCAAAA ACACAGCGCC GUAAGGCGCG

301  AGCCGUAAGG CUAACGACCG GAGCGCUGUC UCGGCGAGGG ACCCGGUGAA

351  AUUGAACUGG CCGUGAAGAU GCGGCCUACC CGUGGCAGGA CGAAAAGACC

401  CCGUGGAGCU UUACUGCCGU AAGGCAGUUU GACUGGGGCG GUCGGCCGUA

451  AGGCUAAAAG UUACCCCGGG GAUAACAGGC UGAUCGCCGU AAGGCGGUUU

501  GGCACCUCGA UGUCGGCUCG UCGCGCCGUA AGGCUUGGGC UGUUCGCCCA

551  UUAAAGCGGC ACGCGAGCUG GGUUCAGAAC GUCGUGAGAC AGUUCGGUCU

601  CUAUCCGCCA CGGGC
```

P4-P6 (C109G, G212C) gene sequence (sense)

```
1    GTGGGAATTC TAATACGACT CACTATAGGG GAATTGGGGG AAAGGGGTCA

51   ACAGCCGTTC AGTACCAAGT CTCAGGGGAA ACTTTGAGAT GGCCTTGCAA

101  AGGGTATGGT AATAAGCTGA CGGACATGGT CCTAACCCCC AGCCAAGTCC

151  TAAGTCAACA GATCTTCTGT TGATATGGAT GCAGTTCAAG CTTGGTG
```

P4-P6 (C109G, G212C) RNA sequence

```
1    GAAUUGGGGG AAAGGGGUCA ACAGCCGUUC AGUACCAAGU CUCAGGGGAA

51   ACUUUGAGAU GGCCUUGCAA AGGGUAUGGU AAUAAGCUGA CGGACAUGGU

101  CCUAACCCCC AGCCAAGUCC UAAGUCAACA GAUCUUCUGU UGAUAUGGAU

151  GCAGUUC
```

P4-P6 WT gene sequence (sense)

```
1    GTGGGAATTC TAATACGACT CACTATAGGG GAATTGCGGG AAAGGGGTCA

51   ACAGCCGTTC AGTACCAAGT CTCAGGGGAA ACTTTGAGAT GGCCTTGCAA

101  AGGGTATGGT AATAAGCTGA CGGACATGGT CCTAACCCGC AGCCAAGTCC

151  TAAGTCAACA GATCTTCTGT TGATATGGAT GCAGTTCAAG CTTGGTG
```

P4-P6 WT RNA sequence

```
1    GAAUUGCGGG AAAGGGGUCA ACAGCCGUUC AGUACCAAGU CUCAGGGGAA

51   ACUUUGAGAU GGCCUUGCAA AGGGUAUGGU AAUAAGCUGA CGGACAUGGU

101  CCUAACCCGC AGCCAAGUCC UAAGUCAACA GAUCUUCUGU UGAUAUGGAU

151  GCAGUUC
```

P4-P6 UNFLD gene sequence (sense)

```
1    GTGGGAATTC TAATACGACT CACTATAGGG GAATTGCGGG AAAGGGGTCA

51   ACAGCCGTTC AGTACCAAGT CTCAGGGGAA ACTTTGAGAT GGCCTTGCAA

101  AGGGTATGGT AATUAGCTGA CGGACATGGT CCTAACCCGC AGCCAAGTCC

151  TAAGTCAACA GATCTTCTGT TGATATGGAT GCAGTTCAAG CTTGGTG
```

P4-P6 UNFLD RNA sequence (sense)

```
1    GAAUUGCGGG AAAGGGGUCA ACAGCCGUUC AGUACCAAGU CUCAGGGGAA

51   ACUUUGAGAU GGCCUUGCAA AGGGUAUGGU AAUUAGCUGA CGGACAUGGU

101  CCUAACCCGC AGCCAAGUCC UAAGUCAACA GAUCUUCUGU UGAUAUGGAU

151  GCAGUUC
```

APPENDIX B

DESCRIPTION OF DATASET 1

Dataset 1 has been uploaded as a supplement to this document. It is provided in form of a spreadsheet file which contains the SHAPE datasets generated in chapters 3-5, as well as $\Delta Mg^{2+}/\Delta Fe^{2+}$ values and subsets of the acquired data utilized in analyses of GNRA tetraloop SHAPE reactivities (chapter 6).

APPENDIX C

ADDITIONAL SHAPE FIGURES

These figures are provided for easier visualization of the LSU rRNA SHAPE data presented in chapters 4 and 5. Mapped data for the Na^{+}/*low* Mg^{2+} dataset for a-rRNA referenced in chapter 5 is also provided (numerical data is supplied in Dataset 1). RiboVision-generated figures are provided here in larger versions which include nucleotide, helix, and domain numbering.

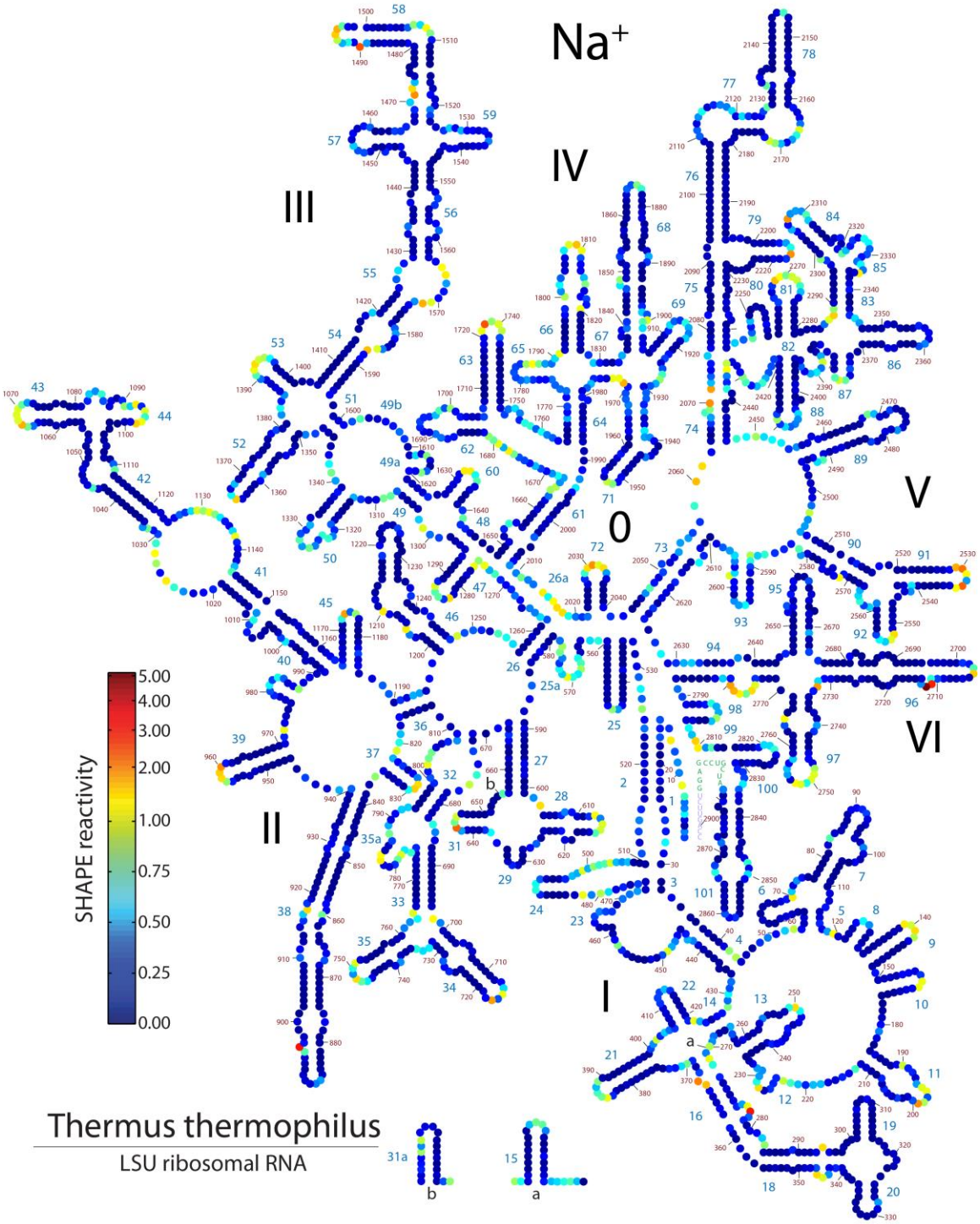


Figure C.1. SHAPE reactivities for the *T. thermophilus* LSU rRNA in Na^+. SHAPE reactivities are mapped onto LSU rRNA secondary structure. Samples contained 200 mM NaOAc, 50 mM NaHEPES, pH 8. Helix numbers (blue), domain numbers, and insertion points are indicated. Nucleotide numbers are denoted every ~10 nt (*E. coli* numbering). Regions where SHAPE data is not available (5' and 3' ends) are displayed as sequence only. This is a full-page version of Figure 4.3A. Figure generated with RiboVision.

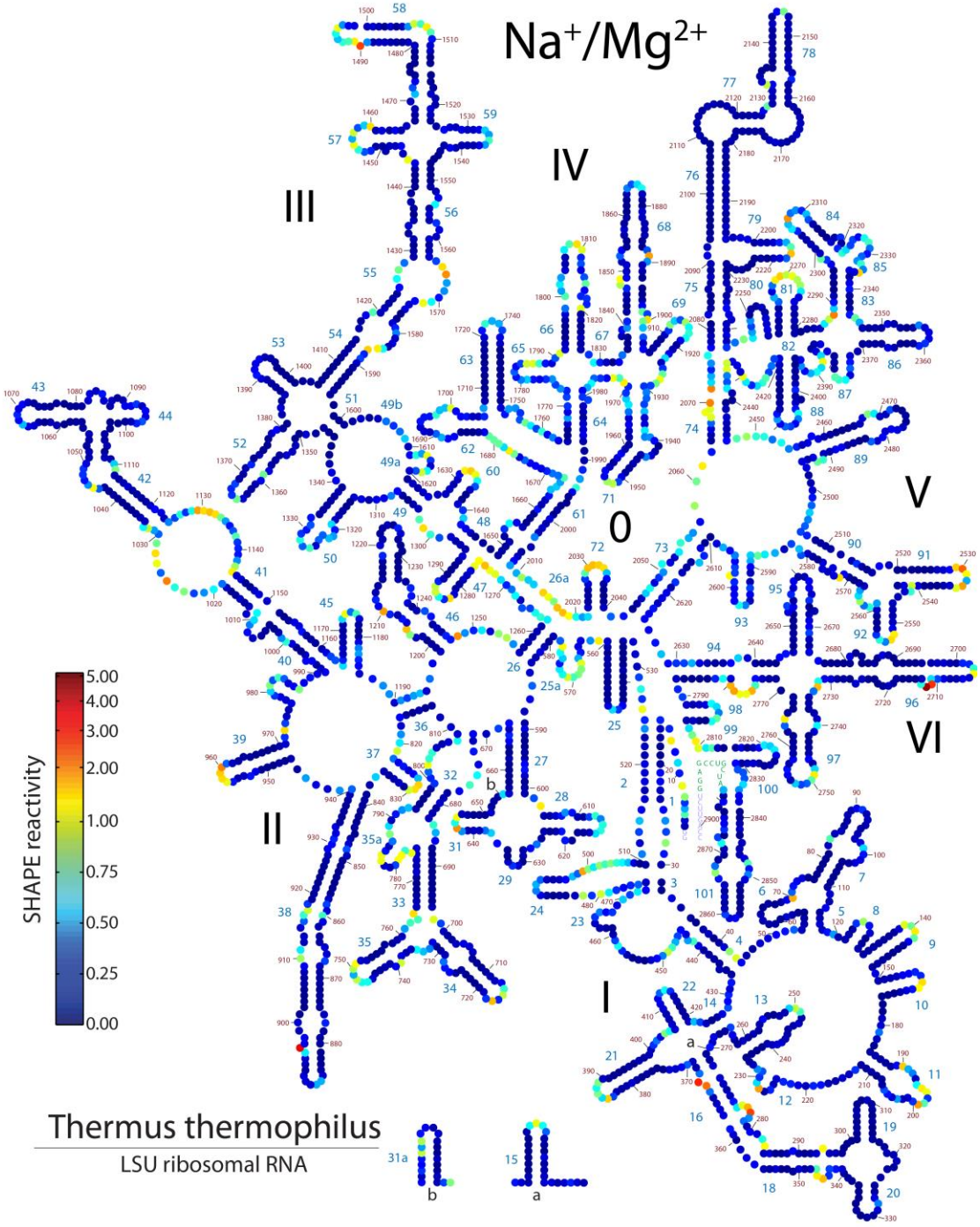


Figure C.2. SHAPE reactivities for the *T. thermophilus* LSU rRNA in Na^{+}/Mg^{2+}. SHAPE reactivities are mapped onto LSU rRNA secondary structure. Samples contained 200 mM NaOAc, 50 mM NaHEPES, pH 8, and 10 mM $MgCl_2$. Helix numbers (blue), domain numbers, and insertion points are indicated. Nucleotide numbers are denoted every ~10 nt (*E. coli* numbering). Regions where SHAPE data is not available (5' and 3' ends) are displayed as sequence only. This is a full-page version of Figure 4.3B. Figure generated with RiboVision.

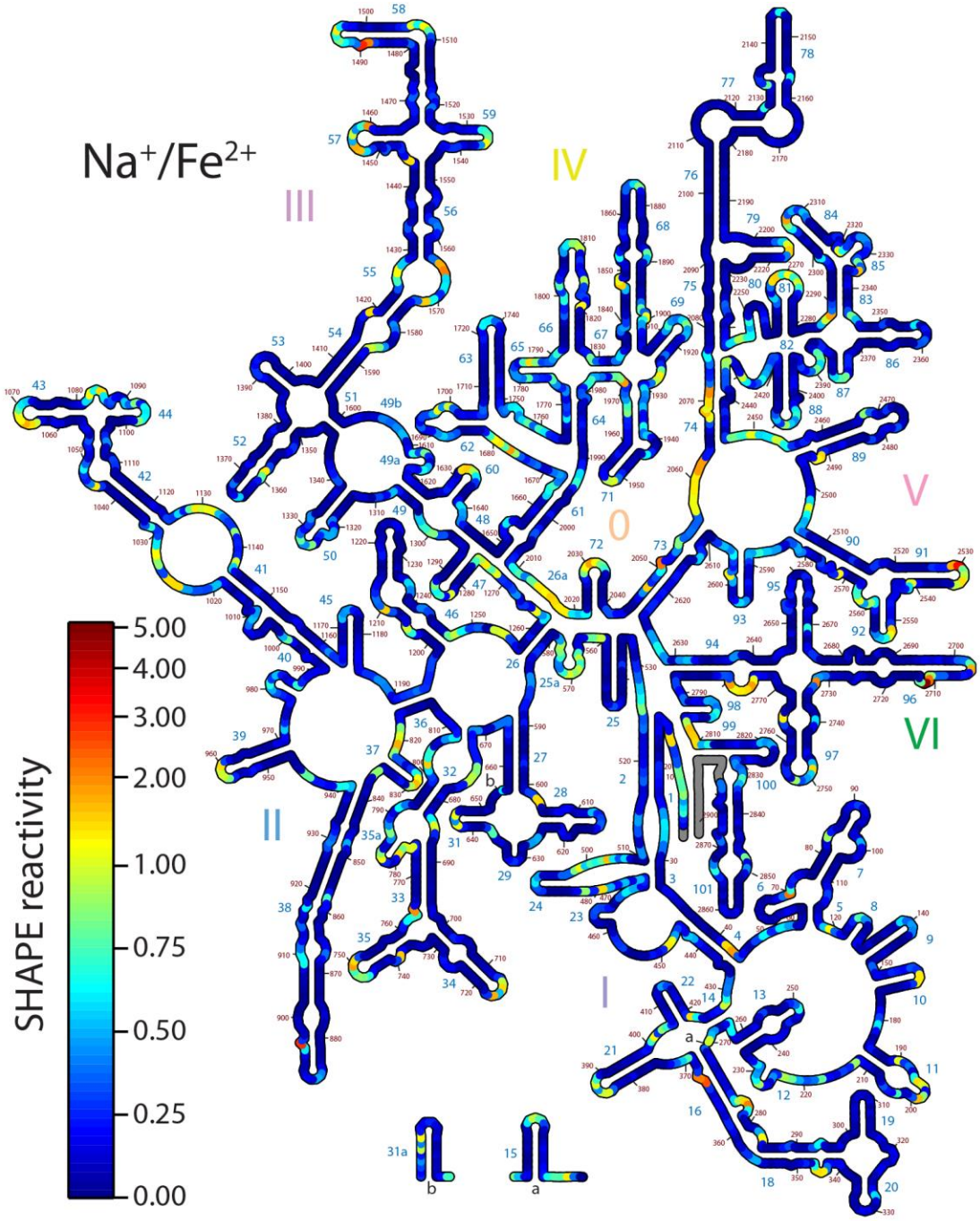


Figure C.3. SHAPE reactivities for the *T. thermophilus* LSU rRNA in Na^{+}/Fe^{2+}. SHAPE reactivities are mapped onto LSU rRNA secondary structure. Samples contained 200 mM NaOAc, 50 mM NaHEPES, pH 8, and 2.5 mM $FeCl_2$. Helix numbers (blue), domain numbers, and insertion points are indicated. Nucleotide numbers are denoted every ~10 nt (*E. coli* numbering). Regions where SHAPE data is not available (5' and 3' ends) are grey. This is a full-page version of Figure 5.3B. Line segment length is not proportionate to number of nucleotides due to inconsistent nucleotide spacing in the secondary structure. Figure generated with RiboVision.

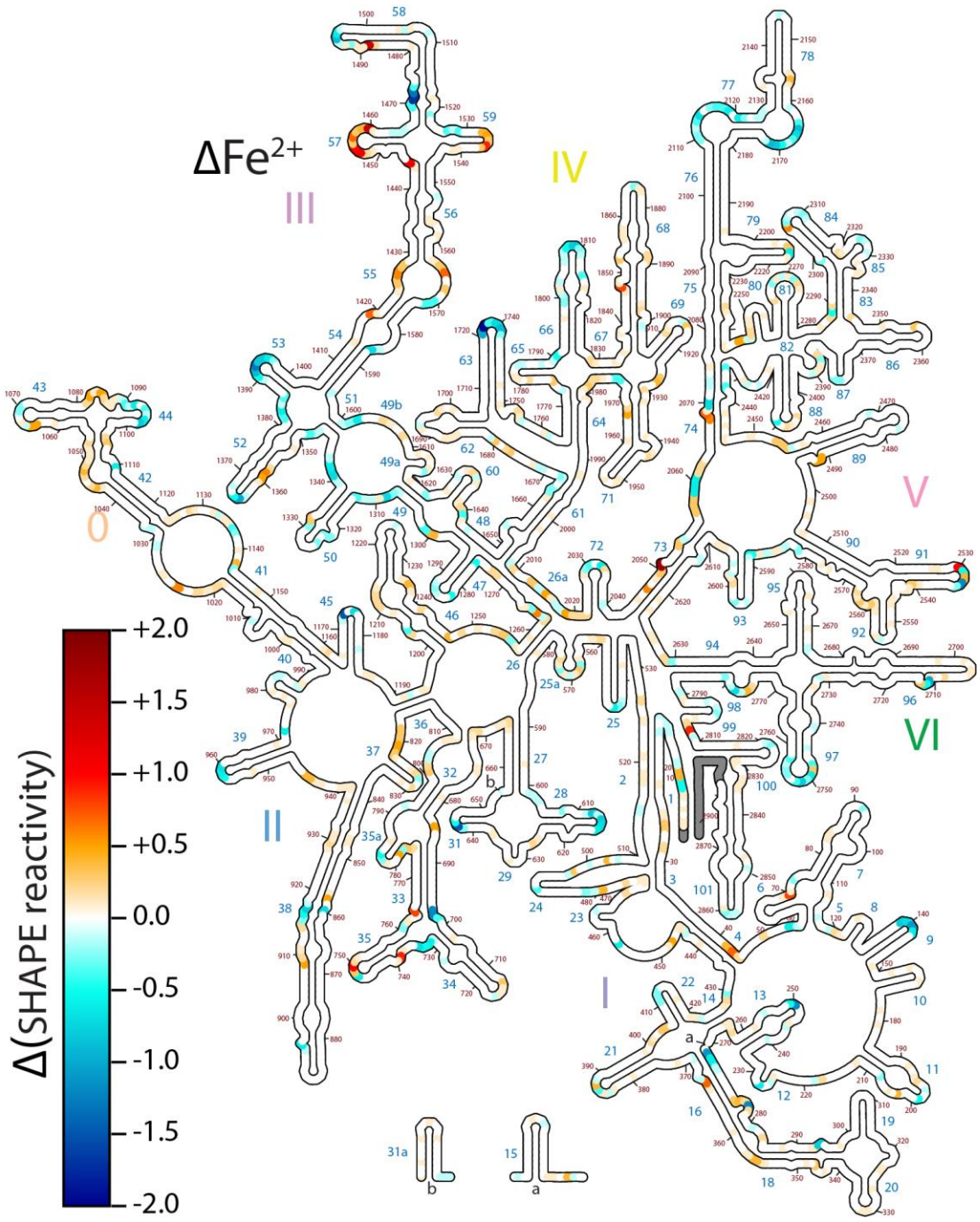


Figure C.4. Fe^{2+}-induced changes in SHAPE reactivity for the LSU rRNA mapped onto secondary structure (compared against SHAPE reactivities in Na^{+}). Positive values indicate nucleotides with increased SHAPE reactivity in presence of 2.5 mM Fe^{2+}, while negative values denote decreased reactivity. Helix numbers (blue), domain numbers, and insertion points are indicated. Nucleotide numbers are denoted every ~10 nt (*E. coli* numbering). Regions where data are not available (5' and 3' ends) are grey. Line segment length is not proportionate to number of nucleotides due to inconsistent nucleotide spacing in the secondary structure. This is a full-page version of Figure 5.3D. Figure generated using RiboVision.

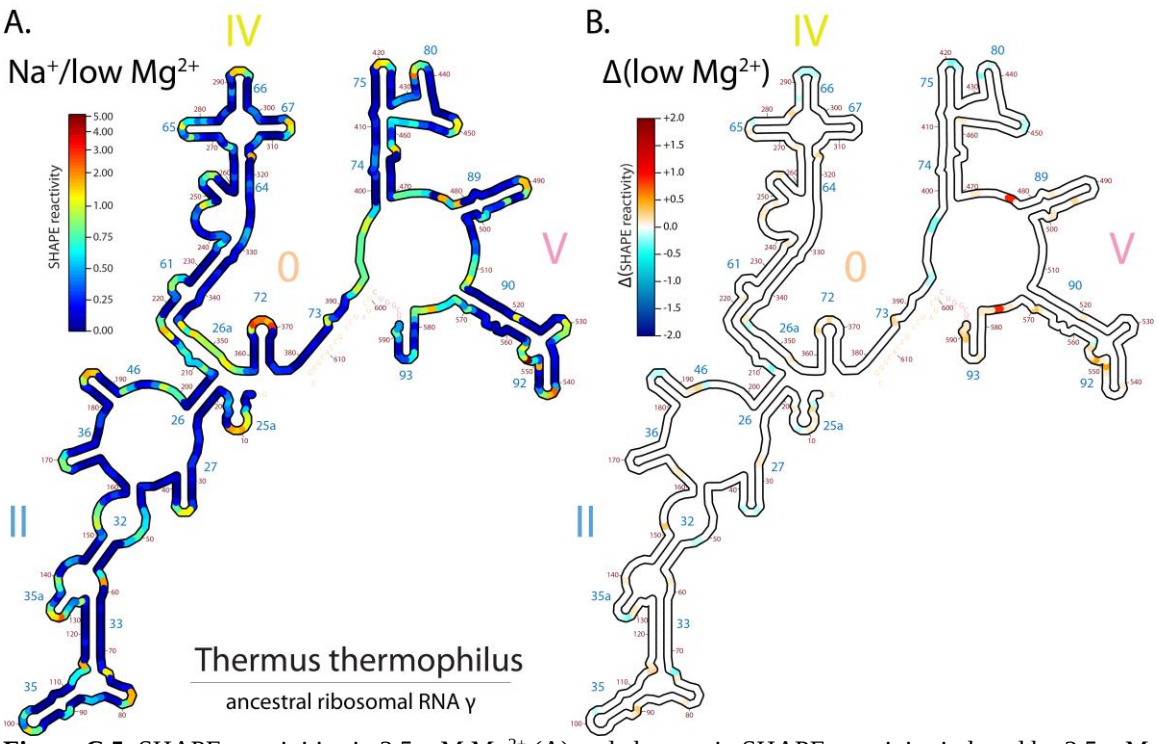


Figure C.5. SHAPE reactivities in 2.5 mM Mg^{2+} (A) and changes in SHAPE reactivity induced by 2.5 mM Mg^{2+} for a-rRNA (B). A) SHAPE reactivities mapped onto predicted a-rRNA secondary structure in *Na^{+}/low Mg^{2+}*. All samples contained 200 mM NaOAc, 50 mM NaHEPES, pH 8, and data was collected under anoxic conditions. B) Induced changes in SHAPE reactivity by 2.5 mM Mg^{2+} mapped onto predicted a-rRNA secondary structure, compared against samples with no divalent cations. Positive values indicate nucleotides with increased SHAPE reactivity in presence of 2.5 mM Mg^{2+}, while negative values denote decreased reactivity. Regions where data are not available (5' and 3' ends) are shown as sequence only. Domain/nucleotide/helix numberings are indicated. Line segment length is not proportionate to number of nucleotides due to inconsistent nucleotide spacing in the secondary structure. Figures generated using RiboVision.

REFERENCES

(1) Rich, A. In *Horizons in Biochemistry*; Kasha, M., Pullman, B., Eds.; New York: Academic: 1962, p 103.

(2) Crick, F. H. *J. Mol. Biol.* **1968**, *38*, 367.

(3) Orgel, L. E. *J. Mol. Biol.* **1968**, *38*, 381.

(4) Gilbert, W. *Nature* **1986**, *319*, 618.

(5) Cech, T. R. *Cell* **2009**, *136*, 599.

(6) Neveu, M.; Kim, H. J.; Benner, S. A. *Astrobiology* **2013**, *13*, 391.

(7) Robertson, M. P.; Joyce, G. F. *Cold Spring Harb Perspect Biol* **2012**, *4*.

(8) Bernhardt, H. S. *Biol Direct* **2012**, *7*, 23.

(9) Higgs, P. G.; Lehman, N. *Nat. Rev. Genet.* **2015**, *16*, 7.

(10) Khaitovich, P.; Mankin, A. S.; Green, R.; Lancaster, L.; Noller, H. F. *Proc. Natl. Acad. Sci. U. S. A.*. **1999**, *96*, 85.

(11) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 905.

(12) Woese, C. R., Personal communication to Francis Crick (1967).

(13) Woese, C. R.; Fox, G. E. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 5088.

(14) Bowman, J. C.; Lenz, T. K.; Hud, N. V.; Williams, L. D. *Curr. Opin. Struct. Biol.* **2012**, *22*, 262.

(15) Brion, P.; Westhof, E. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 113.

(16) Porschke, D. *Nucleic Acids Res.* **1979**, *6*, 883.

(17) Lynch, D. C.; Schimmel, P. R. *Biochemistry* **1974**, *13*, 1841.

(18) Stein, A.; Crothers, D. M. *Biochemistry* **1976**, *15*, 160.

(19) Lindahl, T.; Adams, A.; Fresco, J. R. *Proc. Natl. Acad. Sci. U.S.A.* **1966**, *55*, 941.

(20) Draper, D. E. *Biophys. J.* **2008**, *95*, 5489.

(21) Auffinger, P.; Grover, N.; Westhof, E. *Met Ions Life Sci* **2011**, *9*, 1.

(22) Zheng, H.; Shabalin, I. G.; Handing, K. B.; Bujnicki, J. M.; Minor, W. *Nucleic Acids Res.* **2015**.

(23) Klein, D. J.; Moore, P. B.; Steitz, T. A. *RNA* **2004**, *10*, 1366.

(24) Butcher, S. E. *Met Ions Life Sci* **2011**, *9*, 235.

(25) Johnson-Buck, A. E.; McDowell, S. E.; Walter, N. G. *Met Ions Life Sci* **2011**, *9*, 175.

(26) Brown, I. D. *Acta Crystallogr. Sect. B.* **1992**, *48*, 553.

(27) Rashin, A. A.; Honig, B. *J. Phys. Chem.* **1985**, *89*, 5588.

(28) Maguire, M. E.; Cowan, J. A. *BioMetals* **2002**, *15*, 203.

(29) Bock, C. W.; Markham, G. D.; Katz, A. K.; Glusker, J. P. *Theor. Chem. Acc.* **2006**, *115*, 100.

(30) Baes, C. F.; Mesmer, R. E. *Hydrolysis of Cations*; Wiley: New York, 1976.

(31) Serdyuk, I. N.; Zaccai, N. R.; Zaccai, J. *Methods in Molecular Biophysics: Structure, Dynamics, Function*; Cambridge University Press: Cambridge, UK, 2007.

(32) Diebler, H.; Eigen, M.; Ilgenfritz, G.; Maass, G.; Winkler, R. *Pure Appl. Chem.* **1969**, *20*, 93.

(33) Kankia, B. I. *Biophys. Chem.* **2003**, *104*, 643.

(34) Kankia, B. I. *Biopolymers* **2004**, *74*, 232.

(35) Hsiao, C.; Mohan, S.; Kalahar, B. K.; Williams, L. D. *Mol. Biol. Evol.* **2009**, *26*, 2415.

(36) Petrov, A. S.; Bowman, J. C.; Harvey, S. C.; Williams, L. D. *RNA* **2011**, *17*, 291.

(37) Glendening, E. D. *J. Am. Chem. Soc.* **1996**, *118*, 2473.

(38) Schenter, G. K.; Glendening, E. D. *J. Phys. Chem.* **1996**, *100*, 17152.

(39) Rulisek, L.; Sponer, J. *J. Phys. Chem. B* **2003**, *107*, 1913.

(40) Petrov, A. S.; Pack, G. R.; Lamm, G. *J. Phys. Chem. B* **2004**, *108*, 6072.

(41) Ditzler, M. A.; Otyepka, M.; Sponer, J.; Walter, N. G. *Acc. Chem. Res.* **2010**, *43*, 40.

(42) Takamoto, K.; Das, R.; He, Q.; Doniach, S.; Brenowitz, M.; Herschlag, D.; Chance, M. R. *J. Mol. Biol.* **2004**, *343*, 1195.

(43) Hsiao, C.; Williams, L. D. *Nucleic Acids Res.* **2009**, *37*, 3134.

(44) Hsiao, C.; Tannenbaum, M.; VanDeusen, H.; Hershkovitz, E.; Perng, G.; Tannenbaum, A.; Williams, L. D. In *Nucleic Acid Metal Ion Interactions*; Hud, N., Ed.; The Royal Society of Chemistry: London, 2008, p 1.

(45) Cate, J. H.; Hanna, R. L.; Doudna, J. A. *Nat. Struct. Biol.* **1997**, *4*, 553.

(46) Juneau, K.; Podell, E.; Harrington, D. J.; Cech, T. R. *Structure.* **2001**, *9*, 221.

(47) Toor, N.; Keating, K. S.; Taylor, S. D.; Pyle, A. M. *Science* **2008**, *320*, 77.

(48) Robertson, M. P.; Scott, W. G. *Science* **2007**, *315*, 1549.

(49) Dixon, N.; Duncan, J. N.; Geerlings, T.; Dunstan, M. S.; McCarthy, J. E.; Leys, D.; Micklefield, J. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 2830.

(50) Athavale, S. S.; Petrov, A. S.; Hsiao, C.; Watkins, D.; Prickett, C. D.; Gossett, J. J.; Lie, L.; Bowman, J. C.; O'Neill, E.; Bernier, C. R.; Hud, N. V.; Wartell, R. M.; Harvey, S. C.; Williams, L. D. *PLoS ONE* **2012**, *7*, e38024.

(51) Hsiao, C.; Chou, I.-C.; Okafor, C. D.; Bowman, J. C.; O'Neill, E. B.; Athavale, S. S.; Petrov, A. S.; Hud, N. V.; Wartell, R. M.; Harvey, S. C.; Williams, L. D. *Nature Chemistry* **2013**, *5*, 525.

(52) Popovic, M.; Fliss, P. S.; Ditzler, M. A. *Nucleic Acids Res.* **2015**.

(53) Anbar, A. D. *Science* **2008**, *322*, 1481.

(54) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. *Nucleic Acids Res.* **2004**, *32*, W665.

(55) Merino, E. J.; Wilkinson, K. A.; Coughlan, J. L.; Weeks, K. M. *J. Am. Chem. Soc.* **2005**, *127*, 4223.

(56) Woodson, S. A. *Acc. Chem. Res.* **2011**, *44*, 1312.

(57) Schlatterer, J. C.; Brenowitz, M. *Methods (San Diego, Calif.)* **2009**, *49*, 142.

(58) Pollack, L. *Annual review of biophysics* **2011**, *40*, 225.

(59) Pabit, S. A.; Qiu, X.; Lamb, J. S.; Li, L.; Meisburger, S. P.; Pollack, L. *Nucleic Acids Res.* **2009**, *37*, 3887.

(60) Chen, H.; Meisburger, S. P.; Pabit, S. A.; Sutton, J. L.; Webb, W. W.; Pollack, L. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 799.

(61) Moghaddam, S.; Caliskan, G.; Chauhan, S.; Hyeon, C.; Briber, R. M.; Thirumalai, D.; Woodson, S. A. *J. Mol. Biol.* **2009**, *393*, 753.

(62) Fiore, J. L.; Holmstrom, E. D.; Nesbitt, D. J. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 2902.

(63) Koutmou, K. S.; Casiano-Negroni, A.; Getz, M. M.; Pazicni, S.; Andrews, A. J.; Penner-Hahn, J. E.; Al-Hashimi, H. M.; Fierke, C. A. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 2479.

(64) Christian, E. L.; Anderson, V. E.; Carey, P. R.; Harris, M. E. *Biochemistry* **2010**, *49*, 2869.

(65) Roh, J. H.; Tyagi, M.; Briber, R. M.; Woodson, S. A.; Sokolov, A. P. *J. Am. Chem. Soc.* **2011**, *133*, 16406.

(66) Anthony, P. C.; Sim, A. Y.; Chu, V. B.; Doniach, S.; Block, S. M.; Herschlag, D. *J. Am. Chem. Soc.* **2012**, *134*, 4607.

(67) Gresh, N.; Sponer, J. E.; Spackova, N.; Leszczynski, J.; Sponer, J. *J. Phys. Chem. B* **2003**, *107*, 8669.

(68) Petrov, A. S.; Lamm, G.; Pack, G. R. *Biopolymers* **2005**, *77*, 137.

(69) Heilman-Miller, S. L.; Thirumalai, D.; Woodson, S. A. *J. Mol. Biol.* **2001**, *306*, 1157.

(70) Chu, V. B.; Herschlag, D. *Curr. Opin. Struct. Biol.* **2008**, *18*, 305.

(71) Tan, Z. J.; Chen, S. J. *Biophys. J.* **2010**, *99*, 1565.

(72) Tan, Z. J.; Chen, S. J. *Methods Enzymol.* **2009**, *469*, 465.

(73) Steitz, T. A. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 242.

(74) Trappl, K.; Polacek, N. *Met Ions Life Sci* **2011**, *9*, 253.

(75) Selmer, M.; Dunham, C. M.; Murphy, F. V.; Weixlbaumer, A.; Petry, S.; Kelley, A. C.; Weir, J. R.; Ramakrishnan, V. *Science* **2006**, *313*, 1935.

(76) Harms, J.; Schluenzen, F.; Zarivach, R.; Bashan, A.; Gat, S.; Agmon, I.; Bartels, H.; Franceschi, F.; Yonath, A. *Cell* **2001**, *107*, 679.

(77) Ben-Shem, A.; Jenner, L.; Yusupova, G.; Yusupov, M. *Science* **2010**, *330*, 1203.

(78) Jenner, L.; Melnikov, S.; de Loubresse, N. G.; Ben-Shem, A.; Iskakova, M.; Urzhumtsev, A.; Meskauskas, A.; Dinman, J.; Yusupova, G.; Yusupov, M. *Curr. Opin. Struct. Biol.* **2012**, *22*, 759.

(79) Yi, Q. M.; Wong, K. P. *Biochem. Biophys. Res. Commun.* **1982**, *104*, 733.

(80) Noller, H. F.; Hoffarth, V.; Zimniak, L. *Science* **1992**, *256*, 1416.

(81) Steitz, T. A.; Moore, P. B. *Trends Biochem. Sci* **2003**, *28*, 411.

(82) Moore, P. B.; Steitz, T. A. *Annu. Rev. Biochem* **2003**, *71*, 813.

(83) Simonovic, M.; Steitz, T. A. *Biochim. Biophys. Acta* **2009**, *1789*, 612.

(84) Lang, K.; Erlacher, M.; Wilson, D. N.; Micura, R.; Polacek, N. *Chem. Biol.* **2008**, *15*, 485.

(85) Nakatogawa, H.; Ito, K. *Cell* **2002**, *108*, 629.

(86) Voss, N. R.; Gerstein, M.; Steitz, T. A.; Moore, P. B. *J. Mol. Biol.* **2006**, *360*, 893.

(87) Fox, G. E.; Tran, Q.; Yonath, A. *Astrobiology* **2012**, *12*, 57.

(88) Orgel, L. E. *Orig. Life Evol. Biosph.* **2003**, *33*, 211.

(89) Crick, F. *Nature* **1970**, *226*, 561.

(90) Seehafer, C.; Kalweit, A.; Steger, G.; Graf, S.; Hammann, C. *RNA* **2011**, *17*, 21.

(91) Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. *Cell* **1982**, *31*, 147.

(92) Cech, T. R. *Biochem. Soc. Trans.* **2002**, *30*, 1162.

(93) Pyle, A. M. *Science* **1993**, *261*, 709.

(94) Pley, H. W.; Flaherty, K. M.; McKay, D. B. *Nature* **1994**, *372*, 68.

(95) Vaidya, N.; Manapat, M. L.; Chen, I. A.; Xulvi-Brunet, R.; Hayden, E. J.; Lehman, N. *Nature* **2012**, *491*, 72.

(96) Lohse, P. A.; Szostak, J. W. *Nature.* **1996**, *381*, 442.

(97) Turk, R. M.; Chumachenko, N. V.; Yarus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 4585.

(98) Noller, H. F. *Cold Spring Harb. Perspect. Biol.* **2010**, *7*, 7.

(99) Lu, Y.; Freeland, S. *Genome Biol* **2006**, *7*, 102.

(100) Khorana, H. G. *Federation proceedings* **1965**, *24*, 1473.

(101) Hassouna, N.; Michot, B.; Bachellerie, J. P. *Nucleic Acids Res.* **1984**, *12*, 3563.

(102) Gerbi, S. A. In *Ribosomal RNA—Structure, evolution, processing, and function in protein synthesis*; Zimmermann, R. A., Dahlberg, A. E., Eds.; CRC Press: Boca Raton, FL, 1996, p 71.

(103) Michot, B.; Qu, L.-H.; Bachellerie, J.-P. *Eur. J. Biochem.* **1990**, *188*, 219.

(104) Wolf, Y. I.; Koonin, E. V. *Biol. Direct* **2007**, *2*, 14.

(105) Fournier, G. P.; Neumann, J. E.; Gogarten, J. P. *PLoS One* **2010**, *5*, e9437.

(106) Schmeing, T. M.; Ramakrishnan, V. *Nature* **2009**, *461*, 1234.

(107) Rodnina, M. V.; Beringer, M.; Wintermeyer, W. *Trends Biochem. Sci* **2007**, *32*, 20.

(108) Mears, J. A.; Cannone, J. J.; Stagg, S. M.; Gutell, R. R.; Agrawal, R. K.; Harvey, S. C. *J. Mol. Biol.* **2002**, *321*, 215.

(109) Bokov, K.; Steinberg, S. V. *Nature* **2009**, *457*, 977.

(110) Belousoff, M. J.; Davidovich, C.; Zimmerman, E.; Caspi, Y.; Wekselman, I.; Rozenszajn, L.; Shapira, T.; Sade-Falk, O.; Taha, L.; Bashan, A.; Weiss, M. S.; Yonath, A. *Biochem. Soc. Trans.* **2010**, *38*, 422.

(111) Hury, J.; Nagaswamy, U.; Larios-Sanz, M.; Fox, G. E. *Orig Life Evol Biosph* **2006**, *36*, 421.

(112) Benner, S. A.; Ellington, A. D.; Tauer, A. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 7054.

(113) Anantharaman, V.; Koonin, E. V.; Aravind, L. *Nucleic Acids Res.* **2002**, *30*, 1427.

(114) Weeks, K. M.; Mauger, D. M. *Acc. Chem. Res.* **2011**, null.

(115) Mortimer, S. A.; Weeks, K. M. *J. Am. Chem. Soc.* **2007**, *129*, 4144.

(116) Mortimer, S. A.; Weeks, K. M. *J. Am. Chem. Soc.* **2008**, *130*, 16178.

(117) Vasa, S. M.; Guex, N.; Wilkinson, K. A.; Weeks, K. M.; Giddings, M. C. *RNA* **2008**, *14*, 1979.

(118) Athavale, S. S.; Gossett, J. J.; Hsiao, C.; Bowman, J. C.; O'Neill, E.; Hershkovitz, E.; Preeprem, T.; Hud, N. V.; Wartell, R. M.; Harvey, S. C.; Williams, L. D. *RNA* **2012**, *18*, 752.

(119) Yoon, S.; Kim, J.; Hum, J.; Park, S.; Kladwang, W.; Das, R. *Bioinformatics* **2011**, *27*, 1798.

(120) Wilkinson, K. A.; Vasa, S. M.; Deigan, K. E.; Mortimer, S. A.; Giddings, M. C.; Weeks, K. M. *RNA* **2009**, *15*, 1314.

(121) Vicens, Q.; Gooding, A. R.; Laederach, A.; Cech, T. R. *RNA* **2007**, *13*, 536.

(122) Mortimer, S. A.; Weeks, K. M. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15622.

(123) Wang, B.; Wilkinson, K. A.; Weeks, K. M. *Biochemistry* **2008**, *47*, 3454.

(124) Hennelly, S. P. H. S. P.; Sanbonmatsu, K. Y. *Nucleic Acids Res.* **2011**, *39*, 2416.

(125) Watts, J. M.; Dang, K. K.; Gorelick, R. J.; Leonard, C. W.; Bess Jr, J. W.; Swanstrom, R.; Burch, C. L.; Weeks, K. M. *Nature* **2009**, *460*, 711.

(126) Athavale, S. S.; Gossett, J. J.; Bowman, J. C.; Hud, N. V.; Williams, L. D.; Harvey, S. C. *PLoS One* **2013**, *8*, e54384.

(127) Purzycka, K. J.; Pachulska-Wieczorek, K.; Adamiak, R. W. *Nucleic Acids Res.* **2011**, *39*, 7234.

(128) Spitale, R. C.; Crisalli, P.; Flynn, R. A.; Torre, E. A.; Kool, E. T.; Chang, H. Y. *Nat. Chem. Biol.* **2013**, *9*, 18.

(129) Deigan, K. E.; Li, T. W.; Mathews, D. H.; Weeks, K. M. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 97.

(130) Low, J. T.; Weeks, K. M. *Methods (San Diego, Calif.)* **2010**, *52*, 150.

(131) Kladwang, W.; VanLang, C. C.; Cordero, P.; Das, R. *Biochemistry* **2011**, null.

(132) Kladwang, W.; Das, R. *Biochemistry* **2010**, *49*, 7414.

(133) Kladwang, W.; Cordero, P.; Das, R. *RNA* **2011**, *17*, 522.

(134) Tian, S.; Cordero, P.; Kladwang, W.; Das, R. *RNA* **2014**.

(135) Barta, A.; Steiner, G.; Brosius, J.; Noller, H. F.; Kuechler, E. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 3607.

(136) Leshin, J. A.; Heselpoth, R.; Belew, A. T.; Dinman, J. D. *Rna Biology* **2011**, *8*, 478.

(137) Lavender, C. A.; Lorenz, R.; Zhang, G.; Tamayo, R.; Hofacker, I. L.; Weeks, K. M. *PLoS Comput Biol* **2015**, *11*, e1004126.

(138) Blouin, S.; Chinnappan, R.; Lafontaine, D. A. *Nucleic Acids Res.* **2011**, *39*, 3373.

(139) Dann, C. E., 3rd; Wakeman, C. A.; Sieling, C. L.; Baker, S. C.; Irnov, I.; Winkler, W. C. *Cell* **2007**, *130*, 878.

(140) Dann, C. E.; Wakeman, C. A.; Sieling, C. L.; Baker, S. C.; Irnov, I.; Winkler, W. C. *Cell* **2007**, *130*, 878.

(141) Hennelly, S. P.; Novikova, I. V.; Sanbonmatsu, K. Y. *Nucleic Acids Res.* **2013**, *41*, 1922.

(142) Tyrrell, J.; McGinnis, J. L.; Weeks, K. M.; Pielak, G. J. *Biochemistry* **2013**, Ahead of Print.

(143) Strulson, C. A.; Boyer, J. A.; Whitman, E. E.; Bevilacqua, P. C. *RNA* **2014**, *20*, 331.

(144) Moss, W. N.; Dela-Moss, L. I.; Kierzek, E.; Kierzek, R.; Priore, S. F.; Turner, D. H. *Plos One* **2012**, *7*.

(145) Lozano, G.; Fernandez, N.; Martinez-Salas, E. *FEBS J.* **2014**, *281*, 3685.

(146) OSHIMA, T.; IMAHORI, K. *Int. J. Syst. Evol. Microbiol.* **1974**, *24*, 102.

(147) http://www.straininfo.net/genomes/13202, 2004; Vol. 2015.

(148) Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. *Cell* **1982**, *31*, 147.

(149) Latham, J. A.; Cech, T. R. *Science* **1989**, *245*, 276.

(150) Butcher, S. E.; Dieckmann, T.; Feigon, J. *EMBO J.* **1997**, *16*, 7490.

(151) Deras, M. L.; Brenowitz, M.; Ralston, C. Y.; Chance, M. R.; Woodson, S. A. *Biochemistry* **2000**, *39*, 10975.

(152) Silverman, S. K.; Deras, M. L.; Woodson, S. A.; Scaringe, S. A.; Cech, T. R. *Biochemistry* **2000**, *39*, 12465.

(153) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1678.

(154) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Szewczak, A. A.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1696.

(155) Hutchison, C. A., 3rd; Phillips, S.; Edgell, M. H.; Gillam, S.; Jahnke, P.; Smith, M. *J. Biol. Chem.* **1978**, *253*, 6551.

(156) Okafor, C. D.; Vander Wood, D.; O'Neill, E.; Williams, L. D.; Hud, N. *TBD* **2015**, *In preparation*.

(157) Hsiao, C.; Lenz, T. K.; Peters, J. K.; Fang, P. Y.; Schneider, D. M.; Anderson, E. J.; Preeprem, T.; Bowman, J. C.; O'Neill, E. B.; Lie, L.; Athavale, S. S.; Gossett, J. J.; Trippe, C.; Murray, J.; Petrov, A. S.; Wartell, R. M.; Harvey, S. C.; Hud, N. V.; Williams, L. D. *Nucleic Acids Res.* **2013**, *41*, 3373.

(158) Woese, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8392.

(159) Woese, C. R. *RNA* **2001**, *7*, 1055.

(160) Williams, D.; Fournier, G. P.; Lapierre, P.; Swithers, K. S.; Green, A. G.; Andam, C. P.; Gogarten, J. P. *Biol. Direct* **2011**, *6*, 45.

(161) Fox, G. E. *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a003483.

(162) Woese, C. R. *The genetic code: the molecular basis for genetic expression*; Harper & Row: N.Y., 1967.

(163) Rost, B. *Fold Des.* **1997**, *2*, S19.

(164) Heinz, D. W.; Baase, W. A.; Zhang, X. J.; Blaber, M.; Dahlquist, F. W.; Matthews, B. W. *J. Mol. Biol.* **1994**, *236*, 869.

(165) Klein, D. J.; Moore, P. B.; Steitz, T. A. *J. Mol. Biol.* **2004**, *340*, 141.

(166) Petrov, A. S.; Bernier, C. R.; Hsiao, C. L.; Okafor, C. D.; Tannenbaum, E.; Stern, J.; Gaucher, E.; Schneider, D.; Hud, N. V.; Harvey, S. C.; Williams, L. D. *J. Phys. Chem. B* **2012**, *116*, 8113.

(167) Bowman, J. C.; Azizi, B.; Lenz, T. K.; Roy, P.; Williams, L. D. In *Recombinant and In Vitro RNA Synthesis: Methods and Protocols, Methods in Molecular Biology*; Conn, G. L., Ed.; Springer Science, LLC: 2012; Vol. 941, p 19.

(168) Stemmer, W. P. C.; Crameri, A.; Ha, K. D.; Brennan, T. M.; Heyneker, H. L. *Gene* **1995**, *164*, 49.

(169) Ouyang, Q.; Kaplan, P. D.; Liu, S.; Libchaber, A. *Science* **1997**, *278*, 446.

(170) Smith, H. O.; Hutchison, C. A., 3rd; Pfannkoch, C.; Venter, J. C. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15440.

(171) Higuchi, R.; Krummel, B.; Saiki, R. K. *Nucleic Acids Res.* **1988**, *16*, 7351.

(172) Prodromou, C.; Pearl, L. H. *Protein Eng.* **1992**, *5*, 827.

(173) Sandhu, G. S.; Aleff, R. A.; Kline, B. C. *BioTechniques* **1992**, *12*, 14.

(174) Singh, P. K.; Sarangi, B. K.; Tuli, R. *J. Biosci. (Bangalore)* **1996**, *21*, 735.

(175) Gurevich, A. I.; Esipov, R. S.; Kayushin, A. L.; Korosteleva, M. D. *Bioorg. Khim* **1997**, *23*, 492.

(176) Wilkinson, K. A.; Merino, E. J.; Weeks, K. M. *Nature protocols* **2006**, *1*, 1610.

(177) Noller, H. F.; Kop, J.; Wheaton, V.; Brosius, J.; Gutell, R. R.; Kopylov, A. M.; Dohme, F.; Herr, W.; Stahl, D. A.; Gupta, R.; Woese, C. R. *Nucleic Acids Res.* **1981**, *9*, 6167.

(178) Mohan, S.; Hsiao, C.; Bowman, J. C.; Wartell, R.; Williams, L. D. *J. Am. Chem. Soc.* **2010**, *132*, 12679.

(179) Tinoco, I., Jr.; Bustamante, C. *J. Mol. Biol.* **1999**, *293*, 271.

(180) Anderson, R. M.; Kwon, M.; Strobel, S. A. *J Mol Evol.* **2007**, *64*, 472.

(181) Krupkin, M.; Matzov, D.; Tang, H.; Metz, M.; Kalaora, R.; Belousoff, M. J.; Zimmerman, E.; Bashan, A.; Yonath, A. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **2011**, *366*, 2972.

(182) Ostergaard, P.; Phan, H.; Johansen, L. B.; Egebjerg, J.; Ostergaard, L.; Porse, B. T.; Garrett, R. A. *J. Mol. Biol.* **1998**, *284*, 227.

(183) Pauling, L.; Zuckerkandl, E. *Acta Chem. Scand.* **1963**, *17*, 9.

(184) Rich, A. In *Chemical Evoltuion and the Origin of Life*; Buvet, R., Ponnamperuma, C., Eds.; North-Holland Publishing Company: Amsterdam, 1971.

(185) Petrov, A. S.; Bernier, C. R.; Hershkovitz, E.; Xue, Y.; Waterbury, C. C.; Grover, M. A.; C., H. S.; Hud, N. V.; Wartell, R. M.; Williams, L. D. *Nucleic Acids Res.* **2013**, *41*, 7522.

(186) Khaitovich, P.; Tenson, T.; Mankin, A. S.; Green, R. *RNA* **1999**, *5*, 605.

(187) Petrov, A. S.; Bernier, C. R.; Gulen, B.; Waterbury, C. C.; Hershkovitz, E.; Hsiao, C.; Harvey, S. C.; Hud, N. V.; Fox, G. E.; Wartell, R. M.; Williams, L. D. *PLoS One* **2014**, *9*, e88222.

(188) Bernier, C.; Petrov, A. S.; Waterbury, C.; Jett, J.; Li, F.; Freil, L. E.; Xiong, b.; Wang, L.; Le, A.; Milhouse, B. L.; Hershkovitz, E.; Grover, M.; Xue, Y.; Hsiao, C.; Bowman, J. C.; Harvey, S. C.; Wartel, J. Z.; Williams, L. D. *Faraday Discuss.* **2014**, *169*, 195.

(189) Sarver, M.; Zirbel, C. L.; Stombaugh, J.; Mokdad, A.; Leontis, N. B. *J. Math. Biol.* **2008**, *56*, 215.

(190) Liiv, A.; O'Connor, M. *J. Biol. Chem.* **2006**, *281*, 29850.

(191) Nikulin, A.; Eliseikina, I.; Tishchenko, S.; Nevskaya, N.; Davydova, N.; Platonova, O.; Piendl, W.; Selmer, M.; Liljas, A.; Drygin, D.; Zimmermann, R.; Garber, M.; Nikonov, S. *Nat. Struct. Mol. Biol.* **2003**, *10*, 104.

(192) Frederiksen, J. K.; Li, N. S.; Das, R.; Herschlag, D.; Piccirilli, J. A. *RNA* **2012**, *18*, 1123.

(193) Schulze, H.; Nierhaus, K. H. *EMBO J.* **1982**, *1*, 609.

(194) Noller, H. F. *J. Bacteriol.* **1993**, *175*, 5297.

(195) Brown, I. D. *Acta Crystallogr. Sect. B.* **1988**, *44*, 545.

(196) Wulfsberg, G. *Principles of Descriptive Inorganic Chemistry*; University Science Books: Sausalito, CA, 1991.

(197) Uudsemaa, M.; Tamm, T. *Chem. Phys. Lett.* **2004**, *400*, 54.

(198) Johnson, D. C.; Dean, D. R.; Smith, A. D.; Johnson, M. K. *Annu. Rev. Biochem* **2005**, *74*, 247.

(199) Noodleman, L.; Lovell, T.; Liu, T.; Himo, F.; Torres, R. A. *Curr. Opin. Chem. Biol.* **2002**, *6*, 259.

(200) Valko, M.; Morris, H.; Cronin, M. T. D. *Curr. Med. Chem.* **2005**, *12*, 1161.

(201) Theil, E. C.; Goss, D. J. *Chem. Rev.* **2009**, *109*, 4568.

(202) Liu, X. W.; Millero, F. J. *Mar. Chem.* **2002**, *77*, 43.

(203) Dupont, C. L.; Butcher, A.; Valas, R. E.; Bourne, P. E.; Caetano-Anolles, G. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 10567.

(204) Dupont, C. L.; Yang, S.; Palenik, B.; Bourne, P. E. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17822.

(205) Harel, A.; Bromberg, Y.; Falkowski, P. G.; Bhattacharya, D. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 7042.

(206) Holland, H. D. *Economic Geology* **1973**, *68*, 1169.

(207) Hazen, R. M.; Ferry, J. M. *Elements* **2010**, *6*, 9.

(208) Holland, H. D. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **2006**, *361*, 903.

(209) Klein, C. *Am. Mineral.* **2005**, *90*, 1473.

(210) Prousek, J. *Pure Appl. Chem.* **2007**, *79*, 2325.

(211) Kozlowski, H.; Kolkowska, P.; Watly, J.; Krzywoszynska, K.; Potocki, S. *Curr. Med. Chem.* **2014**, *21*, 3721.

(212) Aguirre, J. D.; Culotta, V. C. *J. Biol. Chem.* **2012**, *287*, 13541.

(213) Ushizaka, S.; Kuma, K.; Suzuki, K. *Fish. Sci.* **2011**, *77*, 411.

(214) Anjem, A.; Varghese, S.; Imlay, J. A. *Mol. Microbiol.* **2009**, *72*, 844.

(215) Martin, J. E.; Imlay, J. A. *Mol. Microbiol.* **2011**, *80*, 319.

(216) Cotruvo, J. A.; Stubbe, J. *Annu. Rev. Biochem* **2011**, *80*, 733.

(217) Wolfe-Simon, F.; Starovoytov, V.; Reinfelder, J. R.; Schofield, O.; Falkowski, P. G. *Plant Physiol.* **2006**, *142*, 1701.

(218) Torrents, E.; Aloy, P.; Gibert, I.; Rodriguez-Trelles, F. *J. Mol. Evol.* **2002**, *55*, 138.

(219) Dupont, C. L.; Butcher, A.; Valas, R. E.; Bourne, P. E.; Caetano-Anollés, G. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 10567.

(220) Berens, C.; Streicher, B.; Schroeder, R.; Hillen, W. *Chem. Biol.* **1998**, *5*, 163.

(221) Popović, M.; Fliss, P. S.; Ditzler, M. A. *Nucleic Acids Res.* **2015**, epub ahead of print.

(222) Sen, D.; Poon, L. C. *Crit. Rev. Biochem. Mol. Biol.* **2011**, *46*, 478.

(223) Wimberly, B. T.; Brodersen, D. E.; Clemons, W. M., Jr.; Morgan-Warren, R. J.; Carter, A. P.; Vonrhein, C.; Hartsch, T.; Ramakrishnan, V. *Nature* **2000**, *407*, 327.

(224) Schuwirth, B. S.; Borovinskaya, M. A.; Hau, C. W.; Zhang, W.; Vila-Sanjurjo, A.; Holton, J. M.; Cate, J. H. *Science* **2005**, *310*, 827.

(225) Honda, K.; Smith, M. A.; Zhu, X.; Baus, D.; Merrick, W. C.; Tartakoff, A. M.; Hattier, T.; Harris, P. L.; Siedlak, S. L.; Fujioka, H.; Liu, Q.; Moreira, P. I.; Miller, F. P.; Nunomura, A.; Shimohama, S.; Perry, G. *J. Biol. Chem.* **2005**, *280*, 20978.

(226) Petrov, A. S.; Bernier, C. R.; Hsiao, C.; Norris, A. M.; Kovacs, N. A.; Waterbury, C. C.; Stepanov, V. G.; Harvey, S. C.; Fox, G. E.; Wartell, R. M.; Hud, N. V.; Williams, L. D. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 10251.

(227) Blyn, L. B.; Risen, L. M.; Griffey, R. H.; Draper, D. E. *Nucleic Acids Res.* **2000**, *28*, 1778.

(228) Leontis, N. B.; Lescoute, A.; Westhof, E. *Curr. Opin. Struct. Biol.* **2006**, *16*, 279.

(229) Leontis, N. B.; Westhof, E. *Curr. Opin. Struct. Biol.* **2003**, *13*, 300.

(230) Xin, Y.; Laing, C.; Leontis, N. B.; Schlick, T. *RNA* **2008**, *14*, 2465.

(231) Leontis, N. B.; Westhof, E. *RNA* **2001**, *7*, 499.

(232) Leontis, N. B.; Westhof, E. *J. Mol. Biol.* **1998**, *283*, 571.

(233) Correll, C. C.; Swinger, K. *RNA* **2003**, *9*, 355.

(234) Hsiao, C.; Mohan, S.; Hershkovitz, E.; Tannenbaum, A.; Williams, L. D. *Nucleic Acids Res.* **2006**, *34*, 1481.

(235) Apostolico, A.; Ciriello, G.; Guerra, C.; Heitsch, C. E.; Hsiao, C.; Williams, L. D. *Nucleic Acids Res.* **2009**, *37*, e29.

(236) Petrov, A. I.; Zirbel, C. L.; Leontis, N. B. *RNA* **2013**, *19*, 1327.

(237) Gherghe, C. M.; Shajani, Z.; Wilkinson, K. A.; Varani, G.; Weeks, K. M. *J. Am. Chem. Soc.* **2008**, *130*, 12244.

(238) Hershkovitz, E.; Sapiro, G.; Tannenbaum, A.; Williams, L. D. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **2006**, *3*, 33.

(239) Hershkovitz, E.; Tannenbaum, E.; Howerton, S. B.; Sheth, A.; Tannenbaum, A.; Williams, L. D. *Nucleic Acids Res.* **2003**, *31*, 6249.

VITA

TIMOTHY KENNETH LENZ

TIMOTHY was born and raised in Brantford, Ontario, Canada. He attended high school at St. John's College in Brantford, where he took every available science course. He received a Bachelor of Arts & Science from McMaster University in Hamilton, Ontario in 2007, with honours in biochemistry. At McMaster, he participated in a senior biochemistry research project under Dr. Joaquin Ortega, and completed a multimedia thesis project under Dr. Anne Savage entitled "A Mirror Darkly: The Cold War consciousness as reflected in Western science fiction". After taking time off to be a 'professional Canadian' for a year at Walt Disney World, he returned to academics in 2009 to pursue a PhD in biochemistry at the Georgia Institute of Technology. In his first term, Tim joined the NASA Astrobiology Institute's Center for Ribosomal Origins and Evolution under PI Dr. Loren Williams and Co-PI Dr. Nicholas Hud. During his time at Georgia Tech, Tim has greatly enjoyed communicating and conversing about science with his peers at conferences across North America. He served on the organizing committee for the 2013 Astrobiology Graduate Conference, and received a second place award in the 2015 Astrobiology Science Conference student poster competition. Outside of his professional life, Tim has a wide variety of interests including experimenting in the kitchen, collecting comic books, and spending time with his wife, family, and friends.