COMPUTATIONAL BIOINFORMATICS ON THREE-DIMENSIONAL STRUCTURES OF RIBOSOMES USING MULTIRESOLUTIONAL ANALYSIS

A Dissertation
Presented to
The Academic Faculty

by

Chiaolong Hsiao

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

Georgia Institute of Technology
December 2008

COMPUTATIONAL BIOINFORMATICS ON THREE-DIMENSIONAL STRUCTURES OF RIBOSOMES USING MULTIRESOLUTIONAL ANALYSIS

Approved by:

Dr. Loren D. Williams, Advisor
School of Chemistry and 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. Nicholas V. Hud
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Donald F. Doyle
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: August 12, 2008

ACKNOWLEDGEMENTS

I express my deepest gratitude to my advisor, Dr. Loren Dean Williams, for his incredible support, guidance, and encouragement. Without Loren, it would not be possible for me to come this far.

I am grateful to my committee members, Drs. Stephen C. Harvey, Nicholas V. Hud, Roger M. Wartell, and Donald F. Doyle.

People I would also like to thank: Drs. Mark Borodovsky, Eli Hershkowitz, Seiji Komeda, Tinoush Moulai, and Mary Peek; Srividya Mohan, T Maehigashi, and Derrick Watkins.

Lastly I would like to thank my parents, Mao-Hsiung and Su-Hui, and my wife, I-Chun for their unconditional support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ADP	Adenosine Diphosphate
AOCN	Averaged Observed Coordination Numbers
ATP	Adenosine Triphosphate
cryo-EM	Cryo-Electron Microscopy
DevLS	Deviations of Local Structure
DNA	Deoxyribonucleic acid
EDTA	Ethylene Diamine Tetra-Acetic Acid
HM^{23S}	H. marismortui 23S rRNA
LPTL	Lonepair Triloop
LSU	Large Subunit of Ribosome
LUCA	Last Universal Common Ancestor
Mg^{2+}-μc	Magnesium Microcluster
NMR	Nuclear Magnetic Resonance
PTC	Peptidyl Transferase Center
RMSD	Root Mean Square Deviation
RNA	Ribonucleic Acid
rRNAs	Ribosomal RNA
SAAS	Structural Alignment by Anchored Segments
SAP	Segment Alignment Pairs
TT^{23S}	T. thermophilus 23S rRNA

SUMMARY

This thesis presents my work on deciphering, exploring, and discovering the treasure troves of RNA structural bioinformatics, mainly in the areas of multi-resolution analysis of RNA structure.

RNA is amazing. We found that without changing the backbone connectivity, RNA can maintain structural conservation in 3D via topology switches, at a single residue level. I developed a method of representing RNA structure in multiresolution, called the PBR approach (P stands for Phosphate; B stands for Base; R stands for Ribose). In this method, structural data is viewed through a series of resolutions from finest to coarsest. At a single nucleotide resolution (fine resolution), RNA is abstruse and elaborate with structural insertions/deletions, strand clips, and 3,2-switches. The compilation of structural deviations of RNA, called DevLS (Deviations of Local Structure), provides a new descriptive language of RNA structure, allowing one to systematize and investigate RNA structure. At PBR resolution (coarse resolution), fundamental RNA architecture, e.g. A-helix, tetraloop, Kink-turns, E-loop motifs etc., becomes readily observable.

Using PBR analysis, a total of 103 tetraloops within the crystal structures of the 23s rRNA of *H. marismortui* (PDB entry: 1JJ2) and the 70s rRNA of *T. thermophilus* (PDB entry: 2J00 and 2J01) are found and classified. Combining them, I constructed a 'tetraloop family tree', using a tree formalism, to unify and re-define the tetraloop motif and to represent relationships between tetraloops, as grouped by DevLS.

To date, structural alignment of very large RNAs remains challenge due to the large size, intricate backbone choreography, and tertiary interactions. To overcome these

obstacles, I developed a concept of structural anchors along with a '*Divide and Conquer*' strategy for performing superimposition of 23s rRNAs. Here I use tetraloops as structural anchors. The successful alignment and superimpositions of the 23s rRNAs of *T. thermophilus* and *H. marismortui* gives an overall RMSD of atomic positions of 1.2 Å. This superimposition utilizes 73% of RNA backbone atoms (around 2129 residues). This accurate superimposition allows me to identify regions of structural conservation and diversity, to determine relationships between structural and sequence variation, and to investigate structural relationships between RNA to RNA, RNA to ions, ions to ions, at atomic resolution.

By using principles of inorganic chemistry along with structural alignment technique as described above, a recurrent magnesium-binding motif in large RNAs (the 23S rRNAs from *H. marismortui* and *T. thermophilus*, the P4-P6 domain of the tetrahymena Group I intron ribozyme, and a Group II intron ribozyme) is revealed. These magnesium-binding motifs play a critical role in the framework of the Peptidyl Transferase Center of the ribosome by their locations, topologies, and coordination geometries. Features of magnesium-binding motif include (i) bridging phosphate chelation of two magnesium ions in the form of $Mg^{2+}_{(i)}$-(O1P-P-O2P)-$Mg^{2+}_{(j)}$, (ii) 10-membered chelation ringsutilizing phosphate groups of adjacent residues as Mg^{2+} ligands, (iii) crystalline-like Mg^{2+}-Mg^{2+} proximities, (iv) direct Mg^{2+}-phosphate interactions and Mg^{2+} dehydration, (v) undulated RNA surfaces with unpaired and unstacked bases, and (vi) and usually, close proximity to site of catalysis.

CHAPTER 1

INTRODUCTION AND PERSPECTIVES

1.1 RNA

RNA is a remarkable biomolecule. It is amazingly pliable, yet plays important structural and catalytic roles in biology (Kruger *et al.*, 1982). The structural polymorphism of RNA arises from the large number of degrees of freedom of the backbone; six backbone torsion angles plus one angle describing rotation of the base relative to the sugar (Saenger, 1984). The structural polymorphism gives rise, in turn, to a variety of RNA structural motifs, e.g. Tetraloops, Kink-turns, E-loop motifs. RNA pliability, exemplified by its ability to accomodate structural insertions, structural deletions, and strand clips (Figure 1.1A) while maintaining a basic RNA motif has been noted in literature. For example the Kink-turn motif (Klein *et al.*, 2001) accommodates structural insertions and deletions; the strand clips are noticed among other RNA motifs (Klein *et al.*, 2001; Lescoute *et al.*, 2005; Hsiao *et al.*, 2006).

Structural Insertions and deletions are related to but not the same as conventional insertions/deletions (indels) in the primary structure, e.g. an RNA sequence. Indels are insertions or deletions which are required to convert one sequence into another (Durbin, 1998). Structure insertions and deletions are required to convert one structure to another. Definitions of structural insertions/deletions, and strand clips are:

(i) *Structural insertions*: A given conformational state (or a motif) can remain intact and essentially unaltered as residues are inserted.

(ii) *Structural deletions*: A given conformational state (or a motif) can remain intact and essentially unaltered as residues are deleted.

(iii) *Strand Clips*: Some elements of a motif are contributed by residues that are remote in the primary sequence.

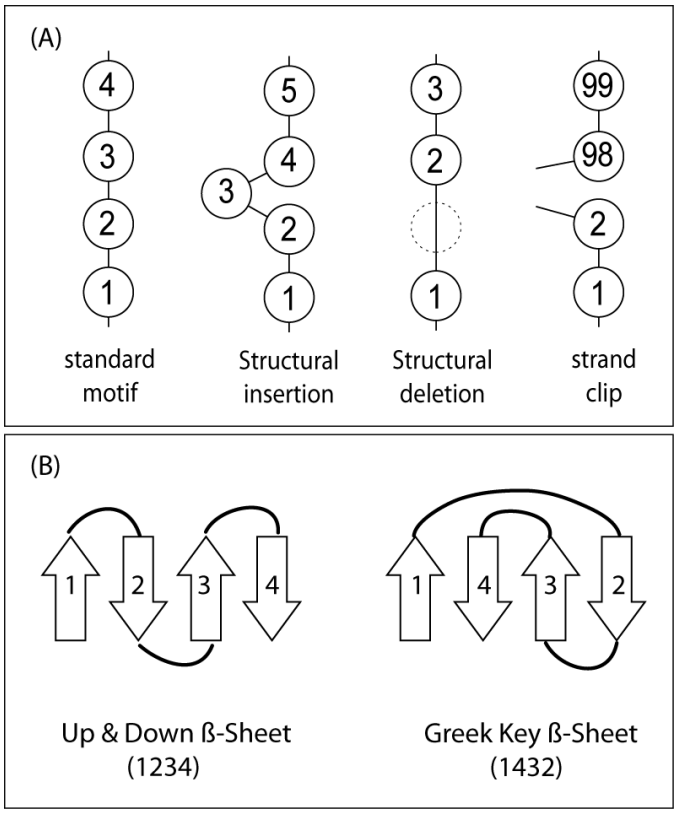

Figure 1.1 Structural variations in RNA and protein structures.
(A) Shown is the schematic of the RNA variations. (B) Shown is the schematic of analogous variations in protein, at secondary structural resolution.

1.1.1 Comparison of RNA with Proteins

It is of interest to compare the pliablility of RNA with that of protein (Table 1.1), the other biological macromolecule with rich structural, folding and catalytic properties. Protein secondary structure readily accommodates topology variations and strand clips. For example anti-parallel β-sheets display up and down topology (in which consecutive strands are ordered i, i+1, i+2, i+3) or Greek key topology (in which consecutive strands are ordered i, i+3, i+2, i+1) [Figure 1.1B]. In addition, domain swapping in proteins is analogous to structural strand clips in RNA. Strand clips are commonly observed in loop

regions, but to our knowledge are not observed within α-helices or β-strands. In addition, to our knowledge, proteins, unlike RNA, cannot accommodate topology switches at the single residue level.

From the perspective of insertions and deletions, proteins are somewhat more restricted than RNA. For α-helices, single-residue insertions cause deformations that are localized to the immediate region of the insertion (Keefe *et al.*, 1993). The carbonyl group of the inserted residue is directed toward the solvent environment and is nearly perpendicular to the helical axis of α-helices. Similar experiments have not been conducted on single amino acid deletions in α-helices.

Apparently, β-sheets cannot accommodate single-residue insertions without global distortions. Single residue insertions, cause β bulges (Richardson *et al.*, 1978), introduce local right-handed strand twist and or cause a register shift of the β-strand (Keefe *et al.*, 1994).

Table 1.1 Comparisons with RNA and Proteins at single residue resolution.

Macromolecule (at single residue resolution)		Structural Insertions	Structural Deletions	Strand clips	Topology Variations[1)]
RNA		**Yes**	**Yes**	**Yes**	**Yes**
Protein	α-helix	Yes	N/A	No[2)]	No
	β-sheet	No	No	No[2)]	No

1) At a single residue level, see Chapter 2 for discussion of RNA topology variations.
2) Strand clips are commonly observed in loop regions, however, to our knowledge, are not observed within α-helices or β-strands.

1.2 Structural Data

The structural databases, e.g. Nucleic Acid Database (Berman *et al.*, 1992), Protein Data Bank (Bernstein *et al.*, 1977), which contains 3D structures of DNA, RNA and protein molecules have grown explosively in the last decade. To date, more than 50 thousand structures have been deposited in the Protein Data Bank, with about four thousand nucleic acid structures in the Nucleic Acid Database.

1.2.1 A Treasure Trove of Structural Bioinformatics

The crystal structures of very large RNA molecules are now available at high resolution (Table 1.2). At 2.4 Å resolution, there is a large subunit (LSU) of the ribosomal RNA from *archaeon H. marismortui* (Ban *et al.*, 2000; Klein *et al.*, 2001) [PDB entry 1JJ2], a halophile from the Dead Sea. At 2.8 Å resolution, there is a ribosomal RNA assembly from *eubacterium T. thermophilus* (Selmer *et al.*, 2006) [PDB entry 2J00, 2J01], isolated from a thermal vent. At 2.3 Å resolution, there is a 160 nucleotide P4-P6 domain of the self-splicing *Tetrahymena thermophila* group I intron (Juneau *et al.*, 2001) [PDB entry 1HR2 , this deletion of C209 mutant gives the best available resolution]. These very large and accurate structures are rich in conformation, folding, tertiary base-base interactions, secondary structures, kinetically trapped intermediates, and sequence information, allowing one to mine structural data, at resolutions from finest to coarsest, in dimensions from 1D sequences to 3D structures, at levels from single nucleotide to multiple nucleotides, to decipher the treasure troves of structural information.

Table 1.2 The data qualities of treasure troves of macromolecular structures

	PDB Entry	Resolution (Å)	R-Value	R-Free
***H. marismortui* 50S**	1JJ2	2.4	0.189	0.222
***T. thermophilus* 70S**	2J00, 2J01	2.8	0.271	0.313
***T. thermophila* group I intron P4-P6 domain**	1HR2	2.3	0.244	0.264

1.3 Decipher the Treasure Troves of Structural Information

Information is everywhere, exhibited in many presentations. A very well known presentation is a polynucleotide of an RNA molecule in xyz coordinates. To find structural patterns of geometries, conformations, inter/intra-molecular interactions,

binding, and folding, it is necessary to treat biological molecules at different levels of presentations. Examples of presentations are Cartesian presentation (xyz presentation), torsion presentation, binned torsion presentation, molecular interaction presentation, pseudo-bond presentation, and sequence presentation, each of these presentations is described below.

1.3.1 Cartesian presentation (xyz presentation)

The structure of molecule is presented by xyz coordinates of each atom of the molecule. A small molecule, under this presentation, shows clearly of all details of physical and chemical properties. However the amount of structural information contained in a macromolecule can be immense. Weaknesses of the xyz presentation for structural pattern deciphering are that it limits the structural comparison between different classifications of molecule, e.g. protein versus RNA, and it lacks internal reference.

1.3.2 Torsion presentation

Bond rotations specify structural conformation without the descriptors of bond lengths and angles (Olson, 1981). Seven torsion angles, (α, β, γ, δ, ε, ζ, χ), define the conformation of a single residue in presenting the conformation of an RNA polynucleotide (Figure 1.2). In RNA, the seven torsion angles are coupled to each other increasing the difficulty of structural pattern deciphering, as the RNA molecule gets larger. Eliminating the degeneracy of the seven torsion angles, by statistical methods, into four torsion angles decreases the complexity of RNA conformation although some structural information may be lost (Sundaralingam, 1969; Sundaralingam, 1973; Srinivasan and Olson, 1980; Hershkovitz *et al.*, 2003). Torsion angles are internally referenced: conformational similarity/dis-similarity can be assayed without superimposition.

1.3.3 Binned torsion presentation

In prior work from our lab, the continuous conformational information (torsion angles) was binned to a limited number of discrete descriptors (Hershkovitz *et al.*, 2003) [Figure 1.2]. For example the α torsion angle falls in one of three frequency envelopes centered at 300°, 165°, and 70°, also see (Murthy *et al.*, 1999). The probability of finding α outside one these three envelopes is vanishingly small [the boundries of the envelopes are 40-90° (envelope 1), 135-190° (envelope 2), 260-330° (envelope 3)]. Therefore the continuous α torsion angle is binned and transformed to one of the three discrete values (300°, 165°, or 70°), called envelops. The envelope widths are determined by natural cutoff parameters to delineate the discrete bins. This presentation reduces the RNA structural conformation to a relatively small number of discrete states which can be symbolically encoded. For example the ASCII code "a" indicates the state of a single nucleotide to be (α_3, γ_1, δ_1, ζ_1) [α_3 indicates torsional α in envelope 3, etc.]. Thus the conformation of an RNA molecule is approximated by an ASCII encoded string. With this presentation, the string "aaoa" or "aaoe" denotes a RNA tetraloop motif.

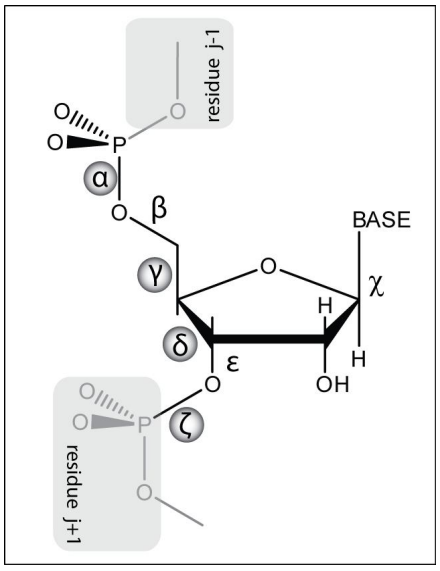

Figure 1.2 RNA torsion angles.
Shown is the RNA seven torsion angles, (α, β, γ, δ, ε, ζ, χ), where the binned torsion presentation of the four descriptors, (α, γ, δ, ζ), are shaded.

1.3.4 Molecular interaction presentation

There are a multitude of molecular interactions. Westhof and coworkers have focused on base-base interactions and proximities to develop the concept of isosteric base-pairs (Leontis and Westhof, 2001; Leontis *et al.*, 2002a; Waugh *et al.*, 2002; Leontis and Westhof, 2003; Yang *et al.*, 2003). Their methods automatically identify and classify types of base pairs, which use various combinations of the three edges; Watson-Crick, Hoogsteen and the Sugar edges. They have surveyed the structural database and compiled statistics for the occurrence of each edge and of each of the 12 base pair families defined by the combinations of the four bases: A, U, C and G.

1.3.5 Pseudo-bond presentation

Pseudo-bonds are vectors between non-bonded atoms. Two pseudo-torsion angles are then defined by series of pseudo-bonds. This approach reduces the complexity of large RNA molecules facilitating pattern recognition and RNA motif searching (Olson, 1975; Duarte and Pyle, 1998; Duarte *et al.*, 2003).

1.3.6 Sequence presentation

As noted in the Section 1.3, RNA structures inhibit in many presentations. From Sections 1.3.1-1.3.5, presentations of RNA structures in three-dimention have elucidated many useful information of RNA. However, another important presentation for study of RNAs is 1D structure (RNA sequence), especially when 3D structure of an RNA molecule is undetermined. RNA sequence, which is given by combinations of four bases, A, U, C and G, is useful, but can conceal enormous treasure troves of structural information. Early phylogenetic comparison of ribosomal RNA sequences led Woese and Fox to the discovery of Archaea, the third kingdom of life (Woese and Fox, 1977; Magrum *et al.*, 1978; Woese *et al.*, 1978). Sequence allows one to decipher RNA secondary structure. For example the covariation analysis method (reviewed by Gutell,

Gutell *et al.*, 2002) was first used by Holley to determine the secondary structure of tRNA (Holley *et al.*, 1965), and by Woese and Fox to determine the secondary structure of 5s rRNA (Fox and Woese, 1975). Covariation was then used to determine secondary structures of the 16 S (Woese *et al.*, 1980) and 23 S rRNA (Noller *et al.*, 1981). More recently, Gutell and coworkers (Goertzen *et al.*, 2003) proposed a secondary structural model for the nuclear ribosomal internal transcribed spacers using a comparative analysis of 340 sequences.

1.4 Structural Alignment on Ribosomal RNAs

More recently, analysis of rRNA has moved from 1D and 2D to 3D. The structural database currently contains structures of ribsomes from five organisms [*Thermus thermophilus*, x-ray, 2.8 Å, Figure 1.3A, (Selmer *et al.*, 2006), *Escherichia coli*, x-ray, 3.2 Å (Berk *et al.*, 2006)*, Haloarcula marismortui*, x-ray, 2.4 Å, LSU only, Figure 1.3B (Ban *et al.*, 2000), *Deinococcus Radiodurans*, x-ray, 3.1 Å, LSU only, (Harms *et al.*, 2001) and *Saccharomyces cerevisiae*, cryo-EM, 11.7 Å (Spahn *et al.*, 2004)]. The structural comparison in between these 3D structural rRNAs, in conjunction with small molecules, reveals the patterns of mechanism, binding, and fundamental inter/intra-molecular interaction.

In a recent work, Steitz and coworkers use a transition analog complex with the large subunit (LSU) of *H. marismortui* to propose a catalysis mechanism for peptidyl transferase activity [(Schmeing *et al.*, 2005)]. They compare the crystal structures of various complexes by the superimposition of phosphorus atoms of domain 5 of 23S rRNA.

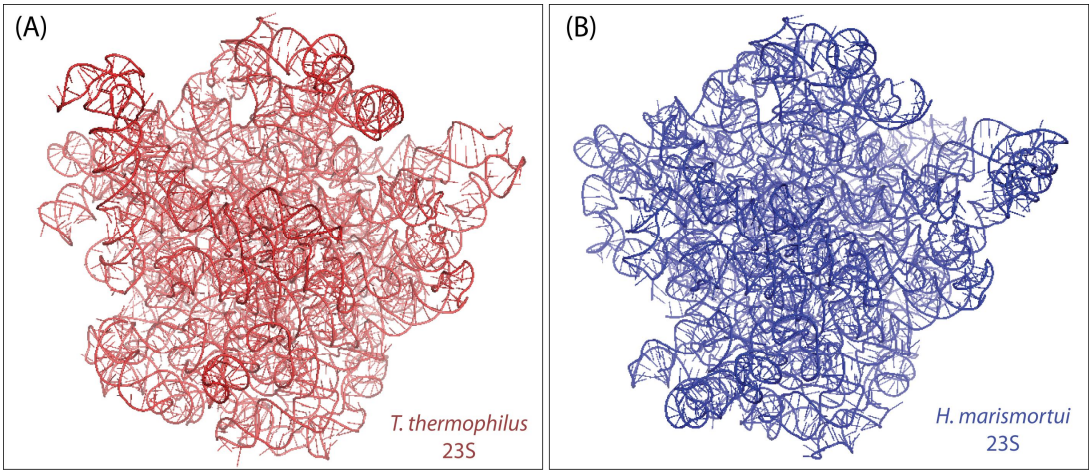

Figure 1.3 3D structures of *T. thermophilus* 23S and *H. marismortui* 23S.
(A) Shown is the 3D structure of *T. thermophilus* 23S (PDB entry: 2J01), cartoon representation. (B) Shown is the 3D structure of *H. marismortui* 23S (PDB entry: 1JJ2), cartoon representation.

One weakness of their superimposition is that only the phosphorous atoms of domain 5 are used, so that the movement of rigid bodies cannot be observed, and errors of structural conservation/divergence information are introduced. To avoid that, alignment must take into account insertions, deletions and sequence variation (see section 1.1, Figure 1.1A, Chapter 2 and 3), while using as many atoms as possible in the superimposition.

An accurate superimposition model allows one to identify regions of structural conservation and diversity, to determine relationships between structural and sequence variation, allows docking of one structure onto another, and study of conformational change, binding, and mechanism in atomic resolution.

1.5 What is an RNA Motif?

The term *RNA Motif* has many definitions. Here is a list of a few commonly applied definitions of the term *RNA Motif*: (i) RNA motifs have known binding or chemical activity (Bourdeau *et al.*, 1999); (ii) RNA motifs are discrete sequences or combinations of base juxtapositions found in high abundance, with a three-dimensional structures that are independent of environment (Moore, 1999); (iii) RNA motifs are

ordered stacked arrays of base pairs with distinctive backbone geometries with ordered arrays of isosteric non-Watson-Crick base pairs (Leontis and Westhof, 2003); (iv) RNA motifs are tRNAs, 5S rRNAs, SRP RNA, C/D box snoRNAs, hammerhead motifs, miRNAs, etc. (Lambert *et al.*, 2004).

These definitions allow one to decipher the contents of the treasure troves of structural information. For example the definition given by Westhof illustrates RNA in the 'molecular interaction presentation' (stacking, pairing) [see section 1.3.4] and in the 'torsion presentation' and 'binned torsion presentation' (backbone geometries) [see sections 1.3.2 and 1.3.3]. The definition given by Moore describes RNA in the 'sequence presentation' (discrete sequences) [see section 1.3.6] and in the 'Cartesian presentation' (three-dimensional structures) [see section 1.3.1].

1.5.1 The RNA Tetraloop Motif

The RNA tetraloop is best known as the simplest, smallest and most frequent RNA motif. Tetraloops are first observed in early phylogenetic comparisons of RNAs (Woese *et al.*, 1983; Tuerk *et al.*, 1988; Woese *et al.*, 1990). Tetraloops were seen to cap A-form stems (Moore, 1999), and to show exceptional thermodynamic stabilities (Tuerk *et al.*, 1988; Cheong *et al.*, 1990; Varani *et al.*, 1991; Antao and Tinoco, 1992). Tetraloops are broadly grouped by sequence into three classes (Woese *et al.*, 1990), which are GNRA (Woese *et al.*, 1990; Jaeger *et al.*, 1994), UNCG (Tuerk *et al.*, 1988) and CUUG (Woese *et al.*, 1990; Jucker and Pardi, 1995b) [where N can be any nucleotide and R is either G or A].

A loop is defined as a short RNA segment that connects one chain of a double helix with the other (Woese *et al.*, 1983). We have redefined the tetraloop. Tetraloops are generally terminal loops, however, not all tetraloops are loops, nor are they terminal (see Chapters 2 and 3).

In sum, tetraloops are thought to (i) initiate folding of complex RNA molecules (Tuerk *et al.*, 1988), (ii) stabilize helical stems (Tuerk *et al.*, 1988; Selinger *et al.*, 1993), and (iii) provide recognition elements for tertiary interactions and protein binding (Michel and Westhof, 1990; Puglisi *et al.*, 1992; Jaeger *et al.*, 1994; Cate *et al.*, 1996a).

1.5.2 The RNA Ion-Binding Motif?

Functional RNAs generally fold into stable states of given conformation (Latham and Cech, 1989; Celander and Cech, 1991). RNA folding is hierarchical (Brion and Westhof, 1997; Tinoco and Bustamante, 1999), beginning with random coil, proceeding through secondary structural units, e.g. stems and loops, and lastly progressing to the final folded state stabilized by tertiary interactions, e.g. short/long range of base-base hydrogen bonding and stacking interactions.

Ions play important roles in the folding pathway. Cations are sequestered from bulk solvent, and held in close proximity to the polynucleotide. Secondary structures are favorable in broad range of ionic conditions. Tertiary structures are favored by divalent cations (Cole *et al.*, 1972; Misra and Draper, 2000). It is generally understandable that *tertiary structure of an RNA* is commonly characterized by short and long range of base-base hydrogen bonding and stacking interactions. And a *folded RNA* must account three dimensional positions of all RNA atoms, including ions, water, etc. However the geometry, conformation, and binding patterns of RNA-ion are intricate and unpredictable. As noted by Sponer, the classical molecular mechanics is inadequate to describing the interactions of ions and RNAs (Gresh *et al.*, 2003; Rulisek and Sponer, 2003; Korostelev *et al.*, 2006). Should a recurrent *RNA ion-binding motif* that possesses general patterns exist among *folded* RNAs? The short answer is *yes* and relies on the Mg^{2+}-RNA interactions. This is discussed in detail in Chapters 4 and 5.

1.6 Objective of Current Studies

In the current studies, I present my work on deciphering, exploring, and discovering the treasure troves of RNA structural informatics, mainly in the areas of multi-resolution analysis of RNA structure, pattern recognition of RNA structural motifs in the PBR space, a new descriptive language of RNA, RNA topology variation at a single residue level, re-defininition and unified description of the RNA tetraloop motif, the construction of tetraloop family tree, a novel approach to superimposing ribosome, and revealing of recurrent magnesium-binding motif among RNAs.

I develop a concept to representing RNA structure in multi-resolution. Groups of atoms (bases / riboses / phosphates / residues / groups of residues, motifs, etc.) are reduced to pseudo-objects, with locations and in some cases, orientations. Larger numbers of atoms in pseudo-objects correspond to lower resolutions. Structural data is viewed through a series of resolutions from finest to coarsest.

During the development of RNA structural data in multi-resolution, a very useful space that I have developed, called PBR space (P indicates Phosphate, B indicates Base, and R indicates Ribose) [see Chapter 2]. A set of unique pattern of RNA motifs, e.g. A-helix, tetraloop, Kink-turn, E-loop motifs etc., are clearly seen in the PBR space, leading us to the discovery of RNA topology variations at a single residue level. Collectively insertions, deletions, strand clips, and 3,2 switches (topology variation) are first introduced by us to systematize and investigate as a general property of RNA.

By comparisons of this method to others (Hershkovitz *et al.*, 2003; Larose *et al.*, 2005), around 30% more of tetraloops have been uncovered and classified. With all, I unified and re-defined the tetraloop motif and represent the relationships of groups of tetraloop in tree formalism, (tetraloop family tree).

Later I noticed that the structural alignment of very large RNA molecules is very challenging due to the intricate backbone and tertiary interactions. To overcome these

obstacles, I developed a concept of structural anchors for performing superimposition of 23S rRNAs (see Chapter 3).

The successful superimposition allows me to investigate the structural relationship between RNA to RNA, RNA to ions, and ions to ions. More importantly, a recurrent *magnesium-binding motif* in RNAs is observed.

CHAPTER 2

SINGLE NUCLEOTIDE RNA CHOREOGRAPHY

2.1 Introduction

Prediction and design of three-dimensional structures of large RNAs are best approached using small structural motifs, with modular and hierarchical characteristics (Moore, 1999; Chworos *et al.*, 2004). The simplest, smallest and most frequent RNA motif is known as the tetraloop. Tetraloops are terminal loops, with characteristic four-residue sequences first observed in early phylogenetic comparisons of RNAs (Woese *et al.*, 1983; Tuerk *et al.*, 1988; Woese *et al.*, 1990). Tetraloops were seen to connect two anti-parallel chains of double-helical RNA, and so cap A-form stems (Moore, 1999). Isolated stem/tetraloops show well-defined structure, and exceptional thermodynamic stabilities (Tuerk *et al.*, 1988; Cheong *et al.*, 1990; Varani *et al.*, 1991; Antao and Tinoco, 1992). Tetraloops are thought to (i) initiate folding of complex RNA molecules (Tuerk *et al.*, 1988), (ii) stabilize helical stems (Tuerk *et al.*, 1988; Selinger *et al.*, 1993), and (iii) provide recognition elements for tertiary interactions and protein binding (Michel and Westhof, 1990; Puglisi *et al.*, 1992; Jaeger *et al.*, 1994; Cate *et al.*, 1996a). Tetraloops have been broadly grouped by sequence into three classes (Woese *et al.*, 1990), which are GNRA (Jaeger *et al.*, 1994; Jucker and Pardi, 1995a; Jucker *et al.*, 1996; Butcher *et al.*, 1997; Correll and Swinger, 2003), UNCG (Tuerk *et al.*, 1988; Cheong *et al.*, 1990; Allain and Varani, 1995; Akke *et al.*, 1997; Williams and Hall, 1999; Ennifar *et al.*, 2000), and CUUG (Jucker and Pardi, 1995b; Baumruk *et al.*, 2001), (where N can be any nucleotide and R is either G or A.)

Here we re-define and expand the RNA motif concept, unifying what previously appeared to be disparate groups of structures. We focus on the tetraloop motif, and

demonstrate increased frequencies, new contexts, unexpected lengths, and novel topologies. The results, with broad implications for RNA structure in general, show that even at this most elementary level of organization, RNA tolerates variation in conformation, topology and molecular interactions. However the variation is not random; it is well-described by four distinct modes, which are insertions, deletions, strand clips and 3-2 switches. Collectively we call theses four modes DevLS (pronounced Devils, Deviations of Local Structure). The four DevLS are shown in Figure 2.1.

RNA structure is commonly understood by analysis of base-base interactions and proximities, which led to the concept of isosteric base-pairs (Leontis and Westhof, 2001; Lemieux and Major, 2002; Leontis *et al.*, 2002b; Waugh *et al.*, 2002; Leontis and Westhof, 2003; Yang *et al.*, 2003; Lee and Gutell, 2004). RNA analysis in torsional space can be simplified and reduced in dimensionality with pseudo-bonds, which are vectors between non-bonded atoms (Olson, 1975). Pattern recognition methods have been applied successfully, by Pyle and coworkers, to geometric relationships between pseudo-bonds (Duarte and Pyle, 1998; Duarte *et al.*, 2003). Finally, phylogenetic covariation allows one to decipher RNA secondary and tertiary structure, and thereby infer three-dimensional structure (Levitt, 1969; Woese *et al.*, 1983; Gutell *et al.*, 1986; Lee *et al.*, 2003). In an example that is relevant to the results described here, Gutell and coworkers have observed the Lonepair Triloop (LPTL) (Lee *et al.*, 2003).

2.1.1 Multi-Resolution Analysis of RNA Structure

We look at RNA at various resolutions (or scales) from the finest to coarsest. Note that we are using the term ‘resolution’ in the sense of signal processing (Leontis *et al.*, 2002b) and it should not be confused with ‘crystallographic resolution’. Resolution is varied by reducing natural groups of RNA atoms (bases / riboses / phosphates / residues / groups of residues, motifs, etc.) to pseudo-objects, with locations and orientations. Larger numbers of atoms in pseudo-objects correspond to lower resolutions. The basic idea is

that important structural features become readily observable only in certain resolution ranges. Therefore resolution is a variable parameter like the tunable magnification of an optical microscope. Analysis of spatial relationships and interactions between RNA pseudo-objects can reveal fundamental RNA architecture that is often obscure at a single resolution. Multi-resolution techniques have been very successful in protein simulations (Monge *et al.*, 1995; Betancourt, 2003) and signal, and data processing (Mallat, 1999). We use the multi-resolution analysis in combination with molecular interactions in an iterative process to develop empirical motif descriptions. Interactions that become evident in multi-resolution analysis are appended to a search model, leading to empirical motif definitions.

The HM 23S rRNA (1JJ2) is our test "database". The crystal structure of the large ribosomal subunit from Haloarcula marismortui has been determined to high resolution by Steitz and Moore (Ban *et al.*, 2000; Klein *et al.*, 2001). At 2.4 Å resolution, the atomic positions of the vast majority of the 23S rRNA of HM LSU are well-characterized, and, as of this writing, are more acutely determined than any other large RNA complex (although error and noise cannot be ignored, Murray *et al.*, 2003). The HM 23S rRNA, with over 2500 residues, constitutes a large database with a rich omnibus of RNA conformation and interactions.

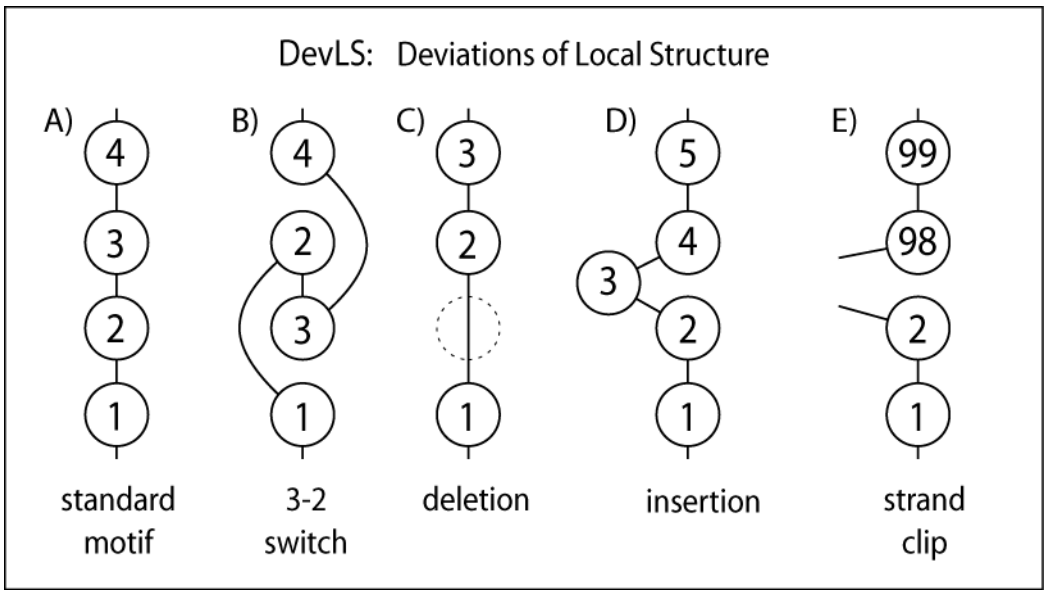

Figure 2.1 RNA DevLS.
(A) A generic RNA motif is represented schematically by four circles, which symbolize four residues. (B) In a motif with 3-2 switch, the positions of two bases, of residues 3 and 2 in the figure, are interchanged. The backbone linkage is maintained. (C) In a deleted motif, a residue is omitted (dashed line). (D) In an inserted motif, a residue is added. (E) In a strand clipped motif one or more residues is contributed from a remote region of the primary sequence. An insertion, if extensive enough can be equivalent to a strand clip. The numbers indicate the covalent ordering of the residues along the polynucleotide strand. These four DevLS arise from common enabling factors, which operate at the single nucleotide level. These factors are the high RNA backbone length per residue (6 bonds separate adjacent residues) and numerous torsional degrees of freedom of RNA nucleotides.

2.2 Materials and Methods

2.2.1 Detection of RNA Tetraloops

To decrease the resolution of RNA, groups of atoms (bases / riboses / phosphates / residues / groups of residues, motifs, etc.) are reduced to pseudo-objects, with locations and in some cases, orientations. Larger numbers of atoms in pseudo-objects correspond to lower resolutions. A very useful space that we have developed, called PBR space (P indicates Phosphate, B indicates Base, and R indicates Ribose) is shown in Figure 2.2. We have defined the center of mass (cm) and orientation of bases, riboses, and phosphates. The relative orientations of adjacent bases are given by the angle Θ_{bpn} which is the angle between the two base plane normals. Information on relative positions of

riboses is provided by Θ_{rcm}. Information on relative positions of phosphates is given by Θ_{ppp}. RNA motifs are detectable by fingerprints in PBR space.

PBR space has reduced ability to distinguish among standard tetraloops and those that have undergone deletions, insertions, strand clips or 3-2 switches: at this scale they have certain equivalencies. This blurring is the point of multi-resolution analysis: successively simplify the search space to find patterns that persist from the finer to coarser scales. If a pattern indeed remains at the coarser resolution it will be much easier to discover.

2.2.2 Molecular Interaction Space, 1st Iteration

The 25 tetraloops identified by torsional analysis (Hershkovitz *et al.*, 2003) were used to devise a minimal molecular interaction definition of a tetraloop. Each of the 25 torsionally-derived tetraloops shows an interaction between the O2' atom of the residue j-1 and the N7 atom of residue j +1. No other hydrogen bonding interaction is conserved. Therefore a search of all j-1(O2') to j+1(N7) interactions was conducted, giving 44 hits. Eleven of those are false positives, 33 are valid tetraloops.

2.2.3 Multi-scale-spaces.

A variety of scale-spaces from fine to coarse grain are in preliminary use in our lab. We followed this path:

(i) A tentative scale-space was defined.

(ii) A preliminary tetraloop fingerprint in that scale-space was established empirically, using the 33 tetraloops identified in torsional spaces and molecular interaction spaces.

(iii) The scale-space was refined, uninformative parameters were discarded, sets of parameters yielding redundant information were consolidated. New parameters were added.

(iv) The new empirical fingerprint for a tetraloop was determined (Scheme 2.1), which in combination with the molecular interaction definition, gave 41 putative tetraloops.

(v) The observed tetraloops are inspected and validated. Two tetraloops were determined to be false positives, leaving 39 tetraloops. Therefore the PBR scale-space revealed 6 tetraloops that had eluded us in torsional and our minimal interaction spaces.

$$(a)\ 60^\circ < \theta^{j}_{ppp}\ ,\ \theta^{j}_{rcm} < 112^\circ$$
$$(b)\ 7.5\mathring{A} < r^{j\,\|\,j-1}_{bcm} < 10.5\mathring{A}$$
$$(c)\ \theta^{j-1}_{ppp} < 162^\circ\ ;\ \theta^{j-1}_{bpn} < 75^\circ$$
$$(d)\ r^{j+1}_{bcm} < 12\mathring{A}$$
$$(e)\ \theta^{j+3\,\|\,j+4}_{ppp} < 138^\circ$$

Scheme 2.1 Tetraloop fingerprint in PBR Space.

2.2.4 Molecular Interaction Spaces, 2nd Iteration.

The additional tetraloops found in the scale-space search allowed us to re-evaluate the molecular interaction definition. The revised tetraloop definition allows either j-1(O2') to j+1(N7) or j-1(O2') to j+2(N7) hydrogen bonds. This definition gives 36 tetraloops, one of which was not found in the PBR space, and 33 false positives. None of the false positives are common to the molecular interaction and PBR space. In combination, PBR space and molecular interaction space reveal 40 tetraloops and exclude all false positives.

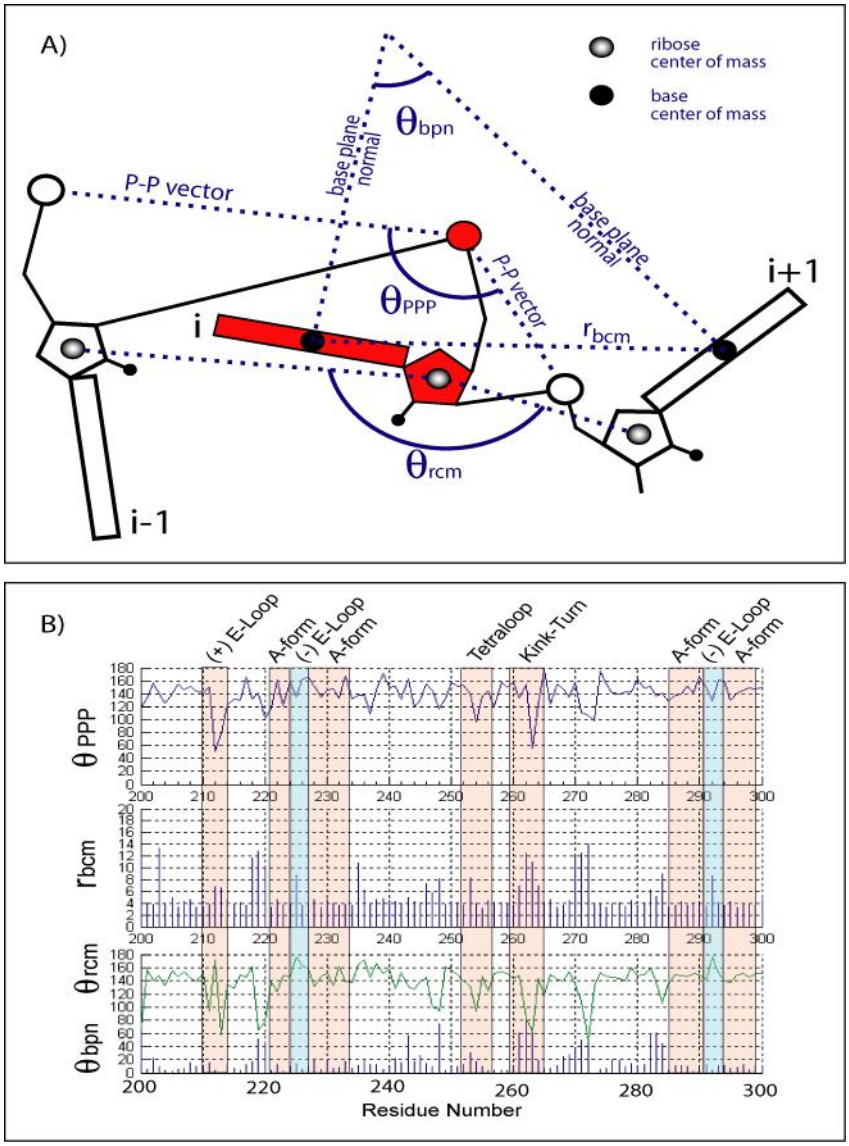

Figure 2.2 PBR space.
(A) PBR space. A decreased resolution view of RNA, where atoms are combined to make pseudo-objects, and special relationships between pseudo-objects are described. (B) Residues 200-300 of 1JJ2 in PBR space. Note that A-helices, E-loop Motifs, Kink-Turns, etc give distinctive fingerprints in PBR space.

2.2.5 Molecular Interaction Spaces, Final Description

A second class of interaction, j-1(base HB donor) to j+2 (O2P) (or less commonly j+1 (O2P)), is observed in 32 of the 40 observed tetraloops, with only one false positive.

2.2.6 Cartesian Spaces.

It is necessary that a general, rigorous, objective and transparent statistical definition of similarity be used to validate that the RNA fragments postulated to be similar are indeed similar, and to define false positive and false negative. For these purposes we use RMSDs of atomic positions.

2.3 Results

We use multi-resolution approaches and molecular interactions to identify motifs in three-dimensional structures of large RNAs. The results show that tetraloops are commonly adorned with four types of DevLS (Figure 2.3). DevLS occur in seventeen of the 40 observed tetraloops.

2.3.1 Tetraloop Family Tree

The incorporation of DEVLS into the tetraloop definition allows us to build a tetraloop family tree (Figure 2.4). Tetraloops fall naturally into eight groups, partitioned by the types and sites of DevLS. We have developed a nomenclature to describe tetraloop groups (Figure 2.3: Tl indicates tetraloop, s indicates standard, d indicates deletion, i indicates insertion, x indicates residue switch, subscripts indicate positions.) The most populated groups are the s-Tl tetraloops (21 members) and d_2-Tl tetraloops (10 members).

2.3.2 Intra-loop Interactions.

A set of consensus molecular interactions characterize tetraloops throughout the family tree, summarized for the 21 s-Tl tetraloops and the 10 d_2-Tl tetraloops in Table 2.1. Observed hydrogen bonding interactions are consistent with expectations for 'GNRA' tetraloops (for example, see Jucker *et al.*, 1996) and U-turns. Hydrogen bonding interactions of O2' of residue (j-1) with cross-loop base atoms are the most enduring throughout the tetraloop family tree. Twenty of 21 s-Tl tetraloops and 9 of 10 d_2-Tl

tetraloops form these hydrogen bonds. s-Tl (1238) is one exception. d_2-Tl (1500), the other exception, has an O2' (j-1) to N7 (j+1) distance of 3.5 Å, which falls nominally outside our hydrogen bonding cutoff.

Although residues j-1 and j+2 appear to be poised to do so, a sheared G-A base pair involving them is infrequent. In s-Tl tetraloops where residue j-1 is a G and residue j+2 is an A, only a single hydrogen bond links them (also see Pley *et al.*, 1994; Correll and Swinger, 2003); the average N3 (j-1) to N6 (j+2) distance for s-Tl tetraloops is 4.7 Å. However for a small subset of tetraloops with DevLS, the distance is considerably shorter [3.4 Å (506), 3.5 (1707), 3.5 (482)], consistent with a true sheared G-A base pair.

As can be seen from Table 2.1, G and U at position j-1 are interchangeable in terms of cross-loop hydrogen bonding interactions. The hydrogen bond donors N1 and N2 of G are roughly replaceable by donor N3 of U in interactions with the O2P of residue j+2. G is preferred over U at j-1 in s-Tl tetraloops and U is preferred over G in d_2-Tl tetraloops (see Sequence Logo: Figure 2.4).

Table 2.1 Consensus Hydrogen Bonding Interactions[1] in s-Tl and d_2-Tl tetraloops

s-Tl	**j+1**	**Frequency**	**j+2**	**Frequency**
j-1	O2'-N7(R)[2]	(19/20)[3]	N1/N2 (G) -O2P[4]	(14/14)
	O2'-N6/O6(R)	(12/20)	N3(U)-O2P	(3/5)
			N2(G)-N7(A)	(13/14)
d_2-Tl	**j+1**		**j+2**	
j-1	O2'-N7 (R)	(8/10)	N1 (G) – O2P	(3/3)
	O2'-N6/O6 (R)[5]	(0/10)	N2 (G) – O5'	(3/3)
			N3 (U) – O2P	(7/7)

1) Hydrogen bonds are determined by geometry (3.4 Å cut off and reasonable angles).
2) This field indicates hydrogen bonding interactions between O2' atoms of residue j-1 (right column) and N7 atoms of purines at residue j+1 (top row).
3) Twenty of 21 s-Tl tetraloops have G or A at position j+1. Nineteen of these show a hydrogen bond from the O2' of residue j-1 to the N7 of residue j+1.
4) Fourteen of 14 s-Tl tetraloops with G at j-1 show a hydrogen bond from either the N1 or the N2 of G(j-1) to the O2P of residue j+2, or both.
5) The hydrogen bond from the O2' of residue j-1 to N6/O6 (R), frequency is 13/20, in s-Tl tetraloops is not observed in d_2-Tl tetraloops.

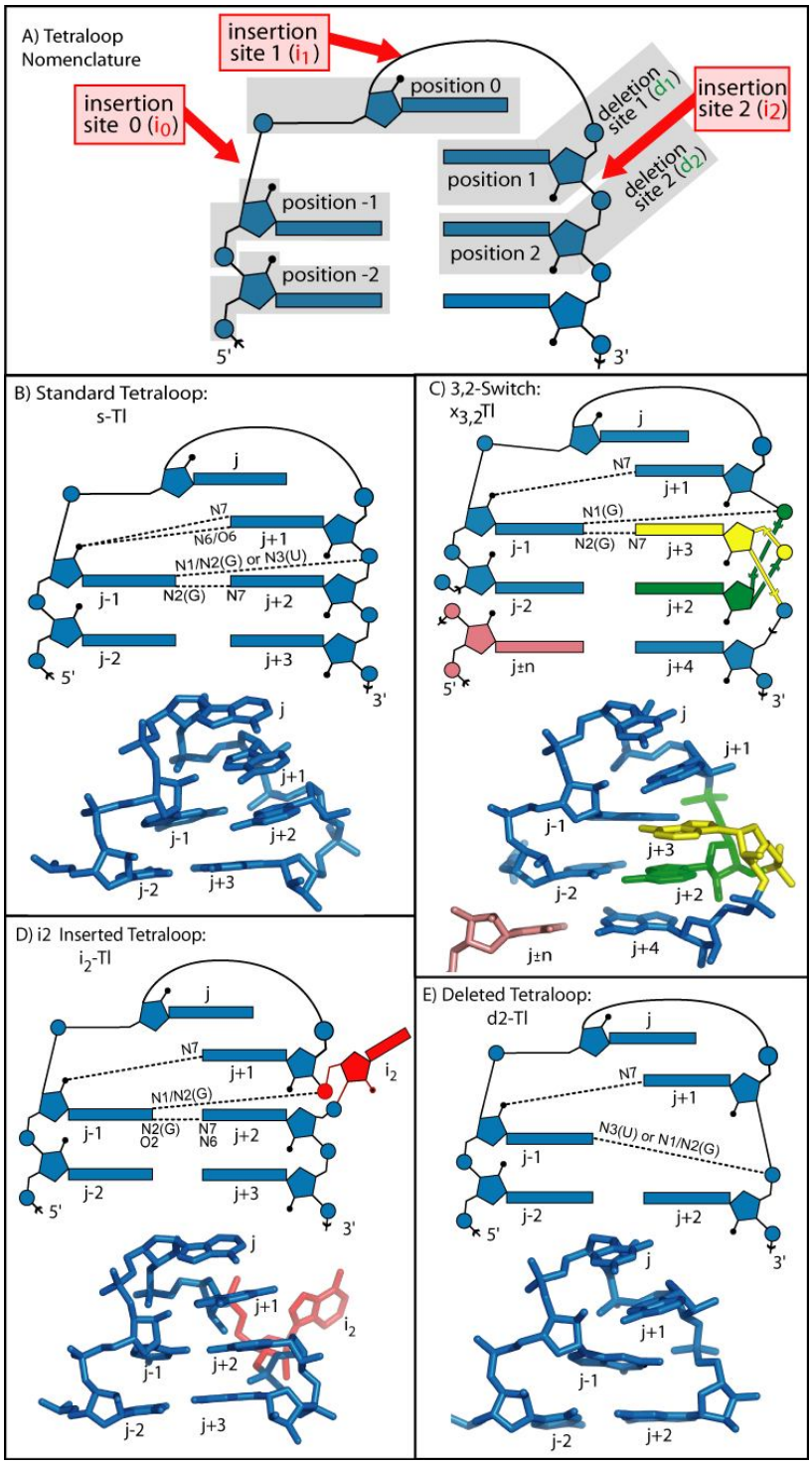

Figure 2.3 Tetraloops adorned with DevLS.
(A) Observed sites of insertions (red text) and deletions (green text) in tetraloops. (B) A standard tetraloop (s-Tl tetraloop 805). (C) A tetraloop with a 3,2-switch ($x_{3,2}$-Tl tetraloop 482). (D) A tetraloop with a residue inserted at the 2 position (i_2-Tl tetraloop 494). (E) A tetraloop with a residue deleted at the 2 position. (d_2-Tl tetraloop 1809). Dashed lines represent consensus hydrogen bonds. Hydrogen bond donors and acceptors are indicated. The top of each panel in B-E shows a consensus schematic representation. The bottom of each panel shows a representative 3D structure from1JJ2.

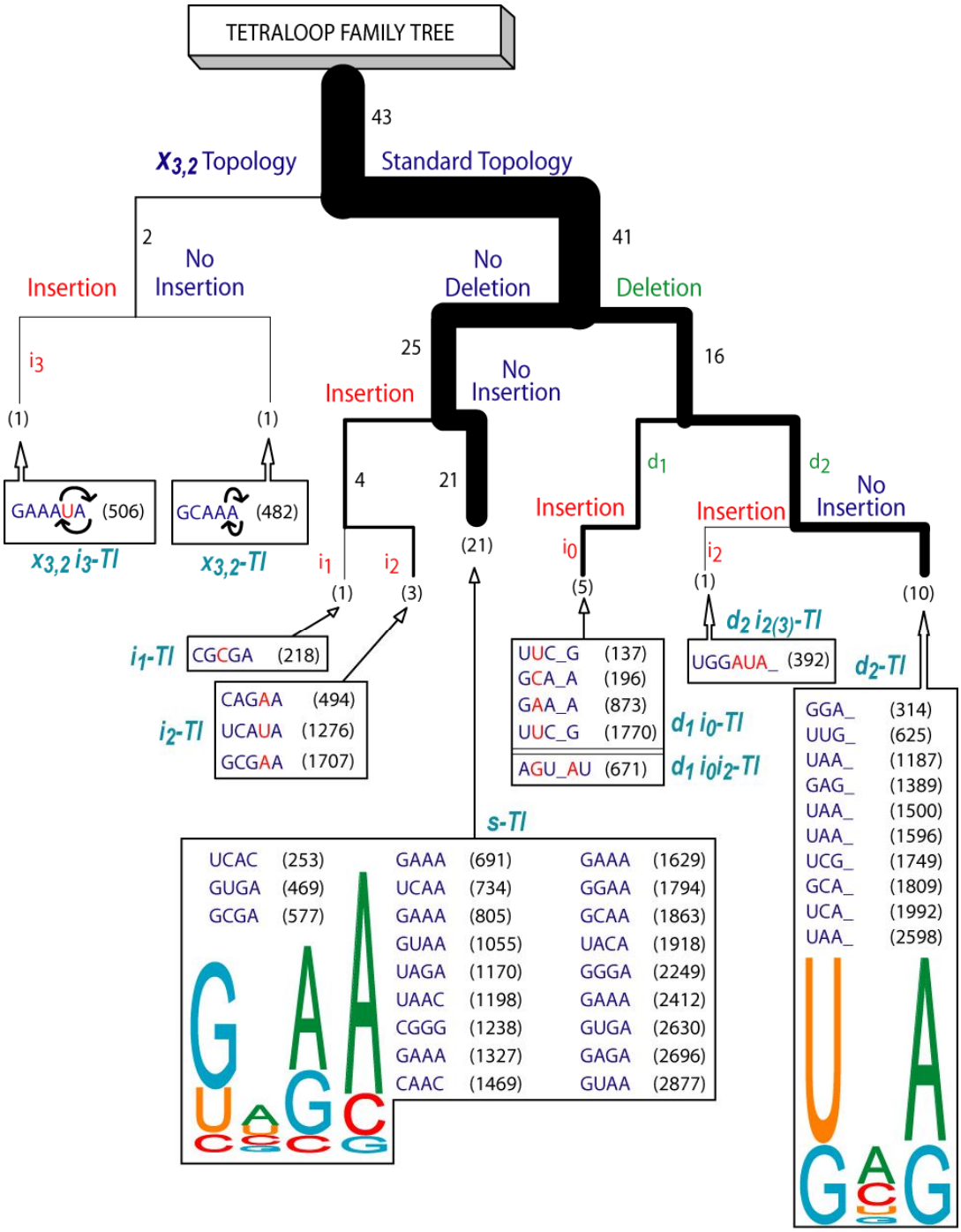

Figure 2.4 Tetraloop Family Tree.
Forty tetraloops of 1JJ2 are distributed by type of position of DevLS. Insertion positions are indicated in red text. Deletion positions are indicated in green text. The positions of deleted residues are marked by underscores. Number of occurrences is indicated in black, with line widths proportional to frequency. There are 8 groups (boxed). The residue number of the first residue and the sequence is given for each tetraloop. The consensus sequence for the s-Tl and d_2-Tl tetraloops are indicated by a sequence Logo representation (Schneider and Stephens, 1990). Entries 196, 671, 873 were described by Huang (Huang *et al.*, 2005). These were not detected by our methods and are outliers in conformation and molecular interactions.

2.3.3 DevLS Influence Helical Capping Function

Seven tetraloops are flanked by strand clips, which are observed adjacent to but not within tetraloops. All observed strand-clipped tetraloops, by definition, cap pseudo–helices, where bases are stacked, and assume a helical form, but are not covalently linked by the backbone. d_2-Tl tetraloops are most frequently associated with clipping (30%). Three d_2-Tl tetraloops are strand-clipped directly on the 5' side of the tetraloop, between residues j+2 and j+3 (d_2-Tl tetraloops 1187, 1809, 2598). One tetraloop is strand clipped between j-2 and j-3 ($x_{3,2}$-Tl 482; Figure 2.3C). s-Tl 1629 is clipped between residues j+3 and j+4. $x_{3,2}$-Tl 506 is clipped between j+4 and j+5. s-Tl 1238 is clipped between residues j-1 and j-2.

Observed tetraloops are mapped onto the secondary structure, and coded by group in Figure 2.5. It can be observed that nineteen of 21 s-Tl tetraloops cap helices (Elgavish *et al.*, 2001) (not 1238 or 1629, which are clipped). All seven standard topology i-Tl tetraloops (tetraloops with insertions but not 3-2 switches) cap helices. None of the d_2-Tl tetraloops cap *unperturbed* A-form stems. An *unperturbed* A-form stem exhibits well-defined molecular interactions such as base pairing and base stacking, with no insertions or strand clipping. Six of 10 d_2-Tl tetraloops cap helices (not 1187, 1749, 1809, or 2598). All non-clipped d_2-Tl associated helices are perturbed by unpaired bases. One d_2-Tl tetraloop (1749) caps neither a helix nor a pseudo-helix. This tetraloop is 'unhinged' in that both terminal resides (j-2 and j+1) crown a cavity, and are not stacked on adjacent helical regions. Neither of the 3-2 switched tetraloops cap helices.

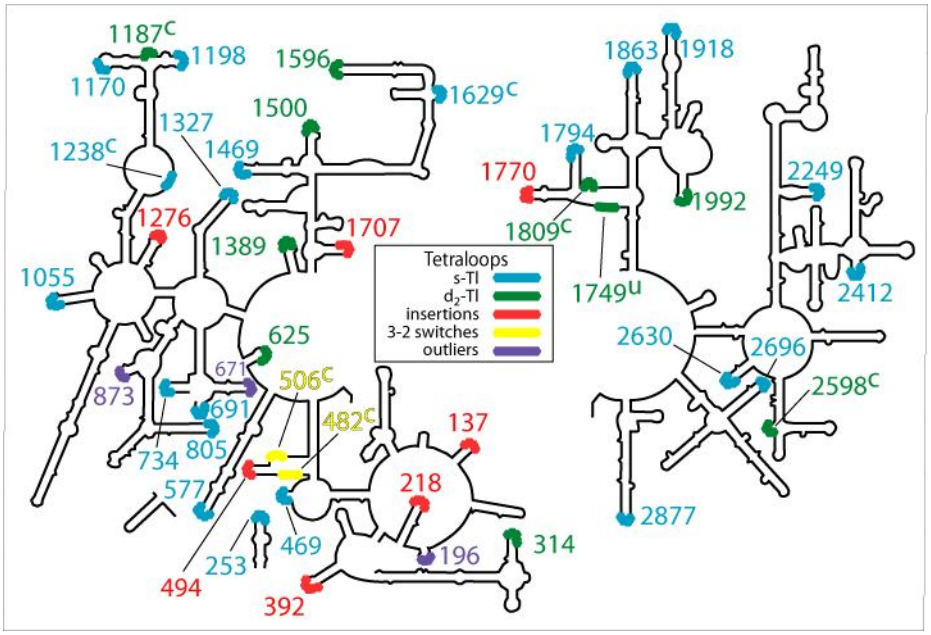

Figure 2.5 Secondary structure of the HM 23s rRNA (1JJ2).
Tetraloop locations and type are indicated by color. Superscripted c's indicate strand-clipped tetraloops. A superscripted u indicates the unhinged tetraloop. The strand clipped tetraloops are in contexts in which they do not cap helical stems, as can be inferred from the secondary structure, but do cap pseudohelical stems. The unhinged tetraloop crowns a cavity. Entries 196, 671, 873 were described by Huang (Huang *et al.*, 2005). These were not detected by our methods and are outliers in conformation and molecular interactions

2.3.4 Group Validation and Similarity Statistics

We believe that all 40 members of the Tetraloop Family Tree are structurally related, and should be described as members of a common motif. This conclusion is supported by *Intra-Group* and *Inter-Group* similarity statistics, and by conservation of molecular interactions. *Intra-group* similarity is characterized by RMSD of atomic positions (RMSD-AP) for atoms that are common *within* a group, generally backbone atoms. *Inter-group similarity* is characterized by RMSD-AP of specified backbone atoms that are common *between* two groups. RMSD-AP is determined after superimposition.

s-Tl tetraloops: The 21 s-Tl tetraloops fit the previous GNRA tetraloop definition. *Intra-Group Similarity*: The RMSD-AP for all backbone atoms [four residues, (j-1), (j), (j+1), (j+2)] is 0.65 Å, giving a natural metric for tetraloop rigidity, and an RMSD-AP

norm for evaluating degree of similarity between and within tetraloop groups. The atoms of residue j+2 show the greatest deviations (Figure 2.6).

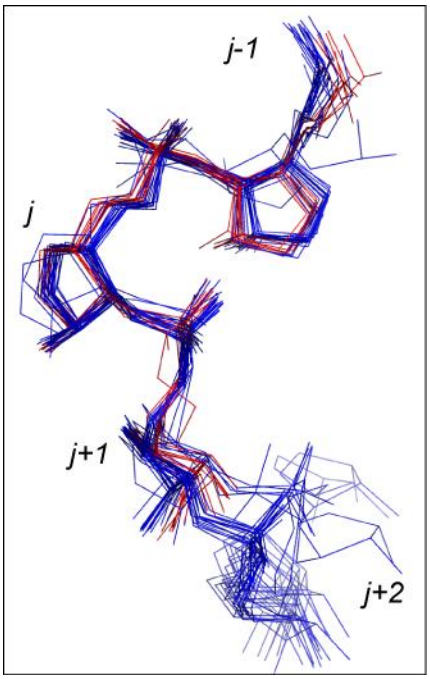

Figure 2.6 Superimposition of 31 tetraloops.
The backbone atoms of the first three residues of all s-Tl and d_2-Tl were superimposed. Bases are omitted for clarity.

d_2-Tl tetraloops: In the 10 members of this group, residue (j+2) of s-Tl is absent. Residue j+3 of s-Tl becomes j+2 of d_2-Tl. *Intra-Group Similarity*: the RMSD-AP is 0.30 Å for all backbone atoms [three residues, (j-1), (j) and (j+1)] of this group. Thus d_2-Tl tetraloops are more restrained in conformation than S-Tl tetraloops. *Inter-Group Similarity*: the RMSD-AP is 0.49 Å for the backbone atoms of the ten d_2-Tl tetraloops and those of the corresponding residues [(j-1), (j) and (j+1)] of the 21 s-Tl tetraloops. This superimposition is shown in Figure 2.6. It can be seen that deletion of residue j+2 does not appreciably change the positions of the remaining backbone atoms of these tetraloops. However deletion at the j+2 position is correlated with adjacent helical distortions such as insertions at position 3 (314, 625, 1387, 1992), clipping at position 2 (1187, 1809, 2598), base pair disruption in the stem (1500, 1596) and unhinging (1749).

i_2-Tl: In the three members of this group, a residue (i_2) is inserted at position 2, between residues (j+1) and (j+2). It should be noted that insertions in tetraloops are evident in the results of Huang et al. (Huang *et al.*, 2005). *Intra-Group Similarity*: the RMSD-AP is 0.42 Å for the three i_2-Tl tetraloop backbone atoms [residues (j-1), (j), (j+1) and (j+2), omitting the inserted residues, which show variable positions]. *Inter-Group Similarity*: the RMSD-AP is 0.76 Å for the common backbone atoms of three i_2-Tl tetraloops and the 21 s-Tl tetraloops. All three members of the i_2-Tl group show the consensus j-1 O2' to j+1 N7 hydrogen bond. One of them (1707) shows hydrogen bonds of j-1 N1(G) to the O2P of residue i_2 and j-1 N2(G) to j+2 N7(A). A second (1276) shows a contact distance just slightly greater than our hydrogen bond cut-off between (j-1) N3 and i_2 O1P. Two, with pyrimidines at the j-1 position, show hydrogen bonds of O2 (j-1) to N6 (j+2). Therefore, insertion of a residue at position 2 does not appreciably change the atomic positions or significantly alter the nature of the interactions.

$d_2i_{2(3)}$-Tl: In this tetraloop, as in the d_2-Tl group, residue (j+2) is deleted. In addition, three residues are also inserted at position 2 (indicated by $i_{2(3)}$). This tetraloop demonstrates deletion simultaneous with multi-residue insertion. *Inter-Group Similarity*: the RMSD-AP is 0.30 Å for common backbone atoms [residues (j-1), (j) and (j+1)] of the $d_2i_{2(3)}$-Tl tetraloop and the d_2-Tl group. The j-1 O2' of $d_2i_{2(3)}$-Tl interacts with the N7 of j+1. The (U) N3 of j-1 interacts with the O1P of j+2. Therefore, the three residue insertion at site 2 does not appreciably change the atomic positions or interactions of the d_2-Tl tetraloop.

i_1-Tl: In this tetraloop, there is a residue inserted at site 1, between residues (j) and (j+1). *Inter-Group Similarity*: the RMSD-AP is 0.87 Å for the backbone atoms of this tetraloop and the **i_2-Tl** tetraloops (omitting the inserted residues). The RMSD-AP is 0.82 Å for the superimposition of common backbone atoms of i_1-Tl(218) and s-Tl(805). In this tetraloop the consensus O2' j-1 to N7 and O6 j+1 interactions are observed. Therefore,

insertion at position 1 does not appreciably change the atomic positions or molecular interactions of this tetraloop.

$x_{3,2}$-Tl: In this tetraloop the positions of the bases of residues j+2 and j+3 are exchanged. *Inter-Group Similarity:* The RMSD-AP is 0.27 Å for the bases of the $x_{3,2}$-Tl(482) and the bases of s-Tl(1863). Since single residue topology variation is one of the most unexpected discoveries of the multi-resolution method, we provide an illustration of this superimposition (Figure 2.7). For the superimposition and the RMSD-AP calculation, the ordering of the residues is switched such that (j-1), (j), (j+1), (j+3) of $x_{3,2}$-Tl(482) were superimposed on (j-1), (j), (j+1), (j+2) of s-Tl(1863). We chose tetraloop s-Tl(1863) for this superimposition because it is the only standard tetraloop with the appropriate sequence. In $x_{3,2}$-Tl(482) the consensus O2' j-1 to N7 and N6 j+1 interactions are observed. In addition the N1 (j-1) to O2P (j+2) interaction is maintained. Finally, the N2 (j-1) to N7 of j+3 (which has replaced j+2) interaction is conserved. In sum, the positions of the bases and the interactions between them and with the backbone are highly conserved even though the connections linking them differ.

$x_{3,2}i_3$-Tl: In this tetraloop the positions of the bases of residues j+2 and j+3 are exchanged, and in addition, a residue is inserted at position 3. *Inter-Group Similarity:* The RMSD-AP is 0.45 Å for the common bases of $x_{3,2}i_3$-Tl and s-Tl(691,805,1327,1629), with the base ordering switched as described above, and the inserted residue omitted. In this group, the topology is the same as $x_{3,2}$-Tl group, and a residue is inserted at position 3, between (j+2) and (j+3). In $x_{3,2}i_3$-Tl(506) the consensus O2' j-1 to N7 j+1 interaction is observed. In addition the N1 (j-1) to O2P (j+2) interaction is maintained. Finally, the N2 (j-1) to N7 of j+3 (which has replaced j+2) interaction is conserved. Therefore the 3-2 switch can accommodate insertions.

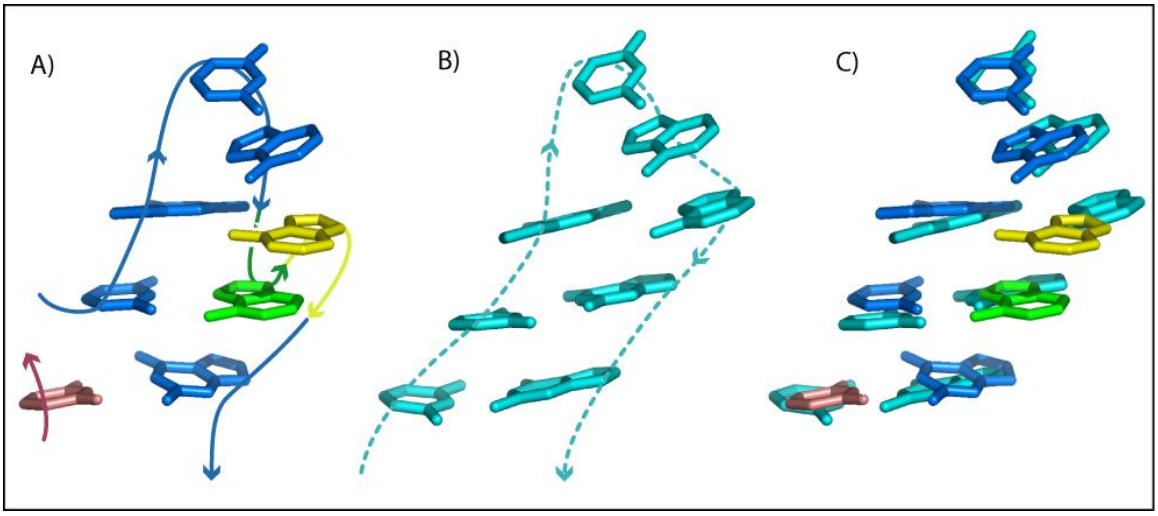

Figure 2.7 Base positions are conserved in standard and 3-2 switched tetraloops.
(A) A tetraloop with a 3-2 Switch ($x_{3,2}$-Tl tetraloop 482). The backbone connectivity is indicated by the arrows. The positions of residue j+2 (green) and j+3 (yellow) are switched relative to standard tetraloop. This tetraloop is clipped between residues j-2 and j-3. (B) A standard tetraloop (s-Tl tetraloop1863), with standard backbone connectivity. (C) Superimposition of the bases of the 3-2 Switch and the standard tetraloops. Backbone atoms were not used in the superimposition and are omitted from the diagram for clarity. All bases shown were used for the superimposition.

d_1i_0-Tl: In these two tetraloops, residue i_0 is inserted between residues (j-1) and (j) and residue (j+1) is deleted. This group is equivalent to the previously described UNCG tetraloop (Tuerk *et al.*, 1988; Cheong *et al.*, 1990; Allain and Varani, 1995; Akke *et al.*, 1997; Williams and Hall, 1999; Ennifar *et al.*, 2000). The 'looped out' N-residue of UNCG is equivalent to i_0. *Intra-Group Similarity*: the RMSD-AP is 0.41. *Inter-Group Similarity*: RMSD-AP is 1.11 Å, for common backbone atoms of the two d_1i_0-Tl tetraloops and s-Tl(805) (which is an average s-Tl tetraloop). It is not clear where this group fits in the family tree (Figure 2.4) because the cross-loop hydrogen bonding pattern is slightly different from the consensus of other tetraloops. The hydrogen bond from the O2' of residue j-1 is with the O6 of j+2, not the N7 of j+1, which is deleted from this group of tetraloops. In addition the O2 of j-1 forms a hydrogen bond with the N1 of j+2, which is a G in both members. It is conceivable that further analysis will lead to reassignment of this group to a new position in the family tree or its removal altogether.

2.4 Discussion

On one level the results here correspond well with expectations, confirming that tetraloops have well-defined conformation (given by atomic positions and torsion angles) and molecular interactions (hydrogen bonding and stacking), and sequence constraints. However we arrive at several conclusions that extend or even contradict previous work.

2.4.1 DevLS.

We propose a classification scheme where all tetraloops, U-turns, and many triloops, pentaloops, etc, are members of a common class (motif) that is elaborated with DevLS - insertions, deletions, strand clips and 3-2 switches. This simplifying scheme can be applied generally to RNA motifs (kink-turns, E-loops, etc.). In fact we observe an E-loop motif in 1JJ2 with two strand clips (residues 911-914, 1045, 1069-1072, and 1293-1294). The commonality of the DevLS between various motifs provides a powerful analytical handle for RNA analysis. One can precisely decompose and describe both polymorphism and the underlying elemental motifs. Approximately one third of the tetraloops in HM 23s rRNA contain DevLS. This significant fraction of tetraloops was not detected in our prior work (Hershkovitz *et al.*, 2003) where DevLS masked tetraloops.

The *3-2 switch* is, to our knowledge, a previously unrecognized conformational element of RNA. We are not, however, the first to observe insertions, deletions and strand clips. Insertions in tetraloops are evident in the results of Huang et al. (Huang *et al.*, 2005). Deletions in tetraloops give the U-turn motif, and some members of LPTL motif of Lee (Lee *et al.*, 2003). Insertions, deletions and strand clips in kink-turns and C-like motifs have been noted (Klein *et al.*, 2001; Lescoute *et al.*, 2005).

2.4.2 The 3-2 Switch

The 3-2 switch (Figures 2.1 and 2.3C) re-orders bases along a roughly helical trajectory such that the effective sequence differs to the primary sequence. Bases with an ordering of 1,2,3,4 in the primary sequence can rearrange, without breaking or altering bonds, to establish a three-dimensional ordering of 1,3,2,4. In a 3-2 switch the RNA backbone skips over one base, then returns to it, then proceeds on in the original direction. We observe three 3-2 switches in the HM 23s rRNA. Two are associated with tetraloops. One is associated with a clipped kink-turn (residues 42-50, 111-115, and 148-149). In addition there are several partial 3-2 switches in which bases 1,3,2 but not 4 are aligned. In sum, RNA accommodates topology variations on the dinucleotide level, whereby the positions and interactions of a series of bases can remain essentially unaltered while the backbone connection linking them varies. We believe 3-2 switches, by partially decoupling covalent sequence from effective sequence, may have significant implications in structure, reactivity and mechanism of evolutionary change.

2.4.3 Tetraloop Triplets

The 3-2 switch appears to facilitate tetraloop-tetraloop interactions. We observe that three tetraloops in 1JJ2 associate to form a tetraloop triplet. This tetraloop triplet consists of tetraloops $x_{3,2}$-Tl(482), $x_{3,2}i_3$-Tl(506) and d_2-Tl(314). Tetraloops 482 and 506, which both contain 3-2 switches, associate via an intimate face-to-face interface, which includes base-pairing interactions of A(486) with A(511) - each of these is a component of a 3-2 switch. s-Tl(314) stacks on the other two, such that all three j-residues interact. We observe a similar tetraloop triplet in the 23s rRNA of *D. radiodurans* (1NKW, ref. Harms *et al.*, 2001). In that structure, tetraloop $x_{3,2}$-Tl (487) and $x_{3,2}i_3$-Tl(510) associate via an intimate face-to-face dimer, which stacks on d_2-Tl(318). We hypothesize that tetraloop triplets play important roles in rRNA folding and stability.

2.4.4 The Tetraloop Family Tree

This tree provides a general, accurate and accessible description of tetraloops and of the relationships among them. The structure-based tree assumes that all tetraloops are members of a single motif class that varies by elaboration with DevLS. To form the tree, the forty observed tetraloops are split first in standard topology and 3-2 switch groups, and are further split by deletions and insertions, according to DEVLS positions. Alternative trees with different branching schemes are possible. The tree allows one to readily observe frequencies, relationships between DEVLS type and sequence, etc. There are many possible family trees. In fact we believe that it may be appropriate, if one were to ignore history, to recast the 10 d_2-Tl tetraloops, which have the greatest conservation of sequence and atomic positions, as the parent motif. In this scheme the current s-Tl group would contain an insertion after residue j+1. With additional data, and more statistically meaningful tree may allow one to infer evolutionary relationships and mechanisms.

2.4.5 Deleted Tetraloops, U-Turns and Lonepair TriLoops

The consensus hydrogen bonding interactions and sequence of s-Tl and d_2-Tl tetraloops are consistent with the U-turn motif (Quigley and Rich, 1976; Jucker and Pardi, 1995a; Auffinger and Westhof, 1999; Gutell *et al.*, 2000). The d_2-Tl tetraloop appears to be essentially identical to the original U-turn of Quigley and Rich (Quigley and Rich, 1976). Gutell and coworkers have used sequence covariation approaches along with visual inspection to detect and describe a motif they refer to as the LPTL (Lee *et al.*, 2003). There is considerable overlap of the LPTL motif of Gutell with the d_2-Tl group described here (Table 2.2). However important distinctions distinguish the two motifs. The d_2-Tl group is characterized by conserved conformation (torsion angles and atomic positions) and molecular interactions, which are also common to the s-Tl group and other tetraloops. By contrast some members of the LPTL group are conformationally distinct

from others, and from standard tetraloops. Some d_2-Tl tetraloops lack closing base-pairs altogether, and so are not consistent with the LPTL definition.

2.4.6 Variation in the Helix Capping Function of Tetraloops

Here, seven of forty tetraloops are strand clipped (Figure 2.5). Strand clipping allows RNA segments that are remote in the primary sequence to join to form a motif (Nagaswamy and Fox, 2002; Lee *et al.*, 2003; Lescoute *et al.*, 2005). Strand clipped tetraloops cap pseudo–helices, which commonly do not appear as stems in secondary structure representations. One observed tetraloop caps neither a helix nor a pseudo-helix, but by all other criteria is an average d_2-Tl tetraloop. This tetraloop is 'unhinged' from any helical regions. None of the d_2-Tl tetraloops cap a clean *unperturbed* helix. In sum, a 'tetraloop' is not necessarily a terminal loop, which by classical definition allows a strand of RNA to fold back on itself to form a helical stem (Moore, 1999).

Table 2.2 Comparison of Lonepair Triloops (LPTL) (Lee *et al.*, 2003) and d_2-Tl tetraloops.

	Class[1]	Number[2]	Number[3]	Group[4]
In common	IA	313:317	314	d_2-Tl
	IA	624:628	625	d_2-Tl
	IA	1388:1392	1389	d_2-Tl
	IB	1186:1190	1187	d_2-Tl
	IB	1808:1912	1809	d_2-Tl
	IB	2597:2601	2598	d_2-Tl
	IIA	505:509	506	$x_{3,2}i_3$-Tl
	IA	481:485	482	$x_{3,2}$-Tl
	IIB	482:486	482	$x_{3,2}$-Tl
Identified in present work, not in Lee et al.	Three residues inserted at insertion site 2.		392	$d_2i_{2(3)}$-Tl
	No LP; no residue j-2 and j+3 unpaired[5].		1500	d_2-Tl
	No LP; no residue j-2 and j+3 unpaired.		1596	d_2-Tl
	No LP; no residue j-2 and j+3 unpaired.		1749u[6]	d_2-Tl
	No LP; no residue j-2 and j+3 unpaired.		1992	d_2-Tl
Identified in Lee et al. (Lee *et al.*, 2003), not in present work	IB	125:129	Variant conformation[7].	
	IB	335:339	Variant conformation[7].	
	IIB	326:330	Variant conformation[7].	
	IB	1651:1655	Variant conformation[7].	
	IB	1966:1970	Variant conformation[7].	
	IB	2482:2486	Variant conformation[7].	

1) Classification scheme of lonepair triloops (LPTLs) (Lee *et al.*, 2003).
2) Initial and final residue numbers from PDB entry 1JJ2.
3) Residue number of position j-1 (Figure 2.3) from PDB entry 1JJ2.
4) Tetraloop group (Figures 2.3 and 2.4), indicating DevLS position and type.
5) LP indicates lonepair.
6) The letter "u" indicates an unhinged d_2-Tl tetraloop.
7) The conformational states and molecular interaction of these loops are not similar to those of tetraloops.

2.5 Acknowledgements

The authors thank Jane Richardson, Laura Murray from Duke University and Steve Harvey from Georgia Tech for helpful discussions.

CHAPTER 3

STRUCTURAL ALIGNMENT BY ANCHORED SEGMENTS: 23 S RRNAS OF THERMUS THERMOPHILUS AND HALOARCULA MARISMORTUI

3.1 Introduction

The ribosome is an ancient evolutionary machine that synthesizes protein. A functional bacterial ribosome (70S particle) is composed of a small 30S subunit (the 16S rRNA plus 21 proteins) and a large 50S subunit (the 23S and 5S rRNAs plus 31 proteins).

Comparison of rRNA sequences is widely used for determination of phylogenetic relationships. Comparison rRNA sequences led Woese and Fox to the discovery of Archaea, the third kingdom of life (Woese and Fox, 1977; Magrum *et al.*, 1978; Woese *et al.*, 1978). Their analysis produced the first phylogenetic tree that included prokaryotes, protozoa, fungi, plants, and animals (Woese, 1987).

Comparison of rRNA sequences is also used for determination of rRNA secondary structure. Understanding the relationships between 1D RNA structure (i.e., sequence) and 2D structure (i.e., stem/loop ensemble) remains a fundamental open problem in molecular biology (Zuker, 2000; Mathews *et al.*, 2004). Recent algorithmic advances in RNA secondary structure prediction include combinations covariation analysis (Gutell *et al.*, 2002; Gardner and Giegerich, 2004), improvements in free energy minimization (Mathews and Turner, 2006), and developments in partition function and statistical sampling approaches (Mathews, 2004; Ding, 2006).

The method of covariation analysis (reviewed by Gutell, Gutell *et al.*, 2002) was first used by Holley to determine the secondary structure of tRNA (Holley *et al.*, 1965), and by Woese and Fox to determine the secondary structure of 5s rRNA (Fox and Woese,

1975). Covariation was then used to determine secondary structures of the 16S (Woese *et al.*, 1980) and 23S rRNA (Noller *et al.*, 1981). Because of their utility in determining phylogentic relationships and secondary structure, greater than 10,000 16S and 16S-like rRNA and 1,000 23S and 23S-like rRNA genes have been sequenced (Cannone *et al.*, 2002).

More recently, analysis of rRNA has moved from 1D and 2D to 3D. The database currently contains structures of ribsomes from five organisms [*Thermus thermophilus*, x-ray, 2.8 Å (Selmer *et al.*, 2006), *Escherichia coli*, x-ray, 3.2 Å (Berk *et al.*, 2006), *Haloarcula marismortui*, x-ray, 2.4 Å, LSU only (Ban *et al.*, 2000), *Deinococcus radiodurans*, x-ray, 3.1 Å, LSU only (Harms *et al.*, 2001), and *Saccharomyces cerevisiae*, cryo-EM, 11.7 Å (Spahn *et al.*, 2004)].

The availability of homologous 3D structures in principle allows one to align ribosomal rRNAs on the basis of conformation and molecular interactions. Structural superimposition would allow one to (i) identify regions of structural conservation and diversity, (ii) determine relationships between structural and sequence variation, (iii) dock elements of one structure onto another, (iv) determine relationships between conformational change, assembly binding, and mechanism, and (v) create in silico chimeras.

However structural alignment must take into account insertions, deletions and sequence variation. The best global superimposition of large RNAs must omit insertions and deletions.

Here we describe a method (SAAS; Structural Alignment by Anchored Segments) to accurately align and superimpose very large RNAs . We apply the method to the structural alignment and superimposition of 23S rRNAs from *Haloacrula marismortui* (HM^{23S}) and *Thermus thermophilus* (TT^{23S}). This method uses tetraloops as anchors to define RNA segments within ribosomes. Homologous segments are then paired, using secondary structure maps as guides. The paired segments are aligned heuristically and

iteratively based on 3D structure, then locally superimposed. An iterative rigid body superimposition is then performed to give the global superimposition of entire ribosomal structures.

3.2 Methods

3.2.1 PBR Space Analysis

In a multi-scaled structural motif pattern recognition approach previously developed in our lab (Hsiao *et al.*, 2006), (see Chapter 2), the atoms from a coordinate file are extracted and transformed into **PBR** space (**P** indicates Phosphate, **B** indicates Base, and **R** indicates Ribose). In PBR space the resolution and complexity are attenuated. This change of scale reveals four RNA **DevLS** (**Dev**iations of **L**ocal **S**tructures), which are 3-2 switches, insertions, deletions and strand clips. With the **PBR** analysis, we are able to identify and classify all tetraloops in the 3D structures of HM^{23S} and TT^{23S} rRNAs.

3.2.2 Structural Motif Anchor

RNA structural motifs are abundant elements that stabilize folded RNA (Moore, 1999). The tetraloop is the simplest, most frequent, most well-known, and best characterized non-helical RNA motif. Tetraloops are useful 'anchors' for analysis and comparison of 3D structures. Large RNAs can be split into tractable segments by in silico cutting at the tetraloops.

3.2.3 Structural Alignment by Anchored Segments (SAAS)

Here we introduce the **S**tructural **A**lignment by **A**nchored **S**egments (**SAAS**) approach for structural alignment and superimposistion. Structural alignment is iterated and optimized within each segment, followed by superimposition. The SAAS approach is

capable of aligning large RNAs on the basis of structure, and (ii) finding the best superimposition of large RNAs. In applying SAAS to HM^{23S} and TT^{23S}, we follow this path:

(i) Identify and classify tetraloops in the 3D structures of HM^{23S} and TT^{23S} rRNAs.

(ii) Mark anchors (tetraloops) on the 2D structures of HM^{23S} and TT^{23S} and establish correspondence of anchors within the two structures.

(iii) Map the anchors between 23S-rRNA of HM and TT and define the paired segment ends. The segments are called Segment Alignment Pairs (SAPs). SAPs are paired homologous segments of RNA of 25 to 189 residues.

(iv) Within each SAP, heuristically find the best structural alignment and determine locations of insertions and deletions.The heuristic process follows this path:

(a) The paired tetraloop anchors are corrected for subfamily differences. Insertions within tetraloops are deleted. Residues paired with deletions within tetraloops are deleted.

(b) SAPs of the same length are superimposed and visually inspected.

(c) For SAPs that differ in length by one residue, each residue is systematically omitted from the longer segment, with a fit performed after each omission. The best fit determines the position of the insertion in the segment of greater length.

(d) For SAPs that differ by two, three or four residues, residues are enumerated from the longer segment, with a fit performed after each omission. For example, if the length difference is two, every pair of residues is stepwise omitted, in all combinations.

(v) Combine SAPS in a stepwise process and superimpose.

(vi) Perform visual inspection of RNA segments that were omitted from the global superimposition, to determine positions of "secondary anchors". The non-SAP RNA is visually inspected to determine if portions of non-aligned segments can be included in the global superimposition. Secondary anchors are non-tetraloop

RNA anchors at regions of where reasonable superimposition terminates, as determined by visual inspection. The goal of this step is to maximize the fraction of RNA used in the superimposition. I use these secondary anchors to increase the amount of RNA in the fit, and iterate the rigid body superimposition (return to step iv).

(vii) The aligned rRNAs are used to globally superimpose the complete LSUs, including non-aligned rRNA, proteins, ions and solvent.

(viii) Determine relationships between sequence conservation, structural conservation and effects of folding and assembly.

3.3 Results

SAAS, in a divide and conquer strategy, utilizes tetraloops as structural anchors to define corresponding segment pairs that are used iteratively, heuristically, and visually in structural alignment. SAAS has been used here successfully to align HM^{23S} and TT^{23S} and LSUs. It allows us to use 73% of RNA backbone atoms (~2129 residues). After the superimposition the Global RMSD of backbone atomic positions is 1.2 Å.

3.3.1 Tetraloop Family Trees of HM^{23S} and TT^{23S}

PBR space allows one to identify and classify tetraloops in the 3D structures of RNA. The classification scheme is based on DevLS (Hsiao *et al.*, 2006) (see Chapter 2). Once classified, tetraloops of both HM^{23S} and TT^{23S} are entered into the tetraloop family (Figure 3.1). The tree branches first at the root of $X_{3,2}$ and standard topology, which and leafs out with insertions and deletions at various positions. The two primary groups fall in the leaves of s-Tl and d_2-Tl. The population of standard tetraloops in TT^{23S} (12 members) is about half that in HM^{23S} (21 members).

3.3.2 Tetraloop Mapping

After identifying and classifying the tetraloops in three dimensions (Figure 3.1), we establish their correspondence in the 2D maps (Figure 3.2) of the two 23S rRNAs. The strong resemblance of the 2D maps of HM^{23S} and TT^{23S} (Figure 3.2) facilitates the establishment of the correspondence of tetraloops. The tetraloop correspondence is illustrated in the 1D map shown in Figure 3.3.

3.3.3 Structural Alignment

The 1D anchor map shows the correspondence of tetraloops as guides to 'divide' the entire 23S rRNAs into manageable segments, of various lengths. Once the paired segments of HM^{23S} at TT^{23S} are aligned, they are termed '**S**egment **A**lignment **P**airs' (SAP). Structural alignment must consider and account for differences in DevLS, which are 3-2 switches, insertions, deletions, and strand clips. Sixteen SAPs, in which the head and tail are capped by tetraloops are defined (entries $1631A^{TT23S}$ and 2205^{TT23S} are the exceptions) [Table 3.1]. I was able to align 16 of 31 segments. The workable length difference for alignment of paired segments is less than five residues. Greater differences in length would consume computational resources beyond that available during the heuristic structural alignment.

3.3.4 Superimposition

In a 'conquer' process, the aligned SAPs are combined with some non-SAP RNA, and fit to give the global superimposition the 3D structures of the rRNAs. The final superimposition (Figure 3.4) uses 73% of RNA backbone atoms (~2129 residues) to give an overall RMSD of backbone atomic positions of 1.2 Å.

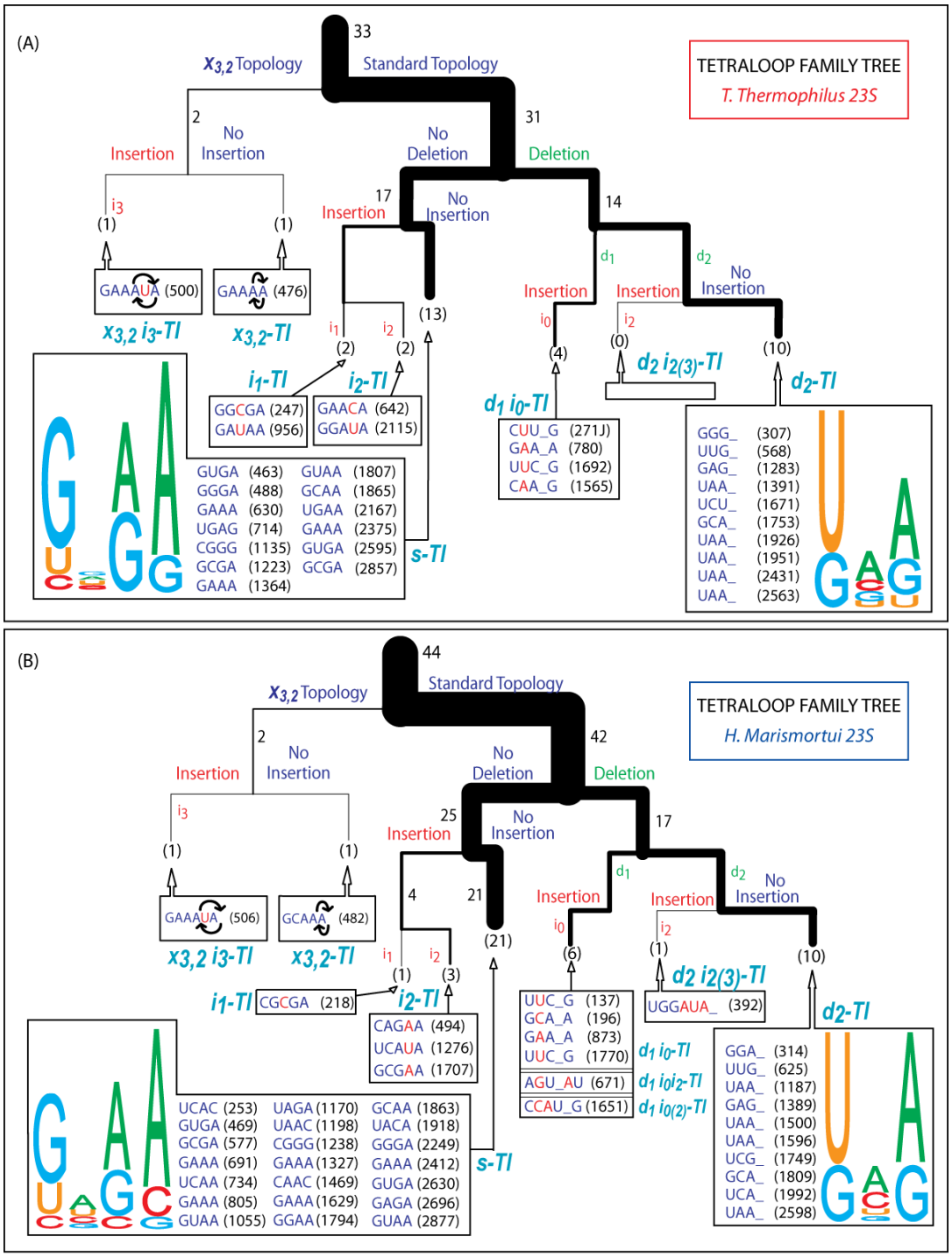

Figure 3.1 Tetraloop Family Trees.
(A) The tetraloop family tree of the TT^{23S} rRNA. Thirty three tetraloops are identified and classified by types of DevLS. (B) The tetraloop family tree of the HM^{23S} rRNA [originally from Hsiao (Hsiao *et al.*, 2006)]. A new discovered tetraloop (entry 1651) is added. The tetraloop totals are indicated at the family tree roots. Insertion positions are indicated in red text. Deletion positions are indicated in green text. The positions of deleted residues are marked by underscores. Number of occurrences is indicated in black, with line widths proportional to frequency.

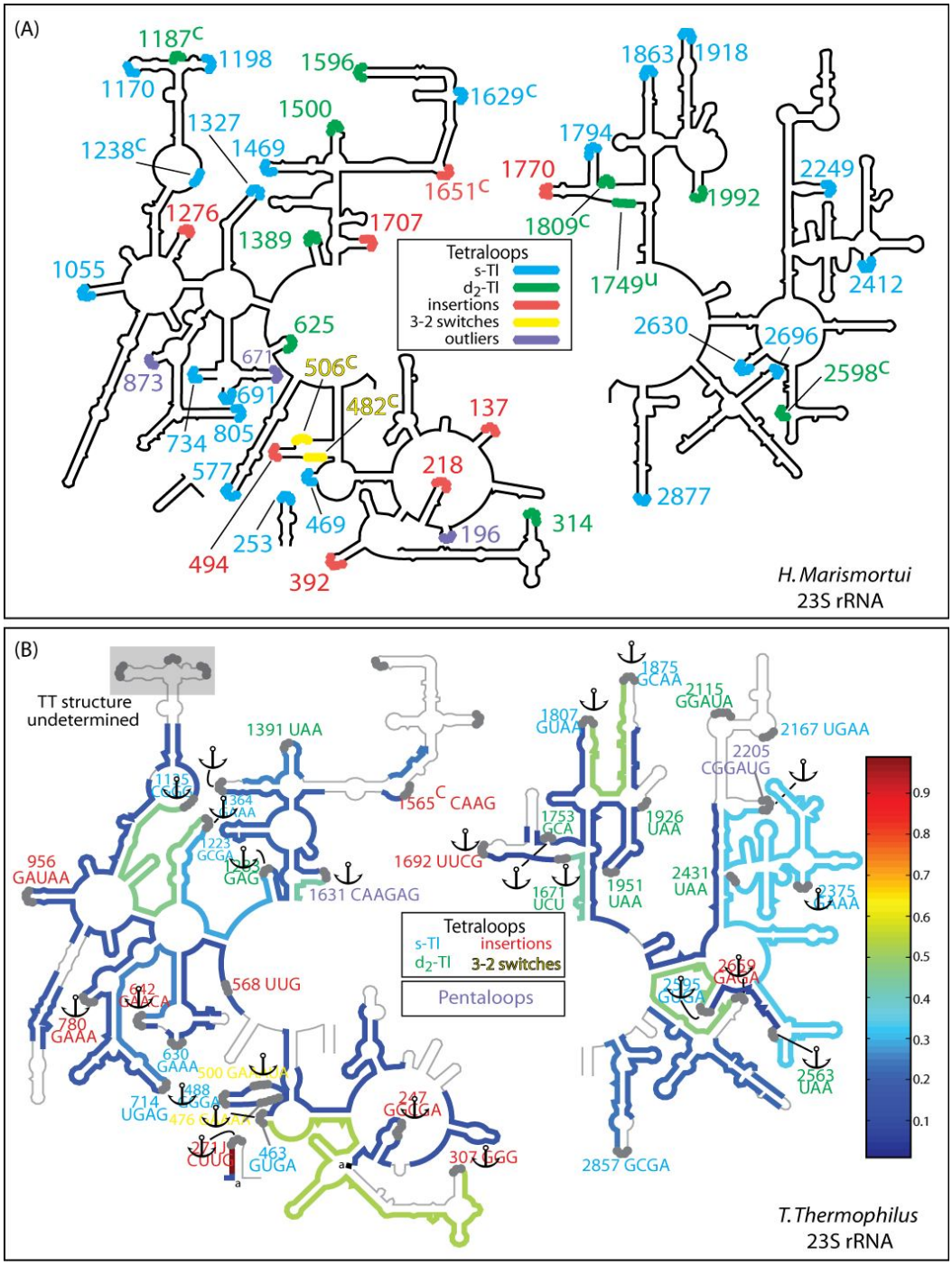

Figure 3.2 Secondary structures of the TT^{23S} and HM^{23S}.
(A) The secondary structure of the HM^{23S} rRNA. Entry 1651, a new subfamily of tetraloop that is elaborated with an insertion and a strand clip, is shown. (B) The secondary structure of the TT^{23S} rRNA. The anchor symbols indicate the tetraloops that are used to define the RNA structural segments. The RNA structural segments are thick lines colored by the RMSDs of the atomic positions between HM^{23S} and TT^{23S} (Table 3.1), deep blue indicates more highly conserved, and deep red indicates more highly divergent. The gray thin line indicates the remaining structural segments that are not used in the superimpositions before any visual inspection. Tetraloop locations are colored by types of DevLS; yellow indicates 3-2 switches, green indicates deletions, red indicates insertions, superscripted c's indicate strand clips, and cyan indicates standards.

3.3.5 Local versus Global- SAP Superimposition

Significance of Local and Global Deviations. The RMSDs of backbone atomic positions after local SAP superimposition indicate the extent of similarity of the conformations of two segments. The difference in the RMSDs of the local and global SAP superimpositions indicates the degree of similarity of global position of two segments in the folded ribosome.

SAP_{14}: Conserved Conformation and Position

The longest SAP (SAP_{14}; Structural Alignment Pair #14) is 189 residues in HM^{23S} and 191 residues in TT^{23S}. The heuristic fit indicates that residues 2431U and 2432A are insertions of TT^{23S} (or are deletions of HM^{23S}) and so were excluded from the alignment. The Local-SAP Superimposition for SAP_{14} gives an RMSD of backbone atomic positions of 1.23 Å. The small magnitude of this RMSD indicates that the backbone positions of the segment are highly conserved in HM and TT. The Global-SAP Superimposition gives an RMSD of backbone atomic positions of 1.28 Å. The similarity of the local and the global RMSDs indicates that both the position and conformation of SAP_{14} of HM^{23S} is unchanged relative to TT^{23S}.

SAP_{10}: Conserved Conformation and Varying Global Position

SAP_{10} is 25 residues in length, and is the shortest segment identified. The length of this segment is conserved in HM^{23S} and TT^{23S}, and the heuristic fit indicates the absence of insertions and deletions. The Local-SAP Superimposition gives an RMSD of backbone atomic positions of 0.33 Å indicating highly conserved conformation of these two segments. However the Global-SAP Superimposition is 0.79 Å. The 0.46 Å difference between the Local and Global SAP Superimpositions indicates that the position of the segment is different in HM^{23S} and TT^{23S}. The difference may arise during folding or during ribosomal assembly. The location of this segment at the LSU/SSU interface, with direct interactions with the 16S rRNA, suggests that the observed shift in

position occurs during assembly. The HM (PDB entry: 1JJ2) is not assembled. The TT (PDB entry: 2J00, 2J01) is a fully assembled ribosome.

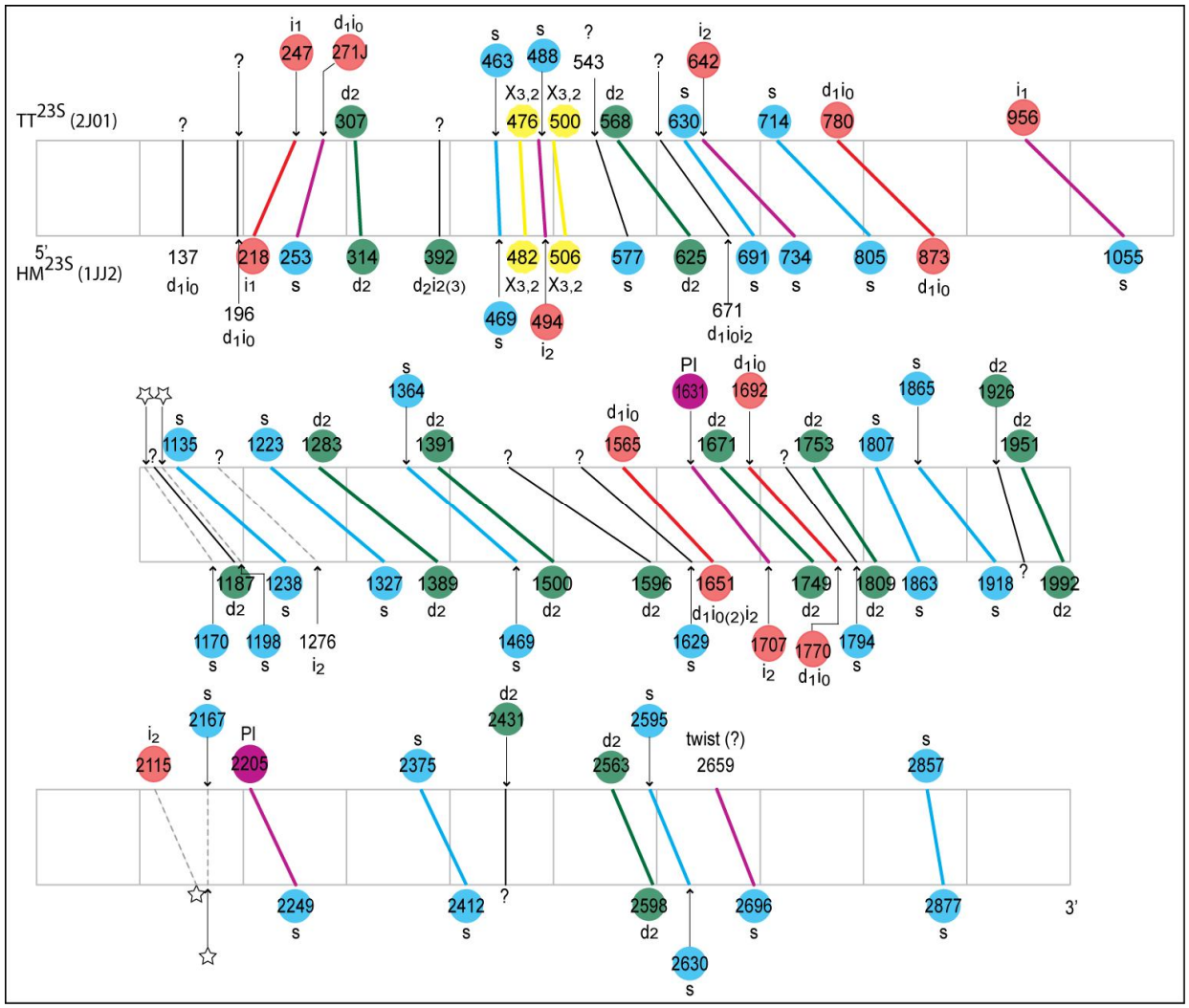

Figure 3.3 1D structural Map.
Anchors are mapped from sequence of the TT^{23S} (PDB entry: 2J01) [top] to that of the HM^{23S} (PDB entry: 1JJ2) [bottom]. The residue number of the first residue is given for each tetraloop. Pl indicates pentaloop. Tetraloops are mapped by location with type indicated by color. The red line maps insertions, the green line maps deletions, the yellow line maps 3,2-switches, the cyan line maps standard tetraloops, and the purple line maps different types of tetraloops or pentaloops. The dotted line indicates lack of corresondence. The question mark indicates non-motif. The star symbol indicates structurally undetermined.

Table 3.1 The SAPs (Segment Alignment Pairs)

SAP #	HM Start:End (# of residues)	TT Start:End (# of residues)	length	Sequence Diff. (%)	Local RMSD	Global RMSD	Δ RMSD	Length Diff. (Δ=)	Alignment Anchors: Tl-(HM)-Tl Tl-(TT)-Tl	Residue Excluded: (1) in HM (2) in TT
1	218:253(36)	247:271J(35)	35	42.9 (28.6)[j]	3.74 (0.5)[j]	0.74	0.24	1	i_1 — s i_1 — d_1i_0	(1) 244(C)[a] (2)
2	314:472(159)	307:466(162)	159	43.4	1.93	2.03	0.1	3	d_2 — s d_2 — s	(1) (2) 346(A)[a] 387(U)[a] 388(G)[a]
3	469:511(43)	463:505(42)	42	28.6	0.62	0.82	0.2	1	s — $x_{3,2}$ s — $x_{3,2}$	(1) 497(A)[a] (2)
4	734:808(75)	642:717(73)	71	40.3	0.99	1.23	0.24	2	s — s i_2 — s	(1) 743(G)[a], 745:746[a] (2) 645(C)[b, h]
5	805:876(72)	714:783(70)	70	41.4	0.66	0.96	0.3	2	s — d_1i_0 s — d_1i_0	(1) 835(U)[a], 825(U)[f] (2)
6	1238:1330(93)	1135:1226(93)	93	62.6	1.72	1.74	0.02	0	s — s s — s	(1) 1288(U)[f], 1279(U)[g] (2) 1205(U)[f], 1220(A)[f],
7	1327:1391(65)	1223:1285(63)	63	52.4	1.07	1.11	0.04	2	s — d_2 s — d_2	(1) 1378(G)[a], 1379(A)[a] (2)
8	1389:1472(84)	1283:1367(85)	84	34.5	0.79	0.94	0.15	1	d_2 — s d_2 — s	(1) (2) 1349(A)[a]
9	1707:1751(45)	1631:1673(43)	43	50.0	1.59	1.66	0.07	2	i_2 — d_2 Pl[i] — d_2	(1) 1707(G)[h], 1710(A)[h], 1730(G)[f], (2) 1634(A)[g]
10	1749:1773(25)	1671:1695(25)	25	16.0	0.33	0.79	0.46	0	d_2 — d_1i_0 d_2 — d_1i_0	(1) (2)
11	1809:1866(58)	1753:1810(58)	58	39.7	0.60	0.74	0.14	0	d_2 — s d_2 — s	(1) (2)
12	1863:1921(59)	1807:1877(62)	59	50.9	1.86	1.90	0.04	3	s — s s — s	(1) (2) 1865(G)[a], 1866(C)[a], 1876(A)[a]

Table 3.1 The SAPs (Segment Alignment Pairs) [cont'd]

SAP #	HM Start:End (# of residues)	TT Start:End (# of residues)	length	Sequence Diff. (%)	Local RMSD	Global RMSD	Δ RMSD	Length Diff. (Δ=)	Alignment Anchors: Tl-(HM)-Tl Tl-(TT)-Tl	Residue Excluded: (1) in HM (2) in TT
13	2249:2415(162)	2205:2378(165)	161	33.1	1.27	1.38	0.11	3	s — s Pl — s	(1) 2339:2344[c), f)], 2291(A)[f)], 2392:2394[g)] (2) 2206(G)[h)], 2218(U)[h)], 2308:2309[a)], 2305:2307[f)], 2310[f)]
14	2412:2600(189)	2375:2565(191)	189	35.1	1.23	1.28	0.05	2	s — d_2 s — d_2	(1) 2468(A)[f)] (2) 2431(U)[a)], 2432(A)[a)], 2433(A)[f)]
15	2598:2633(36)	2563:2598(36)	36	19.4	0.34	0.40	0.06	0	d_2 — s d_2 — s	(1) (2)
16	2630:2699(68)	2595:2662(68)	68	41.2	1.78	1.81	0.03	0	s — s s — twist	(1) 2665(A)[d)], 2666(U)[d)] (2) 2629(A)[e)]

a) The excluded residues are determined by first trial of the heuristic superimpositions.
b) PDB entry 2JJ1: residue 652(C) should be correctedin the PDB entry to 653(C); Residue 652(U), 654(U) and 655(A) are structurally undetermined.
c) PDB entry 1JJ2: residues 2339 to 2343 are structurally undetermined.
d) PDB entry 1JJ2: residue 2665(A) and 2666(U) are structurally undetermined.
e) PDB entry 2JJ1: atoms P, O1P and O2P of residue 2629(A) are excluded during superimpositions.
f) The excluded residues are determined by visual inspection.
g) The excluded residues are determined by the second trial of the heuristic superimpositions.
h) Residues are manually deleted before the heuristic superimposition so that paired tetraloops contain the same number of residues, in spite of insertions or deletions.
i) Pl indicates pentaloop.
j) The RMSD of this segment alignment pair drops after placing the secondary anchors.

3.3.6 A new tetraloop subfamily

A comparison of the 2D maps of HM^{23S} and TT^{23S} (Figure 3.2), marked by tetraloop structural anchors led to my discovery of new tetraloop subfamily. A tetraloop in the TT 2D map appeared to be absent from the HM 2D map. Visual inspection of the HM 3D structure reveals consensus tetraloop molecular interactions as described in our prior work (Hsiao *et al.*, 2006) (see Chapter 2). This tetraloop is adorned by DevLS in a combination not observed previously.

$d_1i_{0(2)}$-Tl tetraloop. This new tetraloop subfamily is *$d_1i_{0(2)}$-Tl* [tetraloop family tree entry 1651^{HM23S}] It has two residues ($C1652^{HM23S}$ and $A1653^{HM23S}$) inserted between residues (j-1) and (j). In addition, residue (j+1) is deleted. Finally there is strand clip just after position 2. In this tetraloop, one of the inserted residues is similar to 'looped out' residue i_0 of the classic *d_1i_0-Tl*. The consensus hydrogen bond from the O2' of residue (j-1) is with the O5' of this 'looped out' residue. The O2' of the 'looped out' residue i_0 is hydrogen bonded to residue (j+2) [residue G]. A remote sequence of RNA stacks on the loop to form the 3' stem of the strand, giving the strand clip. The RMSD of atomic positions of this new tetraloop and the *d_1i_0-Tl* is 0.80 Å. This new tetraloop subfamily is not observed elsewhere in the 23S rRNA of HM or TT, but is in fact equivalent to the previously described UNCG tetraloop (Tuerk *et al.*, 1988; Cheong *et al.*, 1990).

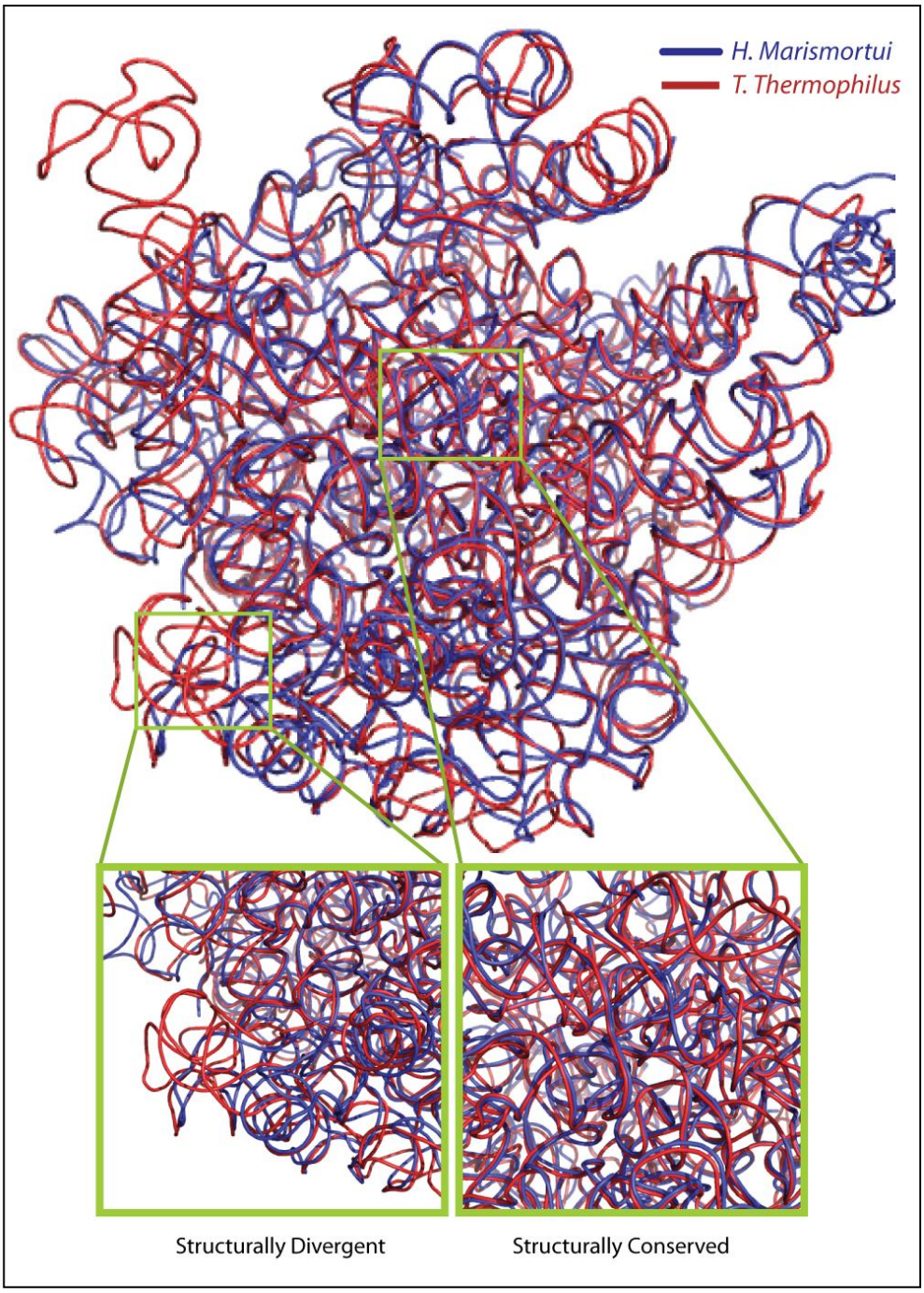

Figure 3.4 3D global view of the superimposition of HM^{23S} and TT^{23S}.
Shown is cartoon representation of TT^{23S} (red) and HM^{23S}(blue). This superimposition used 73% of RNA backbone atoms (~2129 residues); the overall RMSD is 1.2 Å.

3.4 Discussion

The availability of homologous 3D structures allows one to align ribosomal rRNAs on the basis of conformation and molecular interactions. Here we describe the structural alignment of 23s rRNAs and superimpositions of the two corresponding LSUs. Our approach provides an accurate superimposition (aligned 73% of backbone atoms; overall RMSD 1.2 Å) of HM^{23S} and TT^{23S}, providing a bridge between 1D RNA structures (sequence) and 3D RNA structures.

To define RNA structural segments, we identify tetraloops in the 3D structure, locate the tetraloops onto the 1D and 2D maps, and establish their correspondence in the two different 23s rRNAs (Figure 3.3). The RNA structural segments defined by tetraloops allow us to break the complexity of ribosomes into parts. Using this divide and conquer approach we have obtained an accurate superimposition. This superimposition allows us to determine, at atomic resolution, structural relationships between two ribosomes that are widely dispersed in the evolutionary tree. This process has led us to reanalyze and redefine the tetraloop.

3.4.1 RNA Tetraloop

The RNA tetraloop motif, a four-nucleotide terminal / internal loop, is adorned by four DevLS (**D**eviation of **L**ocal **S**tructure) (Hsiao *et al.*, 2006), which are 3-2 switch, insertion, deletion, and strand clip. The tetraloop motif is considered the simplest, the smallest and the most frequent element among all the RNA structures. Early tetraloops are broadly grouped by sequence into three classes, which are GNRA (Woese *et al.*, 1990; Jaeger *et al.*, 1994), UNCG (Tuerk *et al.*, 1988) and CUUG (Woese *et al.*, 1990; Jucker and Pardi, 1995b), and recently re-defined and unified by topology, geometry and DevLS into a tetraloop family tree (Hsiao *et al.*, 2006), a tree formalism that represents the relationships between groups of tetraloop. Among the tetraloop family trees of HM^{23S}

and TT^{23S} (Figure 3.1), the standard tetraloop (s-Tl) is the most populated (observed frequency: 21 out of 44 in HM^{23S} and 13 out of 33 in TT^{23S}) that sequence logo majored in a GNRA class, (where N can be any nucleotide and R is either G or A). In sum (Figure 3.5), the s-Tl group is characterized by (i) there is a cross-strand stack between residue j – 1 (G) and residue j + 3, (ii) the residue j + 2 (A), an unhappy A, is unpaired (1 H bonding) and unstacked on the 3' site, (iii) three consensus intra-loop interactions are *sugar to base* interaction: [O2' (residue j – 1, G) to N7 (residue j + 1, G)], *base to phosphate* interaction: [N1 and N2 (residue j – 1, G) to O2P (residue j + 2, A)], and *base to base* interaction: [N2 (residue j – 1, G) to N7 (residue j + 2, A)], (iv) the backbone torsion angle α of residue j (N) goes to +165° from -65° of A-helix, and (v) a three bases stack, N-R-A stack [residues (j)-(j+1)-(j+2) stack], is commonly seen in the GNRA class tetraloop.

3.4.2 Structural conservation and sequence conservation

Structural alignment and superimposition allows us to determine relationships between structural conservation and sequence conservation between HM and TT. One can observe from Figure 3.6 that local structural divergence generally increases with sequence divergence. The SAP with the lowest sequence divergence shows the smallest local structural divergence. The SAP with the greatest sequence divergence has among the largest structural divergence. However the signal is noisy in that some SAPs with reasonable divergent sequence (40%) show highly conserved structure (RMSD of atomic positions of backbone atoms is 0.6). For this SAP_{11} the structure is more conserved that predicted by the sequence. It appears that 3D structure and sequence evolve at different rates in various regions of the ribosome.

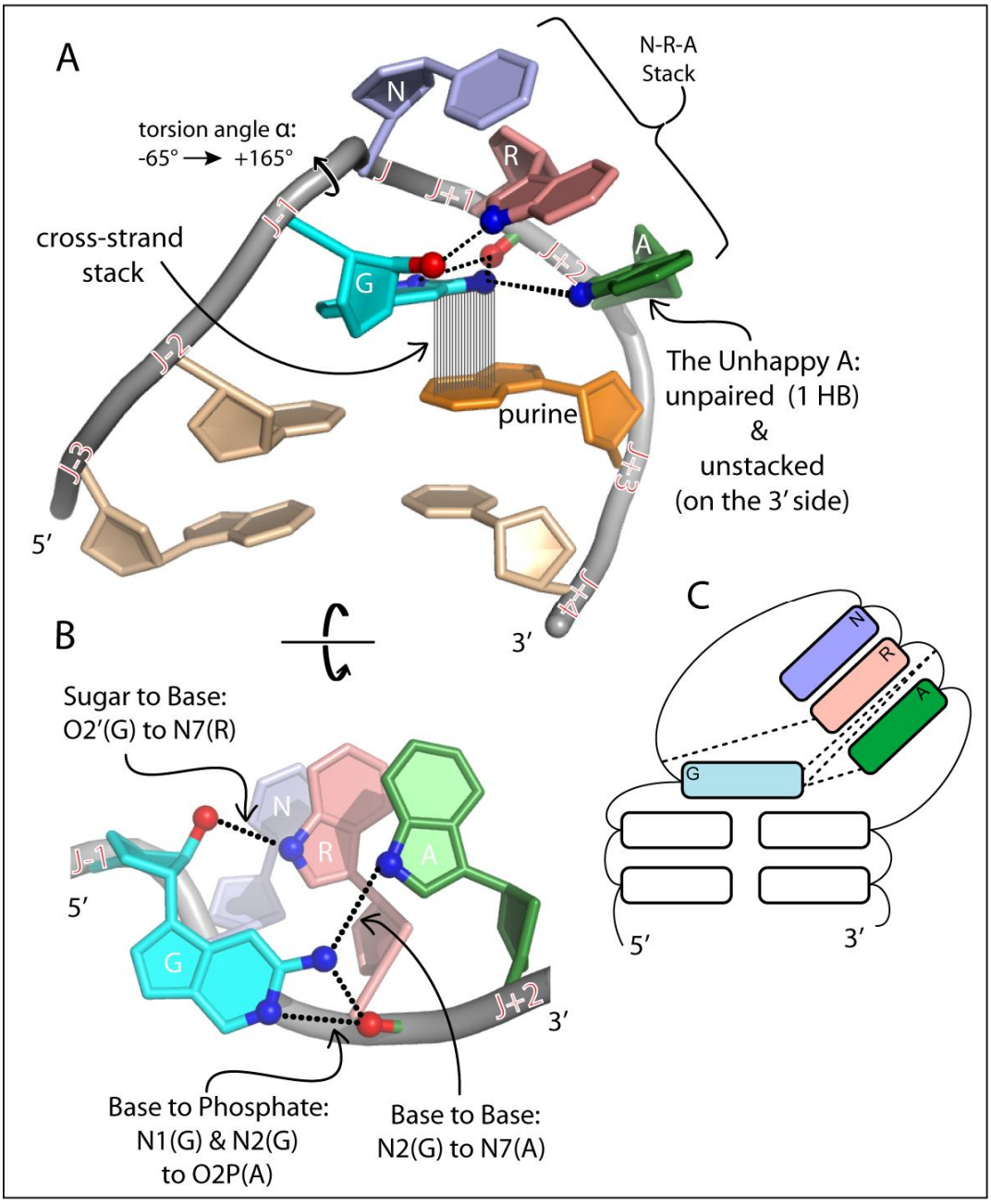

Figure 3.5 The summary of the GNRA tetraloop.
(A) A representative GNRA tetraloop, from the 3D structure from 1JJ2. The backbone is represented by a tube. Stacking is represented by shading. (B) This view is rotated 90° relative to panel A. Three consensus intra-loop hydrogen bonding interactions are shown. (C) A schematic representation of the GNRA tetraloop.

In Figure 3.7 I colored the 2D map of TT^{23S} by RMSD of atomic positions of TT^{23S} versus HM^{23S}. The 2D map also indicates the degree of sequence conservation. However in this case, the sequence conservation is not between HM^{23S} and TT^{23S} but for over 500 23S rRNA sequences, as determined by Gutell. One can observe that both the sequence and the structure are highly conserved within the PTC. In region domain II of A-helix 27 the sequences are highly divergent, while the structures are reasonably conserved (global RMSD: 0.61 Å). In region domain I of A-helix 13 the sequences are conserved, while the structures are divergent (global RMSD: 10.6 Å).

3.4.3 The Tetraloop Triplet assembly

Structural alignment and superimposition allows us to identify and characterize conserved molecular interactions and motifs, and their associations. As I noted (Hsiao *et al.*, 2006), a tetraloop triplet is seen in the LSUs of *H. marismortui* and *D. radiodurans* (1NKW, ref. Harms *et al.*, 2001). Here we report the tetraloop triplet is also seen in the 23s rRNA of *T. thermophilus*, and is in fact a component of an extended, four member tetraloop assembly, called the Tetraloop Triplet assembly (Figure 3.8). Our superimposition indicates that the position and location of the assembly are highly conserved in HM^{23S} and TT^{23S}. The RMSD of backbone atomic positions, for the globally superimposed LSUs, for the 37 residues that form the assembly is 0.91 Å. The RMSD of backbone atomic positions of this assembly for the locally superimposed LSUs is 0.62 Å. The 0.29 Å difference between the local and global superimpositions indicates that the Tetraloop Triplet assembly shifts in HM^{23S} relative to TT^{23S}. This movement may arise during folding or during LSU assembly. Tetraloop Triplet assembly consists of part of the ribosomal exit tunnel. It forms a continuous helical conformation on the surface of the ribosome.

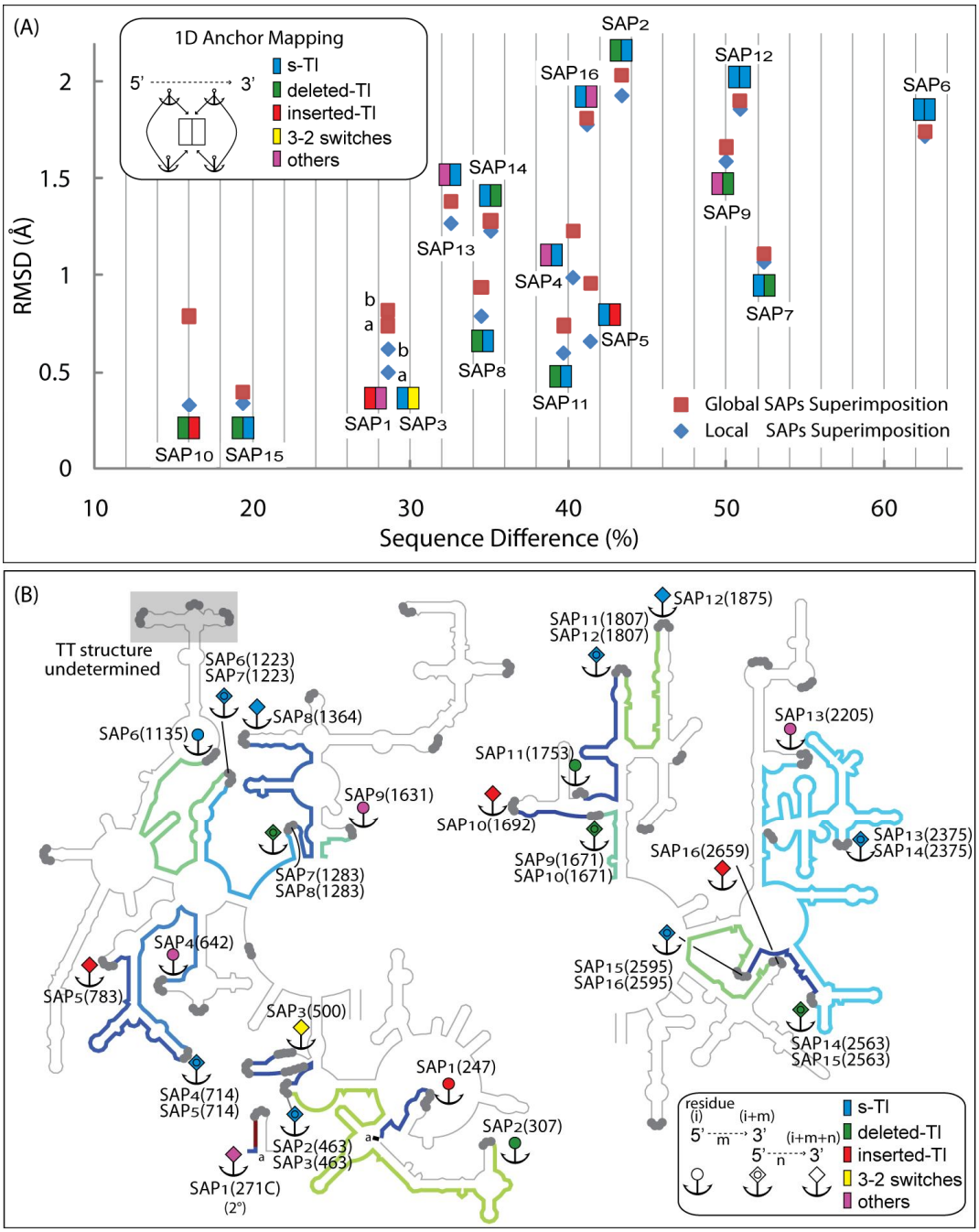

Figure 3.6 The relationship between sequence variation and positional variation.
(A) Shown is the relationship of sequence difference versus RMSD of backbone atomic positions for 16 SAPs, which are locally (blue) and globally (red) superimposed. The left rectangle indicates the anchor at the 5' terminus of the SAP. The right rectangle indicates the anchor at the 3' terminus of the SAP. (B) SAPs colored by the RMSD of backbone atomic positions are shown on the secondary structural map of *T. thermophilus* 23S rRNA. The anchor symbols indicate the termini of the SAPs. Anchors with open circles indicate the head (5') of a SAP. Anchors with open diamonds indicate the tail (3') of the SAP. The hybrid (circle in diamond) anchors indicate the tail (3') of SAP_i and the head (5') of SAP_{i+1}. The 3'-end of SAP_1 is capped by a secondary anchor, as determined by visual inspection. Tetraloop type mapping between HM^{23S} and TT^{23S} is indicated by color. Blue indicates s-Tl maps to s-Tl, green indicates deleted-Tl maps to deleted Tl, red indicates inserted-Tl maps to inserted Tl, yellow indicates 3,2 switch-Tl maps to 3,2 switch-Tl. Purple indicates the corresponding anchors are of differing types in HM and TT.

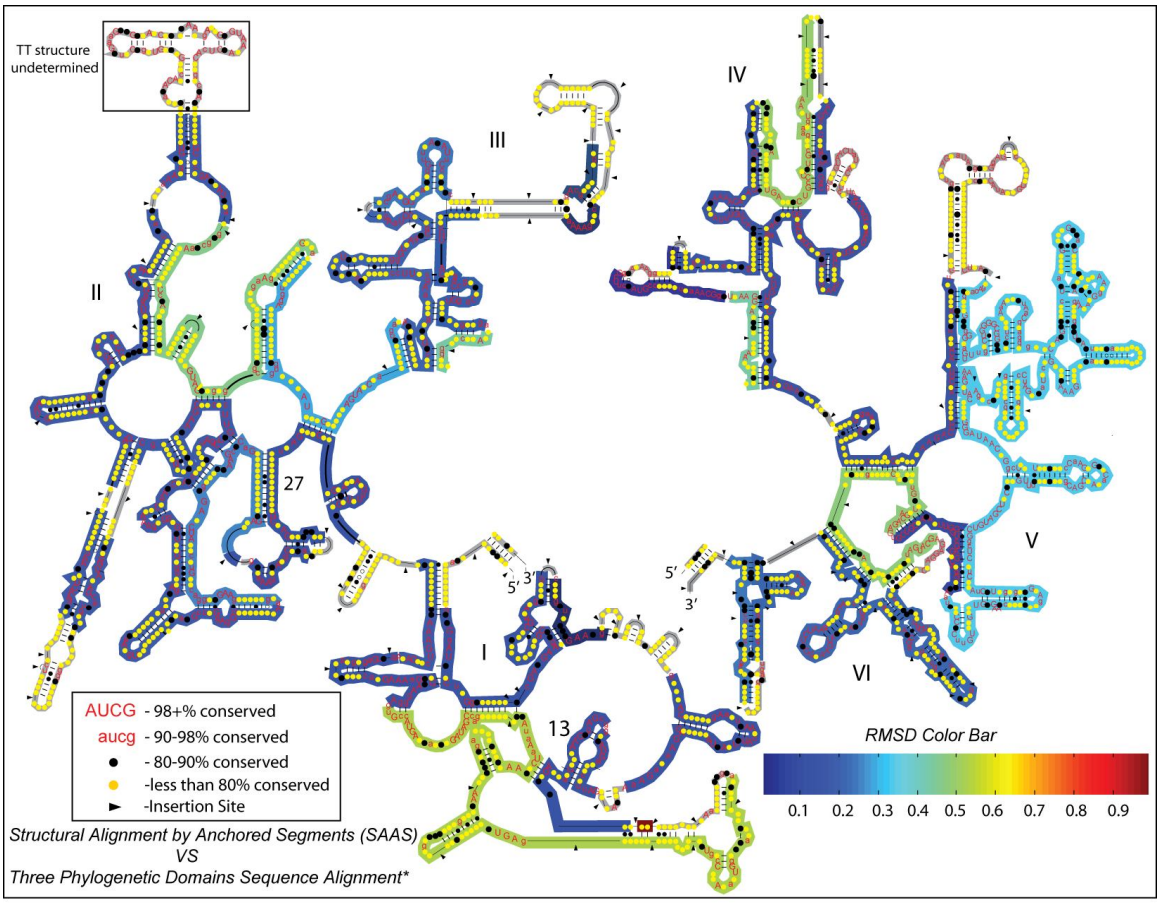

Figure 3.7 Sequence conservations and Structural conservations.
Shown is the relationship of sequence deviation for over 500 ribosomes versus RMSD of backbone atomic positions (HM versus TT) for all superimposed RNA (after global superimposition) mapped onto the secondary structure of TT^{23S}. The degree of structural conservation is indicated by color. Deeper blue indicates higher structural conservation. Deeper red indicates greater structural divergence. The degree of sequence conservation is indicated by colored dots, data obtained from Gutell and coworkers (Cannone *et al.*, 2002).

This assembly consist of tetraloops $x_{3,2}$-Tl(476^{TT23S}), $x_{3,2}i_3$-Tl($500TT^{23S}$), d_2-Tl(307^{TT23S}) and s-Tl(488^{TT23S}). Four layers of base-base hydrogen bonding and stacking plus a ternary interaction form the upper part of the Tetraloop Triplet assembly. The lower part is formed by an short A-helix where the remote standard tetraloop, 488^{TT23S}, caps the terminus (Figure 3.8) [at the relative location, it is an i_2-Tl(492) of HM^{23S}, Figure 3.3]. We hypothesize that the A-helix and the terminal tetraloop assembly along with the tetraloop triplet play important roles in rRNA folding and stability.

3.4.4 Docking and Chimeras

In future work one might combine various SAPs on the basis of RMSDs or functional criteria, and reanalyze their variability. SAP_{11} (Table 3.1) is involved in ribosomal assembly. It forms intermolecular interactions with a K-turn of the SSU. SAP_{14} (Table 3.1) is involved in tRNA binding. It shows '*Key and Lock*' interactions at the E-site.

3.4.5 3D structure prediction

Prediction of 3D structure is dependent on, and is significantly more difficult than, prediction of 2D structure. Folding of small RNAs can be simulated via all atom molecular dynamics. However errors arise from approximations in force-fields and from limitations in simulation times. Ions are important in RNA folding pathways, and must be included in the simulation. Divalent ions are especially difficult to treat. As noted by Sponer, it is unlikely that classical molecular mechanics accurately describes the anion-cation interactions of RNA (Gresh *et al.*, 2003; Rulisek and Sponer, 2003; Korostelev *et al.*, 2006). Atomic-level simulations of folding of large and complex RNAs are beyond current computational resources (Shapiro *et al.*, 2007).

Westhof and coworkers, for example, did propose highly useful 3D models of bacterial RNA P RNAs (Westhof *et al.*, 1998). Currently the best paths to 3D structure prediction uses RNA secondary structures, thermodynamics and phylogeny along with manual manipulation, constraint satisfaction, molecular mechanics, dynamics and structural homology. For example Burks and coworkers (Burks *et al.*, 2005) converted paired regions of 2D structures to helices, and other regions to RNA motifs, using SCOR (Tamura *et al.*, 2004). Models obtained thus far are at the level of folding architecture and relationships between folded elements, and do not provide precise atomic positions. In sum the goal of producing reliable 3D structures, at atomic resolution, using RNA sequence as input data is a fundamental problem has not been realized.

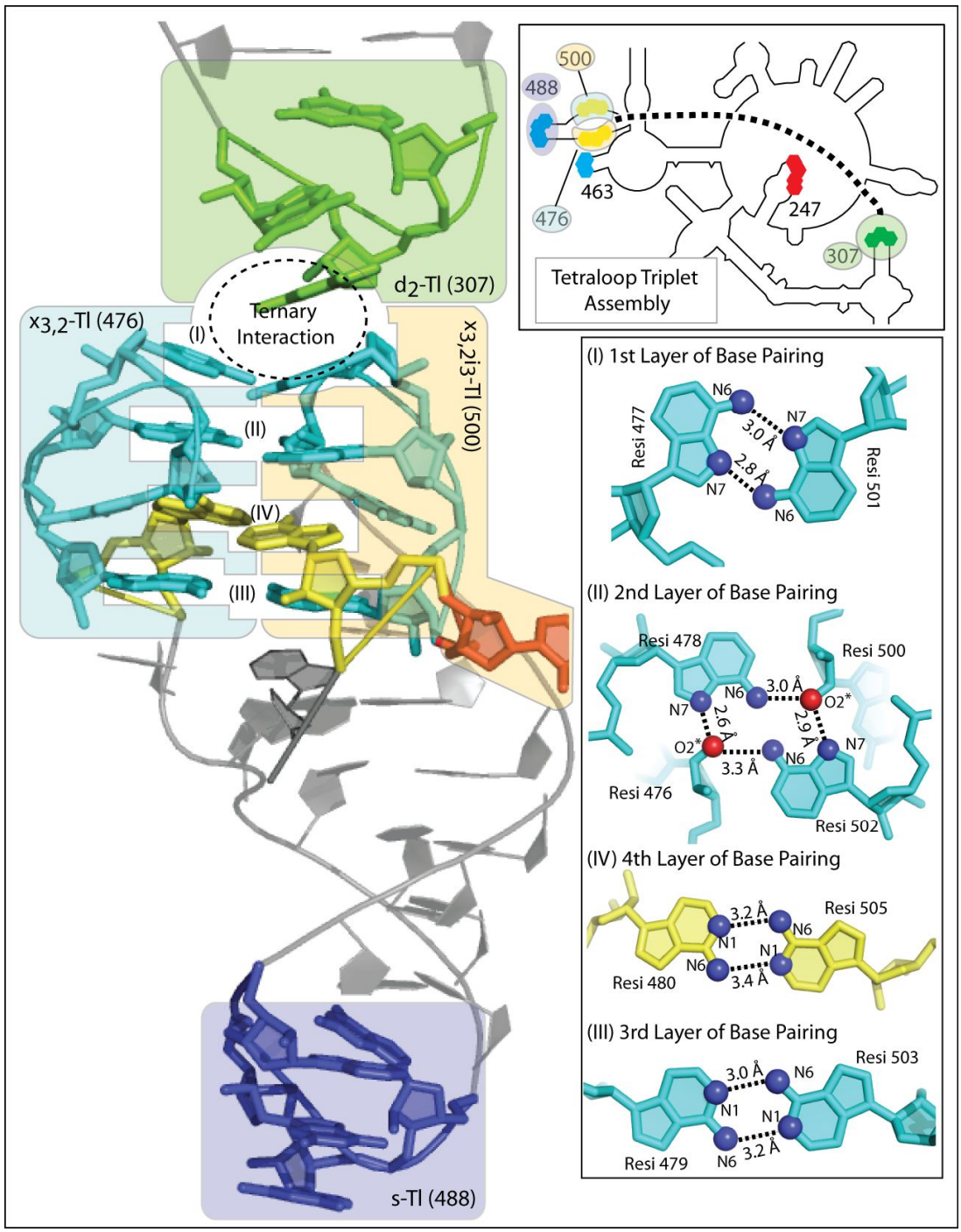

Figure 3.8 The RNA Tetraloop Triplet Assembly.
The RNA Tetraloop Triplet assembly observed in the TT^{23S}. Green indicates the deleted tetraloop; blue indicates the standard tetraloop. The yellow base pairs are both within 3,2-switch tetraloops. The Top Right Panel shows a close up of the relevant region of secondary structure and illustrates how the Tetraloop Triplet facilitates folding. The Bottom Right Panel illustrates important pairing schemes within the Tetraloop Triplet Assembly.

3.4.6 Structural versus sequence analysis

Structural alignment and superimposition in principle allows us to determine relationships between the results of structural analysis and the results of sequence analysis. Do insertions and deletions identified by covariation analysis correspond to insertions and deletions identified by structural alignment? The short answer is yes, but the long answer is more complicated.

Gutell and coworkers have analyzed rRNA sequence by performing covariation analysis of over 500 23S ribosomal sequences from the three phylogenic kingdoms (Cannone *et al.*, 2002). Using their covariation analysis results, one can construct a translation table relating two different ribosomal sequences, e.g. TT sequence versus HM sequence. One can compare insertions and deletions identified by structural alignment (as determined here) to sequence alignment through these two tables.

In many respects the tables are similar. For example both tables show an insertion of residue $C2701^{TT}$. However there are clear differences. As illustrated in Figure 3.9 $U2701^{TT}$ and $U2702^{TT}$ appear to be insertions in three-dimensional space. $C2737^{HM}$ aligns with $C2700^{TT}$ while $G2738^{HM}$ aligns best with $C2703^{TT}$. By contrast, in the sequence alignment, $G2738^{HM}$ aligns best with $U2702^{TT}$.

Structural alignment can provide information that is missing or noisy in the sequence alignment. For example residues 236-242 of the HM are aligned to residues 265-271 of the TT in both translation tables. In the structural alignment, the translation table the six residues beyond residues $A242^{HM}$ and $A271^{TT}$ are aligned. Therefore the alignment of structure is sometimes observed even when there is no obvious alignment of sequence.

Comparisons of translation tables from sequence alignment to structural alignment allow one to apply structural constraints to increase the accuracy of sequence alignment. In addition, the comparison will facilitate the prediction of 3D structure from 1D sequence. Thus the translation table obtained from structural alignment can be useful

in combination with the translation table from sequence alignment when modeling RNA sequences into RNA 3D structures. Finally, the structural alignment and superimposition allow us to identify regions of structural conservation and diversity. For examples, the ribosomal PTC region is structurally the most conserved, and ribosomal surface regions are structurally most divergent.

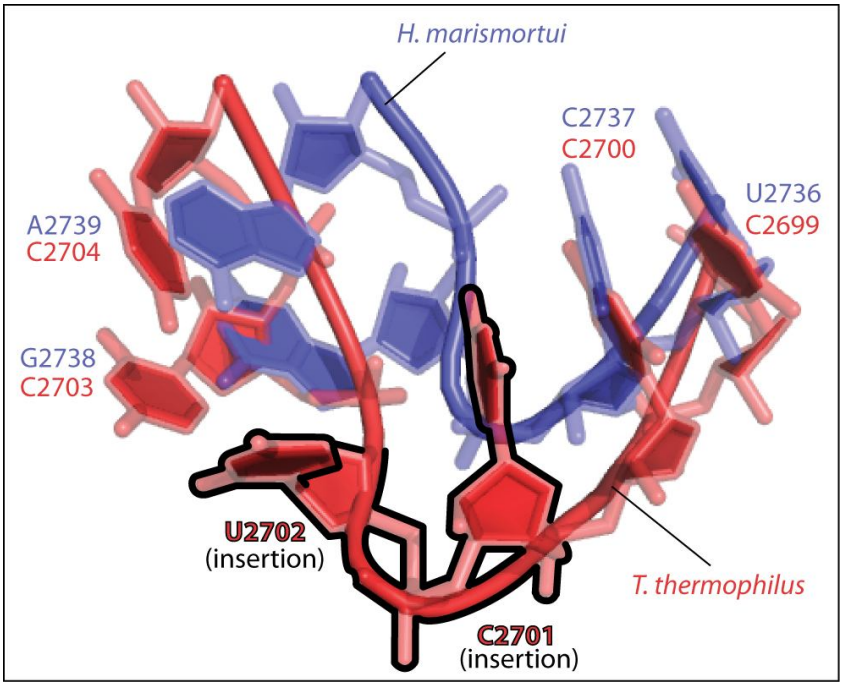

Figure 3.9 RNA sequence alignment in 3D.
The sequence of HM^{23S} (residues 2736-2739), blue, and the sequence of TT^{23S} (residues 2699-2704), red, are structurally aligned in the finished superimposition of 23S rRNAs of HM and TT. Residues are structurally aligned in 3D. Residues $C2701^{TT}$ and $U2702^{TT}$ (outline in black) are identified as insertions in the structural alignment within the finished superimposition.

3.5 Acknowledgements

The authors thank Steve Harvey and Andrew Huang from Georgia Tech for helpful discussions.

CHAPTER 4

COMPLEXES OF NUCLEIC ACIDS WITH GROUP I AND II CATIONS

4.1 Introduction

Recent structures of large RNAs, such as the P4-P6 domain of the tetrahymena ribozyme (Cate *et al.*, 1996a; Cate *et al.*, 1996b; Cate *et al.*, 1997; Basu *et al.*, 1998; Juneau *et al.*, 2001), and larger RNAs such as rRNAs (Cate *et al.*, 1999; Ban *et al.*, 2000; Wimberly *et al.*, 2000; Harms *et al.*, 2001; Yusupov *et al.*, 2001; Klein *et al.*, 2004; Berk *et al.*, 2006; Selmer *et al.*, 2006; Voss *et al.*, 2006), combined with a general increase over time in the sophistication of diffraction experiments, show cations in diverse and sometimes unexpected environments. The interactions of nucleic acids with cations follow basic principles of coordination chemistry. The effects of cations on RNA stability and conformation demonstrate the endurance of these relatively simple principles.

One focus of this chapter is the coordination of Na^+, K^+, Ca^{2+}, and Mg^{2+} by phosphates and nucleic acids. We describe coordination chemistry, electrostatic forces/energetics, conformational effects, and ion-selective binding. We explain the origins of the specific requirement for Mg^{2+} in RNA folding and the tight coupling between Mg^{2+} binding and RNA conformation. We describe crystallographic methods for determining cation positions. We propose a model of RNA folding that is consistent with Mg^{2+} coordination properties of RNA. Previous reviews are available on roles of metals in biology (Williams, 1971; Williams, 1991; Black *et al.*, 1994), in polyelectrolyte theory (Manning, 1978; Record *et al.*, 1978; Sharp and Honig, 1990; Anderson and Record, 1995), in DNA structure (Swaminathan and Sundaralingam, 1979; McFail-Isom *et al.*, 1999; Williams and Maher, 2000; Hud and Polak, 2001; Sponer *et al.*, 2001; Subirana

and Soler-Lopez, 2003), in RNA folding (Draper and Misra, 1998; Draper, 2004; Onoa and Tinoco, 2004; Draper *et al.*, 2005), and in RNA catalysis (Agris, 1996; Scott and Klug, 1996; Pyle, 2002; Been, 2006).

4.1.1 Modern Treasure Troves of Structural Information: Large RNAs

Very large RNA assemblies are now available at high resolution. The largest and most accurate structures are used here in conjunction with smaller structures, down to the level of mononucleotides, to illustrate patterns of interaction of nucleic acids with cations. 23S-rRNA^{HM} refers to the 23s rRNA from the archaeon H. marismortui (Ban *et al.*, 2000; Klein *et al.*, 2001) [2.4 Å resolution, PDB entry 1JJ2], a halophile from the Dead Sea. 23S-rRNA^{TT} refers to the 23s rRNA from eubacterium T. thermophilus (Selmer *et al.*, 2006) [2.8 Å resolution, PDB entry 2J01], isolated from a thermal vent. The fractional sequence identity of the 23S rRNAs from HM and TT is around 60%. RNA$^{P4\text{-}P6}$ refers the 160 nucleotide domain of the self-splicing Tetrahymena thermophila intron [2.3 Å resolution, PDB entry 1HR2, this ΔC209 mutant (Juneau *et al.*, 2001) gives the best available resolution].

4.2 Folding

4.2.1 Cations

During protein folding, water molecules in contact with hydrophobic surfaces are released to bulk solvent. During nucleic acid folding, cations are sequestered from bulk solvent, and held in close proximity to the polymer. Protein sidechains are multifarious, with a variety of shapes and chemical properties. The nucleic acid backbone is intricate, with many accessible rotameric states (Richardson *et al.*, 2008), and carries charge.

Functional nucleic acids generally fold into compact and stable states of given conformation (Latham and Cech, 1989; Celander and Cech, 1991). DNA can form

quadruplexes (Williamson *et al.*, 1989; Hud *et al.*, 1999b; Neidle and Parkinson, 2003; Burge *et al.*, 2006; Gill *et al.*, 2006), triplexes (Moser and Dervan, 1987; Francois *et al.*, 1988; Shafer, 1998), i-motifs (Leroy *et al.*, 1993; Gilbert and Feigon, 1999) etc. Structured RNAs range in size from aptamers and tRNAs to ribosomes. However some functional nucleic acids, such as riboswitches, are conformationally polymorphic (Winkler and Breaker, 2005; Coppins *et al.*, 2007). For our purposes folded nucleic acid structures fall into three general classes (Cate *et al.*, 1996a; Cate *et al.*, 1997): (i) *helical structures* such as A-form, B-form, and triplexes, (ii) *quasi-globular structures* such as tRNA, with base-base tertiary interactions but no buried phosphates (Kim *et al.*, 1974; Jack *et al.*, 1976), and quadruplexes (Williamson *et al.*, 1989; Neidle and Parkinson, 2003), and (iii) *true globular structures* such as the tetrahymena ribozyme (Latham and Cech, 1989) and its P4-P6 domain (Cate *et al.*, 1996a; Cate *et al.*, 1996b; Cate *et al.*, 1997; Basu *et al.*, 1998; Juneau *et al.*, 2001), with base-base tertiary interactions plus buried OP atoms (OP indicates a non-bridging phosphate oxygen). True globular structures have distinct 'insides' and 'outsides'. Folding of helices, quasi-globular structures and true globular structures increases proximities of phosphate groups, and the electrostatic repulsion among them. Therefore folding is intrinsically linked to association with cations. Phosphate-phosphate repulsion must be offset by attraction between phosphates and cations. Cations most strongly associate with regions of DNA and RNA in which phosphate groups assume greatest 'density'.

As will become clear in the following sections, Mg^{2+} stabilizes distinctive conformational and energetic states of nucleic acids. Mg^{2+} shares a special geometric and electrostatic complementarily with phosphate, with a specific coordination and thermodynamic fingerprint. These states are simply not accessible in the absence of Mg^{2+}, even when other cations are at high concentration. The thermodynamic and conformational consequences of first shell OP interactions with Mg^{2+} are different than for neutral ligands or for other cations, with lesser charge or greater size.

4.2.2 The RNA Folding Hierarchy

RNA folding is hierarchical (Brion and Westhof, 1997; Tinoco and Bustamante, 1999). Folding progresses through a series of intermediates are commonly characterized by extents and types of base-base hydrogen bonding and stacking interactions. The unfolded state, the random coil, is a conformationally polymorphic and fluctuating ensemble with few local or long-range base-base interactions. Early intermediates contain double-stranded stems and hairpin loops, interspersed by single-stranded regions. These stems and loops are known collectively as secondary structural units. Late intermediates and the final folded state are stabilized by base-base tertiary interactions, between residues that are remote in the secondary structure. To a first approximation, secondary structure can be conceptually and experimentally separated from tertiary structure. Secondary structure forms before tertiary structure and is favorable in a broad range of ionic conditions. Tertiary structure is favored by divalent cations (Cole *et al.*, 1972; Misra and Draper, 2000). Although compact structures with base-base tertiary interactions can be achieved a very high concentrations of other cations, for true globular structures, the fully folded state is absolutely dependent on Mg^{2+}. It can be useful to realize a distinction between the *tertiary structure of an RNA*, which is a description of short and long range base-base interactions, and *a folded RNA*, which is a description of three dimensional positions of all atoms, including of course the phosphate groups.

A hierarchical model that focuses exclusively on base-base interactions is useful but somewhat limited approximation. In true globular structures, ground states are stabilized by specifically-associated Mg^{2+} ions, each with at least three first-shell OP ligands. The importance of $Mg^{2+}(OP)_3$ coordination complexes is discussed in later sections. Multidentate interactions of OP atoms with Mg^{2+} are generally local along the RNA backbone. A small and important subset of OP ligands of a common Mg^{2+} are remote in the secondary and primary structures, thus forming 'electrostatic tertiary interactions' (Mg^{2+} mediated linkages between remote OP groups). Extensive base-base

tertiary interactions during folding do not necessary imply formation of electrostatic tertiary interactions. At least some globular RNAs fold into compact (but non-native) structures, with extensive base-base tertiary interactions - in the absence of Mg^{2+} (Takamoto *et al.*, 2004). We don't know if the converse is true (i.e., are electrostatic tertiary interactions fully dependent on correct base-base tertiary interactions?). At any rate, to fully understand and describe the structure of a globular RNA, one can extend a conventional tertiary description of base-base interactions to include electrostatic tertiary interactions.

4.2.3 Alternative RNA Folding Hierarchies

To illuminate the underlying dependence of folding on cations, one can re-state the hierarchy of RNA folding using 'phosphate density'. In early folding steps, a subset of phosphate-phosphate distances decreases from >10 Å (P to P) in random coil to around 5.8 to 6.2 Å (in A-form helical regions and loops). In subsequent steps, a subset of P to P distances decreases further, to 5.0 to 4.6 Å. Associated with this group of short P to P distances are tightly packed anionic OP atoms, which are in van der Waals contact with each other ($d_{OP\text{-}OP}$ = 2.8-3.2 Å). This tight packing of anionic oxygen atoms is dependent on multidentate chelation of Mg^{2+} by OP atoms. Neither monovalent cations nor polyamines can substitute for Mg^{2+} in stabilizing structures with such short OP-OP contacts. During folding, some phosphates and associated Mg^{2+} ions become buried in the globular interior.

4.3 Coordination Chemistry

The binding of ligands to Group I and to Group II cations is dictated by the chemical properties of the cations and of the ligands, and to a significant extent, by interactions between ligands (Williams, 1971). Chelators, with covalently linked ligands, create cavities for ions, and bind with greater affinity and selectivity than monomeric ligands. The length of the chelator linker is a critical component of stability. As the linker length increases the entropic cost to assembling the ligands for joint coordination increases. The optimum linker gives a ring size of six atoms [cation $-(L)_5$-cation] as with EDTA and ADP/ATP (below). Hud and Polak previously noted the chelation properties of DNA, calling it an ionophore (Hud and Polak, 2001). Here we illustrate how the phosphate groups of nucleic acids commonly act as chelators of cations.

4.3.1 Group I

Group I cations prefer hard neutral ligands, or one singly charged ligand plus additional neutral ligands. In their associations with nucleic acids, Group I cations are most commonly associated with non-anionic oxygens (i.e., oxygen atoms other than OP) as inner shell ligands (Draper and Misra, 1998; Shui *et al.*, 1998).

The monovalent cations [sodium (Na^+), potassium (K^+), rubidium (Rb^+), cesium (Cs^+), thallium (Tl^+) and ammonium (NH_4^+), excluding lithium (Li^+)] are characterized by relatively large ionic radii, low charge density and modest enthalpies of hydration (Table 4.1). The coordination chemistry of Li^+, with its small atomic radius and high charge density, is distinct from that of other alkali metals. Tl^+ and NH_4^+ are listed here along with the Group I metals because they are well-developed K^+ substitutes with useful spectroscopic and crystallographic signals. NH_4^+ positions are indicated by NOEs in solution (Hud *et al.*, 1998; Hud *et al.*, 1999b; Hud *et al.*, 1999a). Tl^+ positions are indicated in solution by NMR, and in crystals by a distinctive x-ray scattering signal

(anomalous scattering). Tl^+, K^+ and NH_4^+ have similar ionic radii and enthalpies of hydration (Table 4.1).

These monovalent ions, except Li^+, display irregular and variable coordination geometry (Brown, 1988; Brown, 1992). The variability in coordination geometry is associated with non-covalency of interaction, weak ligand-ligand interactions and loose ligand-ligand packing. For a given monovalent ion, the number of first shell ligands can vary from four to over ten. These properties are quantitated by 'average observed coordination numbers' (AOCN, Table 4.1) over a large number of structures within the Cambridge Structural Database as reported by Brown (Brown, 1988).

Table 4.1 Physical Properties of Cations

	Li^+	Na^+	K^+	Rb^+	Cs^+	Tl^+	NH_4^+	Mg^{2+}	Ca^{2+}	Mn^{2+}
Ionic Radius[a] (Å)	0.60	0.95	1.33	1.48	1.69	1.49	(b)	0.65	0.99	0.80
$\Delta H_{hydration}$[c] (kcal mol^{-1})	-127	-99	-77	-72	-66	-78	-78	-458	-358	
AOCN[d]	5.3	6.7	9.0	9.8	10.4	8.3	(e)	5.98	7.3	5.98

a) From I.D. Brown (Brown, 1988).
b) The radius of a non-spherical species such as NH_4^+ is not well-defined. Rashin and Honig (Rashin and Honig, 1985) estimated from geometrical considerations that the effective radius of NH_4^+ is very nearly the same as the radius of K^+.
c) From Honig and Rashin (Rashin and Honig, 1985).
d) Average Observed Coordination Numbers, from I.D. Brown (Brown, 1988).
e) Not reported.

Na^+

The ideal Na^+ to oxygen distance is 2.4 Å (Figure 4.1A). The distance between first-shell ligands of Na^+ is variable, depending on coordination number and coordination geometry. An octahedral arrangement of first shell oxgen ligands is loosely packed. The O to O distance is 3.4 Å, which is significantly greater than twice the van der Waals radius of oxygen (oxygen radius = 1.4 Å). Therefore inner shell ligands of Na^+ are not crowded and the geometry of the Na^+ inner sphere is not determined by ligand-ligand

interactions. A Na^+ ion with ideal octahedral geometry in association with the O6 position of a guanine of DNA with five water molecule inner shell ligands (Komeda *et al.*, 2006) is shown in Figure 4.1A. As can be seen the Na^+ to O distances average around 2.4 Å, while the distance between cis oxygen atoms (adjacent oxygen ligands) averages around 3.4 Å.

K^+

The ideal K^+ to oxygen distance is around 2.7 Å. For an octahedral arrangement of first shell oxygen ligands in the K^+ first shell, the average O to O distance is over 4.0 Å. Thus the inner shell ligands of K^+ are even sloppier than those of Na^+. Specific K^+ binders, that exclude Na^+, are generally composed of stacked, planar arrangements of keto oxygens (Williamson *et al.*, 1989; Hud *et al.*, 1996; Doyle *et al.*, 1998). In these K^+ selective structures, the positions of the keto oxygens (O^{keto}) are fixed such that the enthalpy of dehydration is compensated by O^{keto}-K^+ interactions but not by O^{keto}-Na^+ interactions, which are too long.

4.3.2 Group II

Ca^{2+}

Ca^{2+} often shows irregular coordination geometry and coordination numbers greater than six. The ionic radius of Ca^{2+} is large, the charge density is low, the inner shell ligands are loosely packed, and the magnitude of the hydration enthalpy is small (compared to Mg^{2+}). But like Mg^{2+}, Ca^{2+} prefers and mix of anionic and neutral ligands.

Mg^{2+}

Mg^{2+}, from life's beginning, has been closely associated with some of the central players in biological systems - phosphates and phosphate esters (Westheimer, 1987). Mg^{2+} shares a special geometric, electrostatic, thermodynamic relationship with

phosphates and phosphate esters. In comparison with group I ions, Ca^{2+}, or polyamines, Mg^{2+} has a much greater affinity for OP atoms, and binds to OP with well-defined geometry. Unlike other cations, Mg^{2+} brings OP atoms into direct contact with each other.

The ionic radius of Mg^{2+} is small (0.65 Å), the charge density is high, the six ligands of an octahedral inner first shell are tightly packed (Figure 4.1B), and the magnitude of the hydration enthalpy is large (Table 4.1). The heat of hydration of Mg^{2+} is much greater than for other biological cations (Table 4.1). Mg^{2+} prefers two to four oxyanions (along with a complement of water molecules) over uncharged oxygens and nitrogens as inner shell ligands.

The first coordination sphere of Mg^{2+}, whether water, nucleic acid or protein, assumes octahedral geometry (Brown, 1988; Brown, 1992; Bock *et al.*, 1999; Sines *et al.*, 2000), as shown in Figure 4.1B. The AOCN of Mg^{2+} is 5.98 (Brown, 1988). Because Mg^{2+} is small and highly charged, ligand-ligand crowding is one of the hallmarks of Mg^{2+} complexes, leading to highly restrained ligand-Mg^{2+}-ligand geometry, and strong ligand-ligand repulsive forces. Although probably not germane to nucleic acid structure, four-coordinate Mg^{2+} is observed at high temperature in the gas phase (Rodriguez-Cruz *et al.*, 1998).

Hexa-aquo Mg^{2+} complexes

In hexa-aquo complexes [$Mg^{2+}(H_2O)_6$ or $Mg^{2+}{}_{aq}$], Oxygen -- Mg^{2+} distances are 2.07 Å. The cis O--Mg--O angle is 90° and the cis O to O distance is 2.93 Å (Bock *et al.*, 1994; Markham *et al.*, 2002; Bock *et al.*, 2006). The trans O--Mg--O angle is 180°. Adjacent oxygen atoms in $Mg^{2+}{}_{aq}$ are in van der Waals contact. First shell water molecules are strictly oriented such that their dipole moments are directed in toward the metal, with Mg^{2+}---O-H angles of 120-128°. This orientation prevents hydrogen-bonding between water molecules in the Mg^{2+} first coordination shell.

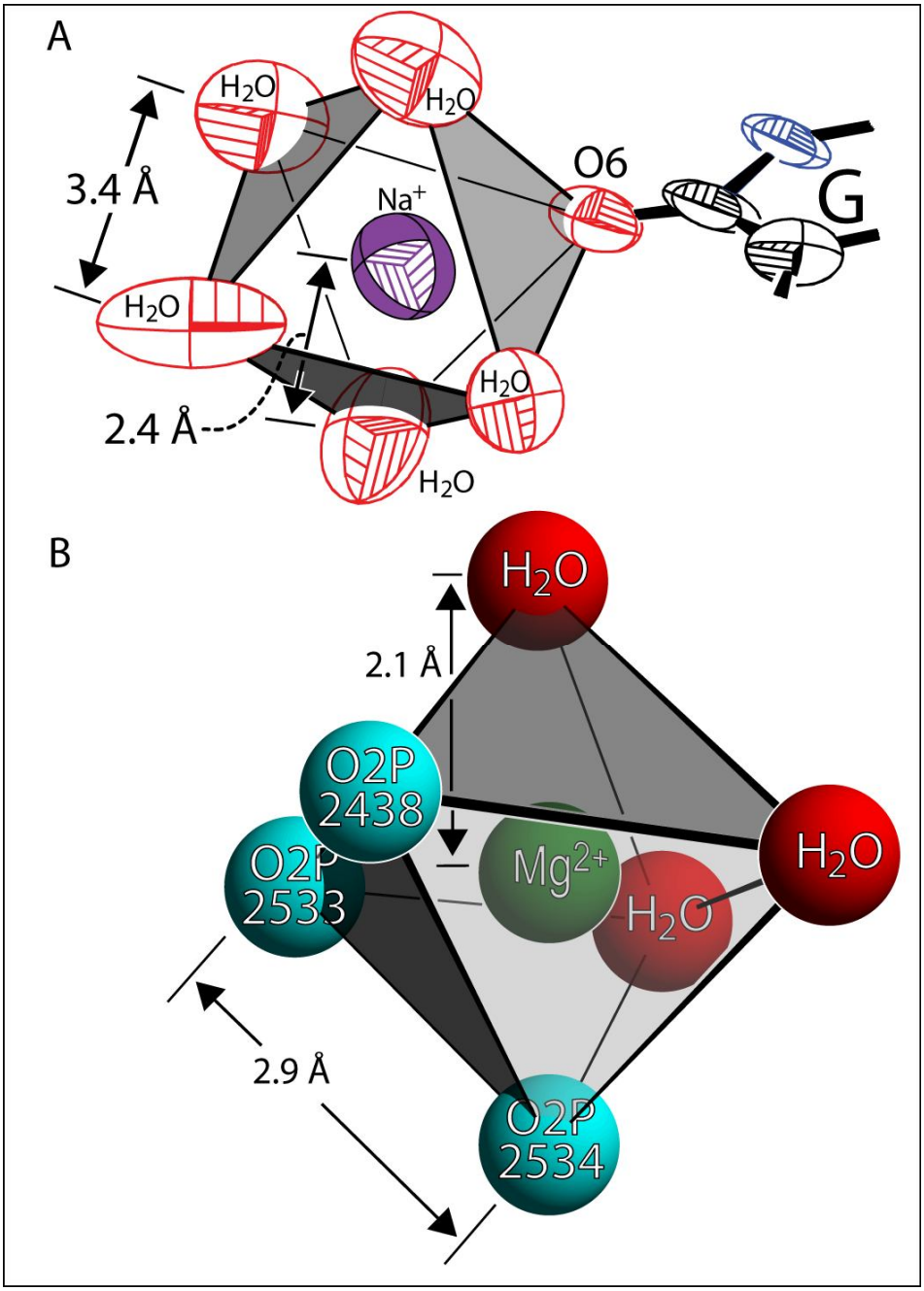

Figure 4.1 Sodium (Na^+), Magnesium (Mg^{2+}) coordination geometry.
(A) A Na^+ ion bound to the O6 of a guanine of DNA (ion 487 from PDB entry 2DYW). The Na^+ ion (purple ellipsoid) is octahedral, with one bond to the floor of the major groove and five water molecules (hydrogens not shown). All first-shell oxygen ligands are represented by red ellipsoids. Na^+-oxygen bonds are around 2.4 Å in length. The inner ligands are beyond van der Waals contact with each other. (B) A trichelated Mg^{2+} ion (ion 8001 from 23S-rRNA^{HM}). This Mg^{2+} ion (green sphere) is octahedral, with three bonds to phosphate groups of RNA (cyan) and three bonds to water oxygens (red). Mg^{2+}-oxygen bonds are around 2.1 Å in length. Mg^{2+} coordination imposes oxygen-oxygen distances of around 2.9 Å in the inner coordination sphere. For clarity the radii of the spheres are reduced from the van der Waals radii of the atoms, and have no physical significance.

ADP-Mg^{2+} complexes

The rules of engagement for complexes of Mg^{2+} with ADP and ATP are highly predictive of Mg^{2+} interactions with other multidentate OP ligands such as RNA. Some ADP-Mg^{2+} and ATP-Mg^{2+} complexes are shown in Figure 4.2. One can observe that ADP and ATP are mono- and multidentate chelators of Mg^{2+}, contributing OP ligands to the inner coordination sphere. In this section we focus on structural aspects of Mg^{2+} complexes with nucleotides. Thermodynamic aspects are discussed in a subsequent section of this chapter.

Mg^{2+} interacts with non-bridging OP atoms, but not with bridging oxygens, base or sugar atoms of ADP/ATP. The OP atoms of ADP bind to Mg^{2+} by either monodentate interactions (40% of structures surveyed, Figures 4.2A and B, Tables 4.2 and 4.3) or by bidentate chelation (60%, Figures 4.2C, 4.2D & 4.2E). Monodentate interactions occur exclusively by $O^{\beta}P$ while bidentate chelation involves $O^{\alpha}P$ and $O^{\beta}P$.

Chelation ring size is an important factor in modulating stability. The bidentate chelation complexes of Mg^{2+} with ADP/ATP are composed of 6-membered rings consisting of atoms Mg^{2+}-O^{α}P-P-O-P-O^{β}P-Mg^{2+}. Bidentate chelation by two OP atoms bound to a common phosphorous atom would require a chelation ring size of four, and is not observed in ADT/ATP complexes. When Mg^{2+} is monochelated by ADP, at least one protein ligand is also found in the Mg^{2+} first shell. The protein ligand is invariably oxygen, but may be charged or neutral. Similarly, most bidentate ADP-Mg^{2+} complexes contain protein first shell ligands. OP ligands are adjacent to each other within the octahedron of first shell ligands of Mg^{2+}. All bidentate Mg^{2+}-ADP complexes assume cis (OP-Mg^{2+}-OP angle of 90°), cis-cis or cis-cis-cis orientation of the ligands surrounding the Mg^{2+} depending on the number of first shell protein ligands. Trans bidentate Mg^{2+}-ADP complexes (OP-Mg^{2+}-OP angle of 180°) are not observed and appear to be stereochemically prohibited.

ATP-Mg^{2+} complexes

Mg^{2+} binds to ATP by bidentate (66%, Figures 4.2F, 4.2G, 4.2H & 4.2I) or tridentate (33%, Figure 4.2J) interactions. Monodentate ATP-Mg^{2+} complexes are not observed. The bidentate complexes are all cis, while the tridentate complexes are all cis-cis. The tridentate complexes, by definition are bicyclic, with two six-membered rings. Each ring consists of six atoms (Mg^{2+}-O^{α}P-P-O-P-O^{β}P-Mg^{2+}) Bidentate complexes most commonly involve O^{β}P and O^{γ}P (88%). One O^{α}P-O^{β}P bidentate complex is observed.

An O^{α}P-Mg^{2+}-O^{γ}P bidentate complex would form an 8-membered ring. The importance of ring size in chelation complexes is underscored by the absence of a O^{α}P-Mg^{2+}-O^{γ}P bidentate complexes in the structural database. The observed bidentate complexes contain from 0 to 3 protein ligands (Table 4.2), which are all oxygens, with no obvious preference in the number or charge of protein ligands. First shell protein ligands are not observed tridentate ATP complexes.

Table 4.2 Mg^{2+} interactions with ADP and ATP[a]

PDB Entry	Number of 1st Shell OP atoms	Total 1st Shell Ligands[b]	Protein Ligands	Orientation	Resolution (Å)	Author
			ADP			
1G6H	1	2	Ser	cis	1.6	Hunt
1L4Y	1	2	Ser	cis	2.0	Ji
1P5Z	1	3	Ser, Glu	cis-cis	1.6	Lavie
1SVL	1	2	Ser	cis	1.95	Chen
1BYQ	2	3	Asn	cis-cis	1.5	Pavletich
1KJQ	2	4	Glu, Glu	cis-cis-cis	1.05	Holden
1OHA	2	2		cis	1.9	Rubio
1PHP	2	3	Asp	cis-cis	1.65	Watson
1RL9	2	2		cis	1.45	Chapman
1Z2N	2	4	Asp, Asp	cis-cis-cis	1.2	Hurley
			ATP			
1G5T	2	4	Glu, Thr	cis-cis-trans	1.8	Bauer
1A82	2	5	Glu, Asp, Thr		1.8	Schneider
1D4X	2	2		cis	1.75	Almo
1F2U	2	4	Ser, Gln	cis-cis-trans	1.6	Craig
1KAX	2	2		cis	1.7	Mckay
1SVM	2	3	Ser	cis-trans	1.94	Chen
2IXE	2	3	Ser	cis-trans	2.0	Gaudet
2IYW	2	3	Ser	cis-trans	1.85	Bartunk
1J09	3	3		cis-cis	1.8	Yokoyama
1OBD	3	3[c]		cis-cis	1.4	Wilson
2A84	3	3		cis-cis	1.55	Eisenberg
2NT8	3	3[c]		cis-cis	1.68	Rayment

a) Structures were obtained from the PDB. Structures were rejected if the Mg^{2+} ion is coordinated by anything other than the nucleotide, water or protein. OP-Mg^{2+} bond lengths greater than 2.25Å, or other bad geometric were excluded. Ten Mg^{2+}-ADP and 12 Mg^{2+}-ATP structures were obtained.
b) The total number of first shell ligands other than water molecules.
c) These Mg^{2+} ions are 4 coordinate presumably because two water molecules were omitted during refinement.

Table 4.3 Mg^{2+} Chelation by Nucleotides[a] and by RNA[b]

Number of 1st shell OP Ligands	0 (hexa-aquo)	1 (mono-dentate)	2 (bidentate)	3 (tridentate)
ADP	NA	4	6	0
ATP	NA	0	8	4
RNA	36	40	31	10

a) The number of nucleotides with OP atoms in the Mg^{2+} first-shell. Protein ligands are excluded.
b) In 23S-$rRNA^{HM}$.

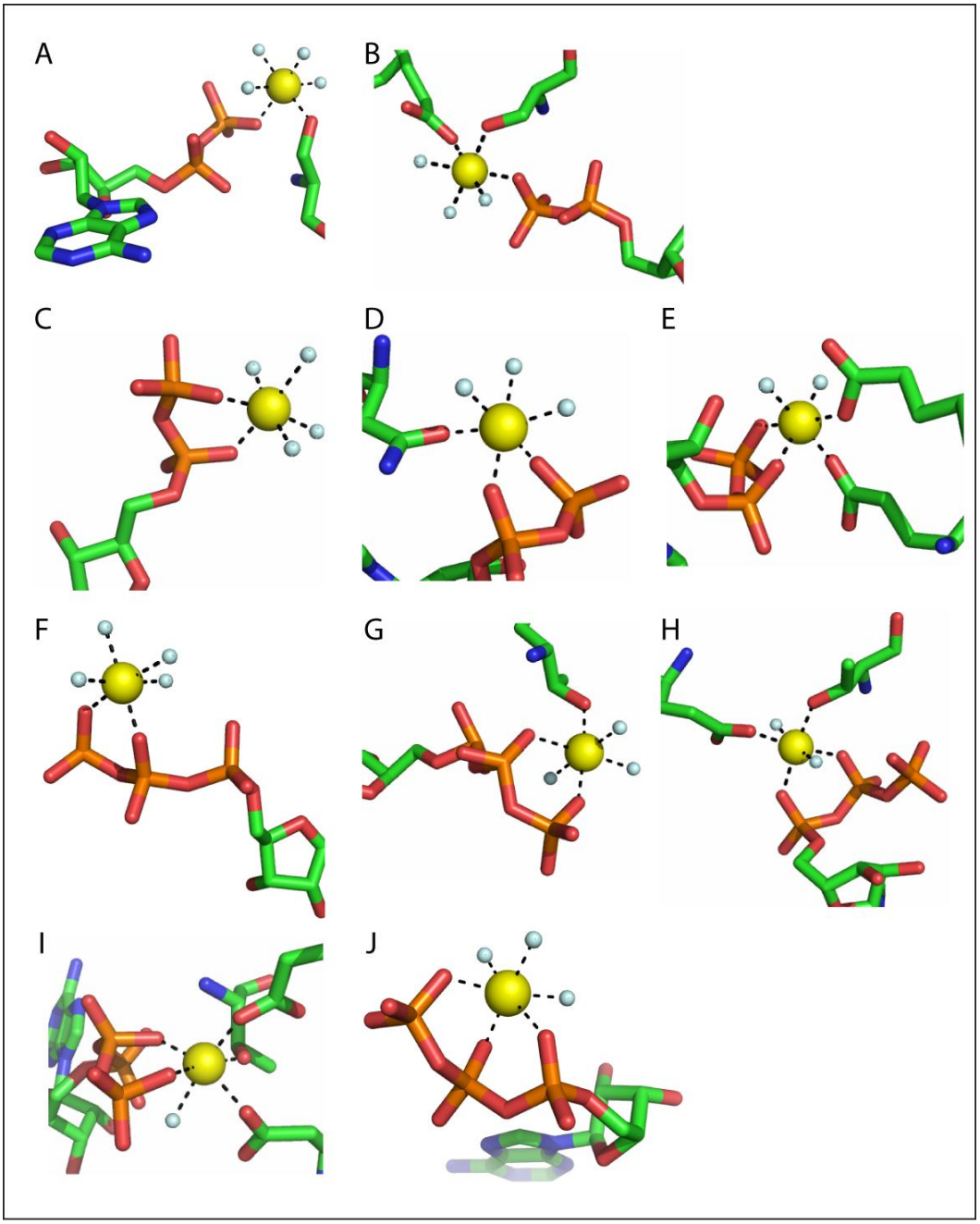

Figure 4.2 ADP and ATP as Mg^{2+} chelators.
Mg^{2+} ions are represented by yellow spheres and water molecules by small white spheres. Nucleotides and protein ligands are represented by sticks. (A) Monodentate chelation by ADP ($O^{\beta}P$) plus one protein ligand. (B) Monodentate chelation by ADP with two protein ligands. (C) Bidentate chelation by ADP ($O^{\beta}P$ and $O^{\alpha}P$) with no protein ligands. The OP ligands are cis. (D) Bidentate chelation by ADP plus one protein ligand. (E) Bidentate chelation by ADP with two protein ligands. (F) Bidentate chelation by ATP ($O^{\beta}P$ and $O^{\gamma}P$) with no protein ligands. (G) Bidentate chelation by ATP plus one protein ligand. (H) Bidentate chelation by ATP plus three protein ligands. (I) Bidentate chelation by ATP ($O^{\alpha}P$ and $O^{\beta}P$) plus two protein ligands. (J) Tridentate chelation by ATP. The OP ligands are cis-cis.

RNA-Mg^{2+}: OP preferred

With RNA, Mg^{2+} associates preferentially with OP atoms over base and ribose atoms, just as it prefers OP atoms of ATP/ADP (above). Interactions with uncharged functional groups of sugars and bases of RNA are infrequent (Table 4.4).

The arrangement of three OP atoms and three water oxygens around one particular Mg^{2+} in a large globular RNA is shown in Figure 4.1B. The 23S-$rRNA^{HM}$ is a tridentate chelator of this Mg^{2+} ion, contributing three OP ligands to the inner coordination shell.

Large rRNAs such as 23S-$rRNA^{HM}$ are associated with many Mg^{2+} ions, which can be visualized by x-ray diffraction. Of the 117 Mg^{2+} atoms associated in some fashion with 23S-$rRNA^{HM}$, 98 are "bonded" to the 23S rRNA, (Mg^{2+}-OP distance < 2.6 Å), with first shell RNA ligands. The most common Mg^{2+} ligands are water oxygens (407 ligands). The vast majority of RNA ligands are OP atoms (129 first shell OP ligands). Mg^{2+} does not express a preference for O1P (63 ligands) over O2P (66 ligands). RNA bases are infrequent ligands, with twelve O6 atoms, ten N7 atoms, five O4 atoms, and three O2 atoms within Mg^{2+} first shells. As anticipated from DNA structures (Shui *et al.*, 1998), RNA amino groups, (N6, N2 and N4) do not enter the first coordination shell of Mg^{2+}. The 5S rRNA associated with 23S-$rRNA^{HM}$ interacts with one Mg^{2+}. There, two OP atoms chelate a Mg^{2+}.

Chelation Ring Size

In contrast to preferential bi- and tridentate chelation of Mg^{2+} by ADP and ATP, RNA generally prefers monodentate chelation of Mg^{2+} (Table 4.3). The primary reason for this preference is that RNA forms 10-membered chelation cycles with Mg^{2+} (Figures 4.3 & 4.4), which are less favorable than the 6-membered chelation cycles of ADP/ATP (Figure 4.2). A second reason is that the relative positions of OP atoms in common RNA secondary structural elements such as A-form helices and tetraloops are not favorable for

multidentate chelation of Mg^{2+}. In such canonical RNA conformations the OP atoms are too far apart and are not optimally oriented for multidentate Mg^{2+} chelation. In regions of irregular RNA conformation OP assume the correct geometric disposition for Mg^{2+} chelation. For example the RNA conformation in the $Mg^{2+}(OP)_2$ complex in Figure 4.3 deviates profoundly from that in A-form helical RNA.

Table 4.4 Paleo-Magnesium ions[(a)] in 23S-rRNA^{HM} and 23s-rRNA^{TT}

Mg^{2+} ID[(b)]	Ligand Type (OP,B,S,P)[(c)]	Mg^{2+} ID Number[(d)]	rRNA residue numbers[(e)]
Dual Mg^{2+} Bicycles			
D1	$(OP)_3$	8001 (122)	2483 (2448), 2533 (2498), 2534 (2499)
D1'	$(OP)_2B$	8002 (123)	2483 (2448), 2534 (2499), 627^{Base} (570^{Base})
D2	$(OP)_3$	8003 (66)	876 (783), 877 (784), 2624 (2589)
D2'	$(OP)_2$	8013 (70)	877 (784), 2623 (2588)
D3	$(OP)_3$	8016 (86)	1504 (1395), 1678 (1603), 1679 (1604)
D3'	$(OP)_2$	8029 (85)	1503 (1394), 1679 (1604)
D4	$(OP)_3$	8005 (124)	1836 (1780), 1838 (1782), 1839 (1783)
D4'	$(OP)_3$	8007 (126)	832 (740), 1839 (1783), 1840 (1784)
Lone Tridenate Mg^{2+}			
4	$(OP)_3$	8026 (130)	2608 (2573), 2609 (2574), 2610 (2575)
5	$(OP)_3$	8008 (127[(f)])	919 (826), 2464 (2427), 2465 (2428), na (2429)[(f)]
6	$(OP)_3B$	8033 (133)	1747 (1669), 1748 (1670), 2585 (2550), 1749^{Base} (1671^{Base})
9	$(OP)_3$	8006 (125)	821 (730), 822 (731), 854 (761)
10	$(OP)_3$	8081 (143)	1420 (1314), 1421 (1315), 1438 (1332)
Ancillary Mg^{2+} Ions[(g)]			
a1	$(OP)_2$	8009 (272)	2611 (2576), 2612 (2577)
a2	$(OP)_2$	8023 (129)	2617 (2582), 2618 (2583)
a3	$(OP)_2BP$	8067 (422)	1845 (1789), 1846 (1790), 1884^{Base} (1828^{Base})
a4	$(OP)_2$	8015 (100)	844 (751), 1689 (1614)
a5	$(OP)_2$	8077 (141)	880 (787), 883 (790)

a) Paleo-Magnesium ions identified by coordination and are conserved in position and coordinating ligands of 23S-rRNA^{HM} and 23S-rRNA^{TT}.
b) Refer to Figure 4.7 for the locations of these ions in the 2D structure and to Figure 4.8 for their locations in the 3D structure.
c) OP indicates non-bridging phosphate; B indicates base atom; S indicates sugar (ribose) atom; P indicates protein atom.
d) These are the numbers in the 1JJ2 coordinate file (23S-rRNA^{HM}). The numbers in parenthesis refer to the numbers in the 2J01 coordinate file (23S-rRNA^{TT}, which follows the E. coli numbering scheme).
e) These are the rRNA residue numbers of the Mg^{2+} first shell in 23S-rRNA^{HM}. The numbers in parenthesis refer to 23S-rRNA^{TT}, which follows the E. coli numbering scheme.
f) This Mg^{2+} does not have conserved coordination in 23S-rRNA^{HM} [$(OP)_3$], and 23S-rRNA^{TT} [$(OP)_4$].
g) These ions are found in close association with Paleo-Mg^{2+} ions and are conserved in the 23S-rRNA^{HM} and 23S-rRNA^{TT} structures.

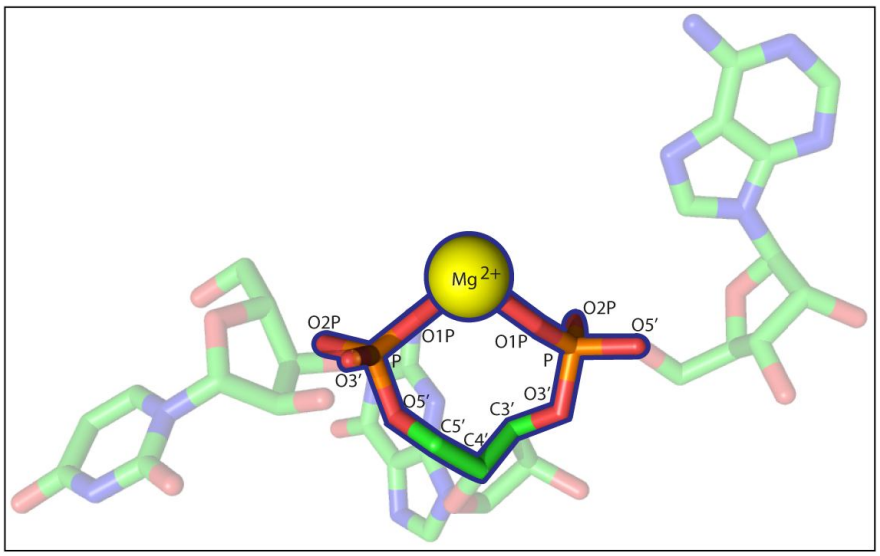

Figure 4.3 RNA as Mg^{2+} chelator.
Shown here is representative bidentate $Mg^{2+}(OP)_2$ complex with a 10-membered chelation cycle. For clarity the base atoms along with ribose atoms C1', C2' O4' and O2' are lightened. This bidentate complex forms a 10-membered chelation cycle (Δ=1). It can be seen that this trinucleotide is in a non-A-helical conformational state. This is Mg^{2+} 8009 from 23S-$rRNA^{HM}$, along with residues U(2610), G(2611) and A(2612).

The most frequent Mg^{2+} chelation motif in RNA is when OP atoms of adjacent residues chelate a common Mg^{2+} (Δ=1, Figure 4.4). Larger ring sizes are less frequent. The Δ=1 is observed at higher frequency that other bidentate complexes (Figure 4.4) because it offers (i) the shortest achievable linker and thus the lowest entropy penalty upon binding, (ii) charged oxygens (OP atoms) as first shell Mg^{2+} ligands, and (iii) favorable RNA bond rotamers. This 10-membered ring system (Mg^{2+}-OP-P-O5'-C5'-C4'-C3'-O3'-P-OP-Mg^{2+}), which requires cis-oriented OP atoms around the Mg^{2+}, is the elemental unit of Mg^{2+} chelation by RNA. Tri- and tetradentate Mg^{2+} complexes nearly always contain at least one of these 10-membered ring systems (Figures 4.5 and 4.6).

Accuracy of 3D structures

The Mg^{2+} positions in 23S-$rRNA^{HM}$ are, as determined by various geometric criteria, highly credible, with a few exceptions. The Mg^{2+} to oxygen distance is an excellent metric in that the predicted and observed (in 23S-$rRNA^{HM}$) frequencies reach

distinct maxima at 2.1 Å and fall to nearly zero by 2.6 Å. As indicated by relative OP positions, essentially all relevant Mg^{2+} ions were added correctly to the model. Other ribosome structures and smaller globular RNAs follow 23S-rRNA^{HM} in the general patterns of Mg^{2+} interaction [for example, compare Figures 4.5 (23S-rRNA^{HM}) & 4.6 (RNA$^{P4\text{-}P6}$ and 23S-rRNA^{TT})]. The same level of confidence does not apply to the monovalent cations of 23S-rRNA^{HM}, which in many cases display unorthodox coordination geometry such as fewer than three first shell ligands, and are judged to be less credible.

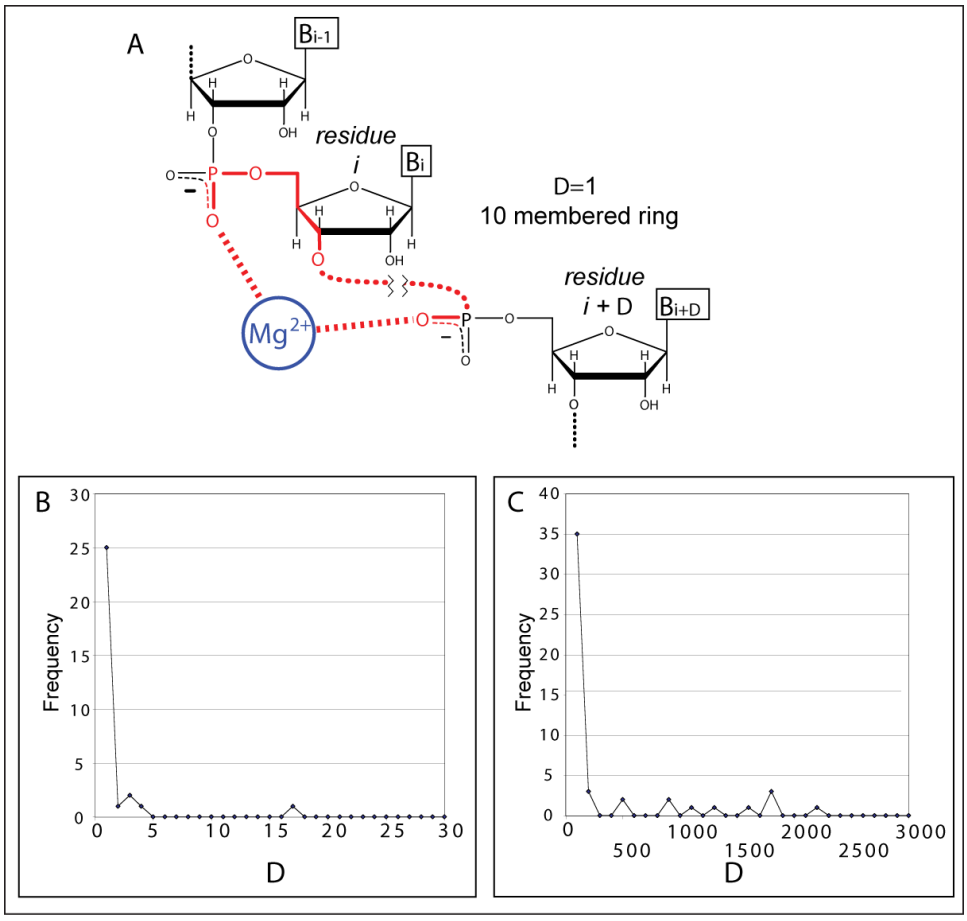

Figure 4.4 Bidentate Mg^{2+} binding is predominantly local.
OP ligands are clustered in along the RNA chain. (A) Definition of Δ, which is the distance in residues between two OP ligands that bind to a common Mg^{2+} ion. Δ = 1 indicates binding by OPs of adjacent residues, giving a 10-membered chelation ring. (B) The observed frequency of occurrence of Δ values in 23s-rRNA^{HM}, expanded to show the distribution at small Δ. Δ = 1 is preferred over all other Δ. (C) The observed frequency of occurrence of Δ values. Small Δ are preferred over large Δ.

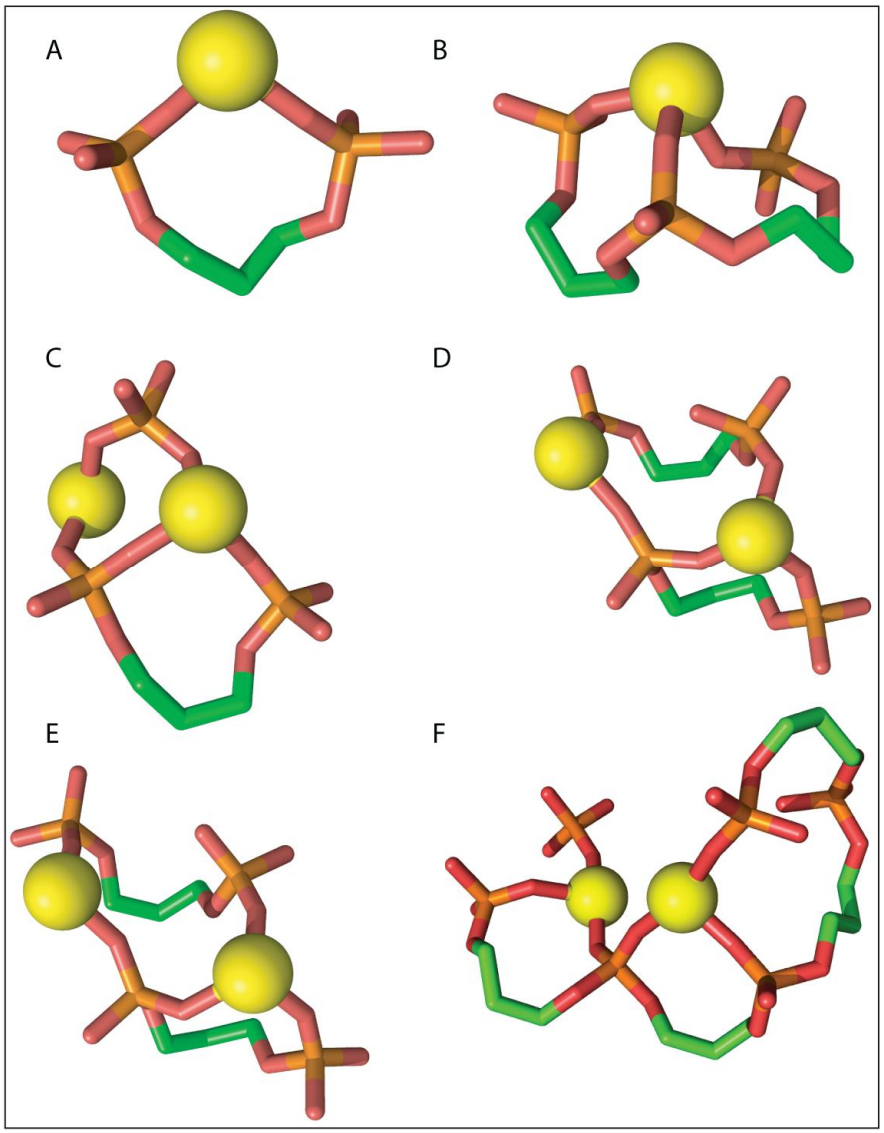

Figure 4.5 RNA as Mg^{2+} chelator.
Shown here are some Mg^{2+} chelation centers in 23S-rRNA^{HM}. The base atoms and ribose atoms C1', C2' O4' and O2' are omitted for clarity. (A) Bidentate chelation by neighboring OP atoms. This common motif is a ten membered-ring with a cis orientation of OP ligands. This is a $\Delta = 1$ complex, indicating that the OP atoms are contributed by reside i and residue i+1 (Mg^{2+} 8009, same as Figure 4.3). (B) Tridentate chelation by three OP atoms from neighboring residues ($\Delta1 = 1$, $\Delta2 = 1$, Mg^{2+} 8026). (C) Dual Mg^{2+} Bicycle Center D1. This is a bicyclic complex with an Mg^{2+}-OP-P-OP-Mg^{2+} bridge head. One Mg^{2+} is involved in tridentate chelation by neighboring and a remote OP atoms (Mg^{2+} 8001, $\Delta1 = 1$, $\Delta2 = 50$). The tridentate complex is coupled to a bidentate complex (Mg^{2+} 8002, $\Delta=51$). (D) Dual Mg^{2+} Bicycle Center D3. This is a bicyclic complex with a Mg^{2+}-OP-P-OP-Mg^{2+} bridge head. One Mg^{2+} is involved in tridentate chelation by two neighboring and one remote OP atoms (Mg^{2+} 8016, $\Delta1 = 1$, $\Delta2 = 175$). The tridentate complex is coupled to a bidentate complex (Mg^{2+} 8029, $\Delta = 176$). (E) Dual Mg^{2+} Bicycle Center D2. This is a bicyclic complex with an Mg^{2+}-OP-P-OP-Mg^{2+} bridge head. One Mg^{2+} is involved in tridentate chelation by two neighboring and one remote OP atoms (Mg^{2+} 8003, $\Delta1 = 1$, $\Delta2 = 1747$). This complex is coupled to a bidentate complex (Mg^{2+} 8013, $\Delta=1746$). (F) Dual Mg^{2+} Tricycle Center D4. This is a tricyclic complex with an Mg^{2+}-OP-P-OP-Mg^{2+} bridge head. One Mg^{2+} is involved in tridentate chelation by two neighboring and one proximal OP atoms (Mg^{2+} 8005, $\Delta1 = 1$, $\Delta2 = 2$). This complex is coupled to another tridentate chelation complex, by two neighboring and one remote OP atoms (Mg^{2+} 8007, $\Delta1 = 1$, $\Delta2 = 1007$).

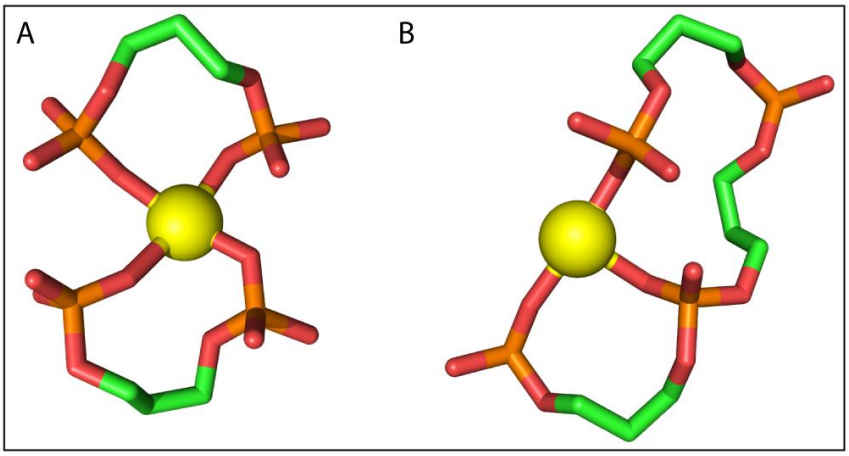

Figure 4.6 RNA as chelator, continued.
(A) Tetradentate chelation in the *Thermus thermophilus* 70S Ribosome. A complex formed by two Δ=1 motifs (Mg^{2+} 245, 2J01). (B) Tridentate Mg^{2+} chelation in $RNA^{P4\text{-}P6}$ (Mg^{2+} 373). Neighboring and next-nearest neighboring phosphate groups (Δ1 = 1, Δ2 = 2) bind to the same Mg^{2+}. Note the similarity of this complex to that in panel F of the previous figure.

Why Mg^{2+}?

The predominant mode of interaction of Mg^{2+} with RNAs large and small is in the form of Mg^{2+}_{aq} and $Mg^{2+}(OP)_1$ complexes (the five water molecules are required to complete the hexacoordinate coordination sphere of $Mg^{2+}(OP)_1$ are not specified for the sake of brevity) . Geometric considerations suggest that these types of Mg^{2+} ions can be substituted by other cations, such as Na^+, polyamines and cationic sidechains of proteins without substantial alteration of RNA conformation.

Specific requirements for Mg^{2+} in RNA folding derive from Mg^{2+} stabilization of distinctive conformational states of RNA. Mg^{2+}- specific states are characterized by bi-, tri- and tetradentate Mg^{2+} complexes with OP atoms [in $Mg^{2+}(OP)_2$, $Mg^{2+}(OP)_3$, $Mg^{2+}(OP)_4$ complexes]. A representative $Mg^{2+}(OP)_2$ complex in $23S\text{-}rRNA^{HM}$ is shown in Figure 4.3. The OP atoms in these multidentate complexes reach the global minimum in OP-OP distances in RNA. These OP atoms are in closer proximity and are more tightly restrained in position than in any other environment, such as when associated with larger ions such as K^+ or Na^+, or polyamines, or cationic sidechains of proteins, or when not

directly associated with ions. Therefore RNA conformation in the vicinity of these Mg^{2+} ions is dependent on and is specific for Mg^{2+}. Such tightly packed OP atoms, in association with Mg^{2+} ions, are found in all globular RNAs including the P4-P6 domain of the tetrahymena ribozyme (Cate *et al.*, 1996a; Cate *et al.*, 1996b; Cate *et al.*, 1997; Basu *et al.*, 1998; Juneau *et al.*, 2001) and ribosomes (Cate *et al.*, 1999; Ban *et al.*, 2000; Wimberly *et al.*, 2000; Harms *et al.*, 2001; Yusupov *et al.*, 2001; Klein *et al.*, 2004; Berk *et al.*, 2006; Selmer *et al.*, 2006; Voss *et al.*, 2006). Only Li^{+} can rival Mg^{2+} in driving the close packing of OP atoms. However the first-shell ligands of Li^{2+} tend to assume tetrahedral rather than octahedral geometry.

Mg^{2+} versus Na^{+}

It has been suggested that high concentrations of Na^{+} attenuate OP-OP repulsion to the extent that globular RNAs can fold, achieving native OP-OP proximities, in the absence of Mg^{2+} (Takamoto *et al.*, 2004). In evaluating such models, one must account for unyielding differences in the coordination chemistry of Na^{+} and Mg^{2+}. The close proximity of adjacent OP atoms in the Mg^{2+} first shell is inconceivable in Na^{+} complexes at any concentration, and the preference of Na^{+} for neutral ligands would drive it to other sites.

It seems that globular RNAs can collapse in the presence of high $[Na^{+}]$ to states stabilized by native-like base-base tertiary interactions, but lacking electrostatic tertiary interactions. One might expect base-base and electrostatic interactions to be somewhat independent of each other because they do not necessarily link the same secondary elements, and do not, on a local level, necessarily act in concert. Where the phosphates come closest together the corresponding bases are remote from each other.

Why chelation? When Chelation?

Multidentate Mg^{2+} binding is commonly under control of local RNA conformation, or conversely, RNA conformation is coupled with multidentate Mg^{2+}

binding. The relationship between local conformation and Mg^{2+} binding is important especially when OP atoms are contributed by adjacent residues (Δ=1, Figures 4.3 and 4.4). Distinctive conformational states are associated with such bidentate complexes. The 10-membered rings push the RNA conformation away from A-helices, tetraloops, etc. Trans bidentate complexes are favored electrostatically (by attenuated OP-OP repulsion) in comparison to cis bidentate complexes, but are disfavored by the entropic cost associated with the necessary increased linker size. The minimum observed linker for a trans $Mg^{2+}(OP)_2$ complex in 23S-rRNA$^{HM/TT}$ is four residues.

What stabilizes cis Mg^{2+} $(OP)_2$ complexes, in which OP atoms are forced into close proximity? Firstly, OP - - Mg^{2+} attraction offsets the OP-OP repulsion. Secondly, repulsion is attenuated by charge transfer from OP to Mg^{2+}. Third, RNA backbone allows formation of 10-membered chelation rings in the absence of rotameric restraints, i.e., in the absence of unfavorable bond rotations. The implications of multidentate Mg^{2+} coordination transcend thermodynamics. Close OP-OP proximities present kinetic barriers to changes in coordination state (Bandyopadhyay and Bhattacharyya, 2003).

Many multidentate $Mg^{2+}(OP)_n$ complexes are found in 23S-rRNA^{HM} as described by Klein, Moore and Steitz (Klein *et al.*, 2004). As noted there, the most frequent chelation motif is $Mg^{2+}(OP)_2$, where phosphate groups from neighboring residues chelate a common Mg^{2+} (Figures 4.3-4.6). Twenty five of these 10-membered bidentate chelation cycles are observed in 23S-rRNA^{HM}. The OP ligands within the 10-membered cycles can be either O1P or O2P atoms, and are invariably in the cis orientation around the Mg^{2+}. For these 10-membered cycles, by definition Δ=1, where Δ is the distance, in number of residues, between the two OP groups (Figure 4.4). The tridentate complex shown in Figure 4.1B yields two different Δ values; 1 and 95. Nine of ten $Mg^{2+}(OP)_3$ centers in 23S-rRNA^{HM} contain at least one Δ=1 complex. The $Mg^{2+}(OP)_3$ center in Figure 4.5B is composed of two Δ = 1 complexes.

The chelating ring size patterns observed in 23S-rRNA^{HM} are general features of large RNAs, in which $Mg^{2+}(OP)_2$ Δ=1 complexes are observed alone, and in combination with other chelation rings. A double Δ=1 complex, with tetradentate chelation is observed linking the 16S RNA with the mRNA in the intact ribosome of Thermus thermophilus (PDB entry 2J01, Figure 4.6A). A homolog of the $Mg^{2+}(OP)_3$ $\Delta_1 = 1$, $\Delta_2 = 2$ complex that forms part of the D4 center of 23S-rRNA^{HM} (Figure 4.5F) is found in RNA$^{P4\text{-}P6}$ (Figure 4.6B).

Bicycles, Tricycles.

Tridentate Mg^{2+}-RNA complexes, although less frequent than monodentate and bidentate complexes, are especially important in structure, stability and function. Ten Δ=1 cycles are fused with secondary cycles to form bicyclic $Mg^{2+}(OP)_3$ structures in 23S-rRNA^{HM}. The bicycles fall into two classes; those composed of RNA and a single Mg^{2+} ion (a Δ=1, Δ=1 example is shown in Figure 4.5B) and those containing Mg^{2+}-OP-P-OP-Mg^{2+} linkages (Figures 4.5C-F). Four Mg^{2+}-OP-P-OP-Mg^{2+} linked bicycles are observed in 23s-rRNA^{HM}. In these dual Mg^{2+} bicycles, called here D1, D2, D3, and D4, both the O1P and O2P of a single phosphate group are first shell Mg^{2+} ligands. In each of these centers, a tridentate $Mg^{2+}(OP)_3$ complex is paired with, and mutually stabilizes a bidendate $Mg^{2+}(OP)_2$ complex. The exception is D4, in which two $Mg^{2+}(OP)_3$ complexes form a tricyclic structure. A dual Mg^{2+} bicycle (or tricycle) is essentially a single extended structural unit, of high rigidity, and with stability greater than the sum of the parts.

The high measure of similarity within subsets of these dual Mg^{2+} bicycles [compare Figures 4.5D (D3) with 4.5E (D2)] suggests general rules of RNA conformation and interaction are discernable from analysis of them.

Paleo-Magnesium Ions

Here we introduce the concept of the paleo-Mg^{2+} ion. Paleo-Mg^{+2} ions play key roles in RNA folding, stability and function, and are conserved over vast evolutionary timescales. They were identified initially by their RNA coordination, then validated by comparison between different rRNAs. Broadly, a paleo-Mg^{2+} ion is in either a $Mg^{2+}(OP)_3$ center, with at least three RNA first shell OP ligands, or is component of a dual Mg^{2+} bicycle or tricycle. Five ancillary $Mg^{2+}(OP)_2$ ions are located in close association with Paleo-Mg^{2+} ions in 23S-rRNA$^{HM/TT}$. Figures 4.5B-F show examples paleo-Mg^{2+} ions from 23S-rRNA^{HM}. The requirement for three OP first shell ligands over base and sugar ligands but is founded on the lower frequency and smaller contribution to stability of base/sugar ligands compared with OP ligands.

In sum 13 paleo-Mg^{2+} ions associate with 23s-rRNA^{HM} (Figures 4.7 & 4.8, Table 4.4). Their importance in RNA folding, stability and evolution is underscored by the following.

(i) Conservation in 3D. The paleo-Mg^{2+} ions are highly conserved in position and mode of interaction between the 23s rRNAs of HM (archea) and TT (bacteria). There is essentially a 1:1 mapping of paleo-Mg^{2+} ions, of the surrounding RNA conformation, and Mg^{2+} coordination geometry between the two 3D structures, which are separated by billions of years of evolution.

(ii) Effect on 3D structure. It is apparent that $Mg^{2+}(OP)_3$ complexes alone and especially dual Mg^{2+} bicycles and tricycles form unique structural entities with rigidity, stability and forced dispositions of functional groups that cannot be approximated by RNA alone or in conjunction with other ions.

(iii) Conservation in 2D. Paleo-Mg^{2+} ions generally associate with the most conserved 2D elements in rRNA, and link these conserved 2D elements by electrostatic tertiary interactions. These elements are remote in secondary structure. The relationship of paleo-Mg^{2+} ions to secondary structural elements of 23S-rRNA^{HM} is shown in Figure

4.7. The secondary elements linked by paleo-Mg^{2+} ions are conserved between bacteria and archea (see next), in the proposed secondary structures of eukaryotic (Cannone *et al.*, 2002) and mitochondrial rRNAs,(Mears *et al.*, 2006) and in a proposed minimal 23S-rRNA (Mears *et al.*, 2002).

(iv) Conservation of Sequence. The base sequence of the RNA surrounding paleo-Mg^{2+} ions is highly conserved between the 23s rRNAs of HM and TT. Where the sequences do differ, only purine to purine or pyrimidine to pyrimidine substitutions are allowed.

(v) Role in Function. The locations of paleo-Mg^{2+} ions appear to lend critical support to function (Figure 4.8). Ten Paleo-Mg^{2+} ions form a loose ring around the peptidyl transfer center. Three paleo-Mg^{2+} ions are located by the exit of the peptide tunnel. Dual Mg^{2+} bicycles and tricycles D1, D2 and D4 flank the peptidyl transfer center, while D3 is located by exit of the peptide tunnel.

(vi) Paleo-Mg^{2+} ions are not coordinated by protein ligands (Table 4.4), suggesting ancestry prior to development of the ribosomal proteins.

(vii) Linkage with RNA conformation. Paleo-Mg^{2+} ions, by nature of their 10-membered chelation cycles, impose constraints on RNA conformation and topology. The relationship between RNA conformation and Mg^{2+} chelation is discussed below.

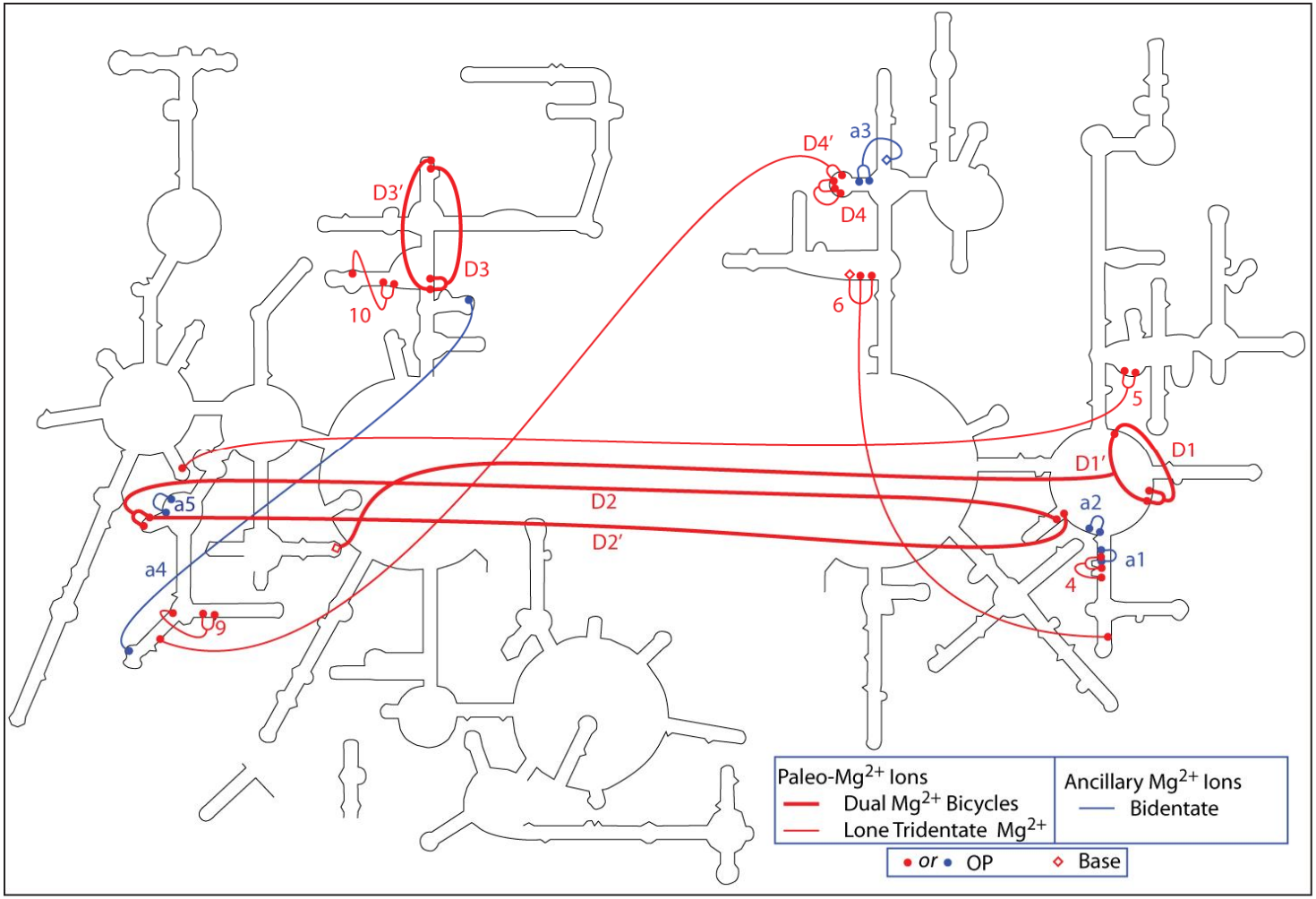

Figure 4.7 The secondary structure of 23S-rRNA^{HM}.
The locations of paleo-Mg^{2+} ions are depicted in red. The thick red lines indicate the dual Mg^{2+} bicycles (D1, D2, D3 and D4). The thin red lines indicate the isolated $Mg(OP)_3$ complexes. The ancillary $Mg^{2+}(OP)_2$ ions, which cluster on the 2D map with paleo-Mg^{2+} ions, are shown in blue. The circles indicate the OP atoms that contact Mg^{2+} ions. Diamonds indicate base atoms that contact Mg^{2+} ions.

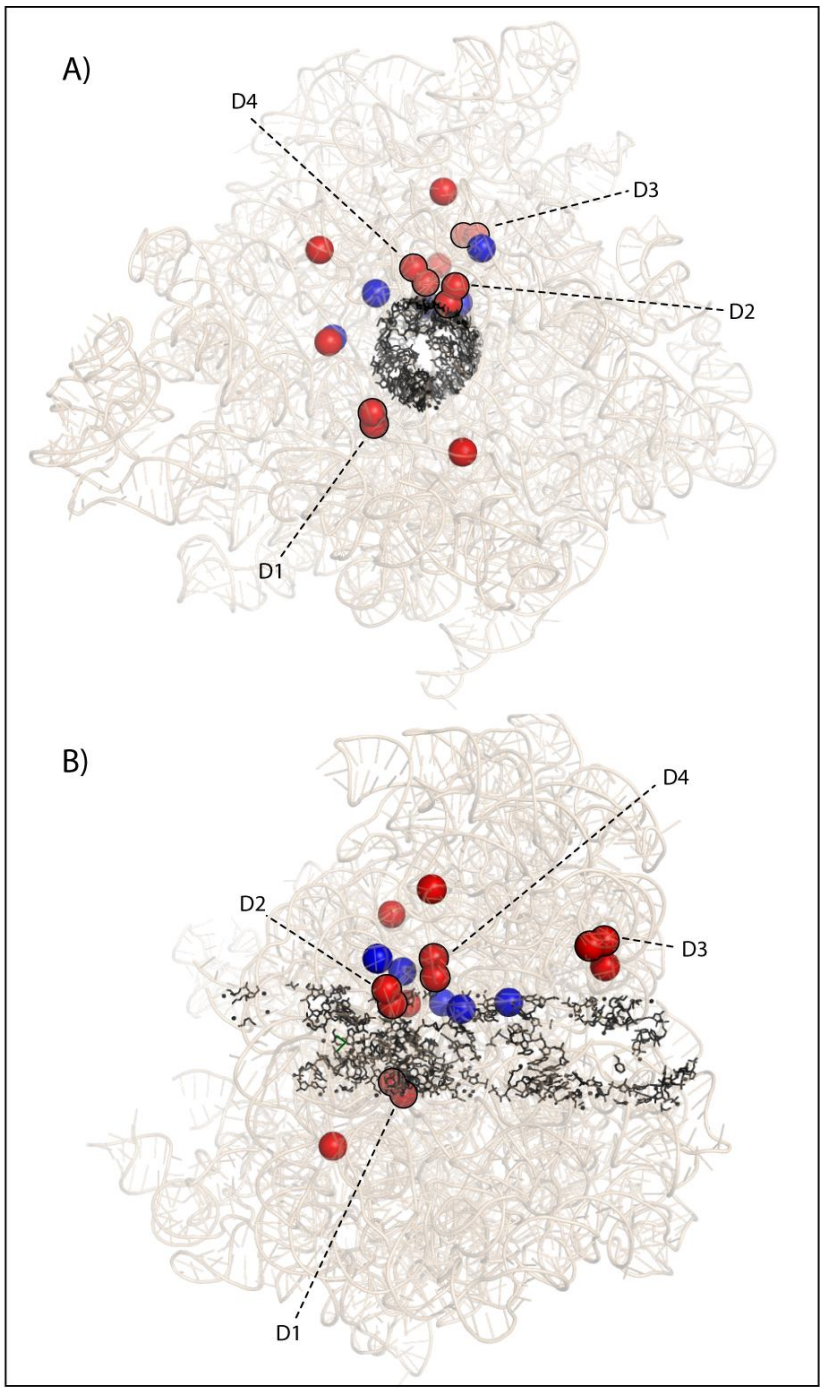

Figure 4.8 The locations of Paleo- / Ancillary- Mg^{2+} ions in 23S-rRNA^{HM}.
(A) View into the Peptidyl Transfer Center of 23S-rRNA^{HM}. The thirteen Paleo-Mg^{2+} ions are shown as red spheres. The Mg^{2+} ions of the four dual Mg^{2+} bicycles/tricycle (D1, D2, D3 and D4) are indicated. Ancillary Mg^{2+} ions are shown in blue. The RNA atoms lining the peptide exit tunnel are accented in black. (B) This view, looking across the peptide tunnel, is rotated by 90° relative to the top panel. The radii of the Mg^{2+} ions are increased over their normal ionic size for clarity and have no physical significance. The proteins and 5S rRNA are omitted for clarity.

Mg^{2+} avoids RNA motifs

Chelation of Mg^{2+} is coupled to RNA conformation. Mg^{2+} ions select against chelation complexes with canonical conformations such as A-form helices and tetraloops. As noted by Moore (Moore, 1999) and others (Leontis and Westhof, 2003), folded RNA is largely composed of a relatively small number of motifs such as A-helices (Saenger, 1984), tetraloops (Tuerk *et al.*, 1988; Woese and Gutell, 1989; Woese *et al.*, 1990), E-loop motifs (Wimberly *et al.*, 1993; Szewczak and Moore, 1995; Leontis and Westhof, 1998; Correll *et al.*, 2003; Vallurupalli and Moore, 2003), kink-turns (Klein *et al.*, 2001; Matsumura *et al.*, 2003; Goody *et al.*, 2004), etc. RNA motifs are essentially equivalent to RNA secondary structural units, that form early in RNA folding processes. We have used multi-resolution data-mining approaches to extend the definition of RNA motifs, to allow for deletions, insertions, strand clips and topology switches (Hsiao *et al.*, 2006). Formation of RNA motifs (secondary structure) is not Mg^{2+}-dependent.

One can observe that Mg^{2+} inner shell complexes select against RNA tetraloops. In Figure 4.9, first-shell OP interactions with Mg^{2+} are mapped onto the 23S-$rRNA^{HM}$ secondary structure, as are the locations of tetraloops (Hsiao *et al.*, 2006). The sites of first shell Mg^{2+} coordination do not in general correspond with the locations of the tetraloops.

This preference is supported by statistical results. Automated methods (Hershkovitz *et al.*, 2003; Hershkovitz *et al.*, 2006; Hsiao *et al.*, 2006) allow one to count conformational states, determine their populations (frequencies), locations and sequences. One can seek correlations between frequency of occurrence of conformational states and locations of site-bound Mg^{2+} ions. 23S-$rRNA^{HM}$ can be partitioned into conformational states, and grouped by frequency.

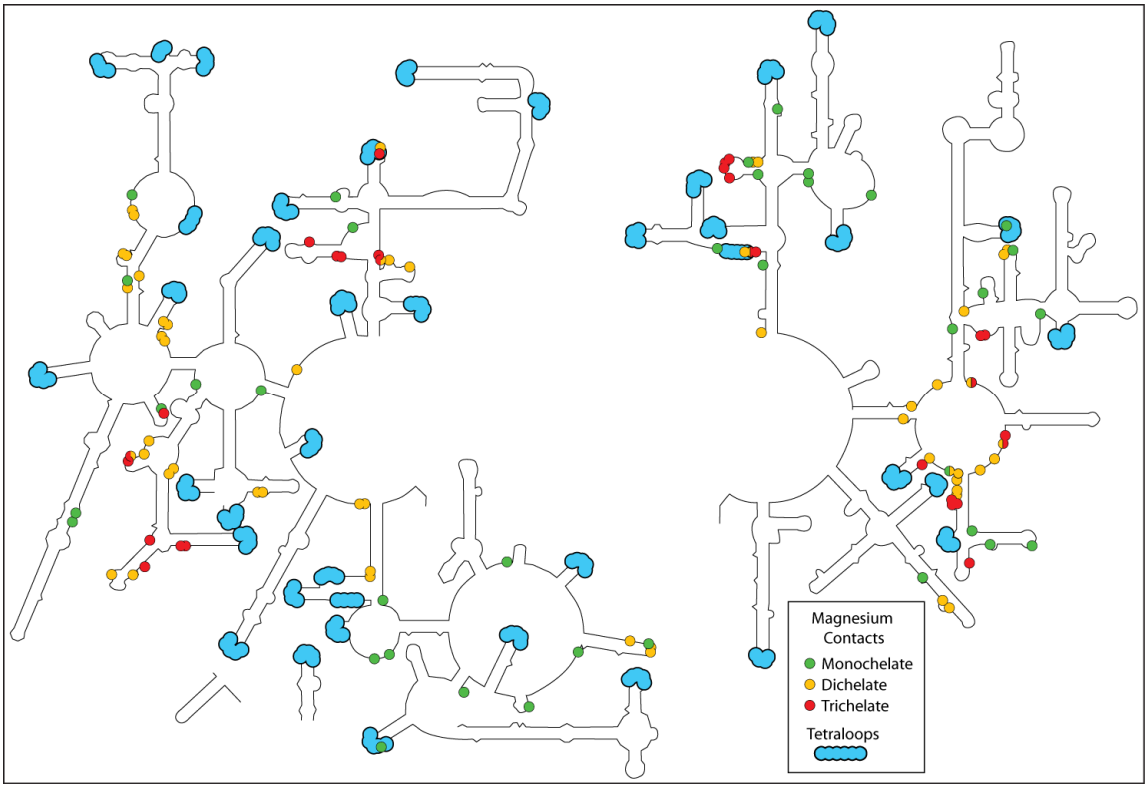

Figure 4.9 Multidenate Mg^{2+} binding selects against tetraloops.
The secondary structure and Mg2+ contacts of the 23s rRNA from HM are shown. Tetraloop positions, as determined by Hsiao,(Hsiao *et al.*, 2006) are indicated in blue. Mg2+-OP interactions are indicated by circles. OP atoms that interact with Mg2+ via monodentate interactions are green circles. OP atoms that interact with Mg2+ via bidentate interactions are yellow circles. OP atoms that interact with Mg2+ via tridentate interactions are red circles.

RNA associated with Mg^{2+} is dispersed in conformational space

The diversity of RNA conformation in the vicinity of $Mg^{2+}(OP)_2$, even with conserved RNA ligands, is illustrated in Figure 4.10. The conformation of RNA acting as bidentate Mg^{2+} ligands is different from the conformation of other RNA, including RNA in the vicinities of other bidentate Mg^{2+} ions. This result is consistent with the observation of Klein Moore and Steitz that Mg^{2+} ions tend to bind to "highly idiosyncratic binding sites in 23S RNA that are un-like any previously reported" (Klein *et al.*, 2004). However there are reports that in some cases Mg^{2+} ions bind to specific motifs such as non-Watson-Crick base pairs in the E-loop of the 5S rRNA (Leontis and Westhof, 1998).

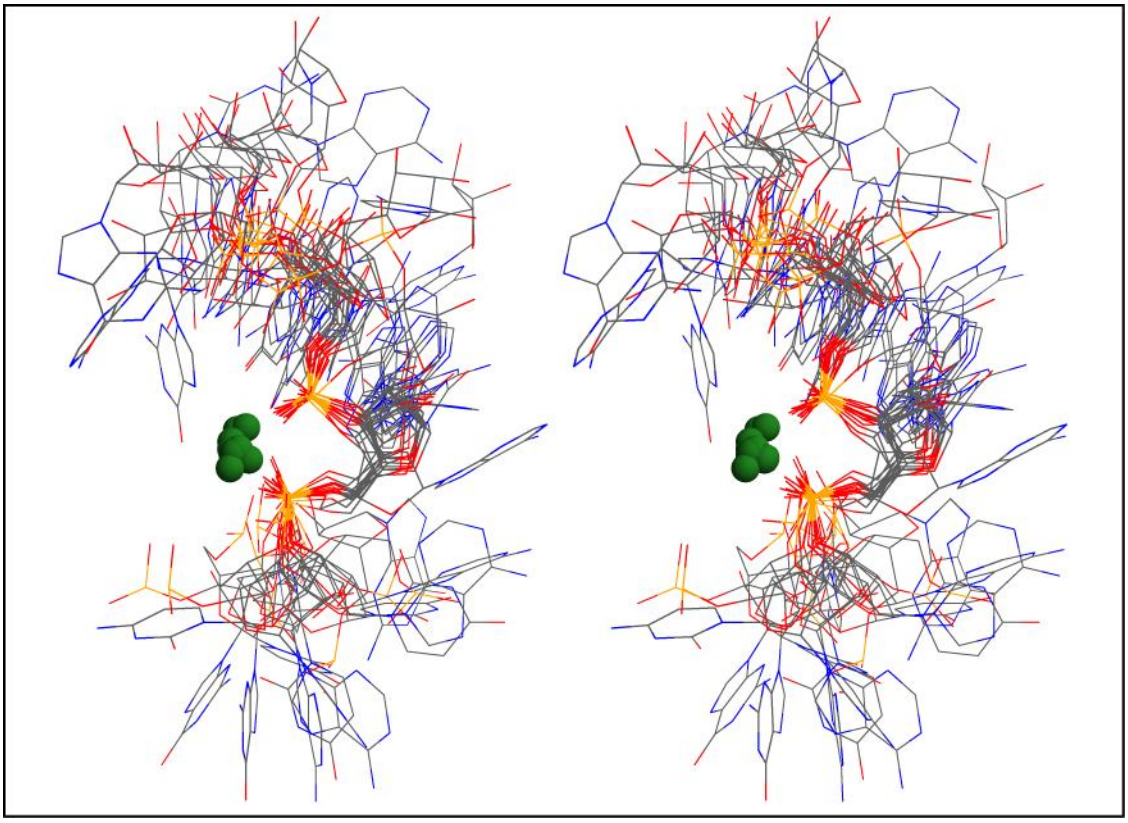

Figure 4.10 Mg^{2+} ions that form multiple bonds to RNA prefer conformational-deviants. Fragments of RNA that bind directly to Mg^{2+} were superimposed. Each of these RNA fragments bind to Mg^{2+} use the same two atoms (O1 of residue i and OP2 of residue i+1). The backbone atoms of the two residues that bind to the Mg^{2+} ions were used for the superimposition. Mg^{2+} ions are represented by green spheres. Other Mg^{2+} ligands (water molecules) are omitted for clarity.

Mg^{2+}-OP Energetics

Cations interact with nucleic acids by both electrostatic and non-electrostatic interactions. Stability is influenced by solvent screening and other ions, and by entropic effects (Sponer *et al.*, 2001; Petrov *et al.*, 2002; Gresh *et al.*, 2003; Rulisek and Sponer, 2003; Petrov *et al.*, 2004). Entropic gain by water release may in some cases be a dominant factor in binding, especially for Mg^{2+} (below). High-level calculations underscore the importance of nonelectrostatic contributions, such as polarization and charge transfer. For example, calculations by Sponer suggest that inner-shell binding Mg^{2+} to the N7 position of a guanine significantly reduces the energy of the outer-shell binding of the same Mg^{2+} to a phosphate group, due to polarization and charge transfer

from Mg^{2+} to guanine (Rulisek and Sponer, 2003). NLPB, molecular dynamics simulation and other computational approaches generally assume that electrostatics are the only significant contribution to energy of binding. In some applications a level of error should be anticipated from these approximations.

The thermodynamic fingerprint for OP interactions with Mg^{2+} is anomalous in comparison with OP interactions with other divalent cations. For example the standard enthalpy of interaction of ATP in aqueous media with Mg^{2+} is small and positive (Khan and Martell, 1966). Other divalent cations (Ca^{2+}, Mn^{2+}, Zn^{2+}, etc.) give negative enthalpies for this reaction. Thus in water, enthalpic factors are less favorable for forming ATP complexes with Mg^{2+} than with other divalent cations. The effect is related to the large heat of hydration of Mg^{2+}. The negative enthalpy of ATP-Mg^{2+} interaction is offset by a larger negative entropy of interaction of Mg^{2+} with H_2O. The magnitude of the entropic effect is much greater for Mg^{2+} than for other divalent cations. Thus the association of ATP with Mg^{2+} is driven by entropy – arising from the release of water molecules. This thermodynamic paradigm is specific to Mg^{2+}, and should be applicable aqueous Mg^{2+} - OP interactions in general.

4.4 Experimental methods for determination of cation positions in x-ray structures

The solvent/ion environment in the vicinity of nucleic acids is difficult to fit unambiguously to x-ray diffraction data. Solvent positions and species identification by x-ray diffraction should be considered to be approximations (Williams, 2005). Monovalent cations in particular present non-trivial analytical challenges. Sodium ions (Na^+), potassium ions (K^+), rubidium ions (Rb^+), cesium ions (Cs^+), ammonium ions and water molecules, and even polyamines and divalent cations, compete for overlapping sites adjacent DNA and RNA (Shui *et al.*, 1998; McFail-Isom *et al.*, 1999; Howerton *et*

al., 2001). A variety of cation species coexist and co-localize, giving partial and mixed occupancies. In addition, Na^+ and NH_4^+ scatter x-rays with nearly the same power as water or partially occupied K^+. Therefore it is not generally correct to exclude K^+ in favor of other species based on thermal factors. One can interchangeably fit H_2O, NH_4^+, Na^+, K^+, etc., to many solvent peaks simply by adjusting or fitting occupancies.

4.4.1 Group I

Tl^+ was initially investigated as a K^+ mimic in biological systems by R.J.P. Williams (Manners *et al.*, 1970; Williams, 1971) and others (Britten and Blank, 1968; Inturris.Ce, 1969b; Inturris.Ce, 1969a; Post *et al.*, 1969; Kayne, 1971; Reuben and Kayne, 1971). Tl^+ can effectively replace K^+ in diol dehydratase, pyruvate kinase, some phosphatases, and other enzymatic systems. In x-ray diffraction experiments, the anomalous signal of Tl^+ renders it a beacon that circumvents the necessity of interpretation of subtle differences in coordination geometry and scattering power of K^+ versus Na^+ versus H_2O. More recently Tl^+ has been used as a K^+ substitute in the catalytic mechanisms of sodium-potassium pumps (Pedersen *et al.*, 1998), fructose-1-6-bisphosphatase (Villeret *et al.*, 1995), and pyruvate kinase (Loria and Nowak, 1998). Tl^+ has been shown to stabilize guanine quadruplexes in a manner analogous to K^+ and ammonium (Basu *et al.*, 2000; Caceres *et al.*, 2004; Gill *et al.*, 2006). We have used Tl^+ as a probe for K^+ in association with B-DNA (Howerton *et al.*, 2001; Moulaei *et al.*, 2005) and in DNA-drug complexes (Howerton *et al.*, 2003). Doudna and coworkers used Tl^+ as a probe for K^+ in the tetrahymena ribozyme P4-P6 domain (Basu *et al.*, 1998). Draper and coworkers used Tl^+ as a K^+ probe in the structure of a fragment of the 23S rRNA (Conn *et al.*, 2002). Correll used Tl^+ as a K^+ probe in the structure of the sarcin/ricin loop of the 23S rRNA. Tl^+ was used by Caspar and coworkers to determine counterion positions adjacent to insulin (Badger *et al.*, 1994b; Badger *et al.*, 1994a) and by Gill and Eisenberg to determine the positions of ammonium ions in the binding pocket of

glutamine synthetase (Gill and Eisenberg, 2001). Rb^+ has also been used, with some success as a K^+ substitute in DNA structures (Tereshko *et al.*, 1999; Tereshko *et al.*, 2001) and in ribosomal structures (Klein *et al.*, 2004).

4.4.2 Group II

In contrast to monovalent cations, Mg^{2+} ions can often be identified by coordination geometry. As noted above, Mg^{2+} is surrounded by an octahedron of first-shell ligands, generally oxygen atoms, with ligand to Mg^{2+} distances of 2.1 and ligand to ligand distances of 2.9 Å (above, Figure 4.1B). No other species found in proximity to DNA/RNA has this geometric fingerprint. Mg^{2+} ions that appear to have coordination numbers of less than six are disordered and/or partially occupied. Mg^{2+} can generally be substituted by manganese (Mn^{2+}) (Eisinger *et al.*, 1962; Eisinger *et al.*, 1965; Reuben and Cohn, 1970; Bock *et al.*, 1999; Feig, 2000), which gives a useful anomalous signal (Cate *et al.*, 1997; Salgado *et al.*, 2005). Cobalt hexamine is a useful NMR probe for fully hydrated magnesium (Gessner *et al.*, 1985; Sen and Crothers, 1986; Braunlin *et al.*, 1987; Cowan, 1993; Rudisser and Tinoco, 2000; Brannvall *et al.*, 2001).

4.5 Reaction Coordinates for RNA Folding

4.5.1 The utility of 3D databases for determining mechanism

Crystal structures, when averaged, can provide excellent predictions of solution behavior. It has been observed that relative populations over a large number of crystal structures reflect populations and relative energies in solution (Allen *et al.*, 1996; Taylor, 2002). Structural databases allow determination of averages and deviations of bond and hydrogen bond lengths, bond angles and dihedrals (Taylor *et al.*, 1983; Taylor *et al.*, 1984). Structural databases also allow determination of coordination sphere geometry (Bock *et al.*, 1994; Markham *et al.*, 2002; Bock *et al.*, 2006), and reaction coordinates

and transition pathways (Burgi, 1973; Burgi *et al.*, 1973; Sundaralingam and Sekharudu, 1989; Allen *et al.*, 2003; Bandyopadhyay and Bhattacharyya, 2003; Hays *et al.*, 2005).

Burgi and Dunitz used data-mining of crystal structures to determine reaction coordinates for simple organic reactions (Burgi, 1973; Burgi *et al.*, 1973; Burgi *et al.*, 1974). Similarly, reaction coordinates for conformational transition reaction coordinates (Vargason *et al.*, 2001) and along folding reaction coordinates (Sundaralingam and Sekharudu, 1989) have been determined for biological polymers.

4.5.2 Mg^{2+}-RNA complexes report on folding intermediates

As noted in previous sections, in ground state crystal structures of large RNAs OP atoms of $Mg^{2+}(OP)_2$ complexes tend to be from neighboring residues (i.e., Δ=1 is most probable). Further, RNA in $Mg^{2+}(OP)_2$ complexes is conformationally polymorphic [i.e., $Mg^{2+}(OP)_2$ complexes are conformational-deviants].

Does one expect to capture such complexes within a large folded RNA? Yes, *if* $Mg^{2+}(OP)_2$ complexes for preferentially with single-stranded regions of RNA folding intermediates. Adjacent residues along a *single-stranded (i.e., flexible) RNA* chain achieve proximity with greatest probability (Flory, 1953), and are conformationally most polymorphic. Adjacent residues along a *double-stranded (i.e., relatively rigid) RNA* chain achieve proximity with lower probability, and are conformationally homogeneous.

Thus the combined data suggest that much of Mg^{2+} binding to folding intermediates (i) is local, (ii) occurs in flexible, single-stranded regions, (iii) dampens flexibility, and decreases the available number of conformational states, and (iv) is fast on the timescale of large-amplitude RNA conformational change. However, a subset of Mg^{2+} ions stitch together RNA elements that are remote in primary sequence (i.e., are distant along the backbone). During RNA folding, such electrostatic tertiary interactions may form after many base-base tertiary interactions. A summary of this RNA folding model is shown in Figure 4.11. Support for this general mechanism is provided by results

of Woodson and coworkers (Koculi *et al.*, 2006), who conclude that Mg^{2+} dampen the dynamics of RNA folding intermediates. The mechanism here is consistent with preferential binding of Mg^{2+} to ssRNA over dsRNA as observed experimentally in solution (Kankia, 2003) even though that preference is counter to the predictions of polyelectrolyte theory (Manning, 1978; Record *et al.*, 1998).

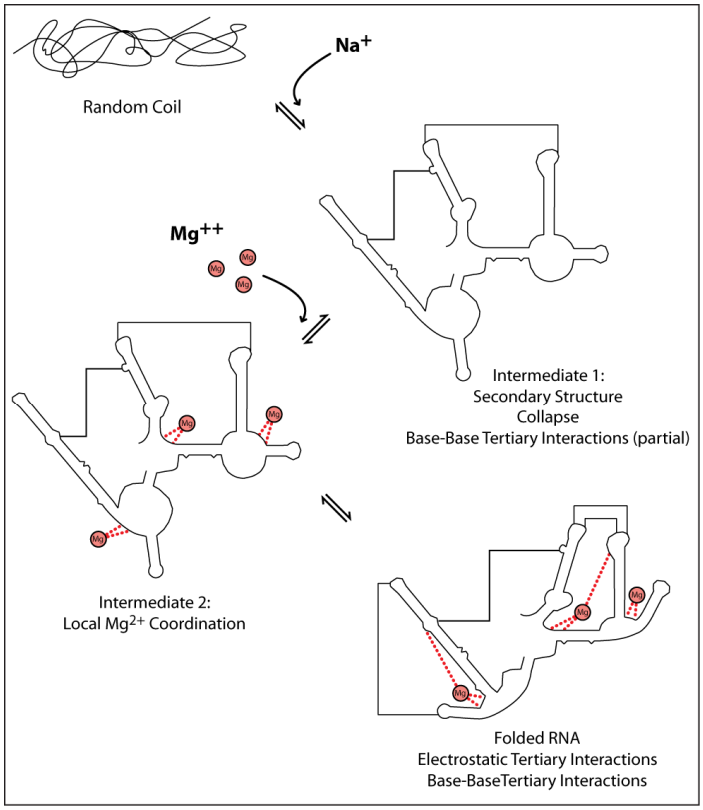

Figure 4.11 Two limiting mechanisms of RNA folding.
Mechansim I involves local RNA binding. Mechanism II involves long-range 'tertiary' interactions.

4.6 Acknowledgements.

The authors thank Drs. J. Michael Schurr, Nicholas Hud and R.J.P. Williams for helpful discussions.

CHAPTER 5

A RECURRENT MAGNESIUM-BINDING MOTIF PROVIDES A FRAMEWORK FOR THE PEPTIDYL TRANSFERASE CENTER

5.1 Introduction

The ribosome is a macromolecular machine responsible for the synthesis of all proteins in all living organisms. Here we demonstrate that the ribosomal peptidyl transferase center (PTC), the site of protein synthesis, is supported by a framework of magnesium microclusters (Mg^{2+}-μc's). We show that the Mg^{2+}-μc is a recurrent motif in large RNAs, with pivotal roles in RNA folding, function and evolution.

5.2 Methods

RNA structures were obtained from the PDB (Berman *et al.*, 1992). First shell Mg^{2+}-ligand interactions are defined by distances less than 2.4 Å. Hydrogen bond distances are less than 3.4 Å between heavy atoms. A exhaustive survey of all Mg^{2+}-RNA interactions was conducted with the MeRNA database (Stefan *et al.*, 2006). L2/L8 Homologs were obtained using Blast. Protein alignment was performed with Clustalw (Larkin *et al.*, 2007).

5.3 Results

5.3.1 Mg^{2+}-μc's (D1, D2, D3, D4)

Four Mg^{2+}-μc's (called Mg^{2+}-μc D1… Mg^{2+}-μc D4, Figure 5.1) are located within the 23S rRNAs of *H. marismortui* [an archaebacterium, PDB entry 1JJ2 (Ban *et al.*, 2000)] and *T. thermophilus* [a bacterium, PDB entry 2J01 (Selmer *et al.*, 2006)]. These Mg^{2+}-μc's appear to predate the last universal common ancestor of life, because they are highly conserved in position, in RNA interactions, and in protein interactions (Figure 5.2) even though bacteria and archea diverged at the LUCA, several billions of years ago (Olsen and Woese, 1997). Mg^{2+}-μc's, including those in the 16S rRNA, the P4-P6 domain of the tetrahymena Group I intron ribozyme (Cate *et al.*, 1997), and a Group II intron ribozyme (Toor *et al.*, 2008), are defined by common features (Figure 5.2A) including (i) Mg^{2+}-(O1P-P-O2P)-Mg^{2+} bridges, (ii) 10-membered chelation rings, utilizing phosphate groups of adjacent residues as Mg^{2+} ligands, (iii) crystalline-like Mg^{2+}-Mg^{2+} proximities, (iv) direct Mg^{2+}-phosphate interactions and Mg^{2+} dehydration, (v) undulated RNA surfaces with unpaired and unstacked bases, and (vi) and usually, close proximity to site of catalysis. Each Mg^{2+}-μc contains two Mg^{2+} ions plus the RNA that engages in first and second shell interactions with the paired Mg^{2+} ions. The position and conformation of Mg^{2+}-μc RNA is constrained by the Mg^{2+} ions. The relative positions of the Mg^{2+} ions are constrained by the Mg^{2+}-(O1P-P-O2P)-Mg^{2+} bridges.

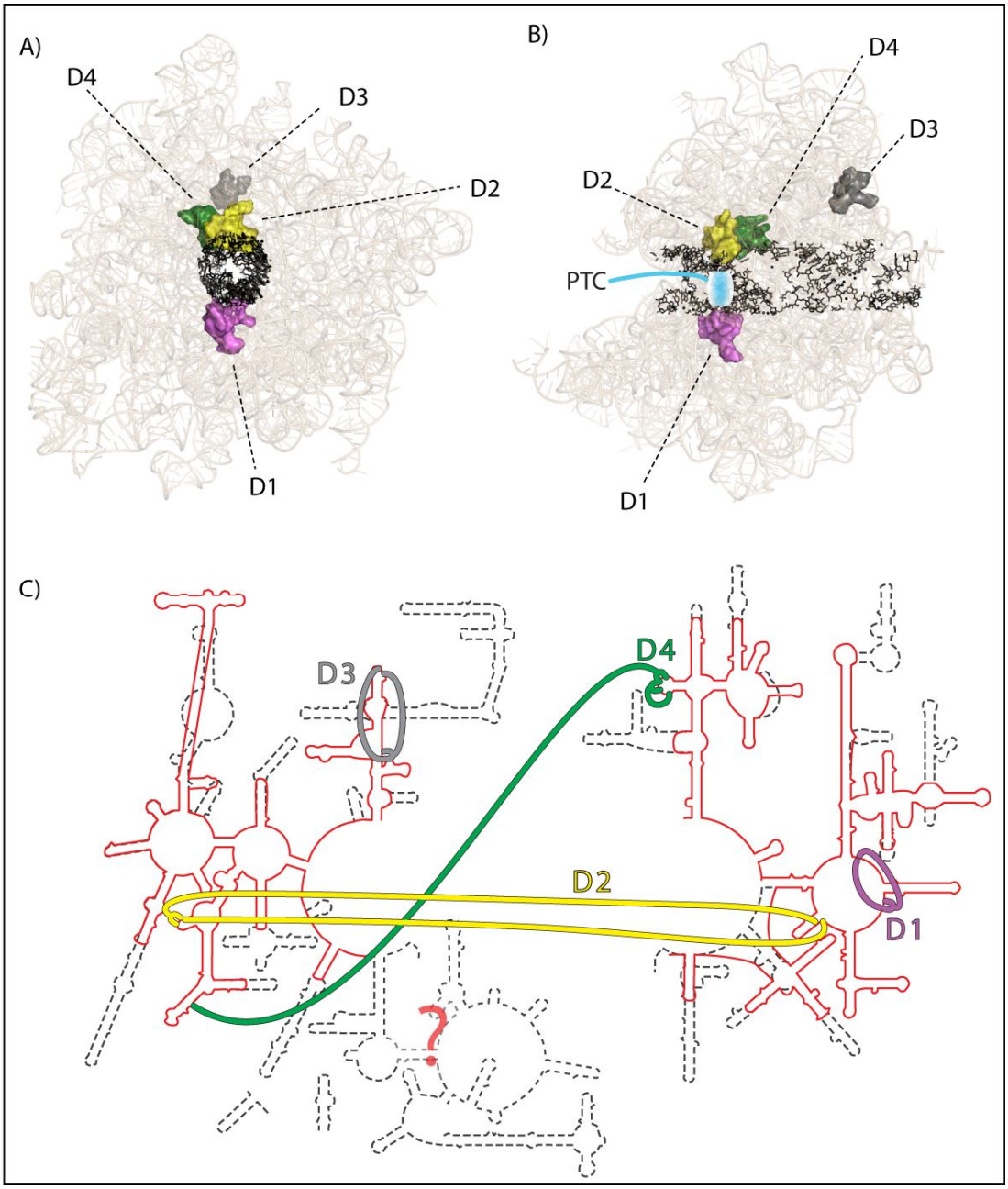

Figure 5.1 Locations of the four Mg^{2+}-μc's in the 2D and 3D structures.
(A) View into the PTC of *H. marismortui* (PDB entry 1JJ2). The four Mg^{2+}-μc's are represented as solid surfaces. The RNA atoms lining the peptide exit tunnel are accented in black. Mg^{2+}-μc's D1, D2, and D4 encircle the PTC. Mg^{2+}-μc's are colored: D1, purple; D2, yellow; D3, gray; D4, green. Ribosomal proteins and the 5S rRNA are omitted for clarity. (B) This view, looking across the peptide tunnel, is rotated by 90° relative to the panel A. (C) The secondary structures of LSU rRNAs of *H. marismortui* [23S rRNA (Ban *et al.*, 2000), dashed black line] and the mitochondrion of *B. taurus* [16S rRNA(Sharma *et al.*, 2003)], red line). Phosphate groups that are linked by electrostatic tertiary interactions within Mg^{2+}-μc's are indicated by colored lines. The secondary structural elements that interact with Mg^{2+}-μc's are conserved. In the *C. elegans* LSU, the rRNA that binds to D3 is absent (Mears *et al.*, 2002). The question mark indicates a small portion of the mitochondria of *B. Taurus* LSU rRNA for which the secondary structure is unknown.

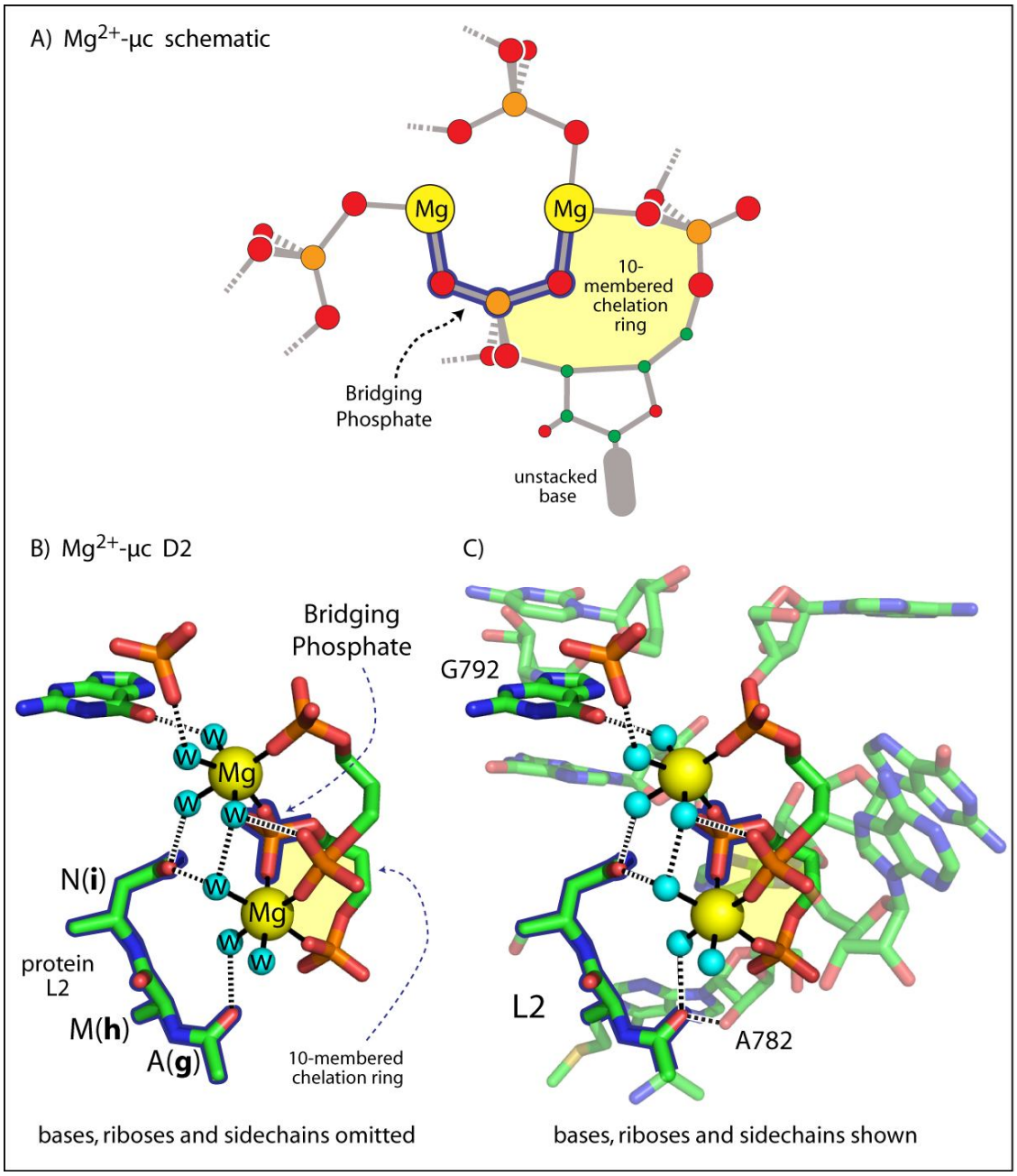

Figure 5.2 Mg^{2+}-µc schematic.
(A) A schematic diagram illustrating the features common to Mg^{2+}-µc's. Mg^{2+}-µc's are characterized by Mg^{2+}-(O1P-P-O2P)-Mg^{2+} bridges (outlined in blue), 10-membered chelation cycles (yellow), unstacked bases, and first shell Mg^{2+}-phosphate interactions. Carbon is green, oxygen is red, phosphorous is orange and magnesium is yellow. (B) Mg^{2+}-µc D2 from *H. marismortui* with bases, oxyriboses and protein sidechains omitted. A three residue fragment of L2 contains universally conserved asparagine (**g**) and methione (**h**). First-shell Mg^{2+} contacts are black solid lines. Hydrogen bonds are dashed lines. This Mg^{2+}-µc contains Mg^{2+} ions 8003 and 8013 of PDB entry 1JJ2. (C) Mg^{2+}-µc D2 from *H. marismortui* with bases, oxyriboses and protein sidechains omitted included. *E. coli* RNA residue numbers are shown. The *H. Marismortui* residue numbers are G885 and A875.

5.3.2 Three Mg^{2+}-μc's (D1, D2 and D4) flank the PTC.

These Mg^{2+}-μc's are not involved in catalysis, and do not form the innermost layer of the PTC, but provide the framework and supporting structure for RNA that does. Mg^{2+}-μc's lend critical support to function by forming convoluted binding surfaces and providing rigid frameworks for attachment and buttressing of catalytic residues. No protein ligands penetrate the interior of the region of the magnesium clusters flanking the PTC, suggesting Mg^{2+}-μc ancestry prior to ribosomal proteins. The fourth (D3) is located near the exit site of the polypeptide exit tunnel. Previously Steitz and Moore (Hansen *et al.*, 2001; Klein *et al.*, 2004) showed that within the large subunit (LSU) of the ribosome, the concentration of Mg^{2+} ions is greatest near PTC. They also described one "magnesium cluster".

Mg^{2+}-μc's are highly conserved in evolution, and are found within the most conserved rRNA secondary structures (Figure 5.1C). Mg^{2+}-μc's link these conserved 2D elements via 'electrostatic tertiary interactions', which are composed of phosphate-Mg^{2+}-phosphate interactions. The 2D elements linked within Mg^{2+}-μc's are conserved between bacteria, archea, and eukarya (Cannone *et al.*, 2002) and mitochondrial rRNAs (Sharma *et al.*, 2003; Mears *et al.*, 2006), and in a proposed minimal 23S-rRNA (Mears *et al.*, 2002). The exception is Mg^{2+}-μc D3, which has been dispensed of in some mitoribosomes (such as that of C. elegans) by conversion of the RNA-based polypeptide exit tunnel to a protein-based tunnel (Sharma *et al.*, 2003).

5.3.3 Mg^{2+}-μc D2 binds to ribosomal protein L2.

Ribosome activity is exquisitely sensitive to mutations in ribosomal protein L2 (Cooperman *et al.*, 1995; Uhlein *et al.*, 1998; Diedrich *et al.*, 2000; Meskauskas *et al.*, 2008). The amino acid sequence of the N-terminal region of ribosomal protein L2 is among the most highly conserved in the phylogenic tree (Table 5.1). An 18 amino acid

loop of L2 (loop-L2/D2) forms a binding pocket for Mg^{2+}-μc D2. Ten amino acid residues of loop-L2/D2 are universally conserved in all cytoplasmic and chloroplast ribosomes and in fungal mitoribosomes. Mutations of other residues of loop-L2/D2 are infrequent and are between analogous amino acids such as aspartic acid and glutamic acid or between valine and threonine.

Prior to the availability of ribosome structures beginning in 2000, a direct role for L2 in catalysis appeared to be consistent with phylogenic, mutagenesis and biochemical data. But a catalytic role for L2 is ruled out by the realization that the ribosome is a ribozyme (Noller *et al.*, 1992; Nissen *et al.*, 2000; Yusupov *et al.*, 2001,). The strict sequence conservation of loop-L2/D2 is now seen to arise from a requirement for complementarity of the L2 protein surface with that of Mg^{2+}-μc D2 (Figures 5.2 and 5.3). Asparagine (**i**) and the carbonyl oxygen of alanine (**g**) bind to the first shell water molecules of the Mg^{2+} ions. Methionine (**h**) and valine (**c**) along with the backbone carbonyl of an alanine (**g**) form a pocket for adenosine 875^{Hm} (782^{Ec}). Histidine (**m**) forms part of the tightly packed core of loop-L2/D2. Because its mutation knocks out PTC activity, histidine (**m**) was previously thought to be part of a serine-protease like charge-relay system.

L2 sequence conservation is required in part by conformational conservation. The conformation of loop L2/D2 is highly conserved between *T. thermophilus* and *H. marismortui* (Figure 5.3). The RMSD of atomic positions of loop-L2/D2 is 0.6 Å (*H. marismortui* versus *T. thermophilus*, using all atoms of 18 residues except four differing sidechains, 110 atoms total).

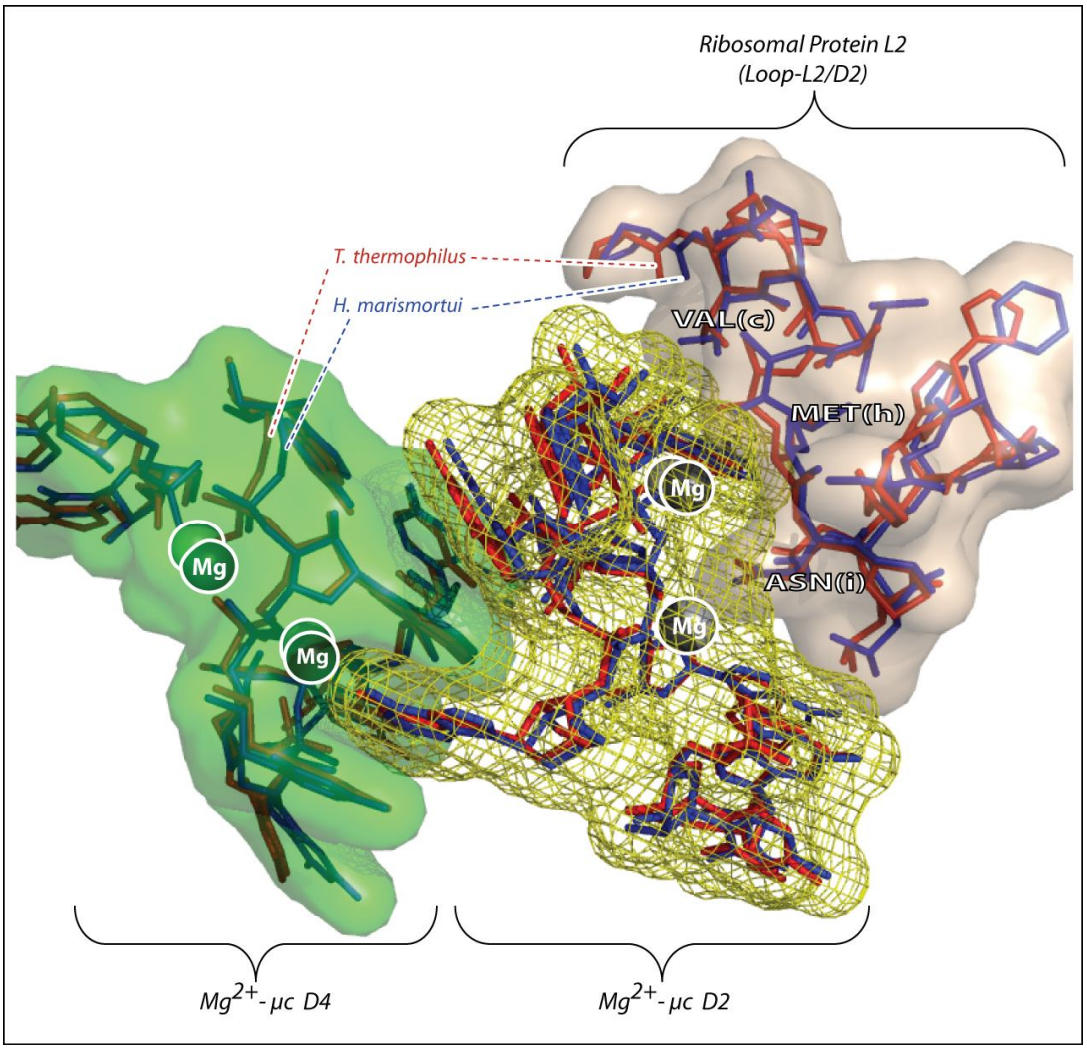

Figure 5.3 The complex formed by Mg^{2+}-µc's D4 and D2 and the loop-L2/D2 of ribosomal protein L2. Structures of both *H. marismortui* and *T. Thermophilus* are shown. Magnesium ions are indicated by spheres. When the 23S rRNAs of *H. marismortui* and *T. thermophilus* are superimposed, the RMSD of atomic positions of the eight Mg^{2+} ions within the four Mg^{2+}-µc's is very small, only 0.4 Å.

Mg^{2+}-µc's D1, D2 and D4 but not loop-L2/D2 are conserved in all mitoribosomes. Mitoribosomes have been substantially remodeled over time, as can be see by comparison of extant mitoribosomes with those of the ancestral endosymbiont (O'Brien, 2003; Sharma *et al.*, 2003; Smits *et al.*, 2007). Mitoribosomes have twice the protein and half the rRNA of the bacterial ribosome. Although rRNA secondary elements that contain Mg2+-µc's D1, D2 and D4 are conserved all mitoribosomes, loop-L2/D2 appears to be absent from mitoribosomes other than those of fungi (Table 5.1). In the human

mitoribosome protein L8 (the mammalian equivalent of L2), the N-terminus has been replaced by a sequence that diverges widely from loop-L2/D2 (Table 5.1). It may be that loop-L2/D2 has been structurally replaced by a nuclear encoded protein with unrelated sequence.

Table 5.1 Mg^{2+}-μc D2 Binding Loop of Ribosomal Protein L2 (loop-L2/D2)[a]

SPECIES	CODE	AMINO ACID SEQUENCE[b] NNNNNNNNNNNNNNNNabcdefghijklmnopqrNNNN[c]
Homo sapien	EAW82048.1	KAGRAYHKYKAKRNCW**PRVRGVAMNPVEHPFGGG**-NHQ
C. elegans	NP_507940.1	KAGRSYHKYKAKRNSW**PRVRGVAMNPVEHPHGGG**-NHQ
S. cerevisiae	P05736.3	KAGRAFHKYRLKRNSW**PKTRGVAMNPVDHPHGGG**-NHQ
H. marismortui[d]	AAA86862.1	KAGNKHHKMKARGTKW**PNVRGVAMNAVDHPFGGG**-GRQ
E. coli	BAE77974.1	KAGAARWRGVR-----**PTVRGTAMNPVDHPHGGG**EGRN
T. thermophilus[e]	AAS81667.1	KAGRSRWLGRR-----**PHVRGAAMNPVDHPHGGG**EGRA
A. thaliana-chloro[f]	NP_051123.1	RAGSKCWLGKR-----**PVVRGVVMNPVDHPHGGG**EGRA
S. cerevisiae-mito[f]	NP_010864.1	KAGRSRWLGIR-----**PTVRGVAMNKCDHPHGGG**RGKS
Homo sapien-mito[g]	NP_057034.2	KAGRNRWLGKR-----*PNSGRWHRKGGWAGRKIRPLPP*

a) The eukaryotic equivalent of L2 is L8. The mitochondrial equivalent of L2 is rml2. Blue text indicates conserved sequences in all ribosomes including mitoribosomes. Red text indicates conserved sequences in all cytoplasmic and chloroplast ribosomes and in fungal mitoribosomes, but not in other mitoribosomes. The ordering of this table was obtained from the complete L2/L8/rml2 sequence alignment with ClustlW (Larkin *et al.*, 2007).
b) Loop-L2/D2 is bold. Conserved residues are in red. Observed sequence changes of loop-L2/D2 are conservative.
c) Positions of loop-L2/D2 are defined by a-r.
d) Loop-L2/D2 contains ribosomal protein L2 residues 187-204 in *H. marismortui*.
e) Loop-L2/D2 contains L2 residues 219-226 in *T. thermophilus*.
f) Loop-L2/D2 contains rml2 residues 331-348 in *S. cerevisiae*. Mitochondrial and chloroplast ribosomes are thought to have undergone major remodeling (O'Brien, 2003; Sharma *et al.*, 2003; Smits *et al.*, 2007) and are the most divergent from other ribosomes.
g) The *Homo Sapien* mitoribosome lacks loop-L2/D2 as do other non-fungal mitoribosomes.

5.4 Discussion

Mg^{2+}-μc's by nature of their Mg^{2+}-(O1P-P-O2P)-Mg^{2+} linkages impose unusual constraints on RNA conformation and force de-stacking of bases. Mg^{2+}-μc's are unique structural entities with rigidity and forced dispositions of functional groups that cannot be approximated by RNA alone or by RNA in association with other ions. The bridging phosphate is, like all phosphates, restricted to tetrahedral geometry. The ligands of Mg^{2+}

ions are restricted to octahedral geometry. Therefore the core of each cluster is rigid and tightly packed. It can be seen that the RNA of Mg^{2+}-μc's form intricate and convoluted surfaces (Figure 5.3).

Mg^{2+}-μc's are observed in other ribozymes. A Mg^{2+}-μc, see appendix xxx, is observed in the P4-P6 domain of the tetrahymena Group I intron ribozyme (Cate *et al.*, 1997). An additional Mg^{2+}-μc was recently described in a Group II intron ribozyme (Toor *et al.*, 2008). A Mg^{2+}-μc in the 16S rRNA of *T. thermophilus* [PDB entry 1FJG (Carter *et al.*, 2000)], is disrupted upon ribosomal assembly [PDB entry 2J00 (Selmer *et al.*, 2006)]. The SARS s2m RNA described by Scott contains a pair of Mg^{2+} ions linked by a single phosphate (Robertson *et al.*, 2005), but those Mg^{2+} are otherwise fully hydrated, with no additional RNA ligands and so do not constitute a Mg^{2+}-μc.

The Mg^{2+}-μc's in the LSU, the Group I intron, and the Group II intron differ from the Mg^{2+} complexes proposed in the two-Mg^{2+} catalyzed phosphoryl-transfer mechanism (Beese and Steitz, 1991; Steitz and Steitz, 1993). In those complexes, a phosphate oxygen (not a phosphate group) bridges two Mg^{2+} ions, with a Mg^{2+}-Mg^{2+} distance of 3.9 Å. Reasonable coordination geometry in a Mg^{2+}-μc results in Mg^{2+}-Mg^{2+} distances of 5.3 to 5.6 Å (except for the doubly bridged D1 cluster, with a Mg^{2+}-Mg^{2+} distance of 4.7 Å).

Mg^{2+}-μc's facilitate RNA folding. The conformational space accessible to RNA is rather limited, and is driven by stacking interactions and double-strand formation. As noted by Noller (Noller, 2004), small molecules can extend the repertoire of RNA structures, and probably performed just that role during early evolution. Magnesium in particular can drive RNA into unusual conformation states (Klein *et al.*, 2004). Mg^{2+}-μc's demonstrate not only Mg^{2+}-driven deviation from canonical stacked conformations, but show how these altered states increase surface undulation, and facilitate highly specific interactions, such as those observed between Mg^{2+}-μc D2 and loop-L2/D2 and between Mg^{2+}-μc's D2 and D4 (Figure 5.3).

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

At atomic resolution (high resolution), large RNAs show a high level of complexity in conformation and in inter-and intra-molecular interactions. A given RNA motif with a conserved basic structure can be adorned by four types of DevLS, which are structural insertions, structural deletions, strand clips, and 3,2-switches. (see sections 2.3.4 and 2.4.2). Once DevLS are recognized the underlying structure is simpler and more tractable to analysis.

6.1 Multi-resolution Analysis of RNA Structure

I developed a multi-resolution technique for representing and analyzing RNA structures (Chapters 2 and 3). In this method RNA is viewed at various resolutions. At PBR resolution groups of RNA atoms (bases / riboses / phosphates / residues / motifs, etc.) are reduced to pseudo-objects, with locations and orientations. At PBR resolution (lower resolution), fundamental RNA architecture becomes readily observable (Figure 2.2, Chapter 2).

I used the multi-resolution approach for analysis of RNA tetraloop frequency and conservation between homologous structures. The method detects 44 tetraloops within the 23S rRNA of *H. marismortui*, and 59 tetraloops within the rRNA of *T. thermophilus*, with 33 within the large subunit (LSU) [Figure 3.1, Chapter 3], and 26 within the small subunit (SSU) [see appendix B, Figures B.1 and B.2]. All 103 dectected tetraloops within the 23S rRNAs of *T. thermophilus* and *H. marismortui* and the 16S rRNA of *T. thermophilus* are combined to build a summary tetraloop family tree (appendix B, figure B.3). This summary tetraloop family tree is a database showing populations, sequences, and structural relationships. One can see for example that nearly half of all tetraloops are

deletants. Combining all tetraloops in the summary tree gives more data points, allowing me to more confidently calculate sequence logos. However we do not believe the current summary tree to be fully unbiased, in that other types of globular RNAs may show different relative frequencies.

In future work I propose to use the multi-resolution approach to provide a method for cross-platform structural comparisons, e.g. protein versus RNA. Since groups of atoms are reduced into pseudo-objects, with locations and orientations, the structural description does not use real atoms (e.g. sidechains and peptide units in protein and, ribose and phosphate in RNA become pseudo-objects). Therefore the specific chemical identity of a molecule can be transcended. This method may make feasible structural comparisons and structural homology searching between molecules of different chemical identities.

6.2 RNA Structure in PBR Space

6.2.1 Tetraloops and Tetraloop Family Tree

I originated the tetraloop family tree, a formalism showing the relationship between tetraloops grouped by DevLS. Three mainly groups are leafed, which are s-Tl, d_2-Tl, and d_1i_0-Tl (UNCG tetraloop). By observing consensus molecular interactions, conformations, and topologies throughout tetraloop family tree, it can be seen that tetraloops are generally but not always terminal loops, i.e., not all tetraloops are loops, nor are they terminal.

In the s-Tl group, GNRA (where N can be any nucleotide and R is either G or A) is the most populated, as shown by the sequence logo diagram (Figures 2.4, 3.1 and B.2, Chapters 2 and 3, Appendix B). A set of consensus molecular interactions and conformation characterize the GNRA class of s-Tls (Figure 3.5, section 3.4.1): (i) there is a cross-strand stack between residue j – 1 (G) and residue j + 3; (ii) there is an unhappy A,

residue j + 2, unpaired (1 H bonding) and unstacked on the 3' site; (iii) there are three consensus intra-loop interactions which are *sugar to base* interaction: [O2' (residue j – 1, G) to N7 (residue j + 1, G)], *base to phosphate* interaction: [N1 and N2 (residue j – 1, G) to O2P (residue j + 2, A)], and *base to base* interaction: [N2 (residue j – 1, G) to N7 (residue j + 2, A)]; (iv) the backbone torsion angle α of residue j (N) goes to +165° from -65° of A-helix; (v) there is a three bases stack, N-R-A stack [residues (j)-(j+1)-(j+2) stack].

d_2-Tl, the second most populated tetraloop family, has the greatest conservation of sequence and conformation. From the results of structural mining, we hypothesize that d_2-Tl-like structures might be intermediates during folding of s-Tls. Solution experiments and simulations are currently in progress to test our hypothesis.

A recurrent four-tetraloop assembly of what we called "Tetraloop Triplet assembly" (Figure 3.8, section 3.4.3) within the *H. marismortui* [PDB entry: 1JJ2, ref. (Klein *et al.*, 2001)], *D. radiodurans* [PDB entry: 1NKW, ref. (Harms *et al.*, 2001)], and *T. thermophilus* [PDB entry: 2J01, ref. (Selmer *et al.*, 2006)] is observed. This four-tetraloop assembly consists of a tetraloop triplet with an extended short A-helix where is capped by a remote tetraloop. The tetraloop triplet assembly forms a continuous pseudo-A helix that is located on the surface of ribosome where it forms of part of the ribosomal exit tunnel. We hypothesize that the Tetraloop Triplet assembly plays an important role in the ribosome folding and stability.

6.2.2 Helical Junctions

With the PBR approach, helical junctions of RNA are readily observable. Within the 23S rRNA of *H. marismortui* and the 16S rRNA of *T. thermophilus,* I detect a total of 31 helical junctions (see appendix C, Table C.1). Two junctions are blunt, 26 are 3'(i) stacked junction [i indicates the number of 3' single-stranded stacked bases, i=1…6] (for an example 3'(1) stacked junction, see appendix C, Figure C.1). Three junctions are 5'(j)

stacked junction [j indicates number of 5' single strand stacked base, j=1||4]. Based on PBR analysis of helical junctions of RNA structures and on available thermodynamic information, we have proposed an atomic resolution reaction mechanism of RNA helix propagation, called the stack-ratchet (Mohan *et al.*, 2008). In the stack-ratchet, stacking of the 3' strand of the junction precedes base pairing. A 3' stack leads the single-strand / double-strand helix junction. We propose that, during RNA folding, double helices propagate via the stack-ratchet mechanism.

6.2.3 Kink-turns and E-loop motifs

In PBR space, A-helices, E-loop Motifs, Kink-Turns, etc., give distinctive fingerprints (see Chapter 2, Figure 2.2). In future work I propose to use PBR analysis to identify and re-define the Kink-turns and E-loop motifs. The tree formalism of these two RNA motifs will give an insight into structural conformation, consensus molecular interaction, sequence information, and structure variation with insertions, deletions, strand clips, and topology switches.

6.3 Structural Alignment by Anchored Segment

I developed an approach to structure-based alignment of RNA allowing 3D superimposition of very large RNA molecules. I use tetraloops as structural anchors in a '*divide and conquer*' strategy. During the '*divide*' steps, the RNA is split into segments, defined by tetraloops. During 'conquer' steps, DevLS locations are determined to optimize the structural alignment. I applied this method and describe the structure relationships between *H. marismortui* and *T. thermophilus*. The successful alignment and superimposition gives an overall RMSD of 1.2 Å utilizing 73% of RNA backbone atoms (around 2129 residues) [see Chapter 3].

By this approach, very accurate superimpositions of large RNAs of homologous structure can be obtained. An accurate superimposition allows one to identify regions of structural conservation and diversity, to determine relationships between structural and sequence variation, to dock elements of one structure onto another, and to study conformational change, binding, and mechanism in atomic resolution. I propose to use the finished superimposition (overall RMSD 1.2 Å) to model and explore the catalytic mechanism for peptidyl transferase activity.

Currently there are challenges in unraveling the catalytic mechanism. The main restraint is that currently available structures lack the acceptor arm of the A site tRNA. The acceptor arm of this tRNA is disordered and is not contained in the refined structure of the assembled *T. thermophilus* ribosome. To model the A site tRNA and the accommodation of tRNA at the A site, it will require more structural comparisons (superimpositions) between other ribosomal crystal structures in the structural database, and maybe performing directed molecular dynamic simulations through a series of crystal structures.

6.4 A Recurrent Magnesium-Binding Motif

There are a total of 117 Mg^{2+} ions identified in the LSU of *H. marismortui* (PDB entry: 1JJ2) and a total of 741 Mg^{2+} ions identified in the 70S of *T. thermophilus* (PDB entry: 2J00, 2J01). Using SAAS (see Chapter 3), the very accurate superimpositions of ribosomes has allowed me to elucidate the correspondence of magnesium ions. Using principles of inorganic chemistry along with structural alignment technique, I have identified a recurrent magnesium-binding motif, that we have now discovered is a general feature of large folded RNAs (Chapters 4 and 5). I observe that these magnesium-binding motifs play a critical role in the framework of the Peptidyl Transferase Center of the ribosome by their locations, topologies, and coordination geometries (Figures 4.5 and 5.1,

Chapters 4 and 5). The characteristics of the magnesium-binding motif are (i) a bridging phosphate chelating two magnesium ions in the form of $Mg^{2+}_{(i)}$-(O1P-P-O2P)-$Mg^{2+}_{(j)}$, (ii) 10-membered chelation rings, utilizing phosphate groups of adjacent RNA residues as Mg^{2+} ligands, (iii) crystalline-like Mg^{2+}-Mg^{2+} proximities, (iv) direct Mg^{2+}-phosphate interactions and Mg^{2+} dehydration, (v) undulated RNA surfaces with unpaired and unstacked bases, and (vi) and usually, close proximity to site of catalysis (Figure 5.2, Chapter 5).

APPENDIX A

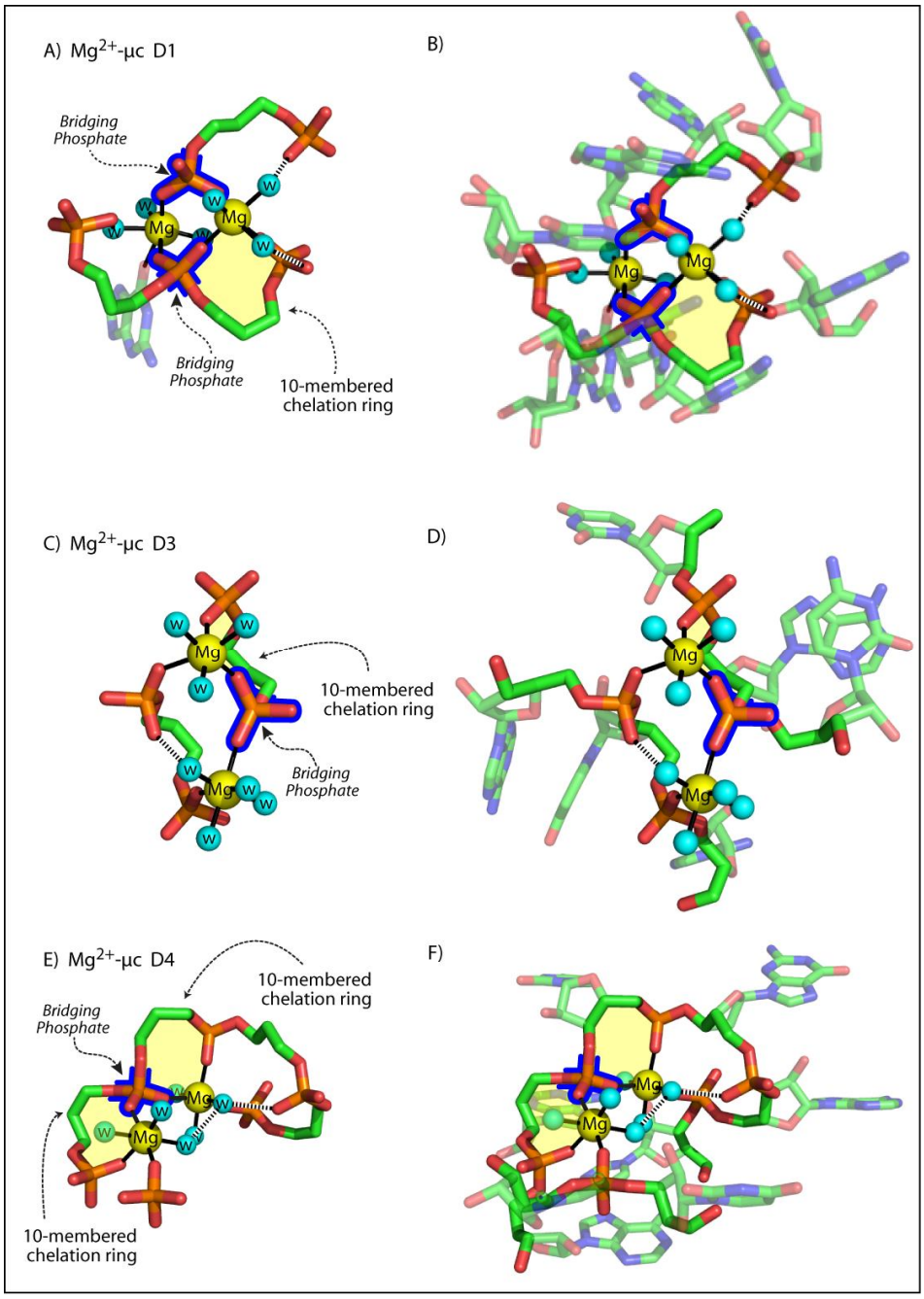

Figure A.1 Mg^{2+}-µc's from the 23s rRNA of *H. marismortu* (PDB entry 1JJ2).
Each Mg^{2+}-µc has at least one bridging phosphate (blue outline) and one 10-membered chelation ring (shaded yellow). (A) Mg2+-µc D1 with bases and oxyriboses omitted. This Mg^{2+}-µc has two Mg^{2+}-O1P-P-O2P-Mg^{2+} bridges. (Mg^{2+} ions 8001 and 8002.) (B) Mg^{2+}-µc D1 with bases and oxyriboses included. (C) Mg^{2+}-µc D3 with bases and oxyriboses omitted. (Mg^{2+} ions 8016 and 8029.) (D) Mg^{2+}-µc D3 with bases and oxyriboses included. (E) Mg^{2+}-µc D4 with bases and oxyriboses omitted. This Mg^{2+}-µc has two 10-membered chelation rings. (Mg^{2+} ions 8005 and 8007.) (F) Mg^{2+}-µc D4 shown with bases and oxyriboses included. Carbon is green, oxygen is red, phosphorous is orange and magnesium is yellow.

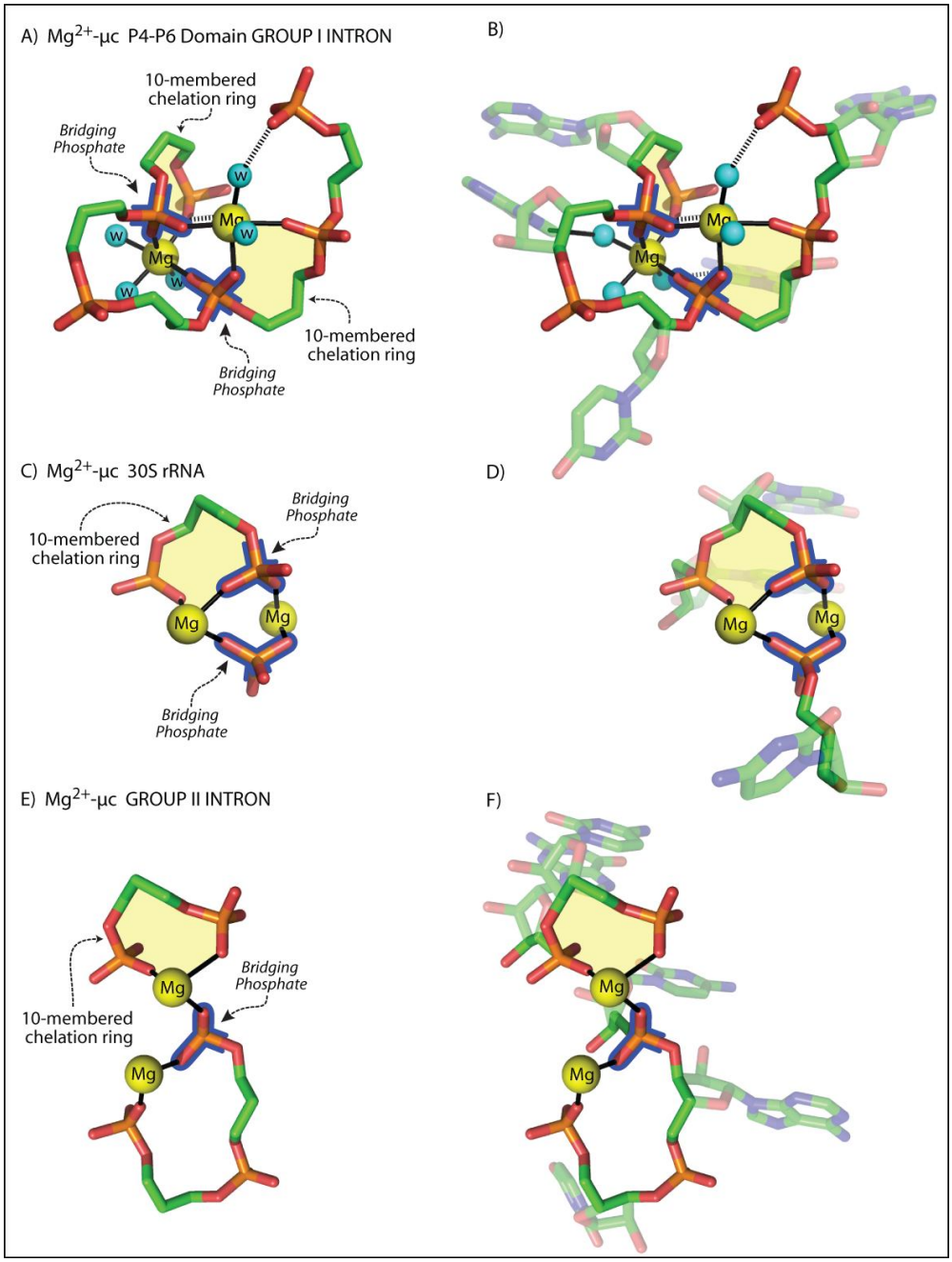

Figure A.2 Mg^{2+}-µc in among RNAs.
(A) Mg^{2+}-µc from the P4-P6 domain of the tetrahymena group I intron ribozyme (Mg^{2+} ions 2 and 4, PDB entry 1HR2). This Mg^{2+}-µc has two Mg^{2+}-O1P-P-O2P-Mg^{2+} bridges and two 10-membered chelation rings. Bases and oxyriboses omitted. (B) Bases and oxyriboses included. (C) Mg^{2+}-µc from the *Thermus thermophilus* 30S ribosomal subunit. This Mg^{2+}-µc has two Mg^{2+}-O1P-P-O2P-Mg^{2+} bridges. Water molecules are not contained in this model therefore second shell interactions are not defined. (Mg^{2+} ions 95 and 96, PDB entry 1FJG). (D) Bases and oxyriboses included. (E) Self-spliced group II intron from *Oceanobacillus iheyensis*. Water molecules are not contained in this model therefore second shell interactions are not defined. (Mg^{2+} ions 413 and 414, PDB entry 3BWP). (F) Bases and oxyriboses included.

APPENDIX B

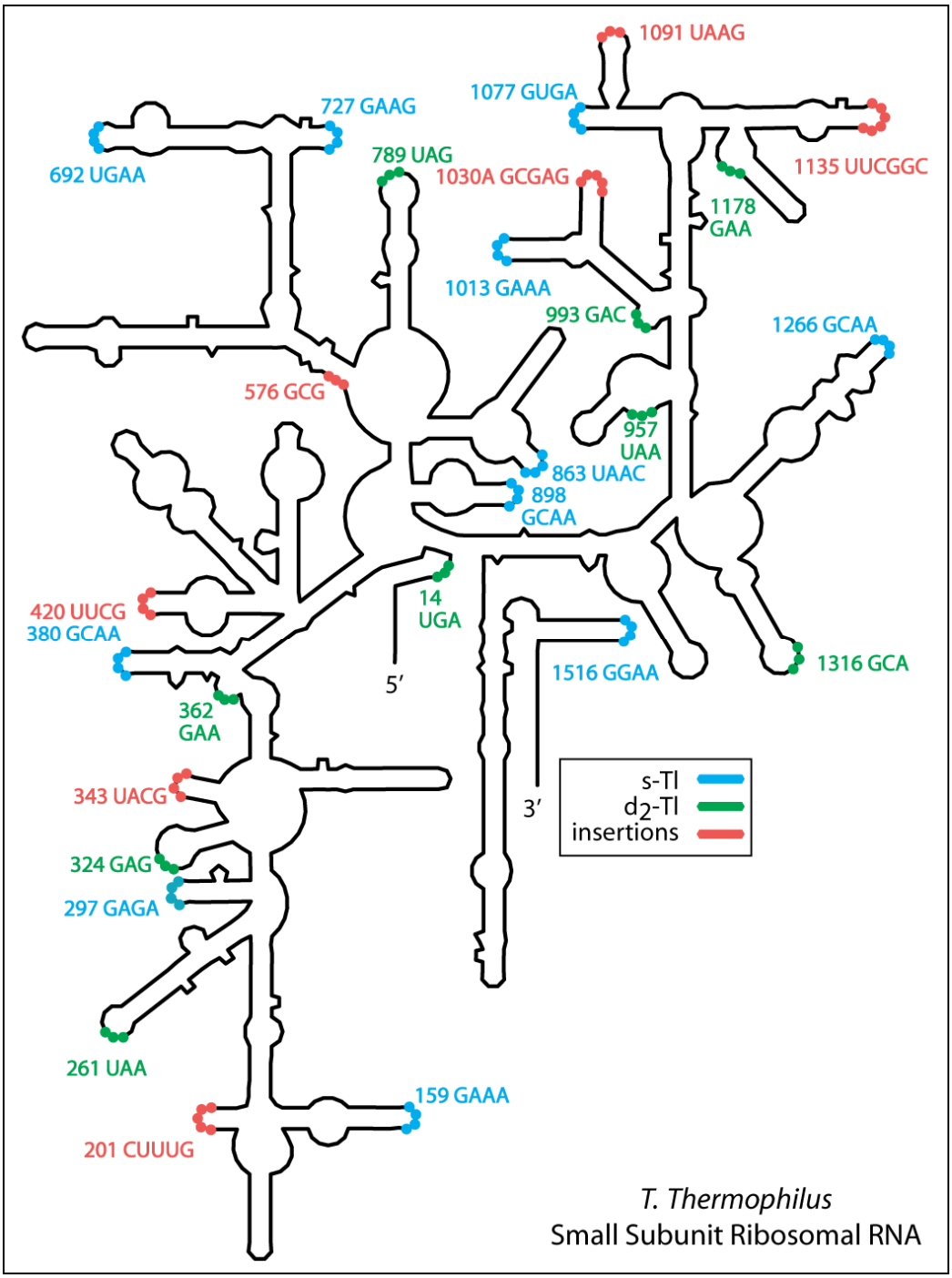

Figure B.1 Secondary structure of the *T. thermophilus* 16S rRNA (PDB entry: 2J00).
Tetraloop locations and type are indicated by color, standard tetraloop in cyan, deletions tetraloop in green, and insertions tetraloop in red.

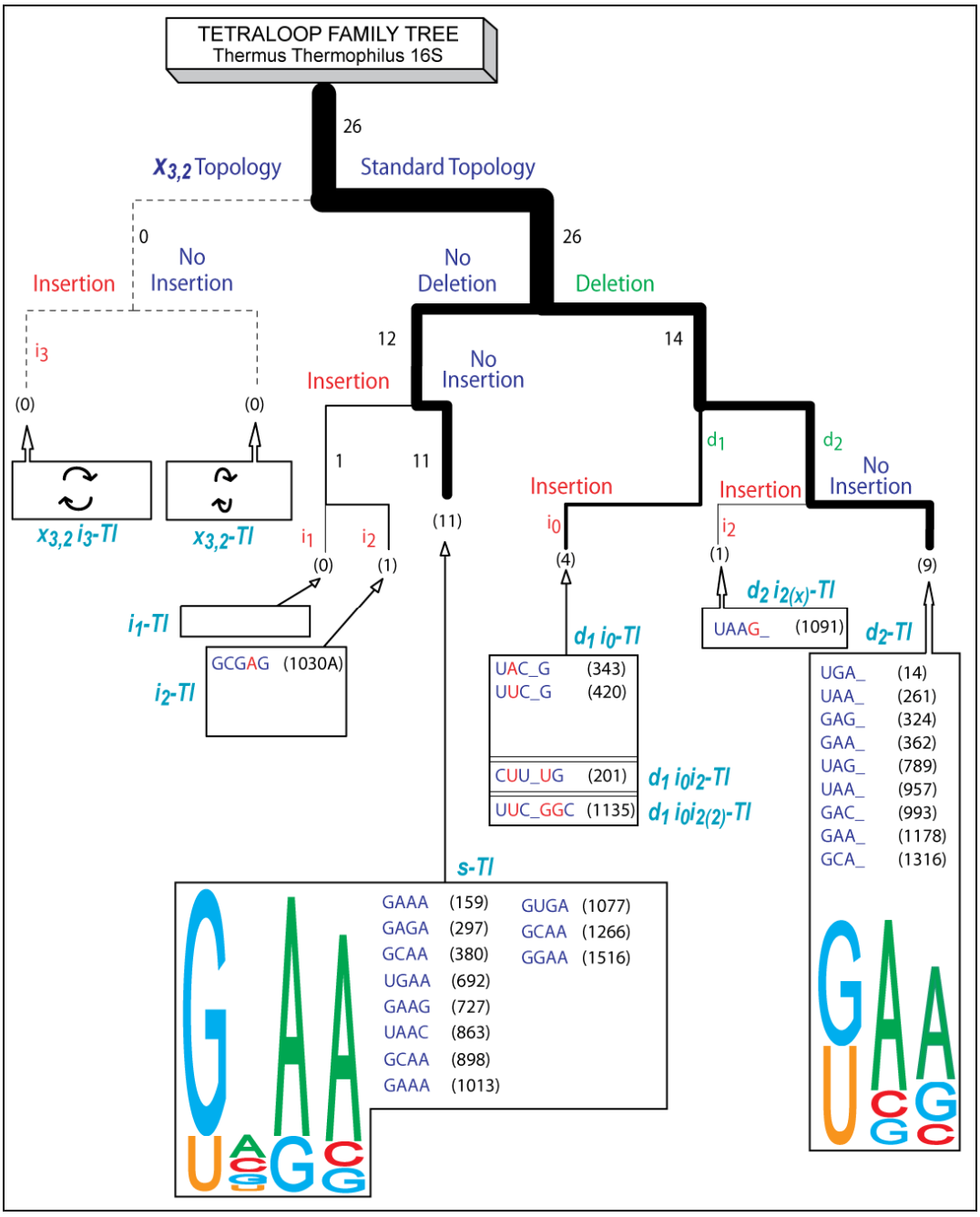

Figure B.2 Tetraloop family tree of the 16S rRNA of *T. Thermophilus*.
Twenty six tetraloops of 2J00 are distributed by type of DevLS. Insertion positions are indicated in red text. Deletion positions are indicated in green text. The positions of deleted residues are marked by underscores. Number of occurrences is indicated in black text, with line widths proportional to frequency. There are 5 groups (boxed) with contents, no 3,2-switch tetraloops were found. The residue number of the first residue and the sequence is given for each tetraloop. The consensus sequence for the s-Tl and d_2-Tl tetraloops are indicated by a sequence Logo representation (Schneider and Stephens, 1990).

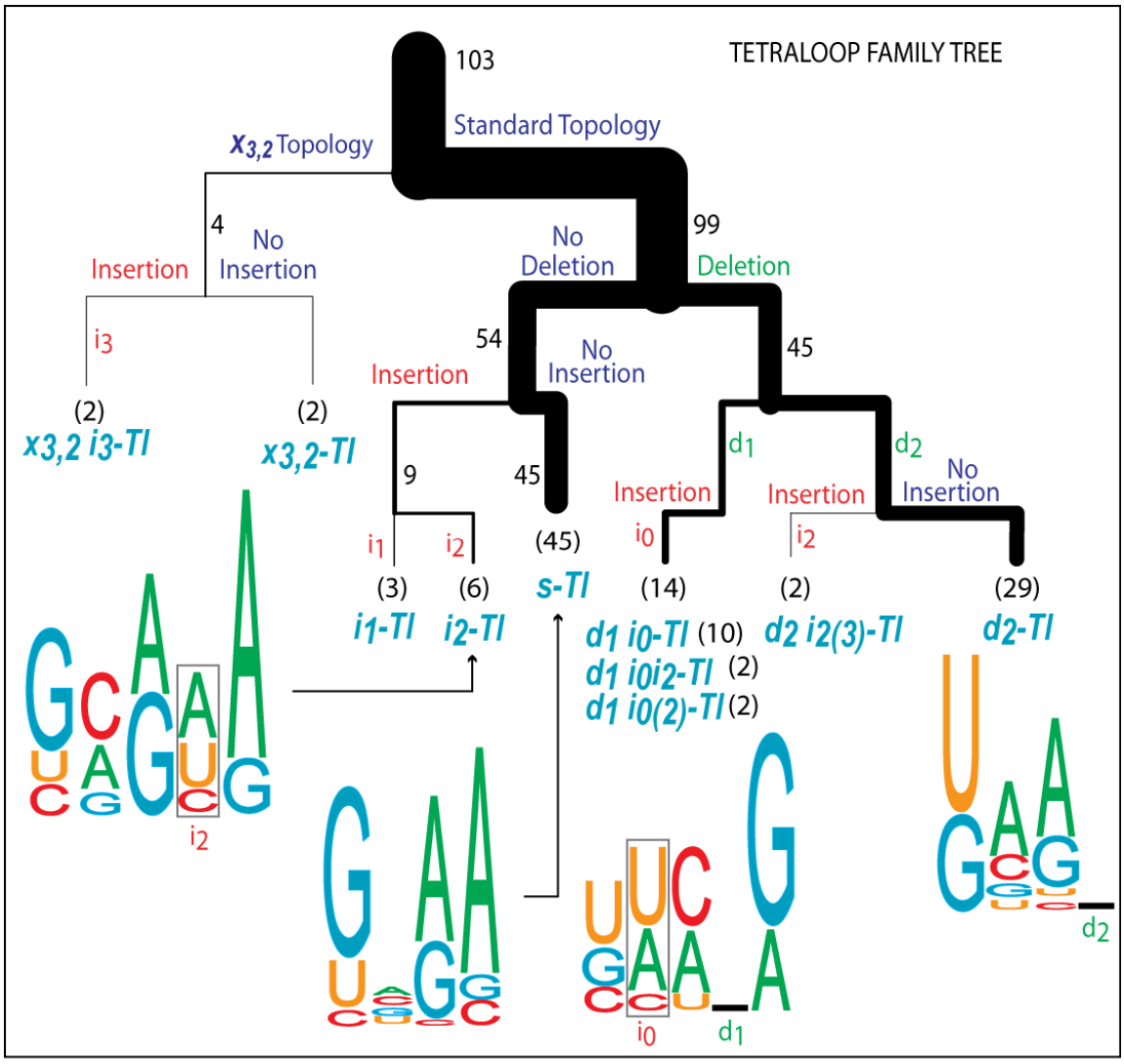

Figure B.3 The summary tetraloop family tree.
Shown is the summary tetraloop family tree that combines all 103 detected tetraloops within the 23S rRNAs of *H. marismortui* and *T. thermophilus* and the 16S rRNA of *T. thermophilus*. Tetraloops are distributed by types of DevLS. The statistic sequence bioinformation of s-Tl, d_2-Tl, d_1i_0-Tl and i_2-Tl are indicated by a sequence logo representation (Schneider and Stephens, 1990). The underscore in the sequence logo indicates a structural deletion sites. A gray box indicates a structural insertions sites. The number of occurrences is indicated in black, with line widths proportional to frequency. The thick line at the deletion branch indicates that structural deletions are common in tetraloops. The most populated tetraloop, s-Tl, shows the characteristic of sequence expected for GNRA type tetraloops.

APPENDIX C

Table C.1 Helical junctions within the 23S of *H. marismortui* and the 16S of *T. thermophilus.*

Helical Junctions	Type	Frequency	Total Number
Blunt	Blunt	2	**2**
3' stack	One residue: 3'(1)	9	**26**
	Two residue: 3'(2)	9	
	Three residue: 3'(3)	2	
	Four residue: 3'(4)	3	
	Five residue: 3'(5)	2	
	Six residue: 3'(6)	1	
5' stack	One residue: 5'(1)	1	**3**
	Four residue: 5'(4)	2	

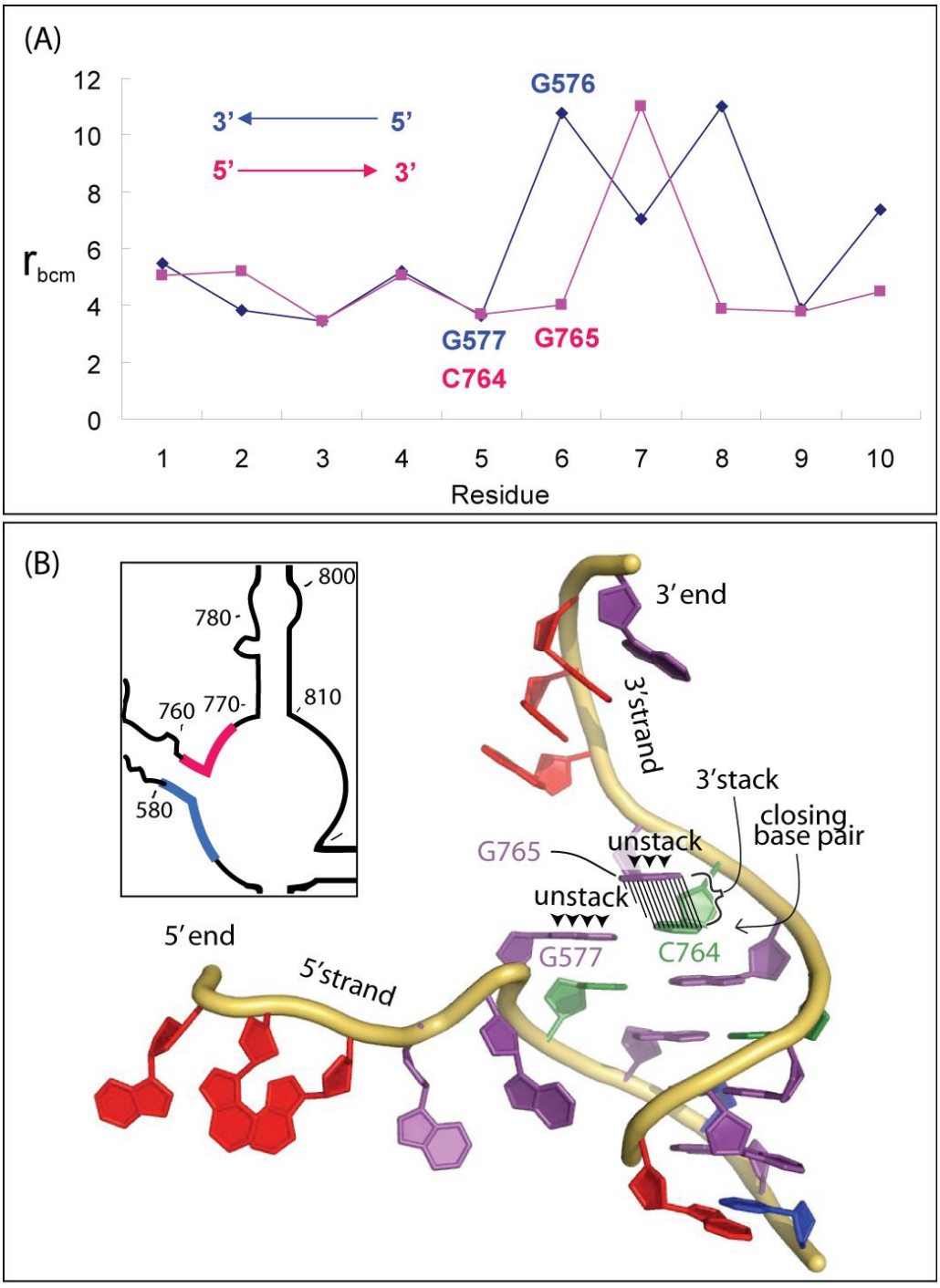

Figure C.1 Representative of 3'(1) stacked junction in PBR space and in 3D.
(A) Helical junction in PBR space. Base pair (G577:C764) is the closing base pair of A-helix. (B) A 3'(1) stacked junction in 3D. Left Panel shows the close up of the relevant regions of secondary structure, red: 3' strand, blue: 5' strand. Adenosine is red, guanosine is violet, uridine is blue and cytidine is green.

REFERENCES

Agris, P.F. (1996) The importance of being modified: roles of modified nucleosides and Mg2+ in RNA structure and function. *Prog Nucleic Acid Res Mol Biol*, 53**,** 79-129.

Akke, M., Fiala, R., Jiang, F., Patel, D., and Palmer, A.G., 3rd (1997) Base dynamics in a UUCG tetraloop RNA hairpin characterized by 15N spin relaxation: correlations with structure and stability. *Rna*, 3**,** 702-9.

Allain, F.H. and Varani, G. (1995) Structure of the P1 helix from group I self-splicing introns. *J Mol Biol*, 250**,** 333-53.

Allen, F.H., Harris, S.E., and Taylor, R. (1996) Comparison of conformer distributions in the crystalline state with conformational energies calculated by ab initio techniques. *J Comput Aided Mol Des*, 10**,** 247-54.

Allen, F.H., Mondal, R., Pitchford, N.A., and Howard, J.A.K. (2003) Mapping the geometry of metal three-coordination using crystal structure data: Reaction pathway for ligand addition to linear Hg-II species. *Helvetica Chimica Acta*, 86**,** 1129-1139.

Anderson, C.F. and Record, M.T., Jr. (1995) Salt-nucleic acid interactions. *Annu Rev Phys Chem*, 46**,** 657-700.

Antao, V.P. and Tinoco, I. (1992) Thermodynamic Parameters for Loop Formation in Rna and DNA Hairpin Tetraloops. *Nucleic Acids Research*, 20**,** 819-824.

Auffinger, P. and Westhof, E. (1999) Singly and bifurcated hydrogen-bonded base-pairs in tRNA anticodon hairpins and ribozymes. *J Mol Biol*, 292**,** 467-83.

Badger, J., Li, Y.L., and Caspar, D.L.D. (1994a) Thallium Counterion Distribution in Cubic Insulin Crystals Determined from Anomalous X-Ray-Diffraction Data. *Proceedings of the National Academy of Sciences of the United States of America*, 91**,** 1224-1228.

Badger, J., Kapulsky, A., Gursky, O., Bhyravbhatla, B., and Caspar, D.L.D. (1994b) Structure and Selectivity of a Monovalent Cation-Binding Site in Cubic Insulin Crystals. *Biophysical Journal*, 66**,** 286-292.

Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science*, 289**,** 905-20.

Bandyopadhyay, D. and Bhattacharyya, D. (2003) Different modes of interaction between hydrated magnesium ion and DNA functional groups: database analysis and ab initio studies. *J Biomol Struct Dyn*, 21**,** 447-58.

Basu, S., Szewczak, A.A., Cocco, M., and Strobel, S.A. (2000) Direct detection of monovalent metal ion binding to a DNA G-quartet by Tl-205 NMR. *Journal of the American Chemical Society*, 122**,** 3240-3241.

Basu, S., Rambo, R.P., Strauss-Soukup, J., Cate, J.H., Ferre-D'Amare, A.R., Strobel, S.A., and Doudna, J.A. (1998) A specific monovalent metal ion integral to the AA platform of the RNA tetraloop receptor. *Nat Struct Biol*, 5**,** 986-92.

Baumruk, V., Gouyette, C., Huynh-Dinh, T., Sun, J.S., and Ghomi, M. (2001) Comparison between CUUG and UUCG tetraloops: thermodynamic stability and structural features analyzed by UV absorption and vibrational spectroscopy. *Nucleic Acids Res*, 29**,** 4089-96.

Been, M.D. (2006) HDV ribozymes. *Curr Top Microbiol Immunol*, 307**,** 47-65.

Beese, L.S. and Steitz, T.A. (1991) Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *EMBO J*, 10**,** 25-33.

Berk, V., Zhang, W., Pai, R.D., and Cate, J.H. (2006) Structural basis for mRNA and tRNA positioning on the ribosome. *Proc Natl Acad Sci U S A*, 103**,** 15830-4.

Berman, H.M., Olson, W.K., Beveridge, D.L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.H., Srinivasan, A.R., and Schneider, B. (1992) The nucleic-acid database - a comprehensive relational database of 3-dimensional structures of nucleic-acids. *Biophysical Journal*, 63**,** 751-759.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) Protein data bank - computer-based archival file for macromolecular structures. *Journal of Molecular Biology*, 112**,** 535-542.

Betancourt, M.R. (2003) A reduced protein model with accurate native-structure identification ability. *Proteins*, 53**,** 889-907.

Black, C.B., Huang, H.W., and Cowan, J.A. (1994) Biological Coordination Chemistry of Magnesium, Sodium, and Potassium-Ions - Protein and Nucleotide-Binding Sites. *Coordination Chemistry Reviews*, 135**,** 165-202.

Bock, C.W., Kaufman, A., and Glusker, J.P. (1994) Coordination of Water to Magnesium Cations. *Inorganic Chemistry*, 33**,** 419-427.

Bock, C.W., Katz, A.K., Markham, G.D., and Glusker, J.P. (1999) Manganese as a replacement for magnesium and zinc: Functional comparison of the divalent ions. *Journal of the American Chemical Society*, 121**,** 7360-7372.

Bock, C.W., Markham, G.D., Katz, A.K., and Glusker, J.P. (2006) The arrangement of first- and second-shell water molecules around metal ions: effects of charge and size. *Theoretical Chemistry Accounts*, 115**,** 100-112.

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

Brannvall, M., Mikkelsen, N.E., and Kirsebom, L.A. (2001) Monitoring the structure of Escherichia coli RNase P RNA in the presence of various divalent metal ions. *Nucleic Acids Res*, 29**,** 1426-32.

Braunlin, W.H., Anderson, C.F., and Record, M.T., Jr. (1987) Competitive interactions of Co(NH3)6(3+) and Na+ with helical B-DNA probed by 59Co and 23Na NMR. *Biochemistry*, 26**,** 7724-31.

Brion, P. and Westhof, E. (1997) Hierarchy and dynamics of RNA folding. *Annu Rev Biophys Biomol Struct*, 26**,** 113-37.

Britten, J.S. and Blank, M. (1968) Thallium Activation of (Na+-K+)-Activated Atpase of Rabbit Kidney. *Biochimica Et Biophysica Acta*, 159**,** 160-&.

Brown, I.D. (1988) What Factors Determine Cation Coordination Numbers. *Acta Crystallographica Section B-Structural Science*, 44**,** 545-553.

Brown, I.D. (1992) Chemical and Steric Constraints in Inorganic Solids. *Acta Crystallographica Section B-Structural Science*, 48**,** 553-572.

Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K., and Neidle, S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res*, 34**,** 5402-15.

Burgi, H.B. (1973) Chemical Reaction Coordinates from Crystal-Structure Data .1. *Inorganic Chemistry*, 12**,** 2321-2325.

Burgi, H.B., Dunitz, J.D., and Shefter, E. (1973) Geometrical Reaction Coordinates .2. Nucleophilic Addition to a Carbonyl Group. *Journal of the American Chemical Society*, 95**,** 5065-5067.

Burgi, H.B., Dunitz, J.D., Lehn, J.M., and Wipff, G. (1974) Stereochemistry of Reaction Paths at Carbonyl Centers. *Tetrahedron*, 30**,** 1563-1572.

Burks, J., Zwieb, C., Muller, F., Wower, I., and Wower, J. (2005) Comparative 3-D modeling of tmRNA. *BMC Mol Biol*, 6**,** 14.

Butcher, S.E., Dieckmann, T., and Feigon, J. (1997) Solution structure of a GAAA tetraloop receptor RNA. *EMBO J*, 16**,** 7490-9.

Caceres, C., Wright, G., Gouyette, C., Parkinson, G., and Subirana, J.A. (2004) A thymine tetrad in d(TGGGGT) quadruplexes stabilized with Tl+/Na+ ions. *Nucleic Acids Research*, 32**,** 1097-1102.

Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Muller, K.M., Pande, N., Shang, Z., Yu, N., and Gutell, R.R. (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, 3**,** 2.

Carter, A.P., Clemons, W.M., Brodersen, D.E., Morgan-Warren, R.J., Wimberly, B.T., and Ramakrishnan, V. (2000) Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature*, 407**,** 340-8.

Cate, J.H., Hanna, R.L., and Doudna, J.A. (1997) A magnesium ion core at the heart of a ribozyme domain. *Nat Struct Biol*, 4**,** 553-8.

Cate, J.H., Yusupov, M.M., Yusupova, G.Z., Earnest, T.N., and Noller, H.F. (1999) X-ray crystal structures of 70S ribosome functional complexes. *Science*, 285**,** 2095-104.

Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R., and Doudna, J.A. (1996a) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, 273**,** 1678-85.

Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Szewczak, A.A., Kundrot, C.E., Cech, T.R., and Doudna, J.A. (1996b) RNA tertiary structure mediation by adenosine platforms. *Science*, 273**,** 1696-9.

Celander, D.W. and Cech, T.R. (1991) Visualizing the higher order folding of a catalytic RNA molecule. *Science*, 251**,** 401-7.

Cheong, C., Varani, G., and Tinoco, I., Jr. (1990) Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC. *Nature*, 346**,** 680-2.

Chworos, A., Severcan, I., Koyfman, A.Y., Weinkam, P., Oroudjev, E., Hansma, H.G., and Jaeger, L. (2004) Building programmable jigsaw puzzles with RNA. *Science*, 306**,** 2068-72.

Cole, P.E., Yang, S.K., and Crothers, D.M. (1972) Conformational changes of transfer ribonucleic acid. Equilibrium phase diagrams. *Biochemistry*, 11**,** 4358-68.

Conn, G.L., Gittis, A.G., Lattman, E.E., Misra, V.K., and Draper, D.E. (2002) A compact RNA tertiary structure contains a buried Backbone-K+ complex. *Journal of Molecular Biology*, 318**,** 963-973.

Cooperman, B.S., Wooten, T., Romero, D.P., and Traut, R.R. (1995) Histidine 229 in protein L2 is apparently essential for 50S peptidyl transferase activity. *Biochem Cell Biol*, 73**,** 1087-94.

Coppins, R.L., Hall, K.B., and Groisman, E.A. (2007) The intricate world of riboswitches. *Curr Opin Microbiol*, 10**,** 176-81.

Correll, C.C. and Swinger, K. (2003) Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 A resolution. *Rna*, 9**,** 355-63.

Correll, C.C., Beneken, J., Plantinga, M.J., Lubbers, M., and Chan, Y.L. (2003) The common and the distinctive features of the bulged-G motif based on a 1.04 A resolution RNA structure. *Nucleic Acids Res*, 31**,** 6806-18.

Cowan, J.A. (1993) Metallobiochemistry of RNA. Co(NH3)6(3+) as a probe for Mg2+(aq) binding sites. *J Inorg Biochem*, 49**,** 171-5.

Diedrich, G., Spahn, C.M., Stelzl, U., Schafer, M.A., Wooten, T., Bochkariov, D.E., Cooperman, B.S., Traut, R.R., and Nierhaus, K.H. (2000) Ribosomal protein L2 is involved in the association of the ribosomal subunits, tRNA binding to A and P sites and peptidyl transfer. *EMBO J*, 19**,** 5241-50.

Ding, Y. (2006) Statistical and Bayesian approaches to RNA secondary structure prediction. *Rna*, 12**,** 323-31.

Doyle, D.A., Morais Cabral, J., Pfuetzner, R.A., Kuo, A., Gulbis, J.M., Cohen, S.L., Chait, B.T., and MacKinnon, R. (1998) The structure of the potassium channel: molecular basis of K+ conduction and selectivity. *Science*, 280**,** 69-77.

Draper, D.E. (2004) A guide to ions and RNA structure. *RNA*, 10**,** 335-43.

Draper, D.E. and Misra, V.K. (1998) RNA shows its metal. *Nat Struct Biol*, 5**,** 927-30.

Draper, D.E., Grilley, D., and Soto, A.M. (2005) Ions and RNA folding. *Annu Rev Biophys Biomol Struct*, 34**,** 221-43.

Duarte, C.M. and Pyle, A.M. (1998) Stepping through an RNA structure: A novel approach to conformational analysis. *J Mol Biol*, 284**,** 1465-78.

Duarte, C.M., Wadley, L.M., and Pyle, A.M. (2003) RNA structure comparison, motif search and discovery using a reduced representation of RNA conformational space. *Nucleic Acids Res*, 31**,** 4755-61.

Durbin, R. (1998) Biological sequence analysis : probalistic models of proteins and nucleic acids.).xi, 356 p., Cambridge University Press, Cambridge, UK New York.

Eisinger, J., Shulman, R.G., and Szymanski, B.M. (1962) Transition Metal Binding in DNA Solutions. *Journal of Chemical Physics*, 36**,** 1721-1729.

Eisinger, J., Fawazest.F, and Shulman, R.G. (1965) BINDING OF MN2+ TO NUCLEIC ACIDS. *Journal of Chemical Physics*, 42**,** 43-&.

Elgavish, T., Cannone, J.J., Lee, J.C., Harvey, S.C., and Gutell, R.R. (2001) AA.AG@helix.ends: A:A and A:G base-pairs at the ends of 16 S and 23 S rRNA helices. *J Mol Biol*, 310**,** 735-53.

Ennifar, E., Nikulin, A., Tishchenko, S., Serganov, A., Nevskaya, N., Garber, M., Ehresmann, B., Ehresmann, C., Nikonov, S., and Dumas, P. (2000) The crystal structure of UUCG tetraloop. *J Mol Biol*, 304**,** 35-42.

Feig, A.L. (2000) The use of manganese as a probe for elucidating the role of magnesium ions in ribozymes. *Met Ions Biol Syst*, 37**,** 157-82.

Flory, P.J. (1953) Principles of polymer chemistry.).672 p., *George Fisher Baker non-resident lectureship in chemistry at Cornell University*, Cornell University Press, Ithaca,.

Fox, G.E. and Woese, C.R. (1975) The architecture of 5S rRNA and its relation to function. *J Mol Evol*, 6**,** 61-76.

Francois, J.C., Saison-Behmoaras, T., and Helene, C. (1988) Sequence-specific recognition of the major groove of DNA by oligodeoxynucleotides via triple helix formation. Footprinting studies. *Nucleic Acids Res*, 16**,** 11431-40.

Gardner, P.P. and Giegerich, R. (2004) A comprehensive comparison of comparative RNA structure prediction approaches. *BMC Bioinformatics*, 5**,** 140.

Gessner, R.V., Quigley, G.J., Wang, A.H., van der Marel, G.A., van Boom, J.H., and Rich, A. (1985) Structural basis for stabilization of Z-DNA by cobalt hexaammine and magnesium cations. *Biochemistry*, 24**,** 237-40.

Gilbert, D.E. and Feigon, J. (1999) Multistranded DNA structures. *Curr Opin Struct Biol*, 9**,** 305-14.

Gill, H.S. and Eisenberg, D. (2001) The crystal structure of phosphinothricin in the active site of glutamine synthetase illuminates the mechanism of enzymatic inhibition. *Biochemistry*, 40**,** 1903-1912.

Gill, M.L., Strobel, S.A., and Loria, J.P. (2006) Crystallization and characterization of the thallium form of the Oxytricha nova G-quadruplex. *Nucleic Acids Res*, 34**,** 4506-14.

Goertzen, L.R., Cannone, J.J., Gutell, R.R., and Jansen, R.K. (2003) ITS secondary structure derived from comparative analysis: implications for sequence alignment and phylogeny of the Asteraceae. *Molecular Phylogenetics and Evolution*, 29**,** 216-234.

Goody, T.A., Melcher, S.E., Norman, D.G., and Lilley, D.M. (2004) The kink-turn motif in RNA is dimorphic, and metal ion-dependent. *Rna*, 10**,** 254-64.

Gresh, N., Sponer, J.E., Spackova, N., Leszczynski, J., and Sponer, J. (2003) Theoretical study of binding of hydrated Zn(II) and Mg(II) cations to 5 '-guanosine monophosphate. Toward polarizable molecular mechanics for DNA and RNA. *Journal of Physical Chemistry B*, 107**,** 8669-8681.

Gutell, R.R., Noller, H.F., and Woese, C.R. (1986) Higher order structure in ribosomal RNA. *EMBO J*, 5**,** 1111-3.

Gutell, R.R., Lee, J.C., and Cannone, J.J. (2002) The accuracy of ribosomal RNA comparative structure models. *Curr Opin Struct Biol*, 12**,** 301-10.

Gutell, R.R., Cannone, J.J., Konings, D., and Gautheret, D. (2000) Predicting U-turns in ribosomal RNA with comparative sequence analysis. *J Mol Biol*, 300**,** 791-803.

Hansen, J.L., Schmeing, T.M., Klein, D.J., Ippolito, J.A., Ban, N., Nissen, P., Freeborn, B., Moore, P.B., and Steitz, T.A. (2001) Progress toward an understanding of the structure and enzymatic mechanism of the large ribosomal subunit. *Cold Spring Harb Symp Quant Biol*, 66**,** 33-42.

Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F., and Yonath, A. (2001) High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell*, 107**,** 679-88.

Hays, F.A., Teegarden, A., Jones, Z.J., Harms, M., Raup, D., Watson, J., Cavaliere, E., and Ho, P.S. (2005) How sequence defines structure: a crystallographic map of DNA structure and conformation. *Proc Natl Acad Sci U S A*, 102**,** 7157-62.

Hershkovitz, E., Sapiro, G., Tannenbaum, A., and Williams, L.D. (2006) Statistical analysis of RNA backbone. *IEEE/ACM Trans Comput Biol Bioinform*, 3**,** 33-46.

Hershkovitz, E., Tannenbaum, E., Howerton, S.B., Sheth, A., Tannenbaum, A., and Williams, L.D. (2003) Automated identification of RNA conformational motifs: theory and application to the HM LSU 23S rRNA. *Nucleic Acids Res*, 31**,** 6249-57.

Holley, R.W., Apgar, J., Everett, G.A., Madison, J.T., Marquisee, M., Merrill, S.H., Penswick, J.R., and Zamir, A. (1965) Structure of a Ribonucleic Acid. *Science*, 147**,** 1462-5.

Howerton, S.B., Nagpal, A., and Williams, L.D. (2003) Surprising roles of electrostatic interactions in DNA-ligand complexes. *Biopolymers*, 69**,** 87-99.

Howerton, S.B., Sines, C.C., VanDerveer, D., and Williams, L.D. (2001) Locating monovalent cations in the grooves of B-DNA. *Biochemistry*, 40**,** 10023-31.

Hsiao, C., Mohan, S., Hershkovitz, E., Tannenbaum, A., and Williams, L.D. (2006) Single nucleotide RNA choreography. *Nucleic Acids Res*, 34**,** 1481-91.

Huang, H.C., Nagaswamy, U., and Fox, G.E. (2005) The application of cluster analysis in the intercomparison of loop structures in RNA. *Rna*, 11**,** 412-23.

Hud, N.V. and Polak, M. (2001) DNA-cation interactions: The major and minor grooves are flexible ionophores. *Curr Opin Struct Biol*, 11**,** 293-301.

Hud, N.V., Schultze, P., and Feigon, J. (1998) Ammonium ion as an NMR probe for monovalent cation coordination sites of DNA quadruplexes. *Journal of the American Chemical Society*, 120**,** 6403-6404.

Hud, N.V., Sklenar, V., and Feigon, J. (1999a) Localization of ammonium lone in the minor groove of DNA duplexes in solution and the origin of DNA A-tract bending. *Journal of Molecular Biology*, 286**,** 651-660.

Hud, N.V., Smith, F.W., Anet, F.A., and Feigon, J. (1996) The selectivity for K+ versus Na+ in DNA quadruplexes is dominated by relative free energies of hydration: a thermodynamic analysis by 1H NMR. *Biochemistry*, 35**,** 15383-90.

Hud, N.V., Schultze, P., Sklenar, V., and Feigon, J. (1999b) Binding sites and dynamics of ammonium ions in a telomere repeat DNA quadruplex. *J Mol Biol*, 285**,** 233-43.

Inturris.Ce (1969a) Thallium Activation of K+-Activated Phosphatases from Beef Brain. *Biochimica Et Biophysica Acta*, 173**,** 567-&.

Inturris.Ce (1969b) Thallium-Induced Dephosphorylation of a Phosphorylated Intermediate of (Sodium + Thallium-Activated) Atpase. *Biochimica Et Biophysica Acta*, 178**,** 630-&.

Jack, A., Ladner, J.E., and Klug, A. (1976) Crystallographic refinement of yeast phenylalanine transfer RNA at 2-5A resolution. *J Mol Biol*, 108**,** 619-49.

Jaeger, L., Michel, F., and Westhof, E. (1994) Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. *J Mol Biol*, 236**,** 1271-6.

Jucker, F.M. and Pardi, A. (1995a) GNRA tetraloops make a U-turn. *Rna*, 1**,** 219-22.

Jucker, F.M. and Pardi, A. (1995b) Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif. *Biochemistry*, 34**,** 14416-27.

Jucker, F.M., Heus, H.A., Yip, P.F., Moors, E.H., and Pardi, A. (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J Mol Biol*, 264**,** 968-80.

Juneau, K., Podell, E., Harrington, D.J., and Cech, T.R. (2001) Structural basis of the enhanced stability of a mutant ribozyme domain and a detailed view of RNA--solvent interactions. *Structure*, 9**,** 221-31.

Kankia, B.I. (2003) Binding of Mg2+ to single-stranded polynucleotides: hydration and optical studies. *Biophys Chem*, 104**,** 643-54.

Kayne, F.J. (1971) Thallium (I) activation of pyruvate kinase. *Arch Biochem Biophys*, 143**,** 232-9.

Keefe, L.J., Sondek, J., Shortle, D., and Lattman, E.E. (1993) The alpha aneurism: a structural motif revealed in an insertion mutant of staphylococcal nuclease. *Proc Natl Acad Sci U S A*, 90**,** 3275-9.

Keefe, L.J., Quirk, S., Gittis, A., Sondek, J., and Lattman, E.E. (1994) Accommodation of insertion mutations on the surface and in the interior of staphylococcal nuclease. *Protein Sci*, 3**,** 391-401.

Khan, M.M. and Martell, A.E. (1966) Thermodynamic quantities associated with the interaction of adenosine triphosphate with metal ions. *J Am Chem Soc*, 88**,** 668-71.

Kim, S.H., Suddath, F.L., Quigley, G.J., McPherson, A., Sussman, J.L., Wang, A.H., Seeman, N.C., and Rich, A. (1974) Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science*, 185**,** 435-40.

Klein, D.J., Moore, P.B., and Steitz, T.A. (2004) The contribution of metal ions to the structural stability of the large ribosomal subunit. *Rna*, 10**,** 1366-79.

Klein, D.J., Schmeing, T.M., Moore, P.B., and Steitz, T.A. (2001) The kink-turn: a new RNA secondary structure motif. *Embo J*, 20**,** 4214-21.

Koculi, E., Thirumalai, D., and Woodson, S.A. (2006) Counterion charge density determines the position and plasticity of RNA folding transition states. *J Mol Biol*, 359**,** 446-54.

Komeda, S., Moulaei, T., Woods, K.K., Chikuma, M., Farrell, N.P., and Williams, L.D. (2006) A third mode of DNA binding: Phosphate clamps by a polynuclear platinum complex. *J Am Chem Soc*, 128**,** 16092-103.

Korostelev, A., Trakhanov, S., Laurberg, M., and Noller, H.F. (2006) Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell*, 126**,** 1065-77.

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982) Self-splicing RNA - auto-excision and auto-cyclization of the ribosomal-RNA intervening sequence of tetrahymena. *Cell*, 31**,** 147-157.

Lambert, A., Fontaine, J.F., Legendre, M., Leclerc, F., Permal, E., Major, F., Putzer, H., Delfour, O., Michot, B., and Gautheret, D. (2004) The ERPIN server: an interface to profile-based RNA motif identification. *Nucleic Acids Research*, 32**,** W160-W165.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, 23**,** 2947-8.

Larose, M., Gendron, P., and Major, F. (2005) MC-Search: a three-dimensional RNA pattern matching tool. *RNA Society Meeting*, **Poster #356**.

Latham, J.A. and Cech, T.R. (1989) Defining the inside and outside of a catalytic RNA molecule. *Science*, 245**,** 276-82.

Lee, J.C. and Gutell, R.R. (2004) Diversity of base-pair conformations and their occurrence in rRNA structure and RNA structural motifs. *J Mol Biol*, 344**,** 1225-49.

Lee, J.C., Cannone, J.J., and Gutell, R.R. (2003) The lonepair triloop: a new motif in RNA structure. *J Mol Biol*, 325**,** 65-83.

Lemieux, S. and Major, F. (2002) RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire. *Nucleic Acids Res*, 30**,** 4250-63.

Leontis, N.B. and Westhof, E. (1998) A common motif organizes the structure of multi-helix loops in 16 S and 23 S ribosomal RNAs. *J Mol Biol*, 283**,** 571-83.

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

Leontis, N.B. and Westhof, E. (2003) Analysis of RNA motifs. *Curr Opin Struct Biol*, 13**,** 300-8.

Leontis, N.B., Stombaugh, J., and Westhof, E. (2002a) The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res*, 30**,** 3497-531.

Leontis, N.B., Stombaugh, J., and Westhof, E. (2002b) Motif prediction in ribosomal RNAs Lessons and prospects for automated motif prediction in homologous RNA molecules. *Biochimie*, 84**,** 961-73.

Leroy, J.L., Gehring, K., Kettani, A., and Gueron, M. (1993) Acid multimers of oligodeoxycytidine strands: stoichiometry, base-pair characterization, and proton exchange properties. *Biochemistry*, 32**,** 6019-31.

Lescoute, A., Leontis, N.B., Massire, C., and Westhof, E. (2005) Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. *Nucleic Acids Res*, 33**,** 2395-409.

Levitt, M. (1969) Detailed molecular model for transfer ribonucleic acid. *Nature*, 224**,** 759-63.

Loria, J.P. and Nowak, T. (1998) Conformational changes in yeast pyruvate kinase studied by 205Tl+ NMR. *Biochemistry*, 37**,** 6967-74.

Magrum, L.J., Luehrsen, K.R., and Woese, C.R. (1978) Are extreme halophiles actually "bacteria"? *J Mol Evol*, 11**,** 1-8.

Mallat, S.G. (1999) A wavelet tour of signal processing. (2nd ed).xxiv, 637 p., Academic Press, San Diego, Calif.

Manners, J.P., Morallee, K.G., and Williams, R.J.P. (1970) Thallium(I) as a Potassium Probe in Biological Systems. *Journal of the Chemical Society D-Chemical Communications***,** 965-+.

Manning, G.S. (1978) The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q Rev Biophys*, 11**,** 179-246.

Markham, G.D., Glusker, J.P., and Bock, C.W. (2002) The arrangement of first- and second-sphere water molecules in divalent magnesium complexes: Results from molecular orbital and density functional theory and from structural crystallography. *Journal of Physical Chemistry B*, 106**,** 5118-5134.

Mathews, D.H. (2004) Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *Rna*, 10**,** 1178-90.

Mathews, D.H. and Turner, D.H. (2006) Prediction of RNA secondary structure by free energy minimization. *Curr Opin Struct Biol*, 16**,** 270-8.

Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M., and Turner, D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A*, 101**,** 7287-92.

Matsumura, S., Ikawa, Y., and Inoue, T. (2003) Biochemical characterization of the kink-turn RNA motif. *Nucleic Acids Res*, 31**,** 5544-51.

McFail-Isom, L., Sines, C.C., and Williams, L.D. (1999) DNA structure: cations in charge? *Curr Opin Struct Biol*, 9**,** 298-304.

Mears, J.A., Cannone, J.J., Stagg, S.M., Gutell, R.R., Agrawal, R.K., and Harvey, S.C. (2002) Modeling a minimal ribosome based on comparative sequence analysis. *J Mol Biol*, 321**,** 215-34.

Mears, J.A., Sharma, M.R., Gutell, R.R., McCook, A.S., Richardson, P.E., Caulfield, T.R., Agrawal, R.K., and Harvey, S.C. (2006) A structural model for the large subunit of the mammalian mitochondrial ribosome. *J Mol Biol*, 358**,** 193-212.

Meskauskas, A., Russ, J.R., and Dinman, J.D. (2008) Structure/function analysis of yeast ribosomal protein L2. *Nucleic Acids Research*, 36**,** 1826-1835.

Michel, F. and Westhof, E. (1990) Modeling of the 3-dimensional architecture of group-i catalytic introns based on comparative sequence-analysis. *Journal of Molecular Biology*, 216**,** 585-610.

Misra, V.K. and Draper, D.E. (2000) Mg(2+) binding to tRNA revisited: the nonlinear Poisson-Boltzmann model. *J Mol Biol*, 299**,** 813-25.

Mohan, S., Hsiao, C., Van Deusen, H., Gallagher, R., Krohn, E., Kalahar, B., Wartell, M.R., and Williams, D.L. (2008) Mechanism of RNA Double Helix-Propagation at Atomic Resolution**,** (submitted).

Monge, A., Lathrop, E.J., Gunn, J.R., Shenkin, P.S., and Friesner, R.A. (1995) Computer modeling of protein folding: conformational and energetic analysis of reduced and detailed protein models. *J Mol Biol*, 247**,** 995-1012.

Moore, P.B. (1999) Structural motifs in RNA. *Annu Rev Biochem*, 68**,** 287-300.

Moser, H.E. and Dervan, P.B. (1987) Sequence-specific cleavage of double helical DNA by triple helix formation. *Science*, 238**,** 645-50.

Moulaei, T., Maehigashi, T., Lountos, G.T., Komeda, S., Watkins, D., Stone, M.P., Marky, L.A., Li, J.S., Gold, B., and Williams, L.D. (2005) Structure of B-DNA with cations tethered in the major groove. *Biochemistry*, 44**,** 7458-7468.

Murray, L.J., Arendall, W.B., 3rd, Richardson, D.C., and Richardson, J.S. (2003) RNA backbone is rotameric. *Proc Natl Acad Sci U S A*, 100**,** 13904-9.

Murthy, V.L., Srinivasan, R., Draper, D.E., and Rose, G.D. (1999) A complete conformational map for RNA. *Journal of Molecular Biology*, 291**,** 313-327.

Nagaswamy, U. and Fox, G.E. (2002) Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. *Rna*, 8**,** 1112-9.

Neidle, S. and Parkinson, G.N. (2003) The structure of telomeric DNA. *Curr Opin Struct Biol*, 13**,** 275-83.

Nissen, P., Hansen, J., Ban, N., Moore, P.B., and Steitz, T.A. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science*, 289**,** 920-30.

Noller, H.F. (2004) The driving force for molecular evolution of translation. *RNA*, 10**,** 1833-7.

Noller, H.F., Hoffarth, V., and Zimniak, L. (1992) Unusual resistance of peptidyl transferase to protein extraction procedures. *Science*, 256**,** 1416-9.

Noller, H.F., Kop, J., Wheaton, V., Brosius, J., Gutell, R.R., Kopylov, A.M., Dohme, F., Herr, W., Stahl, D.A., Gupta, R., and Waese, C.R. (1981) Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Res*, 9**,** 6167-89.

O'Brien, T.W. (2003) Properties of human mitochondrial ribosomes. *IUBMB Life*, 55**,** 505-13.

Olsen, G.J. and Woese, C.R. (1997) Archaeal genomics: an overview. *Cell*, 89**,** 991-4.

Olson, W.K. (1975) Configurational statistics of polynucleotide chains. A single virtual bond treatment. *Macromolecules*, 8**,** 272-5.

Olson, W.K. (1981) Theoretical studies of nucleic acid conformation: potential energies, chain statistics, and model building. (Neidle, Stephen ed*Topics in nucleic acid structure*, Macmillan, London.

Onoa, B. and Tinoco, I., Jr. (2004) RNA folding and unfolding. *Curr Opin Struct Biol*, 14**,** 374-9.

Pedersen, P.A., Nielsen, J.M., Rasmussen, J.H., and Jorgensen, P.L. (1998) Contribution to Tl+, K+, and Na+ binding of Asn776, Ser775, Thr774, Thr772, and Tyr771 in cytoplasmic part of fifth transmembrane segment in alpha-subunit of renal Na,K-ATPase. *Biochemistry*, 37**,** 17818-27.

Petrov, A.S., Lamm, G., and Pack, G.R. (2002) Water-mediated magnesium-guanine interactions. *Journal of Physical Chemistry B*, 106**,** 3294-3300.

Petrov, A.S., Pack, G.R., and Lamm, G. (2004) Calculations of magnesium-nucleic acid site binding in solution. *Journal of Physical Chemistry B*, 108**,** 6072-6081.

Pley, H.W., Flaherty, K.M., and McKay, D.B. (1994) Three-dimensional structure of a hammerhead ribozyme. *Nature*, 372**,** 68-74.

Post, R.L., Kume, S., Tobin, T., Orcutt, B., and Sen, A.K. (1969) Flexibility of an Active Center in Sodium-Plus-Potassium Adenosine Triphosphatase. *Journal of General Physiology*, 54**,** S306-&.

Puglisi, J.D., Tan, R.Y., Calnan, B.J., Frankel, A.D., and Williamson, J.R. (1992) Conformation of the Tar Rna-Arginine Complex by Nmr-Spectroscopy. *Science*, 257**,** 76-80.

Pyle, A.M. (2002) Metal ions in the structure and function of RNA. *J Biol Inorg Chem*, 7**,** 679-90.

Quigley, G.J. and Rich, A. (1976) Structural domains of transfer RNA molecules. *Science*, 194**,** 796-806.

Rashin, A.A. and Honig, B. (1985) Reevaluation of the born model of ion hydration. *Journal of Physical Chemistry*, 89**,** 5588-5593.

Record, M.T., Jr., Anderson, C.F., and Lohman, T.M. (1978) Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Q Rev Biophys*, 11**,** 103-78.

Record, M.T., Jr., Zhang, W., and Anderson, C.F. (1998) Analysis of effects of salts and uncharged solutes on protein and nucleic acid equilibria and processes: a practical guide to recognizing and interpreting polyelectrolyte effects, Hofmeister effects, and osmotic effects of salts. *Adv Protein Chem*, 51**,** 281-353.

Reuben, J. and Cohn, M. (1970) Magnetic resonance studies of manganese (II) binding sites of pyruvate kinase. Temperature effects and frequency dependence of proton relaxation rates of water. *J Biol Chem*, 245**,** 6539-46.

Reuben, J. and Kayne, F.J. (1971) Thallium-205 Nuclear Magnetic Resonance Study of Pyruvate Kinase and Its Substrates - Evidence for a Substrate-Induced Conformational Change. *Journal of Biological Chemistry*, 246**,** 6227-&.

Richardson, J.S., Getzoff, E.D., and Richardson, D.C. (1978) BETA-BULGE - COMMON SMALL UNIT OF NONREPETITIVE PROTEIN-STRUCTURE. *Proceedings of the National Academy of Sciences of the United States of America*, 75**,** 2574-2578.

Richardson, J.S., Schneider, B., Murray, L.W., Kapral, G.J., Immormino, R.M., Headd, J.J., Richardson, D.C., Ham, D., Hershkovitz, E., Williams, L.D., Keating, K.S., Pyle, A.M., Micallef, D., Westbrook, J., and Berman, H.M. (2008) RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA*, 14**,** 465-81.

Robertson, M.P., Igel, H., Baertsch, R., Haussler, D., Ares, M., Jr., and Scott, W.G. (2005) The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol*, 3**,** e5.

Rodriguez-Cruz, S.E., Jockusch, R.A., and Williams, E.R. (1998) Hydration energies of divalent metal ions, Ca2+(H2O)(n) (n=5-7) and Ni2+(H2O)(n) (n=6-8), obtained by blackbody infrared radiative dissociation. *Journal of the American Chemical Society*, 120**,** 5842-5843.

Rudisser, S. and Tinoco, I., Jr. (2000) Solution structure of Cobalt(III)hexammine complexed to the GAAA tetraloop, and metal-ion binding to G.A mismatches. *J Mol Biol*, 295**,** 1211-23.

Rulisek, L. and Sponer, J. (2003) Outer-shell and inner-shell coordination of phosphate group to hydrated metal ions (Mg2+, Cu,2+Zn2+, Cd2+) in the presence and absence of nucleobase. The role of nonelectrostatic effects. *Journal of Physical Chemistry B*, 107**,** 1913-1923.

Saenger, W. (1984) Principles of nucleic acid structure.).xx, 556 p., Springer-Verlag, New York.

Salgado, P.S., Walsh, M.A., Laurila, M.R., Stuart, D.I., and Grimes, J.M. (2005) Going soft and SAD with manganese. *Acta Crystallogr D Biol Crystallogr*, 61**,** 108-11.

Schmeing, T.M., Huang, K.S., Strobel, S.A., and Steitz, T.A. (2005) An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature*, 438**,** 520-4.

Schneider, T.D. and Stephens, R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res*, 18**,** 6097-100.

Scott, W.G. and Klug, A. (1996) Ribozymes: structure and mechanism in RNA catalysis. *Trends Biochem Sci*, 21**,** 220-4.

Selinger, D., Liao, X.B., and Wise, J.A. (1993) Functional Interchangeability of the Structurally Similar Tetranucleotide Loops Gaaa and Uucg in Fission Yeast Signal

Recognition Particle Rna. *Proceedings of the National Academy of Sciences of the United States of America*, 90**,** 5409-5413.

Selmer, M., Dunham, C.M., Murphy, F.V.t., Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R., and Ramakrishnan, V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, 313**,** 1935-42.

Sen, D. and Crothers, D.M. (1986) Condensation of chromatin: role of multivalent cations. *Biochemistry*, 25**,** 1495-503.

Shafer, R.H. (1998) Stability and structure of model DNA triplexes and quadruplexes and their interactions with small ligands. *Prog Nucleic Acid Res Mol Biol*, 59**,** 55-94.

Shapiro, B.A., Yingling, Y.G., Kasprzak, W., and Bindewald, E. (2007) Bridging the gap in RNA structure prediction. *Curr Opin Struct Biol*, 17**,** 157-65.

Sharma, M.R., Koc, E.C., Datta, P.P., Booth, T.M., Spremulli, L.L., and Agrawal, R.K. (2003) Structure of the mammalian mitochondrial ribosome reveals an expanded functional role for its component proteins. *Cell*, 115**,** 97-108.

Sharp, K.A. and Honig, B. (1990) Electrostatic interactions in macromolecules: theory and applications. *Annu Rev Biophys Biophys Chem*, 19**,** 301-32.

Shui, X., McFail-Isom, L., Hu, G.G., and Williams, L.D. (1998) The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry*, 37**,** 8341-55.

Sines, C.C., McFail-Isom, L., Howerton, S.B., VanDerveer, D., and Williams, L.D. (2000) Cations mediate B-DNA conformational heterogeneity. *Journal of the American Chemical Society*, 122**,** 11048-11056.

Smits, P., Smeitink, J.A., van den Heuvel, L.P., Huynen, M.A., and Ettema, T.J. (2007) Reconstructing the evolution of the mitochondrial ribosomal proteome. *Nucleic Acids Res*, 35**,** 4686-703.

Spahn, C.M., Gomez-Lorenzo, M.G., Grassucci, R.A., Jorgensen, R., Andersen, G.R., Beckmann, R., Penczek, P.A., Ballesta, J.P., and Frank, J. (2004) Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *Embo J*, 23**,** 1008-19.

Sponer, J., Leszczynski, J., and Hobza, P. (2001) Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases. *Biopolymers*, 61**,** 3-31.

Srinivasan, A.R. and Olson, W.K. (1980) Yeast transfer rna-phe conformation wheels - a novel probe of the monoclinic and orthorhombic models. *Nucleic Acids Research*, 8**,** 2307-2329.

Stefan, L.R., Zhang, R., Levitan, A.G., Hendrix, D.K., Brenner, S.E., and Holbrook, S.R. (2006) MeRNA: a database of metal ion binding sites in RNA structures. *Nucleic Acids Research*, 34**,** D131-D134.

Steitz, T.A. and Steitz, J.A. (1993) A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A*, 90**,** 6498-502.

Subirana, J.A. and Soler-Lopez, M. (2003) Cations as hydrogen bond donors: a view of electrostatic interactions in DNA. *Annu Rev Biophys Biomol Struct*, 32**,** 27-45.

Sundaralingam, M. (1969) Stereochemistry of nucleic acids and their constituents .IV. allowed and preferred conformations of nucleosides, nucleoside mono-, di-, tri-, tetraphosphates, nucleic acids and polynucleotides. *Biopolymers*, 7**,** 821-860.

Sundaralingam, M. (1973) Conformation of biological molecules and polymers, in *The Jerusalem Symp. on Quant. Chem. Biochem.* 5. (Bergmann, E.D. & Pullman, B. ed).pp 417-456, Academic Press, New York.

Sundaralingam, M. and Sekharudu, Y.C. (1989) Water-inserted alpha-helical segments implicate reverse turns as folding intermediates. *Science*, 244**,** 1333-7.

Swaminathan, V. and Sundaralingam, M. (1979) The crystal structures of metal complexes of nucleic acids and their constituents. *CRC Crit Rev Biochem*, 6**,** 245-336.

Szewczak, A.A. and Moore, P.B. (1995) The sarcin/ricin loop, a modular RNA. *J Mol Biol*, 247**,** 81-98.

Takamoto, K., Das, R., He, Q., Doniach, S., Brenowitz, M., Herschlag, D., and Chance, M.R. (2004) Principles of RNA compaction: insights from the equilibrium folding pathway of the P4-P6 RNA domain in monovalent cations. *J Mol Biol*, 343**,** 1195-206.

Tamura, M., Hendrix, D.K., Klosterman, P.S., Schimmelman, N.R., Brenner, S.E., and Holbrook, S.R. (2004) SCOR: Structural Classification of RNA, version 2.0. *Nucleic Acids Res*, 32**,** D182-4.

Taylor, R. (2002) Life-science applications of the Cambridge Structural Database. *Acta Crystallogr D Biol Crystallogr*, 58**,** 879-88.

Taylor, R., Kennard, O., and Versichel, W. (1983) Geometry of the N-H=/=C Hydrogen-Bond .1. Lone-Pair Directionality. *Journal of the American Chemical Society*, 105**,** 5761-5766.

Taylor, R., Kennard, O., and Versichel, W. (1984) Geometry of the N-H-O=C Hydrogen-Bond .2. 3-Center (Bifurcated) and 4-Center (Trifurcated) Bonds. *Journal of the American Chemical Society*, 106**,** 244-248.

Tereshko, V., Minasov, G., and Egli, M. (1999) A "hydrat-ion" spine in a B-DNA minor groove. *Journal of the American Chemical Society*, 121**,** 3590-3595.

Tereshko, V., Wilds, C.J., Minasov, G., Prakash, T.P., Maier, M.A., Howard, A., Wawrzak, Z., Manoharan, M., and Egli, M. (2001) Detection of alkali metal ions in DNA crystals using state-of-the-art X-ray diffraction experiments. *Nucleic Acids Research*, 29**,** 1208-1215.

Tinoco, I., Jr. and Bustamante, C. (1999) How RNA folds. *J Mol Biol*, 293**,** 271-81.

Toor, N., Keating, K.S., Taylor, S.D., and Pyle, A.M. (2008) Crystal structure of a self-spliced group II intron. *Science*, 320**,** 77-82.

Tuerk, C., Gauss, P., Thermes, C., Groebe, D.R., Gayle, M., Guild, N., Stormo, G., d'Aubenton-Carafa, Y., Uhlenbeck, O.C., Tinoco, I., Jr., and et al. (1988) CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc Natl Acad Sci U S A*, 85**,** 1364-8.

Uhlein, M., Weglohner, W., Urlaub, H., and Wittmann-Liebold, B. (1998) Functional implications of ribosomal protein L2 in protein biosynthesis as shown by in vivo replacement studies. *Biochem J*, 331 (Pt 2)**,** 423-30.

Vallurupalli, P. and Moore, P.B. (2003) The solution structure of the loop E region of the 5S rRNA from spinach chloroplasts. *J Mol Biol*, 325**,** 843-56.

Varani, G., Cheong, C.J., and Tinoco, I. (1991) Structure of an Unusually Stable Rna Hairpin. *Biochemistry*, 30**,** 3280-3289.

Vargason, J.M., Henderson, K., and Ho, P.S. (2001) A crystallographic map of the transition from B-DNA to A-DNA. *Proc Natl Acad Sci U S A*, 98**,** 7265-70.

Villeret, V., Huang, S., Fromm, H.J., and Lipscomb, W.N. (1995) Crystallographic evidence for the action of potassium, thallium, and lithium ions on fructose-1,6-bisphosphatase. *Proc Natl Acad Sci U S A*, 92**,** 8916-20.

Voss, N.R., Gerstein, M., Steitz, T.A., and Moore, P.B. (2006) The geometry of the ribosomal polypeptide exit tunnel. *J Mol Biol*, 360**,** 893-906.

Waugh, A., Gendron, P., Altman, R., Brown, J.W., Case, D., Gautheret, D., Harvey, S.C., Leontis, N., Westbrook, J., Westhof, E., Zuker, M., and Major, F. (2002) RNAML: a standard syntax for exchanging RNA information. *RNA*, 8**,** 707-17.

Westheimer, F.H. (1987) Why nature chose phosphates. *Science*, 235**,** 1173-8.

Westhof, E., Masquida, B., and Jaeger, L., 1998: Molecular modeling of nucleic acids. 346-358.

Williams, D.J. and Hall, K.B. (1999) Unrestrained stochastic dynamics simulations of the UUCG tetraloop using an implicit solvation model. *Biophys J*, 76**,** 3192-205.

Williams, L.D. (2005) Between objectivity and whim: Nucleic acid structural biology. *DNA Binders and Related Subjects*, 253**,** 77-88.

Williams, L.D. and Maher, L.J., 3rd (2000) Electrostatic mechanisms of DNA deformation. *Annu Rev Biophys Biomol Struct*, 29**,** 497-521.

Williams, R.J.P. (1971) Biochemistry of Group Ia and Iia Cations. *Advances in Chemistry Series***,** 155-+.

Williams, R.J.P. (1991) The Chemical-Elements of Life. *Journal of the Chemical Society-Dalton Transactions***,** 539-546.

Williamson, J.R., Raghuraman, M.K., and Cech, T.R. (1989) Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell*, 59**,** 871-80.

Wimberly, B., Varani, G., and Tinoco, I., Jr. (1993) The conformation of loop E of eukaryotic 5S ribosomal RNA. *Biochemistry*, 32**,** 1078-87.

Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Jr., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit. *Nature*, 407**,** 327-39.

Winkler, W.C. and Breaker, R.R. (2005) Regulation of bacterial gene expression by riboswitches. *Annu Rev Microbiol*, 59**,** 487-517.

Woese, C.R. (1987) Bacterial evolution. *Microbiol Rev*, 51**,** 221-71.

Woese, C.R. and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A*, 74**,** 5088-90.

Woese, C.R. and Gutell, R.R. (1989) Evidence for several higher order structural elements in ribosomal RNA. *Proc Natl Acad Sci U S A*, 86**,** 3119-22.

Woese, C.R., Magrum, L.J., and Fox, G.E. (1978) Archaebacteria. *J Mol Evol*, 11**,** 245-51.

Woese, C.R., Winker, S., and Gutell, R.R. (1990) Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops". *Proc Natl Acad Sci U S A*, 87**,** 8467-71.

Woese, C.R., Gutell, R., Gupta, R., and Noller, H.F. (1983) Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol Rev*, 47**,** 621-69.

Woese, C.R., Magrum, L.J., Gupta, R., Siegel, R.B., Stahl, D.A., Kop, J., Crawford, N., Brosius, J., Gutell, R., Hogan, J.J., and Noller, H.F. (1980) Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Res*, 8**,** 2275-93.

Yang, H., Jossinet, F., Leontis, N., Chen, L., Westbrook, J., Berman, H., and Westhof, E. (2003) Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res*, 31**,** 3450-60.

Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H., and Noller, H.F. (2001) Crystal structure of the ribosome at 5.5 A resolution. *Science*, 292**,** 883-96.

Zuker, M. (2000) Calculating nucleic acid secondary structure. *Curr Opin Struct Biol*, 10**,** 303-10.

VITA

CHIAOLONG HSIAO

Chiaolong Hsiao was born in Taipei, Taiwan. In 1998 he received a B.Sc. in Pharmacy from Taipei Medical University where he worked in the organic chemistry laboratory of Dr. Hsien-Saw Kuo. In 2000 he received an M.Sc. in Medicinal Chemistry with thesis entitled "Stereoselective Synthesis of 1',3'-Disubstituted Spiro[7-hydroxybenzofuran-3(2H),4'-piperidines] as Novel Opioids" from National Taiwan University, where he worked the laboratory of Dr. Chen-Yu Cheng. In 2003 he obtained a M.Sc. degree in Systems and Information Science from Syracuse University, Syracuse, NY. In 2003 he entered the Georgia Institute of Technology to pursue a doctorate in Bioinformatics. Chiaolong joined Dr. Loren Williams' group in the Fall of 2004. Chiaolong's research work on structural analysis of DNA grooves hydration and DNA-protein interactions has been published in, [Watkins, D., Hsiao, C., Woods, K. K., Koudelka, G. B., and Williams, L. D. "P22 c2 Repressor-Operator Complex: Mechanisms of Direct and Indirect Readout" Biochemistry, 47, 2325-2338 (2008)].