CHAPTER 1

Complexes of Nucleic Acids with Group I and II Cations

CHIAOLONG HSIAO,[a] EMMANUEL TANNENBAUM,[b] HALENA VANDEUSEN,[a] ELI HERSHKOVITZ,[c, d] GINGER PERNG,[c] ALLEN R. TANNENBAUM[c, d] AND LOREN DEAN WILLIAMS[a]

[a] School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332-0400, USA; [b] School of Biology, Georgia Institute of Technology, Atlanta, GA, 30332, USA; [c] School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332–0250, USA; [d] School of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA 30332–0250, USA

1.1 Introduction

Recent structures of large RNAs, such as the P4–P6 domain of the *Tetrahymena* ribozyme,[1–5] and larger RNAs such as rRNAs,[6–14] 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

RSC Biomolecular Sciences
Nucleic Acid–Metal Ion Interactions
Edited by Nicholas V. Hud

Published by the Royal Society of Chemistry, www.rsc.org

the origins of the specific requirement for Mg^{2+} in RNA folding and the tight coupling between Mg^{2+} binding and RNA conformation. We deconstruct the two binding-mode formalism. 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,[15–17] in polyelectrolyte theory[18–22] (see Chapter 9), in DNA structure[22–28] (see Chapter 3), in RNA folding[29–32] (see Chapters 6 and 7) and in RNA catalysis[33–36] (see Chapter 8).

1.1.1 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*[7,37] (2.4 Å resolution, PDB entry 1JJ2), a halophile from the Dead Sea. 23S-rRNA^{TT} refers to the 23S rRNA from the eubacterium *T. thermophilus*[13] (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$^{P4–P6}$ refers the 160 nucleotide domain of the self-splicing *Tetrahymena thermophila* intron (2.3 Å resolution, PDB entry 1HR2, this ΔC209 mutant[5] gives the best available resolution).

1.2 Nucleic Acid Folding

1.2.1 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 side-chains are multifarious, with a variety of shapes and chemical properties. The nucleic acid backbone is intricate, with many accessible rotameric states,[38] and carries charge.

Functional nucleic acids generally fold into compact and stable states of given conformation.[39,40] DNA can form quadruplexes,[41–46] triplexes,[47–49] i-motifs,[50,51] *etc*. Structured RNAs range in size from aptamers and tRNAs to ribosomes. However some functional nucleic acids, such as riboswitches, are conformationally polymorphic.[52,53] For our purposes, folded nucleic acid structures fall into three general classes: (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,[54,55] and quadruplexes,[41–46] and (iii) *true globular structures* such as the *Tetrahymena* ribozyme[40] and its P4–P6 domain,[1–5] 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 complementarity with phosphate, with a specific coordination and thermodynamic fingerprint. These states are simply not accessible in the absence of Mg^{2+} (or Mn^{2+}), even when other cations are at high concentration. The thermodynamic and conformational consequences of first-shell OP interactions with Mg^{2+} are different to those for neutral ligands or for other cations, with lesser charge or greater size.

1.2.2 The RNA Folding Hierarchy

RNA folding is hierarchical.[56,57] Folding progresses through a series of intermediates that 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.[58,59] Although compact structures with base–base tertiary interactions can be achieved at 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 make 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 a useful but somewhat limited approximation. In true globular structures, ground states are stabilized by specifically associated Mg^{2+} ions, each with up to four first-shell OP ligands. The importance of $Mg^{2+}(OP)_3$ and $Mg^{2+}(OP)_4$ 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 the 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+}.[60] We do not know at present if the converse is true (*i.e.* are electrostatic tertiary interactions fully dependent on correct base–base tertiary interactions?). At any rate, to understand and describe fully the structure of a globular RNA, one can extend a conventional tertiary description of base–base interactions to include electrostatic tertiary interactions.

1.2.3 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 >7 Å (P to P) in random coil to around 5.8–6.2 Å (in A-form helical regions and loops). In subsequent steps, a subset of P to P distances decreases further, to 5.0–4.6 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 ($d_{OP-OP} = 2.8$–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.

1.3 Coordination Chemistry

The binding of ligands to Group I and II cations is dictated by the chemical properties of the cations and of the ligands and, to a significant extent, by interactions between ligands.[15] 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 of assembling the ligands for joint coordination increases. Hud and Polak previously noted the chelation properties of DNA, calling it an ionophore.[27] Here we illustrate how the phosphate groups of nucleic acids commonly act as chelators of cations.

1.3.1 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.[29,61]

The monovalent cations [sodium (Na^+), potassium (K^+), rubidium (Rb^+), caesium (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 1.1). The coordination chemistry of Li^+, with its

Table 1.1 Physical properties of cations.

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

[a]From Brown.[65]
[b]The radius of a non-spherical species such as NH_4^+ is not well defined. Rashin and Honig[160] estimated from geometric considerations that the effective radius of NH_4^+ is very nearly the same as the radius of K^+.
[c]From Rashin and Honig.[160]
[d]Average observed coordination numbers, from Brown.[65]
[e]Not reported.

small atomic radius and high charge density, is distinct from that of other Group I 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.[42,62,63] 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 1.1).

These monovalent ions, except Li^+, display irregular and variable coordination geometry.[64,65] 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 10. These properties are quantitated by 'average observed coordination numbers' (AOCN, Table 1.1) over a large number of structures within the Cambridge Structural Database as reported by Brown.[65]

1.3.1.1 Na^+

The ideal Na^+ to oxygen distance is 2.4 Å (Figure 1.1a). The distance between first-shell ligands of Na^+ is variable, depending on coordination number and coordination geometry. An octahedral arrangement of first shell oxygen 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. An Na^+ ion with ideal octahedral geometry in association with the O6 position of a guanine of DNA with five water molecule inner-shell ligands[66] is shown in Figure 1.1a. As can be seen, the Na^+ to O distances average around 2.4 Å, whereas the distance between *cis* oxygen atoms (adjacent oxygen ligands) averages around 3.4 Å. DNA–cation interactions are discussed in Chapter 3.

1.3.1.2 K^+

The ideal K^+ to oxygen distance is around 2.7 Å. For an octahedral arrangement of first-shell oxygen ligands, the average O to O distance is over 4.0 Å.

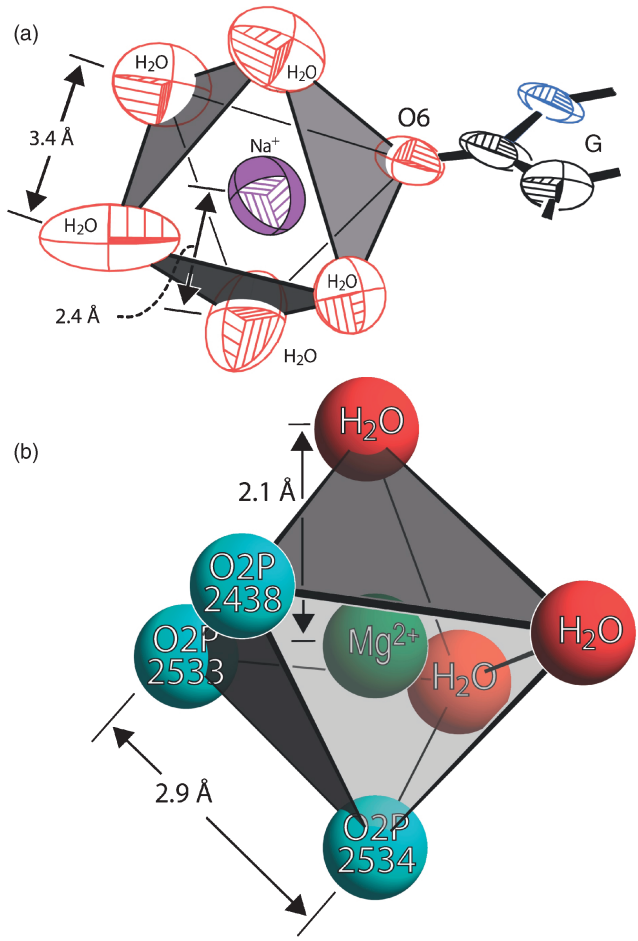

Figure 1.1 (a) An Na^+ ion bound to the O6 of a guanine of DNA (ion 487 from PDB entry 2DYW). The Na^+ ion (purple ellipsoid) is octahedral, with bonds to the floor of the major groove and five water molecules (hydrogens not shown). All first-shell oxygen ligands are represented by red ellipsoids. Na^+–oxygen bonds are around 2.4 Å in length. The first shell ligands are not in 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 first shell. For clarity, the radii of the spheres are reduced from the van der Waals radii of the atoms and have no physical significance.

Hence the inner-shell ligands of K^+ are of less defined geometry than those of Na^+. Specific K^+ binders, that exclude Na^+, are generally composed of stacked, planar arrangements of keto oxygens.[41,67,68] 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.

1.3.2 Group II

1.3.2.1 Ca^{2+}

Ca^{2+} often shows irregular coordination geometry and coordination numbers greater than six. The ionic radius of Ca^{2+} is *ca.* 0.99 Å. Compared with Mg^{2+}, Ca^{2+} is large, its charge density is low, its inner-shell ligands are loosely packed and the magnitude of is hydration enthalpy is small (Table 1.1). However, like Mg^{2+}, Ca^{2+} prefers a mix of anionic and neutral ligands.

1.3.2.2 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.[69] Mg^{2+} shares a special geometric, electrostatic and thermodynamic relationship with phosphates and phosphate esters. In comparison with group I ions, Ca^{2+} or polyamines, Mg^{2+} has a 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 (*ca.* 0.65 Å), the charge density is high, the six ligands of an octahedral inner first shell are tightly packed and the magnitude of the hydration enthalpy is large (Table 1.1). The heat of hydration of Mg^{2+} is much greater than for other biological cations (Table 1.1). 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+} assumes octahedral geometry,[64,65,70,71] as shown in Figure 1.1b. The AOCN of Mg^{2+} is 5.98.[65] 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.[72]

1.3.2.2.1 Hexaaqua Mg^{2+} Complexes. In hexaaqua 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···O distance is 2.93 Å.[73–75] 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 towards 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.

1.3.2.2.2 ADP–Mg^{2+} Complexes. The nature of complexes formed by Mg^{2+} with ADP and ATP are highly predictive of Mg^{2+} interactions with other multidentate OP ligands such as RNA. Some of the ADP–Mg^{2+} and ATP–Mg^{2+} complexes identified in the Protein Data Bank are shown in Figure 1.2 and here we discuss the implications of these and related structures provided by X-ray crystallography. One can immediately appreciate 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.

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; Figure 1.2a and b, Table 1.2 and Table 1.3) or by bidentate chelation (60%, Figure 1.2c–e). Monodentate interactions occur exclusively by $O^{\beta}P$ whereas bidentate chelation involves $O^{\alpha}P$ and $O^{\beta}P$.

Chelation ring size is an important factor in modulating stability (see Section 1.3). The bidentate chelation complexes of Mg^{2+} with ADP/ATP are composed of six-membered rings consisting of atoms Mg^{2+}–$O^{\alpha}P$–P–O–P–$O^{\beta}P$–Mg^{2+} or (Mg^{2+}–$O^{\beta}P$–P–O–P–$O^{\gamma}P$–Mg^{2+}). Bidentate chelation by two OP atoms bound to a common phosphorus atom would require a chelation ring size of four and is not observed in ADP/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 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 180°) are not observed and appear to be stereochemically prohibited.

1.3.2.2.3 ATP–Mg^{2+} Complexes. Mg^{2+} binds to ATP by bidentate (66%, Figure 1.2f–i) or tridentate (33%, Figure 1.2j) interactions. Monodentate ATP–Mg^{2+} complexes are not observed. The bidentate complexes are all *cis*, whereas 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+}–$O^{\alpha}P$–P–O–P–$O^{\beta}P$–Mg^{2+}) or (Mg^{2+}–$O^{\beta}P$–P–O–P–$O^{\gamma}P$–Mg^{2+}). Bidentate complexes most commonly involve $O^{\beta}P$ and $O^{\gamma}P$ (88%). One $O^{\alpha}P$–$O^{\beta}P$ bidentate complex is observed.

An $O^{\alpha}P$–Mg^{2+}–$O^{\gamma}P$ bidentate complex would form an eight-membered ring. The importance of ring size in chelation complexes is underscored by the absence of $O^{\alpha}P$–Mg^{2+}–$O^{\gamma}P$ bidentate complexes in the Protein Data Bank. The observed bidentate complexes contain 0–3 protein ligands (Table 1.2), which are all oxygens, with no obvious preference in the number or charge of protein ligands. First-shell protein ligands are not observed in tridentate ATP complexes.

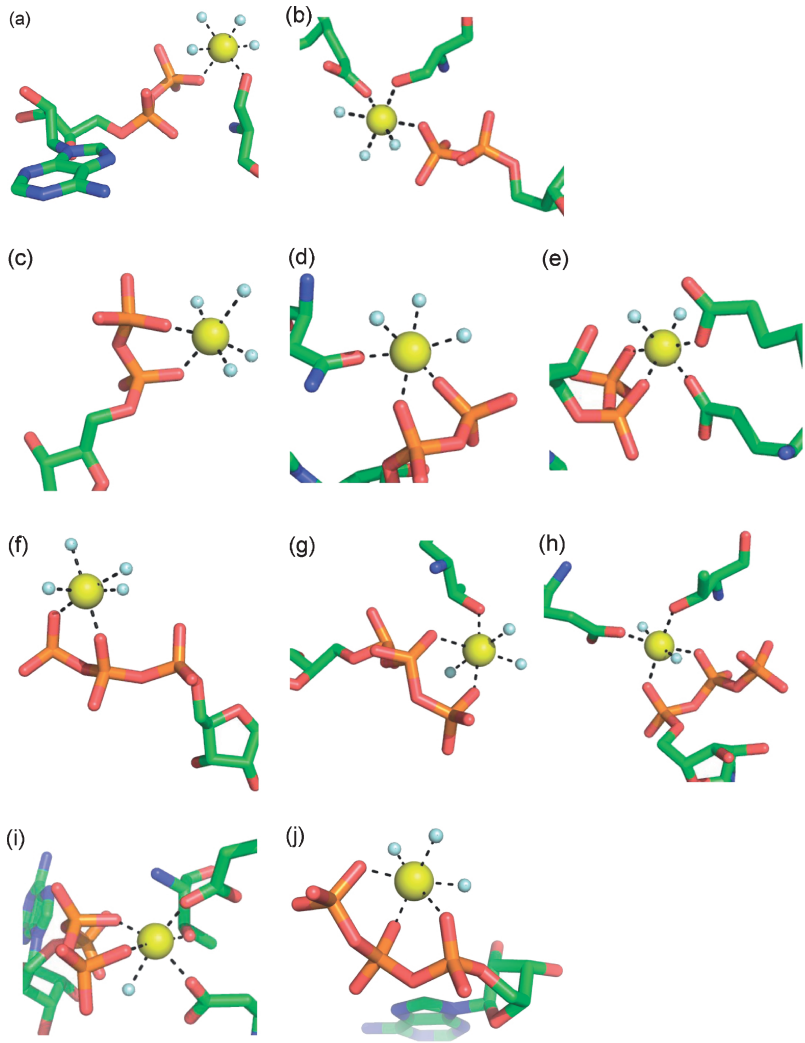

Figure 1.2 ADP and ATP as Mg^{2+} chelators. Mg^{2+} ions are represented by yellow spheres and water molecules by small white spheres. Nucleotides and protein ligands are represented by sticks. (a) Monodentate chelation by ADP ($O^{\beta}P$) with one protein ligand. (b) Monodentate chelation by ADP 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 with 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 ($O^{\beta}P$ and $O^{\gamma}P$). (h) Bidentate chelation by ATP plus three protein ligands ($O^{\alpha}P$ and $O^{\beta}P$). (i) Bidentate chelation by ATP ($O^{\alpha}P$ and $O^{\beta}P$) with two protein ligands. (j) Tridentate chelation by ATP. The OP ligands are *cis*–*cis*. All structures shown were extracted from the Protein Data Bank.

Table 1.2 Mg^{2+} interactions with ADP and ATP.[a]

PDB entry	*No. of first-shell OP atoms*	*Total first-shell ligands*[b]	*Protein ligands*	*Orientation*	*Resolution/Å*	*Corresponding 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 2.25 Å or other bad geometries 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 four-coordinate, presumably because two water molecules were omitted during refinement.

Table 1.3 Mg^{2+} chelation by nucleotides[a] and by RNA.[b]

	No. of first-shell OP ligands			
	0 (hexaaqua)	*1 (monodentate)*	*2 (bidentate)*	*3 (tridentate)*
ADP	NA	4	6	0
ATP	NA	0	8	4
RNA	36	40	31	10

[a]Nucleotides with OP atoms in the Mg^{2+} first shell. Protein ligands are omitted.
[b]In 23S-rRNA^{HM}.

1.3.2.2.4 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 1.4).

The arrangement of three OP atoms and three water oxygens around one particular Mg^{2+} in a large globular RNA is shown in Figure 1.1b. 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, ($Mg^{2+}\cdots$RNA distance <2.6 Å), with first-shell RNA ligands. The most common Mg^{2+} ligands are water oxygens (407 ligands). Most RNA ligands are OP atoms (129 first-shell OP ligands). Mg^{2+} does not exhibit a preference for O1P (63 ligands) over O2P (66 ligands). RNA bases are infrequent ligands, with 12 O6 atoms, 10 N7 atoms, five O4 atoms and three O2 atoms within Mg^{2+} first shells. As expected from DNA structures,[61] 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 an Mg^{2+}.

1.3.2.2.5 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 1.3). The primary reason for this preference is that RNA forms 10-membered chelation cycles with Mg^{2+} (Figures 1.3 and 1.4), which are less favorable than the six-membered chelation cycles of ADP/ATP (Figure 1.2). 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 can assume the correct geometric disposition for Mg^{2+}chelation. For example, the RNA conformation in the $Mg^{2+}(OP)_2$ complex in Figure 1.3 deviates profoundly from that in A-form helical RNA.

The most frequent Mg^{2+} chelation motif in RNA is when OP atoms of adjacent residues chelate a common Mg^{2+} ($\Delta = 1$, Figure 1.4). Larger ring sizes are less frequent. $\Delta = 1$ is observed at higher frequency than in other

Table 1.4 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 No.[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]See Figure 1.7 for the locations of these ions in the 2D structure and Figure 1.8 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 parentheses refer to the numbers in the 2J01 coordinate file (23S-rRNA^{TT}).

[e]These are the rRNA residue numbers of the Mg^{2+} first shell in 23S-rRNA^{HM}. The numbers in parentheses 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.

bidentate complexes (Figure 1.4) 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 1.5 and 1.6).

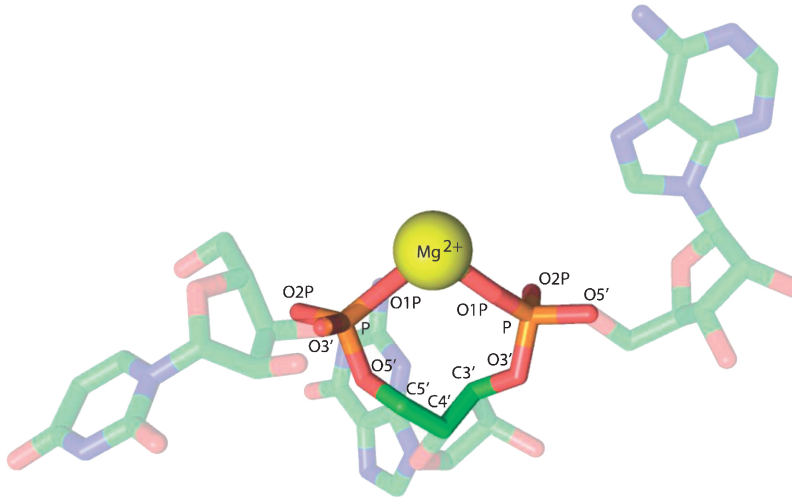

Figure 1.3 RNA as Mg^{2+} chelator. Shown here is a representative bidentate Mg^{2+} $(OP)_2$ complex. For clarity, the base atoms along with ribose atoms C1′, C2′ O4′ and O2′ are lightened. This bidentate complex forms a 10-membered chelation cycle ($\Delta = 1$). It can be seen that this trinucleotide 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).

1.3.2.2.6 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 Figure 1.5 (23S-rRNA^{HM}) and Figure 1.6 (RNA$^{P4–P6}$ 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.

1.3.2.2.7 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 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 side-chains 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

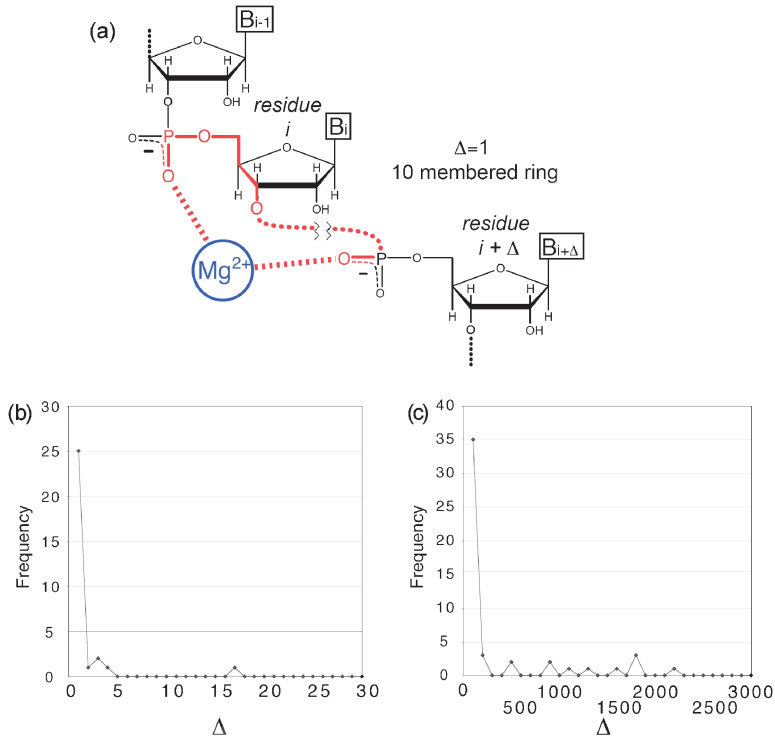

Figure 1.4 Bidentate Mg^{2+} binding is predominantly local. OP ligands are clustered along the RNA chain. (a) Definition of Δ, which is the distance in residues between two OP ligands that bind to a common Mg^{2+} ion. $\Delta = 1$ indicates binding by OPs of adjacent residues, giving a 10-membered chelation ring. (b) The observed frequency of occurrence of Δ values in 23S-rRNA^{HM}, expanded to show the distribution at small Δ. $\Delta = 1$ is preferred over all other Δ. (c) The observed frequency of occurrence of Δ values. Small Δ are preferred over large Δ.

[in $Mg^{2+}(OP)_2$, $Mg^{2+}(OP)_3$, $Mg^{2+}(OP)_4$ complexes]. A representative $Mg^{2+}(OP)_2$ complex in 23S-rRNA^{HM} is shown in Figure 1.3. The OP atoms in these multidentate complexes reach the global minimum in OP $\cdots$ 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 side-chains 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[1–5] and ribosomes.[6–14] Only Li^+ or Mn^{2+} 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.

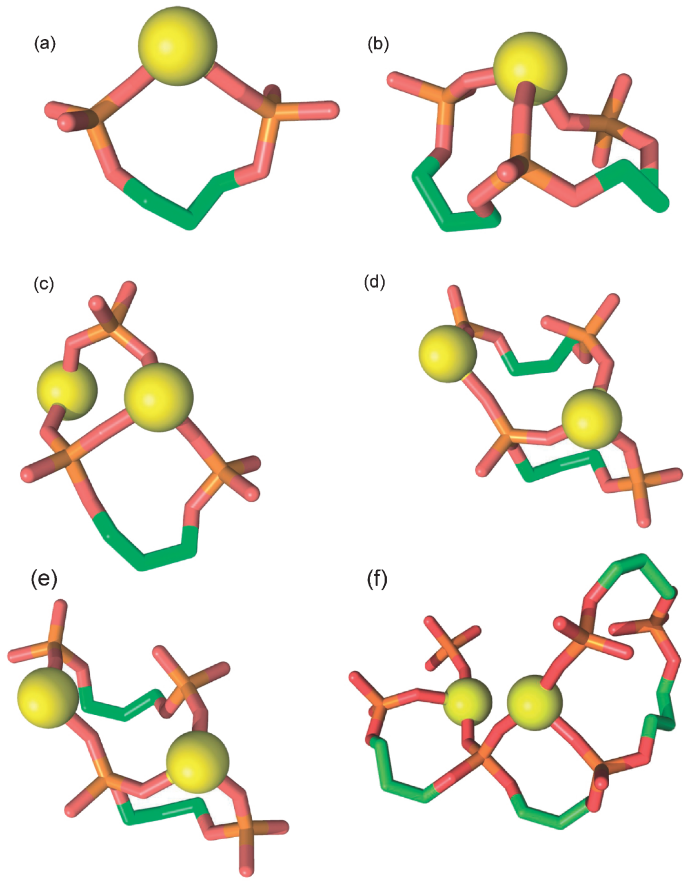

Figure 1.5 RNA as Mg^{2+} chelator. Shown here are some Mg^{2+} chelation complexes in 23S-rRNA^{HM}. The base atoms and ribose atoms C1′, C2′ O4′ and O2′ are omitted for clarity. (a) Bidentate chelation by neighboring OP atoms. This common motif is a 10-membered-ring with a *cis* orientation of OP ligands. 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 1.3). (b) Tridentate chelation by three OP atoms from neighboring residues ($\Delta_1 = 1$, $\Delta_2 = 1$, 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. One Mg^{2+} is involved in tridentate chelation by neighboring and a remote OP atoms (Mg^{2+} 8001, $\Delta_1 = 1$, $\Delta_2 = 50$). The tridentate complex is coupled to a bidentate complex (Mg 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. One Mg^{2+} is involved in tridentate chelation by two neighboring and one remote OP atoms (Mg^{2+} 8016, $\Delta_1 = 1$, $\Delta_2 = 175$). 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. One Mg^{2+} is involved in tridentate chelation by two neighboring and one remote OP atoms (Mg^{2+} 8003, $\Delta_1 = 1$, $\Delta_2 = 1747$). 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. One Mg^{2+} is involved in tridentate chelation by two neighboring and one proximal OP atoms (Mg^{2+} 8005, $\Delta_1 = 1$, $\Delta_2 = 2$). This complex is coupled to another tridentate chelation complex, by two neighboring and one remote OP atoms (Mg^{2+} 8007, $\Delta_1 = 1$, $\Delta_2 = 1007$).

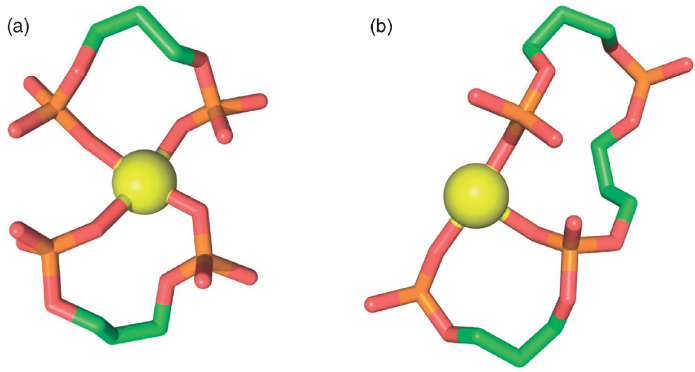

Figure 1.6 RNA as chelator, continued. (a) Tetradetate chelation in the *Thermus thermophilus* 70s ribosome. A complex formed by two $\Delta = 1$ motifs (Mg^{2+} 245, 2J01). (b) Tridentate Mg^{2+} chelation in $RNA^{P4\text{–}P6}$ (Mg^{2+} 3). Neighboring and next-nearest neighboring phosphate groups ($\Delta_1 = 1$, $\Delta_2 = 2$) bind to the same Mg^{2+}. Note the similarity of this complex to that in Figure 1.5f.

1.3.2.2.8 Mg^{2+} *Versus* Na^{+}. It has been suggested that high concentrations of Na^{+} attenuate OP$\cdots$OP repulsion to the extent that globular RNAs can fold, achieving native OP$\cdots$OP proximities, in the absence of Mg^{2+}.[60] 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 this cation to alternate 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 tightly packed OP atoms. 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.

1.3.2.2.9 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 ($\Delta = 1$, Figures 1.3 and 1.4). Distinctive conformational states are associated with such bidentate complexes. The 10-membered rings of RNA that chelate Mg^{2+} are not, with few exceptions, A-form helix or tetraloops. *Trans* bidentate complexes are favored electrostatically (by attenuated OP$\cdots$OP repulsion) in comparison with *cis* bidentate complexes, but are disfavored by the entropic cost associated with the necessarily increased linker size. The minimum observed linker for a *trans* $Mg^{2+}(OP)_2$ complex in $23S\text{-}rRNA^{HM/TT}$ is four residues.

What stabilizes *cis* Mg^{2+} $(OP)_2$ complexes in which OP atoms are forced into close proximity? First, $OP \cdots Mg^{2+}$ attraction offsets the $OP \cdots OP$ repulsion. Second, repulsion is attenuated by charge transfer from OP to Mg^{2+}. Third, the RNA backbone allows the 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 $OP \cdots OP$ proximities present kinetic barriers to changes in coordination state.[76] Chapters 6 and 7 contain detailed discussions of the effects of Mg^{2+} on the kinetics of RNA folding.

Many multidentate $Mg^{2+}(OP)_n$ complexes are found in 23S-$rRNA^{HM}$, as described by Klein *et al.*[11] As these authors noted, the most frequent chelation motif is $Mg^{2+}(OP)_2$, where phosphate groups from neighboring residues chelate a common Mg^{2+} (Figures 1.3–1.6). 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 $\Delta = 1$, where Δ is the distance, in number of residues, between the two OP groups (Figure 1.4). The tridentate complex shown in Figure 1.1b yields two different Δ values; 1 and 95. Nine of 10 $Mg^{2+}(OP)_3$ centers in 23S-$rRNA^{HM}$ contain at least one $\Delta = 1$ complex. The $Mg^{2+}(OP)_3$ center in Figure 1.5b is composed of two $\Delta = 2$ complexes.

The chelating ring size patterns observed in 23S-$rRNA^{HM}$ are general features of large RNAs, in which $Mg^{2+}(OP)_2$ $\Delta = 1$ complexes are observed alone and in combination with other chelation rings. A double $\Delta = 1$ complex, with tetradentate chelation, is observed linking the 16S RNA with the mRNA in the intact ribosome of *Thermus thermophilus* (PDB entry 2J01, Figure 1.6a). 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 1.5f) is found in $RNA^{P4–P6}$ (Figure 1.6b).

1.3.2.2.10 Bicycles, Tricycles. Tridentate Mg^{2+}–RNA complexes, although less frequent than monodentate and bidentate complexes, are especially important in structure, stability and function. Ten $\Delta = 1$ 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 $\Delta = 1$, $\Delta = 1$ example is shown in Figure 1.5b) and those containing Mg^{2+}–O1P–P–O2P–Mg^{2+} linkages (Figure 1.5c–f). Four Mg^{2+}–O1P–P–O2P–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 Figure 1.5d (D3) with 1.5e (D2)] suggests that general rules for RNA conformation and Mg^{2+} interactions are discernible from them.

1.3.2.2.11 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 in ribosomes over vast evolutionary timescales. They were identified initially by their RNA coordination and then validated by comparison between different rRNAs. Broadly, a paleo-Mg^{2+} ion is in either an $Mg^{2+}(OP)_3$ center, with at least three RNA first-shell OP ligands, or is a 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$^{HM/TT}$. Figure 1.5b–f are examples of paleo-Mg^{2+} ions from 23S-rRNA^{HM}. The requirement for three OP first-shell ligands over base and sugar ligands 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 1.7 and 1.8, Table 1.4). 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+}

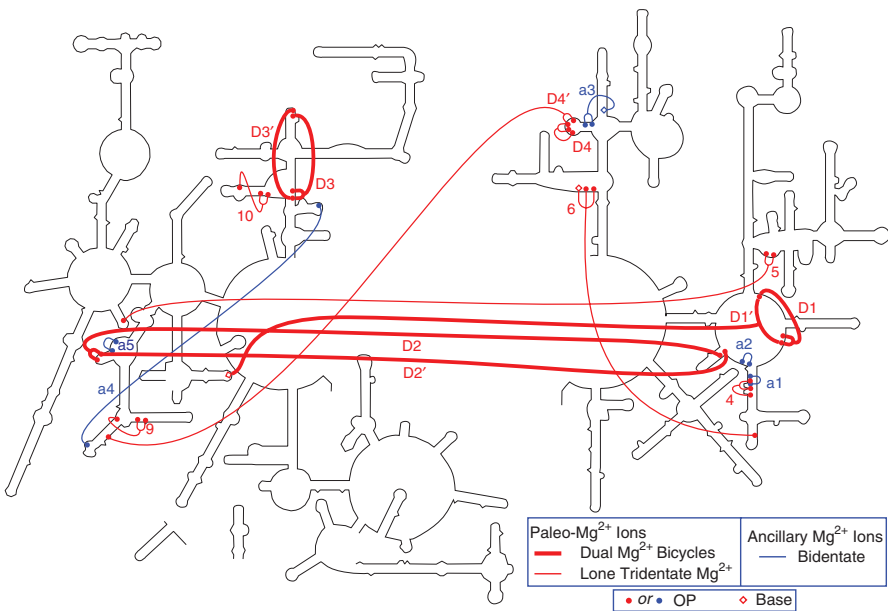

Figure 1.7 The secondary structure of 23S-rRNA^{HM}. The locations of paleo-Mg^{2+} ions are depicted in red. The thick red lines indicate the dual Mg^{2+} bicycles (D1, D2, D3 and D4). The thin red lines indicate the isolated $Mg^{2+}(OP)_3$ complexes. The ancillary $Mg^{2+}(OP)_2$ ions, which cluster 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.

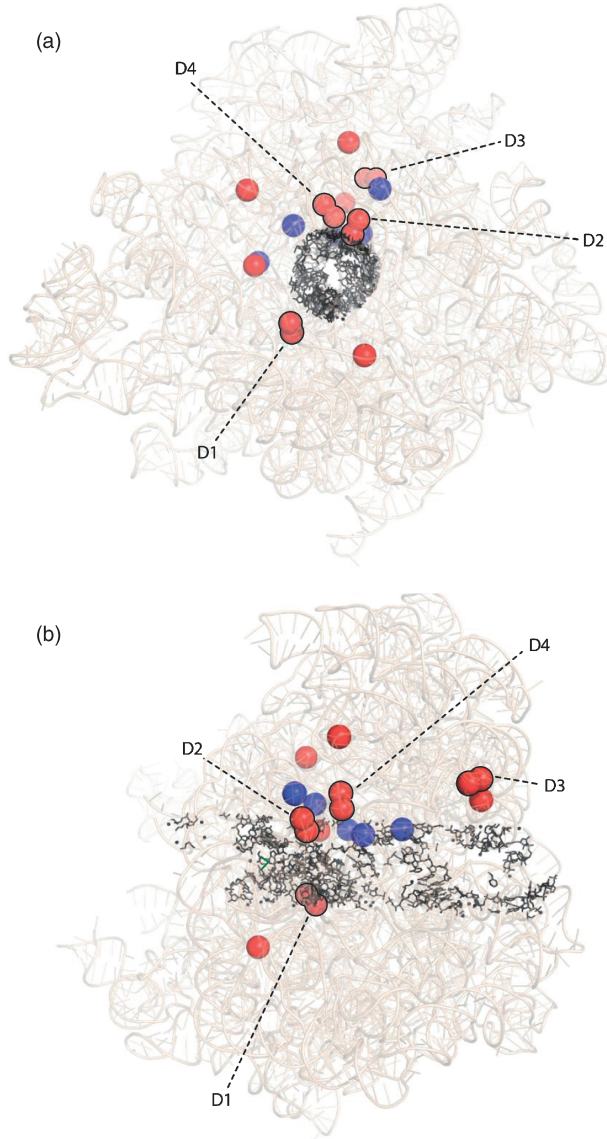

Figure 1.8 (a) View into the peptidyl transfer center of 23S-rRNA^{HM}. The 13 paleo-Mg^{2+} ions are shown as red spheres. The Mg^{2+} ions of the four dual Mg^{2+} bicycles/tricycles (D1, D2, D3 and D4) are indicated. Ancillary Mg^{2+} ions are shown in blue. The RNA atoms lining the peptide exit tunnel 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.

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 1.7. The secondary elements linked by paleo-Mg^{2+} ions are conserved between bacteria and archea (see the next item), in the proposed secondary structures of eukaryotic[77] and some mitochondrial rRNAs,[78] and in a proposed minimal 23S-rRNA.[79]

(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 observed.

(v) *Role in function*. The locations of paleo-Mg^{2+} ions appear to lend critical support to function (Figure 1.8). 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, whereas D3 is located by the exit of the peptide tunnel.

(vi) *Absence of protein ligands*. Paleo-Mg^{2+} ions are not coordinated by protein ligands (Table 1.4), 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.

1.3.2.2.12 Mg^{2+} Avoids RNA Motifs. Chelation of Mg^{2+} is coupled to RNA conformation. Mg^{2+} ions select against canonical conformations such as A-form helices and tetraloops. As noted by Moore[80] and others,[81] folded RNA is largely composed of a relatively small number of motifs such as A-helices,[82] tetraloops,[83–86] E-loop motifs,[87–91] and kink-turns.[37,92,93] 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.[94] Formation of RNA motifs (secondary structure) is not Mg^{2+} dependent.

One can observe that Mg^{2+} inner-shell complexes are apparently incompatible with RNA tetraloops. In Figure 1.9, first-shell OP interactions with Mg^{2+}

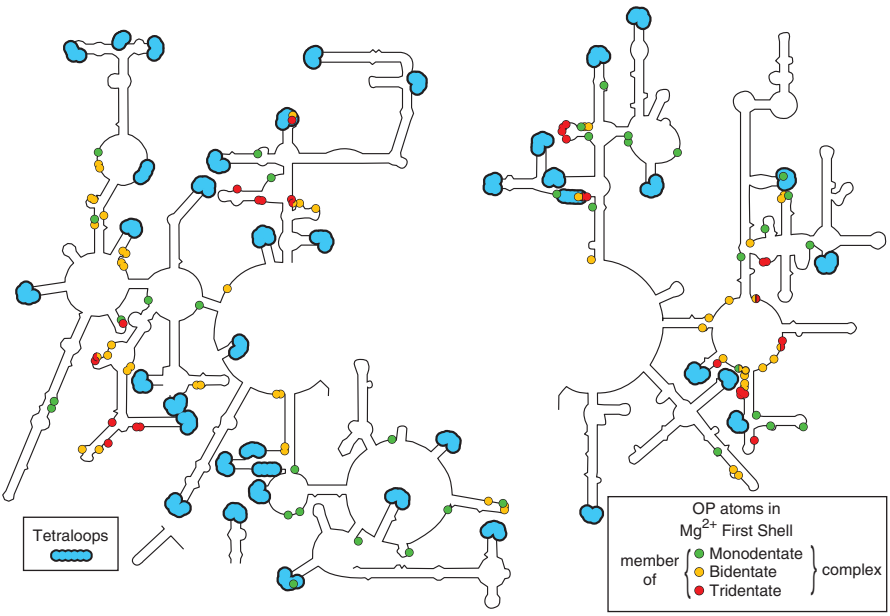

Figure 1.9 Multidenate Mg^{2+} binding selects against tetraloops. The secondary structure and Mg^{2+} contacts of the 23S rRNA from HM are shown. Tetraloop positions, as determined by Hsiao *et al.*,[94] are indicated in blue. Mg^{2+}–OP interactions are indicated by circles. OP atoms that interact with Mg^{2+} *via* monodentate interactions are green circles. OP atoms that interact with Mg^{2+} *via* bidentate interactions are yellow circles. OP atoms that interact with Mg^{2+} *via* tridentate interactions are red circles.

are mapped on to the 23S-rRNA^{HM} secondary structure, as are the locations of tetraloops.[94] The sites of first-shell Mg^{2+} coordination do not in general correspond to the locations of the tetraloops.

This preference is supported by statistical results. Automated methods[94–96] allow one to count conformational states and determine their populations (frequencies), locations and sequences. One can seek correlations between frequency of occurrence of conformational states and locations of Mg^{2+} ions. 23S-rRNA^{HM} can be partitioned into conformational states and grouped by frequency. For example, various conformational states can be differentiated by their torsion fingerprints. States with many occupants are motifs. States with few occupants represent non-motif RNA (conformational-deviants). To help determine if Mg^{2+} interacts preferentially with motifs or with non-motif RNA requires a comparison with a random binding model (see Appendix) that assumes no preference of Mg^{2+} binding to any residue or conformation over another. We have used this method to calculate populations and then compared the calculated with the observed populations. The results, illustrated in Figure 1.10 and tabulated in Table A1 in the Appendix, indicate that Mg^{2+} binds preferentially to non-motif RNA. The observed extent of binding to non-motif RNA exceeds that predicted by the random model. The observed extent

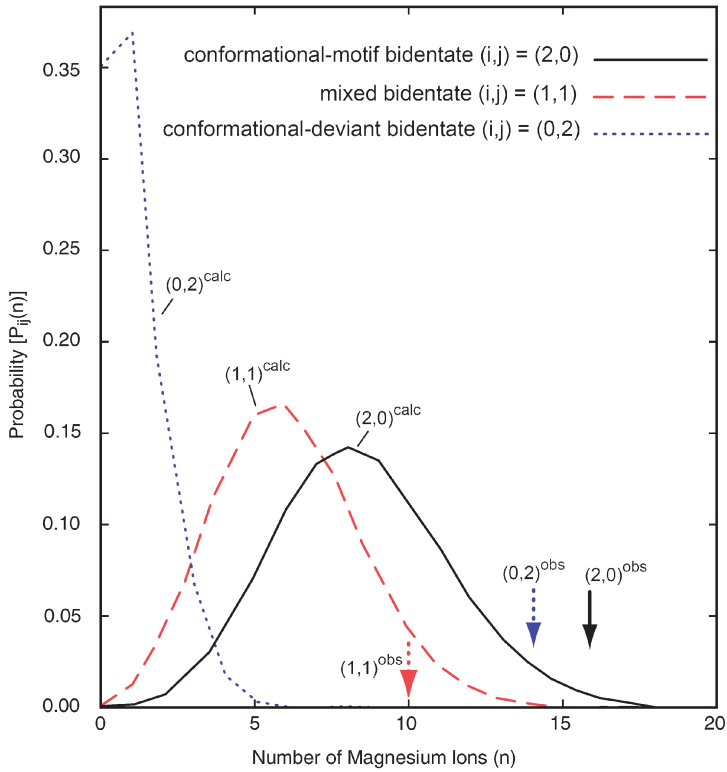

Figure 1.10 Mg^{2+} ions that form multiple bonds to RNA prefer conformational-deviants. The observed populations of various binding states are indicated by vertical arrows. The predicted propensity envelopes, obtained from a random binding model that assumes no preference of Mg^{2+} ion binding for canonical *versus* deviant conformations are shown as solid, dashed and dotted lines. $p_{ij}(n)$ is the probability of n Mg^{2+} ions occupying state (i,j). Predicted envelopes are shown for Mg^{2+} ions in state (2,0) with two bonds to canonical RNA and none to a conformational-deviant, state (1,1) with one bond to canonical and one bond to non-canonical RNA and (0,2) with two bonds to conformational-deviants and none to canonical RNA. The observed populations of states (2,0) and (1,1) are within or very close to the predicted binding envelopes. The observed population in the (0,2) state is far greater than predicted by the random binding envelope.

of binding to motif RNA, such as A-form helices and tetraloops, falls below that predicted by a random binding model.

1.3.2.2.13 RNA Associated with Mg^{2+} is Dispersed in Conformational Space. The diversity of RNA conformation in the vicinity of $Mg^{2+}(OP)_2$, is illustrated in Figure 1.11. The conformation of RNA acting as bidentate Mg^{2+} ligands is different from the conformation of other RNA, including RNA in

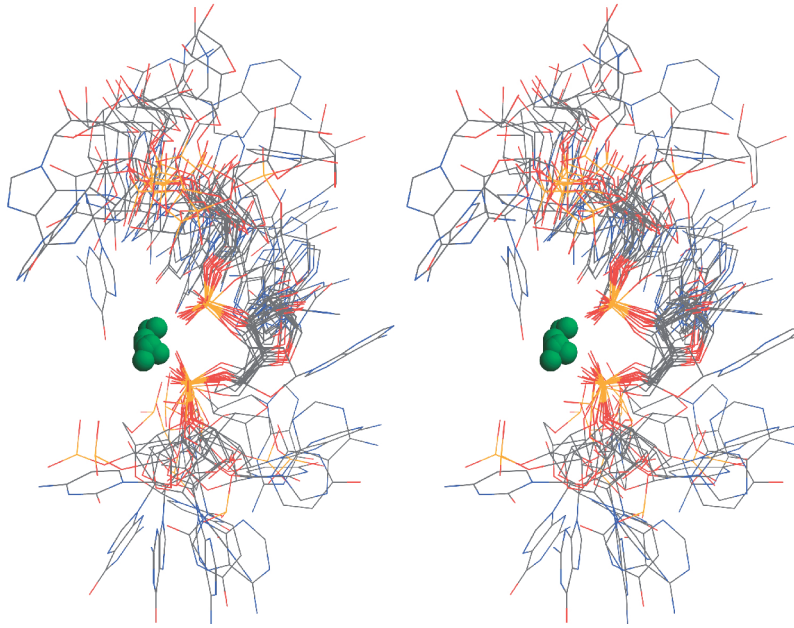

Figure 1.11 Mg^{2+} ions that form multiple bonds to RNA prefer conformational deviants. Fragments of RNA that bind directly to Mg^{2+} were superimposed. Each of these RNA fragments bind to Mg^{2+} using the same two atoms (O1 of residue *i* and OP2 of residue *i*+1). The backbone atoms of the two residues linking 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. From 23S-rRNA^{HM}.

the vicinities of other bidentate Mg^{2+} ions. This result is consistent with the observation of Klein *et al*. that Mg^{2+} ions tend to bind to "highly idiosyncratic binding sites in 23S RNA that are unlike any previously reported".[11] 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.[87]

1.3.2.2.14 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.[26,97–100] Entropic gain by water release may, in some cases, be a dominant factor in binding, especially for Mg^{2+} (see below). High-level calculations underscore the importance of non-electrostatic contributions, such as polarization and charge transfer. For example, calculations by Rulisek and Šponer suggest that inner-shell binding of 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.[98] NLPB, molecular dynamics simulation and other computational approaches generally assume that electrostatic interactions are the only significant contribution to energy of binding. Some level of error should be expected 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.[101] 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. This 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 is applicable to aqueous Mg^{2+}-OP interactions in general.

1.4 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 identified by X-ray diffraction should be considered to be approximations.[102] Monovalent cations in particular present non-trivial analytical challenges. Sodium ions (Na^+), potassium ions (K^+), rubidium ions (Rb^+), caesium ions (Cs^+), ammonium ions and water molecules, and even polyamines and divalent cations, compete for overlapping sites adjacent on DNA and RNA.[24,103,104] A variety of 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 occupancy levels.

1.4.1 Group I

Tl^+ was initially investigated as a K^+ mimic in biological systems by R.J.P. Williams *et al.*[15,105] and others.[106–111] 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 for 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,[112] fructose-1–6-bisphosphatase[113] and pyruvate kinase.[114] Tl^+ has been shown to stabilize guanine quadruplexes in a manner analogous to K^+ and ammonium.[45,115,116] We have used Tl^+ as a probe for K^+ in association with B-DNA[104,117] and in DNA–drug complexes.[118] Doudna and co-workers used Tl^+ as a probe for K^+ in the tetrahymena ribozyme P4–P6 domain.[1] Draper and co-workers used Tl^+ as a K^+ probe in the structure of a fragment of the 23S rRNA.[119] Correll *et al.* used Tl^+ as a K^+ probe in the structure of the sarcin/ricin loop of the 23S rRNA.[86] Tl^+ was used by Caspar and co-workers to determine counterion positions adjacent to insulin[120,121] and by Gill and Eisenberg to determine the positions of ammonium ions in the binding pocket of glutamine synthetase.[122] Rb^+ has also been used, with some success, as a K^+ substitute in DNA structures[123,124] and in ribosomal structures.[11]

1.4.2 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 Å and ligand to ligand distances of 2.9 Å (see Figure 1.1b). 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+}),[70,125–128] which gives a useful anomalous signal.[2,129] Cobalt hexamine is a useful NMR probe for fully hydrated magnesium.[130–135]

1.5 The Two Binding-mode Formalism

Cation associations with nucleic acids are commonly ascribed to one of two modes, site-bound or non-specific.[18,19,29,35,58,97,119,136–142] This formalism is, at its core, a thermodynamic construct. Yet one constantly strives to interpret and extend this model in terms of three-dimensional structures and theoretical models. In fact, the definitions of the site-bound and non-specific modes are variable, depending on the application and context. It is important to understand that that the various definitions are not necessarily self-consistent; an ion can be site bound by one definition and diffusely bound by another. It seems clear that further progress will be aided by concise definitions of terms. Here we list what we believe are the consensus definitions of site-bound and non-specific ions in various contexts.

1.5.1 Thermodynamic/Kinetic Definition

Non-specific cations are found in diffuse ‘ion clouds’ and ‘ion-atmospheres’. Ions on the outer reaches of an atmosphere interact weakly with the nucleic

acid, whereas the inner cations interact strongly. Their integrated contribution to stability is large. These ions retain mobility and are exchangeable for a broad array of alternative ion types. Alternatively, site-bound cations localize in confined volumes with highly favorable enthalpies but at the cost of an entropic penalty (note that the contributions from water release can switch the signs of both effects). Site-bound ions show significantly reduced mobility compared with bulk aqueous ions and cannot be replaced by alternative ionic species.

1.5.2 Structural Definition

Non-specific cations retain water molecules as inner-sphere ligands. These cations are not closely associated with specific nucleic acid functional groups. Non-specific cations are not dependent on specific geometric dispositions of DNA/RNA functional groups. Site-binding is associated with inner-sphere coordination and especially with multidentate interactions (chelation). Site bound ions have lost one or more first-shell water molecules. Multidentate complexes, particularly with Mg^{2+}, require specific geometric dispositions of the RNA/DNA functional groups and can impose special conformational restraints on RNA/DNA. Multidentate Mg^{2+} interactions can stabilize conformational states that are otherwise unfavorable. *Cis*-bidentate Mg^{2+} binding brings OP atoms into van der Waals contact. Draper has discussed an intermediate binding mode called 'water-positioned' cations.[32]

1.5.3 Computational Definition

The non-specific and site-bound categories have specific definitions and practical consequences in NLPB theory,[20,22] in which electrostatic energies are computed from detailed atomic structures. Chapter 9 discusses computational approaches to understanding long-range electrostatic forces. Chapter 6 discusses the application of NLPB to Mg^{2+} and RNA folding. Here it is sufficient to know that NLPB is a continuum solvent model for treating electrostatic interactions. The electrostatic free energy is computed for empty sites plus the free cations and again for filled sites. Non-specific ions are treated non-atomistically, as a continuum that is distinct from the RNA/DNA, with intrinsically incorporated entropy. Specifically bound ions are generally treated as discrete entities. The entropy of site binding is not incorporated automatically. A specific binding site is distinguished from a region of diffuse binding by a small confinement volume and a large occupation probability per unit volume. A specific binding site is restricted to two occupational states, namely occupied or empty, with no partial occupancy permitted. A bound cation is regarded as fixed and part of the fixed charge array of the polyion, rather than part of the diffuse ion atmosphere.

1.5.4 Breaking the Two Binding-mode Formalism

The two binding-mode formalism may provide reasonably satisfactory descriptions of the interactions of Mg^{2+} with nucleic acids. However, for group I cations we believe that this formalism is not generally useful and is probably misleading. For group I cations the distinction between site- and non-specific binding is commonly blurred. At one limit, monovalent cations fully conform to the conventional site-bound description in association with G-quadruplexes,[41–46] with Draper's 58 nucleotide rRNA fragment[119] and with the AA platform of the RNA tetraloop receptor.[1] At the other limit, a subset of fully hydrated group I cations is held in loose association with double helical RNA and DNA.

However, group I ions commonly exhibit intermediate behavior, with a mix of site-bound and non-specific characteristics. Ions appear site bound in that they form inner-shell complexes, with up to four nucleic acid ligands.[66,103,104,123,124] Some of these ions appear non-specific in that they retain significant mobility, switching readily between coordination environments, with residence times from 10 ns to 100 μs,[143] and are exchangeable between a broad array of alternative ion types, including polyamines.[61,103,104,123,144]

DNA ligands of group I ions are uncharged base oxygens and nitrogens, which fall in overlapping arrays along the floors of helical grooves. Therefore, the activation energies and the changes in free energies for the transitions from one site to the next are nominal. Group I cations can transit along the major groove, from say N7/O6 to N7/O6 of adjacent guanines, or along the minor groove B-form A-tract without much of an activation barrier or change in free energy (discussed in Chapter 3). The free energies required to alter coordination geometry and coordination number are small in magnitude. Nucleic acids offer competing isoenergetic binding sites with varying numbers and geometries of first-shell ligands. In sum, differences in coordination chemistry suggest a continuum between site-bound and diffuse modes for Na^+ and K^+ and a clear demarcation for Mg^{2+}.

1.6 Reaction Coordinates for RNA Folding

1.6.1 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.[145,146] Structural databases allow determination of averages and deviations of bond and hydrogen bond lengths, bond angles and dihedrals.[147,148] Structural databases also allow the determination of coordination sphere geometry,[73–75] reaction coordinates and transition pathways.[76,149–153]

Burgi and Dunitz used data-mining of crystal structures to determine reaction coordinates for simple organic reactions.[149,150,154] Similarly, reaction

coordinates for conformational transition reaction coordinates[155] and along folding reaction coordinates[152] have been determined for biological polymers.

1.6.2 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.* $\Delta = 1$ 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 form preferentially with single-stranded regions of RNA folding intermediates. Adjacent residues along a *single-stranded (i.e. flexible) RNA chain* achieve close proximity with greatest probability,[156] and are conformationally most polymorphic. Adjacent residues along a *double-stranded (i.e. relatively rigid) RNA chain* achieve close proximity with lower probability and are conformationally homogeneous.

Hence, 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 time-scale 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 1.12. Support for this general mechanism is provided by results of Woodson and co-workers,[157] who concluded that Mg^{2+} ions dampen the dynamics of RNA folding intermediates. Chapter 6 discusses restriction of RNA conformational fluctuations by Mg^{2+} ions. The mechanism here is consistent with preferential binding of Mg^{2+} to ssRNA over dsRNA as observed experimentally in solution,[158] even though that preference is counter to the predictions of polyelectrolyte theory.[18,159]

Appendix

Statistical Analysis of Mg^{2+}–RNA Interactions: Conformational Motifs *Versus* Conformational Deviants

An Mg^{2+} with i bonds to conformational motif(s) and j bonds to conformational deviant(s) [*i.e.* $Mg^{2+}(RNA)i+j$] is said to be in state (i,j). With a geometric algorithm we determined n_{ij}^{obs}, the number of observed Mg^{2+} ions in 23S-$rRNA^{HM}$ in each state (i,j). The results, summarized in Table A1, indicate, for example, that 25 Mg^{2+} ions are observed in state (0,0) [*i.e.* $Mg^{2+}(RNA)_{n=0}$]. These Mg^{2+} ions are fully hydrated and do not form bonds to RNA in either canonical or non-canonical conformations. Thirty-three Mg^{2+} ions form a single bond to RNA [$Mg^{2+}(RNA)_{n=0}$]. Twenty of those Mg^{2+} ions bond to RNA in

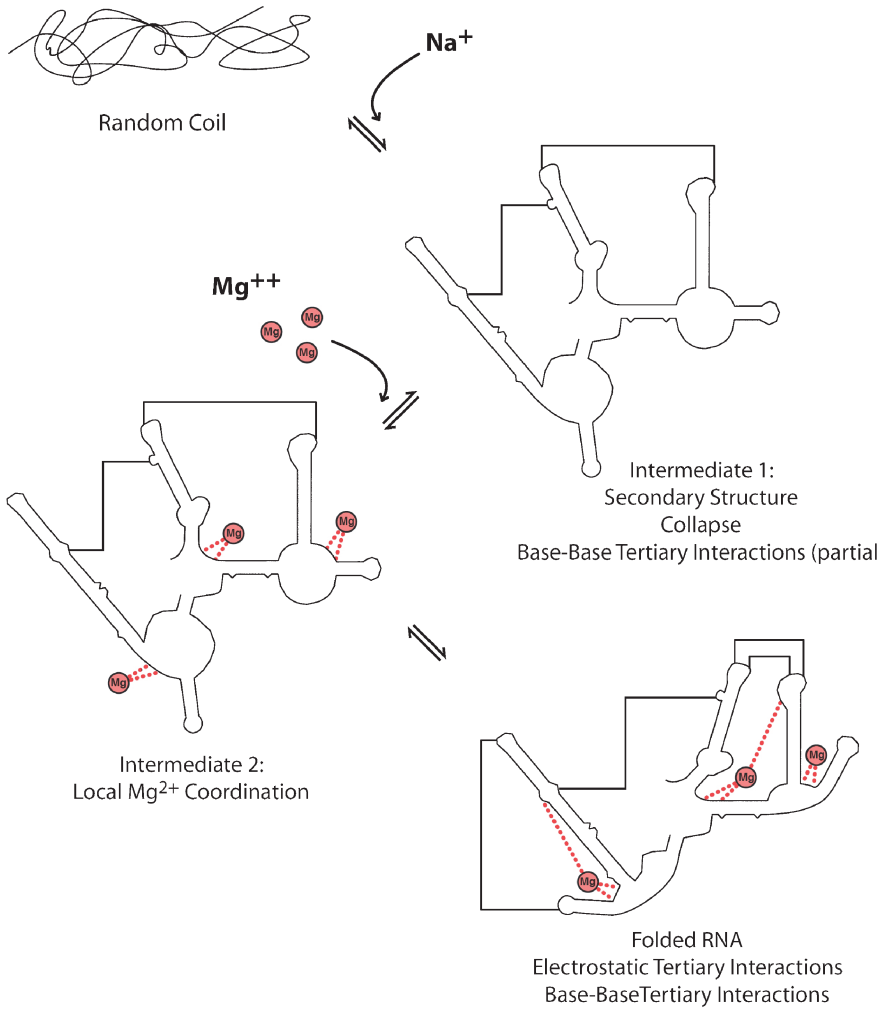

Figure 1.12 Simplified mechanism of RNA folding. The first step of RNA folding is Na^+- (or K^+-) dependent. In this step, secondary structure forms, along with some tertiary structure (solid lines). The second step is Mg^{2+}-dependent. In this step, local $Mg^{2+}(OP)_2$ chelation occurs (short dashed lines), preferentially at single-stranded regions. This Mg^{2+} binding dampens conformational fluctuations of the single-stranded regions. The third step forms electrostatic tertiary interactions (long dashed lines), which are Mg^{2+} mediated interactions between OPs that are remote in primary structure. At the same time, the final tertiary interactions are locked into place.

canonical conformations ($n_{10}^{obs} = 20$) and 13 of them bond to RNA in non-canonical conformations ($n_{01}^{obs} = 13$), *etc.*

To help determine if Mg^{2+} interacts preferentially with conformational motifs or with conformational deviants, we formulated a random binding

Table A1 Observed and calculated probabilities of Mg^{2+} interaction with conformational motifs and conformational deviants.

		State[a]	*Observed*	*Calculated*		
			n_{ij}^{obs}	$n_{ij,ave}^{calc}$	σ_{ij}^{calc}	$\Delta_{ij}/\sigma_{ij}^{calc}$ [b]
Monodentate	$Mg^{2+}(RNA)_0$	(0,0)	25	25	4.4	0.0
	$Mg^{2+}(RNA)_1$	(1,0)	20	15	3.6	1.5
	$Mg^{2+}(RNA)_1$	(0,1)[c]	13	5	2.2	3.6
Bidentate	$Mg^{2+}(RNA)_2$	(2,0)	16	8	2.8	2.9
	$Mg^{2+}(RNA)_2$	(1,1)	10	6	2.4	1.7
	$Mg^{2+}(RNA)_2$	(0,2)[c]	14	1	1.0	12.7
Tridentate	$Mg^{2+}(RNA)_3$	(3,0)	0	5	2.2	−2.3
	$Mg^{2+}(RNA)_3$	(2,1)	3	5	2.2	−1.0
	$Mg^{2+}(RNA)_3$	(1,2)	4	2	1.3	1.6
	$Mg^{2+}(RNA)_3$	(0,3)[3]	5	0	0.45	10.4

[a]State (i,j) contains i bonds to conformational motifs and j bonds to conformational deviants.
[b]$\Delta_{ij}/\sigma_{ij}^{calc}=(n_{ij}^{obs}-n_{ij}^{calc})/\sigma_{ij}^{calc}$, which is the discrepancy between observation and calculation, normalized to the standard deviation of the calculated probability envelope.
[c]Most significant deviations between model and observation.

model that assumes no preference of Mg^{2+} binding to any given residue or conformation over another. We use this model to compute populations of states. We assume that the binding process occurs *via* a series of steps, where at each step a bond is formed with probability p_b. A total of n bonds requires n binding steps, followed by a terminal non-bonding step. The final result is essentially identical with that for a biased one-dimensional random walk. This approach allows us to compute $n_{ij,ave}^{calc}$, the average number of Mg^{2+} ions in state (i,j). In addition, we compute $\sigma_{ij,}^{calc}$, the standard deviation of $n_{ij,ave}^{calc}$. Finally, we compute the statistical significance of the difference between the observed and calculated results $(n_{ij}^{obs}-n_{ij,ave}^{calc})/\sigma_{ij}^{calc}$. The results are tabulated in Table A1 and Figure 1.10.

The combined results indicate that the observations are not well described by the random model. In observed monodentate RNA complexes [*i.e.* $Mg^{2+}(RNA)_1$], Mg^{2+} ions are observed to form bonds with moderate preference for conformational deviants over conformational motifs. In observed multidentate complexes [$Mg^{2+}(RNA)_{n>1}$], Mg^{2+} binds with strong preference for conformational deviants. Multidentate Mg^{2+} [$Mg^{2+}(RNA)_{n>1}$] does not bind randomly, but binds with clear preference to conformational deviants. This preference is indicated by uniformly positive values of $(n_{ij}^{obs}-n_{ij,ave}^{calc})/\sigma_{ij}^{calc}$ for Mg^{2+} ions that form fewer bonds to conformational motifs than conformational deviants [*i.e.* for states (0,1), (0,2), (1,2) and (0,3)]. The (0,2) and (0,3) states show the greatest statistical deviations from the random model, with values of $(n_{ij}^{obs}-n_{ij,ave}^{calc})/\sigma_{ij}^{calc}$ of 12.7 and 10.4, respectively. For the (0,2) state, the random model predicts a population of 1 whereas the database actually contains 14. For the (0,3) state, the random model predicts a population of zero while 23S-rRNA^{HM} contains 5. We excluded Mg^{2+} ions with more than three interactions, since there are very few of them and none of the corresponding states are occupied by more than a single Mg^{2+} ion.

Derivation of Functions Governing Random Mg^{2+} Ion Binding to RNA

We begin by first deriving the probability that a given Mg^{2+} has i canonical and j non-canonical interactions. We say that such an Mg^{2+} is in state (i,j) and we denote the probability of being in state (i,j) by p_{ij}.

Let us assume that the binding process occurs via a series of steps, where at each step a bond is formed with probability p_b. For a total of n bonds to form, there must be n binding steps, followed by a terminal nonbonding step. The total probability is then $p_b^n(1-p_b)$. Assuming that an Mg^{2+} is bonded n times, we can then ask what the probability is of it having i canonical and $j=n-i$ non-canonical interactions. If f denotes the fraction of canonical RNA, then the probability of binding to canonical RNA is f and the probability of binding to non-canonical RNA is $1-f$. In the $n=i+j$ binding steps, i of the steps must be canonical bonds and j must be non-canonical bonds. Hence there are $(i+j)!/i!j!$ distinct binding sequences and since each sequence has a probability of $f^i(1-f)^j$, we obtain a total probability of $(i+j)!/i!j!f^i(1-f)^j$. Multiplying by the probability of forming n bonds gives

$$p_{ij} = (i+j)!/i!j!f^i(1-f)^j p_b^{i+j}(1-p_b)$$

Consider the case of N Mg^{2+} atoms interacting with the RNA molecule. We may denote n_{ij} as the number of Mg^{2+} ions in state (i,j). Let $n=(n_{00}, n_{10}, n_{01}, n_{20}, n_{11}, n_{02}, \ldots)$ denote the vector of n_{ij} values. In principle, n can take on any set of values subject to the restriction that $\sum_{ij} n_{ij} = N$. We set p_b equal to the observed fraction of bonded Mg^{2+}s in order to normalize n_{00} to $n_{00,avg}$.

Let $P(n)$ denote the probability of obtaining the occupancy vector n. Note that there are $N!/\Pi_{ij}n_{ij}!$ ways of assigning states to the N Mg^{2+} ions consistent with the occupancy n. Since each such assignment has a probability $\Pi_{ij}p_{ij}^{n_{ij}}$, we obtain

$$p(n) = (N!/\Pi_{ij}n_{ij}!)\Pi_{ij}p_{ij}^{n_j}$$

With this probability distribution in hand, we can now proceed to compute any statistical quantity of interest to us. Here, we are interested in the following quantities: (1) $P_{ij}(n)$, the probability of n Mg^{2+} ions being in state (i,j), (2) $n_{ave}=(n_{00,ave}, n_{10,ave}, n_{01,ave}, \ldots)$ and (3) $\sigma_{ij}^2 = n_{ij,ave}^2 - n_{ij,ave}^2$. To determine (2) and (3), we consider the general case of evaluating $f(n_{ij})_{,ave}$ over the distribution $P(n)$, for some function f. Note that

$$f(n_{ij})_{,ave} = \sum_n f(n_{ij})P(n) = \sum_{n-ij=0,\ldots,N} f(n_{ij}) \sum_n P(n)$$

where the final summation is taken over all n consistent with the value n_{ij}. Note then that this sum is simply $P_{ij}(n)$ and thus,

$$f(n_{ij})_{,avg} = \sum_{n-ij=0,\ldots,N} f(n_{ij}) P_{ij}(n_{ij})$$

To determine $P_{ij}(n_{ij})$, note that it is obtained by summing over all sequences consistent with having n_{ij} Mg^{2+} ions in state (i,j). We can readily evaluate such a sum as follows: There are $N!/n_{ij}!(N-n_{ij})!$ different ways of placing n_{ij} Mg^{2+} ions into state (i,j). The remaining $N-n_{ij}$ Mg^{2+} ions can each be in any state other than (i,j). Since the probability of being in state (i,j) is p_{ij}, the probability of being in any state other than (i,j) is $1-p_{ij}$. Therefore, given a set of n_{ij} Mg^{2+} ions in state (i,j) and $N-n_{ij}$ Mg^{2+} ions in any other state, the probability of the sequence is $p_{ij}^{n_{ij}}(1-p_{ij})_{ij}^{N-n}$. Multiplying by the number of ways of choosing n_{ij} Mg^{2+} ions to be in state (i,j), we obtain

$$P_{ij}(n_{ij}) = \left[N!/n_{ij}!(N-n_{ij})!\right] p_{ij}^{n_{ij}} (1-p_{ij})^{N-n_{ij}}$$

From this equation, it is a standard result that $n_{ij,ave} = Np_{ij}$ and $\sigma_{ij}^2 = Np_{ij}(1-p_{ij})$. It may be noted that these results are identical with those for a biased one-dimensional random walk.

Acknowledgements

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

References

1. S. Basu, R. P. Rambo, J. Strauss-Soukup, J. H. Cate, A. R. Ferre-D'Amare, S. A. Strobel and J. A. Doudna, *Nat. Struct. Biol.*, 1998, **5**, 986.
2. J. H. Cate, R. L. Hanna and J. A. Doudna, *Nat. Struct. Biol.*, 1997, **4**, 553.
3. J. H. Cate, A. R. Gooding, E. Podell, K. Zhou, B. L. Golden, A. A. Szewczak, C. E. Kundrot, T. R. Cech and J. A. Doudna, *Science*, 1996, **273**, 1696.
4. J. H. Cate, A. R. Gooding, E. Podell, K. Zhou, B. L. Golden, C. E. Kundrot, T. R. Cech and J. A. Doudna, *Science*, 1996, **273**, 1678.
5. K. Juneau, E. Podell, D. J. Harrington and T. R. Cech, *Structure*, 2001, **9**, 221.
6. J. H. Cate, M. M. Yusupov, G. Z. Yusupova, T. N. Earnest and H. F. Noller, *Science*, 1999, **285**, 2095.
7. N. Ban, P. Nissen, J. Hansen, P. B. Moore and T. A. Steitz, *Science*, 2000, **289**, 905.

8. B. T. Wimberly, D. E. Brodersen, W. M. Clemons Jr., R. J. Morgan-Warren, A. P. Carter, C. Vonrhein, T. Hartsch and V. Ramakrishnan, *Nature*, 2000, **407**, 327.
9. J. Harms, F. Schluenzen, R. Zarivach, A. Bashan, S. Gat, I. Agmon, H. Bartels, F. Franceschi and A. Yonath, *Cell*, 2001, **107**, 679.
10. M. M. Yusupov, G. Z. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. Cate and H. F. Noller, *Science*, 2001, **292**, 883.
11. D. J. Klein, P. B. Moore and T. A. Steitz, *RNA*, 2004, **10**, 1366.
12. V. Berk, W. Zhang, R. D. Pai and J. H. Cate, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 15830.
13. M. Selmer, C. M. Dunham, F. V. Murphy, A. Weixlbaumer, S. Petry, A. C. Kelley, J. R. Weir and V. Ramakrishnan, *Science*, 2006, **313**, 1935.
14. N. R. Voss, M. Gerstein, T. A. Steitz and P. B. Moore, *J. Mol. Biol.*, 2006, **360**, 893.
15. R. J. P. Williams, *Adv. Chem. Ser.*, 1971, **100**, 155.
16. R. J. P. Williams, *J. Chem. Soc., Dalton Trans.*, 1991, 539.
17. C. B. Black, H. W. Huang and J. A. Cowan, *Coord. Chem. Rev.*, 1994, **135**, 165.
18. G. S. Manning, *Q. Rev. Biophys.*, 1978, **11**, 179.
19. M. T. Record, C. F. Anderson and T. M. Lohman, *Q. Rev. Biophys.*, 1978, **11**, 103.
20. K. A. Sharp and B. Honig, *Annu. Rev. Biophys. Biophys. Chem.*, 1990, **19**, 301.
21. C. F. Anderson and M. T. Record Jr., *Annu. Rev. Phys. Chem.*, 1995, **46**, 657.
22. J. M. Schurr, 2007.
23. V. Swaminathan and M. Sundaralingam, *CRC Crit. Rev. Biochem.*, 1979, **6**, 245.
24. L. McFail-Isom, C. Sines and L. D. Williams, *Current Op. Struct. Biol.*, 1999, **9**, 298.
25. L. D. Williams and L. J. Maher, *Annu. Rev. Biophys. Biomol. Struct.*, 2000, **29**, 497.
26. J. Sponer, J. Leszczynski and P. Hobza, *Biopolymers*, 2001, **61**, 3.
27. N. V. Hud and M. Polak, *Curr. Opin. Struct. Biol.*, 2001, **11**, 293.
28. J. A. Subirana and M. Soler-Lopez, *Annu. Rev. Biophys. Biomol. Struct.*, 2003, **32**, 27.
29. D. E. Draper and V. K. Misra, *Nat. Struct. Biol.*, 1998, **5**, 927.
30. D. E. Draper, *RNA*, 2004, **10**, 335.
31. B. Onoa and I. Tinoco Jr, *Curr. Opin. Struct. Biol.*, 2004, **14**, 374.
32. D. E. Draper, D. Grilley and A. M. Soto, *Annu. Rev. Biophys. Biomol. Struct.*, 2005, **34**, 221.
33. P. F. Agris, *Prog. Nucleic Acid Res. Mol. Biol.*, 1996, **53**, 79.
34. W. G. Scott and A. Klug, *Trends Biochem. Sci.*, 1996, **21**, 220.
35. A. M. Pyle, *J. Biol. Inorg. Chem.*, 2002, **7**, 679.
36. M. D. Been, *Curr. Top. Microbiol. Immunol.*, 2006, **307**, 47.

37. D. J. Klein, T. M. Schmeing, P. B. Moore and T. A. Steitz, *EMBO J.*, 2001, **20**, 4214.
38. J. S. Richardson, B. Schneider, L. W. Murray, G. J. Kapral, R. M. Immormino, J. J. Headd, D. C. Richardson, D. Ham, E. Hershkovits, L. D. Williams, K. S. Keating, A. M. Pyle, D. Micallef, J. Westbrook and H. M. Berman, *RNA*, 2008, **14**, 1.
39. D. W. Celander and T. R. Cech, *Science*, 1991, **251**, 401.
40. J. A. Latham and T. R. Cech, *Science*, 1989, **245**, 276.
41. J. R. Williamson, M. K. Raghuraman and T. R. Cech, *Cell*, 1989, **59**, 871.
42. N. V. Hud, P. Schultze, V. Sklenar and J. Feigon, *J. Mol. Biol.*, 1999, **285**, 233.
43. S. Neidle and G. N. Parkinson, *Curr. Opin. Struct. Biol.*, 2003, **13**, 275.
44. S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd and S. Neidle, *Nucleic Acids Res.*, 2006, **34**, 5402.
45. M. L. Gill, S. A. Strobel and J. P. Loria, *Nucleic Acids Res.*, 2006, **34**, 4506.
46. P. Podbevsek, N. V. Hud and J. Plavec, *Nucleic Acids Res.*, 2007, **35**, 2554.
47. H. E. Moser and P. B. Dervan, *Science*, 1987, **238**, 645.
48. J. C. Francois, T. Saison-Behmoaras and C. Helene, *Nucleic Acids Res.*, 1988, **16**, 11431.
49. R. H. Shafer, *Prog. Nucleic Acid Res. Mol. Biol.*, 1998, **59**, 55.
50. J. L. Leroy, K. Gehring, A. Kettani and M. Gueron, *Biochemistry*, 1993, **32**, 6019.
51. D. E. Gilbert and J. Feigon, *Curr. Opin. Struct. Biol.*, 1999, **9**, 305.
52. W. C. Winkler and R. R. Breaker, *Annu. Rev. Microbiol.*, 2005, **59**, 487.
53. R. L. Coppins, K. B. Hall and E. A. Groisman, *Curr. Opin. Microbiol.*, 2007, **10**, 176.
54. S. H. Kim, F. L. Suddath, G. J. Quigley, A. McPherson, J. L. Sussman, A. H. Wang, N. C. Seeman and A. Rich, *Science*, 1974, **185**, 435.
55. A. Jack, J. E. Ladner and A. Klug, *J. Mol. Biol.*, 1976, **108**, 619.
56. P. Brion and E. Westhof, *Annu. Rev. Biophys. Biomol. Struct.*, 1997, **26**, 113.
57. I. Tinoco Jr. and C. Bustamante, *J. Mol. Biol.*, 1999, **293**, 271.
58. V. K. Misra and D. E. Draper, *J. Mol. Biol.*, 2000, **299**, 813.
59. P. E. Cole, S. K. Yang and D. M. Crothers, *Biochemistry*, 1972, **11**, 4358.
60. K. Takamoto, R. Das, Q. He, S. Doniach, M. Brenowitz, D. Herschlag and M. R. Chance, *J. Mol. Biol.*, 2004, **343**, 1195.
61. X. Shui, C. Sines, L. McFail-Isom, D. VanDerveer and L. D. Williams, *Biochemistry*, 1998, **37**, 16877.
62. N. V. Hud, P. Schultze and J. Feigon, *J. Am. Chem. Soc.*, 1998, **120**, 6403.
63. N. V. Hud, V. Sklenar and J. Feigon, *J. Mol. Biol.*, 1999, **286**, 651.

64. I. D. Brown, *Acta Crystallogr. Sect. B.*, 1992, **48**, 553.
65. I. D. Brown, *Acta Crystallogr. Sect. B.*, 1988, **44**, 545.
66. S. Komeda, T. Moulaei, K. K. Woods, M. Chikuma, N. P. Farrell and L. D. Williams, *J. Am. Chem. Soc.*, 2006, **128**, 16092.
67. D. A. Doyle, J. Morais Cabral, R. A. Pfuetzner, A. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait and R. MacKinnon, *Science*, 1998, **280**, 69.
68. N. V. Hud, F. W. Smith, F. A. L. Anet and J. Feigon, *Biochemistry*, 1996, **35**, 15383.
69. F. H. Westheimer, *Science*, 1987, **235**, 1173.
70. C. W. Bock, A. K. Katz, G. D. Markham and J. P. Glusker, *J. Am. Chem. Soc.*, 1999, **121**, 7360.
71. C. C. Sines, L. McFail-Isom, S. B. Howerton, D. VanDerveer and L. D. Williams, *J. Am. Chem. Soc.*, 2000, **122**, 11048.
72. S. E. Rodriguez-Cruz, R. A. Jockusch and E. R. Williams, *J. Am. Chem. Soc.*, 1998, **120**, 5842.
73. C. W. Bock, A. Kaufman and J. P. Glusker, *Inorg. Chem.*, 1994, **33**, 419.
74. C. W. Bock, G. D. Markham, A. K. Katz and J. P. Glusker, *Theor. Chem. Acc.*, 2006, **115**, 100.
75. G. D. Markham, J. P. Glusker and C. W. Bock, *J. Phys. Chem. B*, 2002, **106**, 5118.
76. D. Bandyopadhyay and D. Bhattacharyya, *J. Biomol. Struct. Dyn.*, 2003, **21**, 447.
77. J. J. Cannone, S. Subramanian, M. N. Schnare, J. R. Collett, L. M. D'Souza, Y. Du, B. Feng, N. Lin, L. V. Madabusi, K. M. Muller, N. Pande, Z. Shang, N. Yu and R. R. Gutell, *BMC Bioinformatics*, 2002, **3**, 2.
78. J. A. Mears, M. R. Sharma, R. R. Gutell, A. S. McCook, P. E. Richardson, T. R. Caulfield, R. K. Agrawal and S. C. Harvey, *J. Mol. Biol.*, 2006, **358**, 193.
79. J. A. Mears, J. J. Cannone, S. M. Stagg, R. R. Gutell, R. K. Agrawal and S. C. Harvey, *J. Mol. Biol.*, 2002, **321**, 215.
80. P. B. Moore, *Annu. Rev. Biochem.*, 1999, **68**, 287.
81. N. B. Leontis and E. Westhof, *Curr. Opin. Struct. Biol.*, 2003, **13**, 300.
82. W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, 1984.
83. C. R. Woese, S. Winker and R. R. Gutell, *Proc. Natl. Acad. Sci. U. S. A.*, 1990, **87**, 8467.
84. C. Tuerk, P. Gauss, C. Thermes, D. R. Groebe, M. Gayle, N. Guild, G. Stormo, Y. d'Aubenton-Carafa, O. C. Uhlenbeck, I. Tinoco Jr., E. N. Brody and L. Gold, *Proc. Natl. Acad. Sci. U. S. A.*, 1988, **85**, 1364.
85. C. R. Woese and R. R. Gutell, *Proc. Natl. Acad. Sci. U. S. A.*, 1989, **86**, 3119.
86. C. C. Correll and K. Swinger, *RNA*, 2003, **9**, 355.
87. N. B. Leontis and E. Westhof, *J. Mol. Biol.*, 1998, **283**, 571.
88. P. Vallurupalli and P. B. Moore, *J. Mol. Biol.*, 2003, **325**, 843.

89. C. C. Correll, J. Beneken, M. J. Plantinga, M. Lubbers and Y. L. Chan, *Nucleic Acids Res.*, 2003, **31**, 6806.
90. A. A. Szewczak and P. B. Moore, *J. Mol. Biol.*, 1995, **247**, 81.
91. B. Wimberly, G. Varani and I. Tinoco Jr., *Biochemistry*, 1993, **32**, 1078.
92. T. A. Goody, S. E. Melcher, D. G. Norman and D. M. Lilley, *RNA*, 2004, **10**, 254.
93. S. Matsumura, Y. Ikawa and T. Inoue, *Nucleic Acids Res.*, 2003, **31**, 5544.
94. C. Hsiao, S. Mohan, E. Hershkovitz, A. Tannenbaum and L. D. Williams, *Nucleic Acids Res.*, 2006, **34**, 1481.
95. E. Hershkowitz, G. Sapiro, A. Tannenbaum and L. D. Williams, *IEEE/ACM Trans. Comp. Biol. Bioinformatics*, 2006, **3**, 33.
96. E. Hershkovitz, E. Tannenbaum, S. B. Howerton, A. Sheth, A. Tannenbaum and L. D. Williams, *Nucleic Acids Res.*, 2003, **31**, 6249.
97. A. S. Petrov, G. R. Pack and G. Lamm, *J. Phys. Chem. B*, 2004, **108**, 6072.
98. L. Rulisek and J. Šponer, *J. Phys. Chem. B*, 2003, **107**, 1913.
99. N. Gresh, J. E. Šponer, N. Spackova, J. Leszczynski and J. Šponer, *J. Phys. Chem. B*, 2003, **107**, 8669.
100. A. S. Petrov, G. Lamm and G. R. Pack, *J. Phys. Chem. B*, 2002, **106**, 3294.
101. M. M. Khan and A. E. Martell, *J. Am. Chem. Soc.*, 1966, **88**, 668.
102. L. D. Williams. Between Objectivity and Whim: Nucleic Acid Structural Biology. In *Top. Curr. Chem.*, J. B. Chaires, M. Waring, Eds., Springer: Heidelberg, 2005, **253**, 77.
103. X. Shui, L. McFail-Isom, G. G. Hu and L. D. Williams, *Biochemistry*, 1998, **37**, 8341.
104. S. B. Howerton, C. C. Sines, D. VanDerveer and L. D. Williams, *Biochemistry*, 2001, **40**, 10023.
105. J. P. Manners, K. G. Morallee and R. J. P. Williams, *J. Chem. Soc., Chem. Commun.*, 1970, 965.
106. F. J. Kayne, *Arch. Biochem. Biophys.*, 1971, **143**, 232.
107. R. L. Post, S. Kume, T. Tobin, B. Orcutt and A. K. Sen, *J. Gen. Physiol.*, 1969, **54**, 306.
108. C. E. Inturrissi, *Biochim. Biophys. Acta*, 1969, **178**, 630.
109. C. E. Inturrissi, *Biochim. Biophys. Acta*, 1969, **173**, 567.
110. J. S. Britten and M. Blank, *Biochim. Biophys. Acta*, 1968, **159**, 160.
111. J. Reuben and F. J. Kayne, *J. Biol. Chem.*, 1971, **246**, 6227.
112. P. A. Pedersen, J. M. Nielsen, J. H. Rasmussen and P. L. Jorgensen, *Biochemistry*, 1998, **37**, 17818.
113. V. Villeret, S. Huang, H. J. Fromm and W. N. Lipscomb, *Proc. Natl. Acad. Sci. U.S.A.*, 1995, **92**, 8916.
114. J. P. Loria and T. Nowak, *Biochemistry*, 1998, **37**, 6967.
115. S. Basu, A. A. Szewczak, M. Cocco and S. A. Strobel, *J. Am. Chem. Soc.*, 2000, **122**, 3240.

116. C. Caceres, G. Wright, C. Gouyette, G. Parkinson and J. A. Subirana, *Nucleic Acids Res.*, 2004, **32**, 1097.
117. T. Moulaei, T. Maehigashi, G. T. Lountos, S. Komeda, D. Watkins, M. P. Stone, L. A. Marky, J. S. Li, B. Gold and L. D. Williams, *Biochemistry*, 2005, **44**, 7458.
118. S. B. Howerton, A. Nagpal and L. D. Williams, *Biopolymers*, 2003, **69**, 87.
119. G. L. Conn, A. G. Gittis, E. E. Lattman, V. K. Misra and D. E. Draper, *J. Mol. Biol.*, 2002, **318**, 963.
120. J. Badger, Y. Li and D. L. Caspar, *Proc. Natl. Acad. Sci. U.S.A.*, 1994, **91**, 1224.
121. J. Badger, A. Kapulsky, O. Gursky, B. Bhyravbhatla and D. L. Caspar, *Biophys. J.*, 1994, **66**, 286.
122. H. S. Gill and D. Eisenberg, *Biochemistry*, 2001, **40**, 1903.
123. V. Tereshko, G. Minasov and M. Egli, *J. Am. Chem. Soc.*, 1999, **121**, 3590.
124. V. Tereshko, C. J. Wilds, G. Minasov, T. P. Prakash, M. A. Maier, A. Howard, Z. Wawrzak, M. Manoharan and M. Egli, *Nucleic Acids Res.*, 2001, **29**, 1208.
125. J. Eisinger, R. G. Shulman and B. M. Szymansk, *J. Chem. Phys.*, 1962, **36**, 1721.
126. J. Eisinger, F. Fawazest and R. G. Shulman, *J. Chem. Phys.*, 1965, **42**, 43.
127. J. Reuben and M. Cohn, *J. Biol. Chem.*, 1970, **245**, 6539.
128. A. L. Feig, *Met. Ions Biol. Sys.*, 2000, **37**, 157.
129. P. S. Salgado, M. A. Walsh, M. R. Laurila, D. I. Stuart and J. M. Grimes, *Acta Crystallogr., Sect D: Biol. Crystallogr.*, 2005, **61**, 108.
130. S. Rudisser and I. Tinoco, *J. Mol. Biol.*, 2000, **295**, 1211.
131. W. H. Braunlin, C. F. Anderson and M. T. Record Jr., *Biochemistry*, 1987, **26**, 7724.
132. D. Sen and D. M. Crothers, *Biochemistry*, 1986, **25**, 1495.
133. R. V. Gessner, G. J. Quigley, A. H.-J. Wang, G. A. van der Marel, J. H. van Boom and A. Rich, *Biochemistry*, 1985, **24**, 237.
134. M. Brannvall, N. E. Mikkelsen and L. A. Kirsebom, *Nucleic Acids Res.*, 2001, **29**, 1426.
135. J. A. Cowan, *J. Inorg. Biochem.*, 1993, **49**, 171.
136. D. Porschke, *Biophys. Chem.*, 1976, **4**, 383.
137. D. Porschke, *Nucleic Acids Res.*, 1979, **6**, 883.
138. P. R. Schimmel and A. G. Redfield, *Annu. Rev. Biophys. Bioeng.*, 1980, **9**, 181.
139. V. A. Buckin, B. I. Kankiya, D. Rentzeperis and L. A. Marky, *J. Am. Chem. Soc.*, 1994, **116**, 9423.
140. S. L. Heilman-Miller, J. Pan, D. Thirumalai and S. A. Woodson, *J. Mol. Biol.*, 2001, **309**, 57.
141. V. K. Misra and D. E. Draper, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 12456.

142. V. K. Misra and D. E. Draper, *J. Mol. Biol.*, 2002, **317**, 507.
143. F. Cesare Marincola, V. P. Denisov and B. Halle, *J. Am. Chem. Soc.*, 2004, **126**, 6739.
144. K. Woods, L. McFail-Isom, C. C. Sines, S. B. Howerton, R. K. Stephens and L. D. Williams, *J. Am. Chem. Soc.*, 2000, **122**, 1546.
145. F. H. Allen, S. E. Harris and R. Taylor, *J. Comput. Aided Mol. Des.*, 1996, **10**, 247.
146. R. Taylor, *Acta Crystallogr., Sect D: Biol. Crystallogr.*, 2002, **58**, 879.
147. R. Taylor, O. Kennard and W. Versichel, *J. Am. Chem. Soc.*, 1983, **105**, 5761.
148. R. Taylor, O. Kennard and W. Versichel, *J. Am. Chem. Soc.*, 1984, **106**, 244.
149. H. B. Burgi, *Inorg. Chem.*, 1973, **12**, 2321.
150. H. B. Burgi, J. D. Dunitz and E. Shefter, *J. Am. Chem. Soc.*, 1973, **95**, 5065.
151. F. H. Allen, R. Mondal, N. A. Pitchford and J. A. K. Howard, *Helv. Chim. Acta*, 2003, **86**, 1129.
152. M. Sundaralingam and Y. C. Sekharudu, *Science*, 1989, **244**, 1333.
153. F. A. Hays, A. Teegarden, Z. J. Jones, M. Harms, D. Raup, J. Watson, E. Cavaliere and P. S. Ho, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 7157.
154. H. B. Burgi, J. D. Dunitz, J. M. Lehn and G. Wipff, *Tetrahedron*, 1974, **30**, 1563.
155. J. M. Vargason, K. Henderson and P. S. Ho, *Proc. Natl. Acad. Sci. U.S.A.*, 2001, **98**, 7265.
156. P. J. Flory, *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY, 1953.
157. E. Koculi, D. Thirumalai and S. A. Woodson, *J. Mol. Biol.*, 2006, **359**, 446.
158. B. I. Kankia, *Biophys. Chem.*, 2003, **104**, 643.
159. M. T. Record, W. T. Zhang and C. F. Anderson, 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. *In Adv. Protein Chem.*, 1998; Vol. **51**; pp. 281.
160. A. A. Rashin and B. Honig, *J. Phys. Chem.*, 1985, **89**, 5588.