CHAPTER 2. The Tetrahymena Group I Intron

The Tetrahymena Group I intron occupies a special place in RNA biology. It was the first RNA molecule unequivocally shown to demonstrate intrinsic catalytic activity. It uses its catalytic activity to remove itself from its flanking ribosomal RNA, thus it is auto-catalytic, in a process known as self-splicing. Its discovery stimulated a re-investigation of RNA properties in other biological contexts, notably the ribosome and later the spliceosome, as well as smaller autocatalytic RNAs (ribozymes) and RNAs that bind ligands (riboswitches). It has led to attempts to construct all-RNA enzymes such as RNA polymerase that invigorated the RNA World hypothesis. The Tetrahymena Group I intron altered perceptions of RNA molecules and led to a new paradigm of RNA function.

Investigations into the structure and function of the Tetrahymena Group I intron have yielded insights into fundamental properties of all RNA molecules. New experimental methods have been developed to probe its properties, both bench and computational. Despite the enormous number of papers devoted to it, it is not a complete story. But because it is perhaps the most studied of RNA molecules – by biochemistry, genetics, bioinformatics, and biophysics – it provides the perfect basis for a general study of RNA Biophysics.

Biology and Chemistry

The Tetrahymena Group I intron must be removed from its site within ribosomal RNA, since its inclusion would effectively inactivate the rRNA. Leaving aside questions of why a 414 nucleotide insertion should be stuck in the middle of an essential rRNA gene, when the intron was removed, the 5’ and 3’ ends of the flanking rRNA sequence had to be precisely joined

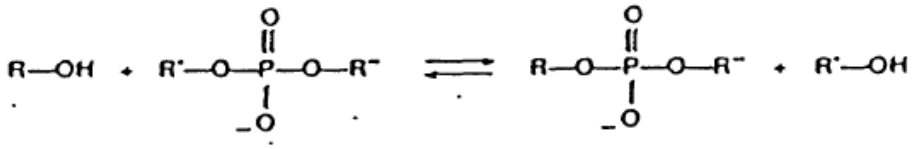
together (ligated) to create the functional form of the rRNA. The scheme could be conceptualized rather simply:

Figure 1. Tetrahymena Group I intron is in the rRNA gene. After transcription, it removes itself, leaving the ligated rRNA and the free intron. Green is upstream rRNA sequence, purple is downstream rRNA sequence. More generally, the flanking sequences are exons, and Group I intron excision is self-splicing.

The devil is in the details, however, and elucidation of the mechanism of intron excision has been the subject of more than 30 years of work. For perspective, this is the RNA that earned Tom Cech a Nobel Prize in 1989. It removes itself after the gene has been transcribed with no assistance from proteins. Although it's autocatalytic, it does require two cofactors: Mg^{2+} and guanosine. These cofactors are common to all Group I introns, independent of their site within a genome and independent of the organism (bacteriophage, dinoflagellate, or fungus)[1–4].

In our scheme, the intron is drawn as an independent element of the transcript, which anticipates its study as an autonomous molecule with catalytic capability. Also implicit in the concept of an autonomous catalytic RNA is that it must adopt a structure that creates its active site. The active site must be responsible for cleavage of the RNA at the junctions of the rRNA/intron. Having mediated the cleavage, the two rRNA strands must be correctly joined (ligated). Does ligation also occur in the active site? What and where is its active site?

The general chemistry of Group I excision/ligation was known long before the active site was identified. The chemistry is transesterification:



Two important points: this reaction requires no exogenous energy source, and it is reversible. This depiction is a classic chemical reaction, but for the intron, the definition of R, R′, and R″ will be unique.

Some definitions: the 5′ splice site is the junction at the 3′ end of the 5’ rRNA strand (the upstream sequence) and the 5′ end of the intron. The 3′ splice site is the junction of the 3′ end of the intron and the 5′ end of the downstream rRNA sequence (Figure 1). Clearly, to bring about rRNA ligation, the 5′ and 3′ splice sites must be in contact. This can only happen if the RNA folds into a 3D conformation that juxtaposes these two sites.

Of course, the classical transesterification reaction is depicted as a bimolecular reaction, and our intron/rRNA system is unimolecular. Here is where the cofactor enters: Guanosine. Guanosine, with its guanine base and ribose sugar, is necessary for the reaction to proceed[5]. A first hint of the mechanism came when the reaction was attempted with deoxy-guanosine, which acted as an inhibitor of cleavage[6]. Therefore, we’ll assign R-OH to rG_{OH} .

Referring to Figure 2, we see that R′ must be the upstream rRNA strand, and R″ must be the intron, thus defining our R′-O-P-O-R″ component. The guanosine cofactor rG_{OH} attacks the phosphodiester bond at the 5′ splice site (to use the language of chemistry, a nucleophile attacks during a nucleophilic displacement reaction). It frees the upstream rRNA (R′-OH), leaving it with a ribose with a normal 2′-OH and 3′-OH on its 3′ terminus. The rG_{OH} becomes covalently attached at the 5′ end of the intron through a normal 3′-5′ phosphodiester bond to create *rG*-O-P-O-R″. But this is only half the reaction!

Now there is another transesterification reaction: this time, the liberated R′-OH becomes our new R-OH, which will attack the 3’ splice site. Again using the scheme of the transesterification reaction, R′-O-P-O-R″, R′ is now [*rG*-intron] and R″ is the downstream rRNA sequence. The ribose 3′-OH at the end of R-OH becomes the attacking nucleophile at the 3′ splice, forming the ligated rRNA and liberating the intron. If this is confusing, Figure 2 might help, as it also illustrates another feature of Group I introns that is nearly invariant: the identity of the U nucleobase at the 5′ splice site and the final G at the 3′ end of the intron.

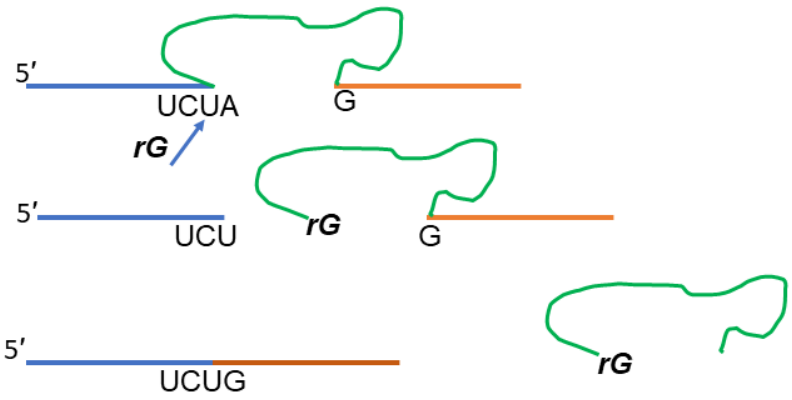


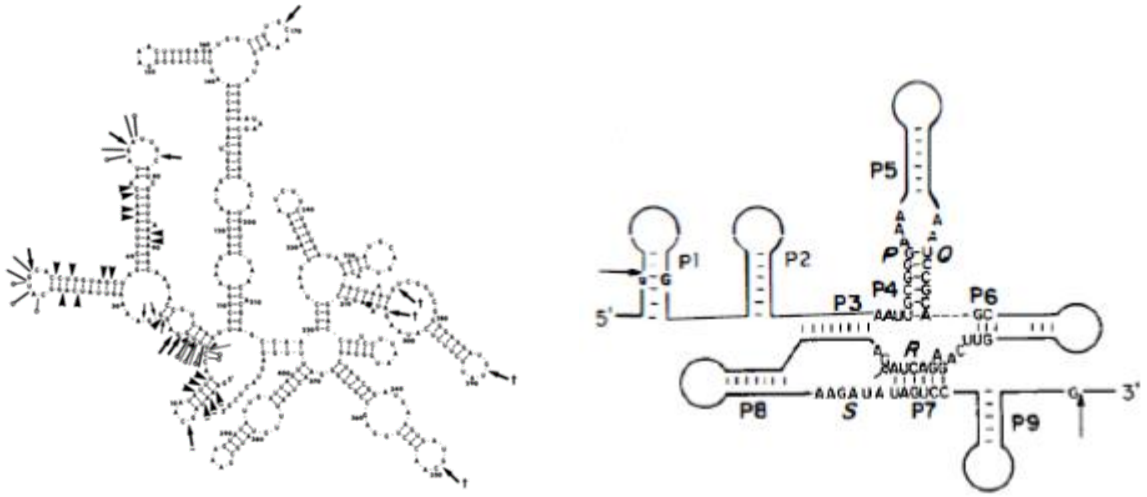
Figure 2. A schematic of the total reaction pathway leading to ligated rRNA and free Group I intron. There are two consecutive transesterification reactions (also known as nucleotidyl transfer reactions): the first results in a free upstream RNA, with concomitant attachment of the *rG* cofactor to the intron; the second produces the ligated rRNA and release of *rG*-intron.

Since guanosine molecules in a cell do not normally attack phosphodiester bonds of random RNA molecules, there must be something special about the active site of the Group I intron that allows (invites?) this reaction to occur. Not only is the reaction specific for Guanine (no other nucleobase will work), but nucleotidyl transfer reactions require a strict geometry of reactants. And, as we will see, the reactions require Mg^{2+} ions. With these necessary strictures, where is the active site of this self-splicing intron?

Group I structure/function

Elucidation of the conformations of the Group I intron, in pursuit of its active site and the mechanism of its reactions, has yielded a treasure-trove of general principles of RNA folding. Such principles include the thermodynamic basis of conformational transitions, the kinetics and thermodynamics of tertiary interactions, the chemical and structural basis of selective ligand binding, conformational selection, folding funnels, and hierarchical assembly. While data strongly suggested that the Tetrahymena Group I intron had structure, there was no guarantee that it had only one structure, or if its structure(s) had an inside/outside analogous to a protein's core rather than an amorphous spaghetti of single strands and short duplexes.

The story of structural elucidation of the Tetrahymena Group I intron can be visualized with these historical depictions (Figure 3). The intron was studied as an independent RNA, transcribed by T7 RNA polymerase from a plasmid and purified by standard methods. This method allows for mutations, insertions, and deletions to be readily introduced into the RNA through the DNA template. The purified intron was then subjected to a variety of solution conditions, where its self-cleavage was used to assay its activity. The assumption was that for self-cleavage to occur, the intron must have adopted its active conformation. It's worth keeping in mind that the structures shown in the Figure represent more than 30 years of work.



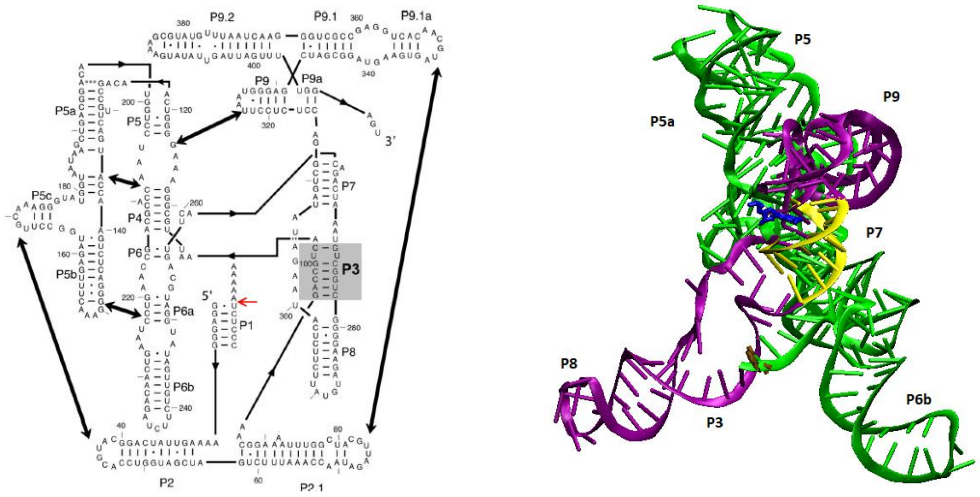


Figure 3. A quick history of Tetrahymena Group I structure determination. A. The combination of biochemical probing experiments with the *mfold* RNA folding algorithm produced this secondary structure model[7] in 1983. B. More probing and biochemistry, coupled with phylogenetic comparisons with known Group I introns produced this model in 1990. Duplexes are P1-P9, *PQRS* are conserved sequences that form the Core of the intron[8]. P1 is formed by the intron and exon, and contains the site of the first transesterification reaction (→). C. In this revised depiction[9] (1994), dark arrows identify the six long-range tertiary contacts. P456 and P5abc are shown as a distinct domain. The cleavage site in P1 is shown (←). D. An X-ray crystal structure of the intron shows how it can be conceptually separated into two domains[10] (2004). P7 (yellow) is the catalytic G-site, in blue is the terminal G414 that occupies the site. This construct lacks P2 and P1.

Tetrahymena intron secondary structure. The first step in the process of structural elucidation is determination of the secondary structure elements, which are then assembled into a 3D model representing the folded tertiary structure. The secondary structure could be probed by a number of biochemical and chemical reactions, under conditions where the intron was shown to be active. As illustrated in Figure 3A, the RNA was subjected to nuclease cleavage by specific RNases. Sites where cleavage occurred are indicated by arrows: cobra venom ribonuclease cleaves double-stranded RNA; ribonuclease T1 cleaves single-stranded RNA after rG's; ribonuclease S1 cleaves single strands; and ribonuclease T2 cleaves after rA's in single strands. Other probing methods were also employed: Fe(II)-EDTA cleavage, photochemical crosslinking, free radical •OH cleavage. And many mutations of the RNA were introduced for

comparative purposes in probing experiments. Results from probing experiments were augmented by secondary structure predictions. Chapter 3 will explain how experimentally generated thermodynamic parameters provide the basis for predicting the thermodynamic stability of RNA duplexes. The secondary structure in Figure 3A is one of several that were first proposed for Tetrahymena Group I intron.

There is a conspicuous simplification of the secondary structure in Figure 3B. By comparison with other Group I intron sequences (identified in a flurry of investigations of organisms from bacteriophage to fungi to vertebrates; they have never been found in vertebrates), a sequence alignment allowed the identification of conserved sites, and a consensus among investigators produced the designations of the duplex regions (P1-P9). Four regions: P, Q, R, and S, have conserved sequences; we see the ΩG at the 3′ terminus of the intron; as well as the 3′ U of the upstream rRNA sequence. Now, we also see that the 3′ U is in the P1 duplex, where it is base paired with a G, to form a G:U wobble pair. The exon sequence contains the conserved U, and the flanking sequences are complementary to the Internal Guide Sequence of the intron. The IGS is unique to each intron, and it can be altered, provided that the G:U pair is preserved. The stability of the duplex is important for efficient splicing[11–13]. The sequence conservation among Group I introns allowed identification of a conserved 'Core', which had immediate implications for their active site.

Comparing the secondary structures in 3B and 3C shows that the helical regions are preserved, but they are re-positioned within the sequence. This reformulation of the secondary structure came about as a result of experiments that physically separated the intron into two pieces: the P456 domain was removed, and the other half was engineered as an independent domain. Phylogenetically, this maneuver was justified, since the P5 domain is variable: many Group I introns lack the P5abc extension. Without P456, the Tetrahymena intron is not active,

but in a striking example of modular RNA construction, addition of P456 to P3789 restored catalytic activity.

These two domains of the Tetrahymena Group I intron will be considered separately for what they reveal about RNA properties. The biophysics of P456 alone are key to its folded structure, stability, and interactions with P3-P9. It will be a model system for tertiary interactions that are common to many folded RNAs and will be explored in greater detail. P1 contains the site of the first nucleotidyl transfer reaction, and how it uses its sequence to identify this site is another lesson in structure/function. P2 in Tetrahymena Group I is very long, and as shown in Figure 4C, there are two long-range tertiary interactions that anchor it to the body of the intron. However, P2 is not necessary for the activity, and it can be deleted. It is not included in constructs for crystallization, since it appears to be unstably attached to the rest of the intron. P9 contains the ΩG, which is the target of the second nucleophilic attack to release the intron. P7 is the binding site for Guanosine and thus it is the active site of catalysis. The two domains will be considered together again in discussions of RNA folding.

The G site. Binding of the Guanosine cofactor is an example of ligand binding by an RNA molecule. The intron forms a binding site that is specific for this ligand cofactor, and it provides the first example we will consider of this phenomenon. Riboswitches, to be considered later, have evolved mechanisms to specifically recognize ligands, and we are anticipating their biology by consideration of the intron.

Two observations provided the rationale in the search for the G binding site of the intron. First, all known Group I introns used Guanosine as a cofactor. There is also a Guanosine at the 3′ terminus of all Group I introns, positioned downstream of P9. Then, phylogenetic comparisons of all known Group I introns found a conserved Core sequence composed of four short duplexes (*P,Q,R,S* in Figure 3B). F. Michel reasoned that the terminal ΩG would be

proximal to J7/9, the Junction between P7 and P9, and that it would bind in a site common to all Group I introns[14]. Within the conserved P7 is the conserved G264:C311 base pair, which mutagenesis had shown was required for cleavage activity. Could this base pair be part of the Guanosine recognition site?

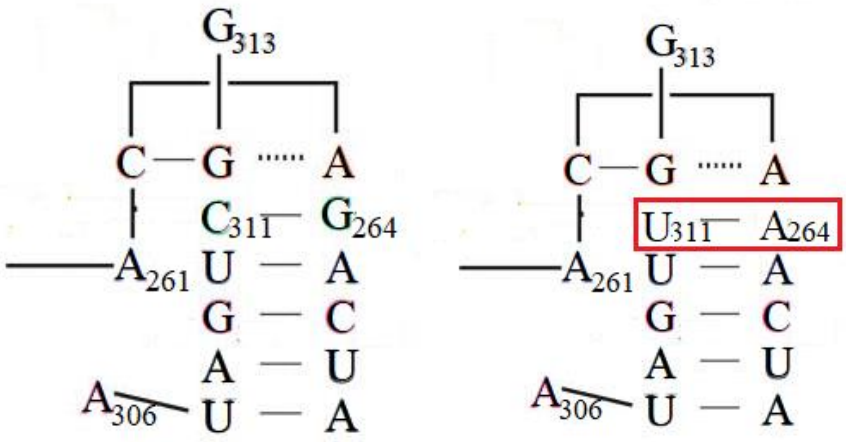
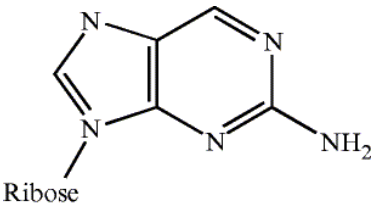


Figure 4. The P7 sequence from Tetrahymena Group I intron. The wild type sequence G264-C311 was replaced with A264-U311 to test a model of ligand specificity[14].
The A264-U311 mutation is not active in the first transesterification reaction when Guanosine is the cofactor.

The critical experiment was to use an alternative cofactor: 2-aminopurine ribonucleoside for the A264:U311 intron. This change in the specificity of cofactor recognition allowed the first step of the splicing reaction to proceed: cleavage at the 5′ splice site was recovered, although cleavage at the 3′ splice site was blocked. (Why?) These results were consistent with recognition of the cofactor by formation of a base triple in the major groove of P7.



2-aminopurine

To Think About: There is more to this story than formation of this base triple interaction, since Guanosines do not typically interact with G-C base pairs in a normal A-form duplex. Much less do they form high affinity binding interactions that lead to phosphodiester cleavage. Are there clues that suggest P7 is in any way unusual as an RNA duplex?

Other experimental results that are important clues: deoxy-G binds only weakly; nucleotide at the position equivalent to A263 is always either an A or C. There is only one binding site: both the exogenous Guanosine cofactor and the terminal ΩG bind to this P7 site.

Thinking about the ligand binding site is the segue to consideration of the three-dimensional structure of the intron. Intron tertiary structure was central to its catalytic ability, but even while its secondary structure had been mapped by experiment and phylogenetic comparison, its tertiary structure was not known. Many investigators have used many methods to probe the tertiary fold of the intron at equilibrium as well as the kinetics of its folding; keep in mind that while there was consensus regarding its secondary structure by 1990, there was no x-ray crystal structure of the intron until 2004, twenty-two years after its discovery. That makes it all the more remarkable that the structure predicted by Michel and Westhof in 1990[15] (Figure 5) has proved to be so accurate in its details.

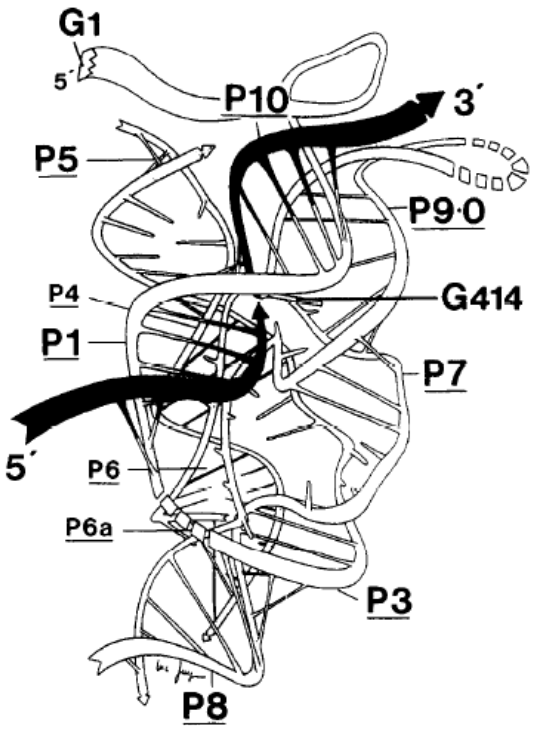


Figure 5. The Michel and Westhof model of the core of Tetrahymena Group I intron. It lacks P5abc.

Exercise for the intrigued scholar: In their 1990 paper, Michel and Westhof explain how they used biochemical results and phylogenetic comparisons to produce their model of the Tetrahymena Group I core structure. They predict previously unknown tertiary interactions and anticipate their general utility in folding RNAs. Their discussion of how the active site must be configured to accommodate the two transesterification reactions is prescient.

Search for the Tertiary structure

Group I introns have a common catalytic mechanism contained within their common core and they all use Guanosine as a cofactor. However equally important in appreciation of their activity is formation of their active conformation, i.e. their tertiary fold. Now is the time to think about how they adopt their folded structure. Two properties of RNA will be critical here: one is how the Intron uses divalent ions to adopt its folded structure, and the other is tertiary interactions between its secondary structure elements that guide and stabilize its fold. We will start with the ion dependence, since it will be a universally applicable to RNAs, and tackle Group I intron folding in a subsequent chapter.

In cells, Mg^{2+} is the most abundant divalent ion, and RNAs use it liberally to achieve their folded conformations. The Intron is no exception in its need for divalent ions to adopt its tertiary structure, but it also needs Mg^{2+} for its catalytic activity. The dual role of Mg^{2+} was discovered early in the history of the Intron, but the mechanics of its interactions during folding of the Intron have been investigated through many subsequent kinetics experiments.

The Intron was known to require Mg^{2+} to undergo splicing, but was this requirement due to its chemistry or formation of its active conformation? To examine ion-dependence of the intron structure, several investigators utilized the chemistry of Fe(II)-EDTA to probe its structure.

In solution, Fe(II)-EDTA was shown to cleave the backbone of DNA by attacking the deoxyribose and thus breaking the strand[16]. The chemistry is generation of a hydroxyl radical that reacts with the sugar at those sites where it is accessible to the reagent. The resulting reaction leads to strand scission. The constraint of physical accessibility led Latham and Cech[17] to use it to address a fundamental property of the Intron: Does it have an inside and outside when it's folded?

After first demonstrating the proof of principle on $tRNA^{PHE}$ with and without Mg^{2+}, (the only known folded RNA structure at the time), they applied it to a Tetrahymena intron construct which lacked the 5' and 3' splice sites. Their results are illustrated in Figure 8, mapped onto the

secondary structure of the intron. Those sequences in shaded areas became protected from Fe(II)-EDTA cleavage in the presence of sufficient Mg^{2+}, representing ~40% of the nucleotides in the Intron. What features of an RNA would protect its ribose from hydroxyl radical cleavage? Based on their tRNA experiments, the authors concluded that a higher order structure of the Intron was present that created an interior where the reagent could not penetrate. They concluded that the folded intron had an inside and outside, and that Mg^{2+} was required for its structure.

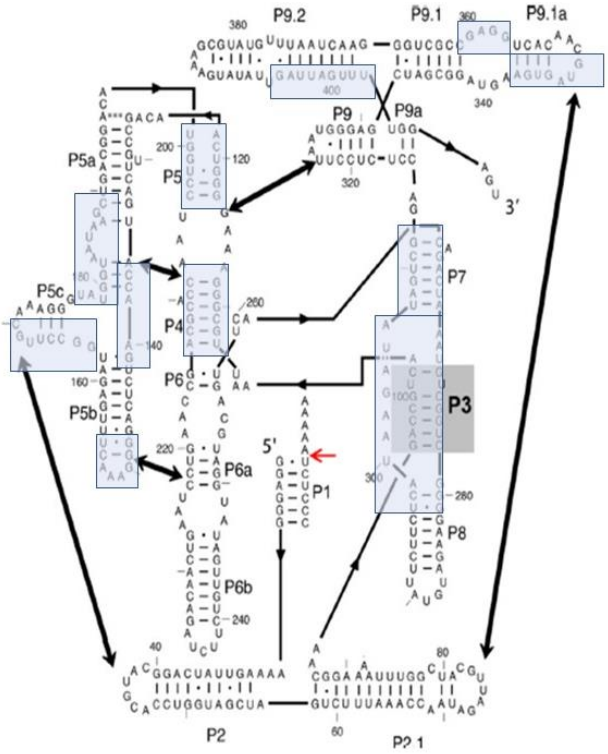


Figure 6. Sequences protected from Fe(II)-EDTA cleavage in the presence of Mg^{2+}. (adapted from Latham & Cech, 1989). Shaded boxes are protected sites, presumably forming the interior. Can you pack these regions into a 3D structure?

A subsequent study of divalent ions and the Intron revealed a general feature of its folding requirements[18]. Again using Fe(II)-EDTA as a probe of the folded structure, Ca^{2+} and Sr^{2+} were found to substitute for Mg^{2+} to stabilize the folded RNA. However, these ions were not equally effective in supporting the tertiary fold. As illustrated in this plot of those experimental

data, the most obvious differences among the ions are the midpoints of the transitions. This was one of the first experiments that mapped the divalent ion dependence of the conformational transition from secondary structure to tertiary structure.

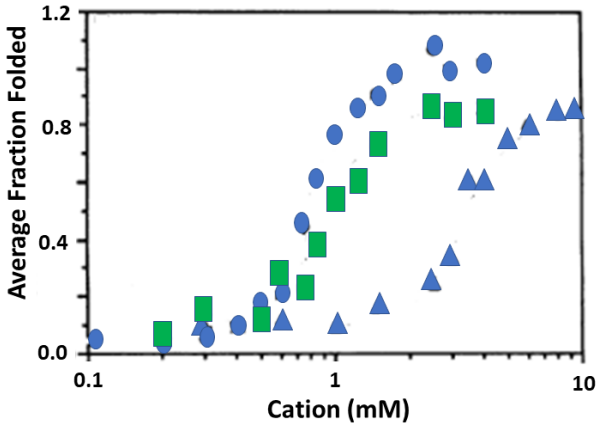


Figure 7. Mapping the transition from secondary to tertiary structure of the Tetrahymena Group I intron as a function of divalent ion concentration. Circles: Mg^{2+}; squares Ca^{2+}, triangles Sr^{2+}. Buffer contained 2 mM Tris-HCl, pH 7.5, 42° C. Adapted from Celander & Cech[18].

This study of ion-dependent tertiary structure folding also revealed a key to the catalytic reaction. Although all three ions were able to fold the intron, it was catalytically active only in the presence of Mg^{2+}, indicating that Mg^{2+} was part of the catalytic chemistry. From the perspective of the tertiary structure, the data indicate that Mg^{2+} has some unique properties that facilitate the transition. Is the difference in the midpoint of the transition a measure of the number of divalent ions required to effect the conformational change? What properties of the ions affect their association with RNA? Such studies are the impetus of a discussion of the unique properties of Mg^{2+} ions and how they interact with RNA.

Producing an X-ray crystal structure of the Tetrahymena Group I intron was very difficult and took many years. To construct an RNA that was amenable to high resolution crystallography, several mutations were introduced that stabilized the global structure[19]. The

long P2 which forms long-range tertiary interactions with both P9 and P5a was removed, for experiments had indicated that its position was not stable within the tertiary structure. Also, the crystallographic construct lacked P1 with its 5′ splice site. This X-ray crystal structure [PDB ID 1X8W] of the Tetrahymena Group I intron is shown.

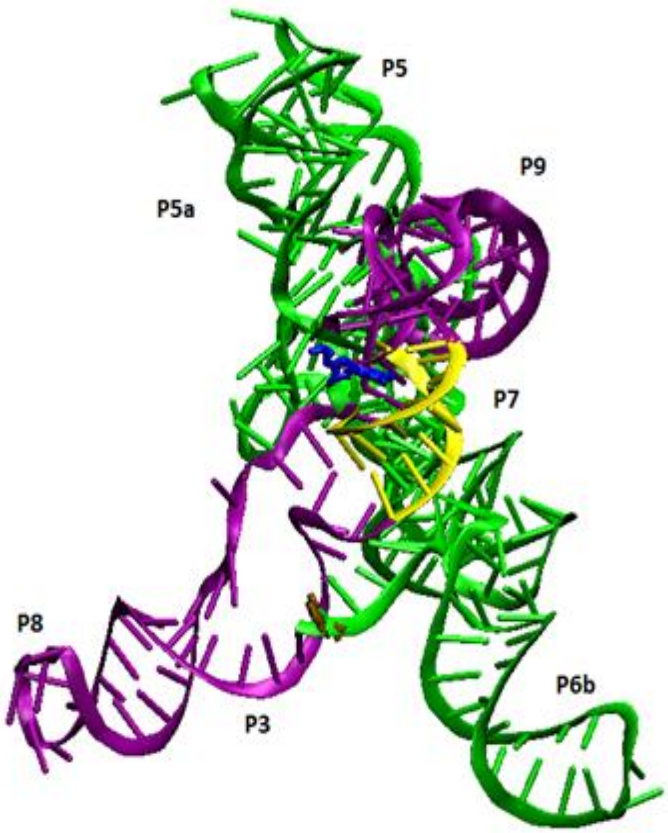


Figure 7. A crystal structure of the Tetrahymena Group I intron at 5 Å resolution [PDB ID 1X8W]. The P456 domain is green, and P3-9 is purple, to emphasize how these two domains are separable. The two domains interact through hydrogen bonding at the interface that includes P7. This crystal structure lacks P1 and P2.[19]

This was the first structure of the intron to reveal the juxtaposition of its ΩG with P7, and indeed it does make a base triple with G264:C311 as predicted. Strikingly it reveals a more intricate ligand binding site: a binding site for ΩG_{414} included stacked nucleobases and a total of three base triples: ΩG414••G264-C311, A263••C262-G312, and A261••U310-A265. In

examination of this structure in Figure 8, note the convoluted trace of the backbone of both strands.

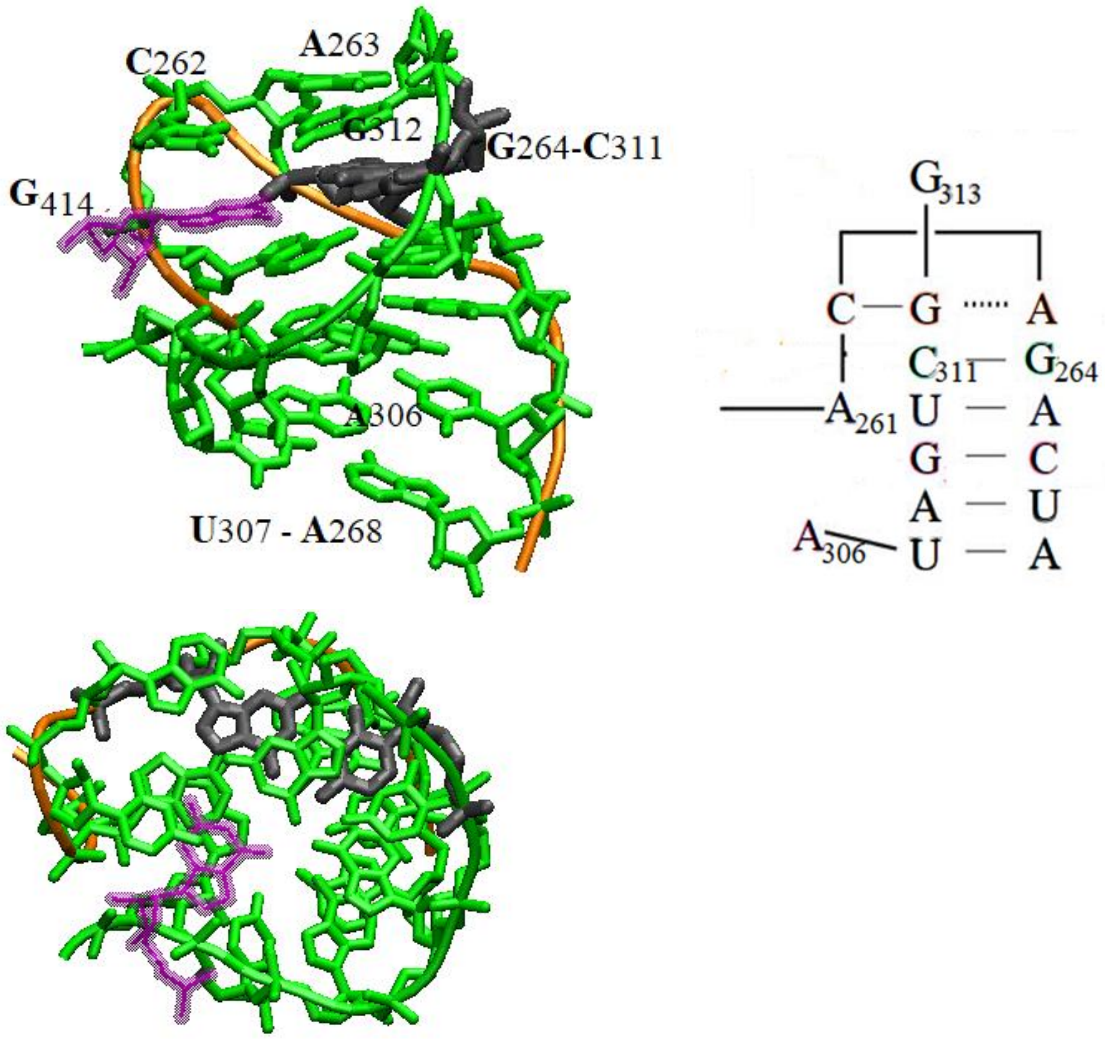


Figure 8. Structures of the Guanosine cofactor bound to the P7 duplex of the Tetrahymena Group I intron (PDB 1x8w), with a diagram for reference. The 261-268 strand backbone is colored orange; the 306-312 strand backbone is green. The C311-G264 base pair is gray, and the ΩG414 is purple. Readers are encouraged to spend time gazing at the intricacies of the G binding site with their favorite visualization tools; this picture was made with VMD.

There are additional Group I crystal structures (reviewed in Vicens & Cech[20]), including a structure of the shorter Azoarcus Group I intron (PDB 1u6b). The Azoarcus crystal structure

differs in several features from the Tetrahymena structure, notably by its inclusion of P1 and sequences from the downstream exon. It also includes P2, revealing tertiary interactions. We will use it is subsequent discussion, but the reader is encouraged to examine it, especially with respect to its Guanosine binding site.

So, to review the Tetrahymena Group I self-splicing. The intron/exon folds into a conformation that includes formation of P1. This duplex then docks into the core of the intron, which is comprised of P3, P7, and P9. P1 is held in place by tertiary interactions, mostly hydrogen bonds from riboses. The base triples help to constrain the exogenous Guanosine for its transesterification reaction with P1. P1 must be released and replaced by ΩG at the junction of P7/P9, which is again held in place by the network of stacked nucleobases. Following the second transesterification reaction, the ligated exons are released and the intron is free. Mg^{2+} ions are required for the chemistry, and divalent ions are needed for tertiary structure formation.

Mg^{2+} ions in RNA structures

Nothing about the interactions of cations with the RNA polyanion is simple. Why are divalent ions necessary, and why don't monovalent ions substitute? This question leads directly to a review of the properties of cations, and because Mg^{2+} is the dominant biological divalent cation for RNA structure/function, we will focus on its properties.

The inorganic chemistry of the Mg^{2+} metal ion is important to appreciate, since it determines its interaction with RNA[21]. Its ionic radius of 0.65 Å is much smaller than its hydrated radius of 4.76 Å, making it unique among divalent ions. Correspondingly, its volumes are dramatically different: 1.2 $Å^3$ vs 453 $Å^3$. In solution, Mg^{2+} is invariably a hexahydrate with octahedral geometry; its six water ligands are held in a rigid orientation. The enthalpy of hydration is large (-450 kcal/mol) but an important feature of Mg^{2+} is its ability to exchange a

water molecule for another ligand. The water exchange rate is estimated at 10^5 s^{-1}, or a lifetime of 10 μs. The ion has an inner and outer coordination shell, and waters in the first coordination shell are acidic. Their oxygens are proximal to the ion with their hydrogens directed outward. The water dipole moments are pointed towards the metal.

Hydrated Mg^{2+} ions interact with RNA in two modes[22]. One is simply electrostatics, where the hexahydrated ion with its +2 charge can efficiently neutralize the -1 charge on a phosphate. In its mode of charge neutralization, the Mg^{2+} ion diffuses around the RNA surface, but can also enter the RNA interior where charge density of the phosphates is high. The Mg^{2+} ions might be necessary to allow the RNA to fold, but these counterion are free to exchange with bulk ions; they are not 'bound' to the RNA[23]. This mode is non-specific screening, which has the effect of reducing the charge density of the RNA (considering the RNA as a strong polyelectrolyte, and the ion as small electrolyte). This regime is dominated by long-range electrostatics and is a function of ionic strength.

A caveat here: while ionic strength and ion concentration are sometimes used interchangeably, they do not describe the same solution conditions. Ion concentration is typically in molar (M), millimolar (mM), or micromolar (μM): a solution might be 150 mM in KCl, 2 mM $MgCl_2$. The same solution would have an ionic strength of $I = \frac{1}{2} \sum_{i=1}^{n} c_i z_i^2$ where c is the molar concentration of each ion and z is its charge. So our solution is $I = \frac{1}{2} [K^+ + Cl^- + Mg^{2+} + 2Cl^-] = \frac{1}{2} [0.15(+1)^2 + 0.15(-1)^2 + 0.002(+2)^2 + 0.004(-1)^2] = 0.156$ M or 156 mM. Not the same conditions!

In contrast, some Mg^{2+} ions can be bound to an RNA with high affinity at specific sites. This binding is ion-specific and cannot be described by ionic strength. The most common example of site-specific Mg^{2+} binding involves a nonbridging phosphate oxygen replacing a water molecule in the first shell of the ion[24]. Exchanging a first shell water with a phosphate

oxygen has no enthalpic cost and is entropically favored by release of the bound water. This molecular exchange is additionally favored since phosphate groups (OP) are more polarizable than water molecules. When a phosphate oxygen enters the first shell of Mg^{2+}, its electron density moves to the metal, leaving its Phosphorus to act as a nucleophile. One Mg^{2+} ion can exchange two waters for two phosphates (a bidentate state) or even three phosphates for three waters (tridentate). Phosphate atoms in the Mg^{2+} first shell are very closely packed (Figure 9). This is possible since charge transfer from O (the lone pairs of the oxyanion) to the cation reduces the charge repulsion between two bound OP atoms, stabilizing the complex. The ability to accommodate OP atoms in very close proximity is unique to Mg^{2+}; Mn^{2+} or Ca^{2+} does not allow this close juxtaposition. A quantum mechanical calculation of bidentate interaction energies* shows decreasing values for Mg^{2+} > Ca^{2+}>Na+[25] and as indicated previously, the close packing of the nonbridging phosphate oxygens reduced the distance from metal ion to ligand.

*(calculations included basis set superposition error (BSSE), using B3LYP/6-311g++(d,p)//B3LYP/6-311g++(d,p)[cpcm].)

When two or more waters are replaced with RNA phosphates, the ion is said to be chelated to the RNA. This is a bound form of the ion. Chelated Mg^{2+} ions are essential for stabilizing folded RNAs.

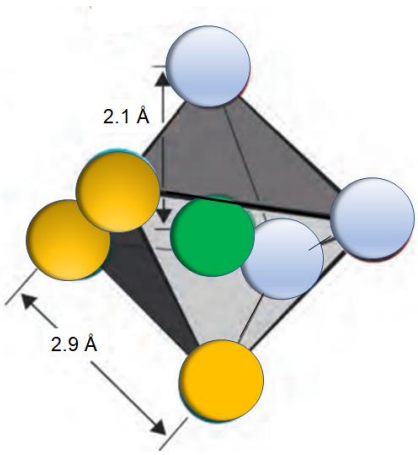


Figure 9. The green Mg^{2+} ion is surrounded by three waters (blue) and three OP atoms (gold) in its octahedral symmetry. This is example of a tri-chelated Mg^{2+} is from a crystal structure of the *H. marismortui* 23S rRNA in its large ribosomal subunit (2.4 Å resolution, PDB id 1JJ2). Adapted from Bowman et al.[22]

Sites of chelated Mg^{2+} ions in an RNA structure are idiosyncratic. Most have been identified from crystal structures of a folded RNA. A caution from these identifications is that in structures with resolution > 2 Å, a Mg^{2+} ion can be confused with a water molecule or a Na^{+} or K^{+} ion. Often the clue to the identify of a Mg^{2+} comes from proximity of the ligand to the ion, based on the distances expected from an atom in the inner shell. A comprehensive comparison of RNA crystal structures has yielded features that contain a chelated Mg^{2+} ion[26]. Zheng et al.[26] found that while the oxygen from a phosphate was the most common ligand, the exocyclic oxygens from Uridine and Guanosine were also present, in both inner and outer sphere coordination (exocyclic amino groups of nucleobases do not bind metals, since their lone electron pair is delocalized into the heterocyclic nucleobase ring). Notably, examples of Mg^{2+} chelated with the N7 atom of Guanine were also detected (Figure 10). The structural context of these interactions included six previously known motifs and seven novel ones.

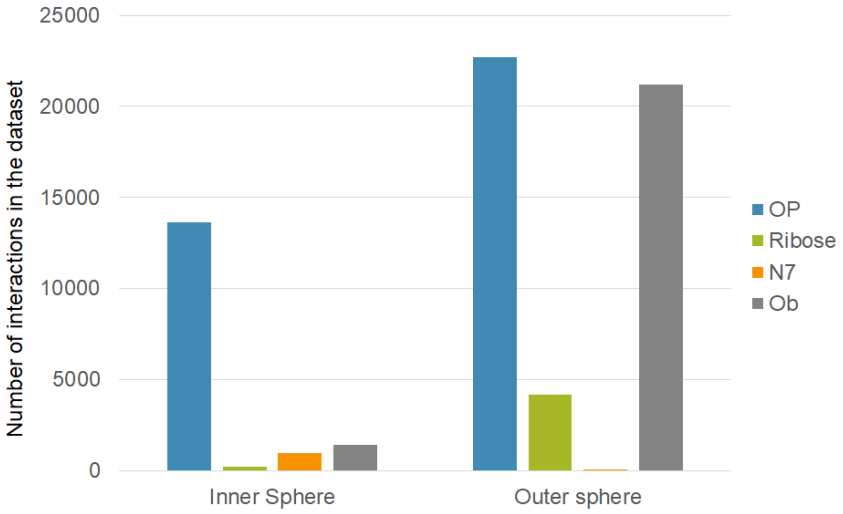


Figure 10. The distribution of RNA interactions with Mg^{2+} ions. The database had a total of 15334 Mg^{2+} sites found in RNA crystal structures from the PDB. OP are nonbridging phosphate oxygens, ribose are 2' oxygens, N7 purine nucleobase position (typically guanosine), and Ob are exocyclic oxygens on nucleobases. Adapted from[26].

It is worthwhile examining several of the most common Mg^{2+} binding motifs, to look for any characteristics that might aid in predicting their presence in an RNA that lacks a crystal structure. The 'magnesium clamp' and the '10-member ring' use the Mg^{2+} ion inner sphere to chelate two phosphate oxygens (Figure 11). Both are examples of bidentate chelation, an arrangement unique to RNA and Mg^{2+} in its geometry and its energetics. It is critical to keep in mind that formation of the Mg^{2+} binding site and the actual binding of the ion are coupled, which complicates anticipation of the interaction in a novel RNA.

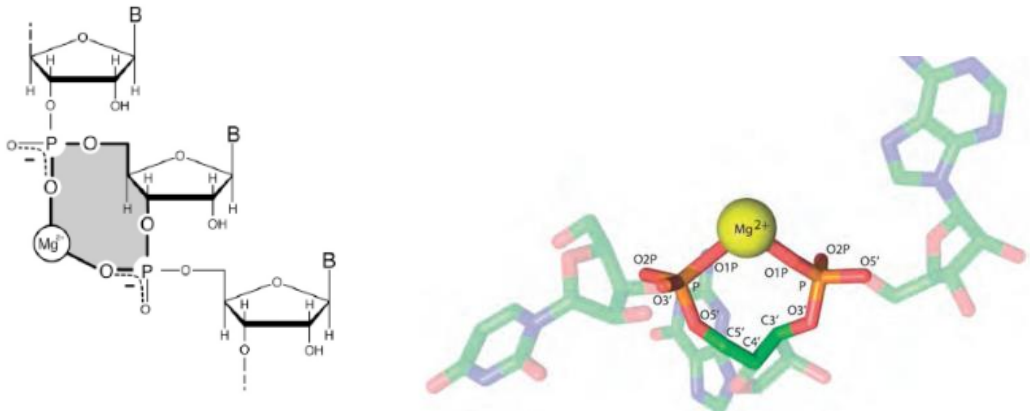


Figure 11. Two examples of Mg^{2+} ions bound to RNA. Left: A 10-member ring [Mg^{2+}-O^--P-O-C-C-C-O-P-O^-]. By definition, the ring encompasses sequential phosphates. Right: a Mg^{2+} clamp can chelate distal phosphates. This complex will be found in non-canonical regions of an RNA structure, not as part of a duplex. The structure shown is from 23S RNA in the crystal of the large ribosomal subunit from *H. marismortui*; taken from Hsiao et al[24].

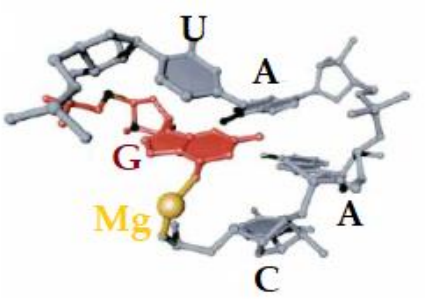


Figure 12. A bidentate interaction between the exocyclic oxygen (6-oxo) of the G and a phosphate oxygen from another strand. The G is hydrogen bonded to the A in a noncanonical arrangement. From Hsiao et al.[24]

The crystal structures of the Tetrahymena Intron do not have sufficient resolution to locate Mg^{2+} ions. The crystal structure of the smaller Azoarcus Group I Intron[27] (3.1 Å resolution) allowed probable identification of Mg^{2+} sites via substitution with Tb^{3+}, Yb^{3+}, Tl^+ and Mn^{2+} ions (Figure 13). These heavy metal ions have anomalous scattering that allows their localization (and phasing of the structure); their positions are assigned to Mg^{2+} or K^+. A

representation of that structure (PDB ID 1u6b) shows the distribution of water molecules and ions.

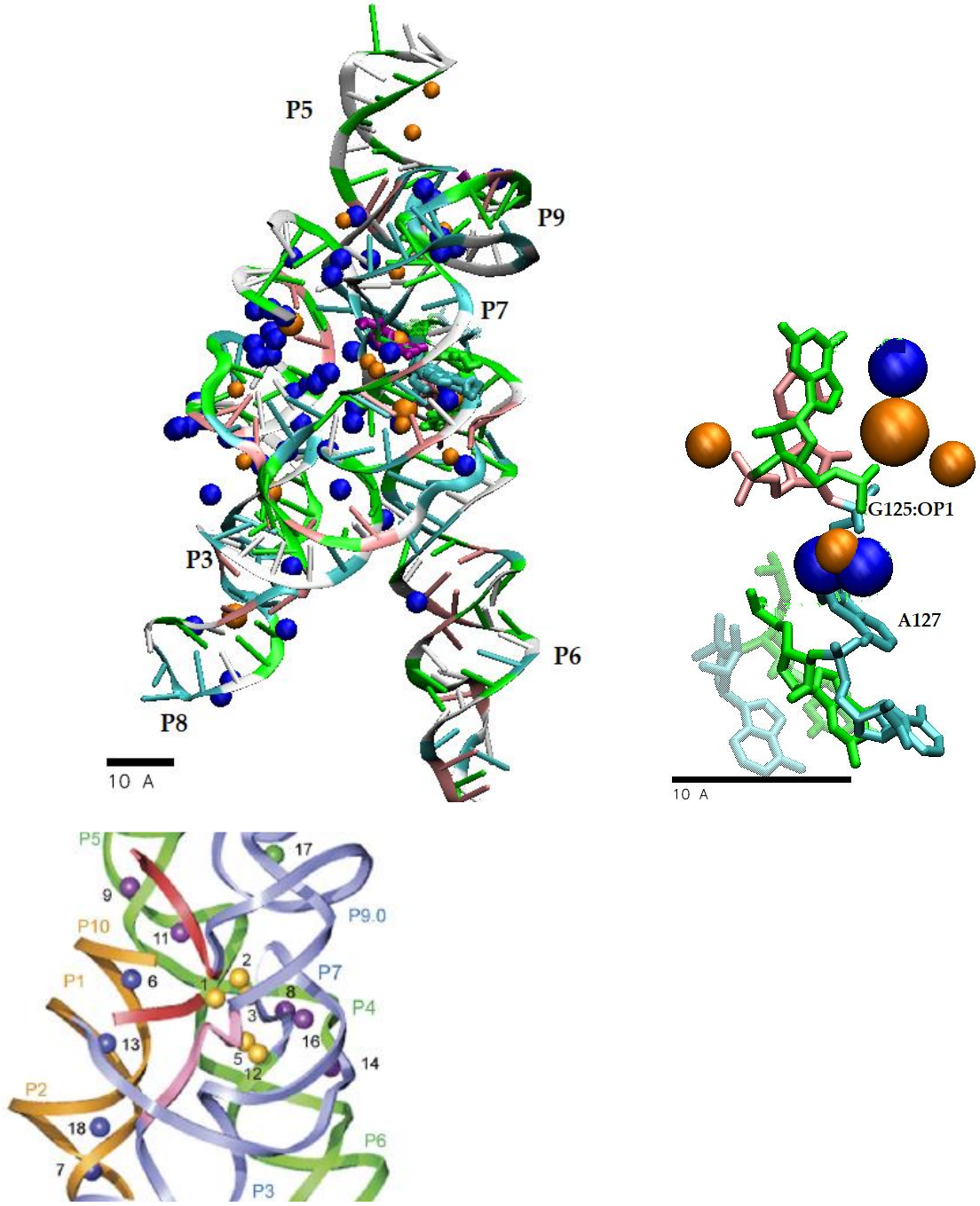


Figure 13. Azoarcus Group I intron was crystallized with both 5′ and 3′ exon sequences. In the crystal, the nucleotides at the end of P6 correspond to the binding site for the U1A protein, and that complex was used to facilitate crystallization (U1A protein not shown). The fragment on the right illustrates one of the Mg^{2+} ion interactions with phosphates in the structure; the corresponding nucleotides flank P7 in the intron structure. At bottom are the five metal ions M1-

M5 that were identified as essential to the core of the intron structure[28]. ΩG is fuschia in the intron structure. Adenosine is cyan, Guanosine is green, Cytosine white, and Uridine pink. Water molecules are blue, ions are gold; large ions are K^+, smaller are Mg^{2+}.

SUMMARY

- Tetrahymena Group I intron has a tertiary structure with a defined interior and exterior.
- The Tetrahymena tertiary structure requires divalent ions.
- Mg^{2+} ion is uniquely able to chelate nonbridging phosphate oxygens to anchor RNA strands together.

Problems.

Q1. How do you explain the statement that 'formation of the Mg^{2+} binding site and the actual binding of the ion are coupled'? What experimental consequences does this coupling present?

Q2. P1 contains a G:U base pair at the site of Guanosine cleavage. This base pair can be substituted by A:C and C:U and still permit the first step of splicing: A G:C pair blocks the first step. How do you explain these observations?

Q3. Draw a hydrogen bonded structure of a G+G-C base triple and a 2-aminopurine + A-U base triple that would be consistent with the experimental data.

Q4. Is it possible to identify those Mg^{2+} ions that are responsible for the divalent ion-dependent folding of the intron, based on their positions in the crystal structure?

Assignment: Read the original paper that presents the Azoarcus Group I crystal structure. (Adams, Stahley, Kosek Wang, & Strobel. 2004. Nature 430:45-50). What general features of ion placement were deduced from this crystal structure? The authors also discuss the role of the ions in the enzymatic reaction (they refer to Group I introns as obligate metalloenzymes). How does their discussion relate to the intrinsic properties of Mg^{2+} ions?

SOURCES and READING

1. Cech, T. R., Zaug, A. J. & Grabowski, P. J. In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **27**, 487–96 (1981).
2. Kruger, K. *et al.* Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31**, 147–57 (1982).
3. Zaug, A. J. & Cech, T. R. The Tetrahymena intervening sequence ribonucleic acid enzyme is a phosphotransferase and an acid phosphatase. *Biochemistry* **25**, 4478–82 (1986).
4. Zaug, A. J. & Cech, T. R. In vitro splicing of the ribosomal RNA precursor in nuclei of Tetrahymena. *Cell* **19**, 331–8 (1980).
5. Bass, B. L. & Cech, T. R. Specific interaction between the self-splicing RNA of Tetrahymena and its guanosine substrate: implications for biological catalysis by RNA. *Nature* **308**, 820–826 (1984).
6. Bass, B. L. & Cech, T. R. Ribozyme inhibitors. Deoxyguanosine and dideoxyguanosine are competitive inhibitors of self-splicing of the Tetrahymena ribosomal ribonucleic acid precursor. *Biochemistry* **25**, 4473–4478 (1986).
7. TR, C. *et al.* Secondary Structure of the Tetrahymena Ribosomal RNA Intervening Sequence: Structural Homology With Fungal Mitochondrial Intervening Sequences. *Proc. Natl. Acad. Sci. U. S. A.* **80**, (1983).
8. Cech, T. R. Self-Splicing of Group I Introns. *Annu. Rev. Biochem.* **59**, 543–568 (1990).
9. Cech, T. R., Damberger, S. H. & Gutell, R. R. Representation of the secondary and tertiary structure of group I introns. *Nat. Struct. Mol. Biol.* **1**, 273–280 (1994).
10. Li, N. *et al.* Cryo-EM structures of the late-stage assembly intermediates of the bacterial 50S ribosomal subunit. *Nucleic Acids Res.* **41**, 7073–83 (2013).
11. Davies, R. W., Waring, R. B. & Towner, P. Internal Guide Sequence and Reaction Specificity of Group I Self-splicing Introns. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 165–171 (1987).
12. Waring, R. B., Ray, J. A., Edwards, S. W., Scazzocchio, C. & Davies, R. W. The Tetrahymena rRNA intron self-splices in E. coli: in vivo evidence for the importance of key base-paired regions of RNA for RNA enzyme function. *Cell* **40**, 371–80 (1985).
13. Zaug, A. J., Been, M. D. & Cech, T. R. The Tetrahymena ribozyme acts like an RNA restriction endonuclease. *Nature* **324**, 429–33 (1986).
14. Michel, F., Hanna, M., Green, R., Bartel, D. P. & Szostak, J. W. The guanosine binding site of the Tetrahymena ribozyme. *Nature* **342**, 391–395 (1989).
15. Michel, F. & Westhof, E. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **216**, 585–610 (1990).
16. Tullius, T. D. & Greenbaum, J. A. Mapping nucleic acid structure by hydroxyl radical cleavage. *Curr. Opin. Chem. Biol.* **9**, 127–134 (2005).
17. Latham, J. & Cech, T. Defining the inside and outside of a catalytic RNA molecule. *Science (80-.).* **245**, 276–282 (1989).
18. Celander, D. & Cech, T. Visualizing the higher order folding of a catalytic RNA molecule. *Science (80-.).* **251**, 401–407 (1991).
19. Guo, F., Gooding, A. R. & Cech, T. R. Structure of the Tetrahymena Ribozyme. *Mol. Cell* **16**, 351–362 (2004).
20. Vicens, Q. & Cech, T. R. Atomic level architecture of group I introns revealed. *Trends Biochem. Sci.* **31**, 41–51 (2006).

21. Maguire, M. E. & Cowan, J. A. Magnesium chemistry and biochemistry. *Biometals* **15**, 203–10 (2002).
22. Bowman, J. C., Lenz, T. K., Hud, N. V & Williams, L. D. Cations in charge: magnesium ions in RNA folding and catalysis. *Curr. Opin. Struct. Biol.* **22**, 262–272 (2012).
23. Draper, D. E. Folding of RNA tertiary structure: Linkages between backbone phosphates, ions, and water. *Biopolymers* **99**, 1105–13 (2013).
24. Hsiao, C. *et al.* Chapter 1. Complexes of Nucleic Acids with Group I and II Cations. in *Nucleic Acid-Metal Ion Interactions* 1–38 (Royal Society of Chemistry, 2008). doi:10.1039/9781847558763-00001
25. Petrov, A. S., Bowman, J. C., Harvey, S. C. & Williams, L. D. Bidentate RNA-magnesium clamps: on the origin of the special role of magnesium in RNA folding. *RNA* **17**, 291–7 (2011).
26. Zheng, H., Shabalin, I. G., Handing, K. B., Bujnicki, J. M. & Minor, W. Magnesium-binding architectures in RNA crystal structures: validation, binding preferences, classification and motif detection. *Nucleic Acids Res.* **43**, 3789–801 (2015).
27. Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J. & Strobel, S. A. Crystal structure of a self-splicing group I intron with both exons. *Nature* **430**, 45–50 (2004).
28. Stahley, M. R., Adams, P. L., Wang, J. & Strobel, S. A. Structural metals in the group I intron: a ribozyme with a multiple metal ion core. *J. Mol. Biol.* **372**, 89–102 (2007).