# Molecular Simulations of $Mg^{2+}$-induced Folding of the Central Domain of the 16S Ribosomal RNA

Naoto Hori,[1] Natalia A. Denesyuk,[1] and D. Thirumalai[1,*]

[1] *Department of Chemistry, University of Texas, Austin, TX, USA*

(Dated: April 8, 2020)

# Abstract

The central domain of the 16S ribosomal RNA folds independently driven either by $Mg^{2+}$ ions or by interaction with a ribosomal protein. In order to provide a quantitative description of ion-induced folding of the ∼350 nucleotide RNA, we carried out extensive simulations using the coarse-grained Three Interaction Site (TIS) model with explicit $Mg^{2+}$ ions at a fixed $K^+$ ion concentration typically used in the Tris buffer. The $Mg^{2+}$ dependence of the radius of gyration shows that globally the RNA folds cooperatively. However, various structural elements order at different $Mg^{2+}$ concentrations, indicative of the hierarchical assembly even within a single domain of the rRNA. Binding of $Mg^{2+}$ ions is highly specific with successive ion condensation resulting in increasing tertiary structure formation. We also predict the $Mg^{2+}$-dependent protection factors, measurable in hydroxyl radical footprinting experiments, which corroborate the specificity of $Mg^{2+}$-induced folding. As a byproduct of the simulations, we provide a molecular basis of the structural rearrangements in the three-way junction in the intact central domain whose folding in isolation has been studied experimentally. In accord with experiments, we find that the three coplanar helices (h20, h21, and h22) with ∼120° interhelical angles at low $Mg^{2+}$ concentration undergo a large conformational transition to a state in which h21 and h22 are coaxially stacked, and h20 forms an acute angle with h21 at high $Mg^{2+}$ concentration. The angle distributions are broad even at high $Mg^{2+}$ concentrations attesting to the assembly heterogeneity. Most of our results are predictions that could be tested.



Keywords: RNA folding, divalent ions, ribosome assembly, three-way junction, coarse-grained simulations.


## INTRODUCTION

The determination of the spectacular ribosome structures has galvanized great interest in dissecting how such a complex structure could assemble so rapidly in vivo. Since ribosomes synthesize proteins in all living organisms, considerable energy and other regulatory mechanisms are used to generate and maintain their homeostasis. Although there have been vast experimental efforts to understand the molecular mechanism of ribosome assembly, which led to some important early discoveries, such as Nomura and Nierhaus maps [1, 2],


* dave.thirumalai@gmail.com

the general principles associated with the assembly process of the ribosome have not been fully resolved. [3, 4]. In order to solve the assembly problem, many beautiful experiments have been systematically performed initially by investigating how the various domains of the rRNA fold and subsequently by the effect of rRNA proteins in reshaping the assembly landscape. These studies, not unexpectedly, have produced an evolving and a complex picture of the assembly process. Along the way it has been pointed out, especially by Woodson [5], that some of the principles of ribozyme folding might form a useful framework for producing a quantitative theoretical model for ribosome assembly. After all rRNA folding problem has to be solved to produce an intact ribosome capable of protein synthesis.

In bacteria, the three ribosomal rRNA chains (∼4,500 nucleotides in total) fold and assemble together with over 50 r-proteins in order to build one functional ribosome. The small subunit, 30S particle, is itself a large ribonucleoprotein complex consisting of a single 16S rRNA chain (approximately) 1500 nucleotides and about 20 proteins. The rRNA chain could be further decomposed into the $5'$ ($\sim$ 560 nucleotides), central ($\sim$ 350 nucleotides), and $3'$ ($\sim$ 625 nucleotides) domains. It is known that the three domains fold independently in the absence of r-proteins. Therefore, a logical approach is to understand how the individual components, especially the various rRNA domains, might fold independent of each other.

Previous simulation studies, focusing on the nuances of the protein (S4) induced structural transitions in the $5'$ domain [6–8] have been most insightful, especially when combined with experiments [8]. These studies showed that S4-guided assembly results in the $5'$ domain reaching the folded state by navigating through multiple metastable states, revealing the rugged nature of the folding landscape of RNA [9, 10]. Inspired by the studies on the complex assembly process, here we investigated using simulations of an accurate model the folding of $Mg^{2+}$-induced folding of the central domain of the 16S rRNA. We view this as a first step in the computational leading to protein-rRNA interactions. Although this work provides no insights into any aspect of ribosome assembly, we hope to use this study as a spring board to a more detailed description of the r-protein interactions with ribosomal RNA.

A major hurdle that has prevented computations of rRNA folding and ribosome assembly is the complexity and the size of the ribosome, in addition to the unsatisfactory quality of force fields for RNA and metal ions. We tackle these issues using coarse grained (CG) models, which we have used with some success in describing ion-induced folding of RNA and the

effects of denaturants on folding transitions in a number of proteins. In particular, we have recently created a model for RNA, which quantitatively captures the folding thermodynamics of group I intron RNA and other RNA constructs, with particular focus on the effects of divalent cations [11].

Divalent cations, $Mg^{2+}$ in particular, are essential in stabilizing tertiary structures of RNA. In the context of ribosome assembly, the importance of $Mg^{2+}$ has been recognized from various kinds of experiments since the 1950s [12, 13]. For instance, it has been recently shown that the entire 23S rRNA from the large subunit of the bacterial ribosome forms a near-native conformation in the presence of $Mg^{2+}$ without any r-proteins [14]. In the central domain of the small subunit, in vitro studies have shown that a three-way junction (3WJ) (Fig. 1) is in dynamic equilibrium between an open and a closed conformation, whose populations depends on the $Mg^{2+}$ concentration. Experiments have shown that the 3WJ undergoes substantial rearrangements upon addition of $Mg^{2+}$ or S15 [15–17].

Here we study $Mg^{2+}$-dependent folding of the central domain of the small subunit of bacterial ribosome. We chose this 346-nt RNA fragment as our starting point because it has been used in experimental studies on the early stage of RNA assembly process involving r-proteins [16–20]. In this paper, we focus on the $Mg^{2+}$-dependent folding of the isolated RNA, aiming to establish a basis for subsequent studies including assembly mediated by r-proteins.

## MODEL AND METHODS

**Three-Interaction-Site model:** We used the Three-Interaction-Site (TIS) coarse-grained RNA model with explicit ions [11] to simulate the folding of the ribosomal RNA. In our previous study we establish that the model quantitatively reproduces thermodynamics of folding of RNA hairpins and pseudoknots as well as folding of the 195 nucleotide *Azoarcus* ribozyme [11]. Because the details of the model have been reported previously [11], we only provide a brief description here. In the TIS model [21], each nucleotide is represented by three coarse-grained spherical beads corresponding to phosphate, ribose sugar, and a base. Ions in the solution, $Mg^{2+}$, $K^{+}$, and $Cl^{-}$, are explicitly treated whereas water is modeled implicitly with temperature-dependent dielectric constant. Briefly, the effective potential energy is taken to be, $U_{\mathrm{TIS}} = U_{\mathrm{L}} + U_{\mathrm{EV}} + U_{\mathrm{ST}} + U_{\mathrm{HB}} + U_{\mathrm{EL}}$, where $U_{\mathrm{L}}$ accounts for chain

connectivity and bending stiffness of the polynucleic acids, $U_{\mathrm{EV}}$ accounts for excluded volume interactions of each chemical group including interactions between RNA sites and ions, $U_{\mathrm{ST}}$ and $U_{\mathrm{HB}}$ are the base-stacking and hydrogen-bond interactions, respectively. All the consecutive bases have the base-stacking interactions, of which the strength depends on the sequence. Any pair of the canonical Watson-Click base pairs (A-U and G-C) and the Wobble base pair (G-U), separated by at least four nucleotides along the chain, can form hydrogen bonds thus contribute $U_{\mathrm{HB}}$. In other words, the model accounts for certain non-native interactions as well. Besides those general stacking and hydrogen-bonding interactions, we generate a list of tertiary stacking and hydrogen-bonding appear in the specific RNA based on the crystal structure (see the next paragraph). The interactions between the charged moieties (phosphate groups and ions), $U_{\mathrm{EL}}$, interact via the standard Coulomb potential. The values of the charges on phosphate groups and ions are $-1$ (phosphate), $+2$ (Mg), $+1$ (K) and $-1$ (Cl). All the force-field parameters used here are the same as in our earlier study [11]. Applications to a wide variety of RNA molecules show that the TIS model is transferable, quantitatively accounting for many aspects of RNA folding, although it would require exhaustive testing to be sure that this is so.

**RNA molecule:** We investigated the folding thermodynamics of the central domain of the small subunit (16S) of *T. thermophilus* ribosome. The central domain of 16S rRNA consists of ∼350 nucleotides (C562–A914 in *T. thermophilus*, PDB entry 1J5E [22]). The secondary structure map and the tertiary structure are displayed in Figure.1C. The list of hydrogen bonds was generated by WHAT-IF server based on the crystal structure [23]. The secondary structure diagram was generated using RNApdbee [24] and Forrna [25]. All three-dimensional graphics of this paper were generated with VMD [26].

**Simulations:** In this study, we examined the effects of $Mg^{2+}$ ions by varying the concentration from 0 to 30 mM in the presence of 50 mM $K^+$, a typical value contained in the Tris buffer. All simulations were conducted using periodic cubic box (each side is 35 nm) containing different number of $Mg^{2+}$, determined by the concentration of divalent ions, and a fixed number of $K^+$ ions. An appropriate number of anions ($Cl^-$) was added to neutralize the entire system. We performed low friction Langevin dynamics simulations at 37°C in order to sample conformations of the system containing the ribosomal RNA and ions [27]. For each condition ($Mg^{2+}$ concentrations), at least 60,000 conformations were collected from

the equilibrated trajectories.

**Calculation of contact $Mg^{2+}$ concentration:** To quantify the affinity of $Mg^{2+}$ at each nucleotide site, we computed a *contact ion concentration* around the $i^{\mathrm{th}}$ nucleotide using [11],

$$c_i^* = \frac{1}{N_{\mathrm{A}} V_{\mathrm{c}}} \int_0^{r_{\mathrm{c}}} \rho_i(r)\, 4\pi r^2 \mathrm{d}r, \tag{1}$$

where $\rho_i(r)$ is number density of the ion at the distance $r$ from the phosphate of the $i^{\mathrm{th}}$ nucleotide, $V_{\mathrm{c}}$ is the spherical volume of radius $r_{\mathrm{c}}$, and $N_{\mathrm{A}}$ is the Avogadro's number to represent $c_i^*$ in molar units. In order to count only tightly bound $Mg^{2+}$, we used a cutoff distance, $r_c = R_{\mathrm{Mg}} + R_{\mathrm{P}} + \Delta r$ where $R_{\mathrm{Mg}}$ and $R_P$ are radii of $Mg^{2+}$ and phosphate sites, respectively, and $\Delta r = 0.15$ nm is a margin for contact formation. Since $\Delta r$ is small, the quantity $c_i^*$ corresponds the local molar concentration of $Mg^{2+}$ at the surface of the phosphate groups.

**Calculation of $Mg^{2+}$ effects on conformational changes:** We computed the free energy contribution of $Mg^{2+}$ binding to conformational changes as $\Delta\Delta G_\alpha$, where $\alpha$ represents a certain conformational change (*e.g.* formation of certain tertiary contacts, TCs). This quantity can be defined for each $Mg^{2+}$ binding site, namely $\Delta\Delta G_\alpha(i)$ for the $i^{th}$ nucleotide. The changes in the stability upon specific $Mg^{2+}$ binding is,

$$\Delta\Delta G_\alpha(i) = \Delta G_\alpha^{\mathrm{Mg}}(i) - \Delta G_\alpha^{\phi}(i), \tag{2}$$

where the superscripts Mg and $\phi$ indicate that $Mg^{2+}$ is bound and unbound to the $i^{th}$ nucleotide, respectively. Each term on the right hand side gives the stability due to contact formation, given that $Mg^{2+}$ is bound or unbound to the $i^{th}$ nucleotide. In other words,

$$\Delta G_\alpha^{\mathrm{Mg}}(i) = G_{\alpha F}^{\mathrm{Mg}}(i) - G_{\alpha U}^{\mathrm{Mg}}(i), \tag{3}$$

where $G_{\alpha\mathrm{F}}^{\mathrm{Mg}}$ and $G_{\alpha\mathrm{U}}^{\mathrm{Mg}}$ are the free energies of the states where the contact $\alpha$ is formed and disrupted, respectively, given that $Mg^{2+}$ is bound to the $i^{th}$ nucleotide. Similarly,

$$\Delta G_\alpha^{\phi}(i) = G_{\alpha F}^{\phi}(i) - G_{\alpha U}^{\phi}(i) \tag{4}$$

is the stability of the contact given that $Mg^{2+}$ is unbound. In our simulations, we calculated the difference in the free energy as a combination of the joint probabilities computed from the ensembles of conformations,

$$\Delta\Delta G_\alpha(i) = -k_{\mathrm{B}} T \ln \frac{P_{\alpha,i}(\mathrm{F},\, \mathrm{Mg}) P_{\alpha,i}(\mathrm{U},\, \phi)}{P_{\alpha,i}(\mathrm{U},\, \mathrm{Mg}) P_{\alpha,i}(\mathrm{F},\, \phi)} \tag{5}$$

where each $P_{\alpha,i}$ is a joint probability. For instance, $P_{\alpha,i}$(F, Mg) is the joint probability that the contact $\alpha$ is formed, and $Mg^{2+}$ is not bound to the $i^{\text{th}}$ nucleotide.

**Angles between helices:** Our previous work had shown that a key element in the self-assembly of *Azoarcus* ribozyme is the establishment of an angle between certain helices, which due to topological frustration, occurs only at high $Mg^{2+}$. To determine if this is the case for this piece of the rRNA we computed the $Mg^{2+}$-dependent angles between certain helices. In the three-way junction (Figure 1C), angles ($\Theta$) between helices h20, h21, and h22 were calculated using the axis of each helix. The helix axes were computed using Kahn's algorithm [28] using nucleotides 577-586 and 755-764 for h20, 588-597 and 643-651 for h21, and 655-672 and 734-751 for h22.

**Footprinting Protecting Factors (FPFs):** In order to calculate the solvent accessible surface area (SASA), we first reconstructed atomistic structures from simulations based on coarse-grained coordinates using a tool developed in-house that employs a fragment-assembly approach and energy minimization by AmberTools [29–31]. Using the reconstructed atomically detailed structures, SASA was computed with Free-SASA version 2.0 [32]. It is known that experimental footprinting data, obtained using hydroxyl radicals, is highly correlated with SASA of the sugar backbone [33, 34]. Considering that hydroxyl radicals preferably cleave C4' and C5' atoms of the RNA backbone [34], we assigned the larger SASA values of C4' and C5' atoms to each nucleotide. From the SASA data, we computed the *protection factor* of $i^{\text{th}}$ nucleotides as,

$$F_p(i) = \frac{\langle SASA(i)\rangle_{\text{Unfolded}}}{\langle SASA(i)\rangle}, \tag{6}$$

where the bracket indicates an ensemble average [35]. We used the conformations obtained in the absence of $Mg^{2+}$ to compute the average SASA of the unfolded state, $\langle \text{SASA}\rangle_{\text{Unfolded}}$.

**Hill equation for $R_g$:** The global transition of the rRNA size, measured by $R_g$ as a function of $Mg^{2+}$ concentration ($[Mg^{2+}]$), was fit to the Hill equation,

$$\theta\left([\text{Mg}^{2+}]\right) = \frac{[\text{Mg}^{2+}]^n}{[\text{Mg}^{2+}]_{\text{m}}^n + [\text{Mg}^{2+}]^n}. \tag{7}$$

The two parameters, the Hill coefficient ($n$) and the midpoint of $Mg^{2+}$ concentration ($[Mg^{2+}]_m$), were obtained by fitting the simulation data to Eq. 7. Before fitting, we converted the average $R_g$ values at each $[Mg^{2+}]$ to $\theta$ using,

$$\theta = -\frac{R_g - \min(R_g)}{\max(R_g) - \min(R_g)} + 1, \tag{8}$$

where $\max(R_g)$ and $\min(R_g)$ were taken from the average $R_g$ values at the lowest and highest $Mg^{2+}$ concentrations, respectively.

## RESULTS

**$Mg^{2+}$-induced collapse of rRNA:** We first focus on global properties that are measurable using Small Angle X-ray Scattering (SAXS) experiments. From equilibrium simulations of the rRNA fragment at various $Mg^{2+}$ ion concentrations, we calculated the average radius of gyration ($R_g$) as a function of $[Mg^{2+}]$ (Figure 2A). The RNA undergoes an apparent two-state folding transition, in terms of $R_g$, as $Mg^{2+}$ concentration increases from 0 to 30 mM. The maximum decrease in $R_g$ occurs in the range of $[Mg^{2+}] = 2{\sim}10$ mM. In the absence or lower $[Mg^{2+}]$ up to $\sim 1$ mM, tertiary interactions of the RNA are disrupted resulting in expanded conformations containing only secondary structures (see the most left structure in Figure 2A). After adding excess amount of $Mg^{2+}$ ($> 10$ mM), the RNA is folded to a compact conformation, with average $R_g \approx 4.3$ nm, that is modestly larger than the $R_g$ of the crystal structure, $R_g^{Native} = 4.0$ nm. This indicates that either or both of the ribosomal proteins and other domains of the rRNA (note that we use only a fragment of the 16S rRNA) may be needed to fully drive the conformation to that found in the crystal structure. To see the similarity between the structures found in the simulations and the crystal structure, we also computed the root-mean-square deviation (RMSD, Figure 2A inset). The RMSD of the entire rRNA fragment is 1.5 nm at 30 mM $Mg^{2+}$, whereas RMSD computed excluding h21 and h27 is 0.7 nm. From this data, we conclude that fluctuations of h21 and h27 contribute to the slightly larger $R_g$ found in simulations. The larger $R_g$ compared to the crystal structure notwithstanding, we conclude that the structure is correctly folded at high $[Mg^{2+}]$, as visually illustrated in Figure 2. Figure 2(B, C) shows the distance distribution functions at several $Mg^{2+}$ concentrations, which may be obtained as the inverse Fourier transform of the scattering function that is measurable by SAXS experiments, thus serving as a testable prediction. We also calculated the shape parameters, asphericity $\Delta$ ($0 \leq \Delta \leq 1$) and prolateness $S$ ($-0.25 \leq S \leq 2$) [36, 37]. Both $\Delta$ and $S$ are unity for rods, whereas $\Delta = S = 0$ for spheres. The calculated distributions show that the shape of the RNA is shifted from being spherical to a prolate ellipsoid as the $[Mg^{2+}]$ concentration increased.

**Helix h19 forms at high $Mg^{2+}$:** In the rRNA fragment, there are nine distinct he-

lices, that are conventionally named h19 through h27 (see Figure 1C). We calculated the probability of formation of the helices as a function of [$Mg^{2+}$] (Figure 3). All the helices except the short h19 are stable over the entire range of $Mg^{2+}$ concentration in the presence of 50 mM KCl at the simulation temperature (37°C). These helices, which are stable at very low [$Mg^{2+}$], contain at least 8 canonical Watson-Crick base pairs. In contrast, h19, connecting the central junction to h27, has only three G-C base pairs. Thus, it is likely that the stability of individual helices determine the order of their formation, as previously suggested [38]. More importantly, as shown in Figure 1C, the two constituent strands of h19 are far separated along the sequence. Consequently, the stability of h19 depends not only on [$Mg^{2+}$] but also on the formation of other helices, especially h25. Note that only after h25 fully forms (at [$Mg^{2+}$] $\approx$1 mM), which brings the two strands of h19 comes into proximity (Figure 1C), does h19 get structured at [$Mg^{2+}$] $\approx$2 mM (Figure 3). The formation of tertiary interactions are predicated on the formation of h19 (see below). The formation of h19 is essential for structuring the central junction (the big loop region surrounded by h19, h25, h24 and h20, shown in Figure 1C).

**[$Mg^{2+}$] midpoints of tertiary contacts formation vary:** In the crystal structure of the rRNA fragment, there are seven regions with major tertiary contacts (TC), where two or more secondary structural elements contact each other (Figure 4A). These contacts are typically clusters of hydrogen-bonding interactions. We label them as in Figure 4, TC1 through TC7. Since we have already seen that all the secondary structures except h19 are stable even in the near absence of $Mg^{2+}$, the formation of the TCs must drive the folding, and consequently compaction (decrease in $R_g$) of the RNA upon addition of $Mg^{2+}$ ions.

In Figure 4B, we show the formations of the seven TCs as a function of [$Mg^{2+}$]. There are three inferences that can be drawn from these results. (i) We first note that TC7, involving between h24 and h27, does not form even at the highest [$Mg^{2+}$]. We surmise that the formation of TC7 may need r-proteins or other domains of the rRNA. (ii) The other six TCs exhibit $Mg^{2+}$-dependent formation. Clearly, the [$Mg^{2+}$] midpoints at which the TCs order vary substantially. The fraction of TC formation sharply changes in the [$Mg^{2+}$] range from 2 to 10 mM. Thus, the hierarchical assembly of rRNA does not occur co-operatively at a sharp value of [$Mg^{2+}$], implying that there is no precise midpoint for folding even though global measures indicate otherwise. (iii) The range of $Mg^{2+}$ concentration over which rRNA orders, coincides with decrease in $R_g$ (Figure 2). The results confirm that the $R_g$ change,

which depends on [$Mg^{2+}$], is a consequence of the tertiary contact formation.

The six TCs can be classified into three pairs in terms of the $Mg^{2+}$ concentration requirements. From the titration curve (Figure 4B), the midpoints of [$Mg^{2+}$] are ∼2.5 mM for TC1 and TC2, ∼4 mM for TC3 and TC4, and ∼5 mM for TC6 and TC7. By mapping the TCs onto the secondary structure (see Figure 4A), we notice that the formations of the TCs occur first at locations near the central junction (TC1 and TC2). Upon further increase in [$Mg^{2+}$], the two interactions, TC3 and TC4, at a moderate distance from the junction are stabilized. Lastly, a much higher concentration of $Mg^{2+}$ ($\sim$ 10 mM) is required to stabilize the long-range contacts between h23 and h24. This hierarchy of tertiary contact formation is revealed in the representative snapshots from the simulations (Figure 2). Helices h25 and h26 (orange) come close to h20 (magenta) in the early part of the transition at low [$Mg^{2+}$], and the formation of contacts between h23 (blue) and h24 (yellow) occurs only at higher [$Mg^{2+}$].

**$Mg^{2+}$ binding to specific sites drives tertiary contact formations:** How do $Mg^{2+}$ ions facilitate the folding of the RNA fragment, in particular, the formation of the TCs? To begin the investigation, we computed the fingerprints of $Mg^{2+}$ binding at the nucleotide resolution as *contact $Mg^{2+}$ concentration* (see Methods). In Figure 5A, the fingerprints are shown at three bulk $Mg^{2+}$ concentrations at which the major conformational changes occur. The data clearly reveal that there are distinct positions where $Mg^{2+}$ ions bind with substantial probability. Although the contact concentrations increase as the bulk $Mg^{2+}$ concentration increases from 2.5 mM to 5.0 mM, many of the distinctive peaks exist even at 2.5 mM. This is a non-trivial finding because, at [$Mg^{2+}$] = 2.5 mM, only TC1 and TC2 form with ∼50% probability but all other contacts are either absent or formed with low probability (Figure 4). This shows that coordination of $Mg^{2+}$ to rRNA is nucleotide specific, and does not occur in a random diffusive manner as often assumed.

We further analyzed the relationship between TC formations and $Mg^{2+}$ binding to specific positions. The effects of specific binding on each TC formation can be revealed by comparing the free-energy change upon formation of each TC, $\Delta\Delta G_\alpha(i)$, which is the difference between bound and unbound states of $Mg^{2+}$ to the $i^{\text{th}}$ nucleotide (see Eq. 2). The subscript $\alpha$ stands for conformational change considered, such as $\alpha$ = TC1, TC2, ..., TC6. The value of $\Delta\Delta G_\alpha(i)$ would be negative if a specific binding of $Mg^{2+}$ to the $i^{\text{th}}$ nucleotide stabilizes the formation of $\alpha$. Figures 5B, C and D show $\Delta\Delta G$ for TC1, TC3 and TC5

formations, respectively, at the midpoint $Mg^{2+}$ concentrations for each TC. Unlike the local $Mg^{2+}$ concentration (Figure 5A), distinct positions where $Mg^{2+}$ binding contributes to the TC formation were extracted. Comparing the three dimensional positions in 5E and the locations of TC in Figure 4, we conclude that the contact formations are associated with specific $Mg^{2+}$ bindings to the same nucleotides.

**Tertiary stacking nucleates folding of the central junction:** The central junction is stabilized by several tertiary base stacking between non-consecutive nucleotide (tertiary stacking, TST). From the crystal structure, we detected six TST interactions around the central junction (Figure 6A, B), that are supposed to contribute the correct folding of the junction. Figure 6C, the fraction of TST formations as a function of $[Mg^{2+}]$, establishes that the TSTs are formed cooperatively at the midpoint $[Mg^{2+}]$ $\sim$2.5 mM. This value corresponds to the midpoint of formation of TC1 and TC2 (compare with Figure 4). Thus, the ordering of the central junction takes place cooperatively with the simultaneous formations of TC1 and TC2. These events resulting in partial folding, as well as the ordering of h19, contribute to the formation of the compact core region, leaving other longer peripheral helices unfolded. Structuring of these helices occur only at higher $[Mg^{2+}]$.

**Three-way junction folds upon tertiary-contact formation of constituent helices:** The three-way junction (3WJ) consisting of h20–22 has been used as a representative folding motif in earlier experimental studies of the rRNA folding [39]. In the unfolded state, because of the electrostatic repulsion, the three helices are expected to be well separating without having major interactions with each other. Indeed, an experimental study using transient electric birefringence indicated that the three angles between helices are nearly equal (roughly 120°) in the absence of $Mg^{2+}$ and r-protein S15 [17]. On the other hand, in the folded state, the two helices h21 and h22 are coaxially stacked, while the other, h20, forms an acute angle with h22 (Figure 1). Although, in the ribosome, S15 binds to the center of the junction and presumably stabilizes the folded form, experiments demonstrated that $Mg^{2+}$ ions alone are sufficient to stabilize the native form [17, 18]. In addition, further analyses showed that the folding of the junction is determined solely by the RNA sequence rather than by binding of S15 [16].

We calculated the three angles between h20, h21 and h22 as a function of $Mg^{2+}$ concentration (Figure 7). At $Mg^{2+}$ concentration below 4 mM, the average of the angle $\Theta_{\text{h20-h22}}$ is

about 80°, whereas the values of the other two angles are around 110°. On an average, the sum of the three angles is smaller than 360° indicating that the helices are not entirely confined to a plane. All the three angles fluctuate in the ensemble of simulated conformations, which is shown as ≈ 20° in the standard deviations (Figure 7). Nonetheless, the three helices are aligned roughly in a radial manner, in qualitative accord with the experiments [17]. At $[Mg^{2+}]$ ∼5 mM or above, the angles change dramatically. Angles $\Theta_{\text{h21-h22}}$ and $\Theta_{\text{h20-h21}}$ are around 150° indicating that two helices are coaxially stacked. Angle $\Theta_{\text{h20-h22}}$ adopts an acute value around 15°, showing that h20 is aligned toward h22, as in the crystal structure. As shown as the shaded regions in Figure 7, the fluctuations in the angles also much less at the higher $Mg^{2+}$ concentrations. The $Mg^{2+}$ concentration at which the 3WJ folds corresponds to the concentration at which TC3 forms, which is reasonable because TC3 is the tertiary interactions between the two ends of h20 and h22.

**Prediction for hydroxyl radical footprinting:** Lastly, we provide predictions for hydroxyl-radical footprinting based on solvent accessible surface area (SASA) of the simulated ensembles at each $Mg^{2+}$ concentration. Footprinting is a powerful experimental technique to probe RNA structures in vitro and in vivo [5, 40]. Much like hydrogen-deuterium exchange experiments using NMR in the context of protein folding, hydroxyl-radical footprinting is used to assess the extent to which each nucleotide is exposed to the solvent. The solvent exposure has been demonstrated to be highly correlated with the SASA of sugar backbone [33, 34]. We first calculated the SASA of each structure generated in the simulations, and then determined the protection factor ($P_F$) of each nucleotide site (see Method) by averaging over the simulated conformational ensemble at each $Mg^{2+}$ concentration.

Figure 9 shows the nucleotide-dependent protection factors at 1, 2.5, and 5 mM $Mg^{2+}$. At $[Mg^{2+}] = 1$ mM, only three nucleotides, G666, C726, and G727, show distinct protection ($P_F > 2$). In the crystal structure, G666 is near a bulge of h22, that interacts with a part of h23 (C726 and G727). Our simulation data indicate that this interaction is formed at relatively low $Mg^{2+}$ concentration (∼1 mM), and could in principle be detected by footprinting experiments. As $Mg^{2+}$ concentration is increased, additional nucleotides are protected. At $[Mg^{2+}] = 2.5$ mM, another tertiary interaction at the center of h25 results in protections of G869, U870, and G874. Above $Mg^{2+} > 5$ mM, nucleotides around the core region are protected (nucleotides colored in green in Figure 9G). Many of these nucleotides are the ones consisting of the central junction, and are involved in tertiary contacts, as shown in Figure

4. For reference we calculated the footprinting profile (Fig. 9F) using the crystal structure coordinates. Note that the protection factors calculated from a single crystal structure is less accurate and would overestimate the protections compared to the values under solution conditions because thermal fluctuations are not included in the crystal structure. Moreover, the crystal packing forces reinforces the packing rigidity. Despite these caveats, the positions of the peaks are consistent with the profile we obtained in the simulations at high $Mg^{2+}$ concentration (Fig. 9E, $[Mg^{2+}] = 20$ mM). The predictions in Figure 9 can be experimentally tested using the footprinting technique.

## DISCUSSION

We investigated folding of the central domain of the 16S rRNA with particular emphasis on on how $Mg^{2+}$ drives structure formation. To our knowledge experiments have not investigated the folding of the intact domain although the expectation is that the central domain could self-assemble autonomously. Consequently, many of our results are predictions that are amenable to tests using standard experimental techniques. The central domain of 16S rRNA unfolds at low $Mg^{2+}$ concentrations (roughly below 2 mM in our simulations), where only the secondary structures are intact. As $Mg^{2+}$ concentration is increased, tertiary interactions form in a hierarchical manner in three distinct stages. Considering that the typical $Mg^{2+}$ concentration of bacterial cytoplasm is $\sim$ 1 mM [41], our data suggest that the rRNA should not spontaneously form a stable compact structure in the absence of r-proteins (Figure 2). This conclusion is not definitive because folding *in vivo* occurs in a crowded milieu, which would lower the effective midpoint for folding [42–44]. The detailed simulations, which currently cannot be done using any other computational method for a large RNA construct studied here, allow us to provide a few additional remarks.

**Fate of the three-way junction:** The independent folding of the 3WJ and related constructs in the central domain (Fig.1) has been studied extensively, principally by Williamson and coworkers, over two decades ago by a variety of experimental methods [16–18]. In a quest to understand the influence of r-proteins on the assembly of 30S particle they focused initially on the folding of the 3WJ, which folds either in the presence of S15 or $Mg^{2+}$ [17]. Using transient electric birefringence and model to analyze the data [17] they inferred that the angles between the three helices (h20, h21, and h22) are roughly 120°, which implies

that they adopt a planar structure. Upon addition of 1.5 mM $Mg^{2+}$ (or S15) the 3WJ is structured in which h21 is coaxially stacked with helix 22 and h20 forms an acute angle with h22. The results of our simulations, carried using the entire central domain, are in excellent agreement with experiments (Fig.7). We find that 3WJ folds cooperatively with the formation of tertiary interactions between h20 and h22 (TC3). The good agreement between simulations and experiments not only validates the model, which has been adequately justified previously [11, 45], but also shows that the global structure of the 3WJ does not change significantly when embedded in the intact central domain.

To illustrate the fluctuations in the 3WJ as function of $Mg^{2+}$ concentration, we calculated the two dimensional distributions of specific distances between the helices and the angles, which could be used to qualitatively compare with single molecule experiments [15, 46] on a related but different 3WJ construct. We find that at low $Mg^{2+}$ concentrations the distributions are very broad (Fig. 8) but does become narrower when $Mg^{2+}$ increases. The distribution of distance between G577 and C735 ($d_{\text{h20-h22}}$) is unusually broad at low $Mg^{2+}$ concentrations, which is reminiscent of the FRET efficiency distributions [15]. The substantial width of the angle distributions, which are difficult to directly measure experimentally, also show that upon folding there are substantial conformations in the 3WJ even in the intact central domain.

**Specificity of $Mg^{2+}$ binding:** The role of $Mg^{2+}$ ions is particularly difficult to probe experimentally because monitoring the binding to RNA is likely to be cooperative. Indeed, we find that the Hill coefficient extracted from the dependence of $R_g$ on $Mg^{2+}$ condensation is roughly three, an indication of cooperative binding (Fig. 2A). Our simulations show in no unequivocal terms that $Mg^{2+}$ ion binding is discrete and highly specific. Even at the lowest concentration of $Mg^{2+}$ there are specific nucleotides where the local concentrations of the divalent ions is significantly higher than the bulk value. It is possible that these discrete and specific binding nucleates the folding of the RNA, which was an idea that was proposed long ago based on the crystal structure of the P4-P6 domain of the *Tetrahymena* ribozyme [47]. Indeed, a new theme that is emerging through simulations that we have performed is that $Mg^{2+}$ ion binding, which can occur directly or mediated by a single water molecule [48], is specific and is dictated by the architecture of the folded state [11, 45]. The present study adds one more example of this new concept.

**Transition midpoint is not unique:** A corollary of specificity of $Mg^{2+}$ ion binding implies that the midpoint concentration of $Mg^{2+}$ at which the RNA folds is not unique. This is evident in Fig. 4 which shows that the important tertiary interactions that stabilize RNA order at different $Mg^{2+}$ concentrations. The variations likely suggests the differences in stabilities in the different regions with the least stable one requiring higher $Mg^{2+}$ concentration. A recent study has explicitly demonstrated this idea in temperature induced folding of several RNA pseudoknots [38]. The non-uniqueness of ordering temperature or denaturant concentration has previously been predicted for proteins [49] in which (typically) the secondary structural elements are not stable, this making it difficult to validate the predictions experimentally.

**Estimates of $Mg^{2+}$ midpoints:** A few remarks on the measurement or calculation of the global $Mg^{2+}$ midpoint (referred to as $c_{\rm m}$), the $Mg^{2+}$ concentrations at which individual secondary or tertiary interactions order, are worth making. The absolute values of $c_{\rm m}$ are not highly significant except as a qualitative assessment of the ion-induced folding reactions. There are three principle reasons for making this assertion. (1) The $c_{\rm m}$ values depend on the RNA concentration, $c_{\rm RNA}$. For example, experiments show that the value of $c_{\rm m}$ for *Azoarcus* ribozyme is $\approx 0.34$ mM at $c_{\rm RNA} = 6.3$ $\mu$M [50] and increases to $c_{\rm m} \approx 0.88$ mM at $c_{\rm RNA} = 15.8$ $\mu$M [51]. We expect that at $c_{\rm m}$ the concentration of the free ion, $c_{\rm m,0}$, is almost (not exactly) negligible [11], which implies that $c_{\rm m,0} = c_{\rm m} - m \cdot c_{\rm RNA} \approx 0$ where $m$ is the number of $Mg^{2+}$ ions bound to one RNA molecule. This relationship shows that roughly $c_{\rm m} \propto c_{\rm RNA}$. (2) In our simulations $c_{\rm RNA} = 38.7$ $\mu$M, which larger than the typical value used in experiments by a factor of $\sim 2$. In light of the linear dependence of $c_{\rm m}$ on $c_{\rm RNA}$ noted above, we expect the predicted $c_{\rm m}$ values to be about a factor of two larger than experimental measurements. (3) The estimate of $c_{\rm m}$ also depends on the order parameter used to assess the extent of folding. Different probes could and do give different values of $c_{\rm m}$. The value of $c_{\rm m}$ obtained using $R_{\rm g}$ would be different if native gel assay or average FRET efficiency is used even if $c_{\rm RNA}$ is fixed. Of course, the $c_{\rm m}$ values would not differ significantly (at most a factor 2-3 say) if reasonable values (range in which no inter RNA interactions occur) of $c_{\rm RNA}$ is used and an appropriate order parameter is considered. In light of the arguments given here we conclude that $c_{\rm m}$ should be treated as a qualitative measure of the efficiency of ions to fold RNA. For instance, for a fixed $c_{\rm RNA}$ the $c_{\rm m}$s are measures of how efficient a particular ion is in folding the RNA.

## CONCLUSION

Currently there appears to be no alternative than to use the TIS model with explicit ions for simulating $Mg^{2+}$-dependent folding of large RNA molecules. The present study on the central domain of rRNA, which yields results that are consistent with experiments on the 3WJ, has produced a number of predictions that could be tested. We are encouraged that the very general framework could treat the folding the full length rRNA. We have also created a protein model, which has produced results for the thermodynamics and kinetics of folding that are in near quantitative agreement with experiments for several proteins (see for example [52–55]). Naturally, our immediate aim is to combine these models to simulate the assembly of RNA-protein complexes, and examine how the ribosomal proteins modulate the shape fluctuations of the rRNA.

## ACKNOWLEDGMENTS

Much of this work was carried out while the authors were in the Institute for Physical Science and Technology at the University of Maryland. N.H. is grateful to Hung T. Nguyen and Mauro Mugnai for valuable discussions. This work was supported in part by a grant from the National Science Foundation (CHE 19-00093). D.T. also acknowledges additional support from the Collie-Welch Regents Chair (F-0019) administered through the Welch Foundation.

**Abbreviations:** TC, tertiary contact; TST, tertiary stacking; SASA, solvent accessible surface area; 3WJ, three way junction.

**Author contributions:** N.H., N.A.D., and D.T. designed research; N.A.D. performed the simulations; N.H. and D.T. analyzed data; and N.H. and D.T. wrote the manuscript.

---

[1] S. Mizushima and M. Nomura, Assembly mapping of 30S ribosomal proteins from E. coli, Nature **226**, 1214 (1970).

[2] M. Herold and K. Nierhaus, Incorporation of six additional proteins to complete the assembly map of the 50S subunit from Escherichia coli ribosomes., J. Biol. Chem. **262**, 8826 (1987).

[3] S. A. Woodson, RNA folding and ribosome assembly, Curr. Opin. Chem. Biol. **12**, 667 (2008).

[4] Z. Shajani, M. T. Sykes, and J. R. Williamson, Assembly of bacterial ribosomes, Annu. Rev. Biochem. **80**, 501 (2011).

[5] S. A. Woodson, RNA folding pathways and the self-assembly of ribosomes, Acc. Chem. Res. **44**, 1312 (2011).

[6] J. Lai, K. Chen, and Z. Luthey-Schulten, Structural intermediates and folding events in the early assembly of the ribosomal small subunit, J. Phys. Chem. B **117**, 13335 (2013).

[7] K. Chen, J. Eargle, J. Lai, H. Kim, S. Abeysirigunawardena, M. Mayerle, S. A. Woodson, T. Ha, and Z. Luthey-Schulten, Assembly of the five-way junction in the ribosomal small subunit using hybrid MD–Gō simulations, J. Phys. Chem. B **116**, 6819 (2012).

[8] H. Kim, S. C. Abeysirigunawardena, K. Chen, M. Mayerle, K. Ragunathan, Z. Luthey-Schulten, T. Ha, and S. A. Woodson, Protein-guided RNA dynamics during early ribosome assembly, Nature **506**, 334 (2015).

[9] D. Thirumalai and S. Woodson, Kinetics of folding of proteins and RNA, Acc. Chem. Res. **29**, 433 (1996).

[10] D. Thirumalai and C. Hyeon, RNA and protein folding: common themes and variations, Biochemistry **44**, 4957 (2005).

[11] N. A. Denesyuk and D. Thirumalai, How do metal ions direct ribozyme folding?, Nat. Chem. **7**, 793 (2015).

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

[13] K. H. Nierhaus, $Mg^{2+}$, $K^{+}$, and the ribosome., J. Bacteriol. **196**, 3817 (2014).

[14] T. K. Lenz, A. M. Norris, N. V. Hud, and L. D. Williams, Protein-free ribosomal RNA folds to a near-native state in the presence of $Mg^{2+}$, RSC Adv. **7**, 54674 (2017).

[15] T. Ha, X. Zhuang, H. D. Kim, J. W. Orr, J. R. Williamson, and S. Chu, Ligand-induced conformational changes observed in single RNA molecules, Proc. Natl. Acad. Sci. U.S.A. **96**, 9077 (1999).

[16] R. T. Batey and J. R. Williamson, Effects of polyvalent cations on the folding of an rRNA three-way junction and binding of ribosomal protein S15, RNA **4**, 984 (1998).

[17] J. W. Orr, P. J. Hagerman, and J. R. Williamson, Protein and $Mg^{2+}$-induced conformational changes in the S15 binding site of 16 S ribosomal RNA, J. Mol. Biol. **275**, 453 (1998).

[18] S. C. Agalarov, G. Sridhar Prasad, P. M. Funke, C. D. Stout, and J. R. Williamson, Structure of the S15, S6, S18–rRNA complex: Assembly of the 30S ribosome central domain., Science **288**, 107 (2000).

[19] M. I. Recht and J. R. Williamson, RNA tertiary structure and cooperative assembly of a large ribonucleoprotein complex, J. Mol. Biol. **344**, 395 (2004).

[20] T. Lavergne, R. Lamichhane, D. A. Malyshev, Z. Li, L. Li, E. Sperling, J. R. Williamson, D. P. Millar, and F. E. Romesberg, FRET characterization of complex conformational changes in a large 16S ribosomal RNA fragment site-specifically labeled using unnatural base pairs, ACS Chem. Biol. **11**, 1347 (2016).

[21] C. Hyeon and D. Thirumalai, Mechanical unfolding of RNA hairpins., Proc. Natl. Acad. Sci. U.S.A. **102**, 6789 (2005).

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

[23] G. Vriend, WHAT IF: a molecular modeling and drug design program, J. Mol. Graph. **8**, 52 (1990).

[24] T. Zok, M. Antczak, M. Zurkowski, M. Popenda, J. Blazewicz, R. W. Adamiak, and M. Szachniuk, RNApdbee 2.0: multifunctional tool for RNA structure annotation, Nucleic Acids Res. **46**, W30 (2018).

[25] P. Kerpedjiev, S. Hammer, and I. L. Hofacker, Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams, Bioinformatics **31**, 3377 (2015).

[26] W. Humphrey, A. Dalke, and K. Schulten, VMD – Visual Molecular Dynamics, J. Mol. Graph. **14**, 33 (1996).

[27] J. D. Honeycutt and D. Thirumalai, The nature of folded states of globular proteins, Biopolymers **32**, 695 (1992).

[28] P. C. Kahn, Defining the axis of a helix, Computers & Chemistry **13**, 185 (1989).

[29] N. Hori, D. Chakraborty, T. N. Hung, and D. Thirumalai, TIS2AA (2017), doi:10.5281/zenodo.581485.

[30] E. Humphris-Narayanan and A. M. Pyle, Discrete RNA libraries from pseudo-torsional space, J. Mol. Biol. **421**, 6 (2012).

[31] D. Case, I. Ben-Shalom, S. Brozell, D. Cerutti, T. Cheatham III, V. Cruzeiro, T. Darden,

R. Duke, D. Ghoreishi, M. Gilson, *et al.*, Amber 2018: San francisco (2018).

[32] S. Mitternacht, FreeSASA: An open source C library for solvent accessible surface area calculations, F1000 Res **5**, 189 (2016).

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

[34] B. Balasubramanian, W. K. Pogozelski, and T. D. Tullius, DNA strand breaking by the hydroxyl radical is governed by the accessible surface areas of the hydrogen atoms of the DNA backbone, Proc. Natl. Acad. Sci. U.S.A. **95**, 9738 (1998).

[35] P. L. Adams, M. R. Stahley, M. L. Gill, A. B. Kosek, J. Wang, and S. A. Strobel, Crystal structure of a group I intron splicing intermediate., RNA **10**, 1867 (2004).

[36] J. A. Aronovitz and D. R. Nelson, Universal features of polymer shapes, Journal de physique **47**, 1445 (1986).

[37] C. Hyeon, R. I. Dima, and D. Thirumalai, Size, shape, and flexibility of RNA structures, J. Chem. Phys. **125**, 194905 (2006).

[38] S. S. Cho, D. L. Pincus, and D. Thirumalai, Assembly mechanisms of RNA pseudoknots are determined by the stabilities of constituent secondary structures, Proc. Natl. Acad. Sci. U.S.A. **106**, 17349 (2009).

[39] A. Lescoute and E. Westhof, Topology of three-way junctions in folded RNAs., RNA **12**, 83 (2006).

[40] R. M. Hulscher, J. Bohon, M. C. Rappé, S. Gupta, R. D'Mello, M. Sullivan, C. Y. Ralston, M. R. Chance, and S. A. Woodson, Probing the structure of ribosome assembly intermediates in vivo using DMS and hydroxyl radical footprinting, Methods **103**, 49 (2016).

[41] E. A. Groisman, K. Hollands, M. A. Kriner, E.-J. Lee, S.-Y. Park, and M. H. Pontes, Bacterial $Mg^{2+}$ homeostasis, transport, and virulence, Annu. Rev. Genet. **47**, 625 (2013).

[42] D. Kilburn, J. H. Roh, L. Guo, R. M. Briber, and S. A. Woodson, Molecular crowding stabilizes folded RNA structure by the excluded volume effect, J. Am. Chem. Soc. **132**, 8690 (2010).

[43] N. A. Denesyuk and D. Thirumalai, Crowding promotes the switch from hairpin to pseudoknot conformation in human telomerase RNA, J. Am. Chem. Soc. **133**, 11858 (2011).

[44] N. A. Denesyuk and D. Thirumalai, Entropic stabilization of the folded states of RNA due to macromolecular crowding, Biophys. Rev. **5**, 225 (2013).

[45] N. Hori, N. A. Denesyuk, and D. Thirumalai, Ion condensation onto ribozyme is site specific and fold dependent, Biophys. J. **116**, 2400 (2019).

[46] H. D. Kim, G. U. Nienhaus, T. Ha, J. W. Orr, J. R. Williamson, and S. Chu, $Mg^{2+}$–dependent conformational change of RNA studied by fluorescence correlation and FRET on immobilized single molecules, Proc. Natl. Acad. Sci. U.S.A. **99**, 4284 (2002).

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

[48] H. T. Nguyen and D. Thirumalai, Charge density of cation determines inner versus outer shell coordination to phosphate in RNA, bioRxiv 10.1101/2020.03.10.986091 (2020).

[49] E. P. O'Brien, B. R. Brooks, and D. Thirumalai, Molecular origin of constant m-values, denatured state collapse, and residue-dependent transition midpoints in globular proteins, Biochemistry **48**, 3743 (2009).

[50] S. Chauhan, G. Caliskan, R. M. Briber, U. Perez-Salas, P. Rangan, D. Thirumalai, and S. A. Woodson, RNA tertiary interactions mediate native collapse of a bacterial group I ribozyme, J. Mol. Biol. **353**, 1199 (2005).

[51] J. H. Roh, L. Guo, J. D. Kilburn, R. M. Briber, T. Irving, and S. A. Woodson, Multistage collapse of a bacterial ribozyme observed by time-resolved small-angle X-ray scattering, J. Am. Chem. Soc. **132**, 10148 (2010).

[52] Z. Liu, G. Reddy, E. P. O'Brien, and D. Thirumalai, Collapse kinetics and chevron plots from simulations of denaturant-dependent folding of globular proteins, Proc. Natl. Acad. Sci. U.S.A. **108**, 7787 (2011).

[53] G. Reddy, Z. Liu, and D. Thirumalai, Denaturant-dependent folding of GFP, Proc. Natl. Acad. Sci. U.S.A. **109**, 17832 (2012).

[54] H. Maity and G. Reddy, Thermodynamics and kinetics of single-chain monellin folding with structural insights into specific collapse in the denatured state ensemble, J. Mol. Biol. **430**, 465 (2018).

[55] H. Maity and G. Reddy, Transient intermediates are populated in the folding pathways of single-domain two-state folding protein L, J. Chem. Phys. **148**, 165101 (2018).

[56] A. S. Petrov, C. R. Bernier, B. Gulen, C. C. Waterbury, E. Hershkovits, C. Hsiao, S. C. Harvey, N. V. Hud, G. E. Fox, R. M. Wartell, and L. D. Williams, Secondary structures of rRNAs from all three domains of life, PLoS One **9**, e88222 (2014).

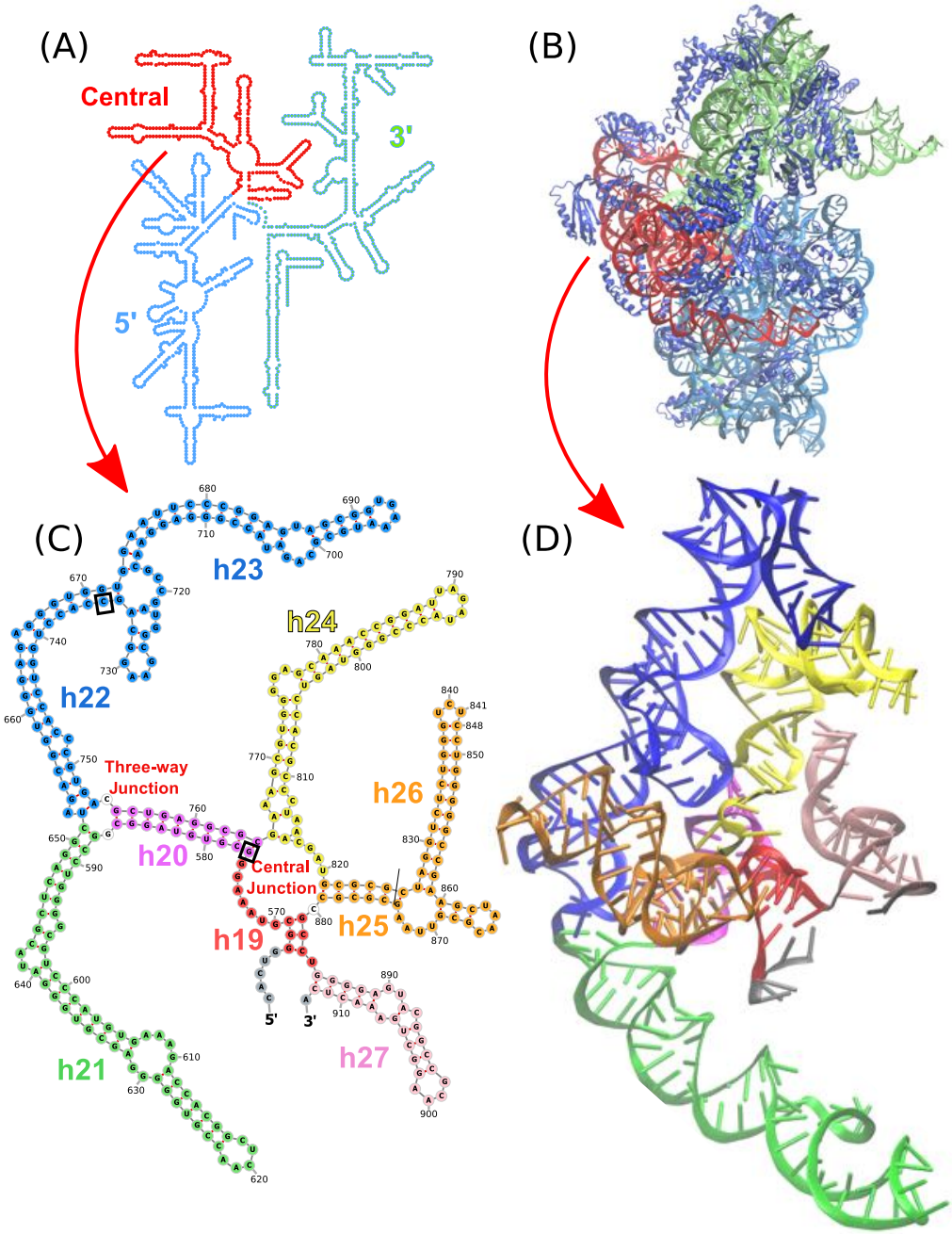



FIG. 1. Sequence and structure of the ribosomal RNA. **(A, B)** The entire small subunit of *T. Thermophilus* ribosome is illustrated as the secondary structure of the 16S rRNA (A) and the crystal structure including ribosomal proteins (B). The central domain simulated is in red and labeled as *Central* in (A). The secondary structure is adopted from [56] and the tertiary structure was taken from PDB entry 1J5E [22]. **(C, D)** Secondary and tertiary structures of the 346-nucleotide fragment central domain of the 16S ribosomal RNA.

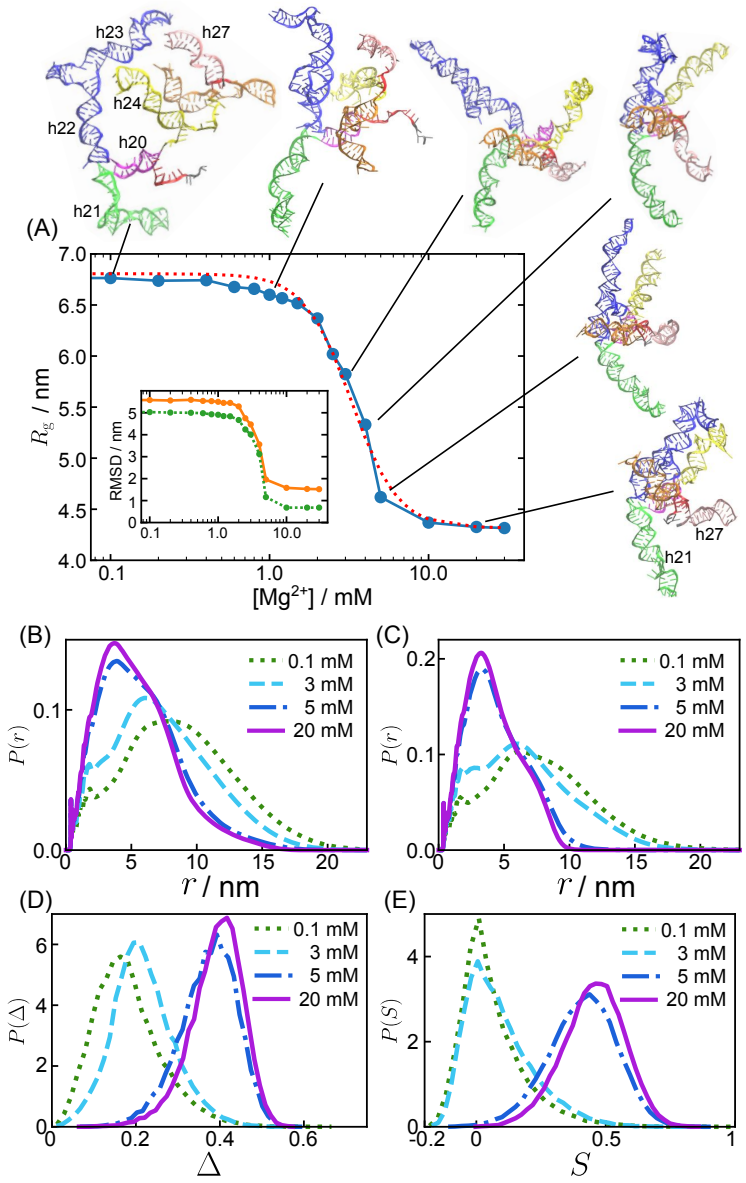



FIG. 2. **(A)** Radius of gyration ($R_g$) as a function of $Mg^{2+}$ concentration. Representative conformations are shown for $[Mg^{2+}] = 0.1$, 1.0, 3.0, 4.0, 5.0, and 20.0 mM (left to right). The inset shows the average RMSD from the crystal structure. The orange solid line is RMSD for the entire RNA fragment, and the green dotted line is calculated excluding h21 and h27 (helices indicated in the bottom right structure). The red dotted line is the fit to the Hill equation with $n = 2.96$ and the midpoint $[Mg^{2+}]_m = 3.3$ mM. **(B, C)** Pair distance distributions at $[Mg^{2+}] = 0.1$, 3, 5, and 20 mM (labeled in the panel) calculated using (B) the entire fragment, and (C) the fragment excluding h21 and h27. **(D, E)** Probability distributions of the shape parameters, (D) asphericity $\Delta$ and (E) prolateness $S$, at the same concentrations as (B, C).

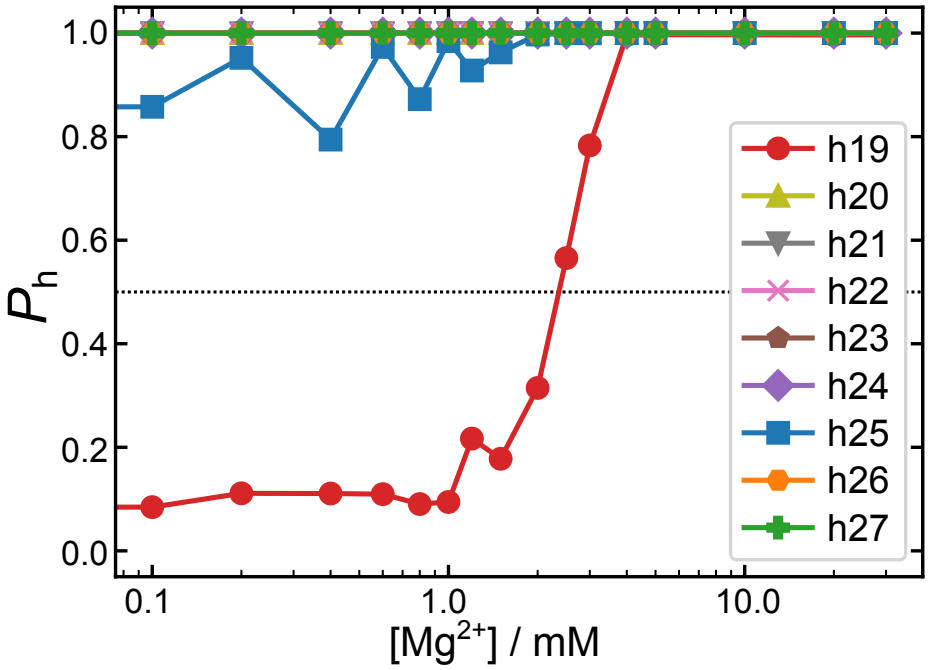



FIG. 3. Fractions of helix formation ($P_h$) as a function of $Mg^{2+}$ concentration. Besides h19 (red) and h25 (blue), all other lines overlap with h27 (green) because these stable helices remain folded even at extremely low $Mg^{2+}$ concentrations. See Figure 1 for positions of the helices. Both h19 and h25 are near the central junction (Figure 1C).

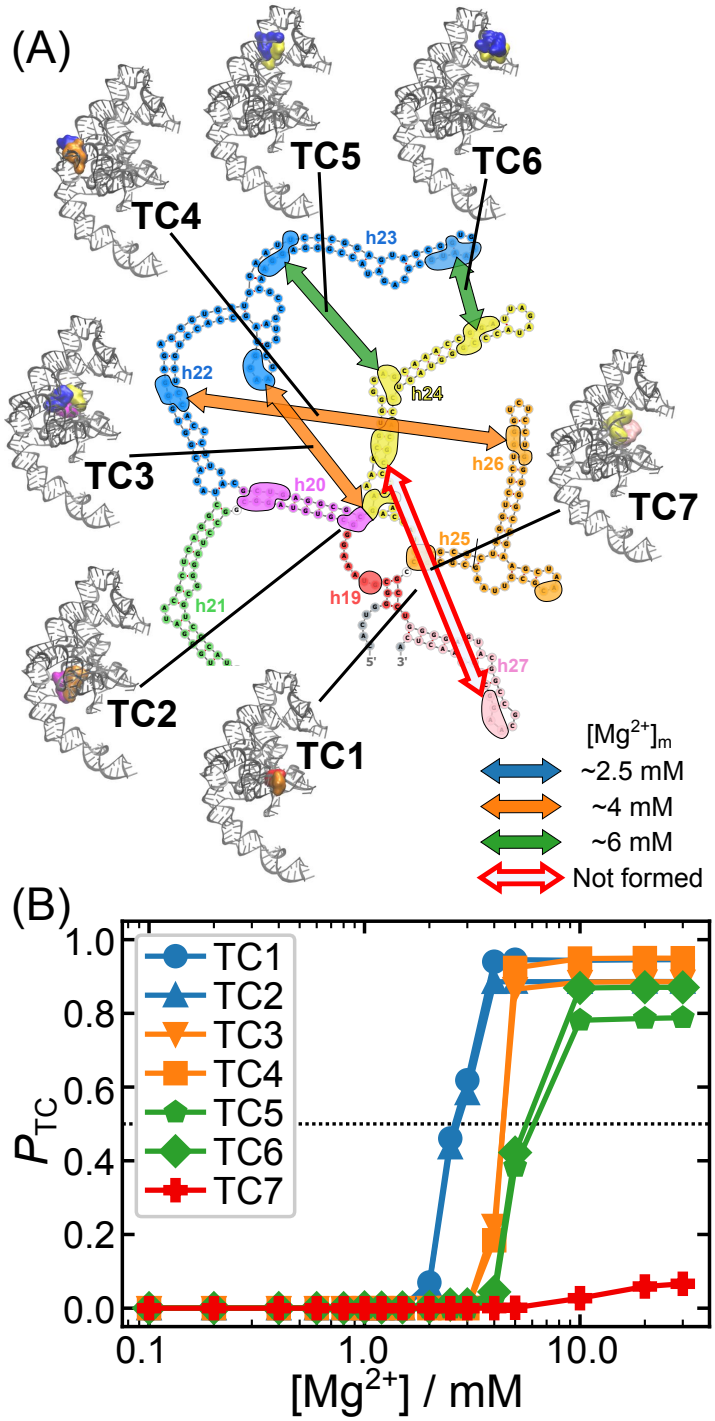



FIG. 4. Key clusters of tertiary contacts across the structure. **(A)** Locations of the seven clusters of tertiary contacts (TC) are mapped onto the secondary structure diagram. On the periphery, tertiary structures are shown using a surface representation for the corresponding regions. **(B)** Fraction of various TC formation as a function of $Mg^{2+}$ concentrations. The $[Mg^{2+}]$ midpoints calculated using $P_{\mathrm{TC}}([Mg^{2+}]_m) = 0.5$ (shown as horizontal line) change substantially depending on the TC. The color code reflects the midpoint of $[Mg^{2+}]$ for each TC as shown on the middle.

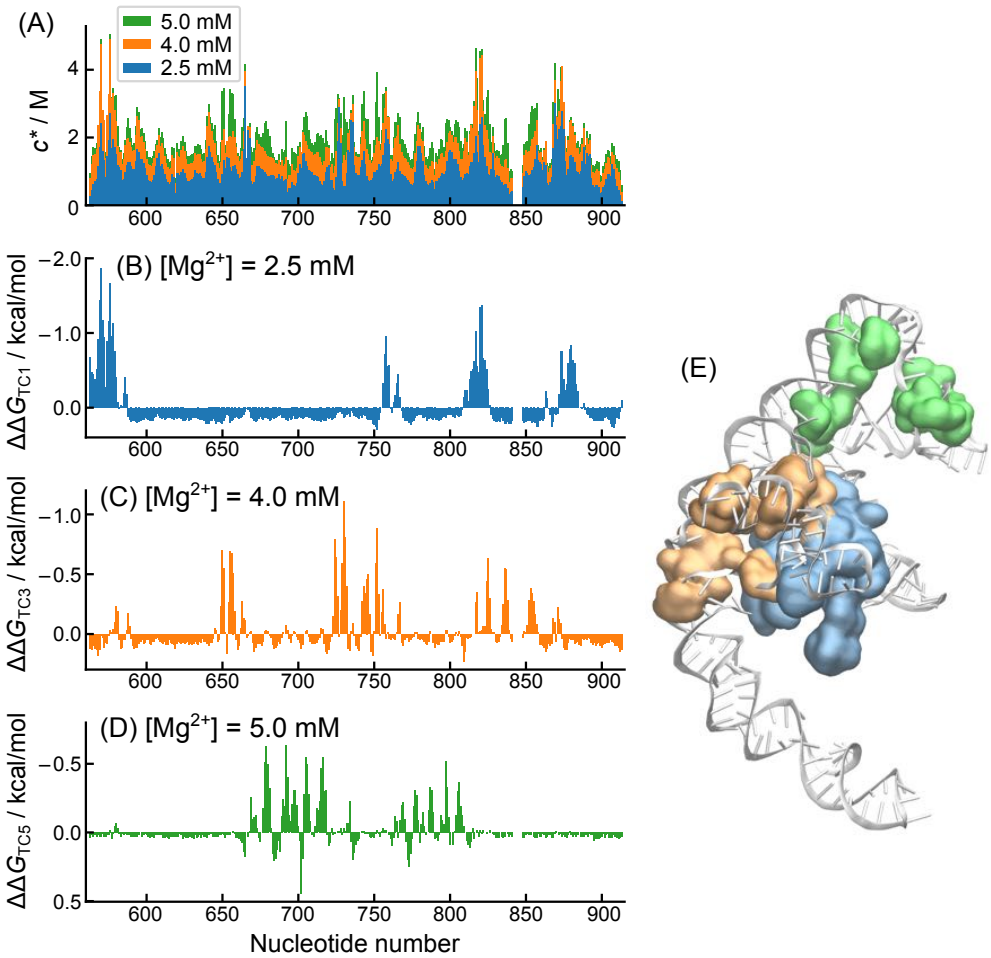



FIG. 5. Fingerprints of $Mg^{2+}$ binding leading to tertiary contact (TC) formations. **(A)** $Mg^{2+}$ binding to each nucleotide is quantitated as the contact ion concentrations defined in Eq. (1). The data is shown for three $Mg^{2+}$ concentrations that correspond to the midpoints at which different TCs form (see Figure 4). Fig. S1 in the Supplemental Information shows similar plots for other solution conditions. **(B-D)** $\Delta\Delta G$ for (A) TC1, (B) TC3 and (C) TC5. The vertical axes are flipped for the ease of comparison to (A). Note that a part of data in A-D is not continuous because nucleotide 842 through 847 do not exist in *T. Thermophilus*. **(E)** Three dimensional positions of nucleotides with distinct peaks, $\Delta\Delta G < -0.5\,k_{\mathrm{B}}T$ ($= -0.31$ kcal/mol), are shown using surface representation in the same color code as (B-D).

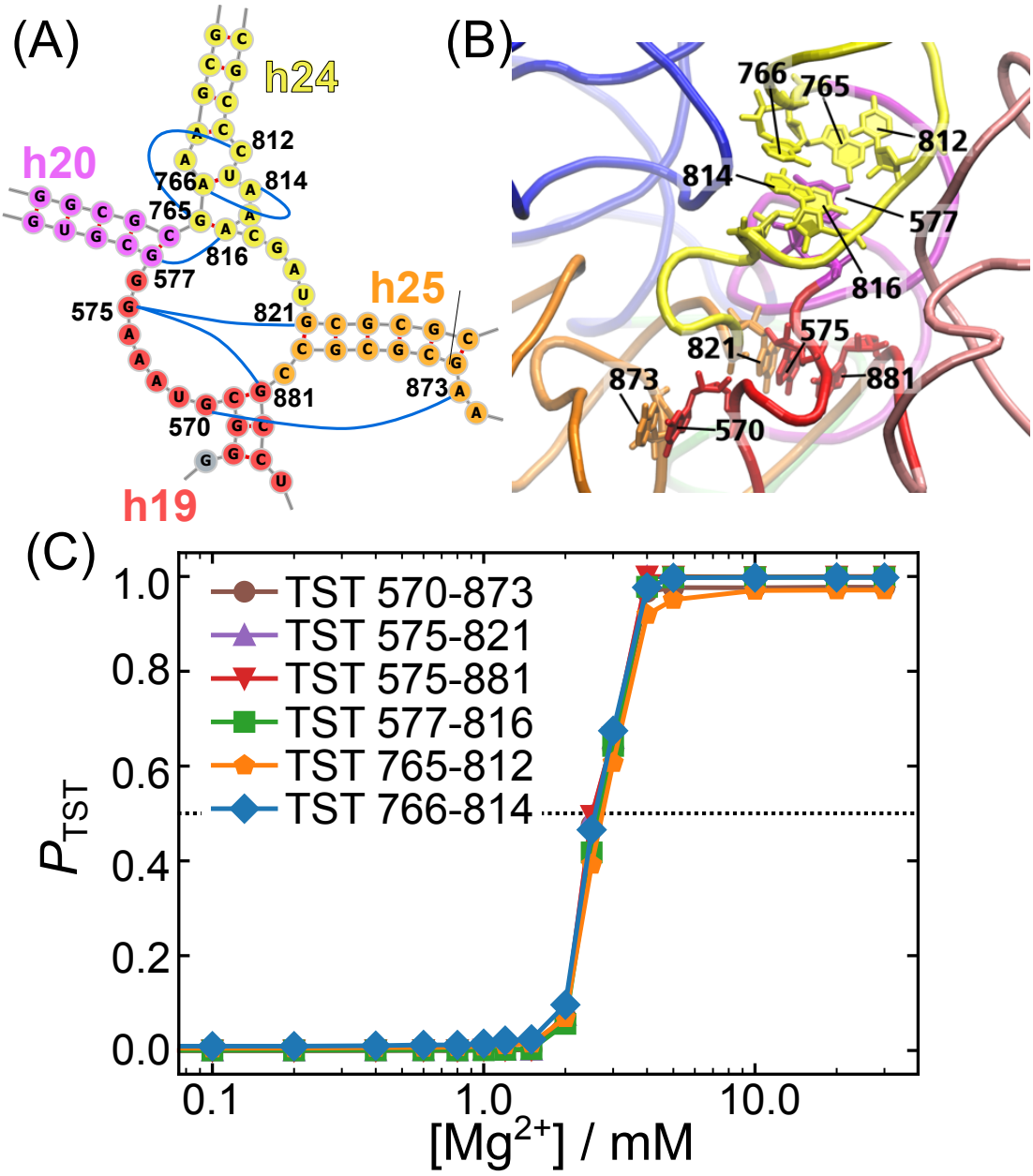



FIG. 6. Formation of tertiary stacking (TST) around the central junction. **(A, B)** Magnified views of the secondary and tertiary structures of the central junction. The TSTs are shown as blue lines. The bases contributing to the TSTs are indicated by the nucleotide numbers. In the tertiary structure, nucleotides involving TSTs are shown by stick representation, and the rest is shown as only the backbone. **(C)** Fraction of TST formation ($P_{\mathrm{TST}}$) as a function of the $Mg^{2+}$ concentration. Because the TSTs fold in cooperative manner, most data points overlap.

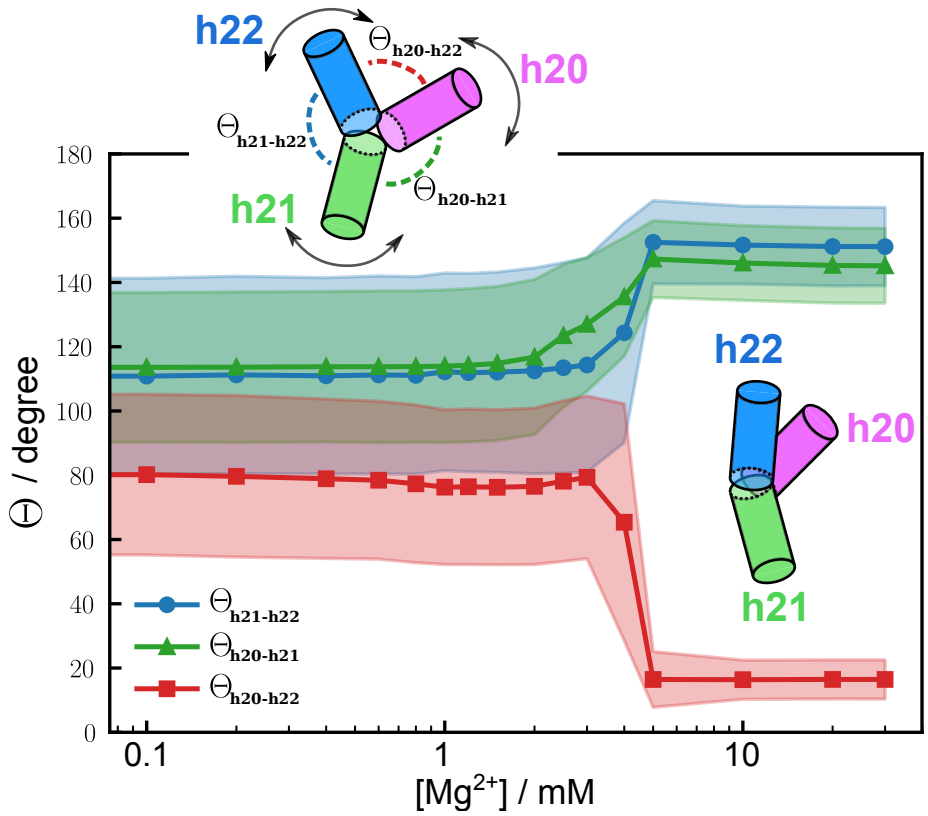



FIG. 7. Changes in the angles between helices at the three-way junction as functions of $Mg^{2+}$ concentration. The angle, $\Theta$, between the specific helix pair is given in the Figure. The filled regions represent the ranges of standard deviations in $\Theta$.

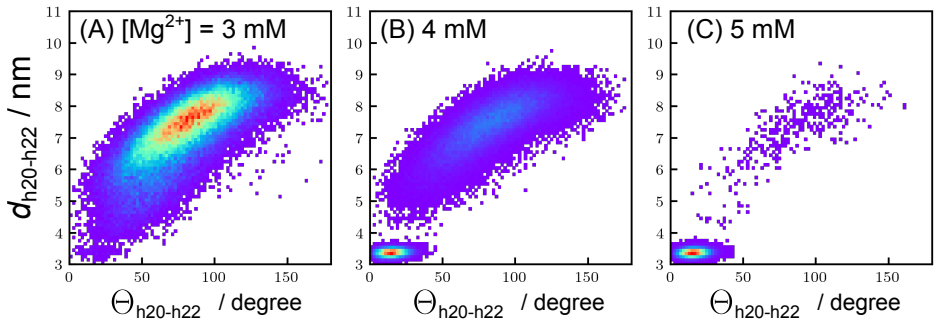



FIG. 8. Two dimensional distributions of the distance ($d_{\text{h20-h22}}$) and the angle $\Theta_{\text{h20-h22}}$ between the helices h20 and h22. The distributions are shown at three $Mg^{2+}$ concentrations around the transition point, (a) 3 mM, (b) 4 mM, and (c) 5 mM. The distance $d_{\text{h20-h22}}$ was measured using the positions of the sugars of G577 and C735. The red color indicates the highest probability and the purple is the lowest probability regions. Fig. S3 in the Supplemental Information shows similar plots for the other two angles, $\Theta_{\text{h20-h21}}$ and $\Theta_{\text{h21-h22}}$.

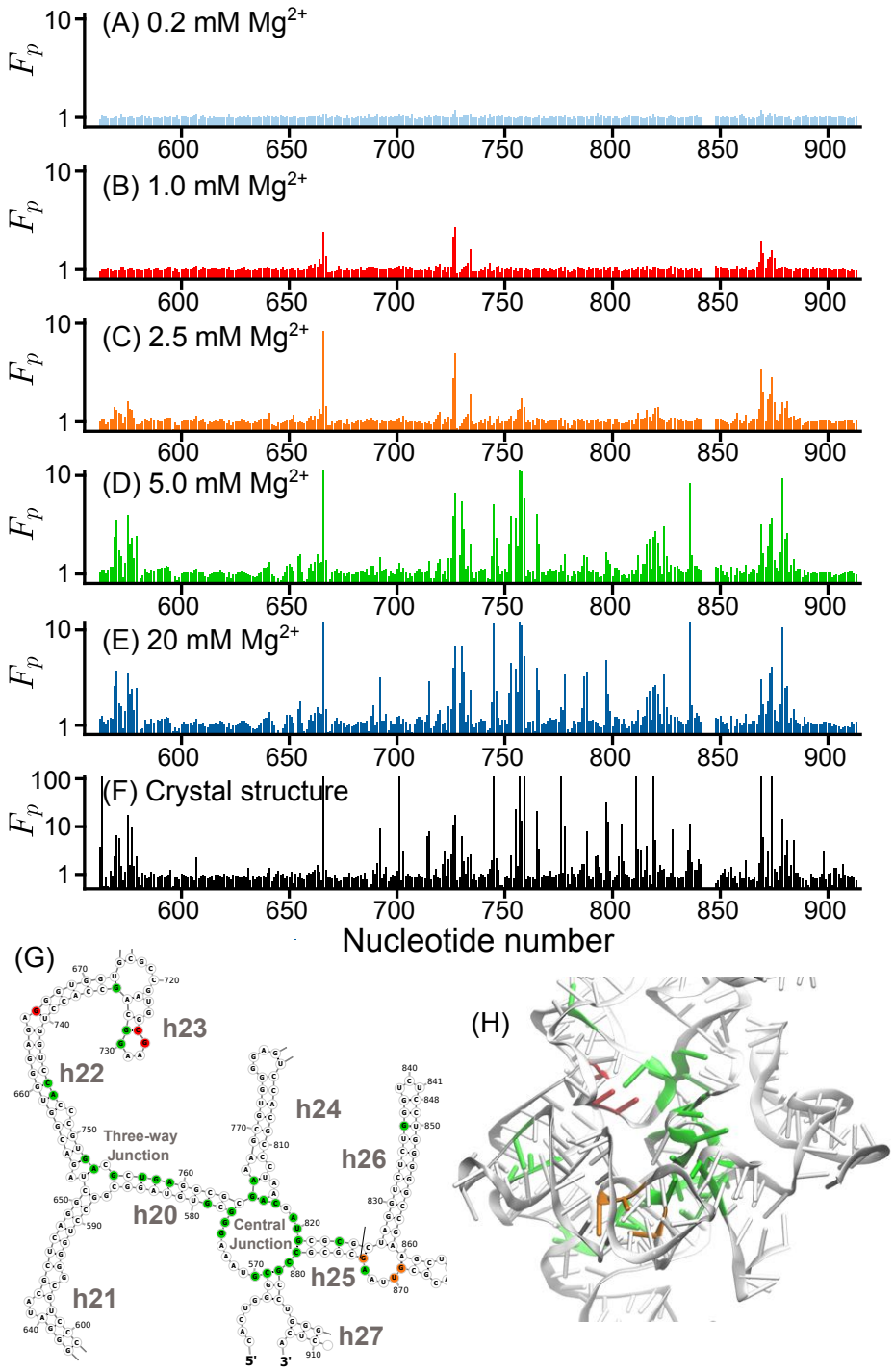
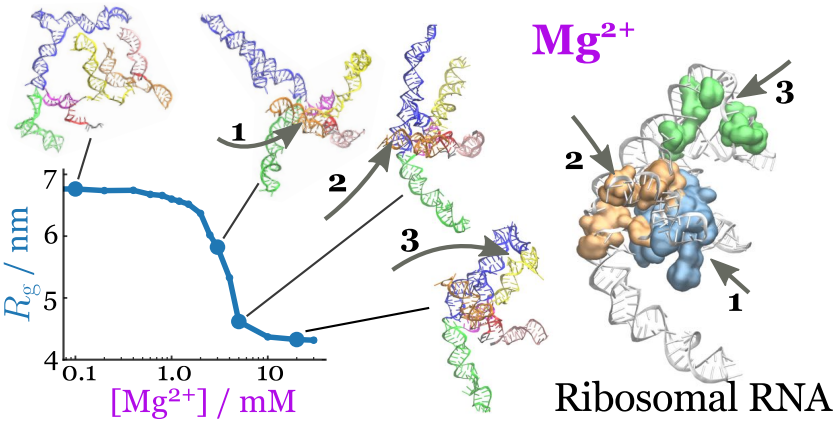



FIG. 9. **(A-E)** Footprinting protection factors estimated based on SASA at nucleotide resolution. $Mg^{2+}$ concentrations is labeled on the top of each panel. **(F)** Protection factors estimated based on SASA calculated for the crystal structure (PDB: 1J5E). A few peaks exceeding $F_p > 100$ were truncated for the sake of clarity. As noted in the text these values are likely to be larger than those under solution conditions. **(G, H)** Secondary and tertiary structures with protected nucleotides highlighted with the same color scheme as in (B-D).




Graphical Abstract

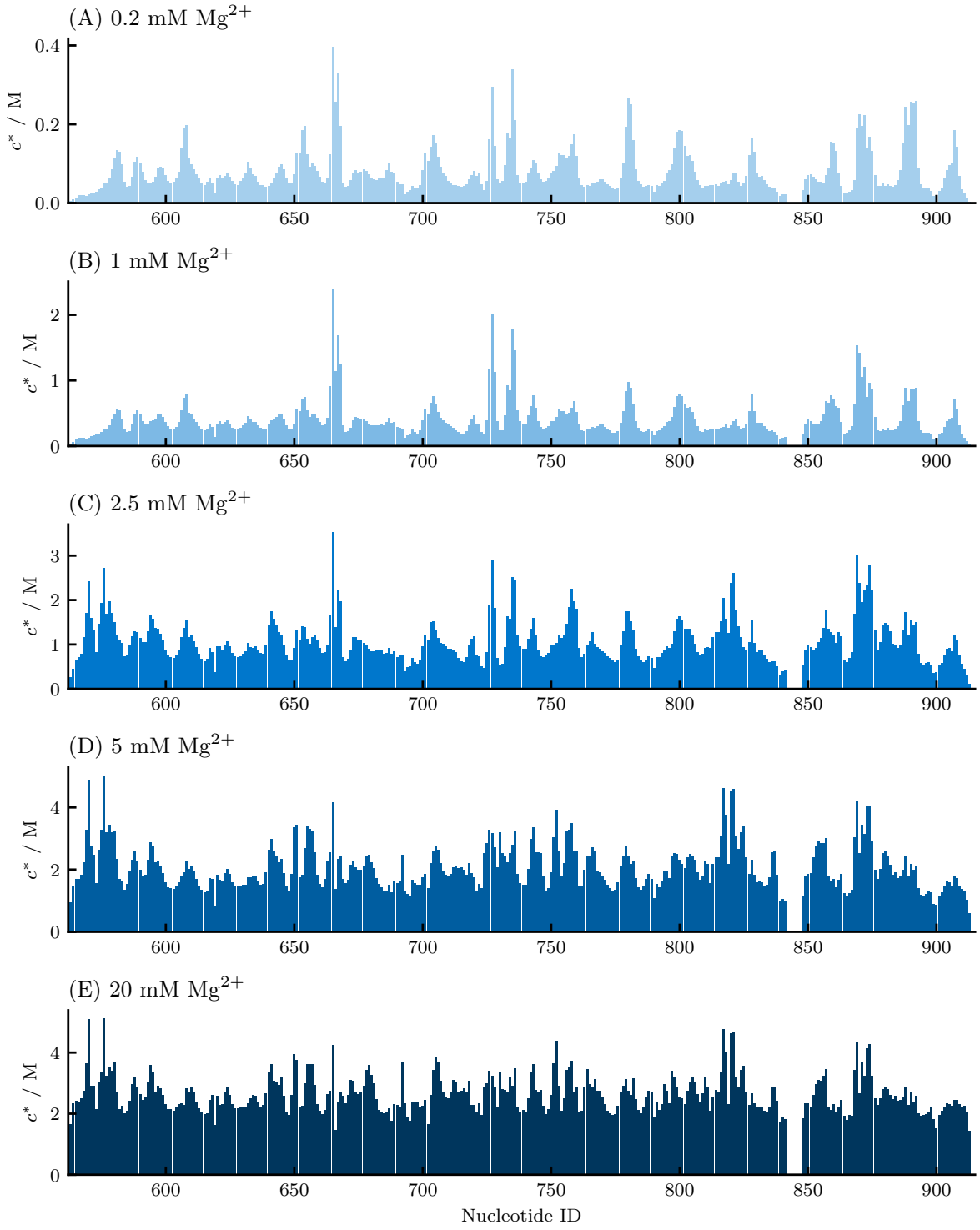



FIG. S1. $Mg^{2+}$ fingerprints at various [$Mg^{2+}$]. The averaged contact ion concentration ($c^*$) of $Mg^{2+}$ around each nucleotide is plotted for various solution condition, [$Mg^{2+}$] = (A) 0.2, (B) 1, (C) 2.5, (D) 5, (E) 20 mM.

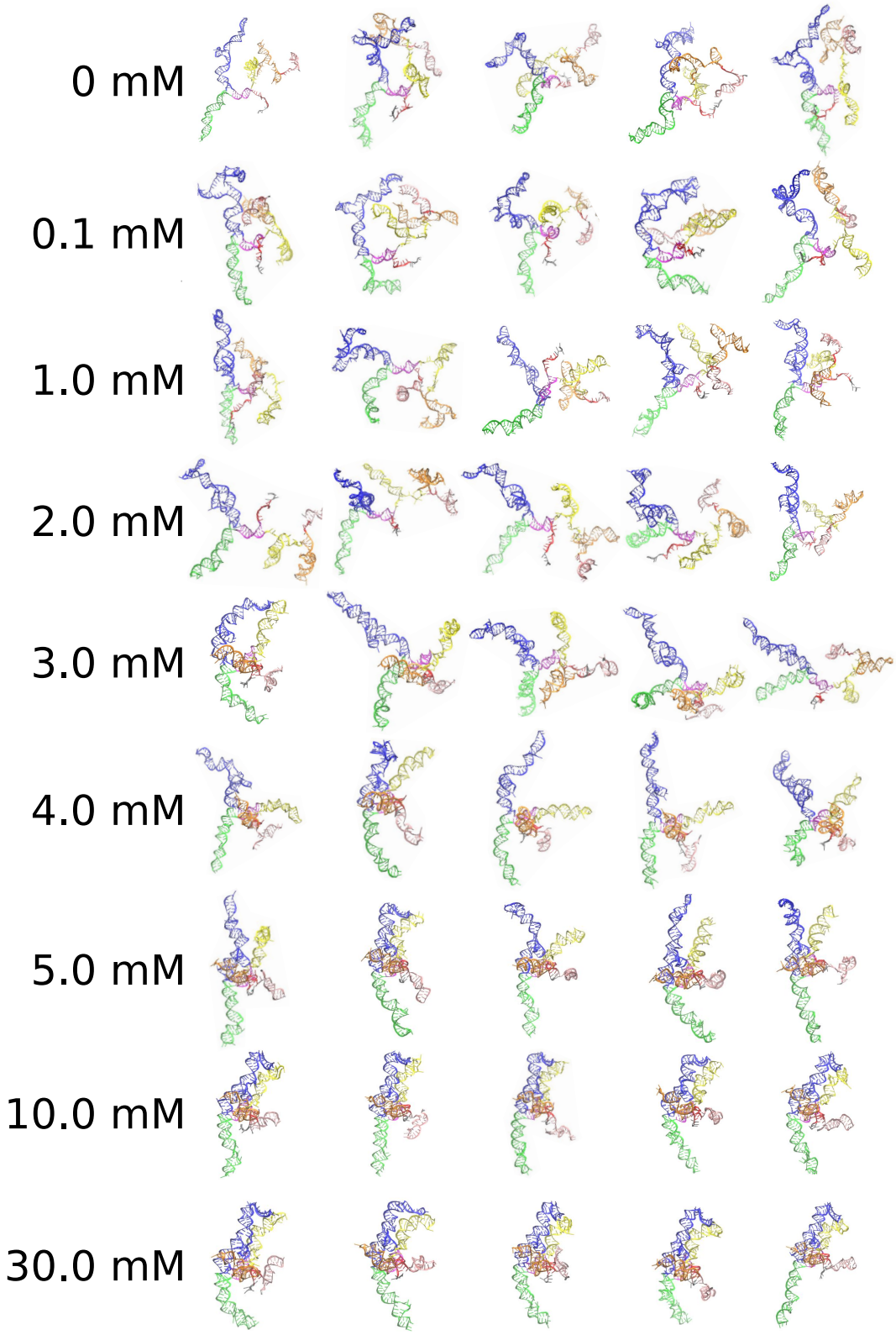



FIG. S2. Variations of the RNA conformations depending on [$Mg^{2+}$] is shown by randomly-chosen five structures from each ensemble.

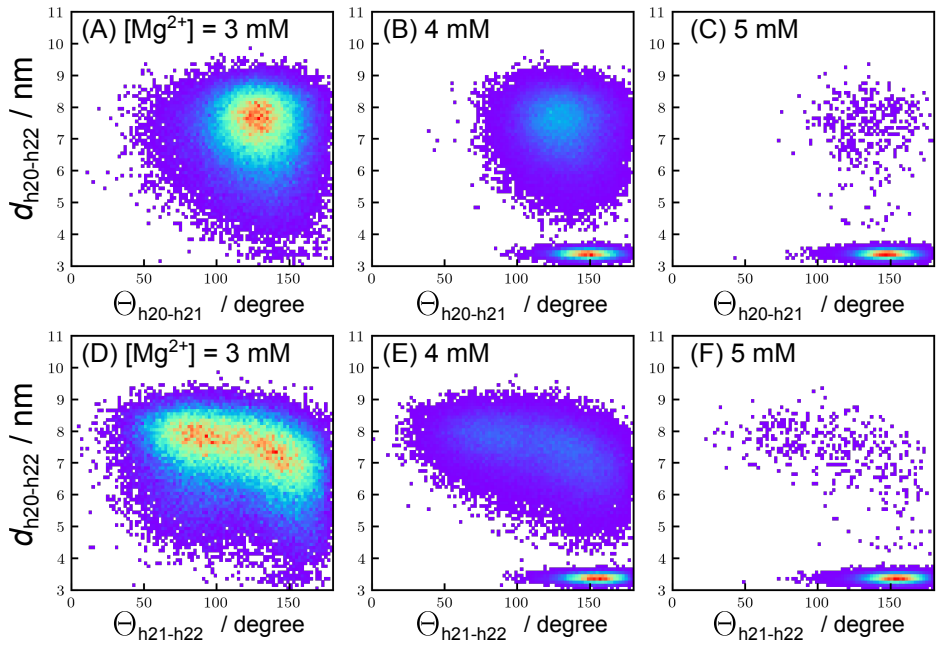



FIG. S3. Two dimensional distributions of the distance ($d_{\text{h20-h22}}$) and angles $\Theta_{\text{h20-h21}}$ (A-C) and $\Theta_{\text{h21-h22}}$ (D-F). The distributions are shown at three $Mg^{2+}$ concentrations around the transition point, (A, D) 3 mM, (B, E) 4 mM, and (C, F) 5 mM. The red color indicates the highest probability and the purple is the lowest probability regions.