The structural stabilization of the κ three-way junction by Mg(II) represents the first step in the folding of a group II intron

Daniela Donghi, Maria Pechlaner, Cinzia Finazzo, Bernd Knobloch and Roland K. O. Sigel*

Institute of Inorganic Chemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

Received June 27, 2012; Revised October 11, 2012; Accepted October 26, 2012

ABSTRACT

Folding of group II introns is characterized by a first slow compaction of domain 1 (D1) followed by the rapid docking of other domains to this scaffold. D1 compaction initiates in a small subregion encompassing the κ and ζ elements. These two tertiary elements are also the major interaction sites with domain 5 to form the catalytic core. Here, we provide the first characterization of the structure adopted at an early folding step and show that the folding control element can be narrowed down to the three-way junction with the κ motif. In our nuclear magnetic resonance studies of this substructure derived from the yeast mitochondrial group II intron Sc.ai5, we show that a high affinity Mg(II) ion stabilizes the κ element and enables coaxial stacking between helices d and d”, favoring a rigid duplex across the three-way junction. The κ-element folds into a stable GAAA-tetraloop motif and engages in A-minor interactions with helix d’. The addition of cobalt(III)hexammine reveals three distinct binding sites. The Mg(II)-promoted structural rearrangement and rigidification of the D1 core can be identified as the first micro-step of D1 folding.

INTRODUCTION

Self-splicing group II introns belong to the class of large ribozymes. Recognized as an independent class only in 1982 (1), they are found primarily in organellar genes of plants, fungi and lower eukaryotes, but also in many bacteria (2,3). It is thought that they were crucial in modeling the genome of almost all terrestrial life forms: a common ancestry has been proposed for group II introns and most nuclear introns, components of the telomerase and the abundant LINE–element form of transposons (4), as well as the eukaryotic spliceosome (5–9). Moreover, they have been found to be able to reinsert into RNA and DNA via a retro-homing mechanism (10–12), thus being mobile genetic elements responsible for the continuous diversification of different genomes.

Group II introns as all other RNA molecules are characterized by a secondary structure usually stabilized by monovalent metal ions, whereas the tertiary structure requires the presence of divalent metal ions (13). The latter are eventually also involved in catalysis (14–17). Despite a lack of primary structure conservation, all group II introns present a similar secondary structure, consisting of six domains originating from a central wheel (2,18,19). Domain 1 represents the largest domain. It is an independent folding unit (20,21) and provides a scaffold for the other domains to dock. Domain 2 and 3 stabilize the folding and enhance the catalytic activity of the whole intron, respectively. Domain 4 can contain an open reading frame encoding for a maturase protein useful for intron activity in vivo (10). Domain 5 is the genetically most conserved one and shares many features with the spliceosomal RNA U6 (22). It represents, together with domain 1, the minimal core to enable catalysis (3,5). Finally, domain 6 contains the bulged adenosine implicated as nucleophile in the first step of splicing (23). Several studies have been devoted to understanding and describing both the folding and the catalytic mechanism of group II introns. One of the best studied systems is the Saccharomyces cerevisiae intron Sc.ai5γ from baker’s yeast (24–29).

In particular, a shorter version of the whole intron containing domains 1, 3 and 5 (D135) has been extensively used for folding studies (28–32). The folding of large RNAs usually involves a rapid collapse on metal ion addition (33,34). The collapse can be unspecific or can be initiated by the formation of specific RNA substructures. Pyle et al. (34) showed that D135 folds in an apparent two state model: the first stage is slow (in the order of seconds) and requires the formation of specific
structural features within D1. This stage is then followed by the fast docking of all the other domains. Both stages require above-physiological Mg(II) concentrations 

*in vitro*, the first one has a $K_{D,Mg}$ of $\sim 13$ mM (21), the second of 20–40 mM (30).

A small subregion in the central core of D1 comprising the three-way junction around the $\kappa$ motif (25) and an 11-nt tetraloop receptor ($\zeta$) was proposed to act as folding control element ($D1\kappa\zeta$, Figure 1) (24,25,28). In the presence of Mg(II), this region adopts a specific structure, which allows the whole intron to attain the active native state (29,34). Moreover, it was recently shown that Mss116p, a DEAD box protein, which promotes folding *in vivo* substituting in part for the Mg(II) (35,36) acts by stabilizing this small substructure (37). In a later stage of folding, $D1\kappa\zeta$ also offers the docking site for D5, as supported by biochemical studies and a recent crystal structure of a group IIC intron from *Oceanobacillus iheyensis* (38–40).

We have now investigated the structure of $D1\kappa\zeta$ by means of nuclear magnetic resonance (NMR) spectroscopy, and demonstrate the central role of Mg(II) in its stabilization. Our data show that the helices next to the three-way junction are destabilized in the absence of Mg(II). Coordination of Mg(II) leads to a rigidification of the central core of this substructure, resulting in the coaxial stacking of two helices as was suggested by Costa et al. (41) from comparative sequence analysis.

The bulged nucleotides of the $D1\kappa\zeta$ junction form a classical GAAA tetraloop motif interrupted between the third and the forth nucleotide by a short helix. In addition, adenosines in this motif are engaged in A-minor interactions with the adjacent helix. Interestingly, this A-minor motif is not present in the only available crystal structure of a group IIC intron (38–40), where the two adenines are interacting with the docked D5.

Coaxial stacking of helices and A-minor motifs have been recognized as among the most important interactions governing RNA tertiary structure (42) and are here combined together within this small subregion. Helices next to multibranch junctions tend to maximize coaxial stacking to reduce overall folding free energy (43), and the stacking has been shown to depend on di- or multivalent ions (44). Our detailed NMR data allow us to locate a possible binding pocket at the junction that explains the requirement of multivalent metal ions for stabilization of this region. Which helices in a junction stack and how they are oriented is strongly influenced by the bulged nucleotides. Geary et al. (45) have highlighted the versatility of A-minor interactions in determining the architecture of junctions and the relative orientation of the helices. To the best of our knowledge, the structural arrangement of the A-minor interactions in the three-way junction present in $D1\kappa\zeta$ has not been described previously.

This NMR solution structure of $D1\kappa\zeta$ adds to the scarce structural data on group II introns (23,27,38–40,46–48). Moreover, we reveal $\kappa$ to be a flexible, bimodal structure, which keeps its characteristic fold both in the context of a minimal A-minor junction motif to the adjacent helix (as observed in the $D1\kappa\zeta$

---

**Figure 1.** (a) Predicted secondary structure of the group II intron $Se.ai57$ from *S. cerevisiae* (the $\kappa$-$\zeta$ region studied in this work is highlighted). (b) The wild-type $\kappa$-$\zeta$ region (top) and the 49 nucleotide construct used for NMR studies (bottom). Nucleotides that were added at the ends to stabilize helix $d''$ and helix d are drawn as contours.
construct), as well as in the tertiary interaction with D5 (in the full group II intron).

MATERIALS AND METHODS

All chemicals were purchased from either Fluka-Sigma-Aldrich (Buchs, Switzerland) or Brunschwig Chemie (Basel, Switzerland) at puriss p.a. or biograde. All solutions were prepared by using autoclaved and filtered ultrapure water. All DNA oligonucleotides were purchased from Microsynth (Balgach, Switzerland) and purified by polyacrylamide gel electrophoresis before usage. The nucleotide 5′-triphosphates (NTPs) were obtained from Cambridge Isotope Laboratories (Andover, USA). The RNA samples were puriss p.a., and purified by polyacrylamide gel electrophoresis before usage. The nucleotide 5′-triphosphates (NTPs) were obtained from GE Healthcare (Glattbrugg, Switzerland), with the exception of uridine-5′-triphosphate (UTP) that was purchased from Acros-Organics (Geel, Belgium). 13C- and 15N-labeled NTPs were obtained from Silantes Chemie (Basel, Switzerland) at puriss p.a. or biograde. All chemicals were purchased from either Fluka-Sigma-Aldrich (Buchs, Switzerland) or Brunschwig Chemie (Basel, Switzerland) at puriss p.a. or biograde.

RNA transcription and sample preparation for NMR experiments

The RNA samples were transcribed and purified following standard procedures (49). The reaction mixture contained 5 mM of each NTP, 0.5 μM of the double-stranded DNA template, 0.1% Triton X-100, 40 mM TRIS-HCl [Tris(hydroxymethyl)-aminomethan-HCl, pH 7.5], 40 mM DTT, 2 mM spermidine and 30 mM MgCl₂. The amount of polymerase depended on the effectiveness of the polymerase itself and was optimized for each polymerase batch. To avoid 3′-end inhomogeneity due to run-off transcription by the T7-RNA polymerase, we used a home-made T7 polymerase (49) and purified through denaturing polyacrylamide gel electrophoresis (15–18%). The RNA samples were transcribed and purified following standard procedures (49). The reaction mixture contained 5 mM of each NTP, 0.5 μM of the double-stranded DNA template, 0.1% Triton X-100, 40 mM TRIS-HCl [Tris(hydroxymethyl)-aminomethan-HCl, pH 7.5], 40 mM DTT, 2 mM spermidine and 30 mM MgCl₂. The amount of polymerase depended on the effectiveness of the polymerase itself and was optimized for each polymerase batch. To avoid 3′-end inhomogeneity due to run-off transcription by the T7-RNA polymerase, we used a home-made T7 polymerase (49) and purified through denaturing polyacrylamide gel electrophoresis (15–18%).

All solutions were prepared by using autoclaved and filtered ultrapure water. Usage of KCl is critical, as all the TRIS must be removed from the solution, as it can compete with RNA for metal ion binding (54). The samples were lyophilized and re-dissolved in 300 μl 100% D₂O or H₂O/D₂O (90:10) and transferred into Shigemi™ tubes before the acquisition of the NMR spectra. All the samples contained 60 mM KCl, 10 μM EDTA and were of pH 6.5–6.9.

NMR spectroscopy

D1κε was investigated by solution NMR methods generally used in RNA structural studies (55–57). The NMR spectra were recorded on a Bruker AV700 MHz spectrometer equipped with a CP-TXI z-axis pulsed field gradient CryoProbe®, on a Bruker AV600 MHz equipped with a TCI z-gradient CryoProbe® and on a Bruker AV500 MHz equipped with a TCI z-gradient CryoProbe® at CERM in Florence. The spectra were acquired either on D1κε constructs prepared with isotopes in natural abundance or constructs prepared by using partially deuterated nucleotides or 13C-, 15N-labeled nucleotides. The use of partially deuterated samples was crucial to avoid the massive overlap in the sugar regions: the sequential walk could be followed in the H6/H8-H1′ region and in the H6/H8-H2′ region. The sugar pucker of the nucleotides was determined by Total Correlation Spectroscopy (TOCSY) experiments with a sample with natural isotope abundance. Non-exchangeable protons were assigned from 2D [1H, 1H]-NOESY (Nuclear Overhauser Effect Spectroscopy) spectra in D₂O, recorded at various temperatures (290, 295, 300, 305, 310 K) and with different mixing times (60, 150, 250, 350 ms); water suppression was either achieved by using a water saturation pulse or an excitation sculpting pulse scheme. The same general spectral features were observed between 290 K and 310 K, but 300 K was found to be the optimum. The pH was kept between 6.5 and 6.9 without significant modification in the spectra. 2D [1H, 1H]-TOCSY spectra (50 ms mixing time) were acquired to assess the sugar pucker. Exchangeable protons were assigned from [1H, 1H]-HSQC experiments recorded in H₂O/D₂O at different temperatures (275, 280, 282, 285 K) with 150 ms mixing times, employing a 1-1 spin-echo pulse sequence or the excitation sculpting pulse scheme for water suppression. [1H, 13C]-HSQC experiments were employed to attribute sugar and aromatic carbons. Two separate experiments were usually recorded for the aromatic and the aliphatic range of the 13C resonances. [1H, 15N]-HSQC experiments were used to attribute the imino nitrogens. band-Selective Optimized Flip-Angle Short-Transient (SOFAST) sequences were used for these experiments (58). 3D-[1H, 1H]-NOESY-[1H, 13C]-HSQC were also performed (59).

The base-pairing pattern was established by recording a 2D J₆,₇ HNN-COSY. Such experiment correlates the imino nitrogens of uracil and guanine across the H-bond to the N1 of adenine or the N3 of cytosine on the other side of the double helix (60).
Titration with K+, Mg(II) and Co(NH₃)₃⁺

A 0.2 mM sample in H₂O/D₂O was titrated with increasing amounts of KCl (60, 100, 150, 200, 250, 300 and 400 mM). 1D ¹H NMR experiments were recorded at each step to evaluate the changes of the imino protons. The solution became increasingly turbid on addition of large amounts of salt.

A 0.25 mM sample in H₂O/D₂O was titrated with increasing amounts of Mg(II) (0.5, 1, 1.5, 2, 2.5, 3, 4, 5, 7, 9 and 12 mM) at 300 K. [¹H,¹H]-NOESY experiments were recorded at each step. The chemical shift changes were plotted in bar diagrams using Origin® and gnuplot (http://www.gnuplot.info/). [¹H,¹H]-HSQC on the aromatic region and 2J[¹H,¹⁵N]-HSQC (in the region of N7 of guanosine and adenosine and N1 and N3 of adenosine) experiments were recorded with a 1.1 mM ¹⁵N-, ¹³C-labeled sample in the presence of increasing amounts of Mg(II) (0.5, 1, 1.5, 2, 3, 4, 6, 8 and 10 mM).

A 0.3 mM partially deuterated sample in D₂O was titrated with increasing amounts of Co(NH₃)₃⁺ (0.6, 1.2, 1.8, 2.4, 3, 3.6, 4.2 mM) at 300 K. [¹H,¹H]-NOESY experiments were recorded at each step.

Structure calculation

Distance restraints were derived from [¹H,¹H]-NOESY spectra measured in D₂O at 60 and 250 ms mixing time and from [¹H,¹H]-NOESY spectra measured in 90%H₂O/10%D₂O at 150 ms mixing time. Not or only moderately overlapping peaks were integrated in Sparky (http://www.cgl.ucsf.edu/home/sparky/), and distances were calibrated to pyrimidine H5H6 and H1’H2’ peak volumes using DYANA’s CALIBA macro (64). Based on these, all assignable peaks were categorized as strong (1.8–3 Å), medium (1.8–4.5 Å), weak (3–6 Å) and very weak (4–7 Å). In addition, dihedral angle restraints were used: For bases in stable base pairs and helical regions, chi was restrained to anti-conformation (-160° ± 20°). G32 was restrained to syn conformation (60° ± 20°) owing to its strong intranucleotide H8-H1 peak, and U29, G15, A16, A17 and A18 were restrained to anti according to literature data (27,65). For all other bases, chi was left unrestrained. Sugar puckers were restrained to 3’-endo conformation (delta = 85°, n1 = -25°, n2 = 37.5°, ±10°) in helical regions, to 2’-endo conformation for residues showing strong H1’H2’ and medium H1’H3’ cross-peaks in 50 ms [¹H,¹H]-TOCSY spectra (delta = 145°, n1 = 25°, n2 = -35°, ±10°) (A6, A16, U30, C31 and A43) (27,66–68), and they were left unrestrained in all other cases. Backbone angles were restrained to standard A-RNA values in regions where the NOESY peak pattern and base-pairing strongly suggest standard helical conformation (alpha = -68°, beta = 178°, gamma = 54°, epsilon = -153°, zeta = -71°, ±20°), and left unrestrained in other regions. Hydrogen bonds in Watson–Crick base pairs were kept stable by short distance restraints between the hydrogen and the acceptor, as well as between the donor and the acceptor and by planarity restraints.

Starting from an extended conformation, an initial ensemble of 100 structures was calculated using CNS 1.21 (69,70) (40 ps torsion angle dynamics at 20000 K, followed by slow cooling for 90 ps in torsion angle space and 30 ps in cartesian space). This was followed by 60 ps of refinement using XPLOR-NIH 2.24 (71,72), slowly cooling from 3000 to 50 K, during which 35 single-bond ¹H-¹³C RDC restraints were introduced (with a force constant increasing from 0.01 to 1.0 kcal mol⁻¹ Hz⁻²). Axial component and rhombicity of the alignment tensor were determined as -20.8/0.2 in a gridsearch using XPLOR-NIH (73). RDC error limits were set to generous 1.5 Hz, to allow for the uncertainty introduced by the facts that peaks are broadened in the presence of Mg(II) and the Pf1 phages.

The 20 lowest energy structures (none of them had NOE violations >0.2 Å, dihedral angle violations >5° or RDC violations >3 Hz) were selected as the final ensemble and analyzed using MolMol (74), Pymol (http://www.pymol.org) and APBS (http://www.poissonboltzmann.org/apbs) (75).

An additional set of structures was refined in XPLOR-NIH including distance restraints (6.0 Å to the central Co³⁺) to three cobalt(III)hexammine molecules.
RESULTS

Design of D1kζ

This work focuses on the solution structural characterization of the central core of domain 1, comprising the κ and ζ elements, of the mitochondrial group II intron Sc.ai5γ from baker’s yeast (Figure 1). To obtain a construct suitable for NMR measurements, a tetraloop needed to be inserted either at the A218-U374 or at the U386-A195 end. The resulting different constructs (Supplementary Figure S1) were transcribed and tested by NMR (see Supporting Information, Supplementary Figures S1–S5). The most promising construct was found to be a 49 nt long molecule (D1-49 in Figures S1–S5). The most promising construct was prepared and used to attribute the resonances belonging to the κζ region and the UUCG tetraloop in our experimental conditions.

Proton attribution of D1kζ

D1kζ contains structural features common to many RNAs (76) like a three-way junction, an internal loop, as well as GAAA and UUCG tetraloops. Literature data from other RNA molecules are available on the chemical shifts of the tetraloop receptor region (ζ region) (66,68), as well as on the GAAA (27) and UUCG tetraloops (65,67). Comparison with literature data supports the assignment for the exchangeable and non-exchangeable protons. Moreover, a 27-nucleotide long construct (D1-27 in Supplementary Figure S1) was prepared and used to attribute the resonances belonging to the κζ region and the UUCG tetraloop in our experimental conditions.

D1kζ was prepared in good yields using wild-type nucleotides, partially deuterated nucleotides and 15N-, 13C-labeled nucleotides (77). In the absence of Mg(II), almost all non-exchangeable protons could be attributed except those between C11 and C14 and between A37 and A39 (located next to the junction, Figure 1), where the sequential walk is interrupted. Moreover, H1′ protons of G19 and G40 are missing (Figure 1). In line with the interruption of the sequential walk, hydrogen bonds of base pairs directly next to the κ region are not visible in a 2D J_NN HNN COSY experiment (data not shown) in the absence of Mg(II). Furthermore, all neighboring base pairs (U10-A39, C14-G19, C25-G36, Figure 1) are characterized by broad imino resonances at 275 K. In total, only 9 of the 15 expected base pairs could be experimentally confirmed (Figure 2, left). The κ extension helix (Figure 1) is hardly formed at all, which is in line also with the absence of G19H1′.

In summary, the resonances next to the three-way junction are either absent or very broad, and we do not observe base-pairing directly neighboring the three-way junction, indicating that no stable structure is achieved in this region in the absence of divalent ions.

Stabilization of D1kζ on Mg(II) addition

On addition of Mg(II) to D1kζ, the imino resonances that were previously missing appear (U20H3, U24H3 and G38H1, Figure 1), and their broad neighbors (U10H3, G19H1 and G36H1) sharpen. In contrast, no new signals are observed on addition of up to 300 mM KCl only (Supplementary Figure S6). The stabilization of H-bonds neighboring the three-way junction was confirmed by a 2D J_NN HNN COSY experiment (data not shown). The newly observed sequential NOE cross-peak between U24H3 and G38H1 suggests that helix d′ and helix d″ form a continuous helix (Figures 1 and 2, right).

Also, non-exchangeable proton resonances appear and increase in intensity, as observed in [1H,1H]-NOESY and can be identified with the help of [1H,13C]-HSQC spectra. At 3 mM Mg(II), the missing G19H1′ appears at 3.79 ppm, and the corresponding sugar Cl′ carbon shows up in the [1H,13C]-HSQC spectrum (Figure 3). These strong upfield shifts are well-known for the nucleotides immediately following a GAAA tetraloop, due to the ring current of the preceding adenine (27). At 4 mM Mg(II), the gap in the sequential walk between C11 and C14, as well as between A37 and A39, is closed. C11H1′ is particularly strongly influenced by the addition of Mg(II). A broad peak in the absence of Mg(II), it broadens more [until 3 mM Mg(II)] and then gains intensity again, all the while moving downfield to an unusually deshielded position at 6.10 ppm (Figure 3). Interestingly, no matter the amount of added Mg(II), G40H1′ is never visible. No peaks for G40H8-H1′ and C41H6-G40H1′ could be identified under any of our experimental conditions, even if the equivalent signals in similar constructs like D1-27 and D1-45 (Supplementary Figures S1 and S2), containing the same region from helix d to d′, but no κ region, are visible and sharp. On the other hand, the missing G40H8-H1′ and C41H6-G40H1′ cross-peaks do become visible in D1kζ on the addition of cobalt(III)hexammine, suggesting a tighter binding of cobalt(III)hexammine with a stronger stabilizing effect. This is also reflected by generally sharper signals in [1H,1H]-NOESY spectra recorded in the presence of cobalt(III)hexammine (Supplementary Figure S7).

Figure 3 shows the appearance of crucial resonances at high Mg(II) concentration, both in the [1H,1H]-NOESY and in the [1H,13C]-HSQC experiments. Even if some of these signals are present also at lower Mg(II) concentration, they are sharper and better defined at 9–12 mM Mg(II) concentration.

In contrast to the other two linker regions of the junction, resonances of A22 and A23 in the third linker are observable both in the absence and in the presence of Mg(II).

In the presence of 4 mM Mg(II), a cross-peak between A18H2 and C9H1′ appears, together with a less-pronounced cross-peak between A18H2 and G8H2′. These cross-peaks would be expected for a cognate GAAA-tetraloop-receptor interaction, which would involve a base quadruple of G15-A18-G8-C41 and an AA platform involving A43 and A44 (78). Neither A18H2-C9H1′ nor A18H2-G8H2′ can be satisfied by an
Figure 2. Imino proton region of NOESY spectra in the absence (left) and presence (right) of 10 mM Mg(II). The asterisks indicate G1 and G2 imino protons from n+1 species (3'-end heterogeneity caused by T7 polymerase). The inset structures denote stable base-pairing that is observed in JNN HNN COSY spectra in green. Base pairs that are less stable, but still display observable imino-cross-peaks, are marked in orange. In the right panel, the NOE indicating coaxial stacking of helices d' and d'' is highlighted in a green box.

Figure 3. [1H,1H]-NOESY (top) and [1H,13C]-HSQC (bottom) in the absence (left) and presence (right) of 9 mM Mg(II) and 12 mM Mg(II), respectively. Resonances that appear only in the presence of Mg(II) are highlighted in green. The same effect is observed with cobalt(III)hexammine (Supplementary Figure S7). All spectra were measured at 300 K, pH 6.6–6.8, 0.4–0.7 mM D1ε, 600 or 700 MHz.
intramolecular interaction (Figure 4). The stronger A18H2-C9H1′ was confirmed to be an intermolecular peak in a double-half-filtered experiment (see Materials and Methods) on a sample containing both wild-type D1κζ and 13C-, 15N-enriched D1κζ and Mg(II). In addition, native gel assays showed that a small amount of the D1κζ dimer is already formed in the presence of 0.5 mM Mg(II) (data not shown).

Although a dimer could also be formed through an asymmetric interaction of one GAAA tetraloop of one molecule with the receptor of another, we have evidence that D1κζ forms a symmetric homodimer, with two symmetric tetraloop-tetraloop receptor (TL-TLR) interactions. In the classical TL-TLR interaction between the 11-nt TLR and a GAAA tetraloop (68), we would expect strong characteristic chemical shift changes in several tetraloop and receptor proton resonances (68), which we do not observe here. On the other hand, we see exchange cross-peaks between aromatic proton resonances of the GAAA tetraloop and resonances corresponding to the receptor-bound form of the tetraloop (68) at 9 mM Mg(II). This suggests that in contrast to the more stable contact between A18 and C9 (indicated by the sharp and pronounced A18H2-C9H1′ peak), interaction of the rest of the TL with the receptor is very transient.

One single TL-TLR interaction is not very strong (KD = 0.4 mM) so that already Davis et al. (68) followed the strategy of a homodimer to study this interaction by solution NMR. While the symmetric dimer of Davis et al. was designed for an optimal interaction, our construct obviously was not. Because the TL and the TLR receptor are not positioned exactly one helix turn apart, there is bound to be some steric hindrance for a symmetric interaction, which will prevent two stable classical TL-TLR receptor interactions. This explains why we see only exchange cross-peaks indicative of a transient interaction.

Nevertheless, the minor presence of the dimer is not expected to influence the main features of D1κζ, which are d′-d″ coaxial stacking and the A-minor motif involving κ (see later in the text).

**D1κζ forms a stable three-way junction in solution**

An ensemble of D1κζ structures with an overall RMSD (Root Mean Square Deviation) of 8.21 Å was calculated using 848 NOE distance restraints collected from spectra recorded in the absence or presence of 9 mM Mg(II) [resonances peripheral from the junction are much sharper and allow for unequivocal assignment in the absence of Mg(II)] (Table 1) in restrained molecular dynamics simulations. In case of experimental evidence for Watson–Crick base pairing from sequential imino proton cross-peaks and JNN HNN COSY spectra (60), and if the peak pattern in [1H,1H]-NOESY spectra agreed with A-helical conformation, hydrogen bonds and standard backbone angles (±20°) were imposed. No backbone angles were restrained in the loop and bulge regions. Including a total of 35 RDC restraints collected in the presence of Mg(II) improved the overall RMSD to 4.3 ± 1.23 Å (Figure 4). Both UUCG and GAAA tetraloop as well as the helical regions converge well (RMSDs, see Table 1), with helix d′ and d″ almost forming a continuous helix. The three adenines of the tetraloop receptor mostly stack into the helix, which is typical for the undocked tetraloop receptor (66) but also other conformations with protruding A6 or A44 are compatible with the NOE restraints that are relatively weak in
RMSDs (for all heavy atoms to mean structure) (Å)

RMSDs are given for the 20 structures of 200 with lowest energies.

Table 1. NMR restraints and structural statistics for the D1κζ structure

<table>
<thead>
<tr>
<th>Restraint statistics</th>
<th>Without RDCs</th>
<th>With RDCs</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOE-derived distance restraints</td>
<td>848</td>
<td>848</td>
</tr>
<tr>
<td>Intranucleotide (i = j)</td>
<td>256</td>
<td>256</td>
</tr>
<tr>
<td>Sequential ([i–j] = 1)</td>
<td>420</td>
<td>420</td>
</tr>
<tr>
<td>Medium range (1 &lt;</td>
<td>i–j</td>
<td>&lt; 5)</td>
</tr>
<tr>
<td>Long range</td>
<td>134</td>
<td>134</td>
</tr>
<tr>
<td>Dihedral restraints</td>
<td>276</td>
<td>276</td>
</tr>
<tr>
<td>Hydrogen bond restraints</td>
<td>83</td>
<td>83</td>
</tr>
<tr>
<td>RDCs</td>
<td>35</td>
<td></td>
</tr>
</tbody>
</table>

Restraint statistics

| Overall                        | 8.21 ± 2.69  | 4.30 ± 1.23  |
| Helix d" (24–37)               | 2.24 ± 0.70  | 1.82 ± 0.45  |
| Helix d and d" (1–11, 24–49)   | 1.56 ± 0.68  | 1.81 ± 0.66  |
| Helix d" and d" (8–12, 21–41)  | 5.00 ± 2.10  | 3.24 ± 1.00  |
| NOE violations >0.2Å            | 7.39 ± 2.82  | 4.11 ± 1.35  |
| Dihedral violations >5°        | 3.44 ± 1.20  | 2.76 ± 0.83  |
| Average NOE r.m.s. deviation (Å)| 0.007        | 0.008       |
| Average dihedral r.m.s. deviation (°)| 0.053 | 0.084 |

Metal ion binding sites: Mg(II) and cobalt(III)hexamine

Metal ions are crucial for stabilizing RNA structure and promoting catalysis (82). By using NMR, it is possible to localize metal ions and also to quantify the strength of the interaction (83). To locate and characterize possible binding sites in D1κζ, we followed 1H, 13C and 15N chemical shift changes and line broadening in [1H,1H]-NOESY, 1J-[1H,13C]-HSQC and 2J-[1H,15N]-HSQC on Mg(II) addition (see ‘Materials and Methods’ section). Furthermore, cobalt(III)hexamine, which is often used as a mimic for the hexaqua complex of Mg(II), was added to observe NOE cross-peaks between the protons of NH₃ and the protons of the RNA (84–86).

Proton resonances of a partially deuterated sample of D1κζ were attributed in [1H,1H]-NOESY spectra, and their chemical shifts plotted against increasing Mg(II) concentration. Some curves are clearly biphasic (Figure 5a), indicating a second effect, which could be structural rearrangement or a second metal ion binding close-by. This behavior mostly involves nucleotides located in the region of the GAAA loop, κ, the starting point but also in helices. As the two effects cannot be clearly separated, only a qualitative picture on metal ion binding sites within the molecule can be drawn.

Figure 5c shows the change in chemical shift experienced by sugar (H1’ and H2’) and aromatic protons (H6/H8/H2) on addition of 9 mM Mg(II). The nucleotides that are most strongly influenced by the presence of Mg(II) are C11 and A12 with resonances shifting by up to 0.4 ppm. C11 and A12 are in a critical position within the three-way junction, similar to A37 and G38, whose resonances also shift noticeably. In the GAAA tetraloop, we observe similar changes for all the sugar protons, without a clear pattern. Helix d" and the UUCG loop are remarkably unaffected by Mg(II), whereas the peaks of nucleotides located in the κ region do shift considerably, get broader and eventually disappear when increasing the amount of metal ions. This could be due to metal ion binding, dimerization (see earlier in the text) or both.
We titrated an RNA sample with increasing amounts of cobalt(III)hexammine (from 0 to 3.6 mM, in six steps, see Materials and Methods) and evaluated both chemical shift variation and NOE cross-peaks. We observe the appearance of the same peaks around the three-way junction that also appear with Mg(II).

At 1 mM cobalt(III)hexammine concentration, it is possible to identify three binding sites through the NOE cross-peaks from the NH3 ligands to D1\textsubscript{kz} protons: (i) the GAAA tetraloop; (ii) the GU wobble pair in the \textit{ζ} region; and (iii) the CG-AU base pair step in helix d\textsubscript{00} (Figure 6a and b). Cross-peaks between cobalt(III)hexammine and U5H3 and U47H3 (Figure 6a) can be explained by the electrostatically favorable interaction of the complex with the major groove of the U5-G45 G-U wobble pair (87). Both the sign and magnitude of chemical shift variations induced by cobalt(III)hexammine are in general very similar to what we observe with Mg(II) (Supplementary Figure S8). However, the same chemical shift change is already reached at much lower concentration, suggesting that cobalt(III)hexammine binds at similar sites but with higher affinity (see, for example, G40H8, Figure 5a and b).

To evaluate possible inner sphere binding of Mg(II), $^2$J\textsubscript{1H,15N}-HSQC spectra were recorded. Thus, direct coordination of the metal ion to the nitrogen of guanines and adenines can be detected (88), the N7 of guanines being generally the most favorable binding site (89). The chemical shift changes experienced by both N7/H8 in guanosines and adenosines as well as N1/N3/H2 in adenines can be followed contemporaneously (88). Usually, an upfield shift of the N7 resonance on metal ion binding indicates a direct binding (88,90,91), and the corresponding H8 should be shifted downfield due to the electron-withdrawing effect of the coordinated metal ion. On the other hand, if a structural rearrangement takes place resulting in increased stacking, the H8 resonances would move upfield.

Figure 7 shows the overlay of $^1$H,$^15$N]-HSQC spectra recorded with increasing amounts of Mg(II). Residues in the helices show a moderate upfield shift of N7 (\textlessthan 1 ppm) and almost no shift of N1 and N3, which is consistent with a general non-specific interaction with the N7 and bad accessibility of N1 and N3. In the \textit{ζ} region, N1 of A6, A43, A44 and N7 of G45 broaden out and eventually disappear on addition of Mg(II). Such line broadening can be due to the ligand exchange rate in the first coordination shell of Mg(II), which is on the NMR timescale (88) and thus suggests inner-sphere coordination in the tetraloop receptor region. On the other hand, also $^1$H and $^{13}$C resonances broaden out, and the cross-peaks observed in $^1$H,$^1$H]-NOESY involving A43, A44 and G45 get broader more quickly in the presence of cobalt(III)hexammine than with Mg(II), and in the first case, this effect cannot be due to inner-sphere coordination. In addition, we have described earlier in the text that there are strong indications for a homodimer formation, which
includes transient interactions of the GAAA tetraloop with the $\zeta$ region and is also expected to lead to a broadening of the signals. We observe a similar behavior for all N7 of the GAAA tetraloop nucleotides: the resonances move upfield more strongly than residues in the helices (>1 ppm) and without broadening. Similarly, also A22N7 and G21N7 in the $\kappa$ tetraloop structure move upfield without broadening. A12N1 and A22N1 move by almost 2 ppm, whereas A17N1 and A18N1 turn around at ~2 mM Mg(II) hinting at the interplay of structural rearrangements and metal ion interactions in this region (Figure 7).

Mg(II) as a hard metal ion prefers oxygen over nitrogen binding. If there is nonetheless a strong direct interaction, it is expected to move the $^{15}$N resonance upfield by ~6 ppm (90,92). None of our observable N1, N3 or N7 moves that much; only the G1N7 moves by 3.4 ppm, and this behavior is explained by the charge of the tri- or di-phosphate present.

DISCUSSION

General structural features: the 11-nt tetraloop receptor and two classical tetraloops

The D1$\kappa\zeta$ construct contains several common structural motifs that have already been described in the literature. Both the GAAA and UUCG tetraloops as well as the tetraloop receptor region ($\zeta$) agree reasonably well with published solution structures [2JYJ (68), 2KOC (65), 1TLR (66)]. The tetraloop receptor in its unbound form is characterized by A43 and A44, a U7-U43 mismatch and a G45-U5 wobble pair (66). Such an arrangement of the tetraloop receptor has been proposed also in the context of the complete D1 before docking of D5 (28) and is now confirmed by our observed NOE pattern in the absence of Mg(II).

In solution, the $\kappa$ region (A12, G21-A23) of D1$\kappa\zeta$ folds into a GAAA tetraloop-like structure similar to the one in the published crystal structure of a group IIC intron from O. iheyensis (38–40) (Figures 8 and 9). The orientation of this motif differs from the one in the fully assembled group II intron (Figures 8 and 9). In the full-length group IIC intron structure, the $\kappa$ region engages in H-bonds with the two G-C pairs of $\kappa'$ in D5 through the Hoogsteen face of A22 (N7), the WC face of A23 (N1) and the sugar face of A12 (N3) (numbering of D1$\kappa\zeta$).

In contrast, in D1$\kappa\zeta$ A22 and A23 form A-minor interactions to helix d (see Results, Figure 10a and b). The interactions can be classified as type 0 (A23N3 and O2' to C11O2', Figure 10a) and type I (A22N1 to U10O2', A22O2' to A39N3 and O2', Figure 10b) according to Nissen et al. (81) and decrease in type order from 5 to 3 (81). An investigation of three-way junctions found in published crystal structures by Lescoute and Westhof (93) suggests that such an interaction as observed here, where adenines inside a junction engage in A-minor interactions to the adjacent helix, is common. Lescoute and Westhof classified the three-way junctions into three families. The D1$\kappa\zeta$ junction belongs to family C, which is the most abundant and appears in various structured RNAs. In this family, J31 is bigger than J23 and J12, and helix P3 (the $\kappa$ extension in D1$\kappa\zeta$) leans towards helix P1 (helix d'') (Figure 10c and d). In addition, the J31 linker can contact the shallow minor groove of helix P2 (helix d') (Figure 10d). Moreover, Lescoute and Westhof predict that the 0-nt linker will most likely be connecting helices that are coaxially stacking, which is also true in our three-way junction. Geary et al. (45) have even used the concept of A-minor interactions from a terminal GNRA-like loop to an adjacent helix to stabilize specific conformations of RNA four-way junctions. The three-way junction in
D1kz is completely homologous to the constructs they used, if the junction is moved by a 1–2 base pairs directly into the GNRA loop (Figure 10d and e). In D1kz, the crossing of the strands takes place directly between A23 and A12, which are the third and the fourth nucleotide of the GAAA-motif (Figure 10d). Although it has been recognized long ago that the GNRA motif is a very favorable fold that can form even when the sequence is interrupted, database searches using FR3D (94) and RNA FRABASE (95) do not find this motif inside a three-way junction in any other structure except the O. iheyensis crystal structures (38–40).

The interaction of k with helix d′ also helps to understand why the H1′ resonance of G40 is not observed in D1kz. At first sight, this is unexpected considering that the residue is within a stable base pair in a helical region, and that the H1′ resonance is present and sharp in constructs lacking the k region (see constructs D1-27 and D1-45 in Supplementary Figures S1 and S2). However, the final structures of D1kz position G40H1′ just below the contact interface of k and helix d′, where it transiently gets into contact with A22, inducing dynamics on a timescale that completely broadens the resonance.

Figure 7. Overlay of 2J-[1H,15N]-HSQC spectra in the region of purine N1, N3 and N7 resonances in the presence of increasing amounts of Mg(II). Resonances disappearing after the first addition of 0.5 mM Mg(II) are labeled in red. (300 K, 0.45 mM D1kz, pH 6.74, 60 mM KCl, 10 μM EDTA, 700 MHz).

Figure 8. (a) Overlay of D1kz (orange) with the corresponding region in the O. iheyensis group IIC intron crystal structure [dark grey; PDB ID: 3G78, (12,40)]. The heavy atoms of the k element were overlaid to visualize the different orientations of k towards the rest of the molecule in the two structures. In green, the coordination loop adjacent to helix d′ of the O. iheyensis group IIC intron interacting with the GAACA loop is shown. (b) Schematic representation of the structures depicted in (a). Tertiary interactions are indicated by asterisks. In D1kz in solution (orange) k interacts with helix d′, which probably favors the docking of D5 (shown in red). The GAAA tetraloop adjacent to k does not interact with helix d′. In the O. iheyensis crystal structure, the loop adjacent to k interacts with the coordination loop, possibly disfavoring the inter-domain interaction observed in D1kz and favoring docking of D5.

Figure 9. (a) The three-way junction as found in the O. iheyensis group IIC intron. Interaction of k (green) with k′ in D5 (red) is indicated. Phosphate groups equivalent to those of A12 and U24 in D1kz are colored. (b) A-minor interaction of k residues A22 and A23 with helix d′ in D1kz. Examples of observed NOE distances are indicated as blue dotted lines.
Our spectra indicate that this A-minor interaction is strong and independent of the presence of Mg(II). In the crystal structure, comprising the whole intron in its native state, however, this A-motif is not present, but κ is positioned differently. There are two plausible reasons for that: (i) the docking of D5 itself switches the κ interaction or (ii) the contact between the capping loop of the κ extension (GAAA in our structure, GAACA in the O. iheyensis intron) and the lower part of the neighboring so-called coordination loop as observed in the crystal structure (Figure 8) is strong enough to disrupt the A-minor interactions. To prove the first hypothesis, we looked at D5 and D1κk in solution, but could not detect inter-molecular NOEs or any other signs of interaction, e.g. in native gel studies (data not shown). The reason could be that the two constructs need to be embedded in the complex framework of the group II intron to adopt the global conformations that allow a favorable interaction. On the other hand, we can also not exclude that D5 in the full intron will be the factor that leads to the release of the A-minor interactions, as also the interactions to the coordination loop (second hypothesis) are not absolutely required for a functional group II intron. When the GAAA tetraloop is replaced by UUCG, full catalytic activity is retained (96). In the context of the full intron, a combination of the two factors might control the final orientation of the κ motif.

Mg(II) promotes coaxial stacking of helices d’ and d”

On addition of Mg(II) and cobalt(III)hexammine, the appearance and sharpening of resonances neighboring the junction indicate that the three-way junction around κ is structurally stabilized in the presence of those metal ions. Effects of the stabilization are noticeable not only directly at the junction but extend also to the second base pairs in all three helices, i.e. d’, d” and the κ extension. The appearance of an NOE cross-peak between U24H3 and G38H1 is strong evidence for a stable coaxial stacking of helix d’ and d”, only in the presence of Mg(II) or cobalt(III)hexammine. This coaxial stacking is observed also in the group IIC intron crystal structure (39) and thus is independent of the orientation of the κ extension, which bulges out between residue 11 and 24 (D1κk numbering) in the crystal and in solution.

In our D1κk structure, stacking between C11 and U24 as well as between A12 and A23 is observed. At the same time the two phosphate-sugar backbones approach each other very closely. Specifically, A12OP2 and U24OP1 are only ~4 A apart in most of the calculated structures (Figure 11), a configuration that is likely to require a charge-compensating metal ion. No Mg(II) ion was found at this location in the crystal structure of O. iheyensis intron, but A12 displayed a strong phosphotriphosphate effect in studies with the D135 ribozyme (28). The distance of the A12OP2 and U24OP1 phosphate oxygens is well within the range of a bridging Mg(II) ion. The fact that cobalt(III)hexammine has a stabilizing effect equivalent to Mg(II) indicates that direct inner-sphere coordination is not a requirement for a cation to be able to stabilize the region.

Obviously, di- or multivalent ions are a prerequisite for this close approach of the two backbones, and thus a
stable coaxial stacking of helix d’ and helix d’’. Coaxial stacking is very common for helices in multibranch junctions in crystal structures (43). In solution, it has been shown for the four-way junction of the hairpin ribozyme that different conformations are sampled at low ionic strength, whereas one dominant coaxially stacked conformation is found under high salt conditions (97,98). We observe here a similar case: the presence of Mg(II) or cobalt(III)hexammine strongly stabilizes coaxial stacking of helix d’ and d’’.

High affinity metal ion binding within D1κζ

D1κζ was titrated with increasing amounts of Mg(II) or cobalt(III)hexammine. As described above, both ions behave similarly, although less cobalt(III)hexammine is needed for structural stabilization, well in line with the higher positive charge. The nucleotides most affected by the metal ion addition are near the tetraloop receptor and the three-way junction. For example, the strong chemical shift changes observed for G8H1 and both the structural motifs are combined in our small D1κζ construct.

In conclusion, the structure of D1κζ in the presence of Mg(II) shows that essential structural features of the final catalytically prone group II intron core are already formed in the absence of any tertiary contacts, as present in the early steps of folding. This suggests that the metal ion-induced rigidification of the three-way junction may be one of the first steps in group II intron folding.

ACCESSION NUMBERS

Atomic coordinates for the refined structures have been deposited with the Protein Data Bank under accession code 2LU0. The NMR chemical shifts have been deposited at the Biological Magnetic Resonance Data Bank (BMRB) under the accession code 18503.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–8 and Supplementary Methods including references.

ACKNOWLEDGEMENTS

Financial support by a Marie Curie Intra European Fellowship (to D.D.) and in the form of access to the Bio-NMR Research Infrastructure within the 7th European Community Framework Programme, by the Roche Research Foundation, the Novartis Stiftung (both to C.F. and R.K.O.S.), the Swiss National Science Foundation (to R.K.O.S.) and the University of Zurich are gratefully acknowledged. R.K.O.S. is holder of an ERC starting grant 2010. The authors thank the Biochemistry Department at the University of Zurich for the use of the ultra-centrifuge.

FUNDING

Marie Curie Intra-European Fellowship (FP7-MC-IEF) [236797 to D.D.]; Bio-NMR Research Infrastructure (FP7/2007-2013) [261863]; Swiss National Science Foundation [200021-124834 to R.K.O.S.]; Roche
Research Foundation (to C.F. and R.K.O.S.); Novartis Stiftung (to C.F. and R.K.O.S.); University of Zurich. Funding for open access charge: University of Zurich.

Conflict of interest statement. None declared.

REFERENCES


