RNA (1997), 3:838–849. Cambridge University Press. Printed in the USA.

An RNA internal loop acts as a hinge to facilitate ribozyme folding and catalysis

ALEXANDER A. SZEWCZAK and THOMAS R. CECH

Howard Hughes Medical Institute, Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309-0215, USA

ABSTRACT

RNA molecules commonly consist of helical regions separated by internal loops, and in many cases these internal loops have been found to assume stable structures. We have examined the function and dynamics of an internal loop, J5/5a, that joins the two halves of the P4–P6 domain of the *Tetrahymena* self-splicing group I intron. P4–P6 RNAs with mutations in the J5/5a region showed nondenaturing gel electrophoretic mobilities and levels of Fe(II)-EDTA cleavage protection intermediate between those of wild-type RNA and a mutant incapable of folding into the native P4–P6 tertiary structure. Mutants with the least structured J5/5a loops behaved the most like wild-type P4–P6, and required smaller amounts of Mg^{2+} to rescue folding. The activity of reconstituted introns containing mutant P4–P6 RNAs correlated similarly with the nature of the J5/5a mutation. Our results suggest that, in solution, the P4–P6 RNA is in a two-state equilibrium between folded and unfolded states. We conclude that this internal loop mainly acts as a flexible hinge, allowing the coaxially stacked helical regions on either side of it to interact via specific tertiary contacts. To a lesser extent, the specific bases within the loop contribute to folding. Furthermore, it is crucial that the junction remain unstructured in the unfolded state. These conclusions cannot be derived from a simple examination of the P4–P6 crystal structure (Cate JH et al., 1996, *Science 273*:1678–1685), showing once again that structure determination must be supplemented with mutational and thermodynamic analysis to provide a complete picture of a folded macromolecule.

Keywords: hydroxyl radical footprinting; native polyacrylamide gel electrophoresis; P4–P6 domain; RNA folding; *Tetrahymena* group I intron

INTRODUCTION

In addition to its role as the genetic messenger, RNA performs a wide variety of functions in nature, many of which depend directly upon chemical reactions that the RNA itself catalyzes (for review, see Atkins & Gesteland, 1993). As is the case for proteins, the biological functions of RNAs depend very much on their macromolecular structures. In almost every instance, it is the three-dimensional conformation of an RNA, specified by its sequence, which enables it to perform a particular biological task. Clearly, in order to fully understand RNA biochemistry, it is imperative to understand RNA folding.

Natural structured RNAs consist of rather short duplex regions (typically one half to one full turn of a helix) separated by internal loops. In secondary structure diagrams, these loops give the appearance of allowing bending or flexibility. In contrast to that expectation, most of the RNA structures solved to date have loop regions that are highly structured. For example, loop E from 5S rRNA is rather structured (Wimberly et al., 1993). Even RNAs with multiple loop regions, such as the sarcin loop (Szewczak et al., 1993), tRNAs (Kim et al., 1974; Robertus et al., 1974; Woo et al., 1980; Westhof et al., 1985; Basavappa & Sigler, 1991), and the P4–P6 domain structure (Cate et al., 1996a) [P denotes a paired region of the group I intron, hence P4–P6 refers to the domain that includes helices P4, P5, P5a, P5b, P5c, P6, P6a, and P6b] have been found to be highly structured. Admittedly, structured RNAs are more amenable to X-ray crystallographic and NMR spectroscopic analysis, and it is still difficult, if not impossible, to predict the conformations of most loop regions in large RNAs. The question remains: What are the structural and dynamic roles of loops? Do most in fact stabilize RNA folding or do they mainly serve as passive linkers? Neither NMR spectroscopic studies (which fail with completely unstructured regions of molecules) nor X-ray crystal-

Reprint requests to: Thomas R. Cech, Howard Hughes Medical Institute, Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309-0215, USA; e-mail: thomas.cech@colorado.edu.

lography (which gives a largely static picture of a macromolecule) is sufficient for study of a dynamic process like RNA folding (Zarrinkar & Williamson, 1994, 1996; Downs & Cech, 1996) or even, at times, RNA catalysis itself (e.g., McKay, 1996).

To further understand the structural and dynamic properties of RNA loop regions, the roles they play in stabilizing the three-dimensional structures of RNAs, and how they facilitate catalysis, we have chosen the P4–P6 domain as a model system. The P4–P6 domain from the *Tetrahymena* self-splicing group I intron is a 160-nt region that forms a tertiary structure independent of the rest of the intron, the folding of which can be monitored using hydroxyl radical footprinting and native PAGE (Murphy & Cech, 1993) (see Fig. 1). The domain stabilizes the remainder of the intron (van der Horst et al., 1991) and contains several highly conserved secondary structure elements and residues that have been modeled as being close to the site of catalysis (Michel & Westhof, 1990). Recently, the P4–P6 domain was shown to reconstitute an active group I intron when mixed with two other portions of the molecule (Doudna & Cech, 1995).

Biochemical experiments and X-ray crystallography have identified two levels of organization of the tertiary structure of the P4–P6 domain, both of which depend on $MgCl_2$ for formation (Murphy & Cech, 1993, 1994; Murphy et al., 1994; Cate et al., 1996a, 1996b). In the absence of $MgCl_2$, only the secondary structure of the molecule forms. At approximately 0.5 mM Mg^{2+}, the P5abc extension of P4–P6 folds into a subdomain structure that protects a small set of residues from Fe(II)-EDTA free radical cleavage (see Fig. 1A, open triangles). In addition, at 2 mM Mg^{2+}, P4–P6 further folds into the full domain tertiary structure, a more global fold that produces Fe(II)-EDTA protections throughout the molecule (Fig. 1A, closed triangles). As illustrated in Figure 1B, this global fold involves coaxial stacking of two sets of helices that in turn pack together, in a configuration not unlike a clothespin. Although interactions that are sufficient to stabilize the P5abc subdomain structure are contained completely within the P5abc extension, residues throughout the molecule are necessary for formation of the whole domain structure. One important domain-stabilizing interaction is the docking of the GAAA tetraloop in L5 of the P5abc extension into the tetraloop receptor located in the J6a/6b internal loop (Murphy & Cech, 1994; Costa & Michel, 1995; Cate et al., 1996a). Another tertiary interaction is the packing of the P5abc subdomain structure onto helices P4 and P6 (Cate et al., 1996a).

We focus here on a set of P4–P6 molecules containing mutations in the junction between P5 and P5a, J5/5a. Biochemical and crystallographic data show that a bend at this junction allows the two coaxially stacked halves of the molecule to pack together (Fig. 1B) (Murphy & Cech, 1993, 1994; Murphy et al., 1994; Cate et al., 1996a). We now ask if the J5/5a region actually directs the bending of the molecule by forming a specific structure or if it instead serves passively as a flexible hinge. Comparative sequence analysis gives no clear expectation, because the 5′ strand of J5/5a has a preferred sequence, whereas the 3′ strand is much more variable (Murphy & Cech, 1994). Our results suggest that the P4–P6 domain exists in an equilibrium between folded and unfolded states. Furthermore, although the bend contributes somewhat to the stability of the final folded structure, it is most crucial that the junction remain unstructured in the unfolded state.

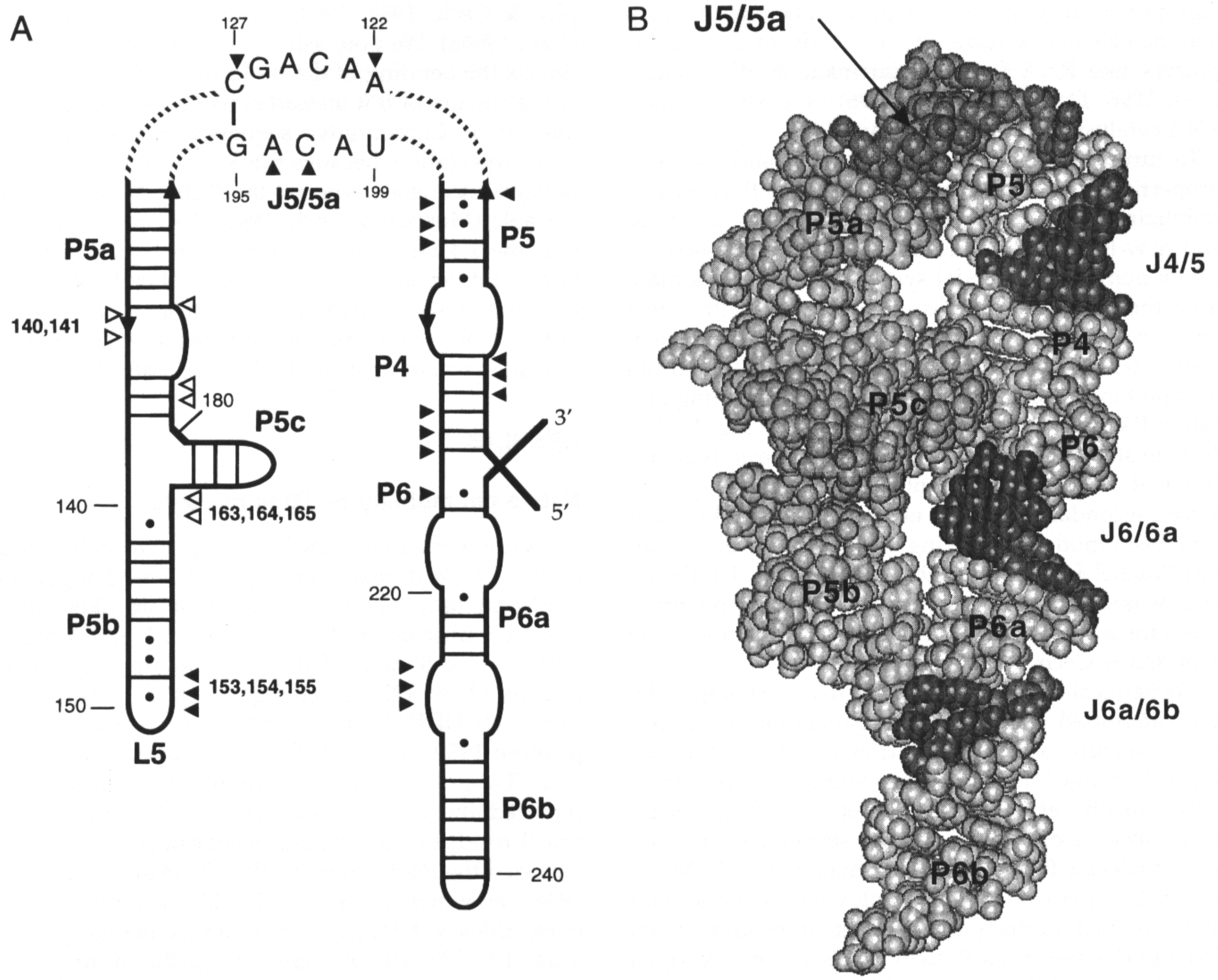
RESULTS

Native gel mobility of J5/5a mutants

To examine the role of the J5/5a region in stabilization of the P4–P6 domain tertiary fold, we changed the wild-type sequence to residues that are likely to be relatively unstructured in solution (see Fig. 1C). In mutant U1, the 5′ side of the J5/5a internal loop from residue 122 to 127 was changed from 5′-AACAGC-3′ to 5′-UUUUUU-3′. In mutant U2, the sequence from position 195 to 199, 5′-GACAU-3′, was replaced with 5′-UUUUU-3′. The double mutant U1+U2 completely replaced the J5/5a sequences on both strands with uracil residues. As a negative control for folding, we included the P4–P6 mutant BP 5/5a (Murphy & Cech, 1994), in which residues 122–127 are replaced with nucleotides with Watson–Crick complementarity to residues 195–199. This mutant is incapable of forming a normal domain tertiary structure (Murphy & Cech, 1994).

To assess the ability of this set of mutant RNAs to fold into a stable domain tertiary structure, we first examined their mobility on native polyacrylamide gels. Previous work had shown that properly folded P4–P6 migrates more quickly than BP 5/5a or other folding mutants in native gels containing 13 mM $MgCl_2$ (Murphy et al., 1994), presumably because a molecule with only secondary structure is much more extended and therefore retarded relative to the compact P4–P6 molecule.

In the absence of $MgCl_2$, all five of the P4–P6 RNAs migrated at the same rate, consistent with a lack of tertiary structure and with the fact that these RNAs are the same length (Fig. 2A). However, in the presence of 10 mM $MgCl_2$, their mobilities were dramatically different (Fig. 2B). As expected, P4–P6 migrated the fastest, consistent with it being fully folded at 10 mM $MgCl_2$ (Murphy & Cech, 1993), and BP 5/5a migrated the slowest, because it cannot fold to form the domain structure. The three U-containing mutants moved at intermediate rates, suggesting that their compactness follows the trend P4–P6 > U1+U2 > U1 >



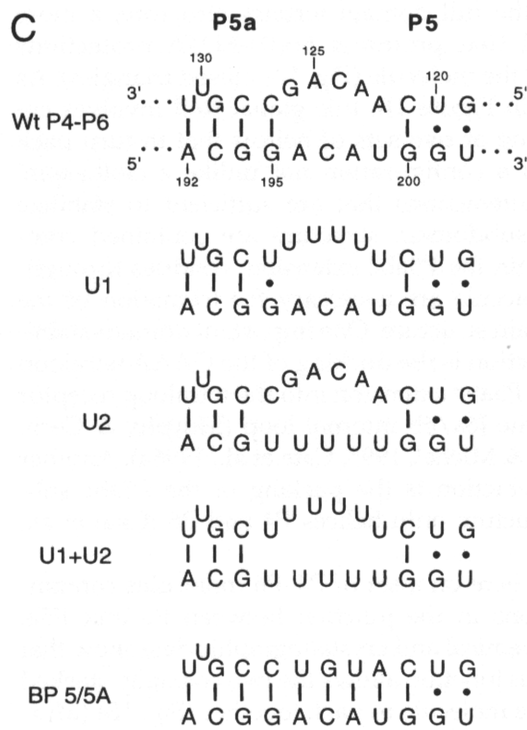


FIGURE 1. Secondary and tertiary structure of the P4–P6 domain. **A:** Schematic diagram of the secondary structure of the P4–P6 domain. Paired regions are labeled P4–P6 and the wild-type J5/5a internal loop sequence is indicated. Residues protected in a Mg^{2+}-dependent fashion in the P5abc subdomain and full domain structures of P4–P6 (Murphy & Cech, 1993) are indicated by open and closed triangles, respectively. **B:** Space-filling representation of the P4–P6 domain as determined by X-ray crystallography (Cate et al., 1996a). Internal loop regions J4/5, J6/6a, and J6a/6b are colored dark gray. The J5/5a internal loop, also in dark gray, is at the top of the molecule marked with an arrow. **C:** Sequences of the P4–P6 J5/5a internal loop mutants studied in this work. Wild-type P4–P6 has the naturally occurring *Tetrahymena* sequence. U1 and U2 mutants were created by replacing the sequence of one strand of the J5/5a junction region with uracil residues. U1+U2 has a J5/5a sequence created by combining the U1 and U2 mutations and BP 5/5a has a fully base paired internal loop.

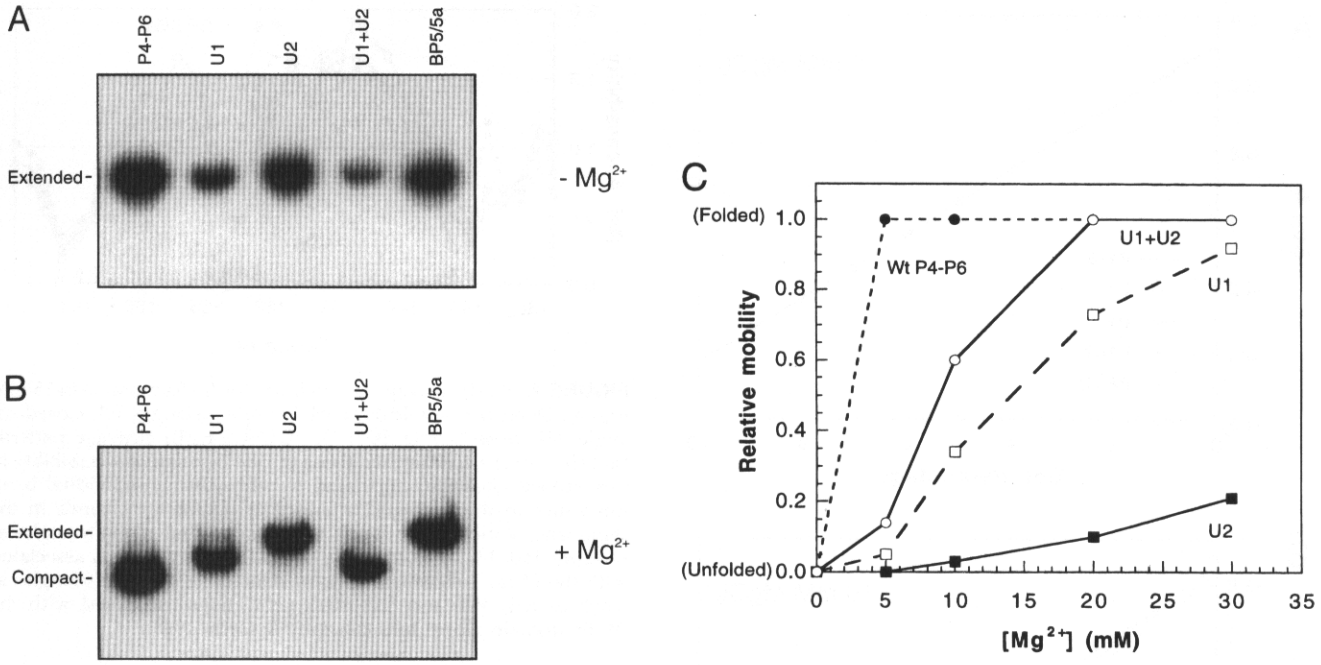


FIGURE 2. Native polyacrylamide gel analysis of the extent of folding of the P4–P6 mutants. **A:** The mobilities of P4–P6 J5/5a variants are identical in the absence of $MgCl_2$. **B:** Native gel mobility of P4–P6 J5/5a internal loop mutants in the presence of 10 mM $MgCl_2$ differentiates between the various RNAs based on their relative compactness. **C:** Relative mobility of P4–P6 mutants plotted as a function of Mg^{2+} concentration. Mobilities are normalized relative to wild-type P4–P6 as described in Materials and Methods.

U2 > BP 5/5a. Thus, these mutations affect the folding of P4–P6 to different extents, at least as judged by the relative electrophoretic mobility of these RNAs on native gels.

To determine if increasing concentrations of Mg^{2+} would help recover folding of the mutants with destabilized tertiary folds, we examined their electrophoretic mobility as a function of $MgCl_2$ concentration. As shown in Figure 2C, increasing the concentration of Mg^{2+} increased the mobility of each mutant relative to wild-type P4–P6. The exception was BP 5/5a, which always showed the same retardation relative to P4–P6 (data not shown). At 5 mM $MgCl_2$, where wild-type P4–P6 is fully folded (Murphy & Cech, 1993), the relative mobilities of the mutants were very low. However, the higher the Mg^{2+} concentration was raised, the more closely the mobility of each mutant approximated that of the wild-type molecule. At 30 mM $MgCl_2$, the relative mobilities of both U1 and U1+U2 approached that of P4–P6, and the relative mobility of U2 also began to increase. Poor resolution at even higher concentrations of $MgCl_2$ prevented us from determining unequivocally that the mobility of U2 also eventually equals that of wild-type P4–P6.

Ferguson plots (Ferguson, 1964) confirmed that the electrophoretic mobility was really a measure of the relative compactness of each RNA (Fig. 3). For a rapidly tumbling molecule, whose size is small relative to the gel matrix, a plot of the logarithm of the mobility of the molecule (log *M*) versus gel concentration gives a line whose slope is a function of the effective radius of the molecule (Rodbard & Chrambach, 1971). P4–P6 RNA mobilities were measured as described in Materials and Methods for all five RNAs as a function of gel concentration, in the presence of 0 mM and 10 mM $MgCl_2$. Under both sets of conditions, all RNAs showed linear plots of log *M* versus gel percentage. In the absence of $MgCl_2$, all RNAs migrated at exactly the same rate at a given gel concentration (Fig. 3A). When 10 mM $MgCl_2$ was included in the gel running buffer and the gel matrix, the relative mobilities of the mutants diverged at every gel concentration, always following the order P4–P6 > U1+U2 > U1 > U2 > BP 5/5a (Fig. 3B). The fact that the mutant P4–P6 RNAs had the same relative mobilities, independent of polyacrylamide concentration, supports the conclusion that the gels give a true indication of relative compactness.

Fe(II)-EDTA footprinting

To examine the local (P5abc subdomain) and global (domain) folding of these RNAs in finer detail, we probed their tertiary structure using free radical cleavage. Solvent-based Fe(II)-EDTA generates free radicals, presumably hydroxyl radicals, that cleave the backbone of nucleic acids (Hertzberg & Dervan, 1984; Tullius & Dombrowski, 1985) in a manner insensitive to secondary structure (Celander & Cech, 1990) and dependent on the surface accessibility of each ribose C4′ atom (Latham & Cech, 1989; Cate et al., 1996a).

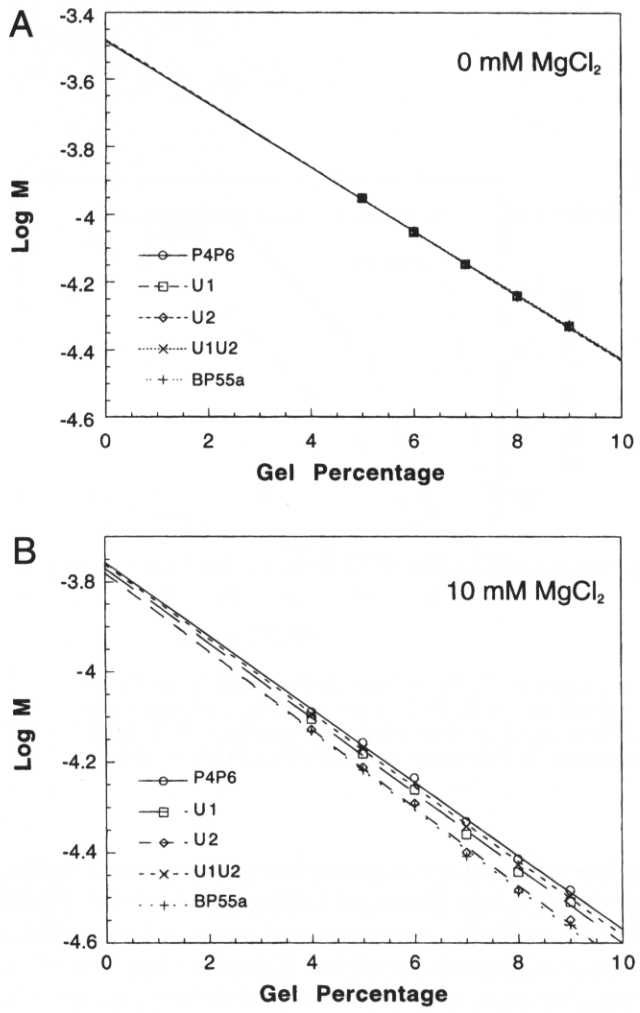


FIGURE 3. Ferguson plot analysis of the electrophoretic mobility of the mutant forms of P4–P6. Log *M*, where mobility, *M*, is defined as the distance migrated (cm) divided by run time in seconds and electric field strength (V/cm), is plotted as a function of gel percentage. **A:** Ferguson plots for mutant and wild-type P4–P6 RNAs in the absence of $MgCl_2$, where all RNAs behave identically. **B:** Ferguson plots for P4–P6 RNAs in the presence of 10 mM $MgCl_2$ demonstrate that the different RNA mobilities are a true indication of relative compactness.

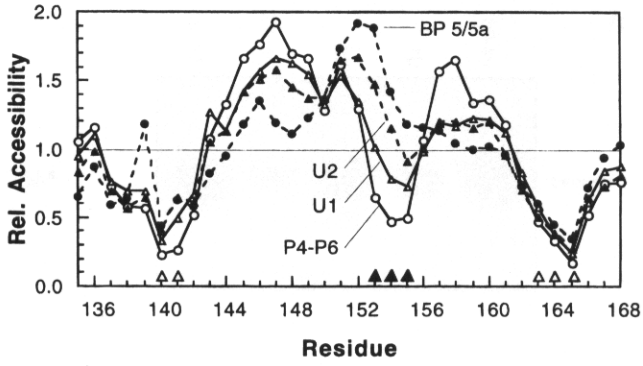


FIGURE 4. Fe(II) cleavage accessibility for J5/5a mutants in 15 mM $MgCl_2$. Open circle, wild-type P4–P6; open triangle, U1; closed triangle, U2; closed circle, BP 5/5a. U1+U2 Fe(II) cleavage patterns (not shown) closely matched those of P4–P6. Relative accessibility to free radical cleavage, calculated by normalizing individual band intensities to the averaged intensity of neighboring bands in the same lane, is plotted for P4–P6 residues 133–166. Open triangles at residues 140, 141, 163, 164, and 165 indicate protections associated with the P5abc subdomain tertiary structure, and solid triangles at residues 153, 154, and 155 mark protections associated with the P4–P6 domain global fold (Murphy & Cech, 1993).

Variations in intensity of the resulting cleavage are therefore an indication of tertiary structure.

Mutant and wild-type P4–P6 RNAs in 15 mM $MgCl_2$ were probed with Fe(II)-EDTA under buffer conditions similar to those used for native gel electrophoresis. A sample plot of relative accessibility for residues 135–168 is shown in Figure 4. Residues marked with open triangles (140, 141, as well as 163–165) show reduced accessibility to free radical cleavage associated with the presence of the subdomain tertiary fold. All mutant and wild-type P4–P6 RNAs show nearly identical protections here, as expected, because the subdomain requires less than 1 mM $MgCl_2$ for folding, and the J5/5a mutations are not expected to disrupt this structure. Solid triangles mark residues normally protected by the whole domain tertiary structure (residues 153, 154, 155). Here mutant P4–P6 RNAs show greater accessibility than is characteristic of the native structure. Mutant U1 is less protected from cleavage than wild-type P4–P6, and mutant BP 5/5a isn't protected at all, whereas mutant U2 shows accessibilities intermediate between U1 and BP 5/5a. Thus, the inaccessibility (protection from cleavage) of residues normally protected in the domain structure follows the order P4–P6 > U1 > U2 > BP 5/5a. The same relative order is seen for tertiary protections over the entire molecule (data not shown). Interestingly, there are also regions over which accessibility is increased when the P4–P6 domain folds (for example residues 144–149, 157–161 in Fig. 4). For these residues, the trend in *accessibility* follows P4–P6 > U1 > U2 > BP 5/5a. In other words, the J5/5a mutations cause less of an enhancement in the reactivity associated with formation of the domain structure.

Additional Fe(II) protection data for all five RNAs were collected over the range of 0–150 mM $MgCl_2$ (Table 1). U1+U2 appears to be half folded at about 5 mM Mg^{2+}, whereas U1 and U2 have $[Mg^{2+}]_{1/2}$ values of 10 and 30 mM Mg^{2+}, respectively. For BP 5/5a, no domain tertiary protections were seen at Mg^{2+} concentrations as high as 150 mM, consistent with its inability to form the domain structure.

Reconstituted intron kinetics

To examine how the stability of the P4–P6 domain affects the catalytic activity of the *Tetrahymena* group I intron, we measured rates of 5′ splice site cleavage in the three-part intron system described originally by Doudna and Cech (1995). The intron has been divided into three component RNAs as follows: "P1–P3" in-

TABLE 1. $MgCl_2$ concentration dependence of Fe(II) protection of P4–P6 J5/5a variants.

RNA	$[Mg^{2+}]_{1/2}$ (mM)[a]
Wt P4–P6	1
U1+U2	5
U1	10
U2	30
BP 5/5a	>100

[a]The $MgCl_2$ concentration at which protection from Fe(II)-EDTA cleavage is half-maximal, estimated from protections at the residues marked with closed triangles in Figure 4.

cludes helices P1, P2, P2.1, and the 5′ half of P3; "P3–P9" includes the 3′ half of P3, P7, P8, P9, P9.1, and P9.2; the P4–P6 domain contributes the remainder of the intron (Fig. 5A). In the presence of 35 mM Tris-HCl, pH 7.5, 5 mM GTP, 80 mM $MgCl_2$, and 5 mM spermidine, the reassembled intron catalyzes attack by GTP on the phosphate at the exon–intron junction within helix P1. Hence, the P1–P3 RNA behaves as a substrate, and reaction of 5′ ^{32}P-labeled P1–P3 can be followed on gels by monitoring the conversion of the 111-nt P1–P3 band into the 8-nt product band.

Cleavage reactions were conducted using 100 nM of each P4–P6 domain RNA. Representative experiments are shown in Figure 5B, and time courses of the reaction are shown for mutant and wild-type P4–P6 RNAs in Figure 5C. The fraction of unreacted substrate (S) was plotted as a function of time, and the resulting graphs were fit to the first-order rate equation. RNase T_1 and alkaline cleavage ladders confirmed that cleavage occurred at the expected site for all mutants (data not shown). In the absence of GTP, the rate of cleavage of each P4–P6 variant was reduced 25-fold, and, in the absence of any one of the three RNA components, absolutely no reaction could be detected. Thus, the observed reactions and differences in activity demonstrated for the J5/5a mutants are both GTP- and RNA-dependent.

The rate we observed for wild-type P4–P6 was 0.35 min^{-1}, about eight times faster than the rate observed previously with a similar system containing 20 nt of 5′ exon sequences (Doudna & Cech, 1995). All four mutant P4–P6 RNAs exhibited significantly reduced activity, their catalytic rates following the trend P4–P6 > U1+U2 ≈ U1 > U2 > BP 5/5a (Table 2). The rate constant for the U2 RNA was more that 25-fold slower than wild-type, whereas U1 activity was reduced only eightfold. U1+U2 RNA reacted at about the same rate as U1 RNA. Interestingly, even the fully base paired mutant, BP 5/5a, showed some measurable activity, although it reacted about 50-fold slower than wild-type P4–P6. The order of reactivity of the P4–P6 mutants was similar to the order of stability of their tertiary folds (compare Tables 1 and 2).

Experiments in which reactivity of the reconstituted *Tetrahymena* intron was followed as a function of the concentrations of the RNA components gave an apparent K_M of 4 nM for P3–P9, and about 30 nM for wild-type P4–P6 (Doudna & Cech, 1995). Preliminary results indicate that the 100 nM concentration used here for P4–P6 RNAs is saturating for the U-containing mutants as well (data not shown).

DISCUSSION

Short helices separated by internal loops are common features of most structured RNA molecules. An internal loop in the self-cleaving hairpin RNA appears to be a simple tether between functional domains (Feldstein & Bruening, 1993). However, in several cases that have been investigated by NMR spectroscopy, internal loops form specific non-Watson–Crick base paired structures either free in solution (Szewczak et al., 1993; Wimberly et al., 1993) or upon binding of a ligand (e.g., Puglisi et al., 1992; Aboul-ela et al., 1995; Dieckmann et al., 1996; Jiang et al., 1996; Yang et al., 1996). Here we investigated the function of an internal loop, J5/5a, that joins the two halves of the P4–P6 domain of the *Tetrahymena* ribozyme. We conclude that this internal loop acts mainly as a flexible hinge, allowing the quasi-helical regions on either side of it to interact via specific tertiary contacts. To a smaller extent, the specific bases within the loop contribute to folding. These conclusions are compared to those that might be derived from a simple examination of the P4–P6 crystal structure (Cate et al., 1996a).

Our investigation of the J5/5a internal loop involved replacement of one or both sides of the loop with uracil residues, which have the least propensity for stacking interactions of the four bases. This replacement should also disrupt interstrand base pairing. We found that these mutations in the J5/5a junction destabilize the global tertiary fold of the P4–P6 domain to varying extents. As judged by native gel electrophoresis at concentrations of Mg^{2+} below 20 mM, wild-type P4–P6 has the most compact time-averaged structure, followed by U1+U2, U1, and U2. All of these have a more compact average structure than BP 5/5a RNA, in which base pairing of the internal loop constrains the molecule to stay unfolded. Additional information about the way that mutations in J5/5a destabilize the domain structure of P4–P6 comes from Fe(II) probing of the J5/5a mutants. In agreement with the native gel experiments, hydroxyl radical domain protections at 15 mM $MgCl_2$ are strongest for P4–P6, somewhat less so for U1+U2, and further decrease in intensity going from U1 to U2. BP 5/5a shows no domain protections. As the concentration of Mg^{2+} is increased, the relative compactness and the Fe(II) protections of the three U-containing mutants smoothly approach those of wild-type P4–P6. The $MgCl_2$ concentration required to re-

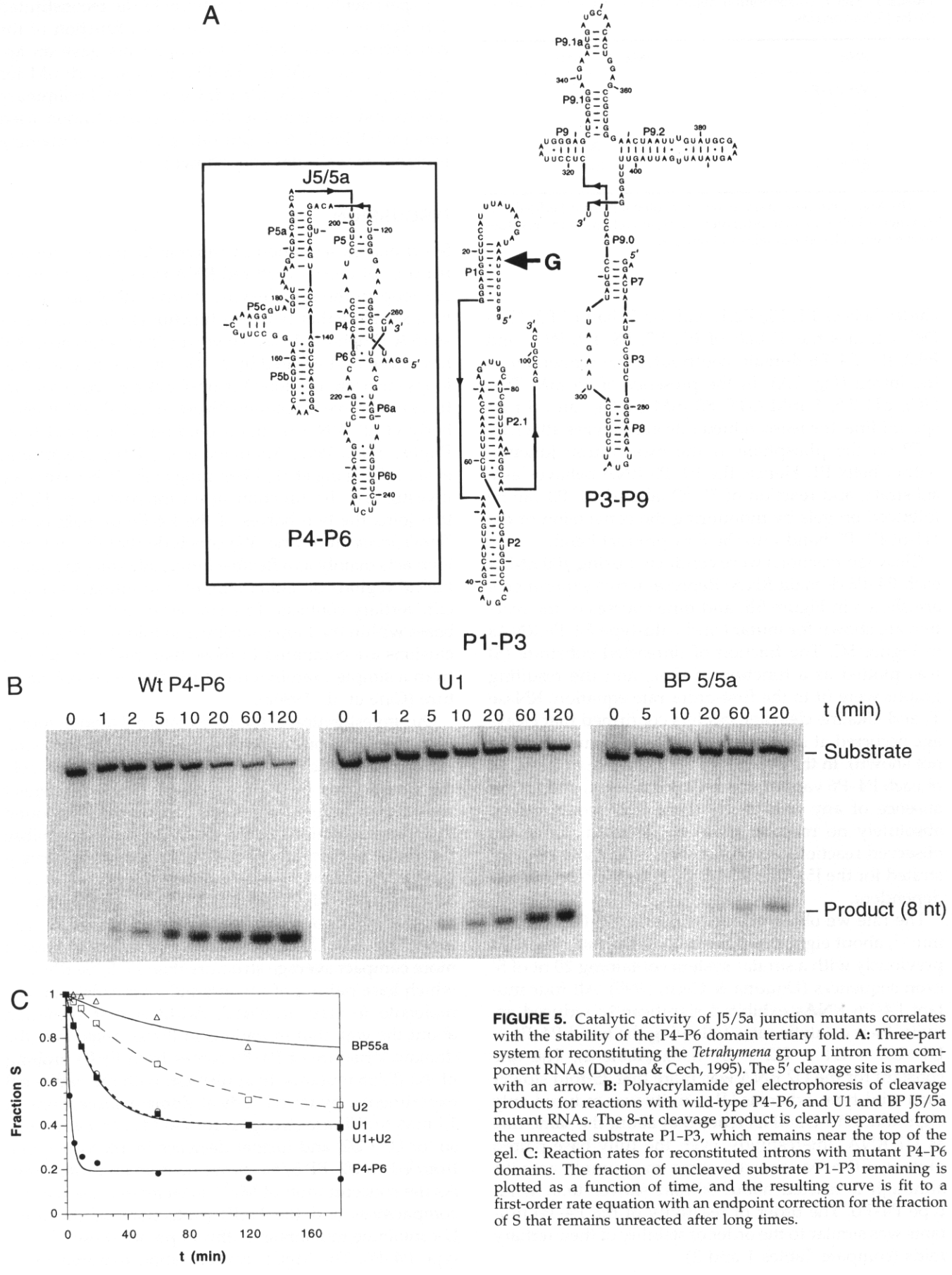


FIGURE 5. Catalytic activity of J5/5a junction mutants correlates with the stability of the P4–P6 domain tertiary fold. **A:** Three-part system for reconstituting the *Tetrahymena* group I intron from component RNAs (Doudna & Cech, 1995). The 5′ cleavage site is marked with an arrow. **B:** Polyacrylamide gel electrophoresis of cleavage products for reactions with wild-type P4–P6, and U1 and BP J5/5a mutant RNAs. The 8-nt cleavage product is clearly separated from the unreacted substrate P1–P3, which remains near the top of the gel. **C:** Reaction rates for reconstituted introns with mutant P4–P6 domains. The fraction of uncleaved substrate P1–P3 remaining is plotted as a function of time, and the resulting curve is fit to a first-order rate equation with an endpoint correction for the fraction of S that remains unreacted after long times.

TABLE 2. First-order rate constants for reconstituted *Tetrahymena* group I introns with mutant P4–P6 domains in 80 mM $MgCl_2$.[a]

Mutant	k_{obs} (min^{-1})	Rel. rate
Wt P4–P6	0.35	(1.00)
U1+U2	0.047	0.13
U1	0.054	0.15
U2	0.014	0.040
BP 5/5a	0.010	0.029

[a]Reaction conditions as described in Materials and Methods.

cover folding is highest for U2, less so for U1, and least for U1+U2 (see Table 1). Increased concentrations of Mg^{2+} can therefore rescue folding for these mutants, consistent with their tertiary structures being destabilized by base changes in the internal loop.

P4–P6 tertiary folding is an equilibrium between folded and unfolded states

The mutant P4–P6 RNAs show intermediate electrophoretic mobilities on native gels and intermediate levels of protection from free radical cleavage. Except for BP 5/5a, which cannot form a stable tertiary structure even at very high $MgCl_2$ concentrations, we were able to follow the Mg^{2+}-dependence of folding. At the other extreme from BP 5/5a is wild-type P4–P6, which is fully folded at low $MgCl_2$ concentrations. The U-containing mutants exhibit intermediate stability. At an $MgCl_2$ concentration of 20 mM, U1+U2 lies closest to the fully folded extreme, U1 less so, and U2 least of all (see Fig. 2C). However, with increasing $MgCl_2$, the mobility for each molecule is shifted more toward the folded state. Mutants U1 and U1+U2 are fully folded at 30 mM Mg^{2+} as observed by native gel mobility experiments. Finally, above 70 mM $MgCl_2$, all three U-containing mutants appear to be fully folded as measured by protections from Fe(II)-EDTA cleavage. Although the native gel experiments say nothing about the details of the structure of the unfolded state, it is presumed to have native secondary structure and, based on the Fe(II)-EDTA cleavage results, the subdomain is properly folded.

These results suggest that, in solution, the P4–P6 RNA is in an Mg^{2+}-dependent two-state equilibrium between folded and unfolded states (see Fig. 6). This equilibrium is rapid on the electrophoretic time scale, as indicated by the presence of single bands with intermediate mobilities on native gels over a wide range of $MgCl_2$ concentrations (Shen & Hagerman, 1994). Also, the formation of the domain structure must be a cooperative process because, for a given mutant at a specific $MgCl_2$ concentration, the Fe(II)-EDTA domain protections match in relative intensity, irrespective of their location in the molecule. These observations argue strongly against two alternative interpretations of

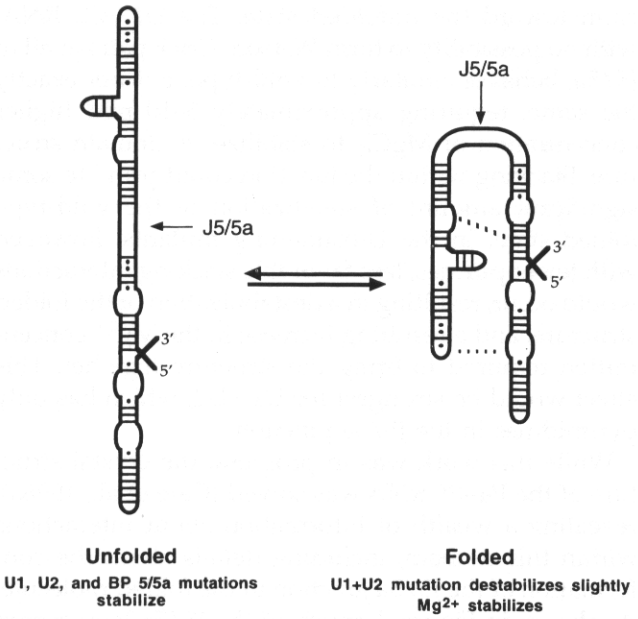


FIGURE 6. P4–P6 folding equilibrium. Schematic representation of the folding equilibrium between the extended and compact forms of the P4–P6 domain. Based on native gel and free radical protection experiments, P4–P6 appears to be in rapid equilibrium between folded and unfolded tertiary states in solution. Local unfolding of the P4–P6 domain tertiary structure does not occur. Mutations in U1, U2, and BP 5/5a primarily stabilize the unfolded form by forming base pairs, whereas increasing Mg^{2+} stabilizes the folded form. Mutations in U1+U2 may also destabilize the folded form due to a lack of purine stacking.

our data—that the mutant intermediate gel mobilities are the result of varying bend angles, or that portions of the mutant molecules might unfold locally. It seems unlikely that the many intermediate gel mobilities such as those seen in Figure 2C could all represent different stable structures of the molecule. Furthermore, it is hard to envision how molecules with a bend angle other than 180° would show the domain tertiary structure Fe(II)-EDTA protections that were observed.

Domain stability and the structure of the J5/5a internal loop

How can the different stabilities of the U-containing mutants be explained? In BP 5/5a, a complete Watson-Crick duplex is sufficient to inhibit formation of the tertiary structure. One difference between U1+U2, U1, and U2 is their potential to form Watson-Crick pairs in the J5/5a junction. Compared to the wild-type P4–P6 sequence shown in Figure 1C, U1 replaces a C-G pair with a possible U-G pair and two possible A-U pairs. In U2, there is the possibility of forming a total of three A-U pairs and a G-U in the joining region, more strongly inhibiting the formation of the bend. It is likely that the formation of alternative secondary structures in the loop stabilizes the unfolded form relative to the folded form (see Fig. 6), and thus moves the equilib-

rium toward the unfolded state. The U1+U2 RNA, with no possibility to form Watson–Crick pairs at all in J5/5a, behaves similarly to wild-type, but not exactly the same, requiring approximately 5–10 mM higher concentration of $MgCl_2$ to stabilize its domain structure. Stacking within the junction could provide some significant amount of stabilization of the wild-type folded state. In the U-containing mutants, however, with fewer purines, less favorable stacking interactions would occur, resulting in a destabilization of the folded structure, and a resulting increase in the Mg^{2+} concentration required to bring the structure together. This effect would be strongest for U1+U2, which has only pyrimidines in the J5/5a junction.

While this work was in progress, the crystal structure of the P4–P6 RNA was solved (Cate et al., 1996a), revealing a wealth of information about interactions within this domain, including details about the conformation of the J5/5a junction in the folded molecule. As shown in Figure 7, much of the J5/5a loop region is in fact structured in the crystal. Beginning with the P5a side of the junction, the bases C127, G126, and C124 are paired with G195, A196, and C197, respectively. A123 and A128 are somewhat far apart and at best may form a weak A-A pair, but A122–U199 is completely broken in this structure. Extensive stacking occurs between nt 195, 196, and 197, and also between residues 124, 126, and 127. Residue 125 is bulged out of the helical stack. Thus, with the exception of A125, most of the P5a side of the junction is stacked coaxially. The actual bend in P4–P6 only occurs at residues 122–123 and 198–199. Interestingly, the mutant U2, the most destabilized RNA next to BP 5/5a, has the potential to form two A-U pairs exactly where the bend occurs.

Because the J5/5a bend has a specific conformation in the crystal structure, and individual stacking and noncanonical base pairing interactions are observed, one might have been led to conclude that these interactions were essential for the bend. However, although the J5/5a junction is well organized in the crystal structure, this is not required for domain folding and the loop does not direct folding strongly. Even single-base mutations in the GAAA tetraloop in L5 continue to disrupt folding at $MgCl_2$ concentrations as high as 50–100 mM (Murphy & Cech, 1994). In contrast, the U1+U2 mutant only requires 20 mM Mg^{2+} to rescue folding completely. In addition, crystallographic *B*-factors suggest that the region may be dynamic (J. Doudna, pers. comm.), as our folding model would indicate. Nevertheless, a small effect due to the lack of correct stacking in the *folded* structure, along with a larger effect due to the increase in Watson–Crick base pairing in the *unfolded* structure, likely explains the destabilization of the U-containing mutants. Thus, it is essential that the J5/5a sequence not form stable structures incompatible with the folded state. It appears to be of secondary importance that the junction be stabilizing in the folded state.

The X-ray structure and our present mutational and thermodynamic analysis can be reconciled easily for two reasons. First, there is no reason to expect that every interaction seen in a crystal structure will be significant energetically. In addition, the crystal structure provides a view of the folded state, but the structure of the unfolded state also contributes to $\Delta G°$ for folding. In the two other regions of contact between the quasi-helical halves of P4–P6 (the tetraloop–receptor interaction and the A-rich bulge–minor groove interaction), the interactions seen in the crystal structure are essential for folding in solution (compare Cate et al., 1996a with Murphy & Cech, 1993 and Murphy et al., 1994).

Activity of the reconstituted intron and the stability of the P4–P6 domain structure

The various P4–P6 mutants produce very different rates of cleavage when included in the reconstituted intron system, with the order of reactivity following the trend P4–P6 > U1+U2 ≈ U1 > U2 > BP 5/5a. The relative strengths of Fe(II)-EDTA tertiary protections follow the trend P4–P6 > U1+U2 > U1 > U2 > BP 5/5a, as do native gel mobilities. Clearly, the folding of the P4–P6 tertiary structure strongly affects the activity of the reconstituted intron, and instability in P4–P6 cannot be rescued completely by the remainder of the intron. The Mg^{2+} concentration in our cleavage assays is 80 mM, above the concentration needed to fold fully all the J5/5a mutants except BP 5/5a. However, cleavage rates measured for the U-containing mutants are still well below the wild-type rate.

It is possible that the entire three-part system itself is rather destabilized, and is therefore especially sensitive to the stability of the P4–P6 domain. Activity has been measured for *Tetrahymena* splicing constructs lacking the P5abc extension; adding back an RNA corresponding to P5abc fully recovered wild-type splicing activity in the presence of 5 mM $MgCl_2$ and 200 mM NH_4Cl (van der Horst et al., 1991). In addition, only 15 mM $MgCl_2$ was required to regain full activity under these conditions, even without the P5abc extension.

Based on native gel and Fe(II)-EDTA experiments, the J5/5a mutants are destabilized but should nevertheless be fully folded at 80 mM Mg^{2+}. However, at this Mg^{2+} concentration, we still see lower than wild-type cleavage rates for the J5/5a mutants. One interpretation of the reduced cleavage rates we observed even at a high Mg^{2+} concentration is that the U-containing P4–P6 mutants do not assemble into the optimally reactive structure or are perturbed in their interaction with the remainder of the intron relative to the situation where P5abc is added in *trans*. Another possibility we have considered is that there is a specific requirement for movement within the P4–P6 domain structure during catalysis, and either the fre-

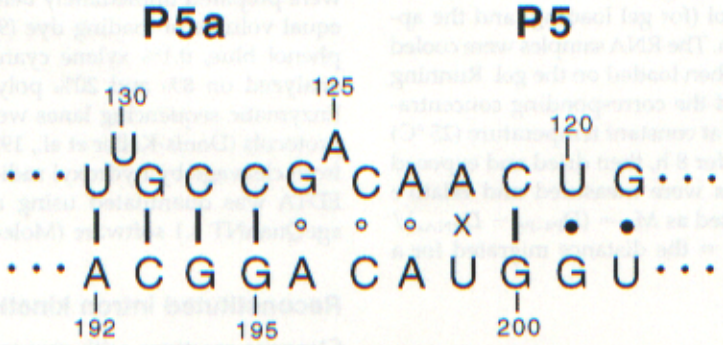


FIGURE 7. Crystal structure of the internal loop between P5 and P5a in the *Tetrahymena* P4–P6 domain (Cate et al., 1996a). Residues 122–127 and 195–199 in the J5/5a internal loop are labeled, along with helices P5 and P5a. The following bases are paired: C127:G195, G126:A196, and C124:C197 (indicated by open circles, G-U wobble pairs are marked with closed circles). A123 and A128 form an A-A pair, whereas A122:U199 (marked with an x) is completely unpaired in this structure. Extensive stacking occurs between nt 195, 196, and 197, and also between residues 124, 126, and 127. Residue 125 is bulged out of the helical stack. The actual bend in P4–P6 occurs at residues 122–123 and 198–199.

quency or type of motion is perturbed in the P4–P6 J5/5a mutants.

MATERIALS AND METHODS

Plasmids

Plasmids pP1–P3 and pP3–P9 (Doudna & Cech, 1995), containing portions of the *Tetrahymena* group I intron sequence positioned behind a T7 RNA polymerase promoter, were provided by J. Doudna. Plasmid pP1–P3 contains the wild-type *Tetrahymena* sequence from residues −6 to +103 (relative to the 5′ splice site), with two Gs added at the 5′ end for efficient T7 transcription. A *Bsa* I restriction endonuclease site follows position 103. Plasmid pP3–P9 contains the *Tetrahymena* sequence from position 262 to 414, in which the wild-type C262:G312 pair has been changed to G262:C312. A naturally occurring *Sca* I site is located 3′ of position 409.

P4–P6 J5/5a internal loop mutants were created from plasmid pP4–P6 (Murphy & Cech, 1993) using Kunkel mutagenesis (Kunkel, 1985; Kunkel et al., 1987) and verified by dideoxy sequencing of the plasmid DNA (Saenger et al., 1977).

RNA preparation

Plasmids pP1–P3, pP3–P9, and mutant and wild-type pP4–P6 were digested with the appropriate restriction enzymes and transcribed using T7 RNA polymerase (generously provided by Anne Gooding). Plasmid pP4–P6 was digested with *Ear* I restriction endonuclease to give a 160-nt runoff transcript that begins with two Gs 5′ of the naturally occurring nucleotide A104 and ends at A261. Plasmid pP1–P3 was digested with *Bsa* I endonuclease to produce a 111-nt RNA that begins with two Gs 5′ of C6 and ends at A103. Plasmid pP3–P9 was digested with *Sca* I to give a 153-nt RNA extending from G262 to U409. Transcription and gel purification of RNAs were performed as described (Latham et al., 1990), except that 8% polyacrylamide–7 M urea gels were used. RNA concentrations were measured by UV absorbance, assuming 1 A_{260} = 40 μg/mL RNA. P4–P6 and P1–P3 RNAs were 5′ end-labeled with ^{32}P as described (Latham et al., 1990). Following labeling reactions, RNAs were again gel purified on 8% polyacrylamide–7 M urea gels.

Native gel electrophoresis

Native 6% polyacrylamide gels (19:1 acrylamide:bisacrylamide) contained 89 mM Tris-borate, pH 8.1, 0.1 mM EDTA (TBE buffer), and varying concentrations of $MgCl_2$, depending on the experiment. Mutant and wild-type P4–P6 RNA samples were annealed by heating the RNA to 50 °C in TBE buffer, along with 5% glycerol (for gel loading) and the appropriate $MgCl_2$ concentration. The RNA samples were cooled on the bench for 10 min and then loaded on the gel. Running buffer contained 1× TBE and the corresponding concentration of $MgCl_2$. Gels were run at constant temperature (25 °C) and constant voltage (150 V) for 8 h, then dried and exposed to film. RNA band positions were measured and relative RNA mobilities were calculated as $M_R = (D_{P4\text{-}P6} - D_{RNAx})/(D_{P4\text{-}P6} - D_{BP55a})$, where D_X = the distance migrated for a given RNA band.

Ferguson analysis

Native polyacrylamide gels with acrylamide concentrations between 4% and 12% (19:1 acrylamide:bisacrylamide) were used to generate Ferguson plots. These gels were poured and allowed to stand overnight to ensure complete polymerization. Samples were treated as for standard native gel experiments. Gels were run at constant temperature (27 °C) and at constant voltage (200 V) for 2–7 h, depending on gel concentration, then dried and exposed.

For rapidly tumbling molecules traveling through a gel matrix, $\log M = \log M_0 + K_R T$, where M_0 is the intrinsic mobility of the molecule in the absence of gel material, M is the observed mobility, K_R is the retardation coefficient, and T is the total gel concentration (Rodbard & Chrambach, 1971). Empirically, K_R is a linear function of molecular mass, $K_R = A + B(MW)$. M_0 is a function of mass, charge, and counter ions, but not of molecular geometry. For a gel composed of a random meshwork of fibers, it can be shown that the retardation coefficient $K_R^{1/2} = c(R + r)$, where R is the molecular radius, r is the gel fiber radius, and c is a constant for a given system that encompasses ionic strength, temperature, pH, extent of crosslinking, degree of polymerization, etc. (Rodbard & Chrambach, 1971).

For each mutant P4–P6 RNA band on native gel autoradiograms, mobility (M) was calculated by dividing the distance migrated (cm) by run time (s) and field strength (V/cm). Log M was then plotted as a function of gel concentration, and the resulting family of curves were simultaneously fit to determine c and R using the program Elphofit (Tietz, 1991; Tietz & Chrambach, 1993), assuming a fiber radius of 1 nm and a dry fiber volume of 0.9 mL/g.

Free radical footprinting

Fe(II)-EDTA reactions were performed as described (Latham & Cech, 1989) with a few modifications. P4–P6 domain RNA labeled at the 5′ end with ^{32}P was incubated at 50 °C for 10 min in 8 μL of 44 mM Tris-HCl, pH 7.5, in the presence of various concentrations of $MgCl_2$, in order to ensure that the RNA adopts the biologically active conformation (Herschlag & Cech, 1990). Samples were then transferred to 42 °C to equilibrate for 5 min. Next, 1 μL of Fe(II)-EDTA reagent [10 mM $(NH_4)_2Fe(SO_4)_2$, 20 mM Na_2EDTA, pH 8.0] and then 1 μL of 50 mM DTT were added to a final sample volume of 10 μL. The final solution contained 35 mM Tris-HCl, pH 7.5, 1 mM $(NH_4)_2Fe(SO_4)_2$, 2 mM Na_2EDTA, pH 8.0, 5 mM DTT, and various concentrations of $MgCl_2$, depending on the experiment. Reactions were incubated at 42 °C for 80 min, then quenched on ice with 1 μL of 100 mM thiourea. Reagents were prepared immediately before use. After addition of an equal volume of loading dye (95% formamide, 0.1% bromphenol blue, 0.1% xylene cyanol), reaction products were analyzed on 8% and 20% polyacrylamide–7 M urea gels. Enzymatic sequencing lanes were prepared using standard protocols (Donis-Keller et al., 1977). The extent of protection from cleavage by hydroxyl radicals generated by the Fe(II)-EDTA was quantitated using a PhosphorImager and ImageQuanNT 4.1 software (Molecular Dynamics).

Reconstituted intron kinetics

Cleavage reactions with reconstituted *Tetrahymena* group I introns (Doudna & Cech, 1995) were conducted as follows. A mixture of 1 nM 5′ ^{32}P-labeled P1–P3, 100 nM P4–P6, and 100 nM P3–P9 were annealed for 10 min at 50 °C in buffer containing 30 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, and 5 mM spermidine, then allowed to equilibrate at 40 °C for 5 min. Cleavage reactions were initiated by adjusting the $MgCl_2$ and GTP concentrations to 80 mM and 5 mM, respectively, using a pre-warmed 5× stock solution. Reactions were performed at 40 °C. Aliquots were removed at various times and quenched on ice with an equal volume of 95% formamide containing loading dyes. Reaction products were analyzed on 8% polyacrylamide gels containing 7 M urea and the labeled substrate and product bands were quantitated on dried gels using a PhosphorImager (Molecular Dynamics). After correcting for background counts, the fraction unreacted substrate [Fraction S = substrate counts/(substrate

counts + product counts)] remaining at each time point was calculated and plotted as a function of time. First-order rate constants were obtained by fitting the resulting plots to the expression Fraction S, where $m0$ = time, $m1$ = first-order rate constant, k_{obs}, and $m2$ = the unreactive substrate fraction.

ACKNOWLEDGMENTS

We are grateful to P. Bevilacqua, J. Sherman, and L. Weinstein for helpful discussions, and to J. Doudna for providing plasmids pP1-P3 and pP3-P9. We thank Anne Gooding for providing T7 RNA polymerase. A.A.S. is a fellow of the Jane Coffin Childs Memorial Fund for Medical Research. T.R.C. is an Investigator of the Howard Hughes Medical Institute and an American Cancer Society Professor.

Received April 7, 1997; returned for revision April 29, 1997; revised manuscript received May 12, 1997

REFERENCES

Aboul-ela F, Karn J, Varani G. 1995. The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J Mol Biol 253*:313–332.

Atkins JF, Gesteland RF. 1993. *The RNA world*. Plainview, New York: Cold Spring Harbor Laboratory Press.

Basavappa R, Sigler P. 1991. The 3Å crystal structure of yeast initiator tRNA: Functional implications in initiator/elongator discrimination. *EMBO J 10*:3105–3111.

Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA. 1996a. Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science 273*:1678–1685.

Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Szewczak AA, Kundrot CE, Cech TR, Doudna JA. 1996b. RNA tertiary structure mediation by adenosine platforms. *Science 273*:1696–1699.

Celander DW, Cech TR. 1990. Iron(II)-ethylenediaminetetraacetic acid catalyzed cleavage of RNA and DNA oligonucleotides: Similar reactivity toward single- and double-stranded forms. *Biochemistry 29*:1355–1361.

Costa M, Michel F. 1995. Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J 14*:1276–1285.

Dieckmann T, Suzuki E, Nakamura GK, Feigon J. 1996. Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *RNA 2*:628–640.

Donis-Keller H, Maxam AM, Gilbert W. 1977. Mapping adenines, guanines, and pyrimidines in RNA. *Nucleic Acids Res 4*:2527–2538.

Doudna JA, Cech TR. 1995. Self-assembly of a group I intron active site from its component tertiary structural domains. *RNA 1*:36–45.

Downs WD, Cech TR. 1996. Kinetic pathway for folding of the *Tetrahymena* ribozyme revealed by three UV-inducible crosslinks. *RNA 2*:718–732.

Feldstein PA, Bruening G. 1993. Catalytically active geometry in the reversible circularization of 'mini-monomer' RNAs derived from the complementary strand of tobacco ringspot virus satellite RNA. *Nucleic Acids Res 21*:1991–1998.

Ferguson KA. 1964. Starch-gel electrophoresis—Application to the classification of pituitary proteins and polypeptides. *Metabolism 13*:985–1002.

Herschlag D, Cech TR. 1990. Catalysis of RNA cleavage by the *Tetrahymena thermophila* ribozyme. 1. Kinetic description of the reaction of an RNA substrate complementary to the active site. *Biochemistry 29*:10159–10171.

Hertzberg R, Dervan PB. 1984. Cleavage of DNA with methidium-propyl-EDTA-iron(II): Reaction conditions and product analyses. *Biochemistry 23*:3934–3945.

Jiang F, Kumar RA, Jones RA, Patel DJ. 1996. Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex. *Nature 382*:183–186.

Kim SH, Suddath FL, Quigley GJ, McPherson A, Sussman JL, Wang AH, Seeman NC, Rich A. 1974. Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science 185*:435–440.

Kunkel TA. 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc Natl Acad Sci USA 82*:488–492.

Kunkel TA, Roberts JD, Zakour RA. 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol 154*:367–382.

Latham JA, Cech TR. 1989. Defining the inside and outside of a catalytic RNA molecule. *Science 245*:276–282.

Latham JA, Zaug AJ, Cech TR. 1990. Self-splicing and enzymatic cleavage of RNA by a group I intervening sequence. *Methods Enzymol 181*:558–569.

McKay DB. 1996. Structure and function of the hammerhead ribozyme: An unfinished story. *RNA 2*:395–403.

Michel F, Westhof E. 1990. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol 216*:585–610.

Murphy FL, Cech TR. 1993. An independently folding domain of RNA tertiary structure within the *Tetrahymena* ribozyme. *Biochemistry 32*:5291–5300.

Murphy FL, Cech TR. 1994. GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain. *J Mol Biol 236*:49–63.

Murphy FL, Wang YH, Griffith JD, Cech TR. 1994. Coaxially stacked RNA helices in the catalytic center of the *Tetrahymena* ribozyme. *Science 265*:1709–1712.

Puglisi JD, Tan R, Calnan BJ, Frankel AD, Williamson JR. 1992. Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science 257*:76–80.

Robertus JD, Ladner JE, Finch JR, Rhodes D, Brown RS, Clark BF, Klug A. 1974. Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature 250*:546–551.

Rodbard D, Chrambach A. 1971. Estimation of molecular radius, free mobility, and valence using polyacrylamide gel electrophoresis. *Anal Biochem 40*:95–134.

Saenger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA 74*:5463–5467.

Shen Z, Hagerman PJ. 1994. Conformation of the central, three-helix junction of the 5 S ribosomal RNA of *Sulfolobus acidocaldarius*. *J Mol Biol 241*:415–430.

Szewczak AA, Moore PB, Chang YL, Wool IG. 1993. The conformation of the sarcin/ricin loop from 28S ribosomal RNA. *Proc Natl Acad Sci USA 90*:9581–9585.

Tietz D. 1991. Analysis of one-dimensional gels and two-dimensional Serwer-type gels on the basis of the extended Ogston model using personal computers. *Electrophoresis 12*:28–39.

Tietz D, Chrambach A. 1993. DNA shape and separation efficiency in polymer media: A computerized method based on electrophoretic mobility data. *Electrophoresis 14*:185–190.

Tullius TD, Dombrowski BA. 1985. Iron(II) EDTA used to measure the helical twist along any DNA molecule. *Science 239*:679–681.

van der Horst G, Christian A, Inoue T. 1991. Reconstitution of a group I intron self-splicing reaction with an activator RNA. *Proc Natl Acad Sci USA 88*:184–188.

Westhof E, Dumas P, Moras D. 1985. Crystallographic refinement of yeast aspartic acid transfer RNA. *J Mol Biol 184*:119–145.

Wimberly B, Varani G, Tinoco IJ. 1993. The conformation of loop E of eukaryotic 5S ribosomal RNA. *Biochemistry 32*:1078–1087.

Woo NH, Roe BA, Rich A. 1980. Three-dimensional structure of *E. coli* initiator $tRNA^{Met}$. *Nature 286*:346–351.

Yang Y, Kochoyan M, Burgstaller P, Westhof E, Famulok M. 1996. Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. *Science 272*:1343–1347.

Zarrinkar PP, Williamson JR. 1994. Kinetic intermediates in RNA folding. *Science 265*:918–924.

Zarrinkar PP, Williamson JR. 1996. The kinetic folding pathway of the *Tetrahymena* ribozyme reveals possible similarities between RNA and protein folding. *Nature Struct Biol 3*:432–438.