A Functional Ribosomal RNA Tertiary Structure Involves a Base Triple Interaction†

Graeme L. Conn,‡ Robin R. Gutell,§ and David E. Draper*,‡

Department of Chemistry, The Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, and Department of Chemistry and Biochemistry, Campus Box 215, University of Colorado, Boulder, Colorado 80309

Received April 13, 1998; Revised Manuscript Received June 30, 1998

ABSTRACT: Comparative sequence analysis reveals a coordinated set of nucleotide exchanges between the base pair 1092/1099 and the unpaired position 1072 [(1092/1099)1072] in the L11 binding domain of 23S ribosomal RNA. This set of exchanges has occurred at least 4 times during evolution, suggesting that these positions form a base triple. The analysis further suggests an important role for positions (1065/1073), adjacent to 1072. The covariation at positions (1092/1099)1072 is studied here by analysis of RNA variants using UV melting and binding of ribosomal protein L11 and thiostrepton to assay for tertiary folding of this domain. The tertiary structure of the RNA is eliminated by alteration of the unpaired nucleotide (C1072 to U mutation), and binding of L11 and thiostrepton are reduced 10-fold compared to the wild type. In contrast, substitution of the base pair (CG1092/1099 to UA mutation) allows formation of the tertiary structure but dramatically alters the pH dependence of tertiary folding. The fully compensated set of mutations, (CG)C to (UA)U, restores the tertiary structure of the RNA to a state almost identical to the wild type. The nature of this base triple and its implications for the folding of the RNA and ligand interactions are discussed.

RNA molecules can form a wide range of tertiary structures which are often crucial to their function in vivo. Base triples are one of the predominant modes of tertiary interaction and have been identified in a variety of RNAs, such as tRNA (*1, 2*), the group I intron (*3*−*6*), RNase P (*7*), ribosomal RNA (*8, 9*), and the human immunodeficiency virus Tat protein binding site, TAR RNA (*10*). Expanding sequence databases and the increasing sophistication of computer algorithms make identification of potential base triples with comparative sequence analysis possible (*11*). Base triples involving the interaction of a single-stranded nucleotide with a base pair over large distances in the RNA secondary structure provide valuable tertiary structure information in large RNA domains currently beyond the scope of high-resolution structural analysis.

We are interested in the tertiary folding of the highly conserved 58 nucleotide fragment, G1051−U1108, of *E. coli* large subunit RNA which forms part of the ribosome structure responsible for elongation factor G-dependent hydrolysis of GTP. While the initial secondary structure model for this region (*12*) has been substantiated and extended with further comparative sequence (*8, 13*) and mutational analyses (Figure 1A), less is known about the details of the tertiary folding. We have previously presented evidence that this RNA contains a defined set of tertiary interactions crucial to its interaction with ligands and ions. The tertiary structure is recognized and stabilized by the conserved ribosomal protein L11 (*14, 15*) and contains the site of interaction of the peptide antibiotic thiostrepton with the ribosome (*16, 17*). Thiostrepton probably makes contacts with nucleotides in both loops of the RNA, suggesting that the RNA is folded so that the loops come into close proximity (*18*). The folding of the RNA sugar−phosphate backbone into such compact three-dimensional structures is invariably sensitive to monovalent and often divalent cation concentration. In addition to the general effects of the ion atmosphere, we have shown that this RNA fragment has binding sites specific for magnesium and ammonium or potassium ions in preference to any other di- or monovalent ion (*19, 20*).

In this study, we have used a comparative analysis of over 550 aligned 23S and 23S-like sequences to identify 2 sets of covariations in the 58 nucleotide RNA domain. One of these covariances, (1092/1099)1072, is clearly substantiated by UV melting experiments and filter binding assays with RNA sequence variants revealing a base triple interaction within the L11 binding region. This interaction represents the first experimentally substantiated base triple to be involved in the folding of a ribosomal RNA tertiary domain and is crucial for the interaction of both ribosomal protein L11 and the thiazole antibiotic thiostrepton.

MATERIALS AND METHODS

Mutagenesis and RNA Preparation. The preparation of a pUC18 derivative containing a T7 promoter immediately followed by DNA encoding the 58 nucleotide RNA and an *Rsa*I restriction site for runoff transcription has been described previously (*19, 21*). RNA transcripts from this plasmid contained an additional 5′-G nucleotide resulting from a *Stu*I site used during cloning. This additional nucleotide was deleted by unique site elimination (USE) mutagenesis (*22*) to generate the wild-type *E. coli* G1051−

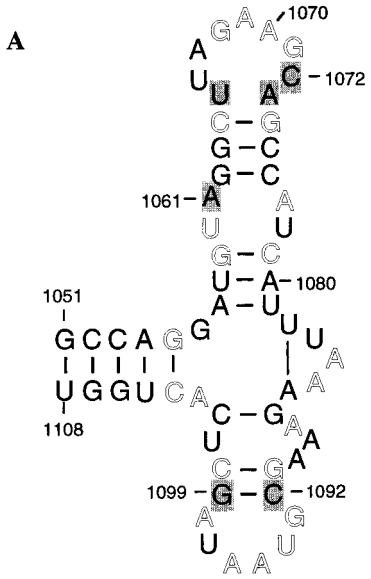
† This work was supported by a Wellcome Trust (UK) International Prize Traveling Fellowship (G.L.C.) and by National Institutes of Health Grants GM29048 (D.E.D.) and GM48207 (R.R.G.).

* To whom correspondence should be addressed (e-mail: draper@jhunix.hcf.jhu.edu).

‡ The Johns Hopkins University.

§ University of Colorado.

S0006-2960(98)00825-3 CCC: $15.00
Published on Web 08/05/1998



B

RNA	(CG)C	(UA)C	(CG)U	(UA)U
1072	C	C	U	U
(1065/73)	(UA)	(UA)	(UA)	(UA)
1092/99	CG	UA	CG	UA

RNA	(GC)C	(GC)U
1072	C	U
(1065/73)	(UA)	(UA)
1092/99	GC	GC

RNA	AC(CG)C	AC(UA)U
1072	C	U
(1065/73)	(AC)	(AC)
1092/99	CG	UA

FIGURE 1: (A) Secondary structure of the 58 nucleotide RNA fragment (*E. coli* numbering). Nucleotides conserved (>95%) in the three primary phylogenetic domains [Archaea, (eu)Bacteria, and Eucarya] and chloroplast (http://pundit.colorado.edu:8080; Gutell, 1993) are shown in outline typeface. Sites of mutation are shadowed (the A1061 mutation was present in all RNA variants). (B) RNA sequence variants used in this study.

U1108 sequence (the *Rsa*I cut gives the correct 3′ end). All further sequence variants were prepared by USE mutagenesis of this plasmid. RNA was synthesized from *Rsa*I-linearized plasmid DNA by T7 RNA polymerase runoff transcription as previously described (*19, 23*). The RNA was purified by gel electrophoresis using 8 M urea and 12 or 20% acrylamide, the excised bands were electroeluted, and the RNA was recovered by ethanol precipitation. ^{35}S-labeled RNA for filter binding assays was transcribed under essentially the same conditions using the *Rsa*I-restricted plasmids and purified on Nensorb 20 columns (Du Pont). Analysis by 12% PAGE and autoradiography showed one major RNA species.

UV Melting Experiments. RNA samples were renatured at 65 °C for 15 min in the melting buffer (10 mM KOAc, pH 4.5 and 5.0, 10 mM MES, pH 5.5 and 6.0, 10 mM MOPS, pH 6.5 and 7.0, 10 mM HEPES, pH 7.5 and 8.0, and 10 mM EPPS, pH 8.5) containing 5 mM $MgSO_4$ and 100 mM NH_4Cl (M_5A_{100}). A cooling curve, from 65 to 10 °C, was collected prior to melting as this was found to give a more reliable lower base line, presumably as it allowed for complete renaturation (although it did not affect the apparent T_m). Between 10 and 65 °C (where the tertiary structure unfolds), the heating and cooling curves were essentially superimposable under these conditions, indicating reversible folding. Data points for the melts were collected every 0.4 °C with a heating rate of 0.8 °C/min as described previously (*21*). The 'melting profiles' obtained for each RNA (see Results) are the first derivatives of the melting curves averaged over a 4 °C window and normalized to the absorbance at 20 °C.

Filter Binding Assays. The preparation of L11-C76 (the C-terminal 76 residues of *B. stearothermophilus* ribosomal protein L11) has been described previously (*24*). Thiostrepton was purchased from CalBiochem. L11−RNA and thiostrepton−RNA complexes were detected by retention on nitrocellulose filters as described previously (*15, 25*), except that L11-C76 did not need to be renatured. The solution conditions for L11-C76 binding assays were 10 mM MOPS, pH 7.5, or 10 mM MES, pH 5.5, 3 mM $MgCl_2$, 175 mM KCl (M_3K_{175}). For thiostrepton, 175 mM NH_4Cl (M_3A_{175}) was used in place of KCl with a final concentration of 5% DMSO.

RESULTS

Phylogenetic Evidence for a Base Triple Interaction in the 58 Nucleotide RNA Domain. In a comparative analysis, the number of sequences and the structural and phylogenetic diversity among them are very important. The present analysis includes over 550 sequences that span the 1051−1108 region of 23S and 23S-like rRNA. These sequences are distributed across the three primary phylogenetic domains [Eucarya, Archaea, and (eu)Bacteria] and the two Eucarya organelles (chloroplast and mitochondria) and include sequences for the majority of the most significant branches of the phylogenetic tree. Except for a few of the mitochondria, these sequences are sufficiently conserved and with few insertions/deletions so their alignment is mostly unambiguous. There is also adequate sequence variation, allowing for the possibility of positional covariations within this region of the rRNA. Thus, our sequence collection is large, and phylogenetically and structurally diverse.

Our analysis of the aligned sequences made use of two correlation algorithms (*11*). The first is a chi-square-based analysis of the triplet frequencies at previously proposed secondary structure base pairs with a third unpaired nucleotide. The second calculates a pseudo-phylogenetic count and score for a collective set of changes that have occurred during the evolution of a base pair and an unpaired position. Two possible covariances within the 58 nucleotide RNA domain were identified. A strong covariance was found between positions 1092/1099 and the unpaired position 1072. A weaker, though still significant, relationship was observed between the positions in the (1092/1099)1072 covariance and positions 1065 and 1073, the first and last nucleotides in the hairpin loop containing the unpaired nucleotide associated with this triplet. The base frequencies for these positions are presented in Table 1.

The relationship between positions (1092/1099) and 1072 is strong, with several examples of coordinated changes

Table 1: Results of Comparative Sequence Alignment Analysis for the Covariances at Positions (1092/1099)1072 and (1065/1073)1072[a]

		% (1092/1099)1072			% (1065/1073)1072		
	no. of sequences	(CG)C	(UA)U	others	(YA)C	(AY)U	others
(eu)Bacteria	166	100	0	0	100	0	0
Archaea	25	100	0	0	100	0	0
Eucarya	105	2[b]	98	0	0	93	7[c]
chloroplast	96	100	0	0	99	0	1
mitochondria	170	85	10	5	85	0	15
(mitochondria)	(144)	(85)	−	−	(98)	(0)	(2)
(mitochondria)	(17)	−	(10)	−	(0)	(0)	(100)
total	562						

[a] The last two rows (in parentheses) show the breakdown of sequences observed at (1065/1073) within the mitochondria for both (1092/1099)1072 triples in more detail. [b] *Physarum polycephalum* and *Didymium iridis* are the only two known Eucarya with a (CG)C triplet. [c] *P. polycephalum* and *D. iridis* both have (AC)C.

occurring during evolution. While all of the known (eu)Bacteria, chloroplast, and Archaea 23S rRNA sequences have a (CG)C at positions (1092/1099)1072, there are examples of (UA)U and (CG)C within the known Eucarya nuclear and mitochondrial sequences (see Table 1). All of the Eucarya have a (UA)U except for the two phylogenetically related protists *Physarum polycephalum* and *Didymium iridis* (the only two members of the Myxomycetes group that we have sequence information for), which have a (CG)C triplet. Thus, with our current sequence dataset, there appears to be a single phylogenetic event leading to these two triplets within the Eucarya phylogenetic group. And since the Myxomycetes group is not at the base of the Eucarya tree, we can infer that another phylogenetic event occurred at the branch between the prokaryotes and the Eucarya, leading to the (CG)C triplet in prokaryotes, and the predominant (UA)U triplet in the Eucarya.

Within the mitochondria, there are several phylogenetic events underlying the distribution of (CG)C and (UA)U triplets. While the (CG)C triplet is present in the majority of the mitochondria within all of the major Eucarya phylogenetic groups (protists, fungi, plants, and animals; this includes some of the Ascomycota fungi and Arthropod animals), the (UA)U triple occurs in several fungi (Ascomycota and Basidiomycota) and animals (Arthropods, Mollusca, and Nematodes). Thus, there has been at least one phylogenetic event in the Ascomycota and another within the Arthropods. Since the mitochondria are thought to have evolved from the (eu)Bacteria, which are composed of the (CG)C triplet, the minimum number of phylogenetic events within and leading to the mitochondria is 2. So all totaled, there are a minimum of four coordinated sets of changes underlying the distribution of (CG)C and (UA)U triplets at positions (1092/1099)1072.

The relationship between the (1092/1099)1072 triplet and positions (1065/1073), the first and last nucleotides of the hairpin, is suggestive although not as strong as that between (1092/1099) and 1072. The identity of position 1072 appears crucial: when this position is C, (1065/1073) is YA (where Y is C or U), and when it is U, (1065/1073) is AY. The correlation between these positions across the phylogenetic domains and Eucarya organelles is shown in detail in Table 1. The relationship is nearly invariant in the (eu)Bacteria, chloroplast, and Archaea. However, there are exceptions to the rule within the Eucarya and mitochondria. In particular, the two Eucarya sequences with a (CG)C triplet at positions (1092/1099)1072 do not have (YA)C at (1065/1073), and none of the mitochondrial sequences with a (UA)U triplet at (1092/1099)1072 have the canonical (AY)U at (1065/1073)1072. Cases of concerted base changes as observed for the (1092/1099)1072 triplet are not as clear. In conclusion, there is a correlation between the two sets of positions (1092/1099)1072 and (1063/1075); however, it is not as pure as the individual (1092/1099)1072 triplet.

Analysis of RNA Sequence Variants. Base mutations in the *E. coli* 58 nucleotide RNA fragment (Figure 1B) were made to determine the structural basis of the covariations identified between C1072 and positions CG1092/1099 and UA1065/1073. All of the mutations were made in the context of an additional mutation, U1061 to A (U1061A; see Figure 1A), which is naturally present in some thermophilic Archaea and greatly stabilizes the RNA tertiary structure compared to the wild-type *E. coli* sequence (*26*). This allowed us to quantify greater destabilizations of tertiary structure caused by disruption of important RNA−RNA interactions.

The unfolding of the 58 nucleotide RNA tertiary structure occurs before the melting of the secondary structure elements and is characterized by a large difference in hyperchromicity between melting curves taken at different wavelengths. A region of strong hyperchromicity in the 260 nm melting profile but with little or none in the 280 nm profile is diagnostic of the melting of the tertiary structure (*26*). The 280 nm data essentially correspond to the unfolding of the various secondary structure elements shown in Figure 1A. Therefore, melting experiments provide an excellent assay for the structural role of the proposed triples in the tertiary folding. As base triple interactions often involve protonation of a functional group, melting profiles were collected for each RNA variant over the pH range 4.5−8.5 (in 0.5 pH unit increments) in M_5A_{100} buffer. The 280 nm melting profiles were compared for each variant to ensure that neither the mutations nor the change in pH was adversely affecting the RNA secondary structure rather than the tertiary structure alone. At pH 4.5, major destabilizations of the RNA tertiary structure and secondary structure were observed in each RNA variant, most probably the result of titration of functional groups in the RNA (most likely cytosine N3 or adenine N1). These melting profiles are excluded from further discussion. In comparing the RNA variants, three main features of the melting profiles are important: the overall shape of the profile, the apparent T_m of tertiary unfolding, and the dependence of both of these upon solution pH.

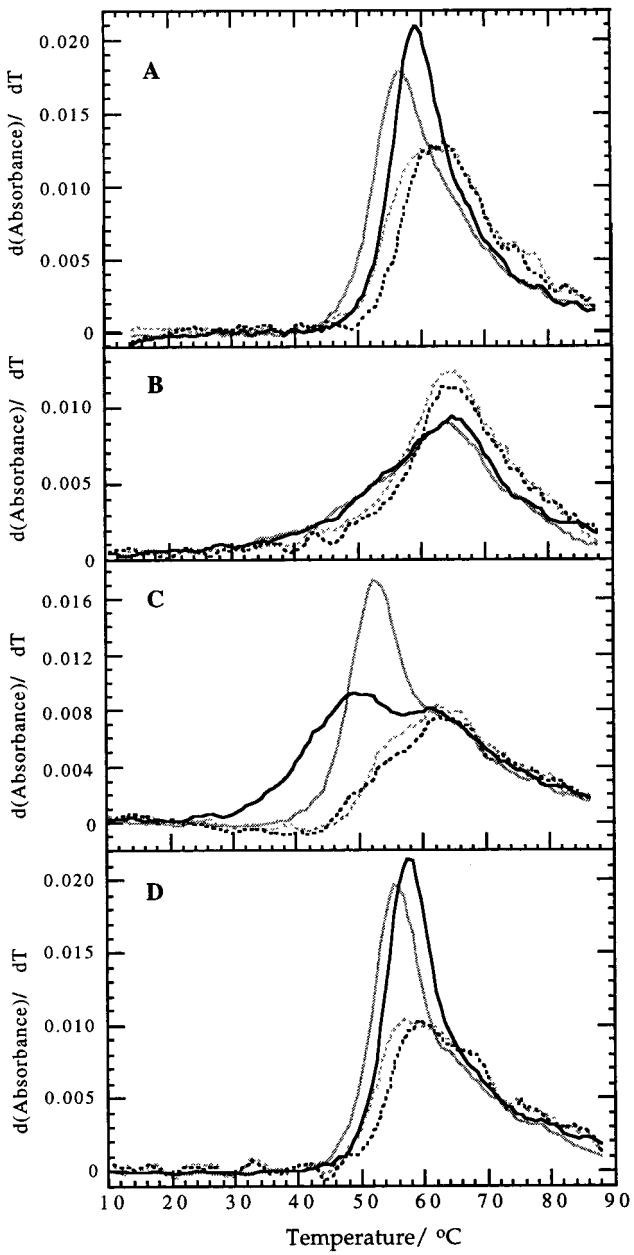


FIGURE 2: UV melting profiles of (A) (CG)C RNA, (B) (CG)U RNA, (C) (UA)C RNA, and (D) (UA)U RNA at pH 7.5 (black) and pH 5.5 (gray) in M_5A_{100} buffer at 260 (solid line) and 280 nm (dashed line).

(CG1092/1099) C1072 Base Triple. For clarity, melting profiles at 260 and 280 nm for each RNA at only two pH points (pH 5.5 and 7.5) are shown in Figure 2. The apparent T_m values for tertiary structure unfolding are given in Table 2. The melting of (CG)C RNA is highly cooperative with a sharp peak in the melting profile, corresponding to the tertiary structure unfolding, and an apparent T_m of 59.4 °C (Figure 2A). The tertiary structure of (CG)C RNA has maximal thermal stability at pH 7.5 but varies little over a range of pH (pH 6−8). At lower pH, there is a small decrease in the hyperchromicity associated with the melting of tertiary structure, and the apparent T_m is decreased significantly only by pH 5.0 (around 5 °C). The same trend with pH is observed for the *E. coli* wild-type sequence without the additional U1061A mutation (not shown). The mutation C1072U, (CG)U RNA, eliminates the tertiary structure but does not affect the secondary structure. Very little difference can be seen between the 260 and 280 nm data, and the melting profile corresponds to the melting of the various secondary structure elements alone (Figure 2B). Alteration of the pH does not affect the melting profiles at either wavelength for this RNA. The unfolding of RNA tertiary structure is still observed in the melting profiles of (UA)C RNA (CG1092/1099 to UA mutation). However, in contrast to the wild-type (CG)C triple, the (UA)C triple shows a significant sensitivity to solution pH (Figure 2C). As the pH is lowered, the tertiary structure transition in the melting profile becomes much sharper, and the apparent T_m increases by around 2−3 °C (Table 2) although the actual change in T_m is probably much greater. In contrast, the part of the melting profile corresponding to the secondary structure elements is not affected by changing pH. The fully compensatory set of mutations, of the base pair 1092/1099 and position 1072 to (UA)U, restores the RNA tertiary structure to a state almost identical to (CG)C RNA. Thus, the melting profile shows a sharp peak corresponding to the unfolding of the tertiary structure (Figure 2D) and with an only slightly reduced apparent T_m (Table 2). Furthermore, the pH dependence caused by the UA1092/1099 mutation has been removed, and this RNA shows maximal stability at around neutral pH. As with the wild-type (CG)C triple, lowering the pH has only a small effect on the hyperchromicity and T_m associated with tertiary structure unfolding.

Table 2: Values of the Apparent T_m for Tertiary Structure Unfolding and Binding of L11-C76 and Thiostrepton for Each RNA Sequence Variant

RNA	apparent T_m (°C)[a] at pH 5.5	apparent T_m (°C)[a] at pH 7.5	L11-C76 K_a (μM^{-1})[b]	thiostrepton K_a (μM^{-1})[b]
(CG)C	56.6	59.4	30 ± 6.6	1.5 ± 0.16
(UA)C	51.5	49[c]	20 ± 2.3	1.6 ± 0.09
(CG)U	−	−	2.0 ± 0.4	0.18 ± 0.02
(UA)U	54.4	56.8	33 ± 3.0	1.7 ± 0.12
(GC)C	52.3	−	11 ± 2.0	1.3 ± 0.19
			26 ± 8.0[d]	1.6 ± 0.02[d]
(GC)U	−	−	8.5 ± 2.7	0.52 ± 0.08
AC(CG)C	35−40[c]	35−40[c]	nd	nd
AC(UA)U	49.6	50.3	nd	nd

[a] The apparent T_m is the maximum in the first-derivative curve corresponding to tertiary structure unfolding. The actual T_m may be slightly higher. [b] Errors shown for K_a values are the standard deviation of values determined in at least three separate assays (nd = not determined). [c] Approximate apparent T_m for broad tertiary structure transition. [d] K_a determined at pH 5.5 (as described under Materials and Methods).

UV melting profiles for (UA)C RNA collected at pH points between 5.5 and 8.0 were used to estimate the pK_a of the protonated site in the base triple. The melting profiles were simultaneously fit to a set of four sequential unfolding transitions (each with an associated T_m, enthalpy, and hyperchromicity) to give a binding polynomial for protonation (*27*). The binding polynomial was best fit using a model with a single site of protonation in both the folded and unfolded forms of the RNA, with pK_a values of 7.6 ± 0.1 and 6.2 ± 0.1, respectively (errors were estimated from the fit using an error of ±0.5 °C in each T_m). Both are significantly shifted from the expected pK_a of cytosine N3 in free nucleotides of around 4.5 (*28*), but the effect is much greater in the folded form. Shifts in pK_a values of functional groups in proteins have been attributed to the effects of microenvironment within the cores of folded proteins (*29*,

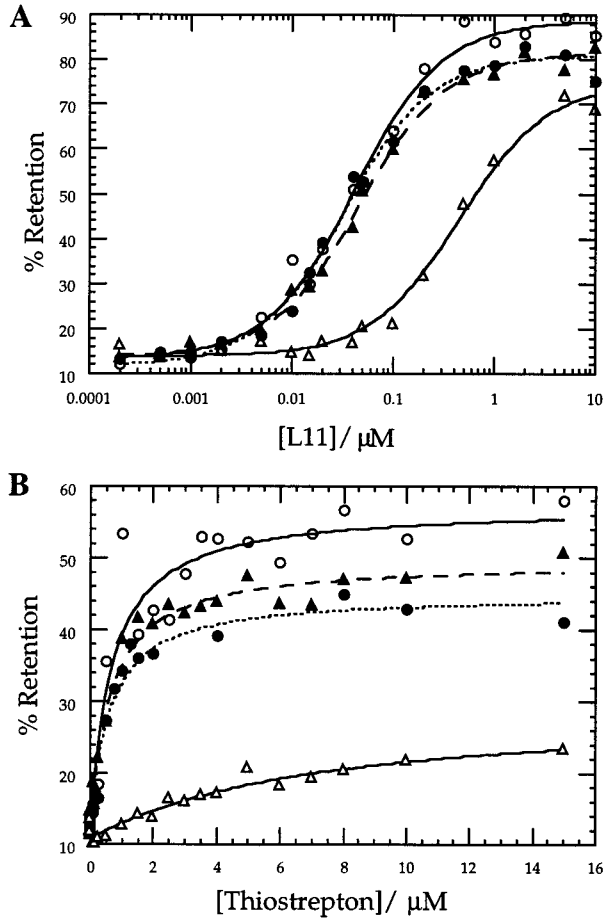


FIGURE 3: Titration of the (1092/1099)1072 RNA variants with (A) L11-C76 and (B) thiostrepton using a filter binding assay; (CG)C (○—○), (CG)U (△—△), (UA)C (▲- - -▲), and (UA)U (●···●).

30). Thus, the shift in pK_a in the folded form is consistent with the putative base triple being involved in the compact folding of the RNA.

The tertiary structure of the 58 nucleotide RNA is specifically recognized by ribosomal protein L11 and the antibiotic thiostrepton (*14*, *16*, *17*). Thus, the binding of these ligands provides a simple and functionally important assay of RNA tertiary structure. Typical binding curves for each (1092/1099)1072 RNA variant with L11-C76 and thiostrepton are shown in Figure 3, and values of the equilibrium association constants, determined from three or more assays, are given in Table 2. The order of the association constants is in agreement with the findings of the UV melting studies. For both ligands, the binding of (CG)U RNA, which showed no tertiary structure transition in the melting profiles, is approximately 10-fold lower than for the other RNA variants. The effect of pH on the binding of (UA)C RNA with thiostrepton was also examined as the tertiary structure is more stable at low pH according to the melting profiles. No difference was found in the association constant with thiostrepton when determined in pH 5.5 M_3A_{175} buffer. This is, however, not too surprising as it is clear from the melting profile that, even at the higher pH, the tertiary structure T_m is well above the temperature used in the filter binding assays.

The UV melting and filter binding experiments clearly demonstrate the presence of a base triple between positions

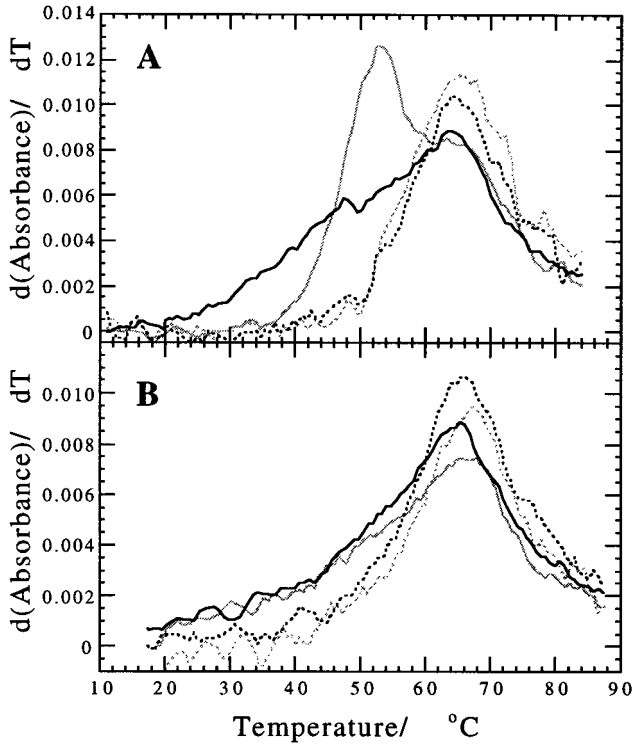


FIGURE 4: UV melting profiles of (A) (GC)C RNA and (B) (GC)U RNA at pH 7.5 (black) and pH 5.5 (gray) in M_5A_{100} buffer at 260 (solid line) and 280 nm (dashed line).

(1092/1099) and 1072 in the 58 nucleotide rRNA domain, as suggested by our correlation analysis. The interaction of these three bases is crucial for the formation of a functional folded tertiary structure and thus for the interaction of L11 and thiostrepton with this rRNA domain. A model of the base triple interaction based on the melting profiles of the (1092/1099)1072 RNA variants is given under Discussion.

To further explore the nature of the base triple interaction, two additional RNA variants were prepared. Here the (1092/1099) CG base pair was changed to GC, in the context of both C1072 and U1072 [(GC)C and (GC)U RNAs, respectively]. The melting profile of (GC)C RNA has a tertiary structure transition most stable at low pH (apparent T_m maximum at pH 5.5; Figure 4A). The melting profile is qualitatively similar to that of the other RNA with a pH-dependent triple, (UA)C RNA, with an apparent T_m of 52.3 °C (compared to 51.5 °C). However, at higher pH the melting profile is very unusual with no clear tertiary structure transition, although there is still a significant difference in hyperchromicity between 260 and 280 nm. The binding of both L11-C76 and thiostrepton to (GC)C RNA is identical to the wild-type (CG)C triple at pH 5.5 (Table 2). At higher pH, the L11 binding is slightly reduced, consistent with the change observed in the melting profile [no pH dependence on binding was found for either (CG)C or the (UA)C mutant]. The (GC)U RNA melting profiles are essentially identical to the (CG)U RNA with no observable tertiary structure transition (Figure 4B). However, the binding of L11-C76 and thiostrepton are not reduced as significantly from the wild type, with association constants 3−4-fold higher than (CG)U RNA. The melting profiles and binding data for these RNA variants appear consistent with the introduction of an equilibrium between the correct secondary and tertiary structure and an alternative misfolded form (see Discussion).

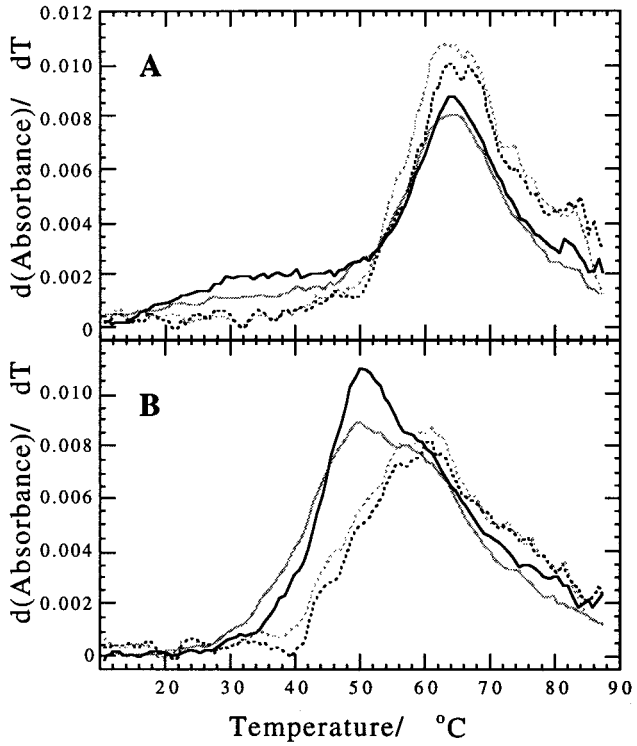


FIGURE 5: UV melting profiles of (1065/1073) RNA variants (A) AC(CG)C RNA and (B) AC(UA)U RNA at pH 7.5 (black) and pH 5.5 (gray) in M_5A_{100} buffer at 260 (solid line) and 280 nm (dashed line).

Effect of Positions 1065/1073. RNA sequence variants were also prepared to examine the role of the 1065/1073 covariation. The mutation UA1065/1073 to AC (i.e., YA to AY) was made in the context of the (CG)C and (UA)U triples at positions (1092/1099) and 1072 (Figure 1B). Melting profiles for these RNAs at pH 5.5 and pH 7.5 are shown in Figure 5, and the apparent T_m of tertiary structure unfolding is noted in Table 2.

It is clear that the AC1065/1073 mutation in the context of the (CG)C triple has a dramatic effect on RNA folding [Figure 5A, AC(CG)C RNA]. The melting profile shows the tertiary structure unfolding as a broad transition centered around 35 °C (a destabilization in T_m of approximately 25 °C). When the UA1065/1073 to AC mutation is made in the context of the Eucaryote-favored (UA)U triple, a much less dramatic change in the melting profile is observed [Figure 5B, AC(UA)U RNA]. However, the effect is destabilizing once again with a broadening of the tertiary structure transition and a reduction in T_m of around 7 °C [compared to (UA)U RNA at pH 7.5]. Both these RNAs show a pattern of dependence on pH similar to the (CG)C and (UA)U RNAs with only a small reduction in hyperchromicity as the pH is lowered to 5.5 and little (if any) effect upon T_m.

The interpretation of the melting profiles for the RNAs with mutations at 1065/1073 in structural terms is less clear than for the (1092/1099)1072 variants. However, it is probable that correct formation of the (1092/1099)1072 triple is dependent on the neighboring bases, as observed in the base triples in tRNA and group I introns (*11*), and so these positions covary with those directly involved in the triple.

DISCUSSION

Comparative sequence analysis has effectively predicted the correct secondary structure for several RNA molecules (*31*). More recently, comparative sequence analysis has been used to predict RNA tertiary structures, and base triples in particular, in large functional RNA molecules (*8*, *11*, *32*). In this study, we have used UV melting and binding assays to demonstrate the presence of a base triple in the tertiary structure of the L11 binding region of large subunit ribosomal RNA. The structural interaction of the loop base C1072 with base pair CG1092/1099 to form a base triple was strongly suggested by comparative analysis of over 550 sequences of this domain. UV melting analysis and binding assays with L11-C76 and thiostrepton clearly supported such an interaction. Single-nucleotide substitution at position 1072 disrupted the tertiary structure, eliminating it from the melting profile, and reduced binding of both ligands 10-fold. The alternative single mutation, of the base pair, did not entirely disrupt the tertiary structure but instead introduced a strong pH dependence on its stability. The fully compensatory mutation, of the base and base pair, restored the tertiary structure to wild-type stability and affinity for ligands. This result represents the first example of a base triple involved in the formation of a ribosomal RNA tertiary domain to be proposed by comparative sequence analysis and experimentally substantiated using site-directed mutagenesis. It is also noteworthy that while L11 and thiostrepton recognize the overall RNA conformation produced by the base triple interaction, their binding is not dependent on the actual identity of the bases within the base triple.

RNA Folding and Interactions with Ions and Ligands. The experimental confirmation of a base triple interaction between the two stem−loops places the first major RNA−RNA contact constraint on the tertiary folding of this domain. The two stem−loop structures must be folded onto each other so that the helices can interact and the hairpin loops come into close proximity. To the best of our knowledge, the joining of two hairpins near the base of each loop represents a novel mode of RNA−RNA interaction. This compact tertiary fold is consistent with studies of ligand interactions: Mutations in either loop, at A1067 and A1095 (*18*), or methylation at A1067 (*33*) can confer resistance to thiostrepton. These observations and the constraints imposed by the base triple interaction provide compelling evidence that thiostrepton acts by binding to both loops simultaneously, stabilizing the RNA and locking the tertiary conformation. The structure of L11 is very similar to the homeodomain proteins, and many of the residues that contact the RNA are on one side of a single α-helix (*34*, *35*). Thus, it is likely that L11 binds primarily through contacts in one of the grooves on the exposed surface of the folded RNA where bulged or mismatched regions could alter groove widths and shapes. For example, the helical region centered around 1060 has two consecutive noncanonical base pairs, contains several positions essential for tertiary folding, and is protected from hydroxyl radicals on binding L11 (*36*). Finally, the folding of this RNA into such a compact structure requires a number of cations to counter the large electrostatic energy barrier: the presence of specific sites for interaction of two magnesium ions (*27*) and at least one ammonium ion (*19*) within this rRNA domain has been demonstrated.

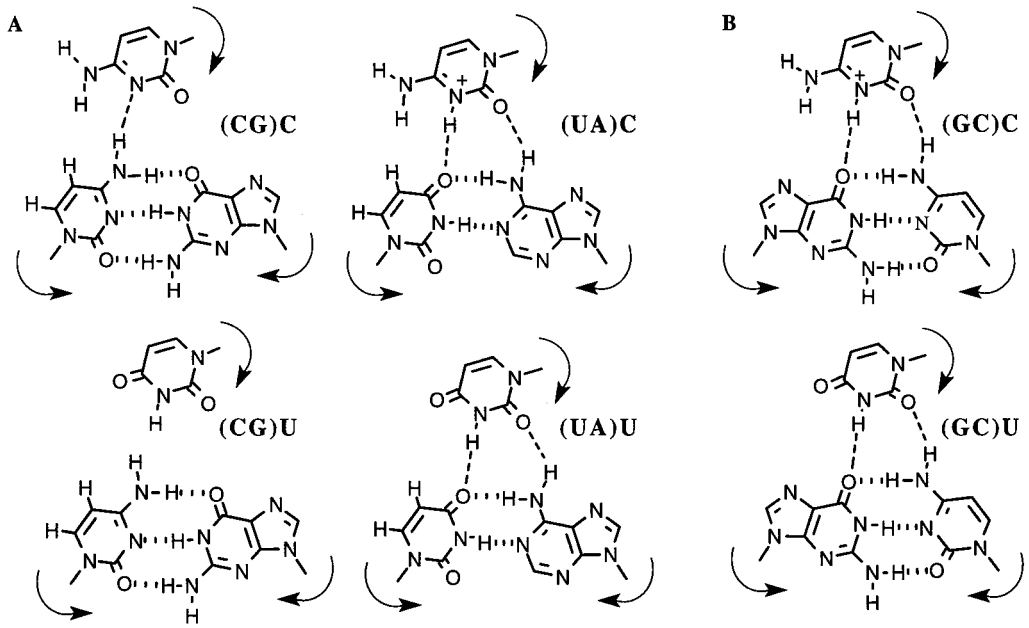


FIGURE 6: (A) Possible base triples consistent with the UV melting and filter binding data. (B) Triples predicted by the model for the two additional variants (GC)C RNA and (GC)U RNA. Hydrogen bonds represented by hatched lines are standard Watson−Crick; dashed lines show possible tertiary hydrogen bonding interactions (see main text). Arrows are used to indicate the polarity of the phosphodiester backbone.

A model of the tertiary structure of the 58 nucleotide RNA has been suggested on the basis of earlier phylogenetic analysis and chemical and enzymatic probing (*37*). The tertiary fold established by the (1092/1099)1072 base triple is generally analogous to this earlier model. However, the base triple now imposes specific constraints on the relative positions of the nucleotides flanking the base triple positions of the distant stem−loops.

A Model of the Base Triple Interaction. The results of the UV melting studies present a number of clues as to the possible nature of the (1092/1099)1072 base triple interaction. Clearly, neither the wild-type (CG)C triple nor the compensatory mutant (UA)U contains protonated species. As a result, the base triple in this RNA cannot involve the usual Hoogsteen interaction, (CG)C^+ triple, found in RNA (or DNA) triple helices (*38*, *39*). The clearest evidence for a contact in the base triple comes from the stabilization of the (UA)C triple at low pH. This result strongly suggests the pyrimidine N3 functionality is the most significant hydrogen bonding interaction as this is the most likely site of protonation. Protonation of the adenine N1 in a (UA^+)C triple cannot be entirely ruled out; however, such a triple would involve non-Watson−Crick base pairing in addition to the tertiary interactions. Furthermore, it is not possible to draw isosteric triples for the other RNAs that can account for both the lack of pH dependence in (CG)C and the formation of a stable tertiary structure for (UA)U but not for (CG)U RNA. The base triples shown in Figure 6A provide the most simple arrangement of a nonprotonated (CG)C triple which allows isosteric arrangements for both (UA)C^+ and (UA)U but not (CG)U. The interaction shown occurs in the major groove with the strands arranged so that the incoming third strand runs antiparallel to that containing C1092 (as suggested by the RNA secondary structure of Figure 1A). For each base triple, except (CG)U, a hydrogen bonding interaction involving the pyrimidine N3 position (dashed line) is possible. An additional interaction of the pyrimidine C2 carbonyl and the base pair adenosine C6 amino group is also possible for (UA)C and (UA)U RNAs.

While we currently favor the model of Figure 6A on the basis of the present data, we note that a similar set of base triples can be drawn that offer additional interactions, in particular for the most stable (CG)C RNA. If the third strand is flipped around the 1072 pyrimidine N3 hydrogen bond, a hydrogen bonding interaction between the C2 carbonyl and the base pair pyrimidine C5 becomes possible for each base triple except (CG)U. Such interactions have been observed in tRNA, where they are involved in stabilizing the anticodon loop (*40*). Alternatively, in this orientation other hydrogen bonds involving the cytosine C4 amino of (CG)C RNA or the uracil C4 carbonyl of (UA)U RNA with the purine N6 or O6 would be possible. However, such interactions would be weak due to the orientation of the amino groups and carbonyl oxygen electron lone pairs. The major drawback of this strand arrangement is that it requires the incoming pyrimidine to adopt a syn conformation about the glycosylic bond or must involve a local reversal of backbone orientation. However, two recent structures show that both these situations can arise in natural RNA hairpins (*41*, *42*). In the latter case, a local reversal of backbone orientation in a ribosomal hairpin loop (*42*), the similarities with the present situation are intriguing. The sites of backbone reversal and the adjacent noncanonical base pair are highly conserved, indicating an important structural role, and the three purines of the loop are oriented such that the bases are turned out of the loop (allowing for tertiary interactions). C1072 also lies adjacent to a structurally important noncanonical base pair (1065/1073), and the sugars of the loop purines 1068−1070 are protected from OH radical cleavage by the tertiary fold of the RNA (G.L.C., unpublished observations).

Two further RNA sequence variants with a reversed 1092/1099 base pair (CG to GC) were made to test the base triple model. The model predicts that with C1072, (GC)C RNA, the triple should form preferentially at low pH (again through

protonation at N3 of the single-stranded cytosine) whereas in the context of U1072, (GC)U RNA, a triple with wild-type stability should be possible (Figure 6B). While the melting profiles of (GC)C RNA did indicate pH-dependent folding of the tertiary structure as predicted (with a maximum apparent T_m at pH 5.5), no tertiary structure transition was observed at any pH for (GC)U RNA. A number of possible reasons for this inconsistency are apparent. First, it may be that the base triples in the different RNAs are not isosteric. The base triple model assumes that (1092/1099) is at least close to a standard Watson−Crick base pair conformation for each RNA. This is the case for the equivalent (CG) base pair in the NMR structure of a hexaloop model of this region of the RNA (*43*). In contrast, however, the equivalent (UA) positions in the NMR structure of the similar stem−loop IIa from yeast U2 RNA were not base paired (*44*). Thus, considerably different structures could be involved in the formation of the triples containing (CG) at 1092/1099 and those containing (UA). A second possibility is that GC1092/1099 forms less favorable stacking interactions with adjacent nucleotides than CG1092/1099. For instance, reversing the adjacent conserved G-A (1093/1098) sheared base pair destabilizes a hairpin containing the hexanucleotide loop by 0.35 kcal/mol (*45*), and it is clear from RNAs with mutations next to 1072 that the structural context of this position is important. Thus, a considerably less stable structure would result, but one in which the correct tertiary hydrogen bonding interactions and therefore ligand binding are possible.

A third possibility for the lack of observable tertiary structure in (GC)U RNA, which we currently favor, is that the reversal of the 1092/1099 base pair allows the formation of an alternative secondary and possibly tertiary structure(s) of similar free energy in equilibrium with the properly folded RNA. This offers a plausible explanation for the unusual melting profile observed for (GC)C RNA at higher pH and the association constants for each variant with L11 and thiostrepton. For (GC)C RNA, there is a dependence of binding upon solution pH, which correlates with the observation of tertiary structure unfolding in the melting profiles. The association constants for (GC)U are unaffected by pH but are significantly higher for both ligands than for (CG)U, even though no tertiary structure is observable in the melting profile. These observations suggest that there is an equilibrium of structures which can be forced in favor of the correct structure by protonation in the case of (GC)C RNA, or by the binding of L11-C76 or thiostrepton for both variants. Previous studies of this rRNA fragment have indicated that it readily adopts competing structures that decrease the apparent stability of the tertiary structure, and that the equilibrium between such structures may be quite sensitive to single nucleotide substitution (*26*). We are currently using a range of chemical and enzymatic probes to assay the secondary and tertiary structures of all these RNAs and determine more fully the exact nature of the triple interaction and its role in the RNA tertiary fold.

ACKNOWLEDGMENT

We thank Dr. Luis Reynaldo for help in the preparation of L11-C76, and Dr. Tom Gluick for many useful discussions and a critical appraisal of the manuscript.

REFERENCES

1. Quigly, G. J., and Rich, A. (1976) *Science 194*, 796−806.
2. Sussman, J. L., and Kim, S.-H. (1976) *Science 176*, 853−858.
3. Green, R., and Szostak, J. W. (1994) *J. Mol. Biol. 235*, 140−155.
4. Jaeger, L., Michel, F., and Westhof, E. (1994) *J. Mol. Biol. 236*, 1271−1276.
5. Murphy, F. L., and Cech, T. R. (1994) *J. Mol. Biol. 236*, 49−63.
6. Tanner, M. A., and Cech, T. R. (1997) *Science 275*, 847−849.
7. Brown, J. W., Nolan, J. M., Haas, E. S., Rubio, M. A. T., Major, F., and Pace, N. R. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 3001−3006.
8. Gutell, R. R. (1996) in *Ribosomal RNA, Structure, Evolution, Processing and Function in Protein Biosynthesis* (Zimmerman, R. A., and Dahlberg, A. E., Eds.) pp 111−128, CRC Press, Boca Raton, FL.
9. Kalurachchi, K., Uma, K., Zimmermann, R. A., and Nikonowicz, E. P. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 2139−2144.
10. Tao, J., Chen, L., and Frankel, A. D. (1997) *Biochemistry 36*, 3491−3495.
11. Gautheret, D., Damberger, S. H., and Gutell, R. R. (1995) *J. Mol. Biol. 248*, 27−43.
12. Noller, H. F., Kop, J., Wheaton, V., Brosius, J., Gutell, R. R., Kopylov, A. M., Dohme, F., Herr, W., Stahl, D. A., Gupta, R., and Woese, C. R. (1981) *Nucleic Acids Res. 9*, 6167−6189.
13. Gutell, R. R., Gray, M. W., and Schnare, M. N. (1993) *Nucleic Acids Res. 21*, 3055−3074 [Database issue].
14. Stark, M. R. J., Cundliffe, E., Dijk, J., and Stöffler, G. (1980) *Mol. Gen. Genet. 180*, 11−15.
15. Xing, Y., and Draper, D. E. (1995) *J. Mol. Biol. 249*, 319−331.
16. Egebjerg, J., Douthwaite, S. R., and Garrett, R. A. (1989) *EMBO J. 8*, 607−611.
17. Lu, M., and Draper, D. E. (1995) *Nucleic Acids Res. 23*, 3426−3433.
18. Rosendahl, G., and Douthwaite, S. (1994) *Nucleic Acids Res. 22*, 357−363.
19. Wang, Y. X., Lu, M., and Draper, D. E. (1993) *Biochemistry 32*, 12279−12282.
20. Laing, L. G., Gluick, T. C., and Draper, D. E. (1994) *J. Mol. Biol. 237*, 577−587.
21. Laing, L. G., and Draper, D. E. (1994) *J. Mol. Biol. 237*, 560−576.
22. Deng, W. P., and Nickoloff, J. A. (1992) *Anal. Biochem. 200*, 81−88.
23. Gurevich, V. V. (1996) *Methods Enzymol. 275*, 382−397.
24. Xing, Y., and Draper, D. E. (1996) *Biochemistry 35*, 1581−1588.
25. Ryan, P. C., and Draper, D. E. (1989) *Biochemistry 28*, 9949−9956.
26. Lu, M., and Draper, D. E. (1994) *J. Mol. Biol. 244*, 572−585.
27. Bukhman, Y. V., and Draper, D. E. (1997) *J. Mol. Biol. 273*, 1020−1031.
28. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, pp 107−110, Springer-Verlag, New York.
29. Yang, A.-S., and Honig, B. (1994) *J. Mol. Biol. 237*, 602−614.
30. Oliveberg, M., and Fersht, A. R. (1996) *Biochemistry 35*, 6795−6805.
31. Gutell, R. R. (1993) *Curr. Opin. Struct. Biol. 3*, 313−322.
32. Michel, F., and Westhof, E. (1990) *J. Mol. Biol. 216*, 585−610.
33. Thompson, J., Schmidt, F., and Cundliffe, E. (1982) *J. Biol. Chem. 257*, 7915−7917.
34. Xing, Y., GuhaThakurta, D., and Draper, D. E. (1997) *Nat. Struct. Biol. 4*, 24−27.

35. Hinck, A. P., Markus, M. A., and Torchia, D. A. (1997) *J. Mol. Biol. 274*, 101−113.
36. Rosendahl, G., and Douthwaite, S. (1993) *J. Mol. Biol. 234*, 1013−1020.
37. Egebjerg, J., Douthwaite, S. R., Liljas, A., and Garrett, R. A. (1990) *J. Mol. Biol. 213*, 275−288.
38. Klinck, R., Liquier, J., Taillandier, E., Gouyette, C., Huynh-Dinh, T., and Guittet, E. (1995) *Eur. J. Biochem. 233*, 544−553.
39. Holland, J. A., and Hoffman, D. W. (1996) *Nucleic Acids Res. 24*, 2841−2848.
40. Auffinger, P., Louise-May, S., and Westhof, E. (1996) *J. Am. Chem. Soc. 118*, 1181−1189.
41. Lynch, S. R., and Tinoco, I., Jr. (1998) *Nucleic Acids Res. 26*, 980−987.
42. Puglisi, E. V., Green, R., Noller, H. F., and Puglisi, J. D. (1997) *Nat. Struct. Biol. 4*, 775−778.
43. Huang, S., Wang, X.-Y., and Draper, D. E. (1996) *J. Mol. Biol. 258*, 308−321.
44. Stallings, S. C., and Moore, P. B. (1997) *Structure 5*, 1173−1185.
45. Serra, M. J., Axenson, T. J., and Turner, D. H. (1994) *Biochemistry 33*, 14289−144296.

BI980825+