In vitro selection of RNAs with increased tertiary structure stability

KARA JUNEAU and THOMAS R. CECHE

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

ABSTRACT

An in vitro selection system was devised to select RNAs based on their tertiary structural stability, independent of RNA activity. Selection studies were conducted on the P4–P6 domain from the *Tetrahymena thermophila* group I intron, an autonomous self-folding unit that contains several important tertiary folding motifs including the tetraloop receptor and the A-rich bulge. Partially randomized P4–P6 molecules were selected based on their ability to fold into compact structures using native gel electrophoresis in the presence of decreasing concentrations of MgCl_2 . After 10 rounds of the selection process, a number of sequence alterations were identified that stabilized the P4–P6 RNA. One of these, a single base deletion of C209 within the P4 helix, significantly stabilized the P4–P6 molecule and would not have been identified by an activity-based selection because of its essential role for ribozyme function. Additionally, the sequence analysis provided evidence that stabilization of secondary structure may contribute to overall tertiary stability for RNAs. This system for probing RNA structure irrespective of RNA activity allows analysis of RNA structure/function relationships by identifying nucleotides or motifs important for folding and then comparing them with RNA sequences required for function.

Keywords: magnesium ion; nondenaturing gel electrophoresis; ribozyme; RNA folding

INTRODUCTION

In vitro selection is a powerful method used for creating and identifying molecular interactions. Such methods have been used to study binding of RNA to proteins (Tuerk & Gold, 1990; Bartel et al., 1991; Doudna et al., 1995), to other nucleic acids (Campbell & Cech, 1995), and to small molecules (Ellington & Szostak, 1990; Jenison et al., 1994). Selection strategies have also been useful for creating new RNA catalysts (Pan & Uhlenbeck, 1992; Bartel & Szostak, 1993; Illangasekare et al., 1995; Zhang & Cech, 1997), altering the activities of existing catalysts (Beaudry & Joyce, 1992; Lehman & Joyce, 1993; Jabri & Cech, 1998), analyzing nucleic acid secondary structure stability (Bevilacqua & Bevilacqua, 1998), and studying the structural and folding interactions required for catalytic RNA activity (Green & Szostak, 1994; Costa & Michel, 1997; Treiber et al., 1998). Studies on catalytic RNAs have relied on activity as a major component of the selection process. Here

we describe in vitro selection of RNA based solely on molecular compactness, independent of RNA function. Selections based only on folding stability may facilitate identification of structural interactions and improve our understanding of structure/function relationships for RNA molecules.

The P4–P6 domain of the *Tetrahymena thermophila* group I intron provides an optimal system for structure probing studies. It is a large self-folding RNA molecule (Murphy & Cech, 1993; 1994) where MgCl_2 -dependent stability of folded wild-type or mutant molecules can be monitored by native gel electrophoresis (Murphy et al., 1994). Native P4–P6 molecules and some site-specific mutants are thought to be in a two-state equilibrium between folded and unfolded forms; the more stable RNA molecules spend a greater percentage of time in the folded state and, thus, achieve a smaller effective radius and exhibit more rapid native gel mobility (Szewczak & Cech, 1997). Additional biochemical analysis of P4–P6 RNA revealed structural motifs such as the A-rich bulge and the tetraloop receptor, which partake in important tertiary interactions (Murphy & Cech, 1994; Costa & Michel, 1995). These and other interactions were ultimately seen at 2.8 Å resolution by X-ray crystallography (Cate et al., 1996).

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

The 11-nt tetraloop receptor is a well characterized motif ideal for testing a system designed to study RNA structure. The receptor has identifiable sequence requirements (Costa & Michel, 1995) and a well understood tertiary structure when combined with its appropriate GNRA tetraloop (Pley et al., 1994; Cate et al., 1996; Butcher et al., 1997). In addition, the tetraloop receptor sequence requirements have been investigated in activity-based selection experiments (Costa & Michel, 1997), so comparisons can be made between functional requirements and structural requirements.

A system to study individual RNA tertiary interactions from a completely structural standpoint was developed by combining in vitro selection with native gel electrophoresis of the P4–P6 molecule. Sequences that stabilize the folded P4–P6 domain of *Tetrahymena thermophila* were identified. We expect this to be a generalizable system for analyzing structural components of RNA folding.

RESULTS

Creating a system to select for RNA folding and stability

The tetraloop-receptor interaction in the P4–P6 domain of *Tetrahymena thermophila* was used to test the RNA-folding selection system. The RNA was randomized at positions C222–A231 and U243–G251 (Fig. 1, randomized region), which include the tetraloop receptor (CCUAAG..UAUGG) plus an additional 4 bp in P6b; the resulting library had a complexity of 10^{11} different sequences. Three different tetraloops and their optimal closing base pairs were chosen for the study based on their frequency of occurrence in structured RNAs (Woese et al., 1990), stability (Antao et al., 1991; Santa-Lucia et al., 1992), and potential for making tertiary interactions in large RNAs. The G(GAAA)C tetraloop¹ at the bottom of P5b was either maintained as a control for the native P4–P6 molecule (Fig. 1), substituted by C(GCAA)G to test for interactions specific to GNRA loops (Jucker et al., 1996), or substituted by G(CUUG)C (Jucker & Pardi, 1995) to study a potentially completely different folding interaction. A structurally inconsequential mutation at the bottom of P6b was made that changes the native pentaloop G(AUCUU)C to G(GAGA)C. This mutation destroys an existing *Bgl*II site, while the folding stability of the RNA is not altered (Golden et al., 1998). The library was constructed using PCR mutagenesis (for details see Materials and Methods) and was run-off transcribed by T7 RNA polymerase for use in the selection procedure.

The selection protocol (Fig. 2) included two selection steps. The first selection was designed to maintain the

P4-P6 Domain

FIGURE 1. The secondary structure of the P4–P6 system used in the selection. The thick line drawing indicates the P4–P6 domain as a highlighted part of the entire *Tetrahymena thermophila* group I intron. Only the P4–P6 domain was used in the selection process. The region randomized contains the tetraloop receptor (CCUAAG..UAUGG), shown in bold, and an adjacent portion of P6b, identified by individual nucleotides. L6b is changed from the wild-type AUCUU loop to a GAGA loop, which creates a structurally inconsequential mutation eliminating a *Bgl*II site (Golden et al., 1998). The wild-type G(GAAA)C loop is pictured at the base of P5b; this loop as well as the loops C(GCAA)G and G(CUUG)C were used in the selection. The double-headed arrow highlights the tetraloop-receptor tertiary interaction, and the 5' splice site is denoted by 5'ss.

nucleic acid length by purification on a denaturing gel; this selection ensured that fast-migrating species isolated from the subsequent native gels were not simply deletion mutants. The RNAs were then separated on native gels and only the RNAs with gel mobility similar to that of wild-type P4–P6 RNA were extracted for amplification by RT-PCR. In the initial pool such RNAs were too rare to be visualized directly (Fig. 3A, generation 0), but nevertheless could be amplified and re-transcribed for the next round.

Ten rounds of selection were carried out, during which the MgCl_2 concentration was sequentially decreased from 10 mM to 1.1 mM to increase selective pressure on the RNAs. After five rounds of selection, the RNA population was completely folded at 5 mM MgCl_2

¹ In this notation, the parentheses flank the loop sequence and the outside nucleotides identify the closing base pair.

FIGURE 2. Flow diagram of selection protocol.

(Fig. 3A). During the subsequent five rounds of selection, the MgCl_2 concentration available for folding the RNAs was decreased to 1.1 mM. The last generation of the RNAs was reverse transcribed and cloned into pUC19, and clonal RNAs were tested individually on 1.1 mM MgCl_2 native gels (Fig. 3B) to ensure competent folding before sequence analysis. Approximately 75% of the RNAs folded as compactly as wild-type P4–P6 RNA. The remaining RNAs had a high degree of variability in gel mobility; some clones migrated faster than wild-type molecules, some slower, and others maintained two or more populations of molecules, each population with a reproducible mobility on native gels. These latter clones were not thoroughly analyzed.

Tetraloop dependence of the selected RNAs

The primary sequences were analyzed, trends were observed, and a few short recurrent sequence motifs were identified. Some selected sequences are shown in Figure 4. The 5' side of the randomized region appears to be pyrimidine rich, with some sequences being entirely composed of cytosine and uracil (example: clone GCAA 69, Fig. 4). Some of the clonal RNAs can be grouped such that, on the 3' side of the randomized regions, 3–5-nt motifs can be identified. The UUGC motif appeared in all three tetraloop selection pools, whereas UUUGA was only prominent in the GAAA selection pool. Additionally, the GNRA loop pool had many sequences that contained the sequence GAA, frequently identified in a larger CAGAA motif. These identified motifs could facilitate RNA folding by forming constructive interactions with the L5b tetraloop, avoid-

ing destructive interactions, or improving folding independent of the tetraloop.

It was necessary to elucidate the relationship between the tetraloop and its potential receptor. The dependence of folding on the tetraloop was tested by substituting a structurally different tetraloop and comparing the folding ability as measured by gel mobility (Fig. 5). The G(CUUG)C loops were substituted with G(GAAA)C tetraloops, whereas the GNRA loops were replaced with G(CUUG)C. The RNAs were run on a 1.1 mM MgCl_2 native gel and the relative mobilities were analyzed. In general, the switching of the tetraloops destabilized the RNAs as anticipated, but the degree of destabilization varied dramatically. The clones from the G(CUUG)C pool exhibited gel mobilities that had little dependence on the identity of the tetraloop, whereas the clones from the GNRA loop pools had either lost some stabilizing interactions or acquired destabilizing interactions upon mutation of their tetraloops. The wild-type P4–P6 folding was disrupted upon tetraloop substitution to a much greater extent than any of the selected RNAs tested. Thus, the wild-type P4–P6 molecule is more dependent upon the native tetraloop-receptor interaction than any of the tested selection clones are upon their potential tetraloop receptors. The data suggest that the selected molecules possess sequences that interact with a specific tetraloop, but there must exist another identifiable region within the RNA that provides additional structural stability.

Mutations in the nonrandomized region of the RNAs contribute to stability

Mutations in the nonrandomized region of P4–P6 arose because of the infidelity of both Taq polymerase and reverse transcriptase. Common mutations in the selected RNAs are shown in Figure 6. Two major families of mutations were identified as candidates for stabilizing tertiary structure. One group of mutations, localized to the P5b stem, changed G•U wobble base pairs into Watson–Crick pairs. Note that these changes would clearly stabilize the secondary structure, but it is less clear how they would affect tertiary folding. A second distinct set of mutations involved deletions of 1–6 nt within the P4 and P6 regions of the molecule.

To test the importance of these nonrandomized region mutations, site-directed mutational analysis was conducted on several selected RNAs and disruptive effects on folding were analyzed on native gels (Fig. 7). Two sets of constructs were made; one set tested the importance of the Watson–Crick pairing in P5b by reverting all the nonnative Watson–Crick pairing back to G•U wobbles (Fig. 7, lanes A), and the other set changed all other opportunistic mutations, which arose during the selection process, back to native P4–P6 sequence (Fig. 7, lanes C). The folding abilities of these two sets of constructs were compared with the original selected

FIGURE 3. Selection progression monitored by native gel electrophoresis. **A:** 5 mM MgCl₂ native gel for the first five generations of each tetraloop library. The unfolded control consists of a P4–P6 molecule with J5/5a base paired (BP), which prevents the RNA from bending at its top hinge region (Murphy & Cech, 1993). The folded control is wild-type P4–P6 (WT). **B:** RNAs from generation 10 were reverse transcribed, cloned, transcribed, and tested by electrophoresis on 1.1 mM MgCl₂ native gels for correct gel mobility prior to sequencing.

	5'-randomized	3'-randomized
WT_P456	<u>AACCACGCAGCCAAGU</u> -CCUAAGUCA-- <u>CAGGAGACUG</u> ---	UUGAUUUGG--- <u>AUGCAGUI</u>
CUUG53	<u>AACCACGC</u> CA--- <u>AAGU</u> -UCCUGGUUCU-- <u>CAGG</u> -GACUG-UCGUC-UUGC----	<u>AUGCAGUI</u>
CUUG79	<u>AACC</u> --- <u>CAGCCAAGU</u> -UACCGUGAUC-- <u>CAGGAGACUG</u> -GA--U-UUGC-U--	<u>AUGCAGUI</u>
CUUG78	<u>A-CCACGCAGC</u> - <u>AAG</u> --CCUUGUAAUA-- <u>CAGGAGACUG</u> -UG--U-UUGC-AG-	<u>AUGCAGUI</u>
CUUG33	<u>AACC</u> ----- <u>GCCAAGU</u> -GCUACCUACG-- <u>CAGGAGACUG</u> -UGCCU-UUGC-A--	<u>AUGCAGUI</u>
CUUG44	<u>AACC</u> ----- <u>GCCAAGU</u> -CUCCGUUAG-- <u>CAGGAGACUG</u> ---	ACA-UUGC-CG- <u>AUGCAGUI</u>
	71% Pyrimidine	69% UUGC
GAAA64	<u>AACC-CG-AGCCAAGU</u> -GAU-GCCGCUU- <u>CAGGAAACUG</u> -GCA-A-UUGC-A--	<u>AUGCAGUI</u>
GAAA19	<u>AAC-ACGCAGCCAAGU</u> -GCUAAC-AACG- <u>CAGGAGACUG</u> -CCACA-UUGC----	<u>AUGCAGUI</u>
GAAA29	<u>AAC-ACGCAGCCAAGU</u> -CUCCA---AU-- <u>CAGGAGACUG</u> -CU---UUUGA-GC-	<u>AUGCAGUI</u>
GAAA39	<u>AAC-ACGCAGCCAAGU</u> -CUUCU-CCUCA- <u>CAGGAGACUG</u> -UC---UUUGA-G-	<u>AUGCAGUI</u>
GAAA79	<u>AAC-ACGCAGCCAAGU</u> -CUCCU---AC-- <u>CUGGAGACUG</u> -C----C-CGAA-CAA	<u>AUGCAGUI</u>
GAAA27	<u>AAC-ACGCAGCCAAGU</u> -CUUUUUGCCAA- <u>CAGGAGACU</u> --UAU--CACGAA-G-	<u>AUGCAGUI</u>
GAAA42	<u>AAC-ACGCAGCCAAGU</u> -CCUGGCCUC-- <u>CAGGAGAC</u> -G-UAA-GAGCCG----	<u>AUGCAGUI</u>
	66% Pyrimidine	25% UUGC 12% UUUGA 25% GAA
GCAA49	<u>AACC-CA-A-CCAAGUCCA</u> UCCUUGC-- <u>CAGGAAACUG</u> ----	GU-UUGC-GACA <u>AUGCAGUI</u>
GCAA60	<u>AACC-CGCAGC</u> -- <u>AGU</u> -CUAC-UUACUU- <u>CAGGAGACUG</u> -CUAC--UUGC-A--	<u>AUGCAGUI</u>
GCAA43	<u>AACC</u> --- <u>CAGCCAAGU</u> --CGAAUCAUUU- <u>CAGGAGACUG</u> --CU--UUGC-AGGA	<u>AUGCAGUI</u>
GCAA69	<u>A-CCACGCAGCCAAGU</u> -CUCCUUU-CUU- <u>CAGGAGACUG</u> -CCAA---UGAA-U-	<u>AUGCAGUI</u>
GCAA64	<u>A-CCACGCACCAAGU</u> -CUCCAUU-AUA- <u>CAGGAGACUG</u> -UA---CACGAA-G-	<u>AUGCAGUI</u>
GCAA74	<u>AAC-ACGCAGCCAAGU</u> -CUUCG-CAAAU- <u>CAGGAGACUG</u> -AC---CACGAA-U-	<u>AUGCAGUI</u>
	73% Pyrimidine	25% UUGC 43% GAA

FIGURE 4. Sequence analysis of a portion of the P4–P6 molecule (A206–A261) that includes the randomized region. Underlined sequences denote nonrandomized regions, plain text shows the randomized region, and dashes represent gaps incorporated in the figure to optimize alignment. The names of the RNAs are located to the far left with wild-type P4–P6 RNA at the top of the list. Potential motifs are identified in bold lettering at the base of each set of sequences, and the frequency at which the motifs appear is represented as a percent [(number of appearances)(100)/total RNAs sequenced].

FIGURE 5. Dependence of RNA folding on the sequence of the L5b tetraloop, analyzed by 1.1 mM MgCl₂ native gel electrophoresis. The folded (WT) and unfolded (BP) controls are the same as in Figure 3. The clones shown are named at the top. Tetraloop dependence was tested by comparing the mobility of selected RNAs to related RNAs in which the tetraloop was switched to a structurally very different tetraloop by mutagenesis (new tetraloop labeled in lowercase letters). Selected RNAs retaining their original tetraloop are symbolized with (•).

clones (Fig. 7, lanes O) and the tetraloop-switched constructs (Fig. 7, lanes B). It was observed that the Watson–Crick pairing in the P5b stem did not have a discernible effect on the tertiary folding stability of the constructs tested (Fig. 7, compare lanes O and A). As already demonstrated (Fig. 5), switching the tetraloops had only a small effect on folding, but what was interesting was that the nonrandomized region mutations, other than those identified in the P5b stem, made a significant contribution to folding stability (Fig. 7, compare lane B with lane C). From this experiment it became obvious that the largest contribution to the folding stability of RNAs from the selection could be attributed to a mutation, or set of mutations, located in the non-randomized region of the RNA molecules.

A mutation that stabilizes the P4–P6 structure is identified

Two specific mutations were tested for their ability to stabilize folding based upon the frequency of their occurrence in the sequence library: the deletion of C209 and a slightly larger deletion of C209–C211 (Fig. 6). A deletion of the cytosine at position 209 was the most frequent mutation found in the library; this deletion occurred 40 times out of 111 sequences (data not shown), which is a relatively high degree of mutation considering that it is located in the “fixed” sequence of the RNA pools. Many of the selected RNAs possessed much larger deletions in the P4 and P6 base pairing regions, so a larger deletion of C209–C211 was also tested. These two deletions were cloned into a P4–P6 folding mutant and assayed for their ability to restore folding. The mutant P4–P6 RNA used contains a tetraloop receptor that is base-paired (see Fig. 8D, rcptr-BP), so folding of the molecule is minimal at low MgCl₂

FIGURE 6. Mutations seen repeatedly in the nonrandomized regions of the RNAs. The P4–P6 RNA is represented as a line drawing. Mutations seen 4–15 times out of 111 sequences are denoted in outlined letters, and mutations seen 20–40 times are labeled with solid black letters. Δ indicates a deletion and Y represents a pyrimidine.

FIGURE 7. Comparison of the contributions to folding of the tetraloop, mutations in P5b, and mutations in the remaining nonrandomized P4–P6 sequence, analyzed on a 1.1 mM MgCl₂ native gel. Lanes O: the original selected RNAs; lanes A: Watson–Crick base pairs in P5b that arose during *in vitro* selection were switched back to native G•U wobbles; lanes B: selected RNAs after their tetraloops were switched to a different loop sequence; lanes C: all fixed region mutations reverted to wild-type P4–P6 sequence. The folded (WT) and unfolded (BP) controls are identical to those used in Figure 3.

concentrations (Murphy & Cech, 1994). Native gel analysis showed that the deletion of C209 stabilized the mutant P4–P6 RNA (Fig. 8A, compare lanes 1 and 2), whereas the larger C209–C211 deletion did not confer additional stability to the P4–P6 RNA containing the base-paired receptor (Fig. 8A, compare lanes 1 and 3).

The deletion of C209 was tested in the context of several other forms of P4–P6 (Fig. 8B) including wild-type RNA, two additional receptor mutants, and the A186U mutant, which has a disruption in its A-rich bulge that is deleterious to folding (Murphy & Cech, 1994). Δ C209 was successful in stabilizing all of the mutant P4–P6 molecules to some degree. Additionally, under very destabilizing conditions (0.5 mM MgCl₂), Δ C209 was shown to confer stability to wild-type P4–P6 RNA as well (Fig. 8C). These data suggest that the stabilization afforded by the C209 deletion is not specific to tetraloop–receptor disruptions; instead, Δ C209 has a more general stabilizing effect on the RNA. In addition, the increased stability conferred to the wild-type P4–P6 molecule suggests that although Δ C209 can stabilize globally, it is not necessarily changing the global structure of the RNA. The stability afforded by Δ C209 appears to be in addition to, not instead of, the existing tertiary interactions that fold the P4–P6 RNA. However the possibility of a refolded secondary or tertiary structure could not be ruled out without conducting structure-probing experiments.

The structure of Δ C209 is indeed similar to wild-type P4–P6

DMS (Fig. 9) and CMCT² modification reactions (data not shown) were carried out on wild-type and Δ C209 P4–P6 RNAs. Qualitative analysis showed that these RNAs have a very similar chemical footprint, suggest-

ing that the secondary structure of the Δ C209 mutant is not different than wild-type P4–P6. One observed difference between the two RNAs was that A210 was more protected from modification in the Δ C209 construct than in the wild-type construct. Such an increase in DMS protection suggests that the normally bulged out A210 (Cate et al., 1996) is now folding into the P4–P6 molecule, creating new stabilizing interactions or removing existing deleterious interactions.

Ribozyme activity of Δ C209 mutants

The effect of the Δ C209 mutation on ribozyme cleavage activity was tested in the context of both wild-type P4–P6 and P4–P6 RNA with a base-paired receptor (rcpt-BP in Fig. 8D). The three-piece system (Doudna & Cech, 1995) was used to expedite kinetic studies; it allows analysis of both the association of the intron domains and the rate of ribozyme cleavage. In the three-piece system, the P4–P6 molecules tested were first allowed to associate with the two other components (P1–P3 and P3–P9) that comprise the rest of the ribozyme. Cleavage activity was assayed by quantitating disappearance of the ³²P-labeled P1–P3, which contains the 5' splice site analog. Kinetic analysis revealed a 100-fold decrease in activity for C209 deletions (Fig. 10). The half-lives of the wild-type ribozyme and the Δ C209–WT mutant were 1.6 min and 150 min, respectively. Similar half-lives were measured for the base-paired receptor construct (Fig. 10, rcpt-BP) and its C209-deleted counterpart, 2.9 min and 270 min, respectively. Because the deletion of C209 could affect association of the P4–P6 domain with P1–P3 and P3–P9, thus affecting cleavage rates, several concentrations of P4–P6 were tested (25 nM, 100 nM, 500 nM; P1–P3 and P3–P9 had concentrations of 1 nM and 150 nM, respectively). The cleavage rates for all three P4–P6 concentrations differed by less than a factor of two, so the P4–P6 concentration used for kinetic analysis (100 nM) was saturating (data not shown). Therefore, the decrease in activity could not be attributed to decreased association of the Δ C209 P4–P6 with the

²Dimethyl sulfate (DMS) modification of N-1 of A and N-3 of C is revealed by primer extension using reverse transcriptase. Similarly, *N*-cyclohexyl-*N'*-[2-(*N*-methyl-4-morpholino)ethyl]carbodiimide *p*-toluene sulfonate (CMCT) modification of N-3 of U and, to a lesser extent, N-1 of G is revealed by primer extension.

FIGURE 9. DMS modification of wild-type P4–P6 RNA compared to the Δ C209 P4–P6 construct. Sequencing lanes flank the DMS modification experiments. DMS modifies the base pairing face of A and C nucleotides and is identified by reverse transcriptase termination at the base preceding the modified nucleotide. Modified nucleotides therefore run one position higher in the sequencing ladder than in the corresponding DMS lanes. DMS modification was quenched at 0, 10 and 20 min. A control was run without any DMS present (far right lane). The position of C209 and the location of the deleted C209 are identified with arrows. (●) marks the modification of A207 and (○) denotes A210.

other components in the three-piece system. These data indicate that while inclusion of C209 destabilizes the RNA, the presence of C209 in the native structure is essential for efficient activity, thereby explaining why C209 is retained in the native ribozyme.

DISCUSSION

We have developed a system for the *in vitro* selection of RNA molecules based on their folding ability, irrespective of catalytic activity. Freedom from functional restraints allowed the RNA selection to find unpredicted solutions for obtaining structural stability by utilizing opportunistic mutations caused by the polymerases used in the selection-amplification protocol. The selected RNAs showed some dependence upon tetraloop identity for maintaining a stable fold, but much of the structural stability could be attributed to mutations in the nonrandomized region of the RNA.

The deletion of C209 was identified and confirmed as a mutation that, by itself, confers significant stability on P4–P6 RNAs. It is possible that some selected molecules could have circumvented the stability requirement for accelerated native gel mobility by con-

forming to an entirely different, yet compact secondary or tertiary structure. In the case of the Δ C209 mutant, however, we have evidence from mutational analysis (Fig. 8) and chemical probing (Fig. 9) that the mutation further stabilizes a structure very similar to that of the native P4–P6 RNA.

Even though the C209 deletion was structurally stabilizing, it was found to be detrimental to ribozyme activity. This is not surprising as C209 is conserved in 80% of the group I introns and the major groove of the C209–G111 base pair is important for proper assembly of the catalytic core (Tanner et al., 1997). Our results emphasize that *in vitro* selection experiments can be directed by various specified selective pressures, but natural selection must be guided by function. Thus, structurally destabilizing nucleotides, essential for activity, are retained and compensated for by a scaffold of additional RNA interactions that stabilize the overall structure of the RNA.

Interestingly, these data also lead to the hypothesis that stabilization of RNA secondary structure can stabilize its tertiary structure. Selected RNAs showed a preponderance of G•U wobbles that mutated to either G–C or A–U Watson–Crick base pairs in P5b. Although

FIGURE 10. Activity of P4–P6 RNAs containing Δ C209 in the context of wild-type and the base paired mutant receptor. Activities were studied in the context of the *Tetrahymena thermophila* three-piece intron system. P4–P6, P1–P3, and P3–P9 RNAs were allowed to assemble in 30 mM Tris-HCl, pH 7.5, and 5 mM MgCl_2 at 50 °C for 5 min, equilibrated to 40 °C for 10 min before initiating reaction with 5 mM GTP and 75 mM MgCl_2 (final concentrations). The observed rate constants were calculated by fitting data to a simple exponential equation [$y = \exp(-kt)$]. The table of symbols appears on the graph. WT: wild-type P4–P6 RNA; Δ C209: deletion of C209; rcpt-BP: the tetraloop receptor is base paired as shown in Figure 8D under rcpt-BP; Δ C209 rcpt-BP: a double mutant combining the C209 deletion and the base-paired tetraloop receptor; BP: the same unfolded control used in all of the pictured native gels in which J5/5a is base paired within an otherwise wild-type P4–P6 construct.

these mutations in isolation had little effect on tertiary folding (Fig. 7), their frequent occurrence in selected RNAs (Fig. 6) argues that they were stabilizing in other sequence contexts. Deletion of C209 causes the adjacent, normally bulged A210 to become less accessible to methylation (Fig. 9), which could indicate that A210 now replaces C209 to give an uninterrupted P4 helix. In support of this idea, site-specific deletion of the bulged A210 in wild-type P4–P6 also results in a more stably folded tertiary structure (our unpubl. data), again supporting the idea that stabilization of RNA secondary structure may indirectly increase its tertiary stability. The specific effects on secondary and tertiary structure caused by the Δ C209 mutation are being investigated in detail; these studies have been greatly facilitated by Δ C209's propensity to crystallize readily.

MATERIALS AND METHODS

Library construction

The initial DNA pool was designed with a randomized region containing 19 nt. This region includes the tetraloop receptor

(CCUAAG..UAUGG) plus an additional 4 bp located in P6a. The pool was created *de novo* using DNA oligos synthesized on an Applied Biosystems DNA synthesizer. The initial PCR reaction contained 0.05 μM of an oligo representing the 3' end of the P4–P6 molecule (U167–A261), including 19 randomized nucleotides ending with an *EcoRI* site followed by an *XbaI* site for cloning, and 0.25 μM of an oligo complementary to the 5' side of the P4–P6 molecule (A104–A184) with the appropriate P5b tetraloop (CUUG, GAAA, or GCAA), a T7 promoter, and an *EcoRI* site for cloning; in addition, two amplification primers were included at concentrations of 1.25 μM . PCR reactions (500 μL) were done in 1 $\times$ PCR buffer (10 mM Tris-HCl, pH 8.3, 1.5 mM MgCl_2 , 50 mM KCl), 1 U/100 μL of reaction of Taq polymerase. The PCR reaction protocol involved denaturation of the DNA for 5 min at 94 °C followed by 6 cycles of 94 °C for 30 s, 60 °C for 60 s, 72 °C for 30 s and finished with a 7 min incubation at 72 °C. Following PCR the DNA for the initial pool was ethanol precipitated and subsequently *EcoRI* digested overnight at 37 °C. The digest was reprecipitated, and 4.0 μg of each DNA template was transcribed in a 1.0 mL reaction containing 0.05 mCi α - ^{32}P ATP, 1 mM ATP, 1 mM GTP, 1 mM CTP, 1 mM UTP, 40 mM Tris-HCl, pH 7.0, 10 mM dithiothreitol (DTT), 20 mM MgCl_2 , 4 mM spermidine, and T7 RNA polymerase for 5 h at room temperature ($\sim$ 22 °C). The RNA for generation 0 was gel

purified on denaturing gels: 8% polyacrylamide, 7 M urea, 1× TBE (100 mM Tris-base, 83 mM boric acid, 1 mM ethylenediaminetetraacetic acid (EDTA)). After excising the RNA from the gel, the gel slices were crushed and soaked in TEN (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 250 mM NaCl) overnight at 4 °C. The supernatant was filtered and the RNA was ethanol precipitated, then resuspended in 60 μ L TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA).

Selection procedure

Denaturing gels were used to select for full-length RNA molecules prior to selecting for folded molecules. The RNAs were separated on 8% polyacrylamide, 7 M urea, 1× TBE gels. The RNA transcripts were visualized by UV shadowing, excised, eluted in TEN, ethanol precipitated, and redissolved in TE.

Native gels were used to separate folded from nonfolded RNAs. Each generation was folded at a specific MgCl_2 concentration, which was decreased throughout the selection: generations 0-2 were folded in 10 mM MgCl_2 , generation 3 in 9 mM MgCl_2 , generation 4 in 5 mM MgCl_2 , generation 5 in 4 mM MgCl_2 , generation 6 in 2 mM MgCl_2 , generation 7 in 1.4 mM MgCl_2 , generation 8 in 1.2 mM MgCl_2 , generation 9 in 1.1 mM MgCl_2 , and generation 10 was sequenced and analyzed. Note that the MgCl_2 concentrations given represent free MgCl_2 concentration in solution; when TBE was used, an additional 1 mM MgCl_2 was included to compensate for MgCl_2 that would bind to the 1 mM EDTA in the TBE. The RNAs were folded at the previously stated MgCl_2 concentration in 1× TBE, 5% glycerol at 50 °C for 10 min, and then equilibrated to 35 °C for 5 min before loading onto the native gels. Prior to loading the RNAs, the 8% polyacrylamide, 1× TBE gels containing the previously stated MgCl_2 concentration were preequilibrated to 35 °C for 1 h with water-cooled coils immersed in a buffer tank contacting the back side of the gels. The running buffer (1× TBE and 10–1.1 mM MgCl_2) contained the same MgCl_2 concentration as the gels. The gels were run at 150–200 V for ~5 h at a constant temperature of 35 °C. The folded RNAs were selected on the basis of correct gel mobility, defined by WT P4–P6 mobility markers. The gel slice corresponding to the mobility of folded P4–P6 RNA was excised and eluted in TEN as described above, ethanol precipitated, and redissolved in 18 μ L of TE.

Isolated RNA was DNase treated, reverse transcribed, and PCR amplified for further rounds of selection. DNase (20 μ L) reactions were carried out with 12 μ L (out of 18 μ L) of selected RNA, 2 μ L of 10× DNase buffer (400 mM Tris-HCl, pH 7.5, 100 mM NaCl, 60 mM MgCl_2 , 100 mM CaCl_2), and 2 μ L of RQ1 DNase (Promega) for 30 min at 37 °C. Reactions were quenched with 1.0 μ L of 500 mM EDTA before phenol-chloroform extraction, ethanol precipitation, and resuspension in 12.5 μ L of TE. Reverse transcription was carried out in three steps. A 20- μ L annealing reaction was set up with 5 μ L of DNased RNA and 14 μ L of MM1 buffer (73 mM Tris-HCl, pH 8.3, 87 mM NaCl, 50 mM DTT, 3.6 μ L primer); this reaction was incubated at 95 °C for 1 min and cooled to room temperature. The entire annealing reaction was added into an extension reaction that contained 0.75 μ L of AMV reverse transcriptase (Life Sciences Inc.) and 30 μ L of MM2

buffer (50 mM Tris-HCl, pH 8.3, 60 mM NaCl, 50 mM DTT, 10 mM MgOAc , 0.67 mM dATP, 0.67 mM dGTP, 0.67 mM dCTP, 0.67 mM dTTP) and incubated at room temperature for 5 min, then 37 °C for 10 min, then 60 °C for 20–30 min. The 50- μ L extension reaction was then mixed with 0.5 μ L of Taq polymerase and 50 μ L of MM3 buffer (1 μ M 5'-primer, 100 mM KCl, 10 mM MgCl_2) and 15 cycles of PCR were carried out: 94 °C for 30 s, 55 °C for 60 s, 72 °C for 30 s. PCR products were ethanol precipitated and *Eco*I digested prior to phenol-chloroform extraction and reprecipitation with ethanol. The final DNA product was brought up in 20 μ L TE, and 10 μ L were used in a 20- μ L transcription reaction identical to the reaction described for preparation of the initial pool of RNA. Gel purification and RNA isolation were also conducted as previously described.

Cloning, sequencing, and analysis

After 10 rounds of selection the DNA from generation 9 was cut with *Eco*R1 and *Xba*I and ligated into a similarly cleaved pUC19 vector. The ligation reaction was ethanol precipitated and brought up in 10 μ L of TE. A 1.5- μ L portion was then used to transform XL1-blue electrocompetent *Escherichia coli* with a Bio-Rad gene pulser (1.8 kV, 400 ohms). Colonies were screened directly using PCR; cells were collected on a toothpick and washed into 20 μ L of LB (10 g/L bacto-tryptone, 5 g/L bacto-yeast extract, 10 g/L NaCl). A 1- μ L portion of the LB plus cells was mixed into a 15 μ L PCR reaction containing 1× PCR buffer (10 mM Tris-HCl, pH 8.3, 1.5 mM MgCl_2 , 50 mM KCl), 0.2 mM of each dNTP, 1 U Taq polymerase, and 0.5 μ M of each primer. The PCR reaction was carried out as follows: 94 °C for 4 min; 94 °C for 30 s; 60 °C for 30 s; and 72 °C for 30 s for 25 rounds, followed by a 7-min incubation at 72 °C. The PCR products were analyzed on 1.5% LE agarose gels containing ethidium bromide. Colonies producing 214-bp PCR products were minipreped, and the DNAs were sequenced using a ABI 377 automated sequencer from Perkin Elmer. Sequences were aligned using Clustal (wPPC). Cloning of additional P4–P6 constructs was carried out in a manner similar to the procedure described above. PCR mutagenesis was used to create the desired constructs that were then sticky-end ligated into pUC19 and transformed into XL1-blue *E. coli*. Automated sequencing verified the mutant sequences.

DMS probing

DMS probing was used to elucidate qualitative differences in structure between wild-type and Δ C209 P4–P6 molecules. RNA (10 pmol) in a 180- μ L reaction with 5 μ M MgCl_2 and 35 mM Tris-HCl, pH 7.5, was folded by incubating for 10 min at 50 °C and cooling to room temperature. The zero time point was quenched before initiating DMS modification with 1.0 μ L of 25% DMS per 100 μ L of reaction mixture. Reactions were quenched at 10 and 20 min on ice with 0.5 vol of 0.5 M β -mercaptoethanol followed by ethanol precipitation. The samples were brought up in TE and reverse transcribed as described (Inoue & Cech, 1985; Zaug & Cech, 1995). Reverse transcription reactions were subject to electrophoresis on 20% polyacrylamide, 7 M urea, 1× TBE denaturing gels.

Ribozyme kinetic analysis using the three-piece system

Kinetics were carried out as described by Doudna and Cech (1995). The RNA components were mixed in a reaction with 1 nM 5' ³²P-labeled P1–P3, 150 nM P3–P9, 100 nM P4–P6, 30 mM Tris-HCl, pH 7.5, 5 mM MgCl₂, then incubated for 10 min at 50 °C and equilibrated to the 40 °C reaction temperature for 5 min. The zero time point was quenched before the reactions were initiated with 0.1 vol of 10× initiation buffer (50 mM GTP, 755 mM MgCl₂, 5 mM spermidine). The reactions were quenched on ice in 1 vol of stop buffer (92% formamide and 30 mM EDTA). Reactions were analyzed on 20% polyacrylamide denaturing gels, as described above.

ACKNOWLEDGMENTS

We are grateful to Barbara Golden for help in initiating these studies and Phil Bevilacqua for useful discussions. This work was supported by National Institutes of Health grant GM28039 to T.R.C. T.R.C. is an investigator of the Howard Hughes Medical Institute and an American Cancer Society Professor.

Received April 7, 1999; returned for revision April 27, 1999; revised manuscript received May 5, 1999

REFERENCES

- Antao VP, Lai SY, Tinoco I Jr. 1991. A thermodynamic study of unusually stable RNA and DNA hairpins. *Nucleic Acids Res* 21:5901–5905.
- Bartel DP, Szostak JW. 1993. Isolation of new ribozymes from a large pool of random sequences. *Science* 261:1411–1418.
- Bartel DP, Zapp ML, Green MR, Szostak JW. 1991. HIV-1 Rev regulation involves recognition of non-Watson–Crick base pairs in viral RNA. *Cell* 67:529–536.
- Beaudry AA, Joyce GF. 1992. Directed evolution of an RNA enzyme. *Science* 257:635–641.
- Bevilacqua JM, Bevilacqua PC. 1998. Thermodynamic analysis of an RNA combinatorial library contained in a short hairpin. *Biochemistry* 37:15877–15884.
- Butcher SE, Dieckmann T, Feigon J. 1997. Solution structure of a GAAA tetraloop receptor RNA. *EMBO J* 16:7490–7499.
- Campbell TB, Cech TR. 1995. Identification of ribozymes within a ribozyme library that efficiently cleave a long substrate RNA. *RNA* 1:598–609.
- Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA. 1996. Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science* 273:1678–1685.
- Costa M, Michel F. 1995. Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J* 14:1276–1285.
- Costa M, Michel F. 1997. Rules for RNA recognition of GNRA tetraloops deduced by in vitro selection: Comparison with in vivo evolution. *EMBO J* 16:3289–3302.
- Doudna JA, Cech TR. 1995. Self-assembly of a group I intron active site from its component tertiary structural domains. *RNA* 1:36–45.
- Doudna JA, Cech TR, Sullenger BA. 1995. Selection of an RNA molecule that mimics a major autoantigenic epitope of human insulin receptor. *Proc Natl Acad Sci USA* 92:2355–2359.
- Ellington AD, Szostak JW. 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature* 346:818–822.
- Golden BL, Gooding AR, Podell ER, Cech TR. 1998. A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. *Science* 282:259–264.
- Green R, Szostak JW. 1994. In vitro genetic analysis of the hinge region between helical elements P5–P4–P6 and P7–P3–P8 in the sunY group I self-splicing intron. *J Mol Biol* 235:140–155.
- Illangasekare M, Sanchez G, Nickles T, Yarus M. 1995. Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* 267:643–647.
- Inoue T, Cech TR. 1985. Secondary structure of the circular form of the *Tetrahymena* rRNA intervening sequence: A technique for RNA structure analysis using chemical probes and reverse transcriptase. *Proc Natl Acad Sci USA* 82:648–652.
- Jabri E, Cech TR. 1998. In vitro selection of the *Naegleria* GIR1 ribozyme identifies three base changes that dramatically improve activity. *RNA* 4:1481–1492.
- Jenison R, Gill SC, Pardi A, Polisky B. 1994. High-resolution molecular discrimination by RNA. *Science* 263:1425–1429.
- Jucker FM, Heus HA, Yip PF, Moors EH, Pardi A. 1996. A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J Mol Biol* 264:968–980.
- Jucker FM, Pardi A. 1995. Solution structure of the CUUG hairpin loop: A novel RNA tetraloop motif. *Biochemistry* 34:14416–14427.
- Lehman N, Joyce GF. 1993. Evolution in vitro of an RNA enzyme with altered metal dependence. *Nature* 361:182–185.
- 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.
- Pan T, Uhlenbeck O. 1992. In vitro selection of RNAs that undergo autolytic cleavage with Pb²⁺. *Biochemistry* 31:3887–3895.
- Pley HW, Flaherty KM, McKay DB. 1994. Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* 372:111–113.
- SantaLucia J Jr, Kierzek R, Turner DH. 1992. Context dependence of hydrogen bond free energy revealed by substitutions in an RNA hairpin. *Science* 256:217–219.
- Szewczak AA, Cech TR. 1997. An RNA internal loop acts as a hinge to facilitate ribozyme folding and catalysis. *RNA* 3:838–849.
- Tanner MA, Anderson EM, Gutell RR, Cech TR. 1997. Mutagenesis and comparative sequence analysis of a base triple joining the two domains of group I ribozymes. *RNA* 3:1037–1051.
- Treiber DK, Rook MS, Zarrinkar PP, Williamson JR. 1998. Kinetic intermediates trapped by native interactions in RNA folding. *Science* 279:1943–1946.
- Tuerk C, Gold L. 1990. Systematic evolution of ligands to bacteriophage T4 DNA polymerase. *Science* 249:505–510.
- Woese CR, Winker S, Gutell RR. 1990. Architecture of ribosomal RNA: Constraints on the sequence of “tetra-loops”. *Proc Natl Acad Sci USA* 87:8467–8471.
- Zaug AJ, Cech TR. 1995. Analysis of the structure of *Tetrahymena* nuclear RNAs in vivo: Telomerase RNA, the self-splicing rRNA intron, and U2 snRNA. *RNA* 1:363–374.
- Zhang B, Cech TR. 1997. Peptide bond formation by in vitro selected ribozymes. *Nature* 390:96–100.