doi:10.1016/S0022-2836(03)00654-5 J. Mol. Biol. (2003) 330, 1015–1025

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Optimization and Generality of a Small Deoxyribozyme that Ligates RNA

Benjamin L. Ricca, Amanda C. Wolf and Scott K. Silverman*

Department of Chemistry University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, IL 61801, USA

In vitro evolution was previously used to identify a small deoxyribozyme, 7Q10, that ligates RNA with formation of a 2′–5′ phosphodiester linkage from a 2′,3′-cyclic phosphate and a 5′-hydroxyl group. Ligation occurs in a convenient "binding arms" format analogous to that of the well-known 10–23 and 8–17 RNA-cleaving deoxyribozymes. Here, we report the optimization and generality of 7Q10 as a 2′–5′ RNA ligase. By comprehensive mutagenesis of its 16-nucleotide enzyme region, the parent 7Q10 sequence is shown to be optimal for RNA ligation yield, although several mutations are capable of increasing the ligation rate approximately five-fold at the expense of yield. The 7Q10 deoxyribozyme ligates any RNA substrates that form the sequence motif UA↓GR (arrowhead = ligation site and R = purine), providing at least 30% yield of ligated RNA in ~1–2 hours at 37 °C and pH 9.0. Comparable yields are obtained in ~12–24 hours at pH 7.5, which may be more suitable for larger RNAs that are more sensitive to non-specific degradation. For RNA substrates that form the related ligation junction UA↓GY (Y = pyrimidine), somewhat lower yields are obtained, but significant ligation activity is still observed. These data establish that 7Q10 is a generally applicable RNA ligase. A plot of log(k_{obs}) *versus* pH from pH 6.9 to 9.0 has a slope of just under 1, suggesting that a single deprotonation occurs during the rate-determining reaction step. The compact 7Q10 deoxyribozyme has both practical utility and the potential for increasing our structural and mechanistic understanding of how nucleic acids can mediate chemical reactions.

Keywords: deoxyribozyme; DNA enzyme; RNA ligation; mutagenesis; cyclic phosphate

*Corresponding author

Introduction

Site-specifically modified RNAs are valuable for studies of RNA structure, folding, and catalysis.[1–19] Modifications are most readily incorporated by solid-phase methods using synthetic ribonucleoside phosphoramidites.[20–23] Because many interesting RNAs are larger than can be prepared directly by solid-phase synthesis, new approaches for RNA ligation are in demand. We have recently begun to develop deoxyribozymes (DNA enzymes)[24–27] for RNA ligation. At least two general strategies may be employed to identify deoxyribozymes that ligate RNA. First, *in vitro* selection may be performed using a random DNA pool as the starting point. Second, a deoxyribozyme that is known to cleave RNA may be evolved to operate "in reverse" as an RNA ligase. In our hands, both approaches have been successful.[28–30] In particular, the latter strategy provided a family of deoxyribozymes (Figure 1(a)) that were evolved from the 8–17 RNA-cleaving DNA enzyme[31] (Figure 1(b)). One of the new RNA ligase deoxyribozymes, designated 7Q10, provides ~30% yield of ligated RNA in 12 hours at pH 7.5 and 40 mM Mg^{2+} with a particular set of RNA substrates.[29] The ligated RNA product was shown to be joined by a non-native 2′–5′ phosphodiester linkage, which should be tolerable or even advantageous under many circumstances.[28,29] Here, we report the systematic optimization of 7Q10 as a general 2′–5′ RNA ligase that can be applied to join a wide range of RNA substrates. Additionally, 7Q10 offers the potential for future structural and mechanistic investigations that should expand our understanding of how nucleic acids can mediate chemical reactions.

E-mail address of the corresponding author: scott@scs.uiuc.edu

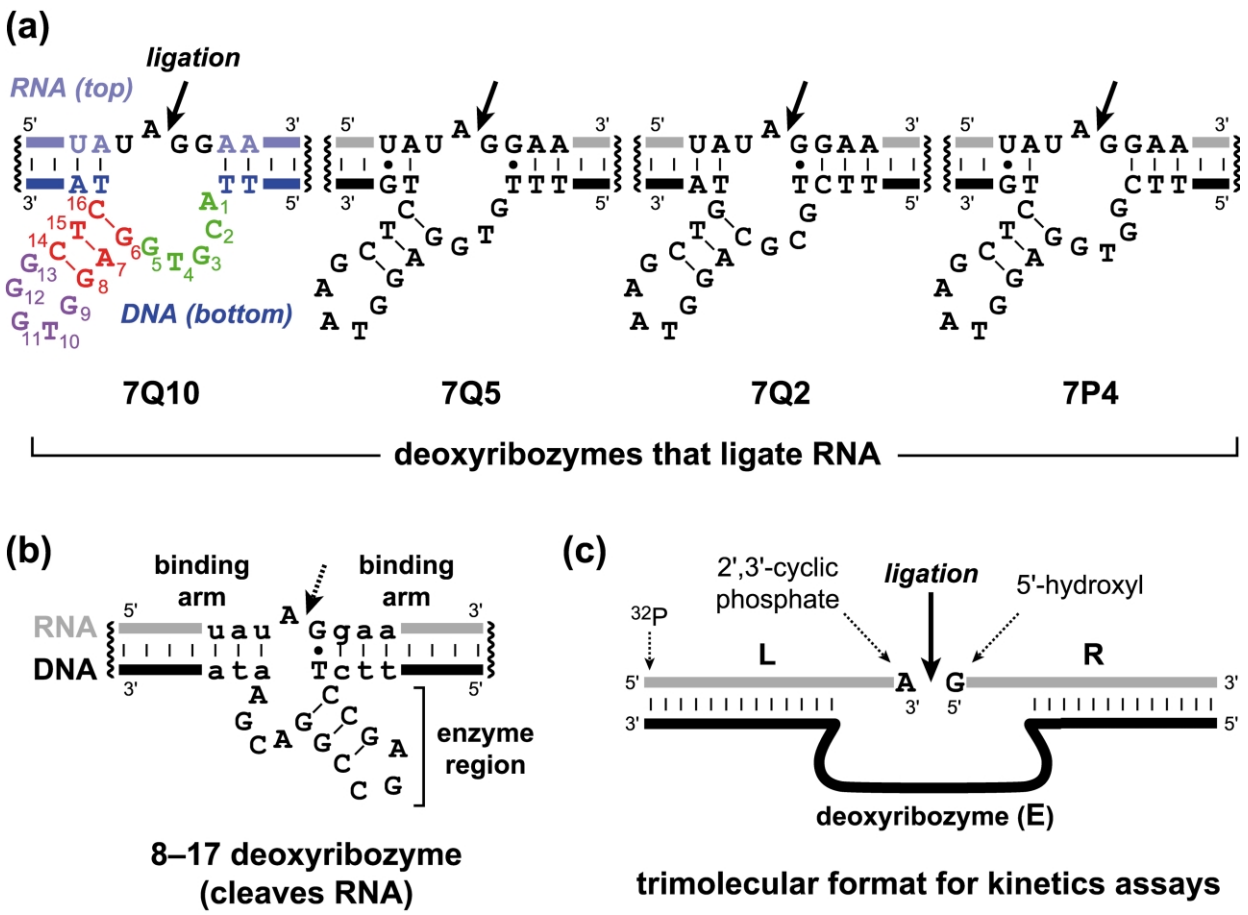

Figure 1. Deoxyribozymes that ligate and cleave RNA. (a) Sequences of the 7Q10 deoxyribozyme that ligates RNA, along with three other deoxyribozymes from the same family.[29] The conjectured secondary structure of the 7Q10 enzyme region is colored green for the single-stranded segment, red for the stem, and purple for the pentaloop. The substrate-binding arms are blue. The arrow indicates the ligation site. (b) Structure of the 8–17 deoxyribozyme that cleaves RNA.[31] The lowercase nucleotides may be changed freely as long as Watson–Crick base-pairing is maintained. The arrow indicates the cleavage site. (c) The trimolecular kinetics assay format used to determine the effects of changing sequence and incubation conditions on the DNA-mediated RNA-ligation reaction.

Results

In Figure 1(a) are shown four deoxyribozymes that ligate RNA. These DNA enzymes were previously identified by *in vitro* evolution from the 8–17 DNA enzyme that cleaves RNA.[29] Here, we used the highest-yielding of these RNA ligase deoxyribozymes, 7Q10, as the starting point for optimizing the RNA ligation reaction and for establishing its generality with a wide range of RNA substrate sequences.

Strategy for optimizing and generalizing 7Q10 as an RNA ligase

The enzyme region of 7Q10 (Figure 1(a)) comprises 16 DNA nucleotides that may be conceptually divided into a single-stranded segment (green), a three base-pair stem (red), and a pentaloop (purple). The DNA–RNA substrate-binding arms are shown in blue. The illustrated secondary structure for 7Q10 is conjectural (i.e. based only on computer folding algorithms), but the structure is consistent with mutagenesis experiments that are reported in the first section below, and the 8–17 DNA enzyme is thought to have a somewhat similar stem–loop structure.[29] Each nucleotide in the enzyme region was investigated for its role in RNA ligation activity by mutagenesis as described below. These experiments used standard assay conditions of 50 mM Hepes (pH 7.5), 150 mM NaCl, 2 mM KCl, and 40 mM $MgCl_2$ at 37 °C, which are the incubation conditions that we have used in related deoxyribozyme selections and assays.[28,29] Further experiments in the second section below examined changes in temperature, pH, and salt concentration to optimize these environmental variables. Previously we showed that the $K_{d,app}$ for Mg^{2+} for 7Q10 is 18 mM at 37 °C and pH 7.5.[29] The divalent metal ion used in all experiments of the second section was 40 mM Mg^{2+}. In the final two sections are described investigations of the RNA substrate sequence dependence of the optimized 7Q10 deoxyribozyme under the most favorable incubation conditions. The overall conclusion is that under optimal conditions, 7Q10

provides at least 30% yield of 2′–5′ ligated RNA in ~1–2 hours with any RNA substrate sequences that can form a UA↓GR ligation junction (R = purine). Furthermore, a UA↓GY linkage (Y = pyrimidine) can be formed with only slightly reduced yield.

Nucleotide requirements in the enzyme region of the 7Q10 deoxyribozyme

In this section are described experiments to optimize the enzyme region DNA sequence of 7Q10 for RNA ligation. The 7Q5 deoxyribozyme was identified during the same selection procedure as was 7Q10.[29] By comparison between 7Q10 and 7Q5 (see Figure 1(a) for sequences and nucleotide numbering), it was noted that the A_1C_2 dinucleotide of 7Q10 is replaced with a single T in 7Q5, and $G_{11}G_{12}$ is replaced with AA. In addition, the A-U base-pair in the 7Q10 left-hand-binding arm is replaced with G-U in 7Q5. As a result of these changes, the observed ligation rate k_{obs} increased fivefold from 7Q10 to 7Q5, but the yield dropped threefold. As summarized in Figure 2, we systematically investigated all possible combinations among these three nucleotide elements. The results reveal that the particular combination of sequence elements present in the parent 7Q10 sequence provides the highest ligation yield, whereas the combination of 7Q5 elements provides the highest k_{obs} at the expense of yield. As expected, the replacement of A-U with G-U in the binding arm mattered little; the binding arms are addressed more systematically below. Replacement of A_1C_2 with T in the 7Q10 single-stranded segment abolished ligation activity, but the same T nucleotide supported a high ligation rate in 7Q5. The only difference other than $A_1C_2 \leftrightarrow$ T is that 7Q10 has $G_{11}G_{12}$ in its pentaloop, whereas 7Q5 has $A_{11}A_{12}$. These results suggest a detailed interplay during the ligation reaction between the single-stranded segment and the pentaloop. Understanding the exact nature of this interplay requires more detailed structural data that is not yet available.

The modified version of 7Q10 with an $A_{11}A_{12}$ pentaloop had a two- to threefold higher ligation rate k_{obs} than the parent 7Q10. Both this modified deoxyribozyme and the parent 7Q10 sequence were used to explore changes at the other DNA nucleotides of the enzyme region, as follows (Figure 3):

Pentaloop nucleotides (Figure 3, purple)

The pentaloop nucleotide T_{10} preceding $G_{11}G_{12}$ was changed to C ($T_{10} \rightarrow C$) with little effect on the 7Q10 rate or yield, suggesting that this nucleotide's identity is unimportant, although the nucleotide could not be deleted without obliterating activity (ΔT_{10}, <4% yield = "inactive"). The $G_9 \rightarrow A_9$ or $G_{13} \rightarrow A_{13}$ mutants were also inactive, indicating critical roles for those nucleotides, which are conserved among all of the deoxyribozymes shown in Figure 1(a). Inserting an additional A into the $A_{11}A_{12}$ pentaloop (i.e. converting $A_{11}A_{12} \rightarrow$ AAA) or changing $G_{11}G_{12} \rightarrow$ G or $G_{11}G_{12} \rightarrow$ A led to inactive or nearly inactive deoxyribozymes, suggesting that a five nucleotide loop is optimal. In contrast, $G_{11}G_{12} \rightarrow$ GA or $G_{11}G_{12} \rightarrow$ AG led to activity essentially equivalent to that of the parent sequence, indicating that the size of the loop is more important than the identities of these particular nucleotides. Pyrimidines were not tested at these positions.

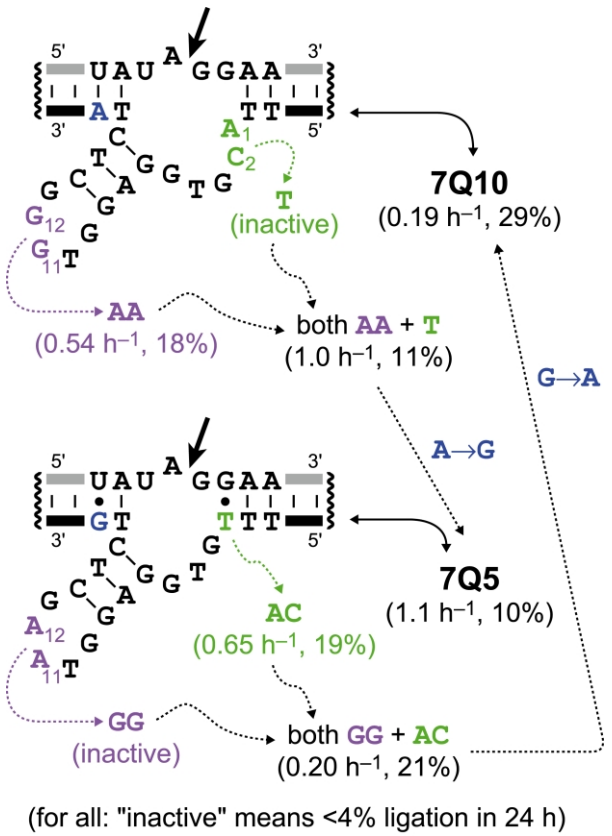

Figure 2. Mutagenesis of three regions of the 7Q10 deoxyribozyme sequence and its effect on ligation activity. In color are highlighted the three sequence elements that differ between 7Q10 (upper sequence) and 7Q5 (lower sequence). All possible combinations of the three sequence elements were tested as described in the text, with the observed ligation rates (k_{obs}) and yields shown (pH 7.5, 37 °C). "Inactive" means <4% yield in 24 hours. The logical connections among the sequences are shown as dotted arrows. Subsequent to the collection of these data and the data in Figure 3, it was found that a slightly longer left-hand substrate modestly improves the rates and yields (see Materials and Methods). Comparison of the relative rates and yields for the various deoxyribozymes should not be affected by this issue.

Single-stranded segment nucleotides (Figure 3, green)

The nucleotides at positions 3–5 of the single-stranded segment are conserved as $G_3T_4G_5$ for three of the four parent deoxyribozymes of

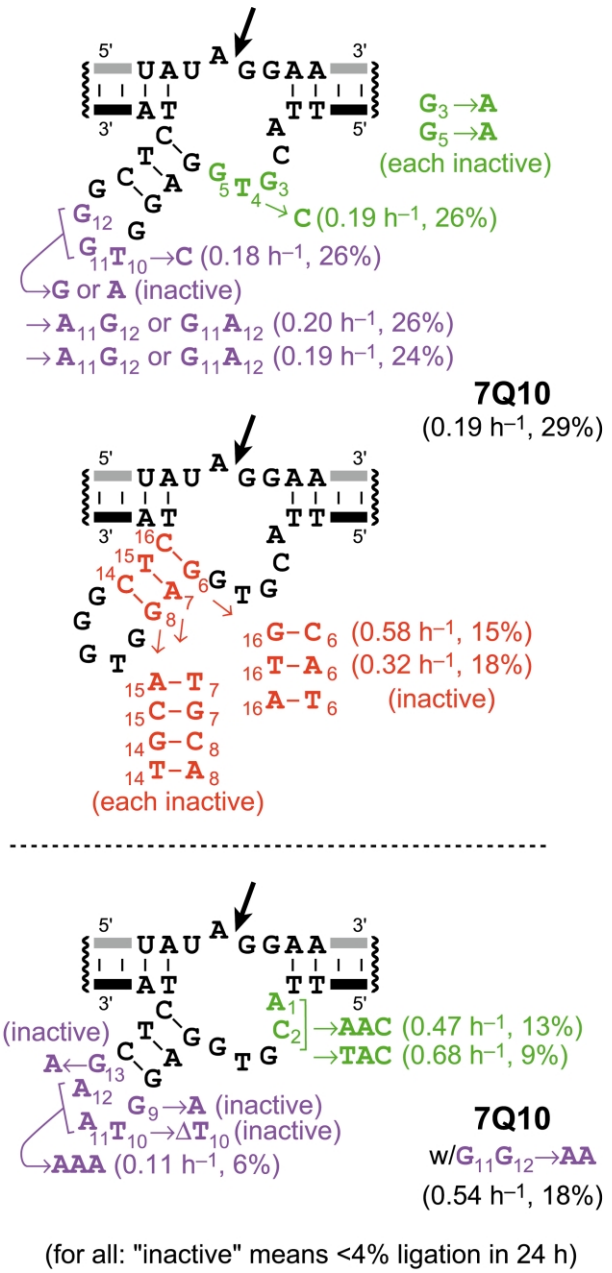

Figure 3. Mutagenesis of the remaining sequence elements of 7Q10 and its effect on ligation activity. The observed ligation rates (k_{obs}) and yields are shown next to each mutation (pH 7.5, 37 °C). "Inactive" means <4% yield in 24 hours. See the text for further explanation.

Figure 1(a), while 7Q2 has $G_3C_4G_5$. As expected based on these sequence comparisons, the transitions $G_3 \rightarrow A$ or $G_5 \rightarrow A$ each abolished RNA ligase activity for 7Q10, while $T_4 \rightarrow C$ was tolerated with little loss of activity. We also examined changes in the first two nucleotides of the single-stranded segment. Because $A_1C_2 \rightarrow T$ (as found in 7Q5) still supports activity as long as $A_{11}A_{12}$ is present, as described above (Figure 2), some flexibility in the length of this sequence was indicated. Therefore, we tested the effect of inserting an additional nucleotide near the A_1C_2. Inserting either an A or T before A_1C_2 gave active deoxyribozymes with equivalent k_{obs} but somewhat reduced yield. This indicates no strict length requirement for the single-stranded segment, at least in some contexts, in contrast to the length requirement for the pentaloop (see above). It is interesting to note that a similar tolerance regarding the length of its single-stranded segment was found for the 8–17 deoxyribozyme.[31]

Stem nucleotides (Figure 3, red)

Various changes to the putative three base-pair stem of 7Q10 were examined. Swapping the nucleotides of either of the bottom two base-pairs was not tolerated at all. At the uppermost base-pair, the 7Q2 deoxyribozyme has G_{16}-C_6, while the other three DNA enzymes of Figure 1(a) (including 7Q10) each have C_{16}-G_6. Curiously, three of the four possible Watson–Crick combinations at this base-pair of 7Q10 gave active deoxyribozymes; only the A_{16}-T_6 base-pair failed to support ligation. The k_{obs} was higher for the two active combinations not found in the parent 7Q10 sequence, although the yield was lower. The data presented here are consistent with the assignment of the putative stem as an element of 7Q10 secondary structure. However, further experiments (e.g. chemical probing or direct structural analysis by X-ray crystallography or NMR spectroscopy) are necessary to demonstrate with confidence the existence of the stem.

A summary of all of the mutagenesis experiments that are depicted in Figures 2 and 3 is that the parent 7Q10 DNA sequence gives the highest ligation yield of any tested sequence. Incorporation of 7Q5 sequence elements (i.e. $G_{11}G_{12} \rightarrow A_{11}A_{12}$ with or without $A_1C_2 \rightarrow T$) or changes to the three-base-pair stem increase the ligation rate up to fivefold, but at the expense of yield. Because the 7Q10 rate is acceptable for practical RNA ligation (see below), and because the highest possible ligation yield is usually desirable even if the incubation time is modestly longer, we used the parent 7Q10 sequence for the remaining assays.

Dependence of 7Q10 Ligation on pH, temperature, and salt concentration

Having established that the parent 7Q10 sequence is optimal for ligation yield, we investigated the effects of changing pH, temperature, and salt concentration. In all assays that are described in this and the following section, the RNA substrate sequences flanking the UA↓GG (i.e. the light blue segments in Figure 1(a)) were kept identical with those used during the original selection of 7Q10,[28,29] and the deoxyribozyme sequences (dark blue segments) were their Watson–Crick complements. A strong temperature dependence between 10 °C and 45 °C was observed at pH 7.5, with faster ligation at higher temperature (Figure 4(a)). Ligation was two- to threefold faster at 45 °C than at 37 °C, but there was substantially more non-specific

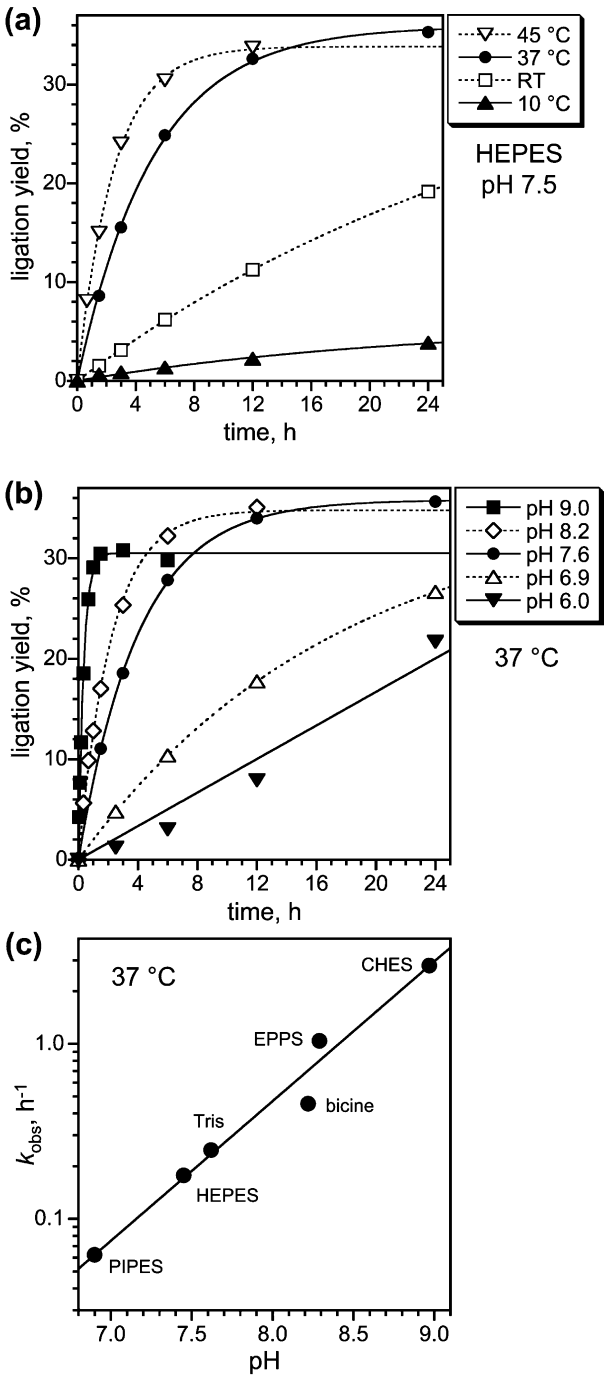

Figure 4. Temperature and pH dependence of 7Q10 ligation activity. (a) Ligation activities observed in 50 mM Hepes (pH 7.5) at temperatures from 10 °C to 45 °C (RT = room temperature, ~23 °C). After 24 hours at 45 °C, significant degradation was observed on the PAGE image (data not shown). k_{obs}: RT, 0.029 h^{-1}; 37 °C, 0.19 h^{-1}; 45 °C, 0.40 h^{-1}. The rates were reproducible to ±10% or better. (b) Representative ligation activities observed at pH values ranging from 9.0 to 6.0 at 37 °C. The buffer compounds (50 mM) for pH 9.0, 8.2, 7.6, 6.9, and 6.0 were Ches, bicine, Tris, Pipes, and Mes, respectively. At incubation times >6 hours at pH 9.0, significant RNA degradation was observed (not shown). Lines through points are first-order exponential fits, except for a linear fit to the pH 6.0 data. A 36-hour data point is not shown for the pH 6.9 assay but was on the fit curve. (c) Plot of log(k_{obs}) *versus* pH derived from data in (b) and similar data. The linear fit has slope 0.80(±0.09). The error bars in both directions are smaller than the sizes of the data points.

degradation at 45 °C at long incubation times (>12 hours; data not shown). At 37 °C, ligation reactions at various pH values from 9.0 to 6.0 gave a wide range of rates for k_{obs} (Figure 4(b)), with significantly higher rates observed at elevated pH. Higher pH values than 9.0 were avoided out of concern for non-specific degradation of the RNA substrates. Between pH 6.9 and 9.0, the value of log(k_{obs}) increased linearly with pH, with a slope of just under 1 (Figure 4(c)). This slope is often ~1 for ribozymes and deoxyribozymes,[32–38] indicating a single deprotonation during the rate-determining step. In contrast to the strong temperature and pH dependence, adjusting both [Na^+] and [K^+] within broad ranges (0–300 mM Na^+ and 0–100 mM K^+) had little effect on the 7Q10 ligation rate or yield (data not shown).

From these data, we reasoned that many practical RNA ligation reactions would likely be performed at pH 9.0, at least for RNA substrates small enough that non-specific degradation would not be problematic. The ligation activity of 7Q10 was reassessed under these conditions at 37 °C and 45 °C. A ligation yield of ~30–40% was obtained in one hour at 37 °C in Ches (pH 9.0) buffer with 40 mM Mg^{2+} (Figure 5). These latter conditions were adopted as the standard conditions for the final assays, which examined changes in the RNA substrate sequences (next two sections).

RNA substrate sequence requirements for 7Q10 near the ligation junction

The results to this point established the optimal DNA sequence (7Q10) and incubation conditions for the RNA ligation reaction. We finally turned our attention to the generality of 7Q10 for RNA ligation with respect to its RNA substrate sequences. Some of the initial experiments in this section were performed at pH 7.5 concurrently with the pH dependence studies of the previous section. Except for these initial experiments, all of the assays in this section were performed at 37 °C with 50 mM Ches (pH 9.0) and 40 mM $MgCl_2$. In all cases, the buffer additionally contained 150 mM NaCl and 2 mM KCl, but the precise monovalent ion concentrations were shown in the previous section to be unimportant.

The original RNA substrates used throughout the selection procedure and in the above assays may be denoted UAUA↓GGAA (see Figure 1(a)), where the arrowhead marks the ligation junction and only four RNA nucleotides are shown on each side of the junction. We methodically replaced these eight nucleotides with others, to explore the tolerance of the parent 7Q10 deoxyribozyme for

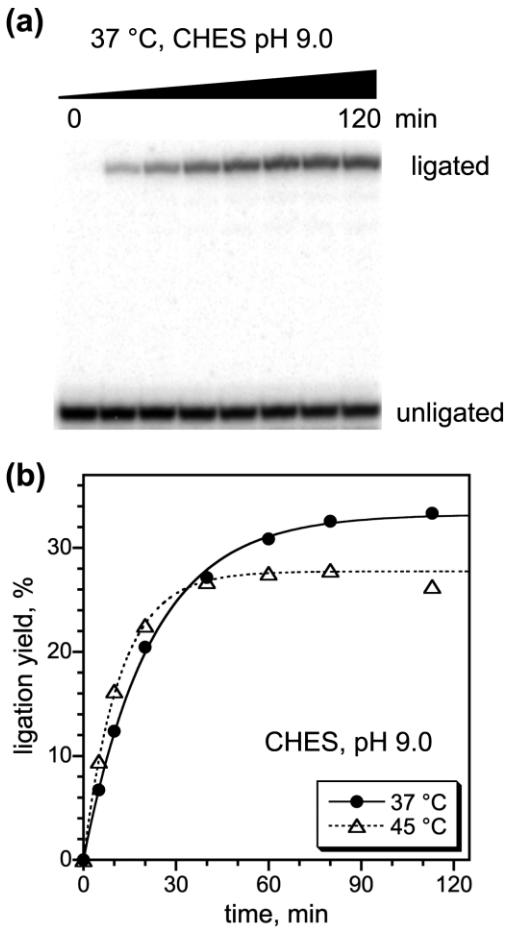

Figure 5. Assays of 7Q10 ligation activity under optimal conditions. (a) PAGE image of a ligation assay performed in 50 mM Ches (pH 9.0) at 37 °C. The yield was reproducible to ±5%. (b) Kinetic plots. k_{obs}: 37 °C, 0.046 min^{-1}; 45 °C, 0.085 min^{-1}. The rates were reproducible to ±10% or better. The final 45 °C timepoint was not fit due to a small amount of degradation clearly visible on the gel image (data not shown). An experiment using 40 mM $MnCl_2$ instead of $MgCl_2$ at 37 °C and pH 9.0 showed a similar ligation rate as with Mg^{2+}, but with only about half the yield (data not shown).

various RNA substrates. These assays were generally performed by taking only three timepoints per reaction, which allowed us quickly to survey ligation activities with a manageable number of gel electrophoresis experiments. Only a summary of the data is provided here; see Supplementary Material for details. In the following descriptions, the altered RNA nucleotides of the UAUA↓GGAA sequence are underlined for identification, and nucleotides not directly relevant to the discussion are omitted.

Changing A↓G to G↓G decreased the activity 14-fold (in terms of yield) to barely detectable levels. Changing A↓G to A↓A, A↓U, or A↓C abolished ligation activity altogether. Changing UA↓G to CA↓G or GA↓G also completely abolished activity (AA↓G was not tested). Together, these data indicate a stringent requirement for UA↓G as a minimal ligation motif of the RNA substrates to be joined. In contrast, the other RNA nucleotides surrounding the ligation junction could be changed more freely. In an initial set of experiments, we used the 7Q10 DNA sequence of Figure 1(a) with no compensatory changes to the nucleotides of its binding arms. Each underlined RNA nucleotide of the UAUA↓GGAA substrates was changed *via* a transition (U to C, A to G, or G to A) with no significant loss of 7Q10 ligation activity, even though the outer two changes on each side appeared to destroy one Watson–Crick base-pair in the putative-binding arm duplex. The implications of these results with respect to changes in the binding arms are explored more systematically below. Based on these initial data, we were encouraged that UA↓G might be the only nucleotides required for ligation, and thus 7Q10 would be quite general as an RNA ligase.

The nucleotide two to the right of the ligation junction (UAUA↓GGAA) provided an interesting situation. By examination of the putative secondary structure of 7Q10 bound with its original RNA substrates (Figure 1(a)), this particular nucleotide appears not to engage in base-pairing interactions. However, such interactions may easily be introduced with appropriate combinations of RNA substrate and deoxyribozyme, and the effects of such interactions were unknown. This was examined more comprehensively by determining the ligation kinetics for every combination of RNA substrate (UAUA↓GXAA) and deoxyribozyme ($A_1 \rightarrow Z$ of the 7Q10 sequence), which required a total of $4 \times 4 = 16$ parallel assays.

The results of these assays are presented in Figure 6. All 16 combinations of nucleotide X in the RNA substrate and Z in the deoxyribozyme gave observable RNA-ligation activity, although some combinations were clearly better than others. The highest yields (>30%) were obtained in the upper left quadrant of Figure 6, with X = R (purine G or A) in the RNA and Z = R′ (the other purine A or G) in the DNA. Conversely, the lowest yields (≤10%) were obtained in the lower right quadrant, where X = Y (pyrimidine C or U) and Z = Y (pyrimidine C or T) in any combination. Along the "Watson–Crick diagonal" (dark shading), yields of >20% were obtained for any Watson–Crick combination of X and Z. These data indicate that the highest ligation yield is achieved for two RNA substrates that form a UA↓GR junction, where the deoxyribozyme nucleotide opposite R is the "other" purine (A or G). If a UA↓GY junction is desired, then this may be achieved with somewhat lower yield by choosing the deoxyribozyme nucleotide opposite Y to be the Watson–Crick complement of Y.

RNA requirements away from the ligation junction: changing the substrate-binding arms

If deoxyribozymes are to be useful for practical RNA ligation, then it must be possible to change

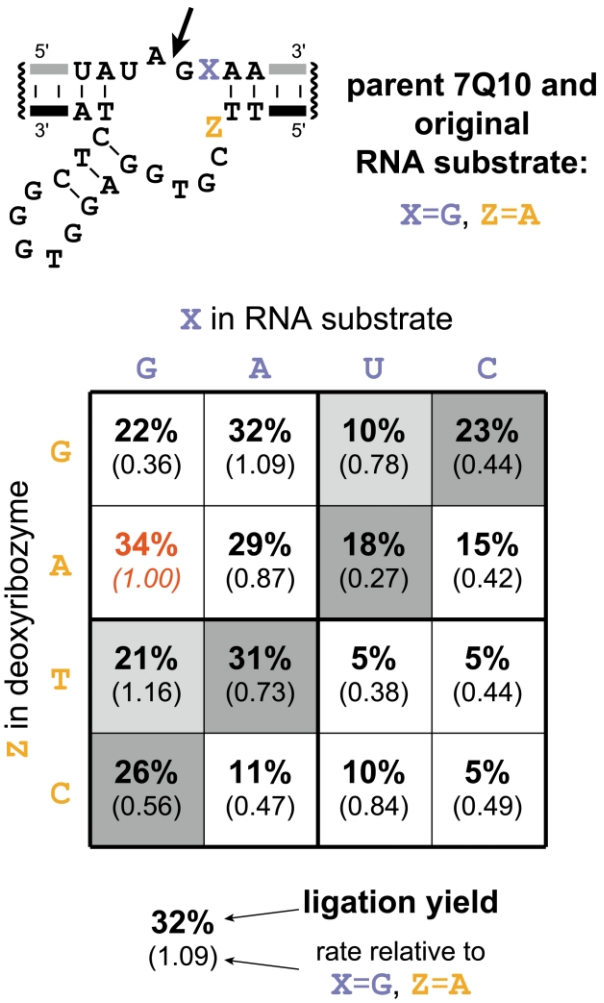

Figure 6. Optimizing the interaction between the 7Q10 deoxyribozyme and its RNA substrates at the position one nucleotide to the right of the ligation junction. The ligation activity of each of the 16 combinations of deoxyribozyme and RNA substrate was determined in 50 mM Hepes (pH 7.5) at 37 °C. Shown are the ligation yields and rates relative to the parent 7Q10 and original RNA substrate sequences (red). In the grid, a dark square denotes a potential canonical Watson–Crick base-pair between the X nucleotide of the DNA and the Z nucleotide of the RNA substrate. A lighter gray square denotes a potential G-U wobble pair.

the RNA substrate sequences away from the ligation site essentially without limitation. For unrelated deoxyribozymes that were identified by *in vitro* selection from completely random DNA pools (unlike 7Q10), the substrate-binding arms were indeed found to have no particular sequence restrictions aside from a limited ligation junction sequence requirement, thereby allowing relatively general RNA ligation.[28] We investigated this issue with 7Q10, using RNA substrates with the common sequence 5′···UA↓GG···3′ to allow particular focus on the effects of changing the surrounding binding arms. As described above, such substrates were already known to provide yields approaching 40% in the special case that their binding arm sequences were identical with those used in the original selection of 7Q10.

These RNA substrate sequences were systematically altered to change every nucleotide flanking the common UA↓GG motif. This was accomplished with a carefully designed set of only six additional substrates, three each for the left and right-hand RNAs, for a total of eight substrate RNAs as shown in Figure 7(a). By scanning vertically down each column of sequence, it may be verified that among the four pairs **1**–**4** of left and right-hand substrates, any particular RNA nucleotide may be found at any position along the sequence (except for the UA↓GG; or in the case of the right-hand substrate of combination **4**, UA↓GA to avoid a long stretch of G nucleotides). Thus, just four ligation reactions with substrate combinations **1**–**4** provide a comprehensive test of the 7Q10 RNA substrate sequence dependence. The results of these assays (Figure 7(b)) reveal that 7Q10 tolerates any RNA substrate nucleotides outside of the UA↓GR region with nearly equivalent ligation rate and yield. Therefore, 7Q10 is truly a general RNA ligase, aside from the modest substrate sequence requirements adjacent to the ligation junction. In control experiments, the additional presence of an RNA substrate not complementary to the deoxyribozyme had no effect on the ligation reaction of the "correct" substrates (data not shown). A deoxyribozyme with binding arms that were not complementary to the RNA substrates also provided no detectable ligation.

Discussion

Our major goals in this study were to optimize the 7Q10 deoxyribozyme for RNA ligation and to explore the extent of its generality for joining various RNA substrate sequences. By determining the activity of 7Q10 and 30 related deoxyribozymes as shown in Figures 2 and 3, we determined that the parent 7Q10 sequence itself is optimal in terms of RNA ligation yield. We further established that outside of a very limited region of the RNA substrates surrounding the ligation site (UA↓G), the 7Q10 DNA enzyme is completely general for creating 2′–5′ RNA linkages. A brief summary of these results is shown in Figure 8. Our findings indicate that 7Q10 has significant promise for utility as a general RNA ligase. Although mechanistic analysis is not the primary focus here, the available data (e.g. Figure 4) also indicate that further biochemical and structural investigations of deoxyribozymes such as 7Q10 will be fruitful for exploring the underlying principles of catalysis by nucleic acids.

Practical considerations on using the 7Q10 deoxyribozyme for RNA ligation

The 7Q10 deoxyribozyme is a general RNA ligase with a relatively simple sequence requirement at the

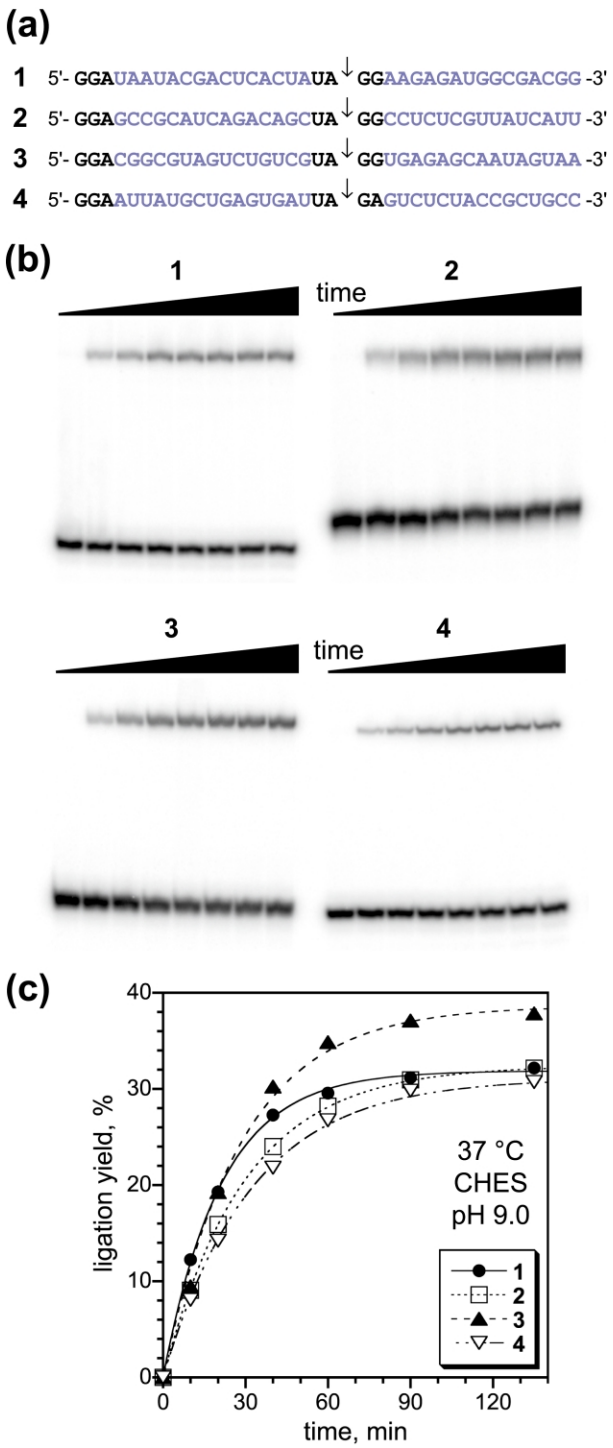

Figure 7. Demonstrating the RNA substrate generality of the 7Q10 deoxyribozyme. (a) RNA substrate sequence combinations **1**–**4** that were tested for ligation by 7Q10. The central UA↓GR region was kept constant, while the flanking nucleotides (blue) were varied. In all cases, the deoxyribozyme-binding arms were changed to maintain Watson–Crick complementarity with the substrates. In sequences **1**-**3**, R = G, while in **4**, R = A to avoid having a series of G nucleotides at the start of the right-hand substrate (which could cause problems during transcription[39]). (b) Ligation activity for the various RNA substrate combinations with their complementary 7Q10 deoxyribozymes (37 °C, Ches, pH 9.0). Timepoints for each at 0, 10, 20 40, 60, 90, 135, and 180 minutes. (c) Kinetic plots. k_{obs}, min^{-1}: **1**, 0.047; **2**, 0.033; **3**, 0.034; **4**, 0.031. For each data set, the 180 minute point was indistinguishable in yield from the 135 minute point and is therefore omitted from the plot for clarity. Note that the data for **1** in this panel are from nominally identical ligation reactions as for the 37 °C data in Figure 5(b). The minor differences in rate and yield between the data sets are representative of the level of reproducibility between experiments.

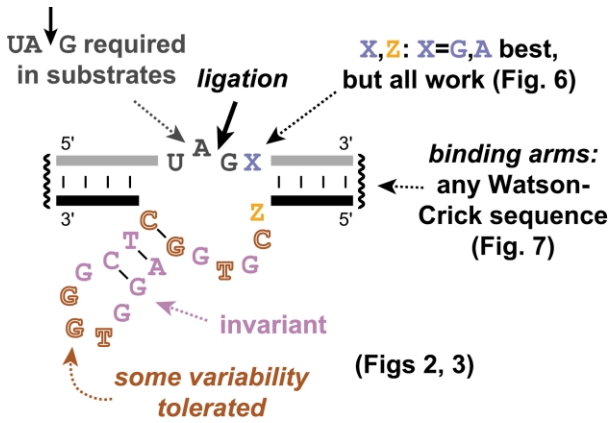

Figure 8. Summary of the nucleotide requirements for the 7Q10 deoxyribozyme and its RNA ligation substrates. This summary is based on the data shown in Figures 2, 3, 6, and 7. It was obviously impractical to test all conceivable mutations of the 16 nucleotide enzyme region of 7Q10. The indication of tolerance to variability in this region is based on the mutations that were experimentally tested.

ligation junction. Specifically, UA↓GR is best; UA↓GY is acceptable; and the flanking RNA nucleotides may be anything at all. On this basis alone, we anticipate that 7Q10 will find immediate use in practical RNA ligations. Many RNA substrates for ligation are prepared by *in vitro* transcription, and these typically begin 5′-GG⋯ or 5′-GA⋯, with the 5′-triphosphate readily removed using a phosphatase. These are precisely the RNA sequences that offer the highest ligation yields using the 7Q10 deoxyribozyme.

The general applicability of the 7Q10 deoxyribozyme for RNA ligation is limited by three factors: (1) the relatively modest ligation yield of ≥30%; (2) the requirement that the RNA substrates have sequence UA↓G at the ligation junction; and (3) the formation of unnatural 2′–5′ linkages *versus* native 3′–5′ linkages. The first two limitations on yield and substrate sequence may not be critical for any given application, and the modest yield is certainly better than nothing at all if other ligation approaches fail. The suboptimal yield may reflect formation of incorrect structures during complexation of the DNA enzyme with its RNA substrates, but the available data are insufficient to evaluate

fully this possibility. Ongoing efforts in our laboratory focus on identification of deoxyribozymes with improved yields and expanded RNA substrate sequence tolerance. The third limitation, formation of non-native linkages, also may not matter in certain situations.[28] The 2′–5′ linkage may even be advantageous if the newly formed bond is otherwise susceptible to nuclease cleavage or if one wishes to test the relevance of a 3′–5′ linkage at a particular site. We are currently pursuing strategies aimed at identifying deoxyribozymes that create native 3′–5′ RNA linkages. All of these efforts will be reported in due course.

If particularly large RNAs are to be ligated but incubation with Mg^{2+} at pH 9.0 is undesirable due to non-specific degradation, then the ligation reactions may instead be performed at pH 7.5. The incubations will need to be left overnight (at least 12 hours) for the best yields (Figure 4), but in our experience this is a routine and acceptable length of time. Because the $K_{d,app}(Mg^{2+})$ for 7Q10 is ~20 mM,[29] whereas all of the experiments reported here were performed at 40 mM Mg^{2+} (i.e. at just twice the $K_{d,app}$), it is also likely that the ligation rate will increase at even higher [Mg^{2+}]. This is useful only if the RNAs of interest do not degrade under the particular incubation conditions, with specific attention to the pH. As usual, it would be prudent to perform small-scale tests before committing large amounts of valuable RNA to a ligation reaction under any incubation conditions.

During certain ligation assays using 7Q10, we have made some observations that are difficult to explain. For example, two particular left-hand RNA substrates that differ only in their 5′-phosphorylation state (monophosphate *versus* triphosphate) provided ligation yields that reproducibly differed by about 5% in side-by-side assays under nominally identical conditions (data not shown). The detailed explanation is not yet known, but the different yields may correlate sensitively with the energy of the DNA–RNA interaction in a way that is difficult to predict. Considering the inexpensive nature of DNA relative to most RNA substrates, for practical use of the 7Q10 deoxyribozyme it would usually be worthwhile to make each binding arm fairly long; e.g. with binding $\Delta G^{o\prime}$ of ≥15 kcal/mol. A similar consideration was advanced regarding RNA-cleaving DNA enzymes.[32] Any minor energetic effect, e.g. that related to the 5′-phosphorylation state of the substrate, would thus be negligible relative to the significant excess of total DNA/RNA binding energy. Increasing the binding arms' lengths may have side benefits in cases where hybridization of the DNA disrupts RNA secondary structure near the ligation site, *via* what is colloquially termed the "disruptor oligo" effect. That is, to the extent that nearby RNA secondary structure interferes with deoxyribozyme activity, disruption of this structure by DNA binding should be advantageous for the ligation reaction.

Relationship of deoxyribozyme-mediated RNA ligation to protein-mediated ligation

RNA ligation mediated by T4 DNA ligase has been the standard approach for joining RNA fragments for over a decade.[33,34] Even so, there are recognized limitations to this protein-mediated approach. In particular, sometimes very low ligation yields are obtained for a given combination of RNA substrates,[15,35–37] and we suspect that such observations are frequently unreported as "negative results". To the extent that protein-mediated ligations fail due to idiosyncracies of the ligase enzyme, rather than failure to form the required DNA:RNA:RNA trimolecular complex, we anticipate that deoxyribozymes will be a valuable counterpart to protein enzymes for RNA ligation. For this to be true, the limitations on yield, substrates, and linkage mentioned above must eventually be overcome, and studies along these lines are in progress in our laboratory. Despite their current limitations, deoxyribozymes like 7Q10 are highly promising as mechanistically interesting and generally useful reagents for RNA ligation.

Materials and Methods

Kinetics assays

All of the kinetics assays used the trimolecular format shown in Figure 1(c). The ^{32}P-radiolabeled left-hand RNA substrate L was the limiting reagent relative to the right-hand substrate R and deoxyribozyme E (the ratio L:E:R was ~1:3:6 to 1:10:30, with the concentration of E equal to ~0.5–3 μM). Increasing the concentration of E or R (or both) did not significantly change the observed kinetics or yields, indicating that the observed yields were not limited by availability of E or R. See our earlier report for a detailed description of the sample preparations and method of analysis.[28] Values of k_{obs} and final yield were obtained by fitting the yield *versus* time data directly to first-order kinetics; i.e. yield(%) = $100 \times Y(1 - e^{-kt})$, where $k = k_{obs}$ and Y = final yield in %. Note that in all of the curve fits, the data clearly level off at a final yield less than 100%, such that the fitted value of Y is meaningful.

Two alternative radiolabeled left-hand RNA substrates were used in the kinetics assays. The assays in the first section of Results were performed using a shorter (5′-^{32}P)-radiolabeled left-hand RNA substrate that had high specific activity but gave slighly reduced overall ligation rates and yields relative to longer left-hand substrates. This shorter substrate was prepared by (5′-^{32}P)-radiolabeling of an RNA oligonucleotide and subsequent cleavage by a 10–23 deoxyribozyme to provide the requisite 2′,3′-cyclic phosphate.[28] Comparisons of relative rates and yields among the various deoxyribozymes should not be impaired for this shorter substrate. In subsequent studies as described in later sections, slightly longer left-hand substrates were used that gave moderately higher ligation yields. These longer substrates were prepared by T7 RNA polymerase transcription from DNA templates,[38] and they incorporated internal ^{32}P-radiolabels from [α-^{32}P]UTP (the 2′,3′-cyclic phosphate was provided by a 3′-terminal HDV ribozyme).

These substrates were longer because they had three additional nucleotides (5′-GGA) on their 5′-end to permit transcription, but they were of lower specific activity due to the source of the ^{32}P and level of incorporation. We estimate that only 12% of transcripts have at least one ^{32}P, *versus* ~100% for 5′-radiolabeling. It was observed that the transcript substrates had generally higher ligation yields (e.g. 30–40%) when compared with the shorter 5′-^{32}P-radiolabeled substrate (e.g. 25–30%; this is a separate issue from the 5% difference in yield for various substrates as mentioned in Discussion). We attribute this difference in yield, in part, to tighter binding of the transcript substrates with the deoxyribozyme due to their increased length, although this has not been proven. In the assays described here, the relatively short left-hand substrate was used to acquire the data for Figures 1–3, and the longer substrate was used for the data in Figures 4–7.

Design of RNA substrates for sequence-dependence experiments of Figure 7

The various RNA substrates shown in Figure 7(a) were designed by taking the parent substrate combination **1** and systematically altering the nucleotides. The substrate combination **2** was obtained from **1** by the interconversions G ↔ U and A ↔ C, except for the leading 5′-GGA (required for T7 RNA polymerase transcription) and the UA↓GG (required for deoxyribozyme ligation). Combination **3** was obtained from **1** by transitions U ↔ C and A ↔ G, and **4** from **3** by the same interconversions as for **2** from **1**. For **3** and **4**, the third and fourth nucleotides to the right of the ligation junction were designed as illustrated to avoid having a stretch of ≥3 G nucleotides at the 5′-end of the substrate.[39] For the same reason, the second nucleotide to the right of the junction for **4** was A and not G. Either purine at this position permits ligation with equivalent yield, as denoted by the UA↓GR motif (Figure 6).

Acknowledgements

This research was supported by the Burroughs Wellcome Fund (New Investigator Award in the Basic Pharmacological Sciences to S.K.S.), the March of Dimes Birth Defects Foundation (Research grant no. 5-FY02-271 to S.K.S.), the National Institutes of Health (GM-65966 to S.K.S.), and the UIUC Department of Chemistry. Acknowledgement is made to the donors of The Petroleum Research Fund, administered by the American Chemical Society, for partial support of this research (38803-G4 to S.K.S.). We thank Amber F. Charlebois and other members of the Silverman laboratory for discussions.

References

1. Musier-Forsyth, K., Usman, N., Scaringe, S., Doudna, J., Green, R. & Schimmel, P. (1991). Specificity for aminoacylation of an RNA helix: an unpaired, exocyclic amino group in the minor groove. *Science*, **253**, 784–786.
2. Pieken, W. A., Olsen, D. B., Benseler, F., Aurup, H. & Eckstein, F. (1991). Kinetic characterization of ribonuclease-resistant 2′-modified hammerhead ribozymes. *Science*, **253**, 314–317.
3. SantaLucia, J., Jr, Kierzek, R. & Turner, D. H. (1992). Context dependence of hydrogen bond free energy revealed by substitutions in an RNA hairpin. *Science*, **256**, 217–219.
4. Paolella, G., Sproat, B. S. & Lamond, A. I. (1992). Nuclease resistant ribozymes with high catalytic activity. *EMBO J.* **11**, 1913–1919.
5. Herschlag, D., Eckstein, F. & Cech, T. R. (1993). Contributions of 2′-hydroxyl groups of the RNA substrate to binding and catalysis by the tetrahymena ribozyme. An energetic picture of an active site composed of RNA. *Biochemistry*, **32**, 8299–8311.
6. Grasby, J. A., Butler, P. J. G. & Gait, M. J. (1993). The synthesis of oligoribonucleotides containing O^6-methylguanosine: the role of conserved guanosine residues in hammerhead ribozyme cleavage. *Nucl. Acids Res.* **21**, 4444–4450.
7. Tuschl, T., Ng, M. M., Pieken, W., Benseler, F. & Eckstein, F. (1993). Importance of exocyclic base functional groups of central core guanosines for hammerhead ribozyme activity. *Biochemistry*, **32**, 11658–11668.
8. Fu, D.-J., Rajur, S. B. & McLaughlin, L. W. (1993). Importance of specific guanosine N^7-nitrogens and purine amino groups for efficient cleavage by a hammerhead ribozyme. *Biochemistry*, **32**, 10629–10637.
9. Hamy, F., Asseline, U., Grasby, J., Iwai, S., Pritchard, C., Slim, G. *et al.* (1993). Hydrogen-bonding contacts in the major groove are required for human immunodeficiency virus type-1 tat protein recognition of TAR RNA. *J. Mol. Biol.* **230**, 111–123.
10. Strobel, S. A., Cech, T. R., Usman, N. & Beigelman, L. (1994). The 2,6-diaminopurine riboside·5-methylisocytidine wobble base pair: an isoenergetic substitution for the study of G·U pairs in RNA. *Biochemistry*, **33**, 13824–13835.
11. Abramovitz, D. L., Friedman, R. A. & Pyle, A. M. (1996). Catalytic role of 2′-hydroxyl groups within a group II intron active site. *Science*, **271**, 1410–1413.
12. Liu, Q., Green, J. B., Khodadadi, A., Haeberli, P., Beigelman, L. & Pyle, A. M. (1997). Branch-site selection in a group II intron mediated by active recognition of the adenine amino group and steric exclusion of non-adenine functionalities. *J. Mol. Biol.* **267**, 163–171.
13. Hamm, M. L. & Piccirilli, J. A. (1997). Incorporation of 2′-deoxy-2′-mercaptocytidine into oligonucleotides *via* phosphoramidite chemistry. *J. Org. Chem.* **62**, 3415–3420.
14. Verma, S. & Eckstein, F. (1998). Modified oligonucleotides: synthesis and strategy for users. *Annu. Rev. Biochem.* **67**, 99–134.
15. Silverman, S. K. & Cech, T. R. (1999). Energetics and cooperativity of tertiary hydrogen bonds in RNA structure. *Biochemistry*, **38**, 8691–8702.
16. Silverman, S. K. & Cech, T. R. (1999). RNA tertiary folding monitored by fluorescence of covalently attached pyrene. *Biochemistry*, **38**, 14224–14237.
17. Silverman, S. K., Deras, M. L., Woodson, S. A., Scaringe, S. A. & Cech, T. R. (2000). Multiple folding pathways for the P4–P6 RNA domain. *Biochemistry*, **39**, 12465–12475.
18. Silverman, S. K. & Cech, T. R. (2001). An early transition state for folding of the P4–P6 RNA domain. *RNA*, **7**, 161–166.

19. Young, B. T. & Silverman, S. K. (2002). The GAAA tetraloop–receptor interaction contributes differentially to folding thermodynamics and kinetics for the P4–P6 RNA domain. *Biochemistry*, **41**, 12271–12276.
20. Scaringe, S. A., Wincott, F. E. & Caruthers, M. H. (1998). Novel RNA synthesis method using 5′-*O*-silyl-2′-*O*-orthoester protecting groups. *J. Am. Chem. Soc.* **120**, 11820–11821.
21. Earnshaw, D. J. & Gait, M. J. (1998). Modified oligoribonucleotides as site-specific probes of RNA structure and function. *Biopolymers*, **48**, 39–55.
22. Scaringe, S. A. (2000). Advanced 5′-silyl-2′-orthoester approach to RNA oligonucleotide synthesis. *Methods Enzymol.* **317**, 3–18.
23. Scaringe, S. A. (2001). RNA oligonucleotide synthesis *via* 5′-silyl-2′-orthoester chemistry. *Methods*, **23**, 206–217.
24. Breaker, R. R. & Joyce, G. F. (1994). A DNA enzyme that cleaves RNA. *Chem. Biol.* **1**, 223–229.
25. Breaker, R. R. (1997). *In vitro* selection of catalytic polynucleotides. *Chem. Rev.* **97**, 371–390.
26. Breaker, R. R. (2000). Making catalytic DNAs. *Science*, **290**, 2095–2096.
27. Lu, Y. (2002). DNAzymes—a new class of enyzmes with promise in biochemical, pharmaceutical, and biotechnological applications. *Chem. Eur. J.* **8**, 4589–4596.
28. Flynn-Charlebois, A., Wang, Y., Prior, T. K., Rashid, I., Hoadley, K. A., Coppins, R. L. *et al.* (2003). Deoxyribozymes with 2′-5′ RNA ligase activity. *J. Am. Chem. Soc.* **125**, 2444–2454.
29. Flynn-Charlebois, A., Prior, T. K., Hoadley, K. A. & Silverman, S. K. (2003). *In vitro* evolution of an RNA-cleaving DNA enzyme into an RNA ligase switches the selectivity from 3′-5′ to 2′-5′. *J. Am. Chem. Soc.* **125**, 5346–5350.
30. Wang, Y. & Silverman, S. K. (2003). Deoxyribozymes that synthesize branched and lariat RNA. *J. Am. Chem. Soc.* **125**, 6880–6881.
31. Santoro, S. W. & Joyce, G. F. (1997). A general purpose RNA-cleaving DNA enzyme. *Proc. Natl Acad. Sci. USA*, **94**, 4262–4266.
32. Pyle, A. M., Chu, V. T., Jankowsky, E. & Boudvillain, M. (2000). Using DNAzymes to cut, process, and map RNA molecules for structural studies or modification. *Methods Enzymol.* **317**, 140–146.
33. Moore, M. J. & Sharp, P. A. (1992). Site-specific modification of pre-mRNA: the 2′-hydroxyl groups at the splice site. *Science*, **256**, 992–997.
34. Moore, M. J. & Query, C. C. (1998). Use of site-specifically modified RNAs constructed by RNA ligation. In *RNA–Protein Interactions: A Practical Approach* (Smith, C. W. J., ed.), pp. 75–108, Oxford University Press, Oxford.
35. Han, H. & Dervan, P. B. (1994). Visualization of RNA tertiary structure by RNA-EDTA.Fe (II) autocleavage: analysis of tRNA (Phe) with uridine-EDTA.Fe (II) at position 47. *Proc. Natl Acad. Sci. USA*, **91**, 4955–4959.
36. Zahler, A. M. & Roth, M. B. (1995). Distinct functions of SR proteins in recruitment of U1 small nuclear ribonucleoprotein to alternative 5′ splice sites. *Proc. Natl Acad. Sci. USA*, **92**, 2642–2646.
37. Strobel, S. A. & Ortoleva-Donnelly, L. (1999). A hydrogen-bonding triad stabilizes the chemical transition state of a group I ribozyme. *Chem. Biol.* **6**, 153–165.
38. Milligan, J. F. & Uhlenbeck, O. C. (1989). Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.* **180**, 51–62.
39. Pleiss, J. A., Derrick, M. L. & Uhlenbeck, O. C. (1998). T7 RNA polymerase produces 5′ end heterogeneity during *in vitro* transcription from certain templates. *RNA*, **4**, 1313–1317.

Edited by J. Doudna

(Received 24 March 2003; received in revised form 13 May 2003; accepted 13 May 2003)

Supplementary Material for this paper comprising two Figures is available on Science Direct