RNA (1999), *5*:158–166. Cambridge University Press. Printed in the USA.

Specificity from steric restrictions in the guanosine binding pocket of a group I ribozyme

RICK RUSSELL and DANIEL HERSCHLAG
Department of Biochemistry, Stanford University, Stanford, California 94305-5307, USA

ABSTRACT

The 3′ splice site of group I introns is defined, in part, by base pairs between the intron core and residues just upstream of the splice site, referred to as P9.0. We have studied the specificity imparted by P9.0 using the well-characterized L–21 *Sca*I ribozyme from *Tetrahymena* by adding residues to the 5′ end of the guanosine (G) that functions as a nucleophile in the oligonucleotide cleavage reaction: $CCCUCUA_5$ (S) + NNG ⇀↽ CCCUCU + $NNGA_5$. UCG, predicted to form two base pairs in P9.0, reacts with a (k_{cat}/K_M) value ~10-fold greater than G, consistent with previous results. Altering the bases that form P9.0 in both the trinucleotide G analog and the ribozyme affects the specificity in the manner predicted for base-pairing. Strikingly, oligonucleotides incapable of forming P9.0 react ~10-fold more slowly than G, for which the mispaired residues are simply absent. The observed specificity is consistent with a model in which the P9.0 site is sterically restricted such that an energetic penalty, not present for G, must be overcome by G analogs with 5′ extensions. Shortening S to include only one residue 3′ of the cleavage site (CCCUCUA) eliminates this penalty and uniformly enhances the reactions of matched and mismatched oligonucleotides relative to guanosine. These results suggest that the 3′ portion of S occupies the P9.0 site, sterically interfering with binding of G analogs with 5′ extensions. Similar steric effects may more generally allow structured RNAs to avoid formation of incorrect contacts, thereby helping to avoid kinetic traps during folding and enhancing cooperative formation of the correct structure.

Keywords: group I intron; ribozyme; RNA catalysis; RNA folding; self-splicing; *Tetrahymena thermophila*

INTRODUCTION

Self splicing of group I introns involves two consecutive phosphoryl transesterification reactions. An exogenous guanosine (G) serves as the nucleophile for the first step. It attacks the 5′ splice site, which is defined by base-pairing with an internal guide sequence (IGS) in the intron (P1 in Fig. 1A). The 3′ splice site is then attacked by the free 3′-OH of the 5′ exon in the second step to generate ligated exons and linear intron. The 3′ splice site is defined, in the context of the folded intron, by a conserved G at the 3′ terminus of the intron, by P10 base-pairing between the IGS and the 3′ exon, and by the P9.0 pairings between residues just upstream of the 3′ terminal G and the intron core (Michel et al., 1989; Burke et al., 1990; Michel & Westhof, 1990).

In the archetypal group I intron from *Tetrahymena thermophila*, these pairing elements are functionally redundant in vitro; disruptions of P9.0 significantly affect splicing only in the presence of a compromised P10 and vice versa (Been & Cech, 1985; Michel et al., 1989). Because of this redundancy and because of the multistep nature of the splicing reaction, it is difficult to dissect individual effects of the P9.0 and P10 pairings in the generation of splice-site specificity. To circumvent these problems, we have studied the effects of P9.0 on specificity in the context of the well-characterized L–21 *Sca*I ribozyme derivative of the intron, which lacks the first 21 and last 5 residues (Fig. 1B; Zaug et al., 1988; reviewed in Cech & Herschlag, 1997). Oligonucleotides terminating in a 3′-G residue, which mimic the 3′ end of the intron, can form P9.0 *in trans* (Fig. 1B) and participate in a reaction in which the 3′ end of a 5′ splice site mimic is transferred to the G (Tanner & Cech, 1987; Moran et al., 1993; Bevilacqua et al., 1996). The effects of base pairs and mismatches in P9.0 can then be assayed by the relative reactivity of 3′ splice site mimics.

We have found that oligonucleotides that form P9.0 are somewhat more reactive than G, consistent with previous results (Moran et al., 1993). Contrary to initial expectations, however, oligonucleotides that form mismatches in P9.0 are less reactive than G, which is simply missing the mispaired residues. The specificity results from steric hindrance within this structured RNA binding site. Appropriate positioning of a steric block

Reprint requests to: Daniel Herschlag, Department of Biochemistry, B400 Beckman Center, Stanford University, Stanford, California 94305-5307, USA; e-mail: herschla@cmgm.stanford.edu.

Abbreviations: IGS, internal guide sequence; K_d, equilibrium dissociation constant; MOPS, 3-(*N*-morpholino)propanesulfonic acid; S, oligonucleotide 5′-splice site mimic.

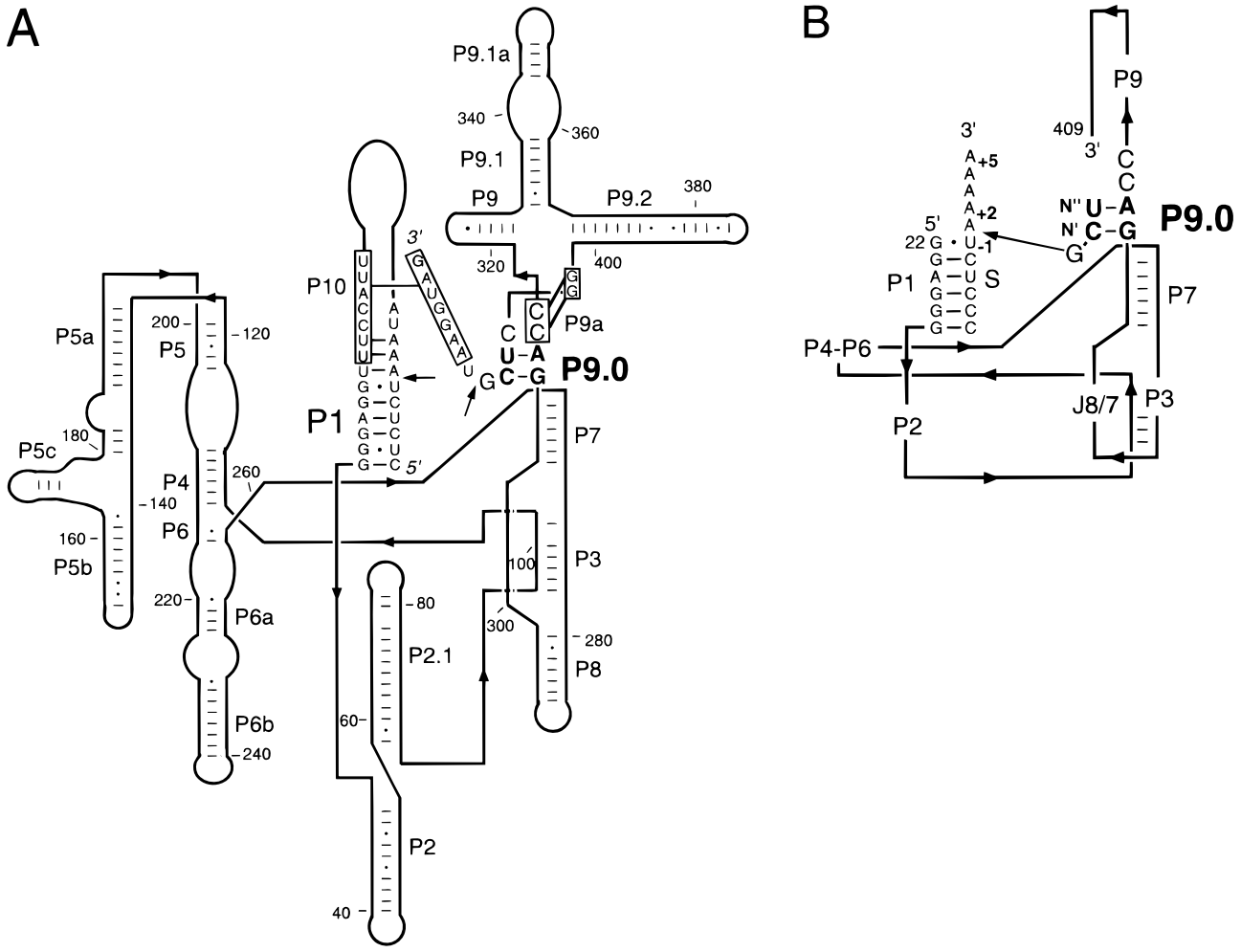

FIGURE 1. Secondary structure of the *Tetrahymena* group I intron and ribozyme. **A**: Secondary structure of the intron, with P9.0 labeled in bold. Residues that pair to form P10 are boxed, and the 5′ and 3′ splice sites are indicated with arrows. The secondary structure is shown in an orientation reflecting the tertiary subdomains of the ribozyme (Cech et al., 1994). **B**: Close-up view of the secondary structure of the L–21 *Sca*I ribozyme with bound substrates. P9.0 pairings are shown between an oligonucleotide 3′ splice site mimic provided *in trans* and residues G313 and A314 of the ribozyme. Positions one and two residues 5′ of the G in the 3′ splice site mimic are denoted N′ and N″, respectively. The cleavage site of the 5′ splice site mimic (S) is indicated with an arrow, and residues on S are numbered.

within a binding cavity may be a general strategy by which large RNAs can discriminate against incorrect ligands and incorrect folding events.

RESULTS

Characterization of specificity mediated by P9.0

To examine the effects of P9.0 on specificity, trinucleotides with 3′ terminal G residues were examined in the ribozyme-catalyzed reaction of Equation 1, in which E refers to the L–21 *Sca*I ribozyme (Zaug et al., 1988).

$$E{\bullet}CCCUCdUA_5 + NNG \xrightarrow{(k_{cat}/K_M)^{Nuc}} E{\bullet}CCCUCdU + NNGA_5 \quad (1)$$

This is identical to the well-characterized reaction in which G serves as the nucleophile, cleaving the 5′ splice site mimic (S) to the shorter product that is a 5′ exon mimic (reviewed in Cech et al., 1992; Cech & Herschlag, 1997; Narlikar & Herschlag, 1997), except that residues 5′ of the G have been added. Under single-turnover conditions with a sufficient concentration of ribozyme, the ^{32}P-labeled CCCUCdUA$_5$ ($-$1d,rSA$_5$) is rapidly and completely bound. The dependence of the observed rate constant for cleavage on subsaturating NNG concentration, or $(k_{cat}/K_M)^{Nuc}$, monitors the free energy barrier for going from the E•S complex, with NNG free in solution, to the chemical transition state (Equation 1). A 2′-deoxyribose residue was included at the cleavage site of the 5′ splice site mimic as depicted in Equation 1 ($-$1d,rSA$_5$) to ensure that the chemical step remained rate limiting under all conditions. This

2′-deoxyribose substitution has been shown to slow the rate of the chemical step ~500-fold without significantly affecting S or G binding (Herschlag et al., 1993b; McConnell et al., 1993; Herschlag & Khosla, 1994). In contrast, with an all-ribose substrate ($CCCUCUA_5$) there is a change in rate-limiting step for $(k_{cat}/K_M)^G$ above pH 7 (Herschlag & Khosla, 1994); this was also found to be the case for reactions of NNG trinucleotides (R. Russell & D. Herschlag, unpubl. results).

The trinucleotide 5′ UCG, predicted to form two P9.0 base pairs (Fig. 1B), reacted ~10-fold faster than G at equivalent subsaturating concentrations (Fig. 2A). The stabilization is in the same range as previously reported (Moran et al., 1993). The observed rate constants in the presence of saturating UCG or G were identical (data not shown), suggesting that the observed stabilization by UCG is because of tighter binding to the E•S complex in the ground state.

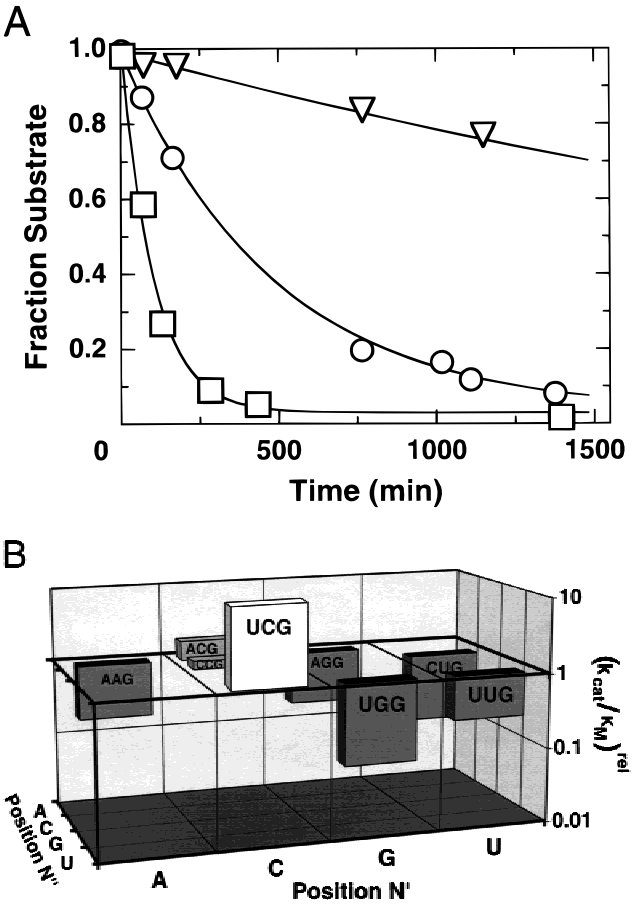

FIGURE 2. Single-turnover reactions of trinucleotide 3′ splice site mimics that form pairs or mismatches in P9.0 with E (200 nM) saturating with respect to ~0.1 nM 5′ end labeled −1d,rSA_5 (50 mM Na•MOPS, pH 6.8, 10 mM $MgCl_2$, 30 °C). **A**: Reactions of 5 μM G (○), UCG (□), and CUG (▽). Fitting first-order curves to the data yielded rate constants of $k_{obs}^{G} = 2.1 \times 10^{-3}$ min^{-1}, $k_{obs}^{UCG} = 9.8 \times 10^{-3}$ min^{-1}, and $k_{obs}^{CUG} = 2.4 \times 10^{-4}$ min^{-1}. First-order rate constants were plotted against the concentration of 3′ splice site mimic to determine $(k_{cat}/K_M)^{Nuc}$ (see Methods), which was compared with the value for G to give $(k_{cat}/K_M)^{rel}$. Seven independent determinations yielded a $(k_{cat}/K_M)^{rel}$ value of 9.0 ± 1.1 for UCG [$(k_{cat}/K_M)^G = 280$ M^{-1} min^{-1}]. Six independent determinations yielded $(k_{cat}/K_M)^{rel,\ CUG} = 0.12 \pm 0.01$. **B**: Second-order rate constants for the reaction of various trinucleotides are shown as $(k_{cat}/K_M)^{rel}$, relative to $(k_{cat}/K_M)^G$. Nomenclature for positions upstream from G: $_{5'}$N″N′G$_{3'}$ (Fig. 1B).

The concentration of CUG required for a half-maximal observed rate constant ($K_{1/2}$) was similar to that for G, but the maximal rate was ~6-fold less than for G (data not shown). The simplest interpretation of these results is that the decreased reactivity results primarily from a slower chemical step, but an inhibitory effect at high concentrations of CUG could instead produce a false plateau in the concentration dependence. Thus, an effect primarily on binding instead of the maximal rate cannot be eliminated. This ambiguity does not affect $(k_{cat}/K_M)^{Nuc}$, which requires low concentrations of CUG and measures the free energy difference between the E•S complex with CUG free in solution and the E•S•CUG ternary complex in the transition state (Fersht, 1985). We therefore focused on effects of changing the identity of the 3′ splice site mimic on $(k_{cat}/K_M)^{Nuc}$ in subsequent experiments.

The values of $(k_{cat}/K_M)^{Nuc}$ for several other trinucleotides were determined (Fig. 2B and Table 1). All trinucleotides assayed in which the residue one position upstream of G, referred to as position N′ (see Fig. 1B), was not C reacted with similar low $(k_{cat}/K_M)^{Nuc}$ values, suggesting that the penalty for mismatches is general. Both ACG and CCG, which can form a base pair at position N′ but not at position N″, reacted significantly faster than trinucleotides that cannot form either base pair, but more slowly than UCG. Thus, base pairs at each position, N′ and N″, contribute stabilization in the transition state.

The reactivity of dinucleotides was also examined to assess the relative contributions of positions N′ and N″ to specificity. CG reacted with a $(k_{cat}/K_M)^{Nuc}$ value similar to that of G [$(k_{cat}/K_M)^{rel} = 1.5$, Table 1]. In contrast, dinucleotides with any other residue at position N′ reacted less efficiently than G, with $(k_{cat}/K_M)^{rel}$ values ranging from 0.20 (UG) to 0.54 (AG). Therefore, a mismatched residue at position N′ is destabilizing relative to its absence.

In contrast, there is little penalty for adding an additional mismatch at position N″ to a mismatched dinucleotide. For example $(k_{cat}/K_M)^{rel}$ for CUG was 0.12, compared with 0.20 for UG (Table 1). Similarly, $(k_{cat}/K_M)^{rel}$ for AGG was 0.20, compared with 0.26 for GG.

TABLE 1. Reactivity of 3′ splice site mimics[a]

3′ splice site mimic[b]	$(k_{cat}/K_M)^{rel}$	$\Delta\Delta G^{\ddagger}_G$ (kcal/mol)[c]	$\Delta\Delta G^{\ddagger}_M$ (kcal/mol)[d]
G	**(1)**	**(0)**	
CG	**1.5**	**–0.2**	**(0)**
AG	0.54	+0.4	+0.6
GG	0.26	+0.8	+1.0
UG	0.20	+1.0	+1.2
UCG	**9.0**	**–1.3**	**(0)**
CCG	1.3	−0.2	+1.1
ACG	1.8	−0.4	+0.9
UUG	0.27	+0.8	+2.1
UGG	0.085	+1.5	+2.8
CUG	0.12	+1.3	+2.6
AGG	0.20	+1.0	+2.3
AAG	0.18	+1.0	+2.3
GUCG	**13**	**–1.5**	**(0)**
AUCG	22	−1.8	−0.3
CUCG	18	−1.7	−0.2
UUCG	10	−1.4	+0.1

[a]50 mM Na•MOPS, pH 6.8, 10 mM $MgCl_2$, 30 °C.
[b]Residues predicted to form P9.0 base pairs are underlined. The completely matched 3′ splice site mimic is shown at the top of each group and is entirely in bold.
[c]The free energy differences for reactions of 3′ splice site mimics compared with that of G [$(k_{cat}/K_M)^G = 280\ M^{-1}\ min^{-1}$; Fig. 2].
[d]Free energy differences for reactions of mismatched 3′ splice site mimics compared with those of the matched 3′ splice site mimic of the same length (in bold).

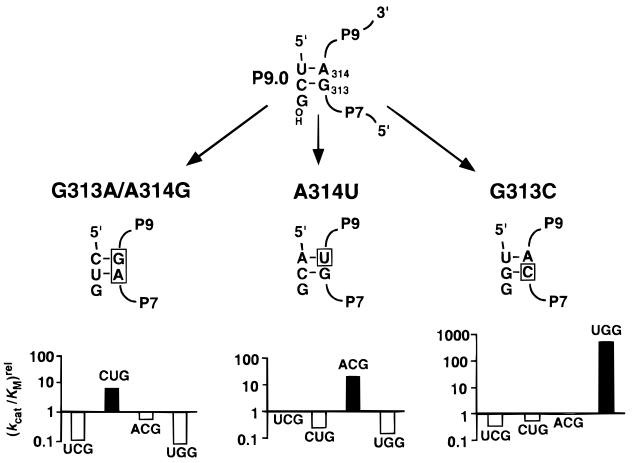

FIGURE 3. Reactivity of ribozymes with altered P9.0 sequences. Residues changed are boxed, and the 3′ splice site mimic predicted to form P9.0 pairings with each variant ribozyme is shown. Reactivity of trinucleotides is shown as $(k_{cat}/K_M)^{rel}$, relative to $(k_{cat}/K_M)^G$ for each mutant. The value of $(k_{cat}/K_M)^{rel}$ for each trinucleotide predicted to form P9.0 is shown as a black bar, whereas values for 3′ splice site mimics that cannot form P9.0 base pairs are shown as open bars.

There is also little penalty for adding a mismatched residue to a matched dinucleotide, as adding A or C to CG had no significant effect on $(k_{cat}/K_M)^{rel}$ (<20% changes, Table 1). Thus, although there is a substantial energetic penalty for adding a mismatched residue at position N′, there is little or no penalty for adding a mismatched residue at position N″.

Specificity switches in variant ribozymes

If the P9.0 pairings were responsible for the observed specificity, then altering the ribozyme core residues involved in P9.0 would be predicted to change the specificity. Three variant ribozymes were constructed (Fig. 3). The two residues were reversed (variant G313A/A314G) or each residue was separately changed to its pyrimidine base-pairing partner (variants A314U and G313C). There were at most small effects on reactions of these variant ribozymes with G. Values of $(k_{cat}/K_M)^G$ and the maximal rates for the variants A314U and G313A/A314G were the same as the wild type within error (data not shown). For the G313C variant, $(k_{cat}/K_M)^G$ was ~3-fold less than for the wild-type ribozyme, primarily because of weaker binding of G (data not shown; experiments were carried out under conditions such that the $K_{1/2}$ for G equals K_d^G; McConnell et al., 1993).

To determine the specificities of these variant ribozymes, $(k_{cat}/K_M)^{Nuc}$ values were determined with a series of trinucleotides (Fig. 3). As expected, the oligonucleotide predicted to form P9.0 reacted more efficiently than G with each ribozyme. The penalty for mismatches relative to G was also maintained. For example, UCG reacted in the context of the G313A/A314G ribozyme with a $(k_{cat}/K_M)^{rel}$ value of 0.10, essentially the same as the value of 0.12 for CUG with the wild-type ribozyme.

Curiously, UGG, the matched oligonucleotide for the variant G313C, reacted with a $(k_{cat}/K_M)^{rel}$ value of 510, whereas $(k_{cat}/K_M)^{rel}$ for UCG with the wild type is 9 (Fig. 3, Table 1). This represents a stabilization of 2.1 kcal/mol from the reversal of the C•G base pair at position N′. This effect is not simply accounted for by stacking with the neighboring U•A base pair, because model studies predict a destabilizing effect for this base pair reversal of 0.5 kcal/mol (Serra & Turner, 1995). The efficient reaction of UGG seems to be primarily a result of extraordinarily tight binding, because K_d was found to be 0.8 μM ($K_d^G = 360\ \mu$M for this variant; for the wild type, $K_d^{UCG} = 15\ \mu$M and $K_d^G = 120\ \mu$M; each determined under conditions such that $K_{1/2} = K_d$; data not shown). This tight binding was also observed in reactions of the dinucleotide GG with the G313C variant ribozyme ($K_d^{GG} = 3\ \mu$M). It is noteworthy that the G313C variant also displays a higher level of specificity for matched versus mismatched trinucleotides than the wild-type ribozyme: ~1,000-fold compared with ~100-fold for the wild type (Fig. 3 and Fig. 2B).

Determination of the number of P9.0 pairings

It was previously suggested that an additional pairing might be formed between a G three residues upstream

from the terminal G (position N‴) and C315 of the ribozyme, although the natural intron does not have a G at the corresponding position (Moran et al., 1993; see Fig. 1A). Consistent with previous results (Moran et al., 1993; Profenno et al., 1997), GUCG reacted more efficiently than UCG (Table 1). However, oligonucleotides with A or C at this position reacted at least as well, indicating that this effect is not specific to a G at position N‴. Likewise, CCUG and GCUG both reacted ~2-fold more efficiently than CUG in the context of the G313A/A314G variant ribozyme (data not shown). Rather than being stabilized by an additional base pair, it appears that a fourth residue contributes via a more general effect, perhaps a stacking interaction. Thus, the P9.0 pairings are likely to be limited to two positions upstream from G. C315 is presumably involved in other interactions or is geometrically constrained such that it is unavailable to base-pair with the residue at position N‴. This result supports a recent three-dimensional model of the ribozyme (Lehnert et al., 1996), in which C315 is proposed to pair with G406, forming one of two base pairs referred to as P9a (Fig. 1A).

Physical origin of specificity

To learn more about the origin of the discrimination against mismatches, the dependence of the energetic penalty on steric bulk of mismatched residues in 3′ splice site mimics was examined. If the guanosine binding pocket were sufficiently tight such that a mispaired residue in P9.0 resulted in unfavorable contacts, then a larger residue could have a larger penalty. 2′ Methoxy groups were therefore added at positions N′ and N″ to CUG, a mismatched 3′ splice site mimic. This oligonucleotide reacted only ~2-fold more slowly than CUG [$(k_{cat}/K_M)^{rel} = 0.069$, compared with 0.12 for CUG], and the effect was similar to that of removing the 2′ hydroxyl groups from CUG to give d(CU)G [$(k_{cat}/K_M)^{rel} = 0.058$]. A decrease in reactivity was also observed upon removal of the 2′-OH groups from UCG (data not shown), similar to previous results on ground-state binding of UCG versus d(TCG) (Moran et al., 1993). To reduce the bulk of the 3′ splice site mimic, the upstream residues were replaced with abasic nucleotides (XXG). $(k_{cat}/K_M)^{rel}$ for XXG was 0.036, an energetic penalty of 2.0 kcal/mol relative to G, and similar to d(CU)G (the abasic residues contained deoxyribose). In addition, pyrimidine mismatches reacted equivalently to purine mismatches (Fig. 2B). Thus, no correlation is observed between the bulk of mismatched residues and $(k_{cat}/K_M)^{rel}$. These results suggest that the discrimination is unlikely to arise from an extraordinarily tight binding pocket that surrounds the 3′ splice site mimic.

Nevertheless, a structural model of the ribozyme (Michel & Westhof, 1990; Lehnert et al., 1996) suggests that bound guanosine and upstream residues may be present in a sterically restricted environment (Fig. 4). Guanosine binds to a "floor" present in the P3–P8 domain. There may be three "walls" that confine the upstream residues (red in Fig. 4). One wall is formed by residues G313 and A314 that form P9.0, another is formed by the P4–P6 domain, and a third may be formed by the 3′-A_5-tail of S. We postulated that if a sterically restricted binding pocket contributed to the discrimination, then removal of one of the walls might reduce the penalty. Reduction of the A_5 tail of S to a methyl group or to a single adenosine (−1d,rSA) entirely eliminated the 1.3-kcal/mol penalty against CUG relative to G (Fig. 5). The reactivity of UCG with −1d,rSA was increased by the same amount as CUG, giving $(k_{cat}/K_M)^{rel} = 100$ for UCG. Thus, the penalty in the presence of the A_5 tail is imposed equivalently on matched and mismatched 3′ splice site mimics, but has no effect on the reactivity of G (data not shown). Systematic variation of the length of the tail showed that the penalty is entirely manifest upon addition of the second A residue (Fig. 5).

These results are most simply explained by a model in which the binding site for the residue upstream from G in the 3′ splice site mimic (position N′) is occluded by A(+2) of S (Fig. 6). A set of 5′ splice site mimics with U tails of varying length behaved identically to the set with A tails, and a 5′ splice site mimic in which A(+2) was replaced by a G (CCCUCdUAGAAA) yielded values of $(k_{cat}/K_M)^{rel}$ for UCG and CUG identical to those obtained with −1d,rSA_5 (data not shown). Thus, a specific base-pairing interaction involving the residue at +2 does not account for the specificity. It is possible that the phosphodiester backbone of the +2 residue contacts the ribozyme, or that no direct contact is made. It is also possible that the +2 residue interacts elsewhere and causes a conformational rearrangement that occludes the P9.0 site. Nevertheless, the finding that the P9.0 site is likely to be occluded by the +2 residue of S is consistent with previously published models (Michel & Westhof, 1990; Lehnert et al., 1996; Fig. 4) and suggests a geometrical constraint for future modeling of the tertiary structure of the ribozyme.

DISCUSSION

Analysis of a series of *Tetrahymena* ribozyme reactions has confirmed that P9.0 can form *in trans* between a 3′ splice site mimic and the ribozyme (Moran et al., 1993), and has extended previous studies by dissecting the contributions of individual base pairs to reactivity. Base pairs at each of two residues upstream from the guanosine nucleophile increase reactivity. Although a third residue provides a modest further increase in reactivity, it does not form a base pair, which is consistent with current structural models (Lehnert et al., 1996). Additionally, these studies have shown that an energetic penalty exists for reactions of guano-

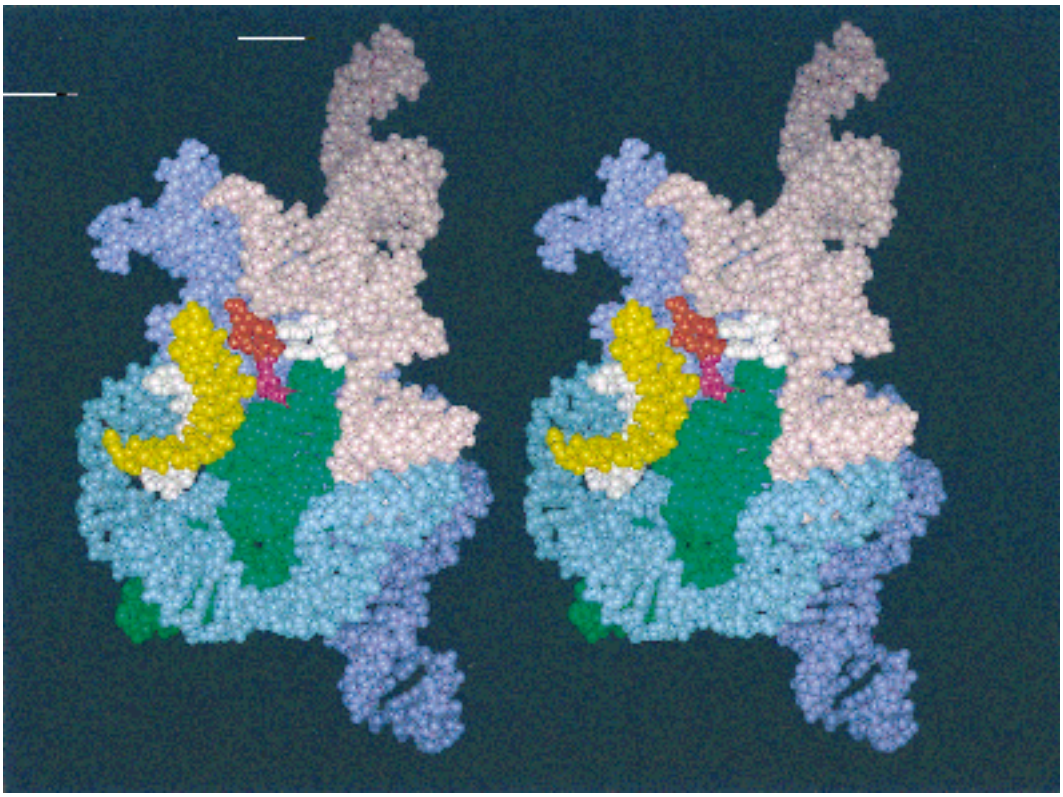

FIGURE 4. The structural model of Lehnert et al. (1996) is consistent with a sterically restricted binding site for an oligonucleotide forming P9.0 (red). The ribozyme residues that form P9.0 are white, and the guanosine nucleophile is magenta. Also shown are the P1 duplex (the 5′ splice site sequence is yellow and the IGS is white), the P3–P8 subdomain (green), the P4–P6 subdomain (blue), the P2 subdomain (cyan), and the P9 subdomain (salmon). The product complex with guanosine covalently attached to the 3′ tail of the 5′ splice site mimic is depicted, as presented by Lehnert et al. (1996).

sine analogs with 5′ extensions that is not present for guanosine alone, apparently resulting from steric occlusion within the restricted P9.0 site (Figs. 4 and 6). As described below, it is possible that folded RNAs more generally obtain specificity from steric occlusion by prepositioned structural elements.

Potential effects of steric blocks by prepositioning are shown in Figure 7 using the P9.0 site as an example. Positioning of a structural element can be imagined as a continuum, and is illustrated by three points along this continuum. For simplicity, we present this structural element as a "wall" in Figure 7. At one extreme, the wall is far enough away from the P9.0 site that it has no effect on binding of either a matched (UCG) or a mismatched (CUG) RNA (Fig. 7A). UCG binds more tightly than CUG because it can form the P9.0 base pairs ($K_A^{UCG} > K_A^{CUG}$; K_A, K_B, and K_C in Fig. 7 represent equilibrium association constants; see legend). Although CUG is expected to occupy more space than UCG because it is not constrained by base-pairing, this does not weaken its binding.

With intermediate positioning of the wall (Fig. 7B), binding of UCG is unaffected because it is held away from the wall by base-pairing interactions ($K_B^{UCG} = K_A^{UCG}$). In contrast, binding of CUG is weakened because the wall restricts the conformational freedom or sterically interferes with bound CUG ($K_B^{CUG} < K_A^{CUG}$).

At the other extreme, the wall is positioned to overlap with the binding site such that it must be moved or distorted to allow binding of either matched or mismatched substrates (Fig. 7C). In the simplest model, the penalties for displacement or distortion of the wall are similar for UCG and CUG, so decreases in their affinities are roughly equal ($K_A^{UCG}/K_C^{UCG} \approx K_A^{CUG}/K_C^{CUG}$). UCG still binds more tightly than CUG because it can form P9.0 after displacing the wall. The situation of Figure 7C describes the results obtained here: the 3′ tail of S decreases the reactivity of matched and mismatched substrates equally, penalizing both relative to the smaller substrate, guanosine.

The steric occlusion described in Fig. 7B provides specificity by affecting the thermodynamics of substrate binding. A steric block that produces an identical penalty against matched and mismatched substrates (Fig. 7C) could nevertheless provide specificity against the mismatched substrate through a kinetic effect, because differences in reactivity between correct and incorrect substrates can be masked if the chemical step is not rate limiting. For example, binding of oligonucleotide 5′ splice site mimics, both matched and contain-

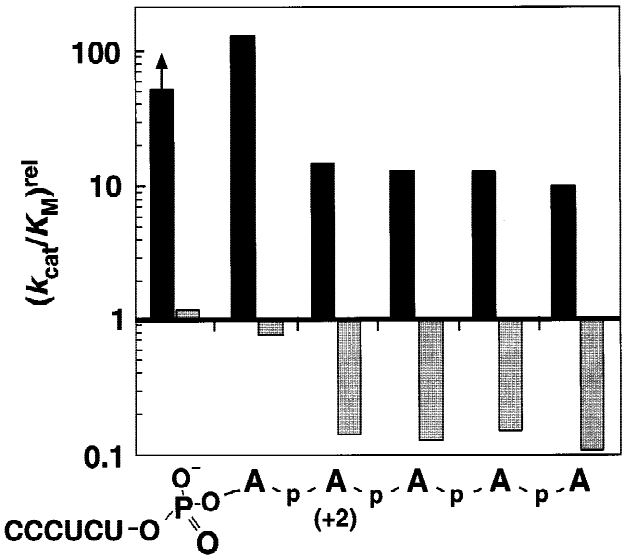

FIGURE 5. Effect of shortening the 3′ tail of the 5′ splice site mimic on reactions with UCG (black) and CUG (gray). Reactivity is relative to G for each substrate (see Methods). The 5′ splice site mimic entirely lacking the 3′ tail contained a methyl group linked to the reactive phosphoryl group; it also contained a ribose residue at position U(−1), rather than a deoxyribose, and reactions were performed at pH 6.0 instead of pH 6.8 to avoid a change to a rate-limiting conformational step (Herschlag & Khosla, 1994). Because $(k_{cat}/K_M)^{UCG}$ with this 5′ splice site mimic was approximately equal to the maximum value imposed by the alternative rate-limiting conformational step, the value of $(k_{cat}/K_M)^{rel}$ represents a lower limit for the effect on the chemical transition state, as indicated by the arrow.

ing a single mismatch, is rate limiting for the L–21 ribozyme reaction, resulting in minimal specificity (Herschlag & Cech, 1990; Herschlag, 1991). Under such circumstances, a structural element that equivalently increases dissociation rates of matched and mismatched substrates can render the chemical step rate limiting, thereby unmasking differences in reactivity. An analogous increase in specificity has been observed for mutations that increase dissociation of 5′ splice site mimics by disrupting the ribozyme structure rather than producing a steric block (Young et al., 1991). A steric penalty that affects matched and mismatched substrates equally (Fig. 7C) may be easier for an RNA enzyme to achieve than a discriminatory one (Fig. 7B), because less precision is required in positioning the steric block.

Although the energetic penalty against G nucleophiles with 5′ extensions observed here with the L–21 *Sca*I ribozyme was uniform for matched and mismatched substrates, this may not hold in the context of the complete intron. In the intact intron, the 3′ tail of S is expected to be constrained in a duplex with the extended IGS, whereas the extended IGS is absent in the L–21 *Sca*I ribozyme (see Fig. 1). It will be of interest to determine whether the extended IGS eliminates the energetic penalty (Fig. 7A), provides a specific penalty against mismatches (Fig. 7B), or gives a uniform effect as observed herein (Fig. 7C).

Specificity in RNA folding

Formation of secondary and tertiary contacts in RNA folding is directly analogous to substrate binding. That is, complexation of an RNA ligand with an RNA enzyme can be considered as an intermolecular model for the

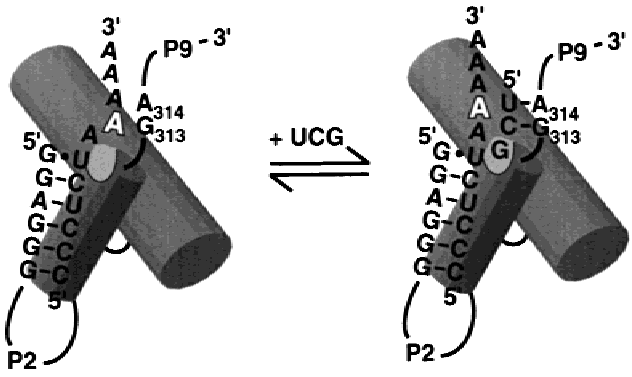

FIGURE 6. Model for steric occlusion of the P9.0 site by the residue at position (+2) of the 5′ splice site mimic (S; numbering as in Fig. 1B). The orientation is similar to Figure 4, with P4–P6 and P3–P8 represented by cylinders. Residue (+2) of S (shown in white) occupies the P9.0 site in the absence of a bound 3′ splice site mimic. The subdomains represented as cylinders are not to scale with respect to the residues shown as letters.

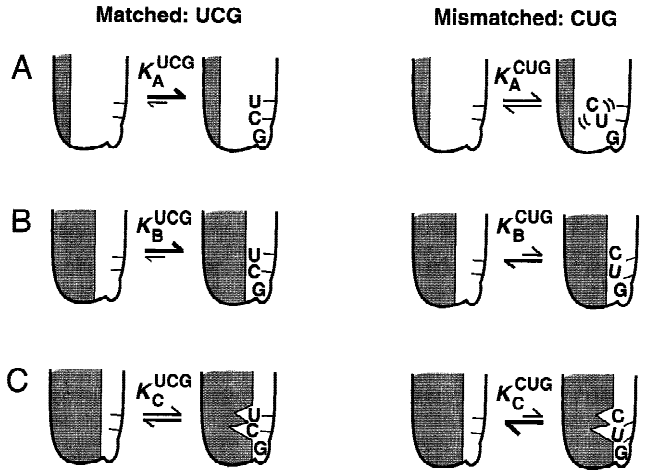

FIGURE 7. A general model for the generation of specificity through steric occlusion, using the P9.0 site as an example. The positioning of one structural element with respect to another may be represented as a continuum, of which three distinct positions are shown. For each position, the binding of UCG and CUG is depicted, and relative affinities are indicated by the thickness and length of arrows. For each substrate, an equilibrium association constant (e.g. K_A^{UCG}) is represented with the subscript indicating the wall position (**A**–**C**). For clarity, free UCG and CUG are not shown explicitly. The structural element providing a steric occlusion is depicted as a rigid stationary wall for simplicity. A less rigid, more mobile element would be expected to produce qualitatively the same effects, with smaller magnitudes and less precisely defined boundaries separating the positions shown in **A**–**C**.

folding of an RNA molecule. Thus, steric effects can play a role in the specificity of folding processes, analogous to the role in substrate binding described herein. These steric effects may increase cooperativity and avoid kinetic traps in folding, as described below.

For a large structured RNA to be functional, a single native structure must be thermodynamically favored over the large ensemble of nonnative structures; that is, folding must be cooperative. A native contact in a folded RNA could position a structural element to provide a steric effect analogous to that described in Figure 7B, interfering with formation of a nonnative contact while allowing an additional native contact to form. Thus, steric occlusion could be utilized by RNA to favor the native state over a nonnative one, thereby enhancing the cooperativity of folding. There may even be RNA subdomains that do not provide significant stabilization of the native state relative to the ensemble of unfolded states, but nevertheless enhance cooperative folding by shielding a site from a potential nonnative interaction, thereby destabilizing a specific family of nonnative states. Analogously, steric occlusion could be used to prevent kinetic traps in RNA folding by destabilizing off-pathway intermediates with a high propensity to misfold. This strategy may be especially important for RNA because it is susceptible to kinetic traps (Herschlag, 1995 and references therein; Uhlenbeck, 1995; Thirumalai & Woodson, 1996; Treiber et al., 1998). Finally, uniform penalties similar to that observed here (Fig. 7C) may also facilitate RNA folding by limiting the length of secondary structural elements or the size and shape of tertiary structural elements.

Evolutionary constraints on P9.0

Surprisingly, the G313C variant ribozyme is ~10-fold more adept than the wild type at discriminating between a 3′ splice site mimic that can form P9.0, UGG, and those that cannot form P9.0. This result suggests that the specificity of the P9.0 interaction in the wild-type ribozyme is not maximized. Additionally, UGG binds to the G313C variant ~20-fold more tightly than UCG binds to the wild type, suggesting that the stability of P9.0 is also not maximized. Residues that form P9.0 in the wild type contribute less to stability than predicted from nearest-neighbor rules (Freier et al., 1986; Serra & Turner, 1995). Furthermore, there is no indication of tertiary interactions from 2′-OH groups in P9.0 (see Results and Moran et al., 1993), unlike the P1 duplex, which uses its 2′-OH groups in tertiary interactions with the ribozyme core (Narlikar et al., 1997; Narlikar & Herschlag, 1997 and references therein).

The absence of maximization of P9.0 binding may reflect a biological need for a transient interaction in self splicing. In contrast to the 5′ portion of the P1 duplex, which is maintained through both steps of the self-splicing reaction and must be sufficiently strong to prevent release of the 5′ exon intermediate, the 3′ splice site must be out of the active site for the first chemical step, but in the active site for the second. If P9.0 were too strong, it could inhibit splicing by holding the 3′ splice site in the active site, blocking access of both exogenous G and the 5′ splice site. Even with the natural P9.0, hydrolysis is observed at the 3′ splice site in the absence of exogenous G (Inoue et al., 1986), demonstrating occupancy of the 3′ splice site prior to the first step. In addition to P9.0, P10 would be expected to favor docking of the 3′ splice site and could exacerbate this potential problem. An evolutionary balance may have been struck, with P9.0, P10, and the guanosine binding site strong enough to generate a discrete 3′ splice site, but sufficiently weak to allow the first splicing step to proceed efficiently.

MATERIALS AND METHODS

Materials

L–21 *Sca*I ribozyme was prepared by in vitro transcription and purified by 4% polyacrylamide/8 M urea gel electrophoresis as described (Zaug et al., 1988) or by using a Qiagen RNeasy column following the manufacturer's instructions. The two purification methods yielded material of identical activity as determined from $(k_{cat}/K_M)^S$ and $(k_{cat}/K_M)^G$ values (data not shown). Plasmids encoding P9.0 variant ribozymes were constructed by PCR site-directed mutagenesis of pT7L–21, and the desired changes, as well as the overall integrity of the ribozyme genes, were confirmed by sequencing. Variant ribozymes were prepared using Qiagen RNeasy columns. 5′ splice site mimics and 3′ splice site mimics were synthesized using standard solid-phase methods by the Protein and Nucleic Acid Facility at Stanford, and 3′ splice site mimics were HPLC-purified using a Dionex PA-100 anion exchange column. Oligonucleotides were eluted from the column with a 300-mL linear gradient of 0–2 M ammonium acetate, concentrated in a Speedvac, and applied to a SepPak (Millipore) column to remove residual salt. 5′ splice site mimics were 5′ end labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase and purified by nondenaturing polyacrylamide gel electrophoresis as described (Zaug et al., 1988; Herschlag et al., 1993a).

Kinetics

All reactions were single turnover, with ribozyme present in excess over ^{32}P-labeled 5′ splice site mimic (S) essentially as described previously (Herschlag et al., 1993a). Reaction conditions were 30 °C, 50 mM Na•MOPS, pH 6.8, 10 mM $MgCl_2$. The ribozyme (200 nM) was preincubated at 50 °C in buffer containing 10 mM Mg^{2+} for 30 min prior to initiation of reaction with S. S was −1d,rSA_5, as shown in Equation 1, unless otherwise indicated. The ^{32}P-labeled reaction product was separated from the labeled substrate using 20% polyacrylamide/8 M urea gel electrophoresis, and the fraction of label present as product was determined using a Molecular Dynamics Phosphorimager.

First-order rate equations were fit to the data using Kaleidagraph (Synergy Software). Progress curves followed good

single-exponential kinetics with endpoints of >97% product. Observed first-order rate constants from reactions performed at three or four subsaturating concentrations of 3′ splice site mimic were plotted as a function of concentration, and $(k_{cat}/K_M)^{Nuc}$ values were obtained from linear fits to the data ($R^2 \geq 0.92$). Values of $(k_{cat}/K_M)^{Nuc}$ from two to four separate experiments were averaged. Variation between experiments was typically 20–30%, and in no cases exceeded twofold. Results are reported as $(k_{cat}/K_M)^{rel}$, which represents $(k_{cat}/K_M)^{Nuc}$ relative to analogous reactions with guanosine [$(k_{cat}/K_M)^{rel} = (k_{cat}/K_M)^{Nuc}/(k_{cat}/K_M)^{G}$]. Values of $(k_{cat}/K_M)^{G}$ for all 5′ splice site mimics that contained a deoxyribose residue at position (−1) and 3′ adenosine tails were the same within error, except for −1d,rSA$_2$, which was ~4-fold lower than the others. 5′ splice site mimics with 3′ uridine tails (−1d,rSU, −1d,rSU$_2$, and −1d,rSU$_5$) yielded $(k_{cat}/K_M)^{G}$ values that were ~3-fold lower than the analog containing an A-tail, except −1d,rSU$_2$, which yielded a similar value to −1d,rSA$_2$. Values of $\Delta\Delta G^{\ddagger}$ were calculated using the relationship $\Delta\Delta G^{\ddagger} = -RT\ln[(k_{cat}/K_M)^{rel}]$, in which $R = 0.00198$ kcal mol^{-1} K^{-1}, and $T = 303$ K (30 °C).

ACKNOWLEDGMENTS

We thank Geeta Narlikar, Rachel Green, and Harry Noller for synthesis of the −1d,rSU and −1d,rSU$_5$ 5′ splice site mimics, and Pehr Harbury, members of the Herschlag lab, and reviewers for helpful comments on the manuscript. This work was supported by NIH Grant GM49243 to D.H. R.R. was supported in part by the Janet M. Shamberger Fellowship Fund.

Received October 1, 1998; returned for revision October 23, 1998; revised manuscript received November 11, 1998

REFERENCES

Been MD, Cech TR. 1985. Sites of circularization of the *Tetrahymena* rRNA IVS are determined by sequence and influenced by position and secondary structure. *Nucleic Acids Res 13*:8389–8408.

Bevilacqua PC, Sugimoto N, Turner DH. 1996. A mechanistic framework for the second step of splicing catalyzed by the *Tetrahymena* ribozyme. *Biochemistry 35*:648–658.

Burke JM, Esherick JS, Burfeind WR, King JL. 1990. A 3′ splice site-binding sequence in the catalytic core of a group I intron. *Nature 344*:80–82.

Cech TR, Damberger SH, Gutell RR. 1994. Representation of the secondary and tertiary structure of group I introns. *Nat Struct Biol 1*:273–280.

Cech TR, Herschlag D. 1997. Group I ribozymes: Substrate recognition, catalytic strategies, and comparative mechanistic analysis. In: Eckstein F, Lilley DMJ, eds. *Catalytic RNA*. Berlin: Springer-Verlag. pp 1–17.

Cech TR, Herschlag D, Piccirilli JA, Pyle AM. 1992. RNA catalysis by a group I ribozyme. Developing a model for transition state stabilization. *J Biol Chem 267*:17479–17482.

Fersht A. 1985. *Enzyme structure and mechanism*. New York: W.H. Freeman and Company.

Freier SM, Kierzek R, Jaeger JA, Sugimoto N, Caruthers MH, Neilson T, Turner DH. 1986. Improved free-energy parameters for predictions of RNA duplex stability. *Proc Natl Acad Sci USA 83*:9373–9377.

Herschlag D. 1991. Implications of ribozyme kinetics for targeting the cleavage of specific RNA molecules in vivo: More isn't always better. *Proc Natl Acad Sci USA 88*:6921–6925.

Herschlag D. 1995. RNA chaperones and the RNA folding problem. *J Biol Chem 270*:20871–20874.

Herschlag D, Cech TR. 1990. Catalysis of RNA cleavage by the *Tetrahymena thermophila* ribozyme. 2. Kinetic description of the reaction of an RNA substrate that forms a mismatch at the active site. *Biochemistry 29*:10172–10180.

Herschlag D, Eckstein F, Cech TR. 1993a. 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.

Herschlag D, Eckstein F, Cech TR. 1993b. The importance of being ribose at the cleavage site in the *Tetrahymena* ribozyme reaction. *Biochemistry 32*:8312–8321.

Herschlag D, Khosla M. 1994. Comparison of pH dependencies of the *Tetrahymena* ribozyme reactions with RNA 2′-substituted and phosphorothioate substrates reveals a rate-limiting conformational step. *Biochemistry 33*:5291–5297.

Inoue T, Sullivan FX, Cech TR. 1986. New reactions of the ribosomal RNA precursor of *Tetrahymena* and the mechanism of self-splicing. *J Mol Biol 189*:143–165.

Lehnert V, Jaeger L, Michel F, Westhof E. 1996. New loop–loop tertiary interactions in self-splicing introns of subgroup IC and ID: A complete 3D model of the *Tetrahymena thermophila* ribozyme. *Chem Biol 3*:993–1009.

McConnell TS, Cech TR, Herschlag D. 1993. Guanosine binding to the *Tetrahymena* ribozyme: Thermodynamic coupling with oligonucleotide binding. *Proc Natl Acad Sci USA 90*:8362–8366.

Michel F, Hanna M, Green R, Bartel DP, Szostak JW. 1989. The guanosine binding site of the *Tetrahymena* ribozyme. *Nature 342*:391–395.

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

Moran S, Kierzek R, Turner DH. 1993. Binding of guanosine and 3′ splice site analogues to a group I ribozyme: Interactions with functional groups of guanosine and with additional nucleotides. *Biochemistry 32*:5247–5256.

Narlikar GJ, Herschlag D. 1997. Mechanistic aspects of enzymatic catalysis: Lessons from comparison of RNA and protein enzymes. *Ann Rev Biochem 66*:19–59.

Narlikar GJ, Khosla M, Usman N, Herschlag D. 1997. Quantitating tertiary binding energies of 2′ OH groups on the P1 duplex of the *Tetrahymena* ribozyme: Intrinsic binding energy in an RNA enzyme. *Biochemistry 36*:2465–2477.

Profenno LA, Kierzek R, Testa SM, Turner DH. 1997. Guanosine binds to the *Tetrahymena* ribozyme in more than one step, and its 2′-OH and the nonbridging *pro*-Sp phosphoryl oxygen at the cleavage site are required for productive docking. *Biochemistry 36*:12477–12485.

Serra MJ, Turner DH. 1995. Predicting thermodynamic properties of RNA. *Methods Enzymol 259*:242–261.

Tanner NK, Cech TR. 1987. Guanosine binding required for cyclization of the self-splicing intervening sequence ribonucleic acid from *Tetrahymena thermophila*. *Biochemistry 26*:3330–3340.

Thirumalai D, Woodson SA. 1996. Kinetics of folding of proteins and RNA. *Acc Chem Res 29*:433–439.

Treiber DK, Rook MS, Zarrinkar PP, Williamson JR. 1998. Kinetic intermediates trapped by native interactions in RNA folding. *Science 279*:1943–1946.

Uhlenbeck OC. 1995. Keeping RNA happy. *RNA 1*:4–6.

Young B, Herschlag D, Cech TR. 1991. Mutations in a nonconserved sequence of the *Tetrahymena* ribozyme increase activity and specificity. *Cell 67*:1007–1019.

Zaug AJ, Grosshans CA, Cech TR. 1988. Sequence-specific endoribonuclease activity of the *Tetrahymena* ribozyme: Enhanced cleavage of certain oligonucleotide substrates that form mismatched ribozyme-substrate complexes. *Biochemistry 27*:8924–8931.