Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 6921–6925, August 1991
Biochemistry

Implications of ribozyme kinetics for targeting the cleavage of specific RNA molecules *in vivo*: More isn't always better

(substrate discrimination/kinetics/antisense/catalytic RNA/base pairing)

DANIEL HERSCHLAG

Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309-0215

Communicated by Thomas R. Cech, May 3, 1991

ABSTRACT **Kinetic and thermodynamic factors that determine specificity of RNA cleavage by ribozymes are illustrated with examples from recent work with a ribozyme derived from the group I intron of *Tetrahymena thermophila* pre-rRNA. The conclusions also apply to other ribozymes, to antisense oligonucleotide experiments, and to RNA and DNA cleavage agents that can recognize a single-stranded or double-stranded region of variable length. At first, adding bases to a ribozyme's recognition sequence is expected to increase cleavage of the target RNA relative to cleavage of other RNAs. However, adding more bases ultimately reduces this discrimination, as cleavage occurs essentially every time the target RNA *or* a mismatched RNA binds the ribozyme. This occurs despite the weaker binding of the mismatched RNA because dissociation becomes too slow (binding is too strong) to allow the ribozyme to "choose" between cleavage of the target RNA and a mismatched RNA. In summary, more (base pairing) isn't always better, because maximal discrimination requires equilibrium binding prior to cleavage. The maximum discrimination that can be obtained is expected to be greater with an A+U-rich recognition sequence than with a G+C-rich recognition sequence. This is because the weaker A·U base pairs (relative to G·C base pairs) allow recognition to be spread over a larger number of bases while preventing binding that is too strong. Finally, creating an A-rich ribozyme rather than a U-rich ribozyme avoids the loss in discrimination expected with U-rich ribozymes from the formation of U·G wobble pairs in addition to the "targeted" Watson–Crick U·A pair.**

The discovery of catalytic RNA has created interest in using RNA enzymes, or "ribozymes," to target the degradation of specific RNA molecules *in vivo* (e.g., refs. 1–3). Targeted destruction of viral or cellular mRNA, to eliminate the formation of a protein that is deleterious in a disease state, has potential therapeutic utility. Targeting also has potential utility for identification of gene function, analogous to experiments using antisense oligonucleotides (e.g., refs. 4–6). Since ribozymes recognize specific sequences in RNA by base pairing (7–10), the design of a therapeutic ribozyme could ultimately be much simpler than the development of new inhibitors directed at protein active sites. Furthermore, the rules for developing a ribozyme for one target can be used to develop a ribozyme for any other target. In contrast, protein active sites are idiosyncratic.

Although a tremendous future obstacle in the development of ribozymes as drugs is a stratagem for their delivery, the initial challenge is to determine whether ribozymes can be made to operate specifically and efficiently once provided *in vivo*. It is estimated that stretches of 11–15 nucleotides define unique sequences for cellular RNA. Thus, formation of a "matched" duplex with a single cellular RNA requires a recognition sequence of an antisense oligonucleotide (or ribozyme) of ≈11–15 nucleotides (6, 11). However, an antisense oligonucleotide that forms a perfect duplex with a single RNA *in vivo* still has the potential to hybridize with other RNAs, forming duplexes with one or more mismatches. Several experiments with antisense oligonucleotides have investigated the affect of mismatches within the recognition sequence, as it is important that nontarget RNAs in the cell not be inactivated (reviewed in ref. 4). In another context, the interplay of kinetic and thermodynamic factors for sensitivity and discrimination in the identification of unique sequences by hybridization and crosslinking has been analyzed in detail (12).

Rational design of ribozymes for *in vivo* targeting will rely upon our understanding of the efficiency and specificity of RNA-catalyzed cleavage. A recent kinetic analysis of a ribozyme derived from *Tetrahymena* pre-rRNA has provided some understanding of factors that determine efficiency and specificity (13). Here these results are applied to discrimination between correct and incorrect sequences in the targeted cleavage of specific RNA molecules. This analysis also holds for other targeting strategies: e.g., antisense binding to a target RNA by an oligodeoxynucleotide followed by cleavage with RNase H or with a chemical reagent or enzyme covalently attached to the oligodeoxynucleotide (10, 14) and triple-strand formation between an oligodeoxynucleotide and a DNA duplex followed by cleavage with a reagent covalently attached to the oligodeoxynucleotide (15). In each case recognition occurs residue by residue along the target and is followed by chemical cleavage. The counterintuitive nature of some of the conclusions presented herein is emphasized by introducing each section with a question.

ANALYSIS OF DISCRIMINATION

Section I. *True or false? The more bases in the recognition sequence of a ribozyme, the better its specificity.* Increasing the length of a recognition sequence (i.e., the number of bases that pair to the target RNA) would appear to increase specificity, because longer recognition sequences lower the probability of finding an identical sequence or even a sequence with only one or two differences. However, beyond a certain number of base pairs, more base pairs actually decrease specificity rather than increase it.

Consider the reaction of two RNA substrates, S_{right} and S_{wrong}, differing in nucleotide sequence (Scheme I). Discrimination between cleavage of these substrates (i.e., specificity) is determined by the ratio of k_{cat}/K_M values, multiplied by the ratio of substrate concentrations (16). k_{cat}/K_M is the second-order rate constant for reaction of free ribozyme (E) with free substrate; this kinetic parameter determines specificity regardless of whether the substrate concentrations are saturating or subsaturating.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

$$\begin{array}{l} S_{right} \xrightarrow{(k_{cat}/K_M)_{right}} P_{right} \\ + \\ E \\ + \\ S_{wrong} \xrightarrow[(k_{cat}/K_M)_{wrong}]{} P_{wrong} \end{array}$$

Scheme I

The solid line of Fig. 1*A* represents the free energy–reaction profile of a hypothetical ribozyme E_1 with S_{right}, which forms a perfectly matched helix; the dotted line represents the free energy–reaction profile for one of many "wrong" substrates that have a single nucleotide difference from the "right" substrate and therefore form less stable (higher energy) $E_1 \cdot S$ complexes. The chemical step is shown as rate-limiting in Fig. 1*A*, because the short recognition sequence of ribozyme E_1 is expected to result in weak binding and fast dissociation. (Weak binding causes the free energy barrier for dissociation of S from E·S to be lower than the barrier for the chemical step, so that the highest barrier to reaction is that for the chemical step.) The higher energy barrier for reaction of S_{wrong} in the transition state (‡) shows that it is discriminated against.

Fig. 1*B* shows the consequences of lengthening the ribozyme's recognition sequence. Although binding of both S_{right} and S_{wrong} are stronger with ribozyme E_2, as it has a longer recognition sequence than E_1, the energy difference between complexes $E \cdot S_{right}$ and $E \cdot S_{wrong}$ is the same with E_2 as with E_1. The difference in energy between S_{right} and S_{wrong} in the transition state of the chemical step is also the same with E_2 and E_1. However, this difference no longer provides discrimination, because the highest barrier to reaction is that for binding of the substrates by E_2 (Fig. 1*B*, ‡). That is, binding rather than chemistry is rate-limiting so that the potential for discrimination from the weaker equilibrium binding of S_{wrong} is no longer realized (17). The values of k_{cat}/K_M for E_2 with S_{right} and S_{wrong} are similar. (A small difference is described in the next paragraph.)

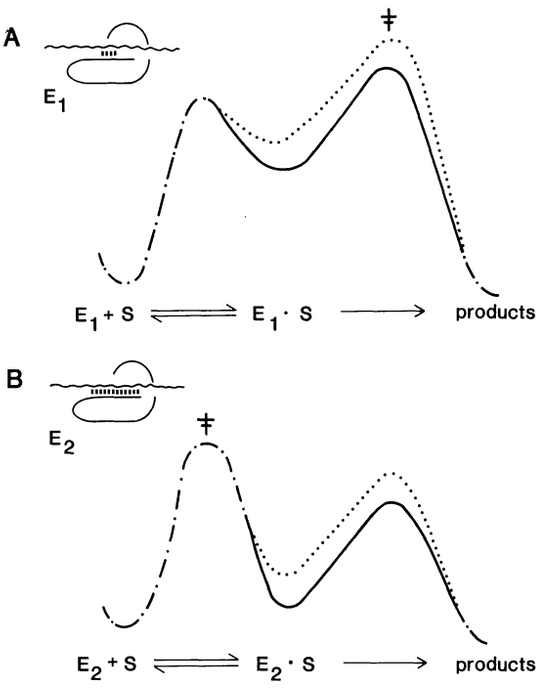

FIG. 1. Hypothetical free energy profiles for reaction of S_{right} (solid line), a substrate that forms a perfect matched duplex with the recognition sequence of the ribozyme, and for reaction of S_{wrong} (dotted line), which forms a duplex with the ribozyme that contains a single mismatch. Reaction profiles are shown for ribozyme E_1, which has a short recognition sequence (*A*), and for ribozyme E_2, which has a longer recognition sequence (*B*). It is assumed that lengthening the ribozyme's recognition sequence strengthens binding, while not changing the rate of the chemical step subsequent to binding or the rate constant for duplex formation, and that mismatches affect only the binding affinity, not the rate of the cleavage step (i.e., once the substrate is bound it reacts at the same rate whether or not there is a mismatch; see ref. 13). In addition, it is assumed for simplicity that the rate constant for duplex formation is the same for S_{right} and S_{wrong}. (A small observed difference is described in the text.) The reaction profiles are drawn for subsaturating or k_{cat}/K_M conditions, such that $[S] << K_M$.

An absence of discrimination caused by rate-limiting substrate binding has been observed for a ribozyme derived from the *Tetrahymena thermophila* pre-rRNA intron (saturating guanosine cofactor, pH 7, 10 mM $MgCl_2$, 50°C; refs. 13 and 18). An RNA substrate that forms a perfect duplex with this ribozyme's recognition sequence (i.e., S_{right}) reacts in the chemical step with a rate constant of $\approx 350\ min^{-1}$, whereas it dissociates with a rate constant of only $0.2\ min^{-1}$. Therefore, every time this substrate binds, it reacts. However, a substrate that forms a mismatched duplex with the recognition sequence (i.e., S_{wrong}) also reacts essentially every time it binds, since the chemical step occurs at $\approx 200\ min^{-1}$ and dissociation at $\approx 50\ min^{-1}$. Thus, despite the 10^3-fold weaker binding of the mismatched RNA the discrimination is only ≈5-fold, and this arises predominantly from the 4-fold slower formation of the mismatched helix with S_{wrong} than of the matched helix with S_{right}.

Discrimination with a pool of potential RNA substrates. A discrimination index representing the rate of cleavage of the target RNA relative to the rate of cleavage of all other RNAs is defined in Eq. **1**:

$$\text{discrimination index} = \frac{(k_{cat}/K_M)_{right}[S_{right}]}{\sum_i (k_{cat}/K_M)_{i,wrong}[S_{i,wrong}]}. \quad [1]$$

Because most cellular RNAs are long molecules, the wrong substrates include those with mismatches within the recognition region of the ribozyme as well as molecules that match the recognition sequence but differ outside this region. Adding bases to a short recognition sequence at first increases the discrimination index. This occurs simply because fewer RNAs in the pool of long RNA molecules have an identical sequence over five residues than over three residues, for example. However, extending the recognition sequence beyond a certain length *decreases* the discrimination index by allowing more RNAs with single and multiple mismatches to react at the same k_{cat}/K_M as the "right" (matched) RNA (Fig. 1*B*). Thus, as the length of the recognition sequence is increased the discrimination index first increases and then decreases.

This maximum in the discrimination index as a function of recognition sequence length is demonstrated with the following numerical example. A pool of all possible RNAs of length $m = 15$ are considered, at equal concentrations. A simple model for binding is used in which each base pair provides a stabilization energy $E = 2$ kcal/mol and there is an energetic cost of 3.5 kcal/mol for helix initiation (Eq. **2**, where n is the length of the recognition sequence and x is the number of mismatched bases).* $(k_{cat}/K_M)_{n,x}$ for each ribozyme of length

*In this model the free energy for a duplex is given solely by the number of base pairs, without an energetic penalty for mismatches. Mismatched bulges within a duplex have been shown to destabilize a helix beyond the the loss in energy from the absence of a single base pair, and "nearest neighbors" also affect duplex stability (19). In addition, cleavage at only a single position along each RNA molecule is considered in this calculation, and the possibility of reaction via duplexes with residues of the substrate or recognition sequence "looped out" is ignored. More complete treatment is expected to give quantitative, not qualitative, differences.

n and each incorrect substrate with x mismatches to the recognition sequence was calculated from Eq. **3** using $k_{on} = 10^8$ $M^{-1}\cdot min^{-1}$ [the rate of duplex formation for substrates with the *Tetrahymena* ribozyme and for simple complementary oligonucleotides (ref. 18 and references therein)], $k_c = 1$ min^{-1} (chosen arbitrarily), and $(k_{off})_{n,x} = (K_d)_{n,x}(k_{on})$. The number of different (incorrect) RNA sequences with x mismatches (N) was calculated from Eq. **4**. The term $n!/x!(n-x)!$ is a statistical factor representing the number of ways to distribute x mismatches over n bases; 3^x arises from three possible mismatches in each of the x mismatched positions; and 4^{m-n} corresponds to the number of sequence possibilities outside of the recognition sequence, where m is the length of the RNA molecules; for $x = 0$, when there are no mismatched bases, this value is reduced by 1 (the correct target sequence).

$$\Delta G^\circ_{n,x} = -(n-x)E + 3.5 \text{ kcal/mol};$$

$$(K_d)_{n,x} = e^{-\Delta G^\circ_{n,x}/RT} \quad [2]$$

$$(k_{cat}/K_M)_{n,x} = k_{on}k_c/[k_c + (k_{off})_{n,x}];$$

$$(k_{off})_{n,x} = (K_d)_{n,x}(k_{on}) \quad [3]$$

$$N = [n!/x!(n-x)!](3^x)(4^{m-n}) \quad \text{for } x = 1 \text{ to } n$$

$$= \{4^{m-n}\} - 1 \quad \text{for } x = 0 \quad [4]$$

$$\text{discrimination index} = \frac{(k_{cat}/K_M)_{n,0}}{\sum_{x=0}^{n} N(k_{cat}/K_M)_{n,x}} \quad [5]$$

The discrimination index (Eq. **5**) for reaction of ribozymes with recognition sequences of length $n = 1$ to 15 with a pool of all RNAs of length $m = 15$ is plotted in the solid line of Fig. 2*A*. There is a maximum in the discrimination index for a recognition sequence of length $n = 7$ in this example, despite the fact that a recognition sequence of length 15 is needed to form a perfect duplex with only one RNA in this pool. There is clearly more to specificity than simply forming a perfect duplex with a unique RNA. The maximum in Fig. 2*A* occurs as the value of k_{cat}/K_M for the target RNA approaches its upper limit of 10^8 $M^{-1}\cdot min^{-1}$, which is set by the rate of binding (Fig. 2*B*, ●). Once the maximum value of k_{cat}/K_M is achieved for the "correct" RNA, further increases in the length of the recognition sequence only enhance k_{cat}/K_M for the incorrect RNAs, thereby lowering discrimination.

The ribozyme with a recognition sequence of length $n = 7$ gives the maximum discrimination index in Fig. 2*A* but still cleaves incorrect RNAs $\approx 10^5$-fold more often than it cleaves the target RNA. These incorrect RNAs fall into two classes: (*i*) about half of the incorrect cleavage arises from RNAs that form matched duplexes with the recognition sequence but contain differences outside of this region and therefore are cleaved at the same rate as the target RNA (Eqs. **2–4**, $x = 0$); (*ii*) the other half of the incorrect cleavage arises from the slower cleavage of RNAs that form duplexes with the recognition sequences containing one or more mismatches (Eqs. **2–4**, $x = 1$ to n).

(*i*) It might be thought at first that class *i* incorrect cleavages could be avoided simply by lengthening the recognition sequence to eliminate these RNAs. However, this lowers the discrimination index, as shown in Fig. 2*A*, because RNAs with mismatches are then cleaved as fast as the target RNA. Thus, to solve this problem, conditions must be found that allow discrimination against mismatched RNAs when the recognition sequence is longer. Such conditions are described below and in *Section II*.

(*ii*) In contrast, class *ii* miscleavage appears to be unavoidable (however, see *Section III.2*). This miscleavage may not be problematic because, despite the large total amount of class *ii* miscleaved RNA, the target RNA is still cleaved faster than any individual incorrect RNA (i.e., k_{cat}/K_M is larger for the target RNA). However, as the reaction proceeds, the preferential depletion of the target RNA results in even more cleavage of the incorrect RNAs relative to the target RNA

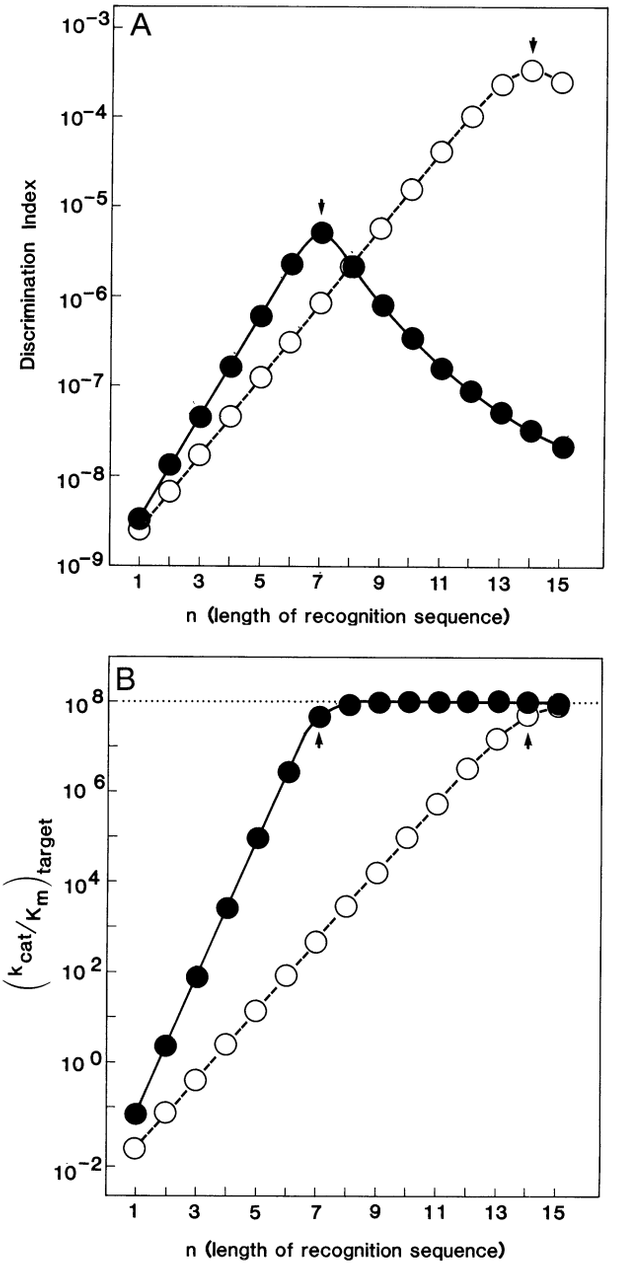

FIG. 2. Discrimination index (*A*) and k_{cat}/K_M (*B*) for cleavage of the target RNA by hypothetical ribozymes with varying length of the recognition sequence (n). The simple model described in the text was used to calculate the discrimination index (Eq. **5**; i.e., the ratio of cleavage of the target RNA relative to cleavage of all other RNAs in a pool of all possible RNAs of length $m = 15$ present at equal concentration) and the value of $(k_{cat}/K_M)_{target}$ for cleavage of the target RNA (i.e., $(k_{cat}/K_M)_{n,0}$ in Eq. **3** for cleavage of an RNA that forms a perfect duplex with the recognition sequence). The discrimination index and k_{cat}/K_M values were calculated with each base pair contributing $E = 2$ kcal/mol (Eq. **2**) (●) or $E = 1$ kcal/mol (○). The dotted line in *B* represents the upper limit for k_{cat}/K_M that is set by the rate of binding ($k_{on} = 10^8$ $M^{-1}\cdot min^{-1}$); the arrows correspond to the maxima in the discrimination index from *A*. A temperature of 37°C was used in these calculations. Note the logarithmic scale for the *y* axis in both panels.

because the relative rates of cleavage are dependent on the relative concentrations of the target and incorrect RNAs (Eq. **1** and Scheme **I**; note that the concentration terms of Eq. **1** are omitted in Eq. **5**, as an equal concentration of all RNAs is assumed for Eq. **5**). This relationship between the extent of cleavage of the target RNA and the amount of miscleavage causes the intracellular concentration of ribozyme and its lifetime to affect the relative amount of target and nontarget RNA that is cleaved over a given time interval. For example, increasing the ribozyme concentration increases cleavage of the target RNA but increases miscleavage by a greater amount. A corollary is that the *observed* specificity of cleavage will be greater for an abundant RNA as compared to a rare RNA, all other factors being equivalent (Eq. **1**). (See ref. 12 for a related discussion of the quantitative effects of probe concentration on the specificity of hybridization. The situations are analogous, as specificity is lost in clearing the target RNA or hybridizing the target DNA to near completion.) A significant loss in the amount of individual RNAs from miscleavage of class *ii* RNAs is expected only when the target RNA is nearly completely removed or when a particular class *ii* RNA is much more abundant than the target RNA. In addition, although decreased levels of some RNAs may be detrimental, the decreased levels of others may be readily tolerated, especially if their intracellular concentrations are tightly regulated.

Is there anything that can be done to allow greater discrimination? As stated above, the specificity problem for ribozymes with long recognition sequences arises because mismatched RNA does not have time to dissociate before cleavage. Thus, changes that allow this dissociation so that substrates can equilibrate between free and ribozyme-bound prior to cleavage will increase specificity. This can be done in two ways: (*a*) by slowing the rate of chemistry or (*b*) by increasing the rate of dissociation (see also *Sections II* and *III*). With respect to Fig. 2*A*, these changes allow discrimination to reach a maximum with longer recognition sequences (n), resulting in a larger value of the discrimination index. For the *Tetrahymena* ribozyme, there are examples of each that give the expected increase in specificity. (*a*) Chemistry has been slowed by lowering the concentration of guanosine, a cofactor in the site-specific endonuclease reaction, resulting in an ≈50-fold increase in discrimination between a matched and a mismatched substrate (13). For the model calculations of Fig. 2*A*, decreasing the rate constant for the chemical step 100-fold to $k_c = 0.01\ min^{-1}$ gives an increase in the maximum discrimination of ≈5-fold, with the maximum occurring for a recognition sequence of length $n = 8$ rather than $n = 7$ as in Fig. 2*A*. (*b*) The rate of dissociation has been increased (and binding weakened) by the introduction of mutations outside of the recognition sequence that interfere with tertiary binding interactions between the ribozyme and substrates. This results in *increases* of up to 70-fold in discrimination between matched and mismatched substrates (B. Young, D.H., and T. R. Cech, unpublished data).

For actual ribozymes, how long is too long? That is, what length recognition sequence gives the maximum discrimination in Fig. 2*A*? As stated above, the wild-type ribozyme derived from the group I *Tetrahymena* pre-rRNA intron does not discriminate between matched and mismatched RNA substrates, even though it recognizes only 6 bases (13, 18). The lack of discrimination for a recognition sequence of only 6 base pairs occurs because tertiary interactions with the *Tetrahymena* ribozyme, which slow dissociation by a factor of 10^4 for both the matched and mismatched substrates, and a fast chemical step (≈200–350 min^{-1}) prevent equilibrium binding prior to cleavage (13, 18). The specificity of other ribozymes has not been reported. It is possible that the hammerhead and hairpin ribozymes (7, 8, 20, 21) lack stabilization from tertiary binding interactions and therefore will be able to provide greater discrimination using longer recognition sequences than the *Tetrahymena* ribozyme. In addition, the formation of two helices between these ribozymes and their substrates, 5′ and 3′ of the cleavage site, may result in less total binding energy than would occur with a continuous helix. This would allow longer recognition sequences before binding became too strong and reduced discrimination.

Section II. *True or false? G·C base pairs, with more energy realized per base pair than A·U pairs, give a ribozyme greater ability to discriminate against RNA molecules of the wrong sequence.* Although more discrimination from more energy sounds reasonable at first, Figs. 1 and 2 showed that beyond a certain amount of recognition energy, binding becomes rate-limiting, specificity is sacrificed, and the discrimination index begins to decrease rather than continuing to increase. Thus, it is not immediately clear whether recognition sequences rich in guanine and cytosine residues or those rich in adenine and uracil residues can give more discrimination.

The following numerical example shows that recognition sequences that are rich in A·U pairs can, in principle, provide more discrimination than those rich in G·C pairs. Consider two series of ribozyme-target RNA complexes, one with all G·C base pairs and the other with all A·U base pairs. Each G·C base pair provides more binding energy than each A·U pair, and for purposes of illustration it is assumed that a G·C pair is worth 2 kcal/mol and an A·U pair is worth 1 kcal/mol. (The actual value depends on the nucleotide sequence; ref. 19.) The discrimination for G+C-rich ribozymes with varying recognition sequence lengths is then estimated by the solid line in Fig. 2*A*, the calculation for which was described above. Performing the analogous calculation for A+U-rich ribozymes with an energy per base pair of $E = 1$ kcal/mol (Eq. **2**) gives the dashed line in Fig. 2*A*. The maximum in the discrimination index is larger for the A+U-rich ribozyme than for the G+C-rich ribozyme (Fig. 2*A*, arrows). In addition, this maximum occurs with a recognition sequence of length $n = 14$, which is long enough to form a matched duplex with a unique RNA sequence *in vivo* (10, 11).

How can the greater discrimination with an A+U-rich ribozyme than a G+C-rich ribozyme be understood? With both the A+U- and G+C-rich series of ribozymes, the discrimination index begins to decrease as the value of k_{cat}/K_M for the target RNA approaches the limiting value set by the rate of binding (Fig. 2*B*; arrows correspond to the recognition sequence length that gives the maximum in the discrimination index). But because there is less binding energy for each base pair with the A+U-rich ribozymes, this limit is not reached until the recognition sequence is longer (Fig. 2*B*, dashed line). The ability to discriminate against more sequences outweighs the disadvantage from less discrimination against each incorrect sequence, giving greater maximal discrimination with the longer A+U-rich ribozymes. Thus, spreading out the recognition energy over more base pairs allows greater discrimination.

Do factors in addition to the A+U vs. G+C content of the recognition sequence affect specificity? Each U residue in a recognition sequence can form the correct U·A base pair with the target RNA or a wobble U·G pair with a different RNA. Compilations of free energies of duplex formation suggest that correct U·A pairs are favored over wobble U·G pairs by only ≈0.5–1.4 kcal/mol (37°C; the value is sequence-dependent; ref. 19). This free energy difference is much less than that of 1.8–4.4 kcal/mol estimated for the change from an A·U base pair to a mismatch (19), and corresponds to only 2.4- to 12-fold discrimination against the wobble RNA. Thus, placing A residues rather than U residues in the ribozyme's recognition sequence is expected to enhance discrimination.

A+U-rich recognition sequences increase discrimination by lowering the amount of energy per base pair (Fig. 2*A*). Discrimination could also be enhanced by lowering the

amount of energy per base by other means. For example, holding the bases of a recognition sequence nonoptimally for helix formation through interactions with other parts of the ribozyme could decrease the binding energy realized upon formation of a base pair. In practice, such subtle changes in base positioning may be hard or impossible to achieve. Creating a ribozyme with alternative nucleotides in the recognition arms could lower the energy per base pair by substitution of sugars such as deoxyribose or glycerol that give less stable base pairs than ribose, or by adding abasic sites or other linkers that create helix defects (22–24).

The experimenter, of course, cannot always find U-rich sequences in the target RNA. To further complicate matters, bound proteins and the structure of target RNAs can interfere with ribozyme cleavage (9, 25, 26).

Section III. *Is there anything else that can be done to increase specificity?* The answer is yes, at least in principle. Two types of improvement are considered.

1. The rate of binding of an RNA substrate to a ribozyme provides the upper limit to k_{cat}/K_M (Fig. 1*B*). If both binding and dissociation were faster, the maximum k_{cat}/K_M would be higher (Fig. 2*B*, dotted line). There could then be a continued increase in specificity for ribozymes with longer recognition sequences, until a new higher specificity limit was reached. If *in vivo* RNA concentrations are subsaturating, there would also be an advantage from faster cleavage of the target RNA.

How could the rate of binding be increased? The *Tetrahymena* ribozyme binds an RNA substrate complementary to its recognition sequence at 10^8 $M^{-1}\cdot min^{-1}$, similar to the rate constant for duplex formation between simple oligonucleotides but well below diffusional encounter (ref. 18 and references therein). The rate of binding to the *Tetrahymena* ribozyme increases with increasing concentrations of Mg^{2+} or Ca^{2+}, because of reduced electrostatic repulsion or because Mg^{2+} and Ca^{2+} hold the ribozyme's recognition sequence in the proper geometry to increase the probability of helix nucleation (T. McConnell, D.H., and T. R. Cech, unpublished results). In addition, there is evidence that proteins that bind both single- and double-stranded nucleic acids can increase the rate of equilibration between single-stranded and duplex nucleic acid (27).

2. A second way to increase specificity would be to design a ribozyme in which discrimination against mismatched RNAs is heightened due to unfavorable interactions of the mismatched RNA with the ribozyme. This may occur with RNA molecules that differ from the target RNA at conserved positions (e.g., ref. 28).

Although it is not clear how to design a ribozyme that can bind RNA faster or can heighten discrimination by unfavorable interactions with "mismatched" RNA substrates, it might be possible to obtain such a molecule through successive rounds of mutation and selection (29–31).

CONCLUSIONS

Adding more bases to a ribozyme's recognition sequence at first increases discrimination but then decreases discrimination. The underlying principle is that discrimination from differential binding affinities requires equilibrium binding prior to the cleavage step. When dissociation of the RNA substrate is too slow or cleavage is too fast to allow equilibrium binding, specificity is sacrificed. This principle for discrimination and the discussions herein also hold for other targeting strategies: in each case binding is followed by cleavage. They can also be applied to targeting that involves inhibition of RNA function by simple duplex formation without subsequent cleavage. In this case, if dissociation of duplexes that contain mismatches is slow on the time scale of the experiment (or the cell's life-span), then the antisense oligonucleotide will remain bound to whatever RNA molecules it finds first, and will never find the RNAs that it binds most stably. Analogous to the conclusions herein, allosteric interactions that *weaken* binding of both cognate and noncognate tRNAs to ribosomes have been suggested to enhance discrimination (32).

General scenarios have been presented in the hope of illustrating principles that can then be applied to specific ribozymes and specific targets. It is hoped that an awareness of basic factors that determine discrimination will help experimental design and analysis.

I thank Tom Cech, Jerry Joyce, Joe Piccirilli, Anna Marie Pyle, Gary Stormo, and Olke Uhlenbeck for helpful discussions and comments and Tim McConnell for writing computer programs and for help in analyzing specificity. The author is a Lucille P. Markey Scholar in Biomedical Science. This work was supported by a grant from the Lucille P. Markey Charitable Trust.

1. Cech, T. R. (1988) *J. Am. Med. Assoc.* **260,** 3030–3034.
2. Haseloff, J. & Gerlach, W. L. (1988) *Nature (London)* **334,** 585–591.
3. Rossi, J. J. & Sarvar, N. (1990) *Trends Biotechnol.* **8,** 179–183.
4. Colman, A. (1990) *J. Cell Sci.* **97,** 399–409.
5. Takayama, K. M. & Inouye, M. (1990) *Crit. Rev. Biochem. Mol. Biol.* **25,** 155–184.
6. Helene, C. & Toulme, J. J. (1990) *Biochim. Biophys. Acta* **1049,** 99–125.
7. Hampel, A., Tritz, R., Hicks, M. & Cruz, P. (1990) *Nucleic Acids Res.* **18,** 299–304.
8. Foster, A. C. & Symons, R. H. (1987) *Cell* **49,** 211–220.
9. Zaug, A. J., Been, M. D. & Cech, T. R. (1986) *Nature (London)* **324,** 429–433.
10. Waring, R. B., Towner, P., Minter, S. J. & Davies, R. W. (1986) *Nature (London)* **321,** 133–139.
11. Miller, P. & Tso, P. (1987) *Anti-Cancer Drug Design* **2,** 117–128.
12. Hearst, J. E. (1988) *Annu. Rev. Phys. Chem.* **39,** 291–315.
13. Herschlag, D. & Cech, T. R. (1990) *Biochemistry* **29,** 10172–10180.
14. Zuckermann, R. N., Corey, D. R. & Schultz, P. G. (1988) *J. Am. Chem. Soc.* **110,** 1614–1615.
15. Griffin, L. C. & Dervan, P. B. (1989) *Science* **245,** 967–971.
16. Fersht, A. (1985) *Enzyme Structure and Mechanism,* (Freeman, New York), 2nd Ed., pp. 111–112.
17. Ninio, J. (1975) *Biochimie* **57,** 587–595.
18. Herschlag, D. & Cech, T. R. (1990) *Biochemistry* **29,** 10159–10171.
19. Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T. & Turner, D. H. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 9373–9377.
20. Uhlenbeck, O. C. (1987) *Nature (London)* **328,** 596–600.
21. Hampel, A. & Tritz, R. (1989) *Biochemistry* **28,** 4929–4933.
22. Walker, G. T. (1988) *Nucleic Acids Res.* **16,** 3091–3099.
23. Martin, F. H. & Tinoco, I., Jr. (1980) *Nucleic Acids Res.* **8,** 2295–2299.
24. Schneider, K. C. & Benner, S. A. (1990) *J. Am. Chem. Soc.* **112,** 453–455.
25. Latham, J. A., Zaug, A. J. & Cech, T. R. (1990) *Methods Enzymol.* **181,** 558–569.
26. Fedor, M. & Uhlenbeck, O. C. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 1668–1672.
27. Pontius, B. W. & Berg, P. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 8403–8407.
28. Ruffner, D. E., Stormo, G. D. & Uhlenbeck, O. C. (1990) *Biochemistry* **29,** 10695–10702.
29. Robertson, D. L. & Joyce, G. F. (1990) *Nature (London)* **344,** 467–468.
30. Tuerk, C. & Gold, L. (1990) *Science* **249,** 505–510.
31. Green, R., Ellington, A. D. & Szostak, J. W. (1990) *Nature (London)* **347,** 406–408.
32. Nierhaus, K. H. (1990) *Biochemistry* **29,** 4997–5008.