Sequence Specificity of a Group II Intron Ribozyme: Multiple Mechanisms for Promoting Unusually High Discrimination against Mismatched Targets†

Qing Xiang,‡ Peter Zhifeng Qin,§ William J. Michels,‡,|| Keri Freeland,⊥ and Anna Marie Pyle*,‡

Department of Biochemistry and Molecular Biophysics, Integrated Program in Molecular, Developmental and Cellular Biology, Columbia University College of Physicians and Surgeons, New York, New York 10032 and Department of Applied Physics, Columbia University, New York, New York 10032

Received October 27, 1997; Revised Manuscript Received January 7, 1998

ABSTRACT: Group II intron ai5γ was reconstructed into a multiple-turnover ribozyme that efficiently cleaves small oligonucleotide substrates *in-trans*. This construct makes it possible to investigate sequence specificity, since second-order rate constants (k_{cat}/K_m, or the specificity constant) can be obtained and compared with values for mutant substrates and with other ribozymes. The ribozyme used in this study consists of intron domains 1 and 3 connected *in-cis*, together with domain 5 as a separate catalytic cofactor. This ribozyme has mechanistic features similar to the first step of reverse-splicing, in which a lariat intron attacks exogenous RNA and DNA substrates, and it therefore serves as a model for the sequence specificity of group II intron mobility. To quantitatively evaluate the sequence specificity of this ribozyme, the WT k_{cat}/K_m value was compared to individual k_{cat}/K_m values for a series of mutant substrates and ribozymes containing single base changes, which were designed to create mismatches at varying positions along the two ribozyme−substrate recognition helices. These mismatches had remarkably large effects on the discrimination index (1/relative k_{cat}/K_m), resulting in values >10 000 in several cases. The $\Delta\Delta G^{\ddagger}$ for mismatches ranged from 2 to 6 kcal/mol depending on the mismatch and its position. The high specificity of the ribozyme is attributable to effects on duplex stabilization (1−3 kcal/mol) and unexpectedly large effects on the chemical step of reaction (0.5−2.5 kcal/mol). In addition, substrate association is accompanied by an energetic penalty that lowers the overall binding energy between ribozyme and substrate, thereby causing the off-rate to be faster than the rate of catalysis and resulting in high specificity for the cleavage of long target sequences (≥13 nucleotides).

Ribozymes are of particular interest as agents for genomic alteration because they combine the ability to recognize target sequences through base pairing with the catalytic capability to modify these sequences through cleavage and/or insertion of new genetic information at the target site (*1*−*3*). The ability to identify and modify individual genes selectively within an organism would result in numerous medical and technological applications (*4*−*7*). Although ribozymes appear to be promising agents for cleaving undesirable RNA molecules (*2*), and even for inserting new genes (or parts of genes) through splicing reactions (*8*−*10*), their efficacy is limited to the extent that they can be sequence specific. Particularly in applications where the genome itself would be modified during the course of reaction (*9*), it is essential that a ribozyme recognize and act only at the correct target sequence. In studies of functional genomics, specificity is important for preventing false-positive signals, or the inappropriate selection of genes that are not involved in a pathway of interest. It is possible that the abundant proteins and environmental complexity of in vivo systems may render ribozymes particularly specific in a cellular milieu. However, in vitro studies on two ribozymes that are currently under development for application in gene targeting have shown that these ribozymes have relatively low levels of sequence specificity (*11*−*13*).

This lack of ribozyme sequence specificity is due to a variety of factors: In the case of the group I intron, the recognition site size is inherently short (rarely greater than six bases), and its specificity for the proper six-nucleotide sequence is further diminished by an abundance of nonspecific tertiary interactions to the substrate backbone (*14*−*16*). Tight substrate binding, combined with a rapid rate for chemical cleavage, causes even mismatched substrates to be cleaved before they can fall off of the ribozyme (*11*). Like the group I intron, the hammerhead ribozyme has low sequence specificity in practice, despite the fact that hammerhead ribozymes can be designed to contain long recognition site sizes (*12, 13*). Again, the high toleration of base mismatches in the hammerhead ribozyme is due to very strong binding energy coupled with a fast chemical cleavage rate. Similar behavior is expected for the hairpin ribozyme (*17*), which has the added problem of intrinsic sequence requirements on the substrate (*18, 19*).

† Supported by grants from the National Institutes of Health (GMRO150313), the National Science Foundation (NYI award to A.M.P.), and the Beckman Foundation.

‡ Department of Biochemistry and Molecular Biophysics, Columbia University College of Physicians and Surgeons.

§ Department of Applied Physics, Columbia University.

|| Present address: Department of Molecular and Cell Biology, University of California at Berkeley.

⊥ Integrated Program in Molecular, Developmental and Cellular Biology, Columbia University College of Physicians and Surgeons.

S0006-2960(97)02661-5 CCC: $15.00
Published on Web 02/27/1998

While many ribozymes can be reconstructed to target and cleave specific RNA sequences, only one depends on sequence-specific genomic targeting for its natural function. Group II introns are mobile genetic elements that migrate to new genomes through a process called "homing", which is related to reverse-splicing (*20, 21*). During homing, free lariat intron attacks and incorporates itself into the sense strand of double-stranded genomic DNA (*9*). The biological function of homing remains unclear, although the introns appear to behave as a form of selfish RNA, or autonomous genetic element, that propagates by insertion into new genomes (*22*). Homing occurs with great site-selectivity during crosses between yeast strains, resulting in 80−100% conversion of intronless alleles (*23, 24*). High target specificity appears to be important for mobility and subsequent self-splicing function in group II introns, since a misincorporation at the wrong site could result in a loss of reading frame (important for translation of intron-encoded proteins), inappropriate codon usage (*20*), and in the loss of interactions that specify the exons for proper splicing (*25*).

In addition to the efficiency and specificity that have been observed for natural homing by group II introns (*20, 21*), there are several other reasons to suspect that group II intron ribozymes might show a high level of specificity for proper target cleavage. First of all, the recognition site size between intron domain 1 (D1) and the substrate is generally long (∼13 base pairs) (*26*), and of sufficient length to target a single mRNA sequence (and potentially a single DNA sequence) within the human genome. Although pairing must be maintained between substrate and EBS (exon binding site) regions of the ribozyme, absolute sequences are not conserved and can be mutationally varied (*26*), suggesting that group II intron ribozymes could be designed to target and cleave any desired sequence. Second, the two substrate recognition elements in D1 (EBS1 and EBS2) are widely separated within the secondary structure. Because these substrate binding sites are not contiguous with each other, it is possible that they form two independent duplexes upon complexation with the substrate, which might tune the overall binding energy and thereby contribute to specificity (*27*). Third, it has been shown that, unlike the group I intron, the group II intron core does not form tertiary interactions with 2′-hydroxyl groups on the substrate backbone (*28*). Because any nucleotide, regardless of sequence, has 2′-hydroxyl residues, interactions with these groups are low in specific information content. Given the studies conducted to date, it appears that the substrate−ribozyme binding energy of group II introns derives primarily from specific base-pairing interactions (*28*).

Before describing the experimental design that is required for monitoring ribozyme specificity, a precise definition of specificity is required: Specificity is defined as the relative preference for enzymatic modification of the correct substrate vs an incorrect one (*29*). In terms of ribozymes, this is the relative preference for cleavage (and/or modification) of a matched substrate vs one that results in mismatched pairings between ribozyme and substrate (*27*). The degree of specificity for a proper substrate is usually expressed in terms of the "discrimination index", or the relative second-order rate constant (k_{cat}/K_m) for cleavage of matched vs mismatched substrates (*27*). The kinetic parameter k_{cat}/K_m is historically known as the "specificity constant", and it is the yardstick

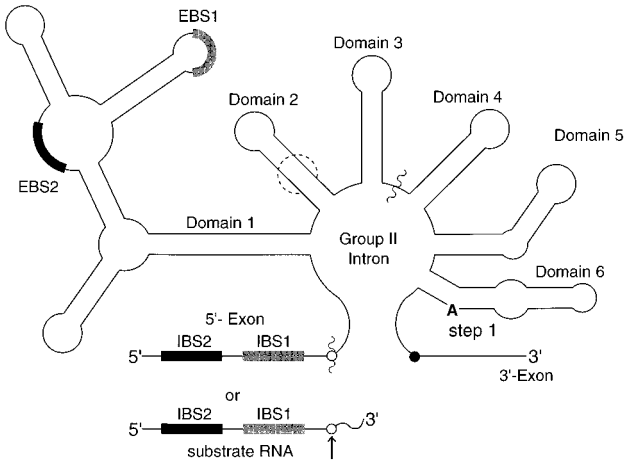

FIGURE 1: Schematic secondary structure of a group II intron ribozyme. The six domains of a group II intron are shown, and the EBS−IBS interactions on the intron and exon, respectively, are shown as bold shaded lines. These pair together in order to define the 5′-splice site for cleavage by the intron. In self-splicing by group II introns, the first step of splicing is initiated by nucleophilic attack of an adenosine 2′-hydroxyl group located in D6 or by a water molecule. During the second step of splicing, the free 5′-exon attacks the 3′-splice site, ligating the exons and liberating free lariat or linear intron. The ribozyme described in the present study consists of domains 1−3 of the intron, with D2 replaced as a short hairpin loop. D5 is added as a short hairpin RNA molecule in trans. Oligonucleotides corresponding to the first 17 nucleotides of the 5′-exon (containing both IBS1 and IBS2, shown below the 5′-exon) and 7 nucleotides of intron sequence are added to the complex of D13 and D5, and are cleaved by attack of a water molecule.

by which the degree of any enzymatic discrimination between substrates is measured (*29*). Therefore, to describe ribozyme specificity in a meaningful way and to reveal the energetic differences underlying apparent preferences for one substrate vs another, it is necessary to measure the k_{cat}/K_m for reaction of ribozyme with a variety of different substrate molecules. It is also essential to note that k_{cat}/K_m is a second-order rate constant: it can only be measured properly if the substrate and the ribozyme are separate molecules.

Therefore, to study the sequence specificity of a group II intron, it was necessary to reconstruct the intron so that second-order rate constants for reaction could be measured. To this end, we created a ribozyme construct in which catalytic domains 1−3 of the intron were placed on one molecule, while the substrate (consisting of 5′-exon sequences) was placed on another (Figure 1). Domain 5, a 34-nucleotide hairpin motif which is required for catalysis by the intron (*30, 31, 32*), is provided separately as a small cofactor (*28, 33*). Because cleavage of the substrate occurs by nucleophilic attack of water, rather than through transesterification, the enzyme itself is not modified during the course of reaction. The use of a hydrolytic reaction in place of transesterification is not a radical departure from the natural behavior of group II introns, as water is commonly used as the nucleophile in the first step of splicing (*34, 35*) and hydrolysis represents a facile competing pathway during the first step of reverse-splicing (spliced-exon reopening) (*36*). In terms of its overall behavior, the trans-cleaving D13/5 ribozyme behaves as a good model system for understanding the first step of reverse-splicing (intron mobility). The two reactions share many features, including the same phosphorothioate diastereomeric preference (R_p)

(*33*, *37*), similar relative R_p- and S_p-phosphorothioate rate effects, similar relative rate constants for cleavage of ribose vs deoxyribose substrates (Q.X., A.M.P., and Justin B. Green, unpublished results), and a similar capability for efficient reaction in the absence of monovalent cations.

Using the trans-cleaving ribozyme system, the effects of mismatches in a variety of different contexts were evaluated. Relative k_{cat}/K_m values were obtained for mismatches at varying positions along both EBS1 and EBS2, for mismatches incorporated by mutating both substrate and ribozyme, and for different types of mismatches incorporated at the same location. To investigate the energetic basis for substrate specificity by group II intron ribozymes, individual k_{cat} and K_m values for mismatched substrate cleavage were also measured. Taken together, the data show that group II intron ribozymes are highly sequence-specific and that this is due to particular conformational features of the ribozyme.

EXPERIMENTAL PROCEDURES

Preparation of RNA Transcripts and Oligonucleotides. The ribozyme D13 RNA was transcribed from plasmid pT7D13, as described previously (*38*). Mutant plasmids pQL19 and pQL20, containing templates for transcription of ribozymes A331:C, A332:C, and A239:G, were constructed by Kunkel mutagenesis of plasmid pT7D13. The D5 cofactor RNA was transcribed from plasmid pJDI5′-75 as previously described (*32*). Oligonucleotide substrates either were transcribed from DNA oligonucleotide templates, using the method of Milligan and Uhlenbeck (*39*), or were synthesized on an ABI 392 DNA/RNA synthesizer using standard procedures (*40*), and worked up accordingly (*41*). All RNA transcripts and synthetic oligonucleotides were purified by denaturing PAGE. The oligonucleotide substrates were 5′-end-labeled with ^{32}P using polynucleotide kinase, followed by purification by PAGE as previously described (*33*).

Ribozyme Reactions and Kinetic Analysis. Prior to reaction, the individual RNA components (D13, D5, and S) were each heated to 95 °C for 2 min in storage buffer (10 mM MOPS, pH 6.5, 1 mM EDTA) to remove inactive conformations that can form at low temperature (*42*). After cooling to 42 °C, the D13 and S RNAs were combined together with reaction buffer (100 mM $MgCl_2$, 40 mM MOPS, pH 7.0) and preincubated for 10 min to allow proper folding. This incubation time was found to be sufficient for determination of the kinetic parameters reported here, having been tested by examining reaction at longer preincubation times (up to several hours). Reaction was then initiated by addition of D5, which had also been incubated separately in reaction buffer for 5 min. It would be ideal in studies of reaction specificity to initiate reaction by addition of substrate, rather than addition of a catalytic effector such as D5. That way, equilibrium between ribozyme and substrate is not being established prior to the initiation of chemistry. However, the two methods of initiating reaction are not expected to differ for reactions (such as those reported herein) where the k_{cat}/K_m for reaction of all substrates is $<1 \times 10^8$ M^{-1} min^{-1}, which is the established second-order rate constant for any type of duplex formation (*43*, *44*). Indeed, side-by-side comparisons of the two methods did not reveal differences in the pseudo-first-order rate constants for reaction of a given substrate (data not shown). Nonetheless, in this study, we chose to initiate by addition of D5, because substrate is important to the overall folding of D1 (*38*). When reaction of D13 is initiated by addition of WT substrate, there is a perceptible lag in the first 10% of reaction, which corresponds to the known rate of D1 folding (*38*). Although plots of the latter 90% of reaction result in a slope that is equal to k_{obs}, lag behavior in early time points can make it more difficult to compute a pseudo-first-order rate with high accuracy. To ensure that the data are as accurate as possible, the alternative methodology, involving initiation with D5, was employed in this study.

Experiments with mutant substrates and ribozymes were all performed at pH 7.0. Certain experiments conducted with the WT ribozyme/substrate combination were performed at pH 6.6, as indicated, to ensure that k_{on} did not become rate-limiting. At pH values greater than 7.0 for the WT ribozyme/substrate combination, k_{chem} becomes very fast (>0.065 min^{-1}) and k_{cat}/K_m becomes equivalent to the on-rate for substrate binding (k_{on}), which is approximately the same in magnitude as the known rate for duplex formation ($\sim 1 \times 10^8$ M^{-1} min^{-1}). Under such conditions, k_{cat}/K_m reaches a maximum value, although changes in k_{chem} can still be measured from pseudo-first-order rates at very high concentrations of substrate.

The total reaction volume was 20 μL, and the final concentrations of the reaction components were 6 μM for D5 (a saturating concentration) and 0.1−1 nM for ^{32}P-S, and the concentrations of D13 were varied as indicated. Samples of reaction were taken at the time points indicated, quenched with denaturing buffer, and loaded on a 20% denaturing polyacrylamide gel as previously described (*33*). Product and substrate bands were quantitated on a Packard Instantimager, and the data were processed according to standard procedures (*28*, *33*). Plots of the logarithm of substrate remaining vs time were used to obtain the pseudo-first-order rate constant for S cleavage at a given D13 concentration (*28*, *33*). Alternatively, progress curves for product formation were plotted and fit to a single-exponential equation, from which the rates were extracted (*34*).

To determine the kinetic parameters k_{cat} and K_m for reaction (which under the conditions reported herein are equal to k_{chem} and K_d), pseudo-first-order rates for substrate cleavage (k_{obs}) were plotted as a function of [D13] ribozyme for each mismatched species tested. The resulting apparent binding curves were fit to a standard 1:1 bimolecular binding isotherm as previously described (*32*), yielding k_{cat} as the horizontal asymptote of the plot and K_m as the [D13] at half-maximal rate. Values of k_{cat}/K_m were obtained by dividing k_{cat} by K_m. The specificity constant (k_{cat}/K_m) was also measured using a second method in each case. In this alternative approach, k_{obs} is plotted against very low [D13], below the K_m for association. This linear range of the rate vs concentration plot can be fit with a line, the slope of which is proportional to k_{cat}/K_m. This was the only means for determining the k_{cat}/K_m of the mismatched oligonucleotide -13C, which was cleaved with extremely low efficiency. It was also a valuable check on the accuracy of the other k_{cat}/K_m determinations, and the k_{cat}/K_m values determined by the two methods were within 2-fold of one another for any particular mismatch studied.

RESULTS

Development of the Ribozyme System. Previous studies on oligonucleotide cleavage by group II intron ribozymes were conducted using a construct in which domain 1 (D1) was combined with domain 5 (D5), to cleave an oligonucleotide substrate (*33*). While it would have been appropriate to use this minimal system for measuring relative k_{cat}/K_m values, D1 was a relatively unreactive ribozyme variant that functioned only under extremely high salt conditions (80 mM MOPS, pH 7.5, 100 mM $MgCl_2$, 1 M KCl). Addition of D3 to D1 radically increases the chemical rate of catalysis and lowers the salt requirements so that oligonucleotide cleavage occurs at a rate comparable to that of reverse-splicing (*28*). Because of its optimized kinetic behavior, D13/D5 was considered a better system for conducting extensive studies of reaction specificity. The D13/D5 ribozyme is one of several constructs with increased reactivity that provide insight into the mechanistic role of individual group II intron subdomains (S. Leuin, Q.X., and A.M.P., unpublished results).

As in the smaller D1 ribozyme system (*33*), the D13/D5 ribozyme cleaves substrate in a reaction that obeys Michaelis−Menten kinetics (S. Leuin, Q.X., and A.M.P., unpublished results), where the rate-limiting step of reaction is chemistry ($k_{cat} = k_{chem}$ and $K_m = K_d$). For this ribozyme, chemistry is rate-limiting at pH values ≤7.0, so that k_{cat}/K_m represents a combination of chemical rate and binding effects. While the D13/D5 ribozyme retains significant levels of activity under low salt conditions (10−25 mM $MgCl_2$), only a fraction of the ribozymes are properly folded at Mg^{2+} concentrations lower than 100 mM (*38*). Therefore, the reaction conditions which were used in this study involved incubation at 42 °C in a buffer consisting of 40 mM MOPS, pH 7.0, 100 mM $MgCl_2$, with no added monovalent ion. For the D13/D5 ribozyme, monovalent ion is not required for efficient reaction.

Effects of Base Mismatches in the Substrate Sequence. Single-base mismatches were incorporated at numerous positions within both IBS1 and IBS2 regions of the substrate (Figure 2). At one position (−4, see legend for nomenclature definition), more than one type of mismatch was incorporated in order to evaluate how different kinds of mismatch affect overall destabilization. The -4A, -7G, -11C, and -13C base changes were chosen in order to introduce the worst possible mismatched pair at the indicated positions. The other mutations, namely, -3G, -4G, -12C, and -15C, were designed to introduce noncannonical base-pairs (G-A and A-C) into the helices, and these were expected to be more stable. Despite these expectations, all of the different types of mismatches were highly destabilizing in a variety of different contexts, resulting in strong effects on reaction rate. At one concentration of D13 ribozyme (10 nM) and in the presence of saturating D5 (6 μM), the pseudo-first-order rate plots for oligonucleotide cleavage span a broad range of reaction efficiencies (Figure 3).

To determine the individual effects of mutations on the transition-state (reflected in k_{cat}) vs effects on ground-state binding (reflected in K_m), reaction rates were measured at varying D13 concentrations. The rates of reaction by each mutant were fit using a standard curve for bimolecular association (Figure 4), from which the kinetic parameters were obtained. Values of k_{cat}/K_m were computed from the individual kinetic parameters k_{cat} and K_m, and also by direct determination of rates at very low concentration of D13 (Figure 5). The resultant parameters were tabulated for comparison (Table 1).

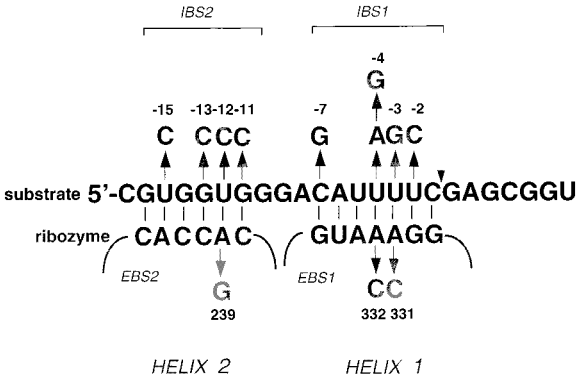

FIGURE 2: Illustration of base mismatches incorporated into the substrate and ribozyme. Note that positions on the substrate are designated by negative numbers (as is the convention for 5′-exon sequences), which become more negative as one proceeds 5′ from the cleavage site. The WT substrate sequence (top) is shown pairing to the EBS recognition elements on the ribozyme (bottom). Mutations are indicated by arrows showing the position at which a particular base change was made. The cleavage site is shown with a small triangle. Helix 1 specifies the IBS1−EBS1 interaction, and helix 2 specifies the IBS2−EBS2 interaction.

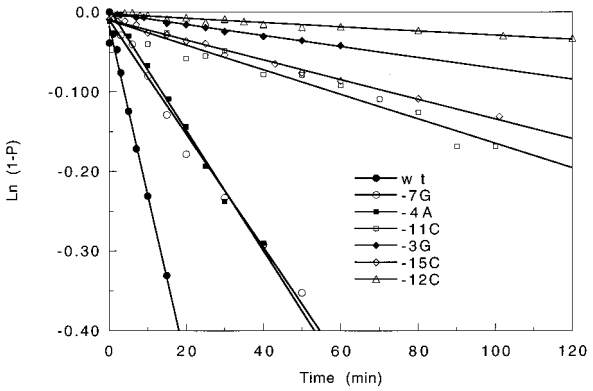

FIGURE 3: Representative pseudo-first-order plots for cleavage of mismatched substrates by the WT D13/D5 ribozyme. The base changes made in the substrate sequence are indicated by the position of the change (a negative number, keyed to Figure 2) and the new base inserted at that position. Pseudo-first-order rate constants for the cleavage reaction using substrate mutants were determined in the presence of D13 (10 nM), S (1 nM), and saturating D5 (6 μM) at 42 °C in 40 mM MOPS (pH 7.0 except for WT at pH 6.6), and 100 mM $MgCl_2$. The k_{obs} values determined from the slope of these plots are 0.021 min^{-1} for matched substrate (WT), 0.0071 min^{-1} for -7G, 0.0076 min^{-1} for -4A, 0.0015 min^{-1} for -11C, 0.0012 min^{-1} for -15C, 0.00027 min^{-1} for -12C, and 0.00069 min^{-1} for -3G (at 7.5 nM D13). No reaction was observed for -13C or -4G under these conditions.

All of the mismatched substrates showed large discrimination indices, representing large perturbations of activation free energy ($\Delta\Delta G^{\ddagger}$, last column, Table 1). This indicates that many types of base mismatches are poorly tolerated by group II intron ribozymes, which are therefore capable of accurate discrimination between matched and mismatched sequences. Interestingly, not all of the effects were on the K_m, or the energy of ribozyme−substrate binding in the ground state. Certain mismatches in both helix 1 and helix 2 resulted in large effects on k_{cat}, suggesting that mismatches

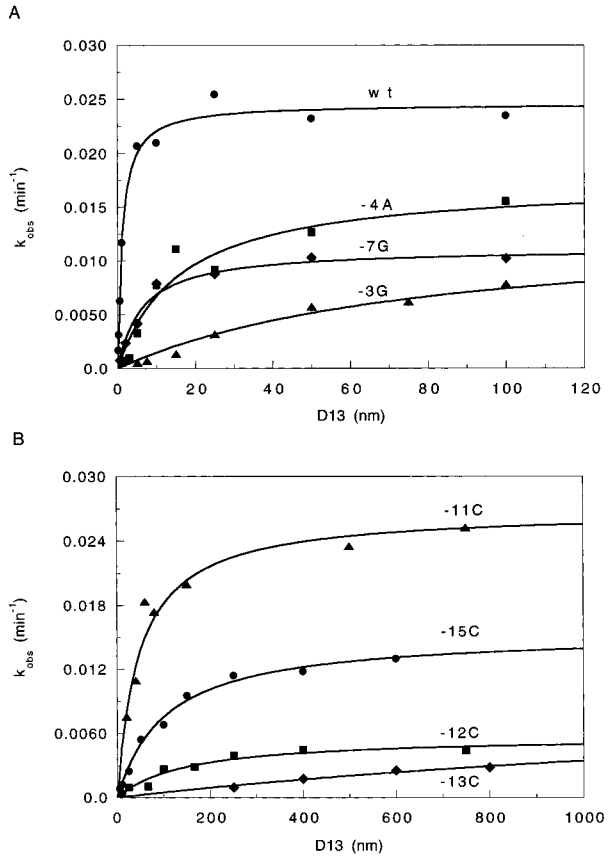

FIGURE 4: Determination of kinetic parameters for mismatched substrates and ribozymes. Shown are representative rate vs concentration plots, from which values of k_{cat} and K_m were derived. Single-turnover, pseudo-first-order kinetics (Figure 3) were performed to obtain k_{obs} values. K_m and k_{cat} values were determined by plotting k_{obs} values as a function of D13 concentration and then fitting the data to a bimolecular binding isotherm that has been published previously (32). (A) Kinetic behavior of WT and of mismatches introduced into helix 1 (the IBS1−EBS1 interaction). All of the data for the mutant substrates was obtained at pH 7.0. For ease of comparison with the mutants, WT rates shown here were measured at pH 6.6, where k_{cat} is ~3-fold slower and the K_m is the same as at pH 7.0. The kinetic parameters for WT, and all mutants, at pH 7.0 are provided in Table 1. (B) Kinetic behavior of selected mutants incorporating mismatches into helix 2. The k_{cat}/K_m value for the -13C mutant, which did not show saturation behavior, was obtained from the slope of the line shown. Values for k_{cat}, K_m, and k_{cat}/K_m, together with standard errors, are reported in Table 1 and described in the text.

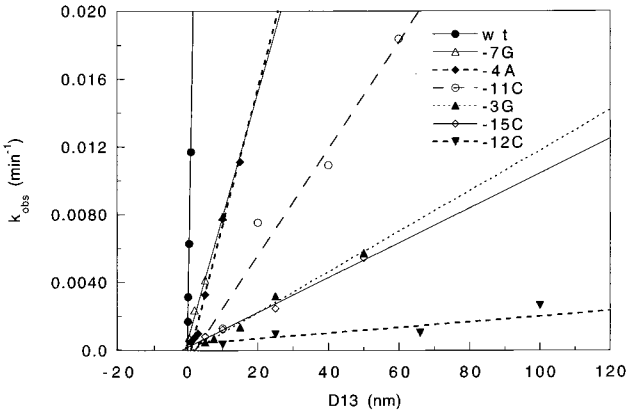

FIGURE 5: Plots of k_{obs} vs low [D13], in the linear range below K_m (k_{cat}/K_m plots). The k_{cat}/K_m values were determined from the slopes of these plots, and are within 2-fold of the values reported in Tables 1 and 2.

can cause perturbations in helical structure and that this has deleterious effects on the chemical environment of the ribozyme active site. The k_{cat} effects of mismatches in helix 2 are particularly intriguing given that this region is expected to be located far from the oligonucleotide cleavage site.

Individual mismatches incorporated into helix 1 resulted in a variety of different effects (Figure 4A, Table 1). The A-A pair introduced by the -4A mutant substrate increased K_m from 1.2 nM to 16 nM. It was expected, based on other studies of internal A-A pairs, that this mispair would be more highly destabilizing (*45, 46*). However, the effects of purine-purine pairs are reported to be extremely context-dependent, making it difficult to predict their behavior (*46*). A striking demonstration of this fact is provided by a comparison with the G-A mismatch introduced by the -4G substrate. Despite the fact that G-A pairs are predicted, in certain limited contexts, to be more stable than A-A pairs (*45*, *46*), the G-A pair at this position increased K_m by 200-fold relative to WT and 15-fold relative to an A-A pair. In addition to this marked K_m effect, the G-A pair resulted in a 10-fold greater effect on k_{cat} than the other mismatches incorporated in helix 1. This suggests that a G-A pair at position −4 introduces particularly damaging conformational perturbations of the ribozyme active site, despite the fact that this position is several nucleotides away from the cleavage site. Such "next to nearest neighbor" effects have been reported in studies of duplexes containing G-A pairs (*46*). Similarly, a G-A pair incorporated at position −3 resulted in large effects on both kinetic parameters.

The mispairs incorporated at the −4 position were placed in the center of helix 1. It was of interest to examine the effects of mismatches introduced at the end of the helix, at position −7. Introduction of a G-G mismatch at this position results in approximately 6-fold effects on both K_m and k_{cat} (Figure 4A, Table 1). In terms of ground-state energetics, the G-G mismatch behaves exactly like a terminal mismatch, resulting in an effect on the $\Delta G°$ of interaction (1 kcal/mol, calculated from the relative K_m value) that is in close agreement with that predicted by calculation of terminal mismatch free energies in this sequence context (1.1 kcal/mol) (*45*).

Individual mismatches introduced into helix 2 resulted in particularly strong effects on ground-state destabilization (Figure 4B, Table 1). This is reasonable considering that, in this particular intron, helix 2 is much stronger than helix 1. The magnitude of free energy perturbations appears to follow a clear trend, getting worse as the mismatch is moved from the terminus of the helix to the center (Figure 2, Table 1). As seen for the mismatch at position −7, a mismatch incorporated at position −11 in helix 2 resulted in ground-state free energy effects very close to those expected for a terminal mismatch, having a value ($\Delta\Delta G°_{GS}$ of 2.3 kcal/mol) that is not radically different from that predicted by nearest-neighbor calculations (1.7 kcal/mol). Mismatches at positions −12 and −15 resulted in very similar perturbations in ground-state binding, having K_d values of 150 and 104 nM, respectively. This is not surprising given that they are both the same type of mismatch, at nearly symmetrical positions around the center of the IBS2 helix. The k_{cat} perturbations introduced by these mismatches are significant and very surprising given that helix 2 appears to be far from

Table 1: Effects of Mismatches Introduced on the Substrate

substrate	IBS[a] region	mismatch introduced	k_{cat} (min^{-1})	K_m (nM)	k_{cat}/K_m (M^{-1} min^{-1})[b]	specificity index, $1/(k_{cat}/K_m)$ (rel)	$\Delta\Delta G^{\circ}_{GS}$ (kcal/mol)[c]	$\Delta\Delta G^{\ddagger}$ (kcal/mol)[d]
WT	–	–	0.065 ± 0.006	1.2 ± 0.20	$(5.4 \pm 1.3) \times 10^7$	–	–	–
-3G	helix 1	**G**-A	0.013 ± 0.001	80 ± 9.7	$(1.7 \pm 0.27) \times 10^5$	325	2.6	3.6
-4A	helix 1	**A**-A	0.017 ± 0.001	16 ± 3.1	$(1.1 \pm 0.27) \times 10^6$	50.0	1.6	2.4
-4G	helix 1	**G**-A	0.0011 ± 0.0001	240 ± 83	$(4.8 \pm 2.0) \times 10^3$	11300	3.3	5.8
-7G	helix 1	**G**-G	0.0110 ± 0.0004	6.3 ± 1.0	$(1.8 \pm 0.34) \times 10^6$	30.6	1.0	2.1
-11C	helix 2	**C**-C	0.027 ± 0.002	47 ± 11	$(5.7 \pm 3.1) \times 10^5$	95.3	2.3	2.8
-12C	helix 2	**C**-A	0.0057 ± 0.0006	150 ± 44	$(3.9 \pm 1.6) \times 10^4$	1390	3.0	4.5
-13C	helix 2	**C**-C	*[e]	*	$(3.4 \pm 0.63) \times 10^3$	15900	*	6.0
-15C	helix 2	**C**-A	0.015 ± 0.001	104 ± 13	$(1.5 \pm 0.24) \times 10^5$	366	2.8	3.7

[a] Helix 1 consists of the IBS1−EBS1 pairing; helix 2 consists of the IBS2−EBS2 pairing. [b] The k_{cat}, K_m, and k_{cat}/K_m values reported here were obtained by monitoring pseudo-first-order, single-turnover reaction rates as a function of ribozyme concentration. Reactions were performed at 42 °C in a buffer of 40 mM MOPS, pH 7.0, and 100 mM $MgCl_2$. Values for k_{cat}/K_m were also determined by a second method, whereby reaction rates were measured at very low concentrations of ribozyme in the linear range below K_m. Values obtained using the second method were within 2-fold of the values shown above, which are likely to be more accurate. [c] Values of $\Delta\Delta G^{\circ}_{GS}$ are differences in the binding free energy between matched and mismatched substrates in the ground state. Binding free energies (ΔG°_{GS}) were determined from the relationship $\Delta G^{\circ} = -RT \ln(1/K_d)$ for each substrate. [d] Values of $\Delta\Delta G^{\ddagger}$ are the differences in the activation free energy of reaction (a combination of ground and transition states) between matched and mismatched substrates. They were determined from the relationship $\Delta\Delta G^{\ddagger} = -RT \ln\{[k_{cat}/K_m(\text{mutant})]/[k_{cat}/K_m(\text{WT})]\}$. [e] The asterisk indicates that these values could not be measured, since the rates did not show saturation behavior. As such, only k_{cat}/K_m could be obtained for this mutant.

the cleavage site. The -13C mismatched substrate results in catastrophic defects in reactivity such that neither k_{cat} or K_m could individually be obtained. The reaction showed an inability to saturate at [D13] as high as 1 μM, and therefore a k_{cat}/K_m value was obtained from the linear fit to a plot of k_{obs} vs [D13] (Table 1, Figure 4B). It is notable that the same type of mutation was introduced at positions −11 and −13, and yet only the latter resulted in extreme effects on substrate reactivity. This is a strong indication that the center and the terminus of helix 2 are very different environments. It appears that the sequence context of helix 2 causes internal mismatches to be particularly destabilizing, disrupting the overall binding free energy to an extent greater than that observed for helix 1.

Another possible explanation for the k_{cat} effects reported here is that mismatches might affect the K_m for D5, which was provided at concentrations that would be saturating for the wild-type ribozyme. If mismatches in helix 1 or 2 disrupt D5 binding, then the D5 component would be present at subsaturation, thereby lowering the maximum rate that could be observed at D13 saturation (the observed value of k_{cat}). To test this idea, the K_m of D5 was measured for two D13/S complexes. The two substrates chosen for study (-4G and -12C) had shown the largest effects on k_{cat} of all the oligonucleotides tested (Table 1). In addition, by testing these two species, the D5 K_m effects of mismatches in both helix 1 and helix 2 could be assessed. The D5 K_m of the -4G mutant was 2 μM, which is 5-fold weaker than that measured for the WT complex (0.4 μM) in a side-by-side assay. This small increase in K_m would have little observable effect on the k_{cat} for oligonucleotide cleavage, however, since D5 was present at 6 μM in all the previous assays and this is well above saturation even for the -4G mutant (Table 1). The D5 K_m of the -12C mutant complex was found to be 5 μM, which is very close to the concentration of D5 added in the previous assays. However, despite the fact that D5 binding has been significantly weakened, the effect on the apparent k_{cat} can be no larger than 2-fold, since D5 was present at a concentration equivalent to its K_m value. Therefore, while it is interesting that the association of D5 is affected by the presence of mismatches in helices 1 and 2, providing further evidence that the active site is sensitive to structural perturbations in both recognition helices, the previous experiments were conducted at a D5 concentration that was sufficiently high to dampen effects on the measured k_{cat} value for oligonucleotide cleavage.

Effects of Ribozyme Mismatches and Their Suppressors. One possible explanation for the effects observed upon substrate mutation is that mismatches per se are not the cause of the dramatic declines in reaction efficiency, but rather that certain sequences inserted into the substrate are not readily tolerated. Although other studies have shown that proper pairing, rather than base identity, is essential for group II intron function (*47*), it was important to conduct controls to show that compensatory mutations restore reaction efficiency of the ribozyme reaction in this study. To this end, mismatches were made in helix 1 by mutating sequences in EBS1 of the ribozyme (Figure 2, Table 2), rather than on the substrate. Just like substrate mismatches, mismatches incorporated on the ribozyme are destabilizing and have dramatic effects on the discrimination index. In both cases where the ribozyme rather than the substrate was mutated, the cleavage rate was slower, indicating that the free energy of activation had been raised by approximately 2 kcal/mol. This indicates that structural and thermodynamic perturbations introduced by mismatches in the helices are important for discrimination of proper target sequences, rather than the choice of particular bases. Furthermore, it matters little whether ribozyme or substrate is mutated. To show further that there is discrimination against mismatches rather than discrimination against particular sequences, mismatches introduced in the ribozyme were corrected by the insertion of compensatory mutations in the substrate (Table 2, rows 3, 5, and 7). Compensatory mutations restored complete function when incorporated in either IBS1 or IBS2 (Table 2). In each case, due to the greater stability of G-C pairs relative to the WT A-U pairs normally found at these positions, binding between ribozyme and substrate was even tighter than observed for WT.

Contribution of an Energetic Penalty to Mismatch Discrimination. The relative k_{cat}/K_m values observed in this study indicate that group II intron ribozymes have an

Table 2: Effects of Ribozyme Mutant Mismatches and Suppressors[a]

mutant	IBS region	resultant pairing	k_{cat} (min^{-1})	K_m (nM)	k_{cat}/K_m (M^{-1} min^{-1})	$1/(k_{cat}/K_m)$ (rel)	$\Delta\Delta G^{\ddagger}$ (kcal/mol)
WT	–	–	0.065 ± 0.005	1.2	$(5.4 \pm 1.3) \times 10^7$	–	–
331C	helix 1	**U**-C	0.016 ± 0.001	4.2 ± 0.79	$(3.7 \pm 0.91) \times 10^6$	14.6	1.7
331C/-3G	helix 1	**G**-C	0.066[b]	<0.5[b]	$>5.4 \times 10^7$	<1	<0
332C	helix 1	**U**-C	0.016 ± 0.001	9.6 ± 2.1	$(1.7 \pm 0.47) \times 10^6$	32.5	2.2
332C/-4G	helix 1	**G**-C	0.058[b]	<0.5[b]	$>5.4 \times 10^7$	<1	<0
-2C	helix 1	**C**-G	0.011 ± 0.003	0.96 ± 0.019	$(1.1 \pm 0.33) \times 10^7$	9.2	–
239G/-12C	helix 2	**G**-C	0.071[b]	<1.0[b]	$>5.4 \times 10^7$	<1	<0

[a] The k_{cat}, K_m, and k_{cat}/K_m values were obtained by monitoring pseudo-first-order, single-turnover reaction rates as a function of ribozyme concentration (see plots in Figure 4). Errors were obtained from the fits to the data. The values for k_{cat}/K_m were also determined by a second method, whereby reaction rates were measured at very low concentrations of ribozyme in the linear range below K_m. Values obtained using the second method were within 2-fold of the values shown above, which are likely to be more accurate. [b] Standard errors for this ribozyme/substrate construct were not computed because K_m was only an estimate due to exceptionally strong binding, which was below the threshold of detection by this assay (0.5 nM). Errors of values for kinetic parameters k_{cat} and K_m rarely vary more than 2-fold, even using different stocks on different days.

Table 3: Free Energy Increments for Substrate Association[a]

ribozyme analog	$\Delta G°$(helix I) (kcal/mol)	$\Delta G°$(helix 2) (kcal/mol)	Σ(helix 1+2) (kcal/mol)	actual $\Delta G°$ of S (kcal/mol)	energetic penalty (kcal/mol)
D13[b]	−6.9	−10.1	−17.0 (−20.4)	−12.8	+4.2 (7.6)
calculated[c]	−7.0	−10.4	−17.4 (−20.8)	−12.8	+4.6 (8.0)

[a] During ribozyme−substrate association, an energetic penalty decreases the total binding energy attained through base-pairing interactions. Binding energy is decreased by ∼8 kcal/mol. The values shown in parentheses were computed in the case where only one entropic penalty was taken upon oligonucleotide association. [b] These values were obtained experimentally from direct binding experiments (P.Z.Q. and A.M.P., unpublished results), using published reaction conditions (*38*). [c] These values were obtained from the sum of nearest-neighbor interactions within the individual duplexes represented in Figure 2. In addition to the nearest-neighbor interactions between base pairs, the thermodynamic contributions of terminal mismatches were included in the calculation, together with a 3.4 kcal/mol penalty for helix initiation in each case. A C-U mismatch flanks the upstream terminal G-C pair in helix 2 (upstream with respect to the substrate). A G-A mismatch flanks the downstream terminal G-C pair in helix 2. An A-A mismatch flanks the upstream C-G pair of helix 1, and a G-U pair flanks the downstream C-G pair of helix 1. In the context of the folded intron, these flanking mispairs may not actually be able to stack, and the calculated value of $\Delta G°$ for each helix would be significantly smaller. However, the fact that the calculated values match the experimental ones suggests that this is not the case.

unusually high degree of sequence specificity. It was of interest to explore the structural and/or conformational basis for this effect. It has been suggested that ribozyme specificity [and perhaps other forms of macromolecular specificity (*48, 49*)] decreases if substrate binding energy becomes too high (*27*). A ribozyme needs many base pairs with its substrate in order to target a unique sequence. However, as the number of base pairs increases, the relative energetic cost of a mismatch declines. Under some conditions, mismatched substrates are bound so strongly that they are cleaved before they can fall off (*27*). One explanation for group II intron specificity is that, despite the lengthy 13-nucleotide recognition site, substrate binding is remarkably weak. Although the K_d of the D13 ribozyme is 1.2 nM, nearest-neighbor parameters for calculating the thermodynamic stability of helix 1 and helix 2 (*45*) predict that the dissociation constant should be in the femtomolar range. One explanation for the reduction in ribozyme−substrate binding energy is that some of the base pairs shown in Figure 2 might not be stable after complex association. However, this explanation is unlikely since mismatches at any position within the recognition helices were shown to have large effects on stability, comparable to effects expected for simple duplexes (Table 1). Instead, it appears that some other feature lowers the net binding energy while base-pairing between ribozyme and substrate is fully maintained.

As a first step in determining the source of the missing binding energy, it was necessary to quantitate the binding energy of short oligonucleotides corresponding to IBS1 and IBS2, respectively, to D13. From these data, the free energies of association for helix 1 and helix 2 were calculated and compared to binding of the full-length substrate (Table 3). The association energies of helix 1 and helix 2 were found to be in close agreement to values calculated using nearest-neighbor calculations (Table 3, columns 1−3). However, their sum is much greater than what is actually measured for substrate binding to the ribozyme (columns 4 and 5), particularly when one considers that a covalent connection would require only a single initiation penalty (values in parentheses, columns 4 and 6) (*45*). The data indicate that, when the IBS1 and IBS2 oligonucleotides are covalently connected and bound to the ribozyme, a substantial energetic penalty of ∼8 kcal/mol is paid upon binding.

DISCUSSION

By measuring the discrimination index for a set of matched vs mismatched substrates and ribozymes derived from a group II self-splicing intron, we have demonstrated that these ribozymes are exceptionally specific for targeting and cleaving their proper, matched substrates. The discrimination is so extreme that, in two cases, a single mismatch results in a 6 kcal/mol diminution in reaction activation energy, which is equivalent to a 16 000-fold rate effect. The smallest effects observed were 2 kcal/mol differences in activation energy for cleavage of matched vs mismatched substrates, which is still considerable by comparison to other ribozymes (*11, 13*) and sequence-specific nucleic acid targeting agents (*50, 51*). Importantly, the effects of substrate mutations were reversed when compensatory changes on the ribozyme restored pairing in the recognition helices and vice versa (Table 2). This confirms that group II introns specifically recognize any sequence with which they can base-pair, rather than conserved sequences containing specific functional groups.

Detailed kinetic analyses reveal that the discriminatory power of the group II intron ribozyme D13/D5 comes from three important sources: First of all, mismatches strongly destabilize the recognition helices, which lowers the ground-state binding energy. Second, mismatches also destabilize the transition state for chemical reaction. By incorporating energetically costly processes into the substrate/ribozyme interaction, the group II intron ribozyme lowers the total binding energy while maintaining the extensive set of base pairs that is required for recognition of a unique substrate.

Selectivity in the Transition State. When this study was initiated, the expectation was that mismatched substrates would simply bind more poorly to the ribozyme. This would result in sequence discrimination because the mismatched substrate would have a faster off-rate, resulting in release before cleavage occurs. Because it appears at first glance that group II introns recognize their substrates through simple base-pairing interactions between EBS and IBS sequences, this was a reasonable if simple model for specificity by group II introns. In other ribozymes, mismatches tended to have only small effects on the chemical step unless they were placed close to the cleavage site, where they caused overt structural distortions in the active site (*12*, *13*). Therefore, when initial results indicated that mismatches resulted in k_{cat}/K_m effects of >1000, it was very surprising since most single mismatches are not generally observed to have duplex destabilizing effects of this magnitude.

To understand the basis for these effects, individual values of k_{cat} and K_m were measured for each ribozyme−substrate combination (Table 1). These experiments showed that almost every mismatch, regardless of its position on helix 1 or helix 2, caused significant effects on k_{cat}. In the case of the -4G substrate, there is a 60-fold effect on k_{cat}, while in the case of -13C, the reaction rate was so damaged that individual kinetic parameters could not be dissected. This indicates that the ribozyme active site is sensitive to distortions in the helical geometry of both helix 1 and helix 2, despite the fact that the latter would appear to be far from the cleavage site. The sensitivity of k_{cat} and K_m to helical shape and distortion, rather than the loss of specific base functional groups, is reminiscent of the mechanism by which polymerases control the fidelity of nucleotide incorporation (*52*). The considerable k_{cat} effects resulting from mismatches in helix 2 suggest that this structure is actually close to the active site in three-dimensional space. In fact, the data imply that helix 2 is bent rather than coaxial with respect to helix 1, potentially swinging back toward the active site. In any case, mismatches along either helix disrupt or open the active site in such a manner that the rate of chemistry is reduced. This serves to increase the probability that a mismatched substrate will fall off before it is cleaved. Therefore, the group II intron ribozyme does not merely rely on the fact that a mismatched substrate will have a faster off-rate; the active site structure (and thereby the chemical step) is sensitive to distortions in helix 1 and helix 2.

Selectivity Due to Lower Ground-State Binding. A longstanding objective in ribozyme research has been to create a sequence-specific ribozyme capable of accurately recognizing a unique stretch of 11−15 base pairs. Indeed, sequence-specific recognition of long polynucleotide target sequences has been the focus of many investigations on synthetic agents that bind DNA (*51*, *53*). However, there are kinetic and thermodynamic reasons why sequence specificity and long recognition elements have been, up to this point, incompatible objectives in ribozyme research. Ribozyme target sites must be long in order to specify unique gene sequences; however, a long target recognition sequence results in extremely tight binding between ribozyme and substrate. When substrate binding becomes very strong, even mismatched substrates fail to fall off the ribozyme before they are cleaved (*27*). This results in a lack of discrimination between matched and mismatched sequences, as reported during studies of the *Tetrahymena* (*11*) and hammerhead ribozymes (*13*). One way that has been proposed to make ribozymes highly specific is to reduce the free energy/base pair in the recognition duplex (*27*). For example, theoretical analyses have shown that a ribozyme containing only A-U pairs (average $\Delta G°$ of 1 kcal/mol per base pair) in the recognition duplex would be able to discriminate mismatched sequences within a 14-nucleotide recognition element, whereas a ribozyme that forms only G-C pairs (average $\Delta G°$ of 2 kcal/mol) could discriminate mismatched substrates only within a 7-nucleotide recognition element. Given that it is impractical to forswear targeting of sequences containing G-C pairs, other means for lowering the free energy per base pair have been of great interest.

Here, we report that a group II intron ribozyme sensitively recognizes mismatches within a 13-nucleotide recognition sequence that contains a variety of different A-U and G-C pairing combinations. Introduction of mismatches at almost any position has large destabilizing effects that are similar in magnitude to those observed in simple duplexes. This indicates that each base pair formed between the group II intron ribozyme and its target substrate is worth the expected amount of energy for that type of pair. However, the ribozyme binds the substrate much less strongly than would be expected from the sum of base-pairing interactions within the recognition duplex. The observed free energy for ribozyme−substrate association is −12.8 kcal/mol (1 kcal/mol per base pair), while the sum of experimentally determined binding energies for helix 1 and helix 2, plus the calculated value for the sum of base-pairing interactions, is −20.8 kcal/mol. Through energetically unfavorable RNA folding events or rearrangements, the intron appears to have evolved a means for paying a costly energetic penalty of ∼8 kcal/mol upon substrate binding. In this way, the overall binding energy is lowered without losing the information content inherent in each base pair. Specificity is maintained by permitting equilibrium binding of long substrates to the ribozyme prior to the chemical step of reaction. Although these equilibrium binding effects are dramatic, it is important to consider that the binding energy and the chemical rate are closely connected in determining the discrimination index. Effects on both the ground state and the transition state are important for promoting sequence-specific binding of the ribozyme to its target. It is important to note that a destabilizing contribution of 5.5 kcal/mol has been shown for the single-stranded core of a hammerhead ribozyme, representing the difference between binding of intact substrates vs individual arms to the ribozyme (*65*). Therefore, the specificity that has been observed for hammerheads with shorter arms (*12*) may be attributable, in part, to destabilization by core folding.

It should be noted that this entire scenario is predicated on the fact that the chemical step of the D13/D5 ribozyme is relatively slow. As the chemical step increases in efficiency, the amount of sequence discrimination would be expected to decrease (*27*). Although it is difficult to anticipate what the chemical step would be for an engineered ribozyme in vivo, it is important to point out that rate enhancements beyond the relatively slow rates typically reported for group II introns and their ribozymes would be disadvantageous for sequence selectivity unless they are counterbalanced by large discriminatory effects on the k_{cat} for cleavage of mismatched substrates. The large transition-state effects reported here suggest that this may, in fact, be the case. However, the discriminatory power of this, or any other, ribozyme will depend on reaction conditions, such as temperature or ionic strength, making it difficult to anticipate the level of specificity under new conditions.

Mismatch Effects Provide Insights into the Active-Site Structure. In addition to information about specificity, the base mismatch effects also provide important insights into the active-site structure of a group II intron ribozyme. Understanding the relative orientation of helix 1 relative to helix 2 is absolutely essential for visualizing core structures and for model building. One possibility is that the two helices are coaxially stacked upon each other. Coaxial stacking is one of the most important determinants in the overall topology of a folded RNA molecule (*54, 55*), and given that there is often no gap between IBS1 and IBS2 in many introns, coaxial stacking of helix 1 and helix 2 is a reasonable model for their orientation. When helices co-axially stack, there are strong nearest-neighbor interactions between base pairs at the helical junctions (*56*). Thermodynamically, stacked helices behave as a continuous duplex, and mismatches in the junction pairs severely weaken helical stability, just as internal mismatches disrupt the stability of any duplex.

However, the mutational data show that mismatches in the center of helices 1 and 2 result in much larger effects than mismatches at positions −7 and −11, which are located at helical termini (Figure 3, Table 1). For example, in helix 2, the specificity index for -11C is 95.3 while that of -13C is 15 900. In helix 1, the specificity index for -7G is 30.6, while that of -4G is 11 300. These observations are consistent with the notion that mismatches in the center of a duplex are more destabilizing than mismatches at the terminus (*46, 57*). Quantitatively, the thermodynamic effects observed for the -7G and -11C mutants are in good agreement with the corresponding values calculated for mismatches at helical termini. Therefore, the data suggest that helix 1 and helix 2 behave as independent, separate duplexes that are not stacked upon each other. In addition to its structural implications, this result is significant because the stability of individual duplexes is radically increased upon coaxial stacking (*55, 56*). A lack of stacking between helix 1 and helix 2 may therefore serve to reduce the total amount of substrate binding energy, thereby increasing ribozyme specificity.

Because helix 1 and helix 2 are not stacked, the helices may be bent with respect to one another. Furthermore, the large k_{cat} effects evident for the helix 2 mutations suggest that the two helices might be acutely bent such that helix 2 is relatively near the cleavage site. The significant k_{cat} effects that are observed for any mispair introduced in helix 1 or 2 suggest that the active site is precisely packed and is therefore sensitive to any perturbation of helical geometry. A mismatch (particularly a purine-purine mispair) is likely to cause a helical distortion that pries open the active-site pocket, loosening important contacts in the transition state, and rendering catalysis less efficient. Although these models fit the data reported in this paper, it is important to emphasize that mismatch effects are highly context-dependent and this is precisely why it has been so difficult to predict their effects.

Insight into the Relative Strength of Different RNA Duplex Mismatches. There is limited published information on the thermodynamic effects of mismatches on duplex stability in RNA, particularly for pyrimidine-pyrimidine pairs. Thermodynamic data have been published on the effects of terminal mismatches, and studies have been conducted on purine-purine mismatches in the context of a G-C-rich duplex (*45, 46*). Most of the other work on mismatches concentrates on effects of tandem mismatches, which have altogether different effects on structure and stability than single mismatches (*58*). Very recently, measurement of single internal mismatches in long RNA helices by temperature gradient electrophoresis was reported, showing that base identity and nearest-neighbor context are important for mismatch stability (*66*). Because, to a first approximation, the group II intron binds its substrate through simple base-pairing interactions, it may be possible to utilize the ground state data reported herein for interpreting the thermodynamic effects of certain base mismatches. Several trends, in particular, are suggested by the data. First of all, G-A pairs in the sequence context of helix 1 are highly destabilizing and, significantly, are more destabilizing than A-A mismatches. It therefore seems possible that unprotonated G-A pairs (the study is conducted at pH 7.0) surrounded by duplex on either side form an arrangement that distorts the helix. The surface area of G is particularly large and, if it is placed accross from A, there may not be much room to accommodate the mispair within the duplex, causing it to open locally in order to accommodate stacking. A second important observation in this study is that the U-C pairs placed in helix 1 resulted in only minor effects on duplex stability, consistent with crystallographic studies of a duplex containing an U-C pair. In that context, the U−C pair induced only minor helical distortions and appeared to be stabilized by a set of direct and water-mediated hydrogen bonds between the bases (*59*). The two paired pyrimidines take up less space in the stack than a purine and a pyrimidine, and there is plenty of room in the duplex to accommodate them. In this sense, they may behave in an opposite manner relative to purine-purine pairs, inducing smaller structural distortions and thereby smaller effects on overall duplex stability.

Implications for the Applicability of Group II Introns as Therapeutics. Despite the fact that group I and group II intron ribozymes have many advantages, including high structural stability, reactivity with DNA, and the capability to perform splicing and reverse-splicing reactions, they have been largely overlooked as potential ribozyme therapeutics. There have been recent successes in ribozyme targeting in cultured cells using the hammerhead and hairpin ribozymes (*1, 2*), despite their apparent lack of sequence specificity in

vitro. This may indicate that ribozymes are more sequence specific in a cellular context, since abundant RNA-binding proteins may weaken ribozyme−substrate duplexes and effectively decrease the free energy per base pair. Alternatively, these studies may not be sensitive to problems of specificity. The ribozymes may, in fact, be attacking many gene products other than the target RNA, and, in cultured cells, the effects are less dramatic than inhibitory effects on damaging disease genes. However, it is possible that in a more complex whole organism, nonspecific effects might become a serious problem, leading to damaging side effects. Furthermore, if ribozymes are used to modify the DNA genome (as only the group I and group II ribozymes can), then nonspecific activity could lead to permanent genomic damage that might be evident only over the long term. Therefore, despite apparent success in developing ribozymes for use in gene therapy, it is important to remember that the development of highly specific ribozymes is an essential goal, just as target specificity is important in the development of any drug.

One factor that would seem to make group II introns less appealing than other ribozymes is the fact that the first introns studied were from yeast mitochondria and these all require high concentrations of Mg^{2+} (and often monovalents, as well) (*60*). It is likely that proteins substitute for high [Mg^{2+}] in the cell, just as the C5 protein reduces the salt requirements of RNase P in vitro (*61*). However, high salt conditions are not universal requirements for group II intron function (*9*, *62*). There are group II introns from other organisms, particularly those with a higher genomic G-C content (*63*, *67*, *68*), that have altered ionic requirements. These introns are likely to have advantages for therapeutic applications once they are reconstructed into ribozyme or trans-splicing constructs.

By placing EBS1 and EBS2 far apart in the secondary structure and requiring energetically costly transitions to occur when they are brought together upon binding the substrate, group II introns may have evolved a distinctive mechanism for ensuring that substrates would be cleaved with high specificity. This capability might have been developed during the evolution of group II introns as mobile elements (*20*), rather than as self-splicing RNAs. Group II introns can insert themselves into the DNA of new genomes through a highly specific process analogous to reverse-splicing (*9*, *62*). A requirement for specificity may have developed for a number of reasons. First of all, the intron must insert itself in-frame in order for encoded proteins to be properly expressed. In addition, the intron must be able to choose splice-sites properly after insertion or it will be unable to splice itself from host pre-mRNA and will thereby become a damaging transpositional lesion. It is likely that certain types of genes are particularly suitable for hosting a group II intron, and there may be genomic positions that are simply more suitable for invasion by mobile elements. In any case, there appears to be a strong requirement for specificity in group II intron mobility, and as a consequence, the structure of the intron may have evolved to facilitate proper substrate discrimination. In the design of therapeutics, it seems particularly reasonable to exploit a ribozyme that is naturally specific rather than trying to force a nonspecific ribozyme to behave better under a particular set of circumstances.

Comparison with Other Studies on Substrate Recognition by Group II Introns. Soon after group II introns were discovered, it was shown that they recognize 5′-exon sequences through a set of two base-pairing interactions: EBS1−IBS1 and EBS2−IBS2 (Figure 2) (*47*). In that study, it was reported that the entire IBS1 or IBS2 sequence could be changed, provided that compensatory changes were made in EBS1 and EBS2. The data reported herein are fully consistent with these early observations, extending the results by showing that even single mismatches in the recognition helices disrupt function, and full activity is restored upon incorporation of compensatory mutations. Recent studies on intron homing in vivo and in systems where reverse-splicing is reconsituted in vitro have shown that EBS−IBS pairings must remain intact in order to promote intron invasion of double-stranded DNA (*21*, *64*). In that work, even single-nucleotide mismatches were found to cause large disruptions in mobility reactions.

While the data reported herein are consistent with those observations, the results are distinct and significant in several important ways: First of all, this work was not designed to define the boundaries of recognition determinants on the exon or substrate; it was designed to quantitatively evaluate the substrate specificity of a group II intron ribozyme. By comparing k_{cat}/K_m values for different types of mismatches, this study shows the degree to which a group II intron ribozyme can discriminate the right target from the many possible wrong targets that it is likely to encounter. The only way that one can evaluate ribozyme specificity is by comparing specificity constants (k_{cat}/K_m values) and determining relative contributions to the free energy of discrimination (*27*, *29*). Having provided these parameters, which are prerequisites for rigorous analysis of ribozyme specificity, this study shows that group II introns are highly specific nucleic acid targeting agents, with discriminating power that exceeds any known ribozyme or synthetic gene-targeting compound described to date. Furthermore, the data suggest that group II introns have evolved a distinctive strategy for preserving high sequence specificity in the recognition of long polynucleotide substrates: An energetically unfavorable conformational state balances the considerable base-pairing energy of substrate binding, to lower the total binding energy and enhance substrate specificity. This suggests that RNA molecules have developed structural features designed not only to promote stability and chemical reaction but also to carry out subtle functions such as the maintenance of specificity.

ACKNOWLEDGMENT

We thank Qiaolian Liu for mutagenesis and cloning in the preparation of mutant ribozymes.

REFERENCES

1. Kijima, H., Ishida, H., Ohkawa, T., Kashani-Sabet, M., and Scanlon, K. J. (1995) *Pharmacol. Ther. 68*, 247−267.
2. Rossi, J. J. (1995) *Trends Biotechnol. 13*, 301−306.
3. Jones, J. T., Lee, S. W., and Sullenger, B. A. (1996) *Nature Med. 2*, 643−648.
4. Roth, J. A., and Cristiano, R. J. (1997) *J. Natl. Cancer Inst. 89*, 21−39.
5. Sokol, D. L., and Gewirtz, A. M. (1996) *Crit. Rev. Eukaryotic Gene Expression 6*, 29−57.

6. Brar, D. S., Ohtani, T., and Uchimiya, H. (1995) *Biotechnol. Genet. Eng. Rev. 13*, 167−179.
7. Fitchen, J. H., and Beachy, R. N. (1993) *Annu. Rev. Microbiol. 47*, 739−763.
8. Sullenger, B. A., and Cech, T. R. (1994) *Nature 371*, 619−622.
9. Yang, J., Zimmerly, S., Perlman, P. S., and Lambowitz, A. M. (1996) *Nature 381*, 332−335.
10. Mikheeva, S., and Jarrell, K. A. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 7486−7490.
11. Herschlag, D., and Cech, T. R. (1990b) *Biochemistry 29*, 10172−10180.
12. Werner, M., and Uhlenbeck, O. C. (1995) *Nucleic Acids Res. 23*, 2092−2096.
13. Hertel, K. J., Herschlag, D., and Uhlenbeck, O. C. (1996) *EMBO J. 15*, 3751−3757.
14. Pyle, A. M., and Cech, T. R. (1991) *Nature 350*, 628−631.
15. Bevilacqua, P. C., and Turner, D. H. (1991) *Biochemistry 30*, 10632−10640.
16. Pyle, A. M., Murphy, F. L., and Cech, T. R. (1992) *Nature 358*, 123−128.
17. Hegg, L. A., and Fedor, M. J. (1995) *Biochemistry 34*, 15813−15828.
18. Berzal-Herranz, A., Joseph, S., Chowrira, B. M., Butcher, S. E., and Burke, J. M. (1993) *EMBO J. 12*, 2567−2573.
19. Joseph, S., Berzal-Herranz, A., Chowrira, B. M., Butcher, S. E., and Burke, J. M. (1993) *Genes Dev. 7*, 130−138.
20. Lambowitz, A. M., and Belfort, M. (1993) *Annu. Rev. Biochem. 62*, 587−622.
21. Eskes, R., Yang, J., Lambowitz, A. M., and Perlman, P. S. (1997) *Cell 88*, 865−874.
22. Lambowitz, A. M. (1989) *Cell 56*, 323−326.
23. Lazowska, J., Meunier, B., and Macadre, C. (1994) *EMBO J. 13*, 4963−4972.
24. Moran, J. V., et al. (1995) *Mol. Cell. Biol. 15*, 2828−2838.
25. Michel, F., and Ferat, J.-L. (1995) *Annu. Rev. Biochem. 64*, 435−461.
26. Michel, F., Umesono, K., and Ozeki, H. (1989) *Gene 82*, 5−30.
27. Herschlag, D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 6921−6925.
28. Griffin, E. A., Qin, Z.-F., Michels, W. A., and Pyle, A. M. (1995) *Chem. Biol. 2*, 761−770.
29. Fersht, A. (1985) in *Enzyme structure and mechanism*, W. H. Freeman, New York.
30. Jarrell, K. A., Dietrich, R. C., and Perlman, P. S. (1988b) *Mol. Cell. Biol. 8*, 2361−2366.
31. Franzen, J. S., Zhang, M., and Peebles, C. L. (1993) *Nucleic Acids Res. 21*, 627−634.
32. Pyle, A. M., and Green, J. B. (1994) *Biochemistry 33*, 2716−2725.
33. Michels, W. J., and Pyle, A. M. (1995) *Biochemistry 34*, 2965−2977.
34. Daniels, D., Michels, W. J., and Pyle, A. M. (1996) *J. Mol. Biol. 256*, 31−49.
35. Podar, M., Chu, V. T., Pyle, A. M., and Perlman, P. S. (1998) *Nature 391*, 915−918.
36. Jarrell, K. A., Peebles, C. L., Dietrich, R. C., Romiti, S. L., and Perlman, P. S. (1988a) *J. Biol. Chem. 263*, 3432−3439.
37. Podar, M., Perlman, P. S., and Padgett, R. A. (1995) *Mol. Cell. Biol. 15*, 4466−4478.
38. Qin, Z. F., and Pyle, A. M. (1997) *Biochemistry 36*, 4718−4730.
39. Milligan, J. F., and Uhlenbeck, O. C. (1989) *Methods Enzymol. 180*, 51−62.
40. Scaringe, S. A., Francklyn, C., and Usman, N. (1990) *Nucleic Acids Res. 18*, 5433−5441.
41. Wincott, F., et al. (1995) *Nucleic Acids Res. 23*, 2677−2684.
42. Uhlenbeck, O. C. (1995) *RNA 1*, 4−6.
43. Herschlag, D., and Cech, T. R. (1990a) *Biochemistry 29*, 10159−10171.
44. Porschke, D., and Eigen, M. (1971) *J. Mol. Biol. 62*, 361−381.
45. Freier, S. M., et al. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9373−9377.
46. Morse, S. E., and Draper, D. E. (1995) *Nucleic Acids Res. 23*, 302−306.
47. Jacquier, A., and Michel, F. (1987) *Cell 50*, 17−29.
48. Roberts, R. W., and Crothers, D. M. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9397−9401.
49. Kehres, D. G., Subramanyan, G. S., Hung, V. S., Rogers, G. W., and Setzer, D. R. (1997) *J. Biol. Chem. 272*, 20152−20161.
50. Pilch, D. S., et al. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 8306−8311.
51. Trauger, K. W., Baird, E. E., and Dervan, P. (1996) *Nature 382*, 559−561.
52. Ahn, J., Werneberg, B. G., and Tsai, M.-D. (1997) *Biochemistry 36*, 1100−1107.
53. Trauger, J. W., Baird, E. E., and Dervan, P. B. (1996) *Chem. Biol. 3*, 369−377.
54. Murphy, F. L., Wang, Y.-H., Griffith, J. D., and Cech, T. R. (1994) *Science 265*, 1709−1712.
55. Walter, A. E., et al. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 9218−9222.
56. Walter, A. E., and Turner, D. H. (1994) *Biochemistry 33*, 12715−12719.
57. Narlikar, G. J., Khosla, M., Usman, N., and Herschlag, D. (1997) *Biochemistry 36*, 2465−2477.
58. Walter, A. E., Wu, M., and Turner, D. H. (1994) *Biochemistry 33*, 11349−11354.
59. Holbrook, S. R., Cheong, C., Tinoco, I., and Kim, S.-H. (1991) *Nature 353*, 579−581.
60. Pyle, A. M. (1996) *Met. Ions Biol. Syst. 32*, 479−519.
61. Reich, C., Olsen, G. J., Pace, B., and Pace, N. R. (1988) *Science 239*, 117−232.
62. Zimmerly, S., et al. (1995) *Cell 83*, 529−538.
63. Ferat, J.-L., Le Gouar, M., and Michel, F. (1994) *C. R. Acad. Sci. Paris, Life Sci. 317*, 141−148.
64. Guo, H., Zimmerly, S., Perlman, P. S., and Lambowitz, A. M. (1997) *EMBO J. 16*, 6835−6848.
65. Hertel, K. J., Herschlag, D., and Uhlenbeck, O. C. (1994) *Biochemistry 33*, 3374−3385.
66. Zhu, J., and Wartell, R. M. (1997) *Biochemistry 36*, 15326−15335.
67. Costa, M. Fontaine, J. M., Goer, S. L., and Michel, F. (1997) *J. Mol. Biol. 274*, 353−364.
68. Matsuura, M., et al. (1997) *Genes Dev. 11*, 2910−2924.

BI972661N