Selection of small molecules by the *Tetrahymena* catalytic center

Michael Yarus*, Mali Illangsekare and Eric Christian

Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309-0347, USA

Received October 29, 1990; Revised and Accepted February 4, 1991

ABSTRACT

The catalytic center in group I RNAs contains a selective binding site that accommodates both guanosine and L-arginine. In order to understand the specificity of the RNA for small molecules, we analyzed 6 RNAs that vary in this region. Specificity for nucleotides resides substantially in G264 rather than its paired nucleotide C311, and is expressed substantially in K_m , with comparatively little variation in k_{cat} . k_{cat} is not notably perturbed even for RNAs with mispairs in the active-site helix. For 5 of 6 sequences, effects of RNA substitutions on arginine binding and GTP reactivity are proportional, confirming that arginine contacts a subset of the groups occupied by G. As a result of particular mutations, reaction with GTP is decreased, and reaction with the natural nucleotides UTP and ATP is enhanced. Molecular modeling of these effects suggests that exceptionally flexible placement of reactants may be an essential quality of RNA-catalyzed splicing. The specificity of the intron can be rationalized by a type of binding model not previously considered, in which the G/arginine site includes adjacent nucleotides (an 'axial' site), rather than a single nucleotide, G264.

INTRODUCTION

The self-splicing reactions of a group I intron are initiated by a guanine nucleoside or nucleotide (collectively called G here) that binds to a highly selective pocket on the RNA (1).

The same pocket selects arginine. Arginine inhibits the self-splicing cascade of the group I RNAs (2). Inhibition is quite specific among sidechains; only arginine is a strong inhibitor (2). Arginine inhibition is kinetically competitive with GTP (2–3) and therefore amino acid site is in, or strongly coupled to, the G site defined in studies of splicing (1). Overlapping locations for G and arginine sites are also suggested by simultaneous effects of the same site-directed changes on both the G and arginine sites (4–5). The two molecules potentially occupy the same cavity as a result of the donation of the same network of hydrogen bonds from the side chain guanidino group of arginine as from the substituents of the 6-membered ring of G (2, 6). In addition, L-

arginine and *anti* guanosine in their most stable conformations are similar in size, and can be placed in a common site so that a favorable interaction for the α -NH⁺₃ of the amino acid is also implied (6).

These findings raise the possibility that selective RNA binding pockets for other small molecules might be possible. Such chemical versatility in RNAs is a central issue in discussions of the origin of life (7), and particularly in the origin of the genetic code (8), as well as in the mechanism of group I splicing.

One of the contacts made by G and arginine on the precursor RNA is known. Michel, *et al.* (5) made the substitution G264A:C311U in the conserved helix P7 (Figure 1) and consequently changed the specificity of the *Tetrahymena* active site from G toward 2-aminopurine and from arginine toward citrulline (5). These observations were accommodated in a binding model in which the incoming G formed a base triple with the conserved G264:C311 base pair.

We have pursued the specificity in this interaction between an RNA and its small-molecular ligands. What follows, accordingly, is an analysis of 6 RNAs that differ at the 264:311 base pair. These RNAs include two standard base pairs, as well as purine-purine (RR), and purine-pyrimidine (RY) mispairs, at the RNA active site. They were chosen in hope of sampling the behavior of varied base appositions. This study has led us to consider another type of binding model for G and arginine.

The phrase 'AC RNA' implies the sequence A264:C311 (Figure 1), RY RNA is purine 264:pyrimidine 311, and N refers to A, C, G, or U.

MATERIALS AND METHODS

Sequences cited

These include published sequences (9), 24 unpublished sequences from the mitochondrial DNA of *Podospira anserina* provided by Don Cummings, an *Ankistrodesmus* sequence from Mitch Sogin, and an *Anabaena azollae* sequence from M-Q. Xu, S.D. Kathe, S.A. Nierzwicki-Bauer, and David Shub.

Splicing Rates

Unlabeled T7 RNA polymerase transcripts (2) with the structure 5'exon-intron-3'exon were reactivated as previously described

* To whom correspondence should be addressed

(6), then briefly incubated at 55° with $\alpha^{32}\text{P}$ NTP, 3000 Ci/mmol, at 0.03–0.3 μM (NEN). When concentrations of GTP up to 560 μM were needed, the same $\alpha^{32}\text{P}$ GTP was diluted with unlabeled GTP (Pharmacia). 5 μl reactions also contained: 0.05 M Tris, pH 7.5, 0.06 M NaCl, 0.01 M MgCl_2 , 0.2 mM EDTA, 5 mM spermidine, and 0.2–0.8 μM precursor RNA. Incubations ranged from 30 sec with wild type to 5 min with less reactive RNAs (eg, AA and AG RNA). Reactions were stopped by addition of urea to 6 M, and EDTA to 0.025 M.

The reaction conditions are highly stabilizing in order to minimize the effect of structural defects on RNA phenotype. In addition, these conditions were selected to minimize the sensitivity of the RNAs to solution changes (eg, on addition of inhibitors). Mutants are often more sensitive to these nonspecific effects than is wild type. This yields potentially misleading inhibitions that are specific to mutated RNAs, yet not active-site-directed. These effects often appear when a complete binding isotherm is done, because inhibition at high concentrations of an ionic inhibitor becomes progressively anomalous.

The major ^{32}P labeled products at early times of reaction are intron-3' exon, and linear intron. These species were summed to measure the total rate of reaction with GTP. These bands were resolved from GTP by electrophoresis in TBE (2) on 5% acrylamide, 8M urea slab gels. The gels were dried at 80° under vacuum onto DEAE paper, and incorporated radioactivity was determined with a computerized 2D scanner (AMBIS Systems, San Diego, CA) at 25% efficiency.

RNA preparations (6) include all macromolecular RNA resulting from transcription, and therefore contain splicing products that have already reacted with GTP during transcription, as well as complete precursor (4).

Accordingly, A_{260} overestimates the intact precursor concentration by a factor up to 3-fold. Therefore if true absolute rates are desired, k_{cat} and k_{cat}/K_m should be multiplied by a similar factor. In fact, when measurements on gel-purified precursor were made, we obtained 0.14 m^{-1} for the k_{cat} of GTP and GC RNA. This should be compared to 0.064 m^{-1} in Figure 4. None of the arguments in the paper would be affected if corrections of this type were applied, but we have presented the data as observed. We elected to use minimally-treated RNA because it shows reproducible kinetic properties.

All values presented are the result of least squares fits to the data, using FitAll (MTR Software, Toronto).

Determinations of K_m & K_i by competition

Under the conditions of these assays, $\alpha^{32}\text{P}$ GTP concentration was 0.1–0.3 μM : that is, $\ll K_m$. With this proviso (2),

$$v_0/v = 1 + I/K$$

where v_0 is the uninhibited and v is the inhibited reaction velocity, I is the concentration of added inhibitor, and K is either K_i (if an added NTP or other inhibitor does not react) or K_m (if an added NTP reacts). Values plotted are least squares fits of the slope of v_0/v vs I plots. Note that tests of nucleotide inhibition demand exceptionally pure inhibiting nucleotides in order to avoid contamination by unlabeled guanosine derivatives. Pharmacia (Uppsala, Sweden) chromatographically purified NTP's were suitable. Sigma (St. Louis, MO) A-6144 ATP, though it contained only 4×10^{-5} mole fraction G, was not.

Interpretation of K_m in this ms is based on the generally valid equation (10):

$$K_m = (\text{enzyme})(\text{substrate}) / \Sigma (\text{all forms of enzyme engaged with substrate}).$$

Thus K_m has a reliable physical interpretation in terms of the total concentration of reaction intermediates, even when its detailed nature, however complicated, is not yet known.

In view of the above, the equivalence of K_i and K_m in the competition analysis is plausible. K_i is usually an equilibrium constant characterizing the disappearance of the RNA enzyme from the GTP reaction pathway by binding to the additional NTP. K_m , while not necessarily an equilibrium constant, is an algebraically similar measure of disappearance of RNA not just into a competitor's NTP-bound state, but into all forms of the ribozyme bound to the inhibiting NTP, however altered by the NTP-ribozyme interaction.

Determination of 5'-terminal triphosphates

RNA was eluted from crushed slices of acrylamide, then ethanol precipitated. P1 nuclease was obtained from BRL (Gaithersburg, MD) and used under the conditions recommended. PEI (plates from Brinkmann, Westbury, NY) chromatography was performed at 25° in 1.0 M KH_2PO_4 adjusted to pH 3.5 with phosphoric acid. Under these conditions, R_f = 0.93 (UTP), 0.79 (CTP), 0.66 (ATP), and 0.45 (GTP).

Molecular modeling

The models shown were drawn and manipulated with CAD software (Generic CADD, Generic Software, Redmond, WA). Atomic positions were obtained for 1- and 9-substituted bases from a crystallographic survey (11). Base pairs and triples with the third base paired to A or G264 were drawn using these positions. Bases were joined by the standard lengths for hydrogen bonds in base pairs (12). If groups approached but were not expected to H-bond, the van der Waal's radii for minimum potential were observed, and failing that, the radii for closest allowable approach (12). These constraints always implied a specific position for the C_1' of ribose, which is used as an indicator for the position of the reactive ribose $-\text{O}_3\text{H}$ of the splicing substrate.

Figure 1. Sequence and 2° structure of the catalytic region of the *Tetrahymena* precursor. Positions 264, 311, and 414 are marked. The bold-faced base pair at 264:311 is the sequence discussed in the text. P1, P2,... are conserved paired regions, or helices.

RESULTS

Reactivity with nucleotides

In displays of specificity data (Figures 2, 3, 4, & 5), we present values for k_{cat}/K_m , k_{cat} , K_m , and K_i on a logarithmic ordinate with faster rate or better binding consistently upward. The log of a rate or equilibrium constant is linearly related to a free energy. That is:

$$\log k = A + B \delta G^\ddagger$$

and

$$\log K = B \delta G^\circ,$$

where $\delta G^\ddagger$ and δG° are standard free energy changes in activation and reaction, respectively, k and K are rate and equilibrium constants, and A and B are numerical constants. Thus because Figures 2, 3, 4, & 5 use the same-sized ordinate scale, the same vertical spacing of points always implies the same free energy difference. This is indicated by the bar marked 1, indicating 1 kcal/mol at 55°C, at the right of each figure. Comparisons are simplified not only within a figure, but also between figures where the same structural change may appear as a shift in equilibrium in one figure, and a change in rate in another.

Figure 2 represents the reactivities of GC, GG, AG, AA, AU, and AC RNAs with low concentrations of the four standard nucleotides triphosphates (NTP's). Each point in the diagram shows the slope of a least squares line fitted to the rate of $\alpha^{32}\text{P}$ NTP incorporation at 5 concentrations. Such slopes, taken at concentrations where rates are first order in NTP, and normalized to moles of precursor RNA calculated from A_{260} , are k_{cat}/K_m , the second order rate constants for reaction of free NTP and the RNA enzyme (10). The $\log_{10} k_{cat}/K_m$ is therefore a linear index

Figure 2. Specificities (k_{cat}/K_m) for the reactions of GC, GG, AA, AG, AU, & AC RNAs with ATP, CTP, GTP, & UTP. The values are the slopes of lines relating the initial velocity of ^{32}P RNA production to nucleotide concentration (at nucleotide concentrations $\ll K_m$). k_{cat}/K_m is in reciprocal minutes and reciprocal micromolar, derived by normalization to the amount of RNA calculated from A_{260} . Filled downward-pointing triangles indicate that no reaction was observable, and are centered on the maximal rate (upper limit) that would have remained undetected. A tick near the center of each bar represents the best fitted value of the k_{cat}/K_m , and the flanking bars show the apparent standard error. Data for the six RNAs are presented in a consistent order in all figures, each linked to its RNA sequence below the plot. The bar marked 1 at the right shows the magnitude corresponding to an effect of 1 kcal/mol.

of the free energy of activation for reaction of precursor RNA with free nucleotide.

These 264:311 substitutions change nucleotide specificity. While CTP did not detectably react with any of the RNAs, the ATP and UTP columns of Figure 2 show faster reactions for certain mutants than for GC RNA (wild type). For examples, AG RNA is ≥ 12 -fold faster than wild type with UTP and GG RNA is ≥ 25 -fold faster than wild type in reacting with ATP. In one sense, AG RNA presents the most striking change; its reactivity with UTP is 1/6 its (decreased) reactivity with the natural substrate, GTP. Figure 3 shows that reactivity changes are usually mediated by changes in the K_m 's for these nucleotides. This strongly supports the identification of the 264:311 pair as a source of discrimination in reactions with nucleotide substrates (5). It further suggests that RNAs reacting rapidly with ATP or UTP might be created by further changes, despite the presently universal use of GTP by group I self-splicing RNAs. Creation of new specificities for the other natural nucleotides may entail new amino acid specificities.

Because these unusual reactions with ATP and UTP might have been attributable to contamination (eg, with GTP) of these $\alpha^{32}\text{P}$ NTP's, each product shown as significant in Figure 2 was characterized by isolating the putative pppN-intron RNA by gel electrophoresis. The 5' terminal label was released by P1 ribonuclease and resolved by thin-layer chromatography on PEI. All RNAs were 5'-labeled with the indicated triphosphate; that is, ATP and UTP were released by P1 nuclease from reaction products where these nucleotides had been used. Therefore Figure 2 presents reactions in which ATP and UTP are incorporated into *bona fide* 3'-5' linkage at the 5' terminus of the intron.

K_m and K_i via inhibition by nucleotides

In Figure 3, each point plotted is the result of a competition experiment, conducted under conditions in which the reaction

Figure 3. K_i and K_m for ATP, CTP, GTP, & UTP. Concentrations of chromatographically purified ATP, CTP, and UTP up to ca 10 mM (containing equimolar Mg^{++}) were added to reactions like those of Figure 2, and K_i was determined from inhibition of the reaction with trace $\alpha^{32}\text{P}$ GTP. Filled downward-pointing triangles represent upper limits for experiments in which no significant inhibition was observed for the concentrations used. Data for GTP are the quotient of k_{cat} (Figure 4) and k_{cat}/K_m (Figure 2). Each value is flanked by bars representing the standard error of the K_i implied by the fitted slope (ATP, CTP, or UTP) or the K_m calculated (GTP). The bar marked 1 at the right shows the magnitude corresponding to an effect of 1 kcal/mol.

velocity is first order in both GTP and RNA. Five concentrations of ATP, CTP, or UTP were added to reactions containing a constant low concentration of $\alpha^{32}\text{P}$ GTP, 0.03–0.3 μM . K_i or K_m was obtained from the slope of a plot of reciprocal fractional velocity versus inhibitor concentration (see Methods). GTP K_m 's were calculated from the quotient of the k_{cat} (Figure 4) and the k_{cat}/K_m (Figure 2).

The ordinate in figure 3 therefore is proportionate to the difference in binding free energies for competitive inhibitors. Differences in the $\log_{10}$ of the enzymatic saturation constant K_m may be considered proportional to apparent free energies in which any state containing enzyme and NTP, however altered by reaction, is counted as equivalently 'bound' (Methods). Competition has not been proven kinetically in all cases, but is likely because all these NTP's are strongly affected by substitutions at the active site.

The most straightforward of these results is the effect of substitution on K_m for GTP. All substituted RNAs require higher concentrations of GTP for saturation, even GG, the least kinetically perturbed RNA (4-fold increase in K_m). Thus all these substitutions alter the G site. In addition, though AU and AC have slightly lower K_m 's, it is clear that the A264 RNAs are a reasonably homogeneous class, having approximately the same apparent free energy of saturation with GTP despite the presence of four different nucleotides at position 311. The column of data for GTP also lies above all other data. Thus, despite substantial effects on specificity, none of these 264:311 substitutions obliterates the preference of the parental RNA for GTP.

It is notable that GG RNA is inhibited by CTP (Figure 3), though it does not measurably react with it (Figure 2). With this single exception, the occurrence of substantial inhibition by ATP, CTP, or UTP in Figure 3 is explicable as the consequence of a reaction with the RNA already detected (Figure 2). ATP shows enhanced competition for GG and AU RNAs and UTP competes

Figure 4. k_{cat} for the reactions. Central ticks represent k_{cat} in reciprocal minutes, calculated as the product of the data in Figure 2 (k_{cat}/K_m) and Figure 3 (K_m) for ATP, CTP, and UTP. GTP values and accompanying standard errors are extrapolated from reactions in which rates were measured up to 560 μM GTP. Flanking bars represent calculated errors, taking account of the errors in the two measured quantities that were multiplied for ATP, CTP, and UTP. Downward-pointing filled triangles are upper limits for cases where no significant reaction was observed. The bar marked 1 at the right shows the magnitude corresponding to an effect of 1 kcal/mol.

for AG and AU RNAs, though the wild type RNA is less sensitive. These alternative reactions decrease the limited amount of RNA catalyst for the $\alpha^{32}\text{P}$ GTP reaction, and allow the measurement of UTP and ATP K_m 's by competition in this assay.

k_{cat} 's

The maximal rates of ATP, UTP, and CTP reactions are obtained from the product of k_{cat}/K_m (Figure 2) and K_m (Figure 3). k_{cat}

Figure 5. Inhibition by added amino acids. These experiments were performed as for nucleotide inhibitors, but with amino acids added at ≤ 62 mM. Central ticks on bars represent measured K_i in millimolar, and flanking bars are the standard errors. Structures for the amino acids descend from their names at the top. Downward-pointing filled triangles represent lower limits for K_i in reactions that showed no significant inhibition over this concentration range. The bar marked 1 at the right shows the magnitude corresponding to an effect of 1 kcal/mol.

Figure 6. Relation between inhibition by arginine and reaction with GTP. The observed K_i of L-arginine (Figure 5) is replotted versus K_m for GTP (Figure 3) of the same RNA, using logarithmic axes. Both axes are therefore linear in free energy. The bars marked 1 at the right and top show the magnitude corresponding to an effect of 1 kcal/mol. The names of the RNAs are placed beside each point, or beside the lower limit (downward pointing triangle) for the K_i of AA RNA by L-arginine. A line of slope 1, characteristic of proportionality, has been put through the data.

may be calculated for observed reactions, or an upper limit may be placed on k_{cat} for unobservable reactions (as for CTP), corresponding to the maximum amount of reaction that would have gone undetected. k_{cat} for GTP reactions was obtained separately from measured rates at varied, high concentrations of $\alpha^{32}\text{P}$ GTP.

k_{cat} is remarkably consistent for all RNA/NTP combinations. The fourteen observed rate constants lie in a relatively narrow range (spanning < 6-fold, Figure 4). Even the upper limits for the slowest reactions (the ten unobserved) are potentially consistent with similar values. This contrasts with a range of more than 10,000-fold in the rate of reaction at low substrate concentrations, k_{cat}/K_m (Figure 2). Thus substitution of the 264:311 pair has little if any effect on the rate-limiting first order

step(s) of the NTP reaction with precursor RNA. For example, GC, GG, and AU RNAs, disparate in structure, show indistinguishable k_{cat} 's with GTP. It is equally clear that the nature of the substrate NTP has no consistent effect. For example, purines ATP and GTP, and pyrimidine UTP show similar k_{cat} 's with AU RNA. The generalization is valid even if the structure of the RNA active site and the substrate are simultaneously aberrant: the mispaired AG RNA reacting with pyrimidine UTP has a k_{cat} near the mean of all values (Figure 4). Neither the conformation of the P7 helix at the 264:311 base pair, nor the disposition of the substrate within this region, greatly affects the rate of reaction of the RNA saturated with substrate.

Figure 7. Geometry of some base triples in the perpendicular model. C_1 positions were derived as described in Materials and Methods. C_1 atoms are shown in Figure 6a around a large filled circle for the C_1 of G_GC, representing the normal substrate interacting with wild type G264:C311 RNA. Labeled small crossed circles represent G_AU, A_AU, U_AU, A_GC, and the hypothetical *syn-anti* pairs: U_A_{syn}G_{anti}, C_G_{syn}G_{anti}, and A_G_{syn}G_{anti}. **7a.** The G_GC triple: This diagram represents the active site configuration for GC RNA reacting with G, under the assumptions of Michel, *et al.* (5). G_GC has a close contact between O_6 of the incoming G, and H_4 of the C in the base pair. The $O_6 \cdots H_4N$ angle is ca 90° , so a substantial bond is unlikely. This contact cannot be resolved in the plane without exceeding allowed hydrogen bond distances, so this triple is likely to be non-planar. **7b.** The G_AU triple: This diagram exemplifies a perpendicular model for the doubly hydrogen-bonded complex with incoming G, potentially formed in AG, AA, AC, and (shown here) AU RNAs.

Figure 8. Axial Models for the G site. Base pairs (as shown in Figure 7) were tilted at 18° , and projected on a plane 4.4 Å from the helix axis. These are standard A form helix credentials (12). No perspective has been used in the representation of the RNA nucleotides A263, G264, or the A265:U310 base pair. Instead, lighter lines represent more distant bases, looking down the major groove from 3' toward 5'. As a result the shaded atoms, suggested for H-bonding interaction with incoming G, are projected to accurate positions relative to the helix axis. In contrast, the diagonally approaching G (upper right) is a perspective illustration, not fully docked in its suggested site, but intended to clarify the fit of the guanosine substrate to the helix groove. Dot-and-dashed lines connect atoms that can form hydrogen bonds when the G substrate is fully engaged. **8a.** Axial I. This drawing shows an axial G in a P7 helix that has an angular interval corresponding to 12 base pairs per helix turn. The perspective drawing of an axial G shows that it can be fitted to a normal A-prime helix. **8b.** Axial II. The RNA backbone has been allowed to depart from A form helical parameters, while still observing the range of backbone angles observed in tRNA (27). The perspective drawing of substrate G shows that phosphate oxygens may be combined with groups in the bottom of the major groove to compose a site.

The amino acid site

Figure 5 shows the K_i 's of 4 amino acids, L-arginine, L-lysine, L-glutamine, and L-isoleucine with these RNAs. Varied amino acid concentrations were added to reactions containing the RNAs and low concentrations of $\alpha^{32}\text{P}$ GTP, and the K_i determined from the observed inhibition. This experiment is analogous to that of Figure 3, but employs amino acids as competitors. L-lysine may be considered an analogue of arginine with a similar charge and charge distribution, but different side chain H-bond potential. Glutamine is a representative uncharged polar amino acid, and isoleucine controls for effects attributable to the α -carbon and aliphatic part of the arginine side chain, if such exist. L-arginine is the major inhibitor, with weak inhibition by lysine detectable for some RNAs. Selectivity for the arginine over the lysine side chain by GC RNA is 44-fold in K_i or 2.3 kcal/mol. This is somewhat greater than the distinction previously measured under different reaction conditions (2), but it is in agreement with a binding model (6) proposing hydrogen bonds to the guanidino group (shared by G) that would not exist for lysine.

In Figure 6, the effects of substitution on inhibition by L-arginine and reaction with GTP are replotted and compared. A correlation is evident; an effect on amino acid binding tends to be accompanied by a closely related effect on reaction with GTP by the same RNA.

DISCUSSION

Changed specificity for nucleotides

GG, AU, and to a lesser extent, AC RNAs react more rapidly with ATP than does the parental GC RNA (Figure 2). AG and AU RNAs show enhanced reactivity with UTP, and AG RNA has reactivities in the same order of magnitude for GTP and UTP. Such absolute increases in reactivity with other normal nucleotides have not been previously reported, and are evidence that G264 was central to the specificity of this RNA.

Quantitatively, these mutants show both decreased interaction with GTP and decreased specificity for GTP. The parental GC (wild type) RNA makes a 5500-fold distinction (k_{cat}/K_m) in favor of GTP over ATP. All mutants are less discriminating. But though AA, AG and AC introns have lost about 1/2 the free energy of discrimination between GTP and ATP possessed by parental GC RNA (Figure 2), they continue to prefer GTP (Figure 2). Thus a usable G site persists.

It seems likely that binding and reactivity is modulated not only by the substrate's fit to the site but also by internal competition. The terminal G of the intron, G414, can enter the G site. For example, inhibition of the GTP reaction has been observed when G414 is freed from the 3' exon (14), and reaction of 5' exon-intron with GTP is facilitated when G414 is modified or removed (15). Moreover, G414 must enter the G site in most models of exon joining, and substitution of G264 alters the relative rate of exon joining in A414 and G414 RNAs (5). Thus when the G site is mutated, internal competition with G414 is altered (Figure 1). This effect may constrain interpretation of differences between RNAs with different sequences. However, all comparisons of different substrates of a single RNA presumably involve the same internal competition.

Reactivities can be interpreted in terms of interactions with N264 rather than N311. Consider the four k_{cat}/K_m data in the GTP column of Figure 2 that represent the pairs, AU/AC RNAs, and AG/AA RNAs. Reactivities vary between pairs but are similar within pairs. Therefore, the substitutions U311C, and

G311A change reactivity with GTP subtly, if at all. Even the difference between GC and GG RNAs is small, and may have another explanation (see below). Figure 4 suggests that these similarities in reactivity rest on a fundamental similarity in K_m . These data strongly suggest that N311 neither contacts incoming G, nor alters internal competition.

The consistent modest difference in activation free energy between RY and RR RNAs may be due to alternate conformations for RR RNAs (Figure 2). A particular example is *syn-anti* isomerization at the RR site. Both *syn* and *anti* conformations exist in solution, and the transition requires only ca 10 nanoseconds for a free purine nucleoside (16). This transition may be fast enough to equilibrate on the time scale of these measurements, even when hindered by the imperfect P7 helix (Figure 1). Isomerization of the free RNA enzyme to a form that interacts with substrate less well will increase K_m , reducing k_{cat}/K_m , as observed (Figure 2).

Implications of a constant k_{cat}

We observe that specificity is due predominantly to changes in K_m , not k_{cat} . In our single-turnover reactions, k_{cat} cannot involve a step after the chemical transformation. Physically, such similar k_{cat} 's may be interpreted in two ways. i) Rate-limiting k_{cat} steps may not engage groups on the 264:311 base pair or on the substrate. For example, an RNA rearrangement centered in a relatively remote location may limit the rate at which the RNA reacts with a nucleoside or nucleotide. ii) Alternatively, the active site could be so flexible that our mutations do not alter the rate of reaction.

We infer that flexibility is likely even if k_{cat} is not the chemical step. While phrased in terms of base triples (5) for concreteness, the following argument is not highly model-dependent.

The base of the substrate must engage the RNA in order to explain the specificity of reaction (Figure 2). ATP, UTP, and GTP engaged by their bases in the same site on AU RNA are unlikely to bring their ribose- O_3H to the same position in any site model, despite similar k_{cat} (Figure 3). Specifically, Figure 7a illustrates the difficulty in devising a ATP_GG RNA model in which the ATP ribose is near that in the normal GTP_GC RNA complex, or in devising a UTP_AU RNA isosteric with an ATP_AU RNA complex. Predicted positions for ribose C_1' (and the linked but unseen reactive ribose $-\text{O}_3\text{H}$) are distributed over a rough triangle with sides of 4–5 Å, which is the length of 3 covalent bonds (Figure 7a). Consequently, the RNA active site brings reactants into juxtaposition starting from divergent equilibrium spacings, and must be flexible. This behavior is unlike familiar catalysts. Similar dimensional tolerance would make an enzyme that accepts alanine also active with substrate methionine. More concretely, a substitution that moves an essential catalytic group by deletion of one C–C bond causes a 1000-fold change in rate for triosephosphate isomerase (17) as a result of a net displacement of about 1 Å (18). In a related system the *Tetrahymena* transesterification has a rate of ca 350 m^{-1} (19). Therefore a 10^4 -fold decrease would make transesterification rate-limiting in this system. Because this does not occur, even for extreme variations in the site and substrate, the site is likely to be flexible.

Flexibility may be required by the splicing mechanism. The NTP reaction, exon joining, as well as circularization may be regarded as the forward and reverse of a single transesterification, carried out by a single catalytic center (14, 20–22, but see 25) which moves with respect to the RNA substrate (23–24).

Because nucleosides reacting at this center include many derivatives of guanosine (1), guanosine at the 3' terminus of a short or long oligonucleotide (26), or internal 3' and 5'-linked G (5), it is plausible that catalysis at the G-site would be tolerant of molecular variation. Further, all substrates for transesterification except free G are bound to the catalytic center by secure external 2° interactions or 1° linkages that may enforce different conformations than for unconstrained G entering the catalytic center. A catalyst with a tolerance for the positioning of reactive groups may be implied by these requirements.

Specificity for Arginine

Comparison of the measured K_i 's for L-arginine (Figure 5) with the K_m 's for GTP (Figure 3) strongly supports the arginine binding scheme (6) in which arginine contacts a subset of the groups contacted by G, and should be similarly affected by most RNA substitutions. Most evidently, because these substitutions alter arginine's binding free energy (Figure 5), N264 provides a significant contact to bound arginine, as it does to G.

Less obviously, K_i for L-arg and GC, GG, AU, AC, and AA RNAs (Figure 5) stand in the same relation as do K_m 's for GTP and the same RNAs (Figure 3). K_i for AA RNA is a lower limit, and therefore can only be said to be consistent with this relationship.

Figure 6 shows the relation explicitly. The diagonal line has a slope of 1.0, showing that the log K_m 's (apparent free energies of saturation) for GTP are linearly related to the log K_i 's (free energy of dissociation) for L-arg with 5 of 6 RNAs. Therefore substitution of the 264:311 pair usually altered the free energy of interaction with L-arginine proportionately to the alteration in the G interaction. The amino acid presumably contacts the same groups, and makes similar bonds with the 264:311 pair as does G. However, G and L-arginine are not identical molecules, and AG RNA is the corresponding exception (Figure 6), which is more inhibited by L-arg than predicted by strict correspondence to G.

Possibilities for the Molecular Nature of the G/arginine site

These experiments seek to define the mechanism of small molecule selection by an RNA site. We now combine and interpret findings that clarify specificity.

Michel, *et al.* (5) showed that when GC RNA was converted to AU RNA, a new affinity for 2-aminopurine, corresponding to a new hydrogen bond donor, appeared within the active site (Figure 7). The only simple interpretation of this finding is that the substituent at C_6 on R264 is hydrogen bonded to the N_1 of an incoming purine. Thus there is good evidence for a G264 $C_6=O \cdots H-N_1$ G bond, placing substrate G in the major groove of helix P7 (Figure 1). However, all other contacts are undefined. For example, rotation of incoming G around the established G264 $C_6=O \cdots H-N_1$ G bond may yield other binding modes.

Figure 8 shows that this is not merely theoretical. Two ways of binding G are shown in which three successive P7 helix nucleotides (A263, G264, & A265) provide 3 normal hydrogen bonds (12) to a G substrate, and in which the base of G is therefore turned to more closely parallel the helix axis than in the base triple suggested by Michel, *et al.* (5). For brevity we will call these axial models, and the base triple (5) a perpendicular model, emphasizing the difference in angle between the G base plane and the helix axis.

Figure 8a and 8b show two axial models, in which the plane of G is related by a 110° rotation. The plane of the G in a perpendicular model would be between the orientations shown,

approximately in the plane of G264. Figure 8a (axial I; G $C_6=O$ into the page) presents an unperturbed A-type helix, and the atoms (shaded) interacting with incoming G in this axial model line up naturally in the major groove. All interactions are with conserved features of the bases of nucleotides 263, 264, and 265.

Figure 8b (axial II; G $C_6=O$ toward the viewer) allows the P7 helix to be more irregular, and have a larger average twist, though these adjustments do not require dihedral angles outside the range usually observed (27). Such adjustments may be facilitated because N263 is unpaired, and the entire site is near a helix terminus. Both conditions should relax structural constraints unavoidable in a continuous helix. In this model, an important contact for the guanine base is on the RNA backbone. Because the G base in axial II is not aligned with the line of sight up the helix, it is more difficult to judge the positions of the hydrogen binding groups when the G is in place, but the fit meets normal constraints (12).

Both axial models accommodate the previous data of Bass & Cech (1) on specificity for analogues of G by providing hydrogen bonds for $C_6=O$, N_1-H , and NH_2 of the guanine base. Axial models also explain the data of Michel, *et al.* (5) in that A264 RNA will lack the contact for N_1-H of G. A264 RNAs will have a site suited instead to 2-aminopurine riboside, as was observed. Finally, both axial I and II explain the k_{cat}/K_m 's measured in this present work; reaction with GTP should not depend entirely on N264, because other specific contacts are made to flanking bases. Therefore, A264 RNAs should continue to prefer G, and can bind arginine weakly, as observed.

There are two arguments, neither definitive, in favor of an axial model. Firstly, the perpendicular (Figure 7a) model provides no clear stabilizing interaction to the $C_6=O$ group of G. However, Bass and Cech (1) found that 6-thioguanosine was a poorer substrate than guanosine, and suggested an H-bond to the $C_6=O$ of G on this and other grounds. Non-nucleotide analogues of G that include an equivalent for the $C_6=O$ bind to the active site with greater avidity than those that do not (cf ethyl guanidinium vs guanylylurea, 6). Marlene Belfort and Renée Schroeder have characterized RNAs containing the *td* intron of bacteriophage T4 with all four nucleotides at the site corresponding to *Tetrahymena* position 263 (pers comm). They find that the K_m for G is increased 20- to 100-fold for the RNAs that correspond to U or G263 *Tetrahymena*, with respect to RNAs corresponding to A or C263 precursor RNA. This is as expected if nucleotide 263 makes a contact that facilitates reaction with G, as shown in both Figures 8a and 8b.

Secondly, GTP preference (Figure 2) is better comprehended in an axial model. Figure 7b shows that a perpendicular model allows base triples between G and A264 RNAs. (Figures 2 and 3). In fact, the base triple provides no clear rationale for the fact that A264 RNAs are a recognizable class notably poorer in their interaction with G than the G264 RNAs (Figures 2, 3, 4). Though the G must shift toward the outside of the helix to match A264 (compare Figure 7a and 7b, in which the base pairs are aligned vertically), the resulting interaction seems equivalent. Consideration of the RNA backbone (not shown) does not obviously change these conclusions. In contrast, the axial models readily accommodate a reduced interaction with G in A264 RNA, because a central contact within the site (G264 $C_6=O$) is withdrawn and not replaced in any A264 RNA.

Both perpendicular and axial models envision only the half-site containing the bonds to the guanine base. The role of the nearby exon-intron junction RNA (6, 13) and potential bonds to the ribose are still to be accounted for. However, the distinction

between the axial and perpendicular models may be made via the critical prediction that specificity can be changed when nucleotides flanking N264 are changed, a prediction that is being tested.

ACKNOWLEDGEMENTS

We thank Don Cummings, Mitch Sogin, and David Shub for their kindness in supplying unpublished group I RNA sequences, and Marlene Belfort for communication of unpublished experiments. MY acknowledges useful conversations with Dan Herschlag and Rob Kuchta. We thank the W.M. Keck Foundation for their support of RNA science in Boulder. These experiments were sustained by NIH research grant GM30881 to MY.

REFERENCES

1. Bass, B.L. and Cech, T.R. (1984) *Nature* **308**, 820–826.
2. Yarus, M. (1988) *Science* **240**, 1751–1758.
3. Hicke, B.J., Christian, E.L., and Yarus, M. (1989) *EMBO Journal* **8**, 3843–3851.
4. Yarus, M., Levine, J., Morin, G.B., and Cech, T.R. (1989) *Nucleic Acids Res.* **17**, 6969–6981.
5. Michel, F., Hanna, M., Green, R., Bartel, D.P., and Szostak, J.W. (1989) *Nature* **342**, 391–396.
6. Yarus, M. (1989) *Biochemistry* **28**, 980–988.
7. Benner, S.A., Ellington, A.D., and Tauer, A. (1989) *Proc. Nat. Acad. Sci.* **86**, 7054–7058.
8. Yarus, M. and Christian, E.L. (1989) *Nature* **342**, 349–350.
9. Cech, T.R. (1988) *Gene* **73**, 259–271.
10. Fersht, A. (1985) in *Enzyme Structure and Mechanism*, Freeman & Company, New York, pp.103–106.
11. Taylor, R. and Kennard, O. (1982) *J Mol Struct* **78**, 1–28.
12. Saenger, W. (1983) in *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, pp. 39, 118, 234.
13. Bass, B.L. and Cech, T.R. (1986) *Biochemistry* **25**, 4473–4477.
14. Inoue, T., Sullivan, F.X., and Cech, T.R. (1986) *J Mol Biol* **189**, 143–165.
15. Kay, P.S. and Inoue, T. (1987) *Nucleic Acids Res.* **15**, 1559–1577.
16. Hemmes, P.R., Oppenheimer, L., and Jordan, F. (1974) *J Am Chem Soc* **96**, 6023–6026.
17. Raines, R.T., Sutton, E.L., Straus, D.R., Gilbert, W., and Knowles, J.R. (1986) *Biochemistry* **25**, 7142.
18. Lolis, E. and Petsko, G., in J.D. Hermes, S.C. Blacklow, and J.R. Knowles (1990) *Proc Nat Acad Sci USA* **87**, 696.
19. Herschlag, D. and Cech, T.R. (1990) *Biochemistry* **29**, 10159–10171.
20. Kay, P.S. and Inoue, T. (1987) *Nature* **327**, 343–346.
21. Szostak, J.W. (1986) *Nature* **322** 83–86.
22. N.K. Tanner and T.R. Cech (1987) *Biochemistry* **26**, 3330–3340.
23. Been, M.D. and Cech, T.R. (1987) *Cell* **50**, 951–961.
24. Doudna, J.A. and Szostak, J.W. (1989) *Nature* **339**, 519–522.
25. Kay, P.S., Menzel, P., and Inoue, T. (1988) *EMBO Journal* **7**, 3531–3537.
26. Inoue, T., Sullivan, F.X., and Cech, T.R. (1985) *Cell* **43**, 431–437.
27. Holbrook, S.R., Sussman, J.L., Warrant, R.W., and Kim, S-H. (1978) *J Mol Biol* **123**, 631–660.