A Specific Amino Acid Binding Site Composed of RNA

MICHAEL YARUS

A specific, reversible binding site for a free amino acid is detectable on the intron of the *Tetrahymena* self-splicing ribosomal precursor RNA. The site selects arginine among the natural amino acids, and prefers the L- to the D-amino acid. The dissociation constant is in the millimolar range, and amino acid binding is at or in the catalytic rG splicing substrate site. Occupation of the G site by L-arginine therefore inhibits splicing by inhibiting the binding of rG, without inhibition of later reactions in the splicing reaction sequence. Arginine binding specificity seems to be directed at the side chain and the guanidino radical, and the alpha-amino and carboxyl groups are dispensable for binding. The arginine site can be placed within the G site by structural homology, with consequent implications for RNA–amino acid interaction, for the origin of the genetic code, for control of RNA activities, and for further catalytic capabilities for RNA.

THE AMINOACYL-tRNA SYNTHETASE REACTION DETERmines the code by covalently attaching an amino acid to a particular RNA sequence, the anticodon. All amino acids are activated for translation by attachment to one adenylic acid (AMP) [in a 5′ aminoacyl-adenylate (*1*)], then transfer to another AMP [the 3′ nucleotide (nt) of a transfer RNA (*1*)]. Thus, the attachment of amino acids to ribonucleotides for activation and coding is a universal quality of modern translation. This association must therefore also be an ancient quality of translation, predating all divergence to modern species. This suggests the working hypothesis: A catalytic RNA performed the original synthetase reaction.

An RNA aminoacyl-RNA synthetase requires specific affinity for its amino acid. A search was therefore performed for amino acid binding sites on RNA's, with the use of a new tool, recently available for such a search. Catalytic RNA's report their conformation through their catalytic constants, which sensitively reflect the state of the RNA throughout the catalytic domain. Therefore, I searched for stimulation or inhibition of RNA catalysis by amino acids, and now describe amino acid inhibition of self-splicing by the group I intron of the *Tetrahymena* ribosomal transcript (*2*).

Assay for amino acid effects. Progress of GTP through the *Tetrahymena* ribosomal RNA splicing pathway, as described by Zaug, Grabowski, and Cech (*3*), was usually measured (*4*) (Fig. 1A). Figure 1, B and C show the effect of amino acids on the formation of the G-labeled splicing intermediates. At the incubation time employed in Fig. 1, the splicing intermediates have reached pseudo steady-state levels (*5*). Therefore radioactivity in intermediate splicing products will be altered if either early steps (reaction with G) or later ones (splicing) are affected by amino acids.

All arguments are made on the basis of the levels or rate of formation of the IVS–3′ exon (*4*) because this product, unlike the 15-nt oligomer (Fig. 1B), is well resolved from the large radioactive background of [α-^{32}P]GTP. Measurements of IVS–3′ exon are also conceptually straightforward because the IVS–3′ exon (Fig. 1A) is the immediate product of G reaction, unlike L − IVS whose formation requires at least two steps. However, this argument could be reformulated, with only numerical changes, for any combination of products (*5*).

Survey of 20 L-amino acids. The effects of the 20 standard L-amino acids at 5 m*M* (save for slightly soluble L-tyrosine, tested at 2 m*M*) are in Fig. 1C (SE = ±10 percent). The ratio of IVS-3′ exon with amino acid to the level in a control without added amino acids is plotted for amino acids grouped according to the chemical character of their side chains. There is one clear inhibitor: L-arginine. It is particularly notable that L-lysine, which has the same positive charge as arginine at *p*H 8, shows much less inhibition. Specificity is also evident from comparison of lanes in the autoradiogram of Fig. 1B, which shows the products of reactions containing other amino acids side-by-side with a reaction containing 5 m*M* L-arginine.

Mechanism of inhibition and binding constants. To determine the mechanism of inhibition, initial velocities for IVS–3′ exon formation were determined for a range of GTP concentrations, in the presence of 0, 2.5, and 5.0 m*M* L-arginine (Fig. 2, A and B; Table 1). Depression of IVS–3′ exon levels by L-arginine is still observed when initial velocities are measured (Fig. 2; compare Fig. 1). Therefore, L-arginine inhibits reaction with GTP instead of, for example, solely stimulating the further reaction of RNA intermediates.

Inhibition also decreases as GTP is increased and is markedly reduced at 2.5 μ*M* GTP, implying competitive inhibition. Indeed, both double reciprocal plots in Fig. 2 converge on statistically indistinguishable maximum velocities with and without arginine, as required for competition of L-arginine with GTP.

Dissociation constants for arginine and GTP are functions of the Mg^{2+} concentration, as might be expected for RNA catalysis. The K_m for GTP is around 2 μ*M* at 10 m*M* Mg^{2+}, and the K_i for L-arginine is about 4 m*M*, as measured at both 2.5 m*M* and 5.0 m*M* L-arginine (Fig. 2A and Table 1). Because further covalent reaction of GTP with the intron is slow, I assume that K_m for GTP reflects GTP binding.

If Mg^{2+} concentration is lowered to 5 m*M*, binding of GTP and of L-arginine improves by a similar factor. For GTP, K_m is 0.63 μ*M*

The author is professor of biology in the Department of Molecular, Cellular, and Developmental Biology, Campus Box 347, University of Colorado, Boulder, CO 80309.

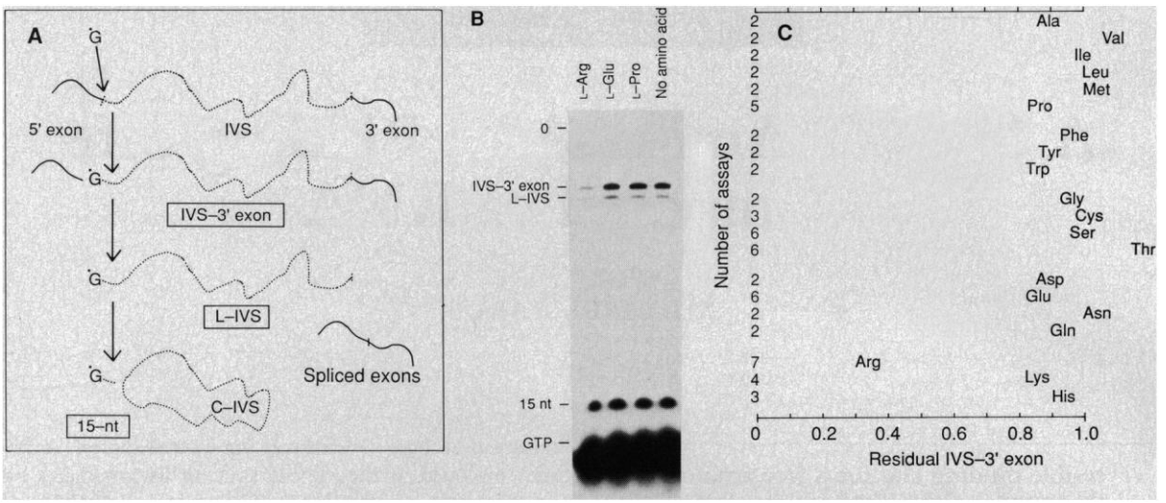


Fig. 1. (**A**) Products expected when labeled G reacts with an unlabeled precursor transcript. The 537-nt precursor RNA is a T7 RNA polymerase runoff transcript consisting of a short 5′ exon, the complete intron, and a short 3′ exon cloned from the *Tetrahymena* ribosomal RNA (rRNA) gene (*26*). An isotopically labeled guanosine, supplied as substrate, first attacks the precursor RNA to form a 5′-labeled IVS–3′exon intermediate. This molecule releases a 5′-labeled L − IVS when the exons are joined. The 5′ radioactivity then appears as a 15-nt riboligomer (or more rarely, 19 nt) when the L − IVS circularizes by transesterification to the 16th (or, more rarely, the 20th) nucleotide. Circularization also yields unlabeled C-IVS. The names of labeled products are boxed, and occur in order of appearance and in order of decreasing size, top to bottom in the panel. (**B**) Gel autoradiogram showing the products of a splicing reaction. The three expected labeled RNA's and unreacted [α-^{32}P]GTP (0.13 μ*M*) appear in this autoradiogram of the RNA products after 20 minutes at 37°C, resolved on a 4 percent acrylamide 8*M* urea gel (*27*). Lanes from left to right comprise reactions containing 5 m*M* L-arginine, 5 m*M* L-glutamate, 5 m*M* L-proline, and no added amino acid. (**C**) Effect on IVS–3′ exon levels of 20 L-amino acids. All except L-tyrosine were tested at 5 m*M*, L-tyrosine was 2 m*M*. Incubation (*4*) was relatively long (20 minutes at 37°C), so as to give steady-state levels of L − IVS and IVS–3′ exon. Each amino acid is represented by its standard three-letter abbreviation, centered on the fraction of control IVS–3′ exon levels measured in its presence. The column of figures at the left shows the number of assays averaged for each amino acid.

at 5 m*M* Mg^{2+}, lower by a factor of 3. For L-arginine K_i is 1.7 to 1.85 m*M* at 5 m*M* Mg^{2+} (Fig. 2B and Table 1), determined as in Fig. 2A by measurements at 0, 2.5, and 5.0 m*M* L-arginine. Thus arginine binding is also two to three times tighter at the lower Mg^{2+} concentration (compare Table 1). At both Mg^{2+} concentrations, about 2000 to 3000 times as much L-arginine as GTP is required to prevent GTP from occupying the catalytic site. However, L-arginine binds roughly as well to the rGTP site as do dGMP (K_i = 1.1 m*M*) and dideoxy G (K_i = 5.4 m*M*), nucleotide competitive inhibitors characterized by Bass and Cech (*6*).

Analysis of different amino acid concentrations. The specificity of the inhibition reaction was determined by measuring initial velocities of IVS–3′ exon production (*4*) at various inhibitor concentrations (Fig. 3). Within the precision of the assay, the reciprocal of the ratio of inhibited to control velocities is linearly related to the concentration of the inhibitor (Fig. 3B), as predicted for competitive inhibition. The reciprocal of the slopes in Fig. 3B is proportional to the K_i (*7*).

With regard to specificity, lysine is a much weaker inhibitor of the G reaction than L-arginine. Furthermore, D-arginine inhibits, but higher concentrations are required than for L-arginine. That is, arginine binding is stereoselective, and the prevalent biological L isomer is favored (Fig. 3, A and B). The derived K_i's (Table 1) show that the inhibitory site makes a significant distinction between D- and L-arginine. The K_i of L-arginine is 4.0 m*M*, indistinguishable from the K_i for L-arginine from Michaelis-Menten analysis (Table 1). D-Arginine is competitive by double-reciprocal analysis (*5*), and the dissociation constant agrees with that implied in Fig. 3B (K_i = 8.4 m*M*; Table 1). Therefore, by two methods of analysis, the stereoisomers of arginine are different by a factor of about 2 in their dissociation constants (0.4 kcal/mole) (Table 1).

This is not the first observed stereoselective reaction between RNA and amino acid derivatives. When DL-alanine imidazolides acylate rIpI, different imidazolides give D- or L-stereoselection at the internal 2′-OH, although the terminal 2′,3′ and 5′ acylations were not measurably stereoselective (*8*). Slightly stereoselective acylation of the 2′,3′-terminal with D-serine has been detected in a related system (*9*).

The site of inhibition. The inhibitory amino acid might plausibly act on either the RNA or the GTP substrate. For example, if an unreactive Arg-GTP complex were formed, so that dissociation had to precede reaction with the intron, this would appear indistinguishable from competitive inhibition of the catalytic RNA in the kinetic analysis above. This possibility is usually ignored for protein enzymes, but is significant in the splicing system because G and precursor RNA are chemically similar.

There are two clear possibilities for interaction of G, the small-molecule splicing substrate, and L-arginine. Perhaps the more evident possibility is that the positively charged side chain (or the α-amino group) of arginine might H-bond to the high negative charge density of the rG triphosphate (complex **1**, Fig. 4A). Although covalent substitution of G in this region does not usually decrease reactivity with the *Tetrahymena* precursor (*10*), this ionic and hydrogen-bonding interaction could play a part in a larger inhibitory complex. A second option for the arginine guanidino group is to hydrogen-bond to the 6-NH_2 and 7-azo groups of G (complex **2**, Fig. 4A), in analogy to the interaction that occurs in specific protein-DNA interactions (*11*). Complex **2** would tie up essential groups for reactivity of G with the intron (*10*). However, experimental data indicate that neither complex **1** nor **2** can be functionally significant.

Inhibition by way of interaction with the phosphate (complex **1**) is eliminated by experiments in which ribonucleosides are used as substrate (Fig. 4, B to D). A uniformly ^{32}P-labeled precursor RNA is spliced in the presence of low concentrations of unlabeled guanosine, 7-methylguanosine (me^7G), or GTP. Summing L − IVS and spliced exon (Fig. 1A) radioactivity yielded the spliced fraction. At low concentrations of substrates the slope of these

curves is V_{max}/K_m, the second order rate constant for reaction of the RNA with GTP, rG, and me^{7}G. Although GTP, rG, and me^{7}G are substrates that have the same order of reactivity with the intron (*10*), Fig. 4C indicates that rG is the best splicing substrate, showing a second order rate about twice that of GTP (Fig. 4B). GTP in turn has about twice the second order rate constant of me^{7}G (Fig. 4D).

At 5 m*M*, L-arginine inhibits splicing with all substrates. The GTP reaction is inhibited (Fig. 4B) to an extent consistent with the above experiments (Figs. 2 and 3), where the same precursor transcript and similar concentrations of GTP are used. Most important, inhibition of the rG and me^{7}G reactions is similar to that for GTP, although neither rG nor me^{7}G is phosphorylated. Thus complex **1** (Fig. 4A) cannot be responsible for splicing inhibition. Nor can complex **2** be the cause of L-arginine inhibition. Complex **2** would be blocked by methylation at the 7 position of G, and the corresponding nucleoside is still inhibited by arginine (Fig. 4D). I speak in terms of an L-arginine–RNA complex for the purpose of presenting further results, and also present further arguments against an inhibitory complex with G below.

Absence of effects on other reactions of the intron. The finding of competitive inhibition places the inhibitory arginine at or close to the G site on the intron. However, arginine might potentially interact with any nucleotide, and in particular, with G's (Fig. 4A). Therefore it seemed useful to look for other effects on the splicing reaction, as a means of detecting other arginine sites, if such exist. Accordingly, IVS–3′ exon and L − IVS labeled at their 5′ termini with [α-^{32}P]GTP were eluted from a gel. These intermediate RNA's were incubated with and without L-arginine in order to observe any effect of the amino acid on further G-independent reactions (Fig. 5, A and B). The IVS–3′ exon shows relatively slow hydrolysis at the IVS junction with the 3′ exon (*12*). L − IVS released in this way, or isolated from the gel, rapidly circularizes and releases 15-nt oligomer (*12*) (Fig. 1A).

These expected reactions are observed when the 5′ ^{32}P-labeled intermediate RNA's are incubated at 42°C (Fig. 5A). L − IVS disappears smoothly, and the radioactivity lost appears quantitatively in 15-nt oligomer (Fig. 5B, compare Fig. 1A). The IVS–3′ exon reaction accumulates a small amount of L − IVS (visible on the original films). The 5 to 10 percent L − IVS in the IVS–3′ exon reactions in Fig. 5A is appropriate for slow release of L − IVS, with subsequent rapid circularization. However, the major product of reincubation of the IVS–3′ exon is 15-nt oligomer, as expected (Fig. 5A, compare Fig. 1A). Most significantly, the progress of both these reactions is unaffected by the presence of L-arginine (Fig. 5, B and C). Therefore, L-arginine inhibition is limited to the initial reaction of G, and the amino acid has no measurable effect on the (rate-limiting) release of the L − IVS from IVS–3′ exon, or on circularization of L − IVS. Accordingly, the arginine that competes with G binding is the only one detectable by these functional assays.

Reversibility of the arginine interaction. The reversibility of arginine binding was tested in two ways. Incubation of L-[^{3}H]-arginine of high specific activity with the precursor gave no detectable bands on a gel autofluorogram, under conditions that produced dark bands with [^{3}H]guanosine of about one-third the arginine's specific activity (*5*). In addition, an arginine-inhibited reaction was ethanol-precipitated, then redissolved in a reaction mixture free of L-arginine. Reaction with GTP resumed at an uninhibited rate (*5*). The first observation shows that arginine is not retained by the RNA during gel electrophoresis, the second that arginine inhibition is reversed by chemically benign conditions that do not reduce reactivity with GTP. Therefore, L-arginine interaction with the precursor RNA is readily reversible.

Binding by arginine analogues. The limits of specificity of an RNA amino acid binding site were tested by measuring the ability of structural analogues of L-arginine to inhibit splicing of the *Tetrahymena* precursor (Fig. 6). The relative ranking and strength of the stronger inhibitors was evident from the intensity of gel bands (*7*). For quantitative purposes, the K_i was derived by least squares fit to plots like that in Fig. 3B and is listed alongside structural formulae for the inhibitors in Fig. 6.

First, the arginine site is stereoselective (Fig. 3), and L-arginine is preferred. In contrast, L-lysine (K_i = 46 m*M*) is one of the weaker inhibitors. Taken together with the absence of inhibition for 18 other amino acids (Fig. 1), this suggests that the guanidino group must be an essential binding element, because it is the only chemical characteristic unique to arginine. L-Lysine, for example, also has two positive charges (side chain and α-amino) spaced at similar distance.

However, the guanidino group alone is not sufficient for strong inhibition, as indicated by a K_i of 190 m*M* for guanidine hydrochloride. This suggests that the distal part of the side chain is also required for efficient inhibition, perhaps to fix the guanidino group in a binding pocket. A related analogue is guanidine acetic acid (K_i = 48 m*M*), which is a better inhibitor than guanidine HCl, despite

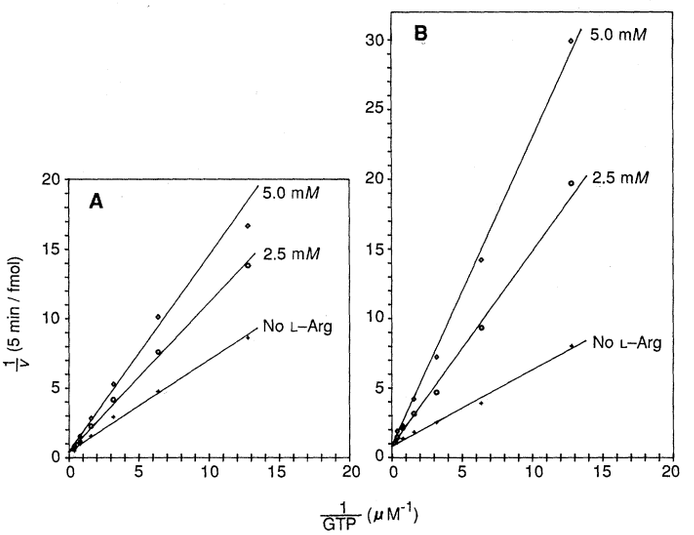


Fig. 2. (**A**) Lineweaver-Burk plot for the initial velocity of IVS–3′ exon formation at 10 m*M* $MgCl_2$, 37°C (*4*), in the presence of 0, 2.5, and 5 m*M* L-arginine. (**B**) Lineweaver-Burk plot for the initial velocity of IVS–3′ exon formation at 5 m*M* $MgCl_2$, 37°C, in the presence of 0, 2.5, and 5 m*M* L-arginine.

Table 1. Summary of kinetic and inhibitory dissociation constants.

Substance	K_m (μ*M*)	K_i (m*M*)	SE*	Arg (m*M*)	Mg^{2+} (m*M*)	Source
GTP	2.1		0.4	0	10	Fig. 2A†‡
L-Arg		3.7	1.4	2.5	10	Fig. 2A
L-Arg		4.7	3.8	5.0	10	Fig. 2A
GTP	0.63		0.05	0	5	Fig. 2B‡
L-Arg		1.7	0.1	2.5	5	Fig. 2B
L-Arg		1.85	0.05	5.0	5	Fig. 2B
L-Arg		4.0	0.3	Varied	10	Fig. 3B§
D-Arg		8.4	0.4	Varied	10	Fig. 3B

*SE is the standard error of the K_i in the same units. †The K_m given is the mean and standard error of four independent experiments at 10 m*M* Mg^{2+}, including Fig. 2A. ‡The value or values shown are not derived from the Lineweaver-Burk plots shown in the text, but are taken from plots of [GTP]/velocity versus [GTP]. To optimize the use of the data in view of the individual variance of the points, these linear plots were fitted with least-squares lines with the use of equally weighted points, a procedure suggested by the statistical analysis of Wilkinson (*28*). §Values given are for least-squares fit to the data represented in Fig. 3B. There are 50 individual measurements on the L- and 35 on the D-arginine curves.

the likely inhibitory effect of an added negative carboxyl charge.

The negative charge effect accounts for much of the data on the carboxylic acid group of L-arginine. When the carboxyl group is removed, as in agmatine (decarboxylated arginine), inhibition is substantially strengthened, K_i = 1.1 mM. The improvement is attributable more to the removal of the carboxyl charge rather than, for example, to a decrease in steric hindrance. Improved binding is also obtained when the carboxyl group is esterified, and enlarged: the methyl ester of Arg, K_i = 3.1 mM, and the Arg ethyl ester, K_i = 2.8 mM. Another derivative in which the size of the carboxyl is slightly increased is Arg-amide (the amide of arginine), the most effective inhibitor yet found. The K_i is 0.7 mM, and efficient inhibition by Arg-amide may reflect the ability of the added NH_2 to find adventitious weak hydrogen bond acceptors, an option that is unavailable to the aliphatic ester groups.

Although steric effects at the carboxyl are not strong, they exist. Neither of the arginine esters is bound as well as agmatine, suggesting a steric effect. The superiority of Arg-amide suggests that there are nearby electronegative atoms. Particularly, a massive carboxyl substitution in L-arginyl–L-leucine, K_i = 14 mM, negates the favorable effect of substitution of the carboxyl seen in the Arg-amide, but yields an inhibitor of similar order to L-arginine itself. The data suggest that the carboxyl group is near the RNA, but not strongly hindered. Rotational freedom in arginine may relieve potential contacts, and the nearby RNA may be easily deformed by additions to the arginine carboxyl group.

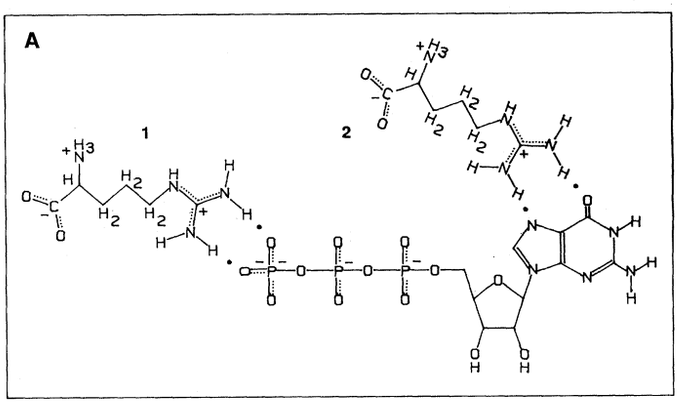
The α-amino group of L-arginine might have been expected to make an indispensable contribution to the RNA interaction, via an electrostatic interaction. Surprisingly, this is not so. L-Arginine in which the α-amino group has been replaced with a hydroxyl (α-hydroxy-Arg, L-α-hydroxy-δ-guanidinovaleric acid) is an improved inhibitor, K_i = 2.2 mM. The α-amino group's contribution to binding is therefore less than that of the uncharged α-hydroxyl, which may H-bond to the RNA.

Closer contact with the α-amino than the α-carboxyl is supported by the effect of the bulky amino substitution in L-leucyl-L-arginine, K_i = 90 mM. Recall that L-arginyl-L-leucine has K_i = 14 mM. This differential effect of leucine addition suggests that substituents of the amino groups are substantially more hindered.

The observed (Fig. 6) differential hindrance of the α-amino and carboxyl groups of L-arginine potentially accounts for the moderate stereoselectivity of the L-arginine site. Arginine, like all amino acids, prefers a given range of rotations at the α,β-carbon [prefers a range of α-amino–carboxyl orientations (*17*)]. Therefore interchange of the smaller α-amino with the larger carboxyl group (that is, a L → D switch) would produce an inhibitor which is more hindered and binds less well, as is observed.

The arginine site is not on free G, but on the precursor RNA. L-Arginine inhibits by binding to the intron, rather than to G. First, two plausible complexes of the amino acid and GTP have been eliminated by study of G analogues. L-Arginine inhibition by interaction with the phosphates of GTP, or by forming more than one hydrogen bond to the guanine base, can be excluded if that arginine inhibition persists when both these interactions with free G are singly or simultaneously prevented (see me^7G in Fig. 4D).

Studies of binding by arginine analogues (Figs. 1C and 6)



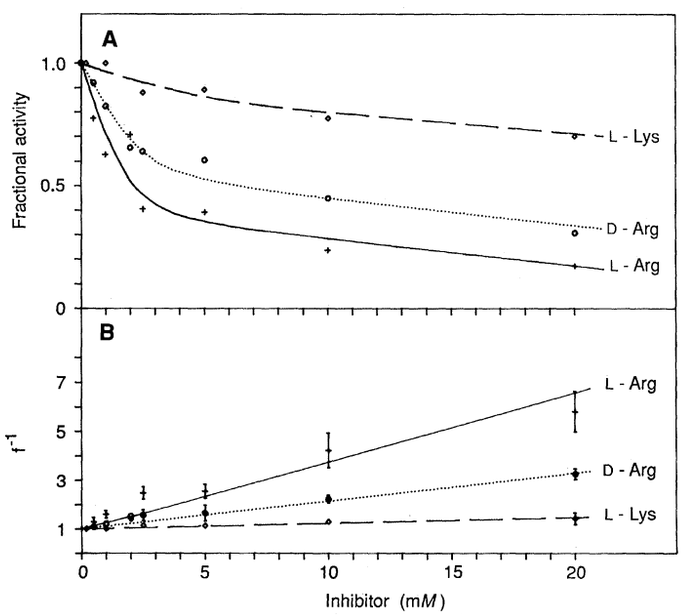


Fig. 3. (**A**) Fraction of the initial control velocity (*4*) for IVS–3′ exon formation in the presence of various concentrations of amino acids. [α-^{32}P]GTP concentration was 0.13 μM. (**B**) The same data as in (A) are plotted as the reciprocal of the fraction of the control velocity, f, as a function of the concentration of inhibitor. The standard error of the reciprocals is shown when the standard error is larger than the plotted symbol.

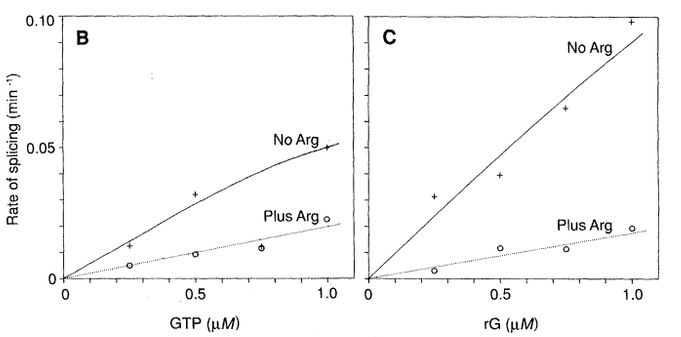


Fig. 4. (**A**) Two possible complexes of L-arginine and GTP are shown. Complex **1** posits hydrogen bonding and charge-charge interaction between the guanidino group of the arginine side chain and the phosphate backbone. Complex **2** is a hydrogen-bonded complex from the guanidino group to the 6-keto and 7-azo atoms of the guanine base. (**B**, **C**, and **D**) Incubation of [α-^{32}P]G, [α-^{32}P]C, uniformly labeled precursor RNA with low concentrations of unlabeled GTP, rG, and 7-methylguanosine (me^7G). Reactions were performed with and without 5 mM L-arginine, and splicing rates were measured by summing Cerenkov radiation from gel sections containing the L – IVS and spliced exons.

indicate the RNA, not the small molecule splicing substrate, is bound by arginine. The guanidino group of arginine must be bound, and the chiral α carbon of the arginine must simultaneously contact its site. No mechanism is evident to produce such a bidentate, simultaneous complex with free G. Removal of the characteristic groups at the chiral center of L-arginine actually improves inhibition instead of weakening the interaction between L-arginine and its site (Fig. 6). Thus the stereoselection must depend on simultaneous contact or near approach to the guanidino, α-amino, and carboxyl groups. An enveloping site is indicated, and the mass and extent required to envelop an arginine exist only on the RNA macromolecule.

Inhibition of splicing requires approach of the L-arginine and a negatively charged center, because of the clear effects of the charge of the carboxyl group (Fig. 6). Agmatine and the amide of L-arginine are more effective inhibitors than L-arginine of [^{3}H]rG reaction with the precursor (*5*). Thus, increased inhibition due to the absence of the carboxyl negative charge persists when the G splicing substrate is no longer negatively charged. That is, the negative center that must be approached to inhibit splicing is the phosphate backbone of the precursor RNA, not GTP.

Finally, previous chemical and biochemical searches for a complex of free G and arginine have not yielded the required complex. Lancelot *et al.*, using PMR chemical shifts for detection, found only a weak complex between an amino- and carboxyl-blocked arginine derivative and 9-ethylguanine [K = 670 mM in dimethyl sulfoxide (DMSO)] (*13*). Saxinger and Ponnamperuma (*14*) used carboxyl-linked amino acids on a permeable resin. They detected small redistributions by 5′-GMP to resin from the aqueous phase. However, redistribution was not specific to arginine, as lysine caused an effect of the same order as arginine, histidine was more effective than arginine, and tyrosine substantially more so. This weak interaction in aqueous solution does not have the side chain specificity of splicing inhibition (Fig. 6). Finally, Bruskov (*15*) measured the effect of L-arginine on the aqueous solubility of guanosine. The effect suggests a complex with a dissociation constant, K = 520 mM, more than two orders of magnitude too weak to explain splicing inhibition. The solubility effect is also not specific, with L-lysine exhibiting a significant K = 1180 mM, and even glycine, K = 1960 mM. Thus, there is no specific complex between rG and L-arginine which would be substantially formed at millimolar concentrations of L-arginine. It is still possible that an interaction with guanine is decisive for L-arginine inhibition (for example, analogous to **2**, Fig. 4A), but the decisive guanine would be a part of the RNA, rather than free G.

Only one arginine site on the RNA is detectable. It may at first be surprising that L-arginine does not inhibit circularization of the IVS (Fig. 5), because that reaction also utilizes the G site. However, the G site for the G substrate, and that occupied by the 3′ terminal G of the IVS during circularization are not identical (*16*). Even more to the point, Tanner and Cech (*16*) estimate that the 3′ terminal G of the intron has an effective concentration of 1 mM because the 3′ G of the RNA is covalently linked to the catalytic center. Because L-arginine inhibition is competitive and requires concentrations about 2000 times higher than G (Fig. 3 and Table 1), inhibition by competition with the 3′ G of the IVS would not be expected in the experiment in Fig. 5. This experiment therefore shows that no L-arginine binding at sites other than the G site can be detected by monitoring the rate of intramolecular circularization reaction of the 3′ G. Release of the IVS from the IVS–3′ exon is also unaffected by L-arginine (Fig. 5C), suggesting that there is no arginine bound near the IVS–3′exon junction.

A similar argument follows from the competitive nature of L-arginine inhibition, which implies (Figs. 2 and 3) that catalysis is quantitatively normal when the inhibitor molecule is excluded from the active site by substrate binding. Thus, at millimolar concentrations, there is no effect (even on the initial reaction of G with the intron) except for that due to L-arginine obstructing the G site.

A particular model for the arginine binding. I propose that L-arginine binds at a precisely homologous subsite within the G site already demonstrated (*6, 10*) on the intron. GMP and L-arginine could occupy the same site with a structural similarity between the atoms of the guanidino group, which exists in both arginine and guanine (Fig. 7). Bass and Cech (*10*) have presented evidence that atoms of the guanidino group are engaged during the catalytic interaction of G with its site. L-Arginine therefore might occupy the G binding site by virtue of a hydrogen-bonding and stacking potential which resembles G (Fig. 7).

The electrostatic potential is probably negative within the arginine side-chain pocket. Thus, the positive charge of the arginine

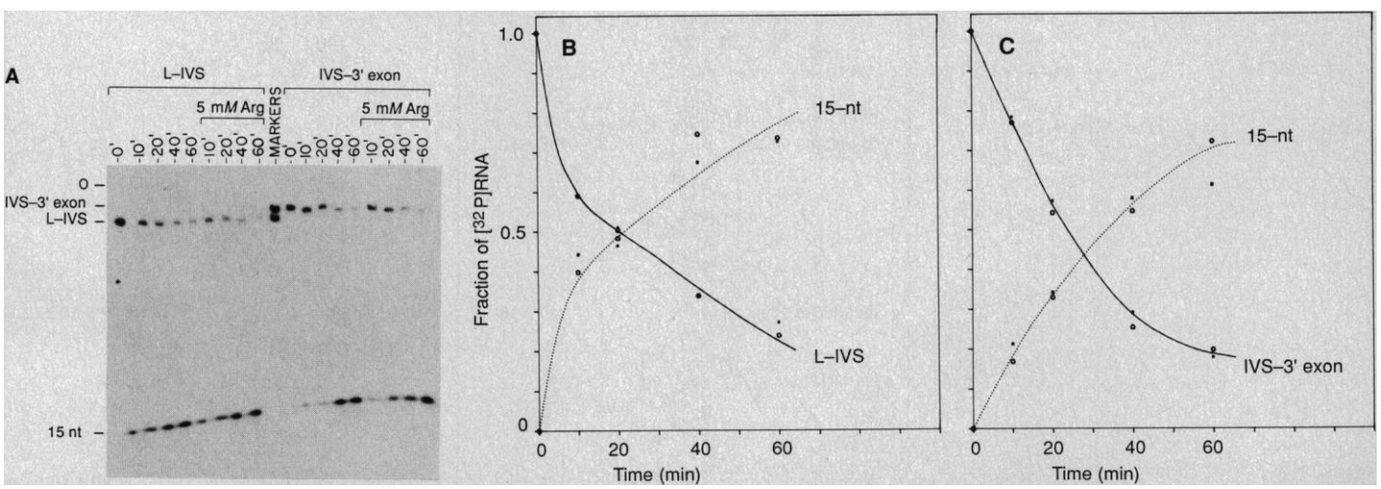


Fig. 5. (**A**) Autoradiogram of 4 percent acrylamide 8M urea gel, showing the products observed when gel-purified L − IVS and IVS–3′ exon labeled with [α-^{32}P]GTP were reincubated (*4*) at 42°C. In each set of lanes the samples consist, left to right, of a "0-minute" incubation control, then a time series without L-arginine, then a series containing L-arginine. The diagonal trend, particularly clear in the 15-nt bands, was produced by loading a gel that ran continuously. (**B**) Fraction of the incubated radioactivity (detected as Cerenkov radiation) recovered in dried gel sections after reincubation of 5′-labeled L − IVS. Small squares, no added amino acid; circles, 5 mM L-arginine. (**C**) Fraction of the incubated radioactivity (detected as Cerenkov radiation) recovered in dried gel sections after reincubation of 5′-labeled IVS−3′ exon. Small squares, no added amino acid; circles, 5 mM L-arginine.

guanidino group probably provides additional stabilization for arginine which does not exist for rG itself. However, arginine has no groups equivalent to the essential ribose hydroxyls (*6*), is a somewhat different shape, and does not form the third hydrogen bond analogous to that directed (*10*) at the (*10*) 6-keto group of G (compare Fig. 7). In fact, 2-aminopurine ribonucleoside, an analogue of rG which, like arginine, lacks the 6-keto group, but also lacks the H-1 atom, has been shown (*10*) to be a splicing substrate with K_m = 2.2 m*M*, which is similar to the *K* of arginine (Table 1). Therefore, a combination of the above differences between arginine and G can easily account for the 2000- to 3000-fold difference in association of L-arginine (Table 1), even though arginine and the nucleotides bind similarly.

The model (Fig. 7) is consistent with the parallel and similar quantitative increase in binding for both arginine and G (Fig. 3 and Table 1) as Mg^{2+} concentration decreases. Both molecules lie in similar orientations within the same site, and might well be affected similarly as the structure of the RNA varies as a result of binding and release of Mg^{2+}.

This mode of binding also accounts for all major qualitative features of the amino acid binding site. Binding and inhibition necessarily are specific to the arginine side chain, as observed. The space required for arginine to bind is assured, even though it is one of the bulkiest amino acids, because it lies along the axis of a space intended for guanosine, which is even larger. Competitive inhibition follows from the extensive overlap between G and arginine within the catalytic site. The observed freedom for substitutions around the α carbon of L-arginine is explicable, because this section of the molecule lies in the portion of the site otherwise occupied by the bulkier ribose (phosphate).

Finer details of the arginine binding can be explained by making a further assumption. In Fig. 7, the α,β and β,γ side chain rotamers are those preferred by arginine side chains in a sample of highly resolved protein structures (*17*). The distal part of the side chain is extended in the trans conformation usually preferred by such aliphatic chains. Thus the drawing arguably shows the most stable L-arginine conformation, but because there are four rotable bonds in the arginine side chain, the conformation shown is not unique. Further, the arginine could be rotated 180° (around the guanidino axis) to produce another similar binding mode.

Nevertheless, the binding mode in Fig. 7 explains the stereoselectivity observed (Fig. 2 and Table 1). The section of the site where the α-amino lies must be closed off, because it is to that side (outward, toward the viewer) that the attack by the 3′-OH of the G ribose on the exon-intron junction takes place during splicing. The exon-intron junction in the RNA precursor must therefore approach the ribose (and the α-amino of arginine) to within atomic dimensions. The ribose hydroxyls are probably both hydrogen-bonded to the precursor RNA (*6*), confirming that the downward and outward direction in Fig. 7 is closed. In contrast, the carboxyl group projects back into the volume otherwise occupied by the ribose or ribose phosphate of the G substrate. The carboxyl should therefore be less

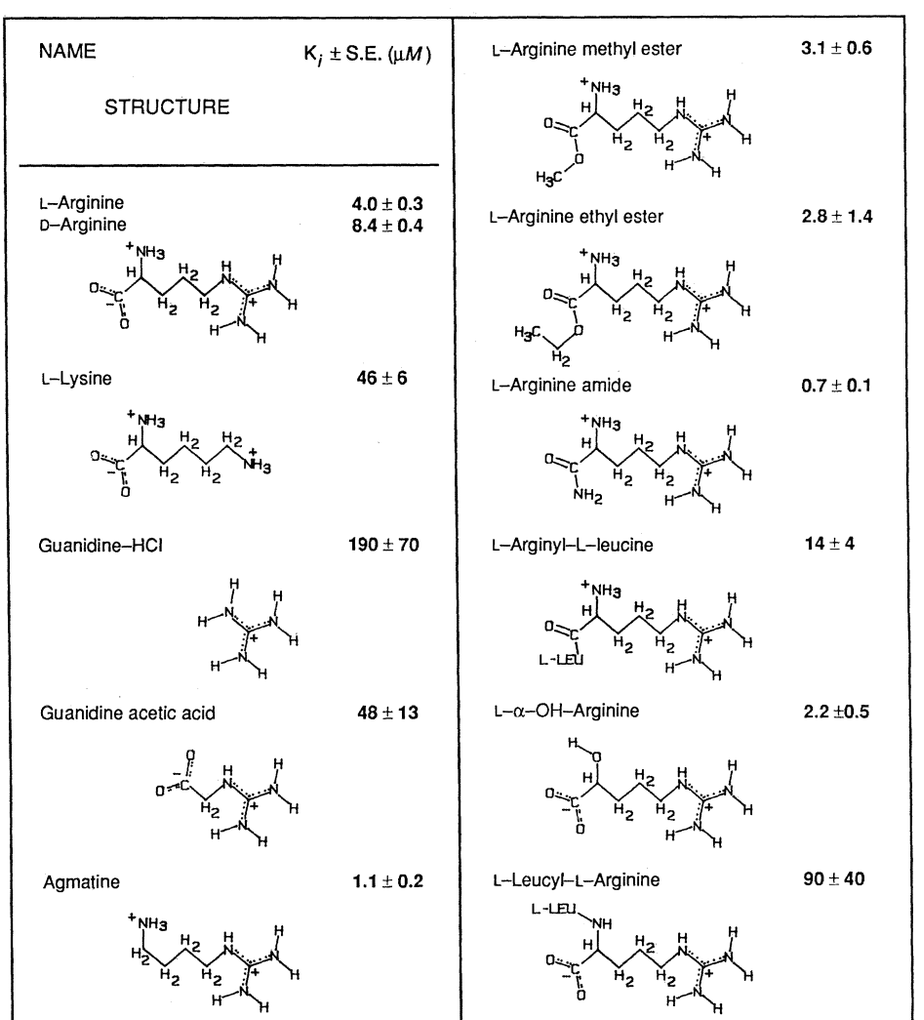


Fig. 6. K_i's measured by depression (*4*) (as in Fig. 3B) of the initial rate of IVS−3′ exon formation at 37°C in the presence of L-arginine and some of its analogues. Extent of inhibition was measured for 6 to 12 concentrations of inhibitor in the range 0 to 25 m*M*, at 10 m*M* Mg^{2+} (Figs. 2 and 3C). [α-^{32}P]GTP was present at 0.13 μ*M*.

hindered, as observed. Thus an asymmetric environment and consequent stereoselection are rationalized by the binding model.

The model suggests that an amino acid derivative might be designed to participate in splicing. The success of such a reaction would verify the relative disposition of the two molecules postulated in Fig. 7, and bring the self-acylating RNA posited in the introduction a step closer. Finally, this hypothesis implies that L-arginine is almost centered in the G site. If the L-arginine site is located, the catalytic G site will simultaneously be defined.

Implications of this work for further amino acid binding sites. The groups attached to the α carbon in free amino acids are asymmetric sterically and electronically, and have a large electric dipole due to their zwitterionic character. Given the conjecture that other amino acid side chains will bind to an RNA pocket, and therefore that these asymmetric groups will approach the negatively charged RNA, binding will probably be stereoselective. In fact, that the environment of the alpha carbon would appear isotropic requires special assumptions. Therefore, given the preference of all amino acids for one of three symmetrical α,β carbon rotamers (*17*), the stereoselectivity of this arginine site is not surprising. The significant finding is instead that stereoselection favors the L-amino acid. Further RNA amino acid sites will probably be as, or more, stereospecific than this one.

Arginine has a side chain that (i) is large and extended, with a varied surface, (ii) is charged, (iii) has a strong planar hydrogen-bonding pattern that resembles a base, and (iv) a pi-electron system suited for interactions based on polarizability (stacking). All these properties seem well suited for occupation of a pocket made of nucleotides. Lysine is similar, but lacks the pi system and multiple hydrogen bond pattern. The K_i for lysine is 46 m*M*, an order of magnitude less tight than the binding of arginine (Fig. 6 and Table 1).

Other side chains seem similar in characteristics that may be important for RNA binding. Aspartate and, particularly, glutamate meet all the criteria in the paragraph above. The relatively large, charged sided chain has an aliphatic extension to keep the negative carboxyl at a distance from the specificity-determining groups. The planar carboxyl group is a pi system that can form strong H bonds. Aspartate or glutamate might, for example, pair as C does, accepting two hydrogen bonds from the H-1 and 2-amino of G, as has been suggested (*18*). A general potential for interaction with nucleic acids is supported by the occurrence of glutamic acid as the second amino acid (with arginine) in the Eco RI nucleotide recognition site (*11*).

However, for the strongest interaction between RNA and free aspartate or glutamate, compensation might be needed for the electrostatic work of bringing a negatively charged side chain into proximity with RNA. For example, the cost of the negative carboxyl charge on arginine is a fourfold increase in the dissociation constant (+0.8 kcal/mole; agmatine compared to arginine; Fig. 6). An evident way to compensate this effect is with a favorable ionic interaction with the α-amino group. Such an ionic bond would be a specific example of the general tendency to stereoselection, pointed out in the first paragraph of this section.

Finally, there is a hint (Fig. 1C) of a substantially different kind of amino acid binding motif. L-Threonine reproducibly stimulates IVS–3′ exon formation. This effect is smaller, and more difficult to study quantitatively, than inhibition by L-arginine. But it will evidently be of interest to extend studies of the effects of amino acids on catalysis to other amino acids, and other RNA's.

Implications for the origin of the genetic code. The existence of the arginine site bears directly on disagreement concerning the origin of the code. Crick (*19*) and adherents suggest that the code is a "frozen accident." That is, were another code to originate elsewhere, it could make different connections between anticodons and amino acids. The opposing argument was made by Woese (*20*), who reasoned that a specific physicochemical interaction must have linked anticodons and amino acids. However, early searches for such nucleic acid–amino acid interactions found little specificity in the weak complexes formed with mono-, di-, and trinucleotides (*21*).

The arginine data suggest that these two points of view might be plausibly combined. Arginine's structure is complementary to several structural features of nucleotides (for example, Fig. 4A). Thus it is likely a priori that a specific nucleotide site might be built for arginine, and it is probable that more than this one will ultimately be identified.

However, a specific site composed of ribonucleotides may require a substantial oligonucleotide that folds stably to produce a particular pocket. The arginine site in the *Tetrahymena* intron is supported by the secondary and tertiary structure of a relatively large RNA. Thus, while a specific physicochemical interaction between RNA and at least one amino acid exists, the nucleotides responsible for binding a particular amino acid side chain in a large primordial RNA can be distinct from the nucleotides which ultimately act as the anticodon. There might therefore be both a highly specific physicochemical basis and an element of chance behind existing coding assignments. This subject may be further clarified when the nucleotide sequences at the *Tetrahymena* arginine binding site can be defined.

In another respect, it is tantalizing that this first RNA amino acid site prefers the L-amino acid, potentially explaining the chiral preference for L-amino acids in modern proteins. The sort of definite but partial selectivity observed for the *Tetrahymena* RNA arginine site might have been extended during evolution to a strong selectivity by successive, easily conceived, structural alterations. However,

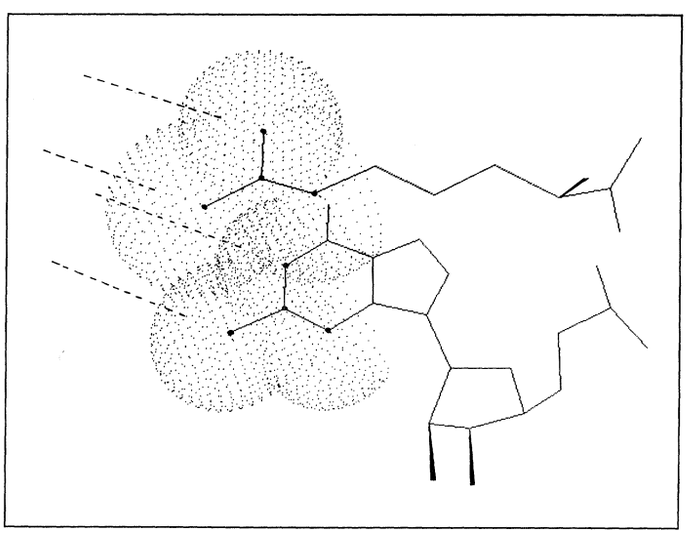


Fig. 7. A model for the relative position of L-arginine in the G site; L-arginine and G can bind alternatively in similar fashions. The picture is a perspective drawing, made by MIDAS molecular graphics software (release of May 1986 from the Computer Graphics Laboratory of UCSF) running on a Silicon Graphics IRIS 2400 computer. The drawing was transmitted from the screen to a Seiko CH-5301 color hard-copy plotter. The L-arginine side chain has been put above and parallel to the plane of the guanine of a GMP molecule. The view is from above, with the α-N atom of arginine projecting toward the viewer at the upper right, and the ribose hydroxyls being downward and slightly outward in the direction of the viewer at the bottom. The forked arginine carboxyl, correspondingly, projects back and slightly downward. The Van der Waal's surface of the guanidino atoms in guanine and L-arginine have been dotted, and the centers of the same atoms have been marked with small black dots. In order to help demonstrate the similarity of the molecules, the dashed lines extending from the two residues toward the upper left suggest analogous hydrogen bonds (from invisible hydrogens) which may be formed by both residues. Structural formulas for G and arginine may be compared in Fig. 4A.

an RNA site which selects D-arginine instead of L-arginine is also easily conceivable. This issue can perhaps be settled if the asymmetric character of the arginine site is a consequence of a frequent and stable feature of RNA structure. Then RNA formed of D-ribose would have a natural preference for L-amino acids. The stereochemical preferences of subsequent RNA amino acid sites will be of great interest for these same reasons.

Potential for regulation of RNA function. The *Tetrahymena* intron functions in a pathway for the processing of ribosomal RNA (*22*), an essential translational component. Ribosome synthesis may be regulated by other translational components (free amino acid) or by negative control proteins analogous to the positively acting protein splicing factor (*23*). The observed dissociation constant in vitro (1.8 m*M* at 5 m*M* Mg^{2+}) (Table 1), suggests that a sensitivity to free amino acids in vivo is possible. But efficient competition by G (Fig. 2 and Table 1) would make arginine ineffective unless it exists as a part of a more strongly binding regulatory peptide or protein. However, even if not in this system, other RNA's will probably exploit this potential for specific, direct regulatory interactions with small molecules like GTP and arginine.

Implications for RNA catalysis. RNA can express a specific affinity for free amino acids. Given binding free energy for amino acids, it is likely RNA sites can exist that distort an amino acid toward its transition state. There is therefore a robust prediction (*24*) that RNA catalysis of some order will extend to amino acid or peptide substrates, in addition to known ribonucleotide substrates. Weiner and Maizels (*25*) have expressed a theory of molecular evolution that includes a similar expectation. Such RNA enzymic activities may yet include one that can plausibly model a primordial aminoacyl-RNA synthetase, justifying the premise of this investigation.

REFERENCES AND NOTES

1. P. R. Schimmel, and D. Soll, *Annu. Rev. Biochem.* **48**, 601 (1979). M. Yarus, *Prog. Nucleic Acid Res. Mol. Biol.* **23**, 195 (1979).
2. K. Kruger *et al.*, *Cell* **31**, 147 (1982).
3. A. J. Zaug, P. J. Grabowski, T. R. Cech, *Nature* **301**, 578 (1983).
4. The reaction mixture used consisted of: 50 m*M* Hepes, 30 m*M* NaCl, 10 m*M* $MgCl_2$, and 0.2 m*M* EDTA, *p*H 8.0, with variations as described. Incubations were usually 5 or 20 minutes at 37°C. Unlabeled precursor transcript was 0.02 to 0.33 μM chains (all A_{260} was attributed to active molecules), and [α-^{32}P]GTP was 0.078 to 2.5 μM, with variations in concentration, label, and form as indicated. Abbreviations for the splicing reactants are as follows: IVS: intervening sequence or intron; L-IVS: linear IVS; C-IVS: circular IVS; IVS–3′ exon: the two-thirds molecule that results from the reaction with G; G: generic term for the small splicing substrate, meaning guanosine, GMP, or GTP.
5. M. Yarus, unpublished data.
6. B. L. Bass and T. R. Cech, *Biochemistry* **25**, 4473 (1987).
7. $f^{-1} = 1 + \{K_m/(K_m + S)\}\{I/K_i\}$, where f is the fraction of the rate of reaction of an uninhibited control, K_m is the Michaelis constant for the substrate, S is substrate concentration, K_i is the dissociation constant of the inhibitor, and I is inhibitor concentration. In these experiments, $\{K_m/(K_m + S)\}$ is about 1.0.
8. A. T. Profy and D. A. Usher, *J. Mol. Evol.* **20**, 147 (1984).
9. A. L. Weber, *ibid.* **25**, 7 (1987).
10. B. L. Bass and T. R. Cech, *Nature* **308**, 820 (1984).
11. Compare Arg^{200} and Glu^{144} in the Eco RI–DNA complex [J. A. McClarin *et al.*, *Science* **234**, 1526 (1986)].
12. T. Inoue, F. X. Sullivan, T. R. Cech, *J. Mol. Biol.* **189**, 143 (1986).
13. G. Lancelot, R. Mayer, C. Helene, *Biochim. Biophys. Acta* **564**, 181 (1979).
14. C. Saxinger and C. Ponnamperuma, *Origins Life* **5**, 189 (1974).
15. V. J. Bruskov, *Stud. Biophys.* **67**, 43 (1978).
16. N. K. Tanner and T. R. Cech, *Biochemistry* **26**, 3330 (1987).
17. J. W. Ponder and F. M. Richards, *J. Mol. Biol.* **193**, 775 (1987).
18. C. Helene and J.-C. Maurizot, *CRC Critical Rev. Biochem.* **10**, 213 (1981).
19. F. H. C. Crick, *J. Mol. Biol.* **38**, 367 (1968).
20. C. R. Woese *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **31**, 723 (1966).
21. J. C. Lacey, Jr., and D. W. Mullins, Jr., *Origins Life* **13**, 1 (1983).
22. N. Din, J. Engberg, W. Kaffenberger, W. Eckert, *Cell* **18**, 525 (1979). T. R. Cech and D. C. Rio, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5051 (1979).
23. R. A. Atkins and A. Lambowitz, *Cell* **50**, 331 (1987).
24. L. Pauling, adapting his Silliman Lecture at Yale, *Am. Sci.* **36**, 51 (1946).
25. A. Weiner and N. Maizels, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7383 (1987).
26. The plasmid transcribed was Eco RI–digested pTT1A3-T7, which is pTT1A3 with a T7 promoter [M. D. Been and T. R. Cech, *Nucleic Acids Res.* **13**, 8389 (1985)].
27. TBE is 0.1*M* tris, 0.1*M* boric acid, 2 m*M* EDTA, *p*H 8.3. Gel electrophoresis was conducted in TBE plus 8*M* urea, in slabs (0.75 by 165 mm) of 4 percent acrylamide (20:1 bis), at 250 V for 70 or 120 minutes, depending on the purpose of the resolution. Gels were dried onto DEAE paper to retain the 15-nt radioactivity. Dried gels were autoradiographed, then sliced for Cerenkov scintillation counting, with the use of film as a template.
28. G. N. Wilkinson, *Biochem. J.* **80**, 324 (1961).

29. I thank T. Cech for expert encouragement; A. Zaug for gifts of materials; J. McSwiggen for helpful markers; A. Braque, D. Deamer, H. Morowitz, D. Usher, and A. Weber for introducing me to chemical evolution; D. Smith for his expertise with MIDAS, which was essential to Fig. 7; B. Bennett for performing the experiment of Fig. 4; J. Curran, L. Folley, L. Gold, D. Prescott, D. Smith, and G. Stormo for helpful comments on a draft manuscript. Supported by NIH research grant R37 GM30881.

12 January 1988; accepted 27 April 1988