Three Small Ribooligonucleotides with Specific Arginine Sites†

Gregory J. Connell, Mali Illangesekare, and Michael Yarus*

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

Received February 16, 1993; Revised Manuscript Received March 30, 1993

ABSTRACT: Arginine-binding RNA motifs are important to protein–RNA interaction and perhaps also for Archean biochemistry. Selection–amplification was used to isolate three RNAs that are eluted by free arginine from an L-arginine affinity column ($K_d \approx 0.2$–0.4 mM). The binding sites contain specific internal and bulge loops, whose sequences can include arginine coding triplets. Binding is highly specific for arginine, but all three motifs, like the self-splicing group I intron, also bind guanosine 5′-monophosphate. One site is stereoselective, somewhat preferring D-arginine.

Two arginine-binding RNA motifs have been previously studied. An arginine site within the self-splicing *Tetrahymena* pre-rRNA was identified by competition with the guanosine cosubstrate (Yarus, 1988, 1989). This binding is stereospecific; free L-arginine has a K_d of 0.4 mM and can be preferred up to 10-fold over D-arginine (Yarus & Majerfeld, 1992). The arginine site overlaps the guanosine binding site in the P7 helix (Michel et al., 1989) and appears within one of four arginine coding triplets (Yarus & Christian, 1989), conserved in all modern group I RNAs. The conserved arginine triplets could have evolutionary significance (Yarus, 1991). The second arginine site appears in TAR, a hairpin at the 5′ end of the HIV (human immunodeficiency virus) mRNAs (Weeks et al., 1990; Calnan et al., 1991). The HIV transactivator protein Tat contacts TAR via a crucial arginine residue. Free arginine also binds TAR with a K_d of 4 mM (Tao & Frankel, 1992), and the conformation of the TAR RNA–arginine complex has recently been characterized using NMR spectroscopy (Puglisi et al., 1992). Gag proteins, antiterminators, ribosomal proteins, and HIV REV contain arginine-rich regions that have been proposed to interact with RNA (Lazinski et al., 1989). To identify other RNA motifs capable of binding arginine, we have used arginine affinity chromatography to select three additional arginine-binding motifs from random ribooligonucleotide pools.

Selection–amplification has been previously used to isolate RNA molecules that bind to specific proteins (Tuerk & Gold, 1990) and to dye affinity columns (Ellington & Szostak, 1990), starting from pools of random sequences. Our pool consisted of 72-mer RNA transcripts containing 25 central randomized positions, encoded within flanking PCR primer sequences containing a T_7 RNA polymerase promoter (Milligan & Uhlenbeck, 1989). The length of the randomized region was chosen to be large enough to form stable TAR-like or *Tetrahymena* P7-helix-like secondary structures as predicted by a RNA folding algorithm (Zuker, 1989; Freier et al., 1986). This small size for the randomized region makes the identification of short consensus sequence elements obvious, and it also explores the minimal length of RNA required for binding to small ligands.

Although the α-amino group of arginine has not been shown to be required for the binding to TAR, it contributes to binding free energy in the group I intron site (Yarus, 1988). Thus the arginine affinity ligand in these selections had an unhindered α-amino group. We amplified pooled RNA eluted by free L-arginine from the affinity column after the bulk RNA had been eluted by buffer. This protocol requires that free arginine compete with the affinity ligand, making it likely that selected RNAs bind both free amino acid and its immobilized derivative and at the same site.

MATERIALS AND METHODS

The Selection Procedure. An L-arginyl-L-cysteine dipeptide was synthesized and linked through the sulfhydryl group to a Thiopropyl Sepharose 6B (Pharmacia) column matrix. This column will be referred to as the "arginine column" for compactness. Two parallel selections were performed on two different 100-μL arginine columns. Column 1 contained 19.6 mM L-arginyl-L-cysteine and was run in column buffer containing 375 mM NaCl, and column 2 contained a 14.7 mM amount of the attached ligand and was run in column buffer containing 325 mM NaCl. The column buffer for both selections also contained 5 mM $MgCl_2$, 1 mM EDTA, and 40 mM Tris-HCl (pH 7.5 at 4 °C).

Approximately 20 μg of random ^{32}P-labeled RNA (≈5 × 10^{14} molecules; transcribed from 5 × 10^{13} DNA templates) in water was heated at 65 °C for 5 min, the salt concentration was adjusted to the appropriate level, and the RNA was cooled to 4 °C over 10 min before the RNA solution in a 25-μL total volume was loaded onto the affinity column. The columns were washed at 4 °C for approximately 8–12 column volumes, and any remaining RNA sticking to the column was eluted with 50 mM L-arginine in column buffer. The pooling of the eluted RNA from the two columns for the initial cycle of selection was less stringent than some of the later cycles so as to minimize the loss of poorly represented transcripts from the RNA populations (Table I). cDNA was synthesized from the pooled arginine-eluted RNA, amplified by PCR, and transcribed into RNA for the next cycle of amplification as previously described (Tuerk & Gold, 1990). After cycle 3, each passage of the RNA through the arginine column was preceded by a passage through an L-cysteine–Thiopropyl Sepharose column, and only the initial 70–80% of the RNA at the void volume was used for subsequent selection on the arginine columns. This preselection minimized the selection of RNAs with affinity for components of the arginine column other than arginine.

Minimal Sequence Requirements for Arginine Binding. The RNA motifs were labeled at either the 3′ or 5′ end with ^{32}P and then partially digested under mild alkaline conditions

† Directly supported by NIH Research Grant GM30881 to M.Y. and by a MRC (Canada) postdoctoral fellowship to G.J.C. We thank the W. M. Keck Foundation for support of RNA science in Boulder.

* To whom correspondence should be addressed.

0006-2960/93/0432-5497$04.00/0

Table I: Percentage of RNA Pooled from the L-Arginine Affinity Columns Relative to the RNA Applied for the Two Separate Selections

cycle no.	column 1 (%)	column 2 (%)
1	3.0	1.8
2	0.9	0.4
3	0.5	0.8
4	5.0	7.5
5	6.8	28.0
6	20.0	25.0
7	60.0	46.0

(Pan & Uhlenbeck, 1992). The digestion products were loaded onto a 100-μL arginine column at room temperature and subsequently washed with 1200 μL of column buffer before being eluted with 50 mM L-arginine in the column buffer. Aliquots of the collected fractions were loaded onto a 10% polyacrylamide gel for autoradiography.

K_d Determinations. The K_d values for the binding of RNA to the arginine columns were determined from $K_d = L_c\{V_n/(V_e - V_n)\}$, where L_c is the ligand concentration on the column, V_e is the median elution volume measured in the absence of free ligand within the column buffer, and V_n is the volume at which an RNA population having no interaction with the column would elute. The elution profiles for homogeneous RNA populations were often broad, possibly because of the differential accessibility of column affinity sites to a large molecule like RNA. Therefore, the median RNA position was used as the value for V_e in these calculations. The value for V_n is dependent on the degree to which the RNA population is excluded from the column, and for the K_d values quoted here, V_n was taken to be 70% of the 100-μL column volume. Uncertainty in V_n may introduce a factor-of-2 uncertainty into the calculations.

The K_d values for the binding of the RNA to arginine in solution were determined from $K_d = L\{(V_{eL} - V_n)/(V_e - V_{eL})\}$, where L is the free ligand concentration used to isocratically elute RNA loaded onto the affinity column, V_{eL} is the median elution volume of RNA both loaded onto the column and eluted in the continuous presence of free ligand, and V_e and V_n are as defined above.

RESULTS

Three Arginine Binding Motifs. RNA from cycle 7 of selection 1 and cycle 6 of selection 2 (Table I) was reverse transcribed, amplified, and cloned in pUC19 for dideoxynucleotide sequencing. Of 60 sequenced clones, 55% were variants of motif 1 (Figure 2); another 5% were altered in sequence but showed a related secondary structure. Motif 2 was defined by 8% of the clones while 12% of the clones defined motif 3. All three motifs were isolated in both selections, done under slightly different ionic conditions (Figure 1). The other 20% of clones were unrelated in sequence and predicted secondary structure to each other and the three major motifs. None of the 60 sequenced clones were identical at all initially randomized positions; thus even when assigned to one of the major motifs, every isolate represents a different starting oligomer.

A representive of each of the three major motifs was further characterized. The mean K_d for the arginine column was determined from analytical affinity chromatography to be approximately 0.2 mM for motif 1, 0.3 mM for motif 2, and 0.4 mM for motif 3. Though these RNAs were counterselected for affinity to the column material, these K_d's could include nonspecific interactions with the column and are also highly dependent on the ionic strength of the column buffer. The

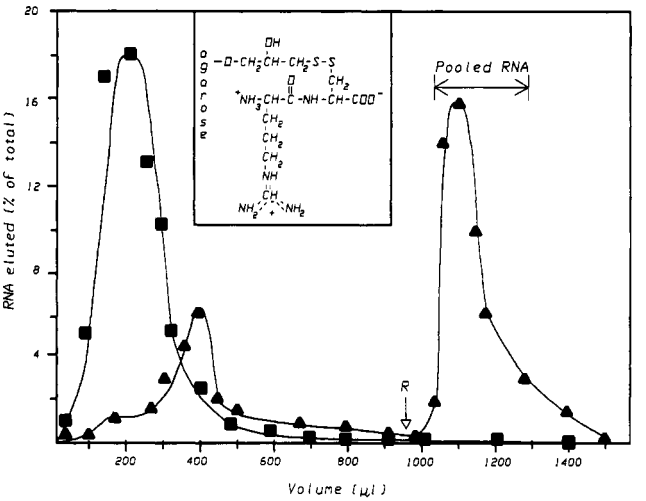

FIGURE 1: The L-arginyl-L-cysteine–Thiopropyl Sepharose 6B column matrix (inset) and the elution profiles for cycles 3 (■) and 7 (▲) from one of the selections for arginine binding motifs are shown. After the columns were washed with approximately 8–12 column volumes, any remaining RNA sticking to the column was eluted with 50 mM L-arginine in column buffer (R). Nonspecific RNAs washed through in the initial few column volumes. The peak at 400 μL of eluant in cycle 7 could be weakly binding motifs or weakly binding alternative conformations.

MOTIF #1

	30					35					40				44
	C	C	G	U	G	C	A	C	U	C	C	U	C	G	A
occurrence			▲	▲											
G	3	-	31	-	31	-	-	-	-	-	-	-	-	32	5
A	4	1	1	-	2	-	33	-	-	-	-	-	-	1	22
U	3	5	1	32	-	2	-	-	33	-	-	33	4	-	4
C	23	27	-	1	-	31	-	33	-	33	33	-	29	-	2

MOTIF #2

	A	U	C	C	G	U	G	C	A	G	G	C	U	C	A	G	A	C	G
occurrence												▲							
G	-	-	-	-	5	-	5	-	-	5	4	-	-	-	2	5	-	1	3
A	5	-	-	-	-	-	-	-	5	-	1	-	-	-	3	-	3	-	1
U	-	4	-	-	-	5	-	-	-	-	-	-	5	-	-	-	2	-	-
C	-	1	5	5	-	-	-	5	-	-	-	5	-	5	-	-	-	4	1

MOTIF #3

	C	C	G	U	G	U	A	A	C	C	C	A	C	U	C	G	A
occurrence				▲													
G	-	-	7	-	7	-	-	-	-	-	-	-	-	-	-	7	-
A	1	-	-	-	-	-	4	5	-	-	-	7	-	-	-	-	5
U	-	-	-	7	-	7	-	-	-	-	-	-	-	7	1	-	1
C	6	7	-	-	-	-	3	2	7	7	7	-	7	-	6	-	1

FIGURE 2: The consensus sequence (bold) for each of the three selected arginine-binding motifs and the observed number and type of variation at each position are indicated. Within each of the motifs, there were clones that had insertions within the consensus, and the location of these insertions are shown (▲). The criterion used for the inclusion of a position within the consensus of motif 1 was that the χ^2 test gave a probability for nonrandomness at that position that exceeded 99%. A position was included within the consensus of motifs 2 and 3 if at that position greater than 50% of the members of the motif had the same identity. Numbering above motif 1 corresponds to that used in Figure 3 for a representative member of this motif.

conditions of these selections, however, were chosen such that our initial pool of random oligomers had a mean K_d of 11 mM (Figure 1), and a 25-nucleotide TAR-like element substituted in place of the random region of our oligomer had a mean K_d of 7 mM.

No clones containing the TAR motif were selected despite this motif's simplicity and, therefore, despite its probable representation in the initial pool. Selected RNAs were eluted with arginine after the majority of the molecules having a K_d near that of TAR would have eluted (Figure 1). TAR, therefore, has too high a dissociation constant to have been preferentially enriched. All three of the selected motifs interact with the arginine α-amino (see below), purposely left free in our column; this may have put TAR-like motifs at a

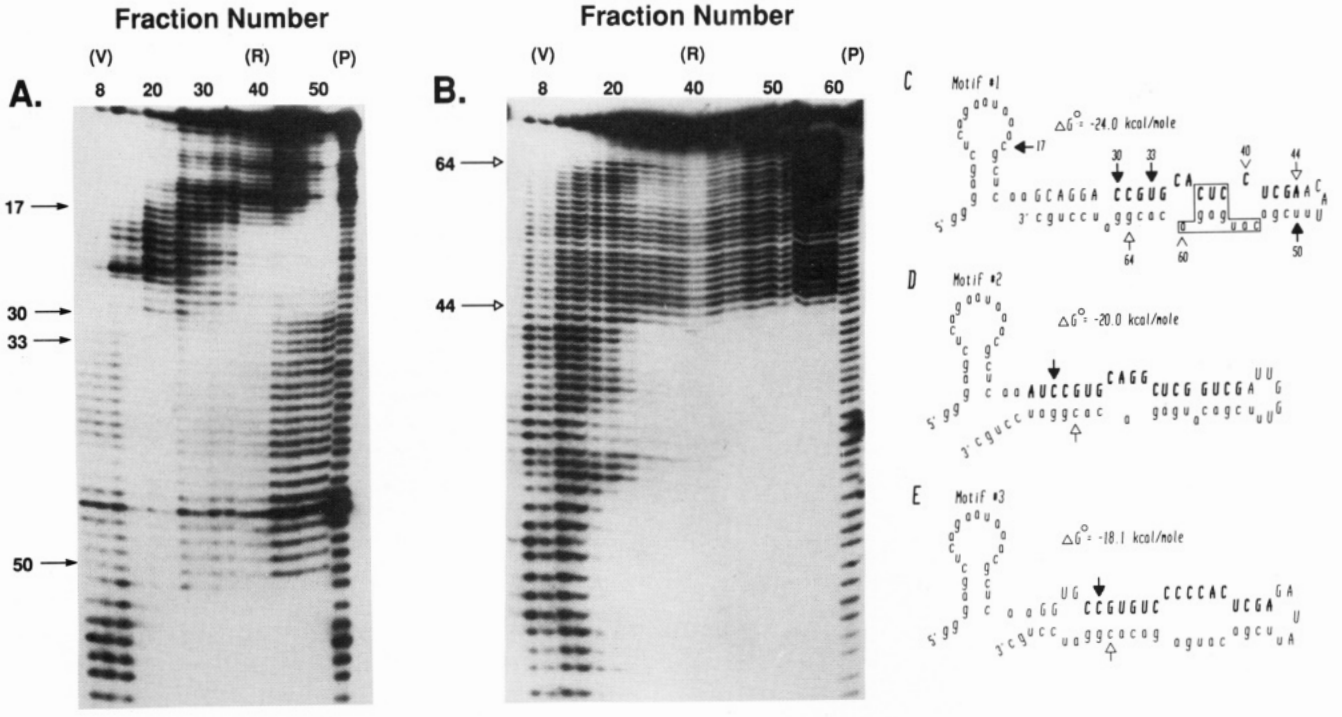

FIGURE 3: Minimal sequence requirement for the binding of motif 1 to the arginine column. RNA was labeled at either its 3′ end (A) or 5′ end (B) with ^{32}P and then partially digested under mild alkaline conditions. The digestion products were loaded onto an arginine column at room temperature and subsequently washed with 1200 μL of column buffer before being eluted with 50 mM L-arginine (R) in column buffer. Fractions collected from the column contained about 30 μL, and aliquots were loaded onto a 10% polyacrylamide gel for autoradiography. An aliquot of the precolumn hydrolysis ladder (P) was also loaded onto the gel, and the approximate column fractions that would be expected to contain the nonbinding RNA are also indicated (V). The positions of the boundaries from the 3′-^{32}P-labeled (closed arrows) and 5′-^{32}P-labeled (open arrows) RNA are marked. The energy-minimized folding of a representative clone from motif 1 (C), motif 2 (D), and motif 3 (E) and the calculated standard free energy of formation ($\Delta G°$) for their most stable conformations are indicated. Capital letters refer to the 25 positions of the RNA that were originally randomized and small letters to the flanking PCR sequences. The consensus sequences are in bold. Boundaries obtained from the 3′-^{32}P-labeled (closed arrows) and the 5′-^{32}P-labeled (open arrows) RNA are indicated. The region of motif 1 that was randomized for subsequent reselection is boxed.

disadvantage. Neither was a clear P7-like motif recovered; perhaps the *Tetrahymena* site requires support from a tertiary structure.

Energy-minimized foldings (Figure 3) suggest that a part of all three conserved sequences was selected to base pair with the 3′ flanking PCR primer sequence. However, the PCR primer sequence probably makes a sequence-specific contribution. Motif 1, for example, always contains a bulged CA three base pairs away from a bulged C, all selected from originally random nucleotides (Figure 3C). These sequence elements are always located over a bulged A adjacent to a bulged 5′-CAU-3′ in the primer sequence. Conserved bulged sequences within the initially random region might have occurred over different primer elements, but this did not occur in 33 independent examples. Thus this combination of loop sequences is selected, though composed of both initially random and initially fixed elements.

Minimal Sequence Requirements. Minimal structures for arginine column binding, and for elution by free arginine, were determined by random degradation and reselection. Ladders of 5′ and 3′-^{32}P-labeled RNA, created by partial alkaline hydrolysis, were examined in eluants from the arginine affinity column. Boundaries of the functional 5′- or 3′-end-containing fragments are indicated in Figure 3.

3′-^{32}P-labeled motif 1 RNA cut 5′ of position 33 flows through in the same early column volumes at which random RNA elutes (Figure 3A). This agrees with the selected phylogeny; the helix containing position 33 is conserved. However, a significant fraction of the [3′-^{32}P-labeled motif 1 RNA that was cut between position 30 and 50 binds strongly to the column. This result superficially conflicts with the experimental phylogeny and, in particular, suggests that the 3′ PCR primer sequence (position 49, 3′ end) binds to the column. Because it is implausible that the RNA form of the primer sequence specifically binds arginine, a new transcript was made encoding only positions 49–72 of motif 1. This shorter RNA did not bind to the arginine column, nor did it bind when denatured and allowed to anneal with full length RNA prior to chromatography. Significantly, the labeled 49–72 3′ transcript did bind to the column when mixed with partially alkaline hydrolyzed motif 1 RNA. This reconstruction suggests that RNA containing only positions 49–72 of motif 1 forms an active base-paired complex with complementary nicked molecules in the hydrolysis mixture. The boundary at position 50, therefore, probably represents the shortest molecule capable of forming a base-paired complex. Similarly, the boundary at position 30 may represent the point at which the stability of the 3′ fragment's self-structure prevents complex formation with the 5′ fragment. The complex between nicked fragments is, therefore, consistent with both the phylogenetic and biochemical data.

3′-^{32}P-labeled motif 1 RNA nicked 5′ of position 17 does not bind as well to the arginine column, though the phylogenetically conserved motif is present (Figure 3A). In reconstruction experiments, an RNA transcribed from an oligo encoding only positions 30–65 of motif 1 binds to the arginine column, but with a higher K_d than that of the full-length RNA. Better binding of RNAs that contain the first 17 5′-nucleotides can probably be attributed to electrostatic interactions. The interaction of arginine with the RNA, though

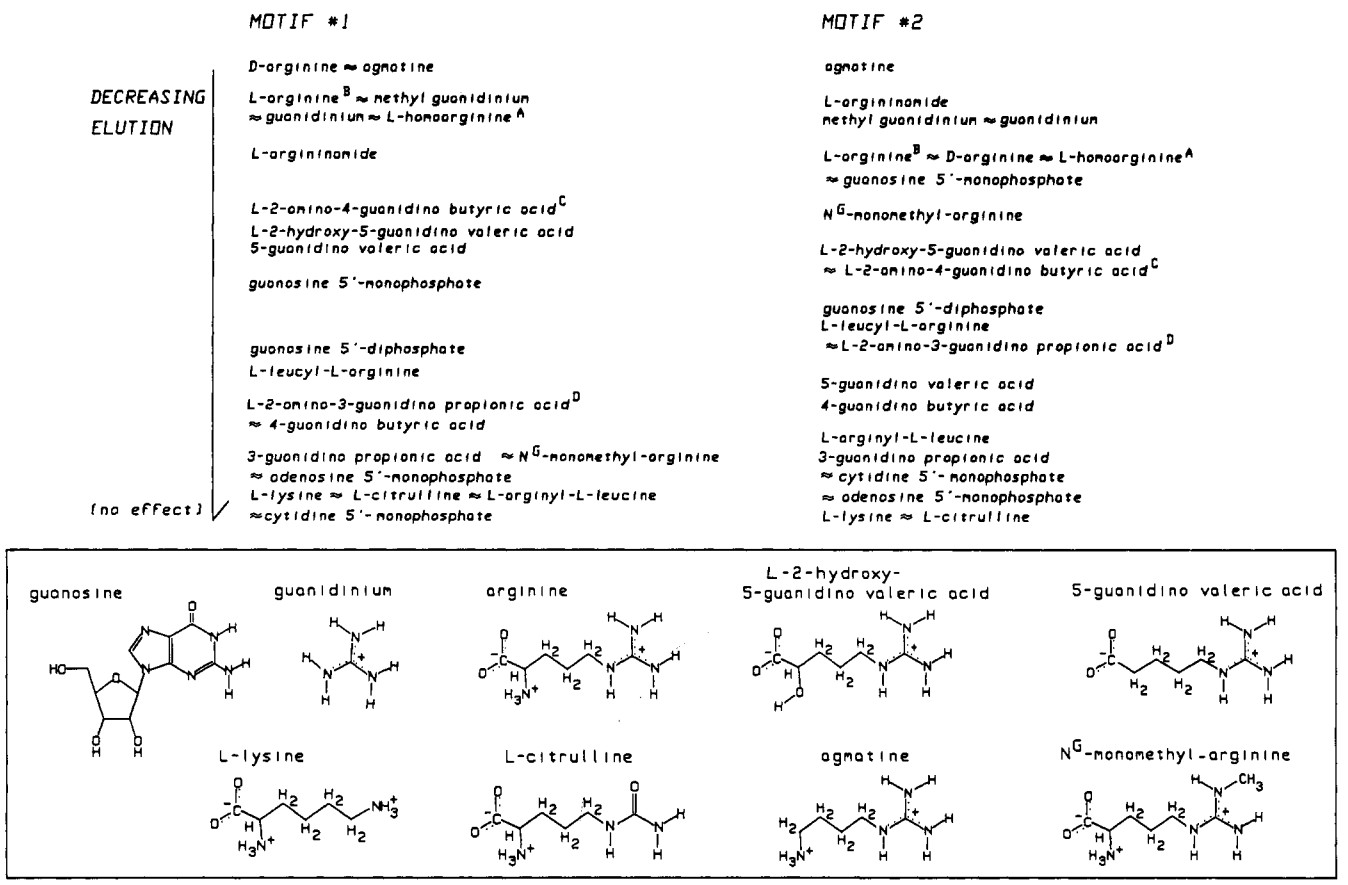

FIGURE 4: The relative ability of arginine and several different analogues of arginine to elute motif 1 and motif 2 RNA from the arginine column is shown. Since the point of elution is determined by the ratio of the concentration of the eluant to its K_d, at the same concentration the more effective eluant also has the lower K_d. Approximately 2 μg of ^{32}P internally labeled RNA was loaded in a 25-μL volume onto the arginine column (see Figure 1). The column was washed with 2 column volumes before being eluted with a series of analogues. The relative effectiveness of the different eluants was compared for about 20 column runs to derive each relative elution series. Analogues, such as citrulline, that do not elute the motifs at all were used at a concentration of 20 mM in column buffer. Analogues that eluted the RNA from the column as well or better than L-arginine were used at a concentration of 3 mM, and all other eluants were used at a concentration of 5 mM. Equimolar Mg^{2+} was added to the GDP solution to compensate for chelation. Capital superscripts highlight the effect of decreasing the size of the aliphatic chain connecting the α-amino to the guanidino group from five (A) to four (B), three (C), and two carbons (D). The structures of arginine and some of the related molecules are also included below. Although the behavior of motif 3 is distinguishable from motif 1, it shares many characteristics, and for that reason its elution series is not shown.

specific, has an electrostatic component (see below); adding nonspecific sequences should strengthen binding.

5'-^{32}P-labeled motif 1 RNA that was nicked 5' of position 64 flows through in the early column volumes, as do nonspecific structures (Figure 3B). This again agrees with experimental phylogeny; the helix at the left of the bulged CA is conserved. RNA, however, nicked between the 3' end and position 44 also binds to the column. We suggest that this result complements the boundary obtained by digesting the 3'-labeled motif (above) and identifies the other partner in functional nicked complexes. Together, nicked complexes argue against a functional role for the right-hand terminal base pair and hairpin loop (Figure 3C) of motif 1, a result also consistent with phylogeny (Figure 2).

The other two motifs, like motif 1, show similar boundaries affected by either complex formation or nonspecific electrostatic interactions. For the sake of clarity only the outer boundaries thought relevant to intact molecules are indicated on the structures (Figure 3D,E).

Specificity of Arginine Binding. All three motifs have specific binding sites and distinguishable specificities. Relative elution by free analogues of arginine was determined, as a way of ordering their dissociation constants (Figure 4). L-Lysine has both an α-amino and a positively charged side chain the same distance from the α-carbon as arginine, and L-citrulline similarly positions both the α-amino group and a side chain capable of hydrogen bonding somewhat like arginine, yet neither elutes these RNAs.

At least two specificities for the arginine side-chain guanidinium appear. N^G-monomethylarginine (Figure 4), elutes all three motifs less efficiently than arginine. However, elution of motif 2 is less affected by methylation than are the other two motifs. Motif 2 is presumably less restrictive in its interaction with the guanidinium.

Strikingly, all three arginine motifs have an unselected affinity for guanosine. All three are effectively eluted from the arginine column by guanosine 5'-monophosphate. Guanosine 5'-diphosphate does not elute the motifs as well as the monophosphate; evidently there is electrostatic or steric hindrance from the additional 5'-phosphate.

All three motifs gain binding free energy from the α-amino group of arginine. Whenever tested (Figure 4), the amino acid is a better eluant than the corresponding carboxylate. Since the improvement in binding occurs at all side-chain lengths, the effect can be attributed in part to nonspecific electrostatic interaction between the α-amino and the negatively charged RNA. However, L-2-hydroxy-5-guanidinovaleric acid (Figure 4) is a better eluant than 5-guanidinovaleric acid. This suggests a hydrogen bond to the α-amino (and α-hydroxyl), in addition to the charge interaction just

discussed. Decreased affinity of all three motifs for the dipeptide L-leucyl-L-arginine relative to arginine further supports the conclusion that the α-amino group closely approaches a part of the site. All three motifs have poor affinity for L-arginyl-L-leucine dipeptide, though the selection was performed on an L-arginyl-L-cysteine column. The larger, more hydrophobic leucine side chain may obstruct binding.

All three motifs may have a primary affinity for guanidinium; guanidinium or methylguanidinium is as effective an eluant as arginine. This suggests that the extra binding energy for the α-amino group compensates for an unfavorable interaction of the carboxyl group. All three motifs also have a preferred spacing between the carboxyl group on the α-carbon and the guanidino group. Shorter guanidino carboxylic acids do not bind as well, suggesting that the binding site may become more restricted closer to the guanidino group and/or the electrostatic repulsion between the carboxyl group and backbone may increase.

Nevertheless, these RNAs interact differently with the arginine α-carboxyl, giving stereoselectivity. The affinity of motif 2 for argininamide is substantially better than that for arginine. This may be explicable because argininamide more closely resembles the affinity ligand. In contrast, argininamide elutes motif 1 less well than L-arginine, possibly because of steric hindrance by the slightly larger amide. This suggests that the L-carboxyl is the most hindered in motif 1, consistent with the observation that L-arginyl-L-leucine is an ineffective eluant for motif 1 but elutes the other two, albeit at low efficiency. Further, agmatine (Figure 4) has significantly greater affinity than L-arginine for all three RNAs, but D-arginine and agmatine are equivalent for motif 1. Thus for motif 1 the L-carboxyl is hindered, but a D-carboxyl group has no detectable repulsive effect.

As one result of this dissimilarity, motif 1 is somewhat stereoselective. Specific interactions by three of the four groups attached to the α-carbon are necessary and sufficient for stereoselection to occur. Since the guanidino and α-amino interactions are significant for all three RNAs, the nonstereoselective motifs 2 and 3 may be unrestrictive to the negatively charged carboxyl. However, the K_d of motif 1 for L- and D-arginine in solution is 1.0 and 0.6 mM, respectively. Clearly, there is no strong chiral complementarity between D-ribose-containing RNA and L-amino acids, as we have now detected sites that prefer both L-arginine (10-fold; Yarus & Majerfeld, 1992) and D-arginine (1.7-fold; motif 1). K_d differences for L-arginine on the column (0.2 mM) and free in solution (1.0 mM) could be the result of a small interaction with the column.

Reselection Using Motif 1. All three motifs contain some fixed primer sequence. To determine the contribution of primer sequences, the boxed region of motif 1 (Figure 3C) was randomized. After four cycles of amplification and selection on the arginine column under conditions similar to those of the original selections (Figure 1), 23 clones from the remutagenized RNA pool were sequenced. K_d values vary for the reselected molecules but are within a factor of 2 of the parental motif 1.

A stem of three nucleotides linking two loops was reselected (Figure 5). Two base pairs connecting internal and bulge loops are present in 21 of the 23 sequenced clones, but these nucleotides are not conserved. These two base pairs conserve the bulged CA at positions 35 and 36. The third base pair between positions 39 and 57 was present in only 6 of 23 clones.

In contrast to this conserved stem structure, new loops are recovered on reselection, suggesting that some loop sequences had been forced by common primer sequences. Reselected RNAs do not conserve either side of the original right-hand loop: bulged C40 and a bulged 5′-CAU-3′ at positions 54–56 (Figure 3C). While no single consensus appears at these positions, preferred sequence combinations recur.

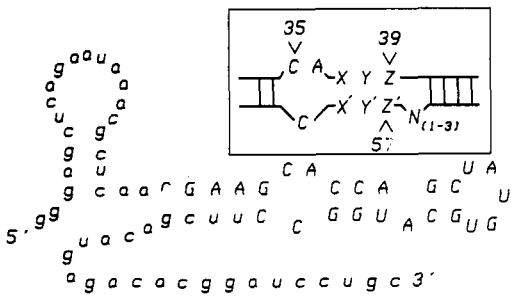

FIGURE 5: The consensus for reselection of randomized motif 1 is shown (upper box). Some of the positions corresponding to motif 1 are numbered (Figure 3C). The two base pairs that were found to be conserved in 21 of the 23 sequenced clones are indicated (X to X′ and Y to Y′). The third base pair (Z to Z′) was present in 6 of 23 sequenced clones. The size of the bulge at position 56 varied from one to three nucleotides in the reselected clones. Base pairs determined to be essential from motif 1 boundary experiments are also shown (vertical lines). One clone obtained from the original selections also contains the reselected consensus (shown below). Capital letters refer to the 25 positions of RNA that were initially randomized and lower-case letters to the flanking PCR sequences. Note that this small motif still cannot be fit entirely into the 25 nucleotide region that was randomized for the initial selection; four positions of the primer sequence are utilized by the structure.

The left-hand internal loop is well conserved. Bulged C35 A36 in the reselected pool is always over a bulged C60. U or G60 are not recovered, presumably because they would disrupt the conserved CA bulge by base pairing. A bulged A60 was the original consensus at this position, but was not reselected. The randomized RNA used for reselection was sequenced, and all four nucleotides were approximately equally represented at this position. A60's absence after reselection may mean that there are fewer restrictions on nucleotides at the other randomized positions with a bulged C60. Fewer sequence restrictions are synonymous with greater abundance in the initial RNA pool and, consequently, greater proportion among a selected population which is only required to exceed a threshold in binding.

DISCUSSION

Of 60 originally sequenced clones, one, initially thought to be unrelated, fits the consensus obtained from the reselection of motif 1 (Figure 5). This clone independently confirms the efficacy of the reselected double loop-and-stem structure, particularly supporting a functional bulged nucleotide 56. In addition, this single isolate of an "optimal" sequence reemphasizes the fact that less frequent structures, for example, those requiring a greater number of exactly specified nucleotides, would not have been detected among 60 clones even if their K_d's were significantly lower than motif 1. Assuming a Poisson sample of 5×10^{13} 72-mers with a 25-mer region containing all four nucleotides without bias, the probability is 0.99 that any 22-mer sequence was present in the starting population. We cannot, however, know that we have isolated the most tightly binding RNAs $\leq$22 nucleotides long; instead, our repeatedly isolated structures are the most frequent (least complex) structures that met our chromatographic criterion. Such structures dominate an arbitrary sample of cloned sequences.

Since most of the selected RNAs are predicted to fold into a structure that utilizes some of the flanking fixed PCR primer

sequence, the 25-nucleotide random region must be close to the lower size limit for an RNA molecule being able to bind to the arginine column within the selected K_d range. It was shown using a RNA folding algorithm (Zuker, 1989) prior to starting these selections that 25-nucleotide-long TAR derivatives could fold into correct and stable structures in the presence of the flanking primer sequences. Presumably, the structural framework required for contact with both the α-amino and guanidino groups of arginine is greater than 25 nucleotides; utilization of the fixed primer sequences makes possible larger structures.

On the basis of phylogeny and boundary determinations, motif 1 seems to require 29 nucleotides in total, only 14 of which must be specified (Figure 3C). A reselected form of motif 1 is even smaller (Figure 5). Nevertheless, motif 1 is both amino acid specific and somewhat stereoselective. Its K_d for arginine is similar to that for the G site of a group I intron (Yarus & Majerfeld, 1992), over an order of magnitude larger in size.

All three motifs rely on specific internal loops that could have related tertiary structures, but the amino acid sites in motifs 1 and 2 have distinct specificities and, therefore, functionally distinct structures (Figure 4). Such structural versatility in short ribooligomers has implications for both evolution and present-day protein–RNA interactions.

A selection for RNA motifs that bind to a tryptophan–agarose column has previously been reported (Famulok & Szostak, 1992). No K_d was reported for binding to tryptophan in solution nor was there any indication that tryptophan could elute this RNA from the tryptophan–agarose column. As a result, it is unclear as to what extent the agarose matrix is contributing to the reported binding constant. Experience has shown that RNAs are easily obtained that incorporate both the ligand and column matrix into the binding site, giving artificially high binding constants and specificity. Thus, such results should be interpreted cautiously.

The phylogenetically conserved region of motif 2 (Figure 3D) contains an arginine triplet in an internal loop, conserved as AGG/A in independent isolates. This sequence was not directly forced by interaction with the primer sequence. There are also four conserved arginine coding triplets within the arginine site of group I RNAs, which are known from mutational evidence to contact the amino acid (Yarus & Christian, 1989; Yarus, 1991). However, there are six arginine codons, and accordingly, they appear frequently by chance. Note also that three of the five (or four of six; see Figure 5) now-known arginine-binding RNAs do not have arginine codons at the binding site. We believe that this work supports RNA–amino acid interaction theories of the code's origin, as follows. There now appear to be many ways to fold a specific RNA site for arginine. Some of these sites contain arginine codons, and some of these directly contact the amino acid. Thus evolution may have chosen among RNA binding sites for some amino acids, and functional sequences within particular sites chosen may have become the modern codons [discussed in Yarus (1993)]. We are studying the conserved arginine codon of motif 2 to see if it contacts the amino acid.

All of these selected arginine-binding motifs, like the group I intron, also bind guanosine 5′-monophosphate. This supports the notion that, because of structural similarities between the two molecules, there will be an overlapping subset of binding motifs which can be utilized by evolution (Yarus, 1991). We have now observed both possible unselected specificities. In the *Tetrahymena* active site, a presumably unselected arginine affinity exists within a guanosine site. Here we find unselected guanosine 5′-monophosphate affinities within arginine sites. Evolution can easily provide both activities by selecting only one.

All five known arginine-binding motifs are, to a greater or lesser extent, guanidinium sites; thus there are at least five ways to fold a specific RNA site for a protein's arginine side chain. These varied structures may have been exploited by evolution as protein interactions with RNA diversified for different roles. The arginine motifs described here are currently being used as binding domains for the construction of novel RNA catalysts.

REFERENCES

Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D., & Frankel, A. D. (1991) *Science 252*, 1167.

Ellington, A. D., & Szostak, J. W. (1990) *Nature 346*, 818.

Famulok, M., & Szostak, J. W. (1992) *J. Am. Chem. Soc. 114*, 3990.

Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., & Turner, D. H. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9373.

Lazinski, D., Grzadzielska, E., & Das, A. (1989) *Cell 59*, 207.

Michel, F., Hanna, M., Green, R., Bartel, D. P., & Szostak, J. W. (1989) *Nature 342*, 391.

Milligan, J. F., & Uhlenbeck, O. C. (1989) *Methods Enzymol. 180*, 51.

Pan, T., & Uhlenbeck, O. C. (1992) *Biochemistry 31*, 3887.

Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D., & Williamson, J. R. (1992) *Science 257*, 76.

Tao, J., & Frankel, A. D. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 2723.

Tuerk, C., & Gold, L. (1990) *Science 249*, 505.

Weeks, K. M., Ampe, C., Schultz, S. C., Steitz, T. A., & Crothers, D. M. (1990) *Science 249*, 1281.

Yarus, M. (1988) *Science 240*, 1751.

Yarus, M. (1989) *Biochemistry 28*, 980.

Yarus, M. (1991) *New Biol. 3*, 183.

Yarus, M. (1993) in *The RNA World* (Gesteland, R., & Atkins, J., Eds.) pp 205–207, Cold Spring Harbor Press, Cold Spring Harbor, NY.

Yarus, M., & Christian, E. (1989) *Nature 342*, 349.

Yarus, M., & Majerfeld, I. (1992) *J. Mol. Biol. 225*, 945.

Zuker, M. (1989) *Science 244*, 48.