Arginine-Binding RNAs Resembling TAR Identified by in Vitro Selection[†]

Jianshi Tao and Alan D. Frankel*

Department of Biochemistry and Biophysics and Gladstone Institute of Virology and Immunology, University of California, San Francisco, San Francisco, California 94143-0448

Received August 8, 1995; Revised Manuscript Received November 30, 1995[⊗]

ABSTRACT: Specific binding of the human immunodeficiency virus Tat protein to its RNA site (TAR) is mediated largely by a single arginine residue located within a basic region of the protein. Many essential features of the interaction can be mimicked by the free amino acid arginine, and an NMR model has been proposed in which the arginine guanidinium group binds to a guanine base in the major groove and to two phosphates adjacent to a bulge, with the RNA structure stabilized by a base triple between a U in the bulge and an adjacent A:U base pair. To compare the TAR structure to other arginine-binding RNAs, we performed in vitro selection experiments and identified RNAs with arginine-binding affinities similar to TAR. About 40% of the selected RNAs contained the same motif found in TAR: two stems separated by a bulge of at least two nucleotides, a U at the 5′ position of the bulge, and G:C and A:U base pairs above the bulge. In many cases, the upper stems contained only the G:C and A:U pairs, located next to small loops. Chemical modification experiments demonstrated that these "TAR-like" RNAs bound arginine in a manner similar to TAR, and in some cases identified nucleotides outside the binding site that contributed to binding. To explore how small loops might help stabilize the structures of adjacent arginine-binding sites, we measured arginine-binding affinities of TAR-like RNAs having all possible three-nucleotide loops. An RNA with a UAG loop bound with highest affinity, and chemical modification and RNase mapping experiments suggested that the RNA changes conformation upon arginine binding, converting a large unstructured loop into a bulge conformation related to that of TAR. The results suggest that the arginine-binding site in TAR is structurally versatile and demonstrate how binding can be modulated by the surrounding RNA context.

The ability of RNAs to fold into diverse tertiary structures enables RNAs to form complex macromolecular assemblies and in some cases to perform catalytic functions. While structural studies of RNAs are still at an early stage, several three-dimensional structures are known and many more are expected given recent advances in NMR (Dieckmann & Feigon, 1994) and crystallization (Doudna et al., 1993). The first crystal structures of tRNAs (Kim et al., 1974; Robertus et al., 1974; Quigley & Rich, 1976) provided examples of nonhelical interactions in RNAs, including non-Watson−Crick base pairs, base triples, and unusual stacking interactions and sugar−phosphate geometries. These local interactions give rise to the overall L-shaped structure of tRNAs. The recent crystal structures of a hammerhead ribozyme and ribozyme analog (Pley et al., 1994b; Scott et al., 1995) reveal a Y-shaped molecule that contains a uridine turn, similar to that found in $tRNA^{Phe}$, and non-Watson−Crick base pairs in its central core. Structural studies of extrastable tetraloops (Cheong et al., 1990; Heus & Pardi, 1991; Varani et al., 1991; Pley et al., 1994a), internal or terminal loops (Puglisi et al., 1990b; Holbrook et al., 1991; Szewczak et al., 1993; Wimberly et al., 1993), bulges (Puglisi et al., 1992), pseudoknots (Puglisi et al., 1990a), and mismatched bases (SantaLucia & Turner, 1993; Cruse et al., 1994) further illustrate the wide range of interesting base−base, base−backbone, and backbone−backbone interactions that can occur in RNA structures.

As anticipated, the shapes of RNAs and the precise positioning of functional groups provide important recognition features for proteins. In some cases, such as in the glutaminyl and aspartyl tRNA synthetase−tRNA complexes (Rould et al., 1989; Ruff et al., 1991), the protein and RNA share an extensive complementary interface, whereas in other cases, such as in the seryl-tRNA synthetase−tRNA complex (Biou et al., 1994), a smaller but relatively unique structural feature of the RNA (a long variable arm) contacts the protein. Cocrystal structures of the R17 coat protein and U1A protein bound to their RNA hairpin binding sites (Oubridge et al., 1994; Valegard et al., 1994) also show contacts to relatively localized regions of RNA structure. Similarly, complexes between RNA-binding peptides from the human immunodeficiency virus (HIV) Tat and Rev proteins and TAR and RRE RNA hairpins, and between a bovine immunodeficiency virus (BIV) Tat peptide and BIV TAR, show interactions with relatively small RNA structural motifs (Calnan et al., 1991b; Weeks & Crothers, 1991; Churcher et al., 1993; Tan et al., 1993; Battiste et al., 1994; Chen & Frankel, 1994; Peterson et al., 1994). Much of the specificity of the HIV Tat−TAR interaction can be modeled by the interaction of the free amino acid arginine with a three-nucleotide bulge region in TAR, even though the affinity of the interaction is much lower than with Tat peptides (Tao & Frankel, 1992). An NMR-derived model of a TAR−arginine complex suggests that arginine hydrogen bonds to a guanine base in the major groove above the bulge and to two phosphates

[†] Supported by NIH Grant GM47478.

* Address correspondence to this author at the Department of Biochemistry and Biophysics, UCSF, 513 Parnassus Avenue, San Francisco, CA. Telephone: 415-476-9994. FAX: 415-502-4315.

[⊗] Abstract published in *Advance ACS Abstracts,* February 1, 1996.

0006-2960/96/0435-2229$12.00/0

below the bulge, and that a base triple interaction between the 5′ U of the bulge and an A:U base pair above the bulge helps to stabilize the complex (Puglisi et al., 1992). Arginine also has been found to bind to the guanosine-binding site of the Tetrahymena group I intron, apparently using a different set of interactions than in TAR (Yarus, 1988, 1989; Yarus et al., 1991), and in vitro selection experiments have identified other RNAs that contain arginine-binding sites (Connell et al., 1993; Connell & Yarus, 1994; Famulok, 1994).

To further explore the structures of arginine-binding sites, and in particular those relevant to recognition by arginine-rich RNA-binding proteins (Mattaj, 1993; Burd & Dreyfuss, 1994), we have identified arginine-binding RNAs by in vitro selection using conditions designed to reflect important characteristics of the TAR−arginine complex, including possible interactions with the phosphate backbone. The arginine-binding RNAs identified are quite different from those previously found (Connell et al., 1993; Connell & Yarus, 1994; Famulok, 1994), and many contain binding sites related to that found in TAR. On the basis of these results, we have begun to define interactions in arginine-binding RNAs that help stabilize the binding conformations, and in particular have identified seemingly simple loop structures that enhance arginine binding. These molecules provide new insights into the structure of the TAR site and may provide good test cases for modeling RNA structure.

MATERIALS AND METHODS

Synthesis and Purification of RNAs. RNAs were synthesized using T7 RNA polymerase and synthetic or PCR-amplified DNA templates (Milligan & Uhlenbeck, 1989). The initial pool of random RNAs was synthesized using an 83-nucleotide template: 5′ GGG AGC CGT GCA CTG CAG (N_{30}) GCC GCG GGA ATT CTC CCT ATA GTG AGT CGT ATT AG 3′. RNAs were purified on 15% polyacrylamide/8 M urea gels as described (Calnan et al., 1991a).

In Vitro Selection of Arginine-Binding RNAs. The randomized pool of 65-nucleotide RNAs (30 μg) was folded by heating at 80 °C for 5 min and then slow-cooling to room temperature in 1 mL of column buffer (10 mM Tris-HCl, pH 7.5, 0.2 mM EDTA) containing 70 mM NaCl. Refolded RNAs were placed on ice and then loaded onto a 0.8 mL lysine−agarose column (Sigma) which had been equilibrated at 4 °C with column buffer. RNAs were fractionated using a 100 mL NaCl gradient from 70 to 500 mM, and the earliest eluting RNA-containing fractions were collected. The NaCl concentration was adjusted to 100 mM, the sample was loaded onto an equilibrated arginine−agarose column (6.3 μmol of L-arginine per milliliter of packed gel; Sigma), and RNAs were fractionated using a 100 mL NaCl gradient from 100 to 500 mM. Tight arginine-binding RNAs were precipitated with ethanol and reverse-transcribed to generate templates for the next round of T7 in vitro transcription (Ellington & Szostak, 1990; Tuerk & Gold, 1990). The lysine column preselection step was omitted during the first two rounds of selection.

Determination of Arginine-Binding Dissociation Constants (K_ds). Twenty micrograms of each 5′-^{32}P-labeled RNA (∼1 × 10^6 cpm) was diluted in 0.5 mL of column buffer containing 140−240 mM NaCl and loaded onto a 1 mL arginine−agarose column. RNAs were eluted under isocratic conditions using the same NaCl concentration as in the loading buffer, and elution volumes of the peak fractions were determined by Cerenkov counting. K_ds were estimated from the following equation: $K_d = [L][(V_t - V_o)/(V_e - V_o)]$, where V_t is the total column volume, V_o is the void volumn, V_e is the elution volume, and [L] is the concentration of ligand bound to the column (Connell et al., 1993; Famulok, 1994). To eliminate possible errors due to column variation, we determined relative K_ds for pairs of RNAs using the following equation: $K_{d2}/K_{d1} = (V_{e1} - V_o)/(V_{e2} - V_o)$, and estimated absolute values by comparing to the known K_d for TAR of ∼4 mM (Tao & Frankel, 1992). Apparent K_ds for TAR-like RNAs containing trinucleotide loops were estimated from the salt concentration required to elute each RNA from an arginine column, using a standard curve relating K_d values measured by isocratic elution to salt-dependent elution (see text). RNAs were eluted in 100 mL NaCl gradients (70−500 mM) at a flow rate of 0.2 mL/min, and isocratic elutions were performed at the same flow rate. Gradient elution experiments are considerably faster to perform than isocratic experiments, and we assume that all RNAs show similar salt-dependent changes in K_d over the range of salt concentrations used. The assumption appears to be valid based on the linear plot of Figure 8 and the observation that relative K_ds for several RNAs did not change when eluted under different isocratic salt conditions.

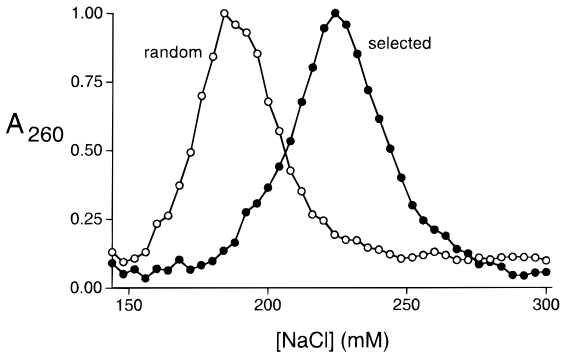

FIGURE 1: Elution profiles of random and selected RNAs. The starting pool of randomized RNAs (10 μg), or the pool after nine rounds of in vitro selection (10 μg), was bound to a 1 mL arginine−agarose column and eluted with a 100 mL NaCl gradient. One milliliter fractions were collected, and A_{260} was measured. Absorbance was normalized to the peak fraction for each pool.

Chemical Modification Interference Assays. Ten micrograms of each 5′-^{32}P-labeled RNA [(2−6) × 10^6 cpm] was modified with diethyl pyrocarbonate (DEPC; Sigma), dimethyl sulfate (DMS; Aldrich), hydrazine (HZ; Sigma), or ethylnitrosourea (ENU; Sigma) as previously described (Chen & Frankel, 1994). Modified RNAs were fractionated on an arginine−agarose column using standard NaCl gradients, and RNAs were ethanol-precipated and cleaved. Equal amounts of RNA, based on radioactivity, were dissolved in deionized formamide and analyzed on 20% polyacrylamide/8 M urea gels or 15% polyacrylamide/7 M urea/40% formamide gels.

RNase Mapping. Ten nanograms of each 5′-^{32}P-labeled RNA (∼50 000 cpm) was digested for 20 min on ice using 0.1 unit of RNase T1 or BC. Reactions contained 10 mM Tris-HCl (pH 7.5), 0.2 mM EDTA, and 70 mM NaCl. To probe arginine-bound conformations, RNAs were preincubated with 10 mM arginine for 10 min prior to RNase addition. Reactions were terminated by adding 10 μg of

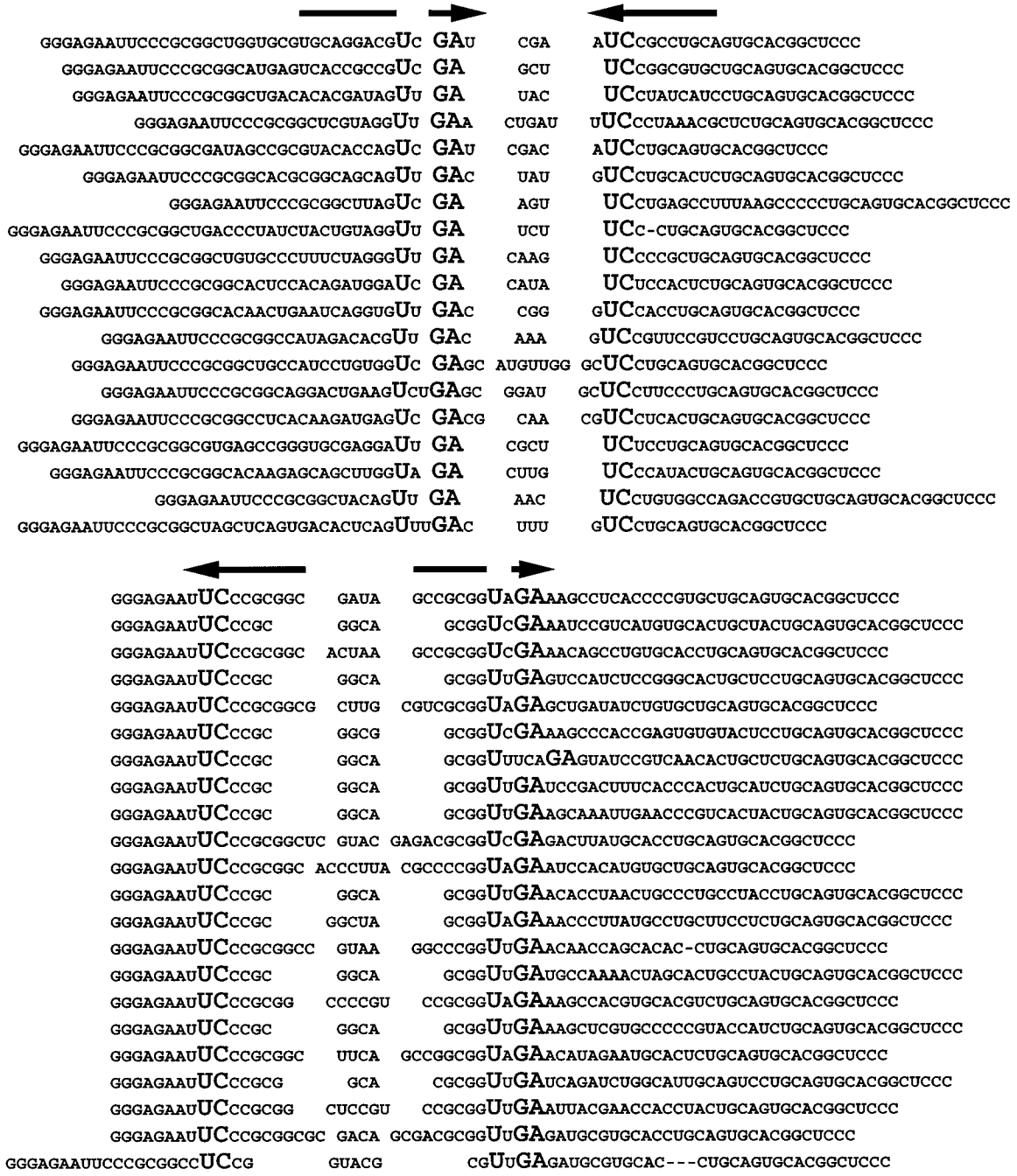

FIGURE 2: Sequences of selected TAR-like RNAs. Shown are 41 sequences (of the first 100 determined) that can be folded into hairpin structures containing 2−5-nucleotide bulges with U at the 5′ position of the bulge (large boldface), and G:C and A:U base pairs immediately 5′ to the bulge (large boldface). The TAR-like motifs in the top set of sequences are located next to terminal loops and in the bottom set are located next to internal loops (see Figure 3). Arrows indicate the stem regions and point toward the adjacent loops.

yeast tRNA (Sigma) and an equal volume of deionized formamide. Digestion products were analyzed on 15% polyacrylamide/8 M urea gels.

RESULTS

In Vitro Selection of Specific Arginine-Binding RNAs. We previously demonstrated that TAR RNA contains a specific arginine-binding site by comparing the relative abilities of wild-type TAR, TAR mutants, and chemically modified TARs to elute from an L-arginine−agarose column in a salt gradient (Tao & Frankel, 1992). The ionic strength required to elute the various RNAs correlated well with the relative binding affinities of Tat protein and Tat peptides, probably in part reflecting important interactions with phosphates near the bulge (Calnan et al., 1991b; Puglisi et al., 1992). To identify other arginine-binding RNAs having similar characteristics, we performed an in vitro selection experiment in which a pool of RNAs containing 30 random nucleotides was fractionated by salt-dependent elution from an arginine−agarose column. The total RNA length, including the PCR primer-binding sites, was 65 nucleotides, and we selected RNAs that eluted at a salt concentration slightly higher than wild-type TAR of the same length. On the particular column used, TAR eluted at 208 mM NaCl whereas a TAR mutant with a deletion of the bulge (considered "nonspecific" RNA) eluted at 160 mM NaCl (data not shown). The pool of random RNAs eluted at 184 mM NaCl (Figure 1), slightly later than the nonspecific control, perhaps reflecting inter-

molecular associations in the pool. Another random pool having different fixed flanking sequences also eluted at 184 mM NaCl (data not shown). We collected RNAs that eluted at 224 mM NaCl in an attempt to find RNAs that would preserve the types of structure found in a natural arginine-binding RNA but that might contain alternative or additional interactions. The difference in salt elution between TAR and RNAs eluting at the selected position corresponds to ~3−4-fold difference in arginine-binding affinity (see below). To avoid enrichment of RNAs that favored simple electrostatic interactions, binding to the amino acid backbone, or binding to the column matrix, we incorporated a negative selection step during each round of selection and amplification in which we collected the earliest eluting fractions from an L-lysine−agarose column. For both arginine and lysine columnns, the amino acid was linked via the α-amino group to a CNBr-activated agarose resin. After 9 rounds of selection, the RNAs reached the desired specificity (Figure 1).

Structures of Selected Arginine-Binding RNAs. cDNAs for the selected arginine-binding RNAs were cloned and sequenced. The secondary structures of the first 100 readable sequences were analyzed by Mulfold or Wisconsin GCG RNA folding programs, and obvious additional base pairs were adjusted manually. Although no 2 sequences were identical, 41 sequences potentially can fold into structures resembling the arginine-binding site in TAR, having a bulge of at least 2 nucleotides with U at the 5′ position, and G:C and A:U base pairs 3′ to the bulge (Figures 2 and 3). There are no obvious structural similarities among the remaining 59 sequences, although 2 may be related to the TAR motif in that they have a U-containing bulge within an extended stem, followed by a G:C pair 3′ to the bulge, followed by a bulged A (see below).

Among the selected "TAR-like" RNAs, shown schematically in Figure 3, the stem 5′ to the bulge often is quite extended, and most are closed by a G:C base pair adjacent to the bulge. In TAR, this base pair is an A:U pair, but when replaced by G:C, arginine-binding affinity increases by ~2-fold (data not shown). The A:U pair in TAR has a tendency to fray (Puglisi et al., 1992), suggesting that the stability of the stem contributes to formation of a good arginine-binding site. This is supported by the observation that some 5′ stems in the TAR-like molecules are closed at the other end by tetraloop sequences expected to adopt extrastable structures (Cheong et al., 1990; Heus & Pardi, 1991). In contrast to the stem 5′ to the bulge, the 3′ stem appears to be surprisingly unstable. Of the 41 TAR-like RNAs, 15 have just G:C and A:U base pairs located immediately adjacent to a terminal loop or internal loop, 18 have just 1 additional base pair, and only 8 have stems of 4 base pairs or more. The significance of the short 3′ stems is addressed below. For most of the molecules with internal loops, nucleotides within the fixed sequences are used to form the binding site.

Most of the TAR-like molecules contain dinucleotide bulges, compared to the trinucleotide bulge of TAR. Although the size of the TAR bulge only minimally affects Tat protein or peptide binding, provided that it contains at least two nucleotides (Weeks & Crothers, 1991; Churcher et al., 1993), we found that the size of the bulge has a more significant effect on arginine-binding affinity, especially when the binding site is placed next to a single-stranded

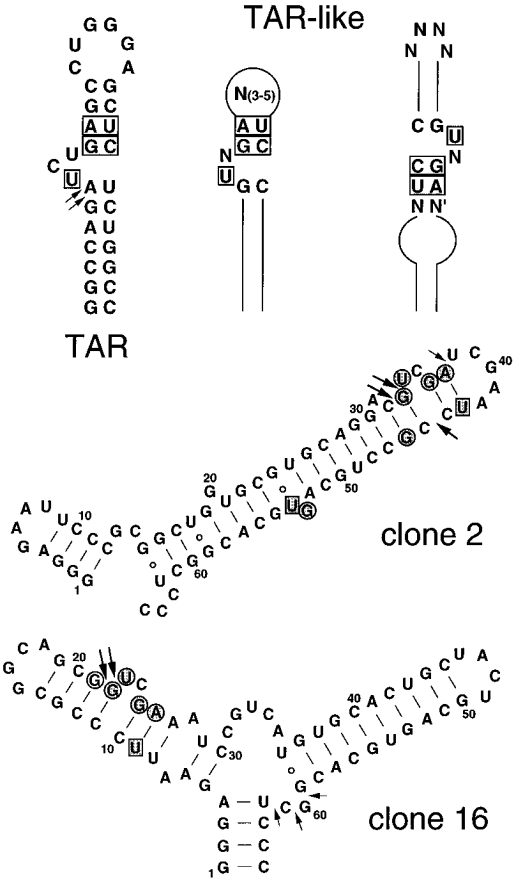

FIGURE 3: Secondary structures of TAR, TAR-like RNAs, and two TAR-like clones. Nucleotides important for arginine binding in TAR are boxed, and the positions of two important phosphates are indicated by arrows. Analogous nucleotides in the TAR-like structures are boxed. Most TAR-like molecules with the arginine-binding site located adjacent to an internal loop (top right structure) have one additional base pair in the 5′ stem (N:N′). Nucleotides and phosphates important for arginine binding in clones 2 and 16 were determined by chemical modification interference experiments (Figure 4) and are indicated by shaded circles (strong interference), open circles (weak interference), shaded boxes (weak enhancement), and large and small arrows (strong and weak phosphate interference).

region as in most of the selected TAR-like molecules (data not shown). We speculate that a smaller bulge may have fewer close phosphate contacts when bound to arginine than a larger bulge, or that less entropy is lost as the site rearranges to its bound conformation. In the Tat/TAR complex, it has been proposed that poor electrostatic contacts in the bulge may be compensated by positively charged side chains surrounding the arginine (Tao & Frankel, 1993).

Chemical and RNase Mapping of Arginine-Binding RNAs. To test whether the TAR-like molecules indeed formed arginine-binding sites similar to that in TAR, we performed chemical modification interference experiments on two representive sequences (clones 2 and 16; Figure 3) that eluted from the arginine−agarose column at 220 and 224 mM NaCl. Base-specific reactions were carried out with hydrazine (HZ), diethyl pyrocarbonate (DEPC), and dimethyl sulfate (DMS), and phosphate-specific reactions were carried out with ethylnitrosourea (ENU) (Figure 4). In both RNAs, modification of the 5′-bulged U, the G and A bases 3′ to the bulge, and two phosphates at the junction of the bulge and 5′ stem interfered with arginine binding, as observed for TAR (Tao & Frankel, 1992; data not shown). Modification of Gs in

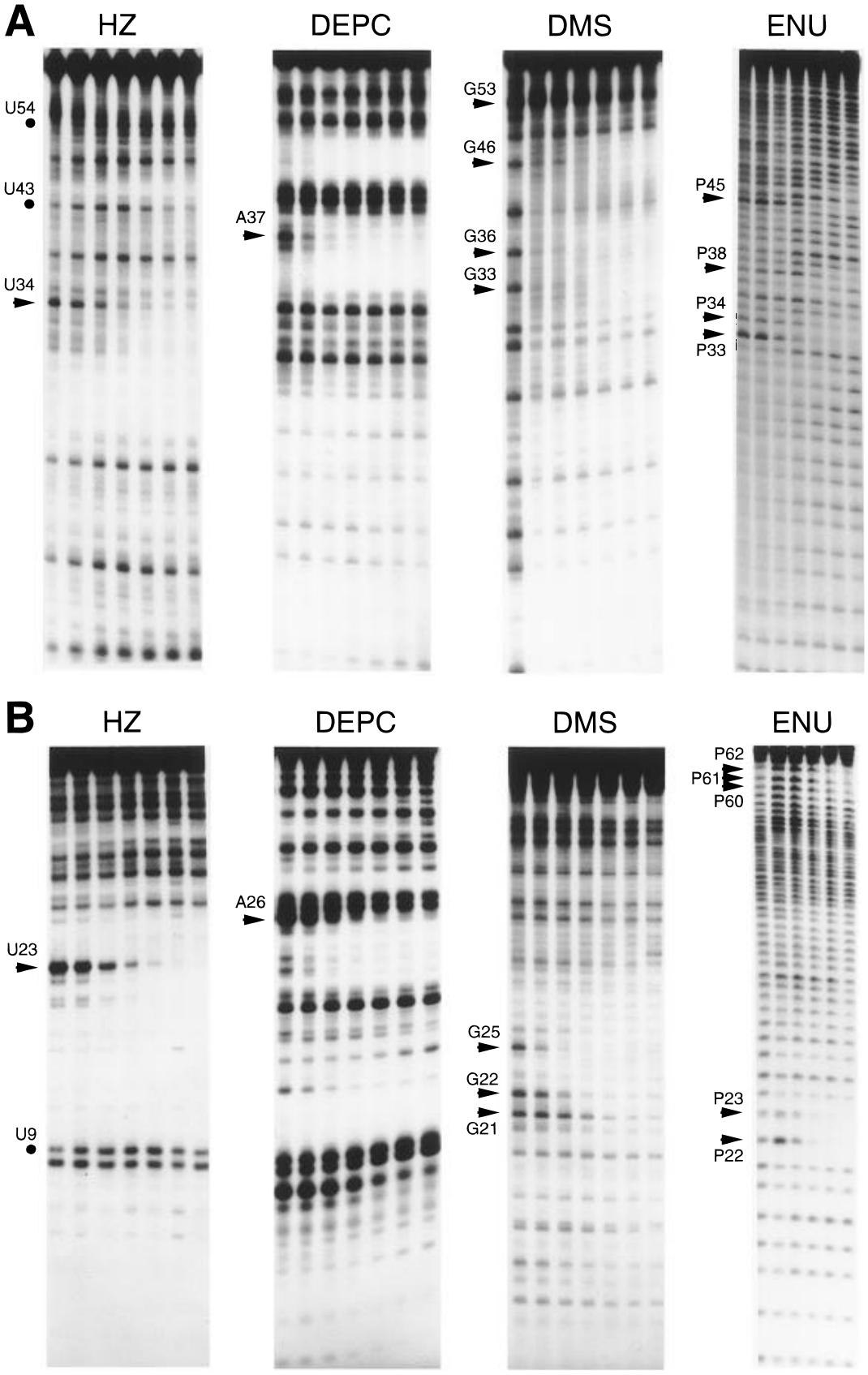

FIGURE 4: Chemical modification interference of TAR-like clones 2 and 16. Clone 2 (A) and clone 16 (B) RNAs were modified with hydrazine (HZ; pyrimidine-specific), diethyl pyrocarbonate (DEPC; primarily A-specific), dimethyl sulfate (DMS; G-specific), and ethylnitrosourea (ENU; phosphate-specific) and were eluted from an arginine−agarose column with an NaCl gradient. Fractions containing modified RNAs were cleaved and run on gels, with salt concentration increasing from left to right. Interference is indicated by the loss of bands in the late-eluting fractions (arrows) and positions of enhancement by an increase in band intensity in the late-eluting fractions (closed circles). Several phosphates 5′ to P33 in clone 2, located in the stem below the TAR-like arginine-binding site (Figure 3), also show enhancement but are not labeled.

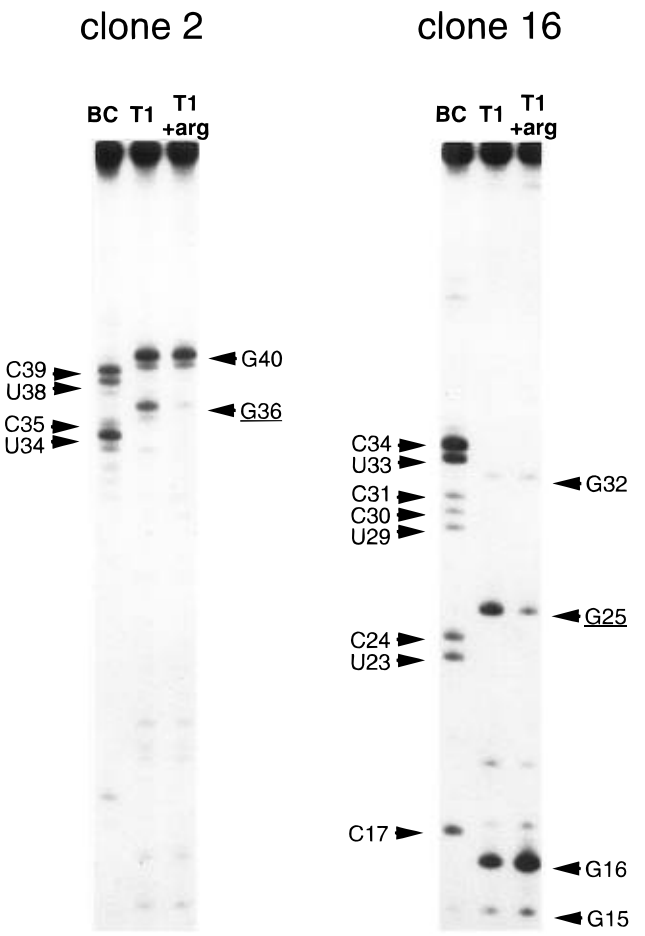

FIGURE 5: RNase mapping of TAR-like clones 2 and 16. RNAs were cleaved with RNase T1, which cleaves at single-stranded G residues, in the absence or presence of 10 mM arginine. Nucleotides protected from cleavage in the presence of arginine are underlined. RNase T1 activity is not generally inhibited at this concentration of arginine. RNase BC, which cleaves at single-stranded pyrimidines, was also used to map the unbound RNA structures.

the 5′ stem also interfered with binding, probably due to destabilization of the stem or steric hindrance by the methyl group. Similar interference has been observed at the analogous A in TAR (Weeks et al., 1990). Unlike TAR, modification of the bulge phosphates did not cause enhanced binding, perhaps because the electrostatic environment of the dinucleotide bulge is more favorable. Interestingly, modification of U43 in clone 2 and U9 in clone 16, expected to be part of the A:U base pair, enhanced binding. As discussed below, partial disruption of the Watson−Crick base pair of the proposed base triple may actually aid in arginine binding.

In addition to the TAR-like interference pattern, modification of a few additional positions in both clones had some effect on arginine binding. In particular, modification of G53 in clone 2 interfered with binding, while modification of the adjacent U, expected to destablize the stem and make the G more accessible, enhanced binding. Modification of several phosphates in the stem between G53 and the TAR-like site also enhanced binding. We suspect that G53 may participate in a long-range interaction that helps stabilize the arginine-binding conformation or prevent formation of alternative structures, although the structural basis for the effect remains to be determined. Modification of nucleotides in single-stranded regions adjacent to the binding site, which might

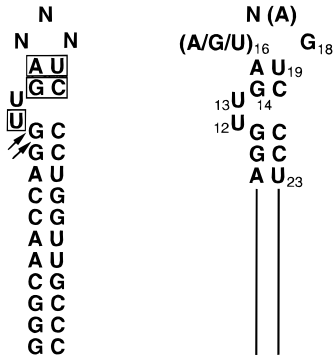

FIGURE 6: Secondary structures of TAR-like RNAs with 3-nucleotide loops. Randomized loops, shown on the left as a trinucleotide loop, were placed directly adjacent to an arginine-binding site. Nucleotides important for arginine binding in TAR are boxed, and the positions of two important phosphates are indicated by arrows. Shown on the right is a consensus sequence for the tightest arginine-binding RNAs containing trinucleotide loops (from Table 1). A, G, or U are favored at the first position of the loop, A is slightly favored at the second position, and G is strongly favored at the third position.

have been expected to contribute stabilizing interactions, did not interfere with binding. However, since DMS and DEPC modify only the N7 positions of G and A, we cannot rule out the possibility that other functional groups may be involved.

RNase T1 digestions were performed on clones 2 and 16 in the absence or presence of arginine to further localize the arginine-binding sites. In the absence of arginine, G36 of clone 2 and G25 of clone 16 were readily accessible to cleavage but were protected in the presence of 10 mM arginine (Figure 5). The corresponding guanine in TAR, which is part of a 4 base pair stem and is inaccessible to RNase T1 in its unbound form (Colvin & Garcia-Blanco, 1992; data not shown), is proposed to directly hydrogen bond to arginine (Puglisi et al., 1992).

Arginine-Binding Sites with Adjacent Loops. Because many of the TAR-like arginine-binding sites are located next to terminal loops or internal loops, we reasoned that particular loop sequences might help stabilize the structure of the binding site and thereby increase affinity for arginine. Placing an extrastable GAAA tetraloop next to the site only slightly increased arginine-binding affinity compared to a CCCC loop, which bound nonspecifically (data not shown). To identify other loops with possible stabilizing interactions, we performed in vitro selection experiments in which a fixed arginine-binding site was located next to randomized loops of 3−5 nucleotides (see Figure 6). After selecting for RNAs with arginine column elution profiles similar to the TAR-like molecules, we found that most had unexpected mutations in other regions of the RNA, presumably generated during PCR amplification, which we suspect may contribute long-range interactions (J. Tao, J. Guthrie, and A. D. Frankel, unpublished data). To limit possible interactions only to the adjacent loops, we synthesized RNAs with each of 64 possible trinucleotide loops (Figure 6) and measured their elution profiles (Table 1). To minimize the number of RNAs synthesized, we first made 16 sets of 4 RNAs each in which the first and third nucleotides were fixed and the middle nucleotide was randomized. Most of these sets eluted as single peaks from the arginine column (for example, the GNU set; Figure 7), suggesting that the trinucleotide loop structures are determined largely by the first and the third

Table 1: Arginine Binding of TAR-like RNAs with Three-Nucleotide Loops[a]

loop	sequence	specificity (Δ[NaCl], mM)	apparent K_d (mM)
TAR		72	4
ANA		12−32	
	AAA	8	48
	ACA	20	28
	AGA	32	17
	AUA	32	17
ANC		52	7.6
ANG		44 and 80	
	AAG	**80**	**2.7**
	ACG	36	14
	AGG	44	10
	AUG	40	12
ANU		48	8.8
CNA		32	17
CNC		48	8.8
CNG		44	10
CNU		52	7.6
GNA		24 and 48	
	GAA	20	28
	GCA	40	12
	GGA	48	8.8
	GUA	48	8.8
GNC		48	8.8
GNG		**80**	**2.7**
GNU		64	4.8
UNA		52	7.6
UNC		52	7.6
UNG		76−92	
	UAG	**100**	**1.4**
	UCG	**76**	**3.1**
	UGG	**80**	**2.7**
	UUG	**88**	**2.0**
UNU		68	4.1

[a] Arginine-binding specificity is defined as the difference in NaCl concentration required to elute each RNA from an arginine−agarose column, relative to bulge-deleted TAR (the nonspecific control). Each RNA was mixed with bulge-deleted TAR and eluted from the column. Apparent K_ds were calculated using the standard curve in Figure 8. RNAs that have higher arginine-binding affinities than TAR are shown in boldface type.

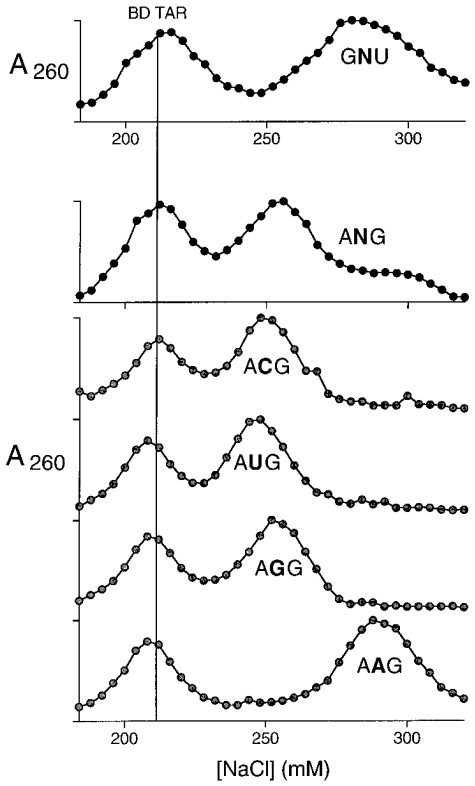

FIGURE 7: Elution profiles of TAR-like RNAs with 3-nucleotide loops. Shown are elutions of GNU and ANG RNAs, where N is any of the four nucleotides. The GNU set gave one peak whereas the ANG set gave two peaks. Elution profiles were measured for individual members of the ANG set to determine which sequences were represented in each peak. In all cases, bulge-deleted TAR RNA (BD TAR) was eluted in the same gradient to mark the position of nonspecific binding.

positions. Four of the 16 sets gave heterogeneous peaks, and individual RNAs in these sets were synthesized and eluted from the arginine column (for example, the ANG set; Figure 7).

To establish the relationship between arginine-binding affinities and salt elution from the arginine column, we measured elution volumes of a series of RNAs run under isocratic conditions. K_ds can be estimated from the following equation: $K_d = [L][(V_t - V_o)/(V_e - V_o)]$, where V_t is the total column volume, V_o is the void volumn, V_e is the elution volume, and [L] is the concentration of ligand bound to the column (Connell et al., 1993; Famulok, 1994). The relative affinities of any two RNAs can be determined from the following equation: $K_{d2}/K_{d1} = (V_{e1} - V_o)/(V_{e2} - V_o)$, and can be estimated for arginine-binding RNAs by comparing to the known K_d for TAR of ∼4 mM (Tao & Frankel, 1992). The logarithmic relationship between apparent K_ds and the concentration of salt required to elute from the arginine column (Figure 8) is consistent with displacement of phosphate-bound counterions upon arginine binding, as observed for DNA-binding proteins (Lohman & Mascotti, 1992), and with the proposed interaction of arginine with a pair of phosphates in TAR (Calnan et al., 1991b; Puglisi et al., 1992). Using this method, the difference in affinity between wild-type TAR and the bulge deletion mutant is ∼20-fold, slightly less than the discrimination observed with Tat protein or Tat peptides (Sumner-Smith et al., 1991; Weeks & Crothers, 1992; Tao & Frankel, 1993). Apparent arginine-binding affinities for all RNAs with trinucleotide loops, estimated from the standard curve in Figure 8, are summarized in Table 1, and a consensus sequence for the tightest binders is shown in Figure 6. The tight binders all contain G at the third position of the loop.

Mapping the Arginine-Binding Site with an Adjacent UAG Loop. The TAR-like RNA containing a UAG trinucleotide loop bound arginine most tightly, with ∼3−4-fold higher affinity than TAR (Table 1). RNase T1 digestions were performed on this RNA in the absence and presence of 10 mM arginine to localize the site of arginine binding and to examine the stability of the putative two base pair stem 3′ to the bulge. In the absence of arginine, G14, corresponding to the contacted base in TAR, and G18 in the UAG loop were completely accessible to cleavage (Figure 9A), suggesting that the two base pair stem does not form in the unbound structure. Upon arginine binding, G14 became completely protected whereas G18 remained accessible, suggesting formation of a TAR-like site. This behavior is similar to that seen for TAR-like clones 2 and 16 (Figure 5), but the difference in cleavage is even more dramatic. Circular dichroism experiments support a conformational change upon binding (data not shown). To attempt to identify determinants that lead to such a stable conformation

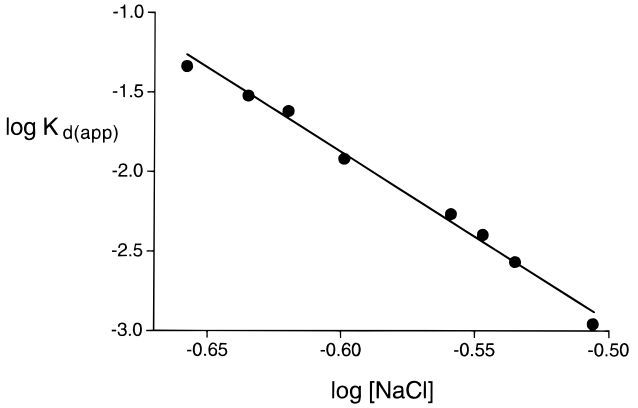

FIGURE 8: Relationship between salt elution profiles and apparent arginine-binding affinities. Apparent arginine-binding dissociation constants [$K_{d(app)}$] were determined for several RNAs by comparing the isocratic elution volume of each RNA to the isocratic elution volume of TAR, which has a known dissociation constant of ~4 mM (Tao & Frankel, 1992). Equations relating relative binding constants to elution volumes are described in the text. log $K_{d(app)}$ is proportional to log [NaCl] required to elute each RNA from an arginine−argarose column. RNAs used to establish this standard curve were (in order of affinities): bulge-deleted TAR, TAR-like RNA with AAA loop, TAR-like RNA with GAA loop, U23→C bulge-substituted TAR, TAR-like RNA with GCA loop, TAR-like RNA with GNU loop, wild-type TAR, TAR-like RNA with UGG loop, and TAR-like RNA with UAG loop. Isocratic elutions were performed at different salt concentrations for some of the RNAs, and relative binding constants were the same under all conditions.

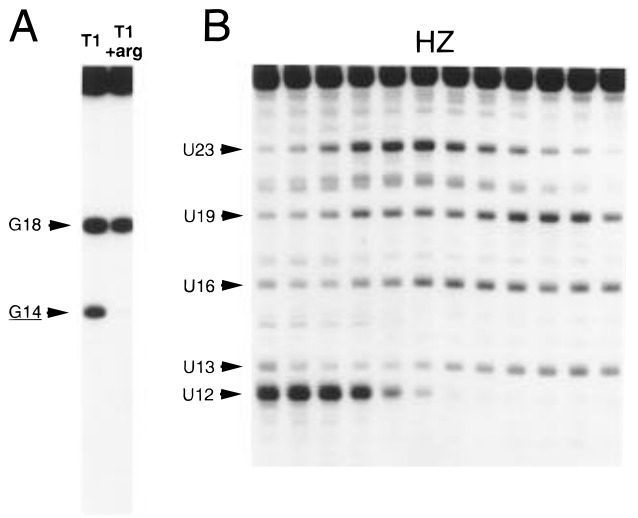

FIGURE 9: RNase mapping and hydrazine interference of a TAR-like RNA with a UAG loop. (A) RNA was cleaved with RNase T1, which cleaves at single-stranded G residues, in the absence or presence of 10 mM arginine. G14 (underlined) is protected from cleavage in the presence of arginine. (B) RNA was modified with hydrazine (HZ) and was eluted from an arginine−agarose column with an NaCl gradient, as described in Figure 4.

when bound to arginine, we used hydrazine interference to probe the importance of uridines in the UU bulge, the UAG loop, and the A:U base pair (see Figure 6). Only modification of U12, proposed to form a Hoogsteen interaction with the A of the base triple, strongly interfered with binding (Figure 9B). Slight interference was observed upon modification of U23, expected to destablize the lower stem. Surprisingly, modification of U19, the Watson−Crick partner of the proposed base triple, did not interfere with binding. The significance of this result is discussed below. The lack of interference at U13, the second nucleotide of the bulge, appears to rule out a model in which U13 pairs with G18 of the loop (the only conserved nucleotide in the consensus; see Figure 6), potentially forming the pseudoknot structure shown in Figure 10. Alternative models are discussed below.

DISCUSSION

In this study, we performed in vitro selection experiments to identify arginine-binding RNAs by using a salt gradient to elute tight-binding RNAs from an arginine column (Tao & Frankel, 1992). This method has been used to study arginine binding to TAR and differs from the ligand-specific elution protocols used in other reports of arginine binders, which typically selected for tighter binding RNAs (Connell et al., 1993; Connell & Yarus, 1994; Famulok, 1994). Consequently, previous selection experiments did not identify TAR-like molecules and our study did not identify previously reported motifs, which may recognize more extensive features of the arginine side chain. Salt-dependent elution also has been used to identify arginine-binding DNAs, and interestingly, no TAR-like molecules were found (Harada & Frankel, 1995). We chose to examine RNAs with similar affinities to TAR and those containing salt-sensitive arginine-binding sites because (i) they might be expected to exist under physiologic conditions, (ii) they are most likely to involve interesting phosphate contacts that presumably would be shielded by high ionic strength, and (iii) guanidinium−phosphate contacts may provide important constraints for use in modeling the structures. The logarithmic relationship between apparent K_ds and salt concentration is similar to that observed when counterions are released from DNA phosphates upon protein binding (Lohman & Mascotti, 1992), and the phosphate interference patterns observed with TAR-like molecules are consistent with the proposed guanidinium−phosphate contacts in TAR (Calnan et al., 1991b; Puglisi et al., 1992). It is unlikely that salt-dependent effects on RNA structure occur over the relatively narrow range of salt concentrations used in these experiments because: (i) we have tested binding of RNAs to arginine columns having different concentrations of bound ligand and have measured similar binding constants even though the RNA elutes at a different ionic strength, (ii) we have performed isocratic elution of several RNAs under different salt conditions and have measured similar binding constants, and (iii) at least for TAR, no salt-dependent changes in RNA structure have been observed.

About 40% of the molecules selected by our protocol contained TAR-like arginine-binding sites, and we have focused our experiments on these RNAs. The remaining molecules contained no obvious common motifs and may be further studied in the future. Based on the selection results, we identified a tight-binding TAR-like RNA in which a UAG trinucleotide loop was placed directly adjacent to the A:U base pair of the proposed base triple in TAR, with only two base pairs between the bulge and loop (see Figure 10). The binding site in this RNA is highly accessible to RNase cleavage and is probably unstructured in the absence of arginine even though it binds arginine with higher affinity than does TAR. Hydrazine interference experiments indicate that the bulged U expected to form a Hoogsteen interaction with the A of the triple is critical for binding but, surprisingly, the U expected to form the Watson−Crick interaction of the triple is unimportant. A similar result was observed with two selected TAR-like clones. By analogy, perhaps only the Hoogsteen interaction of the proposed base triple in TAR

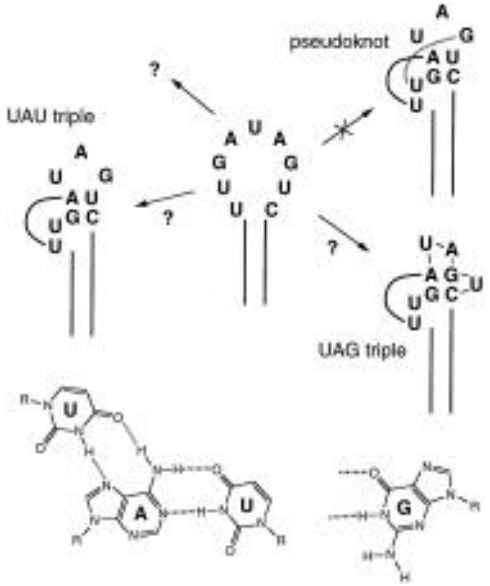

FIGURE 10: Models for the TAR-like RNA with a UAG loop. In the absence of arginine, the bulge and loop region is accessible to nucleases and is probably unstructured (middle). A few possible models for the bound structure are shown and are described in the text. G and U potentially can hydrogen bond to A in similar ways (below). Many other models are equally consistent with the data.

is critical for binding, explaining why substitution of the A with N^6-methyl-dA, which can still maintain the Hoogsteen interaction, has only a small effect on the Tat-binding affinity (Hamy et al., 1993). Mutants of TAR that disrupt the Watson−Crick pair but maintain the Hoogsteen pair also are able to bind arginine (J. Tao and A. D. Frankel, unpublished results), consistent with this interpretation. It has been shown that a TAR mutant in which the proposed A-U:A base triple was replaced by an isosteric C^+-G:C triple also binds arginine and forms a structure similar to wild-type TAR (Puglisi et al., 1993). It will be interesting to perform analogous in vitro selection experiments at low pH (allowing protonation of C residues) to determine whether arginine-binding TAR-like structures can form with a C^+-G Hoogsteen interaction.

To explain why G at the third position of TAR-like RNAs with trinucleotide loops was found in the tightest binding RNAs, we suggest one possible model (Figure 10) in which G replaces U of the Watson−Crick pair to form a U-A-G base triple. It seems plausible that an A-G pair may allow the U-A Hoogsteen interaction to adopt a better geometry for arginine binding than a helical base pair, perhaps optimizing stacking of the arginine guanidinium group under the U (Puglisi et al., 1992). Alternatively, G may interact with groups on the backbone or perhaps form additional interactions with arginine. To explain why U and A of the trinucleotide loop are best, we suggest that they may be the least likely to form alternative structures. We have made variants of the UAG TAR-like molecule in which we expected to create isosteric substitutions or replace unimportant nucleotides (based on the U-A-G base triple or pseudoknot models shown in Figure 10, or on other models), but all bound arginine much less tightly and many formed alternative structures (data not shown). Additional information will be needed to distingish between the many possible models for this RNA.

The TAR-like RNAs and other arginine-binding RNAs identified in this study appear to exhibit many of the complexities observed with much larger RNA structures, including formation of non-Watson−Crick interactions and alternative conformations. There are many cases in which small RNAs removed from their normal RNA contexts adopt conformations important for function or protein binding. For example, an analog of the decoding region of 16S rRNA appears to correctly mimic interactions with antibiotics, tRNA, and mRNA, even outside of its much larger ribosomal scaffold (Purohit & Stern, 1994). Arginine-binding RNAs may contain structures relevant for protein recognition, particularly for arginine-rich RNA-binding proteins, and may provide useful systems for modeling RNA structure. We have obtained a few experimental constraints that lead us to propose a preliminary model for one molecule (Figure 10); however, we have no systematic way to evaluate the many other models that equally statisfy the constraints, or to design the best experiments to distinguish between possible models. We plan to address these problems in a computer-based modeling strategy in which experimental constraints are used to model local elements of RNA structure (B. Walberer and A. D. Frankel, unpublished results). We expect that the localized nature of arginine-binding sites will provide a distinct advantage in initial modeling attempts, but given the complexity already observed, we anticipate interesting and substantial challenges ahead.

ACKNOWLEDGMENT

We thank Jessica Guthrie for technical assistance, members of the laboratory for helpful suggestions, and Jody Puglisi and Dan Herschlag for comments on the manuscript.

REFERENCES

Battiste, J. L., Tan, R., Frankel, A. D., & Williamson, J. R. (1994) *Biochemistry 33*, 2741−2747.
Biou, V., Yaremchuk, A., Tukalo, M., & Cusack, S. (1994) *Science 263*, 1404−1410.
Burd, C. G., & Dreyfuss, G. (1994) *Science 265*, 615−621.
Calnan, B. J., Biancalana, S., Hudson, D., & Frankel, A. D. (1991a) *Genes Dev. 5*, 201−210.
Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D., & Frankel, A. D. (1991b) *Science 252*, 1167−1171.
Chen, L., & Frankel, A. D. (1994) *Biochemistry 33*, 2708−2715.
Cheong, C., Varani, G., & Tinoco, I., Jr. (1990) *Nature 346*, 680−682.
Churcher, M. J., Lamont, C., Hamy, F., Dingwall, C., Green, S. M., Lowe, A. D., Butler, J. G., Gait, M. J., & Karn, J. (1993) *J. Mol. Biol. 230*, 90−110.
Colvin, R. A., & Garcia-Blanco, M. A. (1992) *J. Virol. 66*, 930−935.
Connell, G. J., Illangesekare, M., & Yarus, M. (1993) *Biochemistry 32*, 5497−5502.
Connell, G. J., & Yarus, M. (1994) *Science 264*, 1137−1141.
Cruse, W. B., Saludjian, P., Biala, E., Strazewski, P., Prange, T., & Kennard, O. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 4160−4164.
Dieckmann, T., & Feigon, J. (1994) *Curr. Opin. Struct. Biol. 4*, 745−749.
Doudna, J. A., Grosshans, C., Gooding, A., & Kundrot, C. E. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 7829−7833.
Ellington, A. D., & Szostak, J. W. (1990) *Nature 346*, 818−822.
Famulok, M. (1994) *J. Am. Chem. Soc. 116*, 1698−1706.
Hamy, F., Asseline, U., Grasby, J., Iwai, S., Pritchard, C., Slim, G., Butler, P. J., Karn, J., & Gait, M. J. (1993) *J. Mol. Biol. 230*, 111−123.
Harada, K., & Frankel, A. D. (1995) *EMBO J. 14*, 5798−5811.
Heus, H. A., & Pardi, A. (1991) *Science 253*, 191−194.
Holbrook, S. R., Cheong, C., Tinoco, I., Jr., & Kim, S. H. (1991) *Nature 353*, 579−581.

Kim, S. H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A. H., Seeman, N. C., & Rich, A. (1974) *Science 185*, 435−440.
Lohman, T. M., & Mascotti, D. P. (1992) *Methods Enzymol. 212*, 400−424.
Mattaj, I. W. (1993) *Cell 73*, 837−840.
Milligan, J. F., & Uhlenbeck, O. C. (1989) *Methods Enzymol. 180*, 51−62.
Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., & Nagai, K. (1994) *Nature 372*, 432−438.
Peterson, R. D., Bartel, D. P., Szostak, J. W., Horvath, S. J., & Feigon, J. (1994) *Biochemistry 33*, 5357−5366.
Pley, H. W., Flaherty, K. M., & McKay, D. B. (1994a) *Nature 372*, 111−113.
Pley, H. W., Flaherty, K. M., & McKay, D. B. (1994b) *Nature 372*, 68−74.
Puglisi, J. D., Wyatt, J. R., & Tinoco, I., Jr. (1990a) *J. Mol. Biol. 214*, 437−453.
Puglisi, J. D., Wyatt, J. R., & Tinoco, I., Jr. (1990b) *Biochemistry 29*, 4215−4226.
Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D., & Williamson, J. R. (1992) *Science 257*, 76−80.
Puglisi, J. D., Chen, L., Frankel, A. D., & Williamson, J. R. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3680−3684.
Purohit, P., & Stern, S. (1994) *Nature 370*, 659−662.
Quigley, G. J., & Rich, A. (1976) *Science 194*, 796−806.
Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F., & Klug, A. (1974) *Nature 250*, 546−551.
Rould, M. A., Perona, J. J., Söll, D., & Steitz, T. A. (1989) *Science 246*, 1135−1142.
Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C., & Moras, D. (1991) *Science 252*, 1682−1689.
SantaLucia, J., Jr., & Turner, D. H. (1993) *Biochemistry 32*, 12612−12623.
Scott, W. G., Finch, J. T., & Klug, A. (1995) *Cell 81*, 991−1002.
Sumner-Smith, M., Roy, S., Barnett, R., Reid, L. S., Kuperman, R., Delling, U., & Sonenberg, N. (1991) *J. Virol. 65*, 5196−5202.
Szewczak, A. A., Moore, P. B., Chang, Y. L., & Wool, I. G. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 9581−9585.
Tan, R., & Frankel, A. D. (1992) *Biochemistry 31*, 10288−10294.
Tan, R., Chen, L., Buettner, J. A., Hudson, D., & Frankel, A. D. (1993) *Cell 73*, 1031−1040.
Tao, J., & Frankel, A. D. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 2723−2726.
Tao, J., & Frankel, A. D. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 1571−1575.
Tuerk, C., & Gold, L. (1990) *Science 249*, 505−510.
Valegard, K., Murray, J. B., Stockley, P. G., Stonehouse, N. J., & Liljas, L. (1994) *Nature 371*, 623−626.
Varani, G., Cheong, C., & Tinoco, I., Jr. (1991) *Biochemistry 30*, 3280−3289.
Weeks, K. M., & Crothers, D. M. (1991) *Cell 66*, 577−588.
Weeks, K. M., & Crothers, D. M. (1992) *Biochemistry 31*, 10281−10287.
Weeks, K. M., Ampe, C., Schultz, S. C., Steitz, T. A., & Crothers, D. M. (1990) *Science 249*, 1281−1285.
Wimberly, B., Varani, G., & Tinoco, I., Jr. (1993) *Biochemistry 32*, 1078−1087.
Yarus, M. (1988) *Science 240*, 1751−1758.
Yarus, M. (1989) *Biochemistry 28*, 980−988.
Yarus, M., Illangesekare, M., & Christian, E. (1991) *J. Mol. Biol. 222*, 995−1012.

BI951844B