23S rRNA Similarity from Selection for Peptidyl Transferase Mimicry †

ARTICLE *in* BIOCHEMISTRY · JULY 1997

Impact Factor: 3.02 · DOI: 10.1021/bi963135j · Source: PubMed

3 AUTHORS, INCLUDING:

Mark Welch
DNA2.0
18 PUBLICATIONS **1,057** CITATIONS

23S rRNA Similarity from Selection for Peptidyl Transferase Mimicry†

Mark Welch, Irene Majerfeld, and Michael Yarus*

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

Received December 23, 1996; Revised Manuscript Received March 26, 1997⊗

ABSTRACT: RNAs from a randomized pool were selected by affinity elution for binding to the molecule CCdApPuro, a high-affinity ligand of ribosomal peptidyl transferase designed as a transition-state analogue of peptide formation. The selected RNAs show affinity for CCdApPuro comparable to that of the peptidyl transferase center itself ($K_d \approx 10$ nM). Chemical modification/protection experiments implicate bases completely conserved among the selected RNAs in CCdApPuro interaction, which appears to involve both CCdA and puromycin moieties, that is, both A- and P-site homologues. The apparent selected binding site shows up to 17 nucleotides with similarity to conserved nucleotides of the peptidyl transferase loop domain of 23S rRNA and is conserved when reselected under mutagenesis. Thus, these nucleotides of 23S rRNA likely provide elements of the peptidyl transferase active center that bind the reactants near the site of peptide bond formation. Binding of CCdApPuro by a peptidyl transferase-like motif in the absence of protein strengthens the hypothesis that peptidyl transfer originated in an RNA world.

The ubiquitous integral ribosomal catalyst for peptide bond formation is termed peptidyl transferase. The composition of the active center of this enzyme is only beginning to be unraveled, though many lines of evidence implicate elements of the ribosomal large-subunit RNA [Samaha et al., 1995; Noller et al., 1992; Moazed & Noller, 1989; for review, see Noller (1991)]. Still, little of the structure and participation of rRNA is proven. Recently, a direct rRNA−tRNA contact was shown by Samaha et al. (1995) to occur upon binding of tRNA substrate to the P site (donor site) of peptidyl transferase. A Watson−Crick base pair appears to form between C74 at the acceptor end of the tRNA and G2252 of 23S rRNA. G2252 and several other large-subunit rRNA bases associated with peptidyl tranferase are completely conserved in the rRNAs of organisms and organelles (Gutell et al., 1993) and are thus as ancient as the progenitor of all known species. Elucidation of the structure and function of these rRNA elements may provide insight into the evolution of protein synthesis.

Previously, we described a high-affinity ligand for the peptidyl transferase center that may serve as a probe and/or template for active-site structure (Welch et al., 1995). This molecule (CCdApPuro), designed as a transition-state analogue for peptidyl transfer, consists of the trinucleotide 5′-CCdAp-3′ (an analogue of the ubiquitous 3′-terminal sequence of tRNA) linked at the 3′-position through a phosphoramide to the antibiotic puromycin (an acceptor or A-site substrate of peptidyl transferase; Figure 1). This molecule is proposed to mimic elements of the A- and P-site peptidyl transferase substrates positioned around a tetrahedral charged center as predicted in the transition state of a reaction mechanism involving direct nucleophilic attack (Welch et al., 1995). CCdApPuro was shown to bind strongly to peptidyl transferase; K_d's of 2−90 nM were measured (Welch et al., 1995). Both puromycin and CCdA moieties contribute to binding, and CCdApPuro appears to bridge the P and A peptidyl transferase substrate sites. Upon binding, CCdAp-Puro protects bases of 23S rRNA associated with peptidyl transferase from attack by chemical probes (Welch et al., 1995), supporting the presence of rRNA at the active center and suggesting ribosomal binding at least in part via interaction with the RNA itself. Now we describe the structure of a new RNA selected to bind to CCdApPuro with affinity on the order of that of peptidyl transferase.

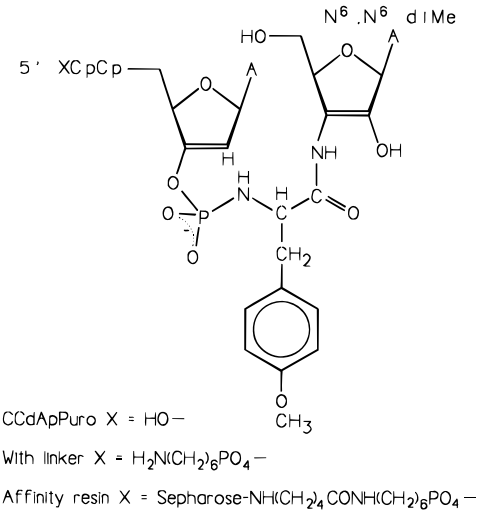

FIGURE 1: Structure of CCdApPuro and attachment to the selection column matrix. Coupling to the column resin was through a linker arm at the 5′-position of the CCdA moiety as shown (see Materials and Methods).

Selection/amplification methods (Tuerk & Gold, 1990; Ciesiolka et al., 1996) were used to enrich RNAs with high affinity for CCdApPuro from a randomized pool. CCdAp-Puro-binding RNAs were enriched from a pool of 10^{14} distinct sequences with 50 contiguous randomized nucleotides, predicted to sample all motifs of 25 contiguous positions, as well as noncontiguous motifs with similar probability (Ciesiolka et al., 1996). Pools with similar

† Research grant support was provided by the NIH (Grants GM30881 and GM48080 to M.Y.).

* Author to whom all correspondence should be addressed.

⊗ Abstract published in *Advance ACS Abstracts,* May 15, 1997.

S0006-2960(96)03135-2 CCC: $14.00

complexity have yielded diverse RNA sites, including those for amino acid binding (Connell et al., 1993; Connell & Yarus, 1994; Famulok et al., 1994; Majerfeld & Yarus, 1994), nucleotide binding (Connell & Yarus, 1994; Sassanfar & Szostak, 1993), divalent cation binding (Ciesiolka et al., 1995), and catalysis of RNA aminoacylation (Illangasekare et al., 1995). Among RNAs capable of binding CCdApPuro, we hoped for similarities to 23S rRNA that would address its role at the peptidyl transferase center.

MATERIALS AND METHODS

Preparation of the Selection Matrix. $H_2N(CH_2)_6PO_3$-O-CCdApPuro (LCCdApPuro) was synthesized from $H_2N(CH_2)_6$-PO_3-O-CCdAp (Macromolecular Resources) and puromycin and purified by C18 reverse-phase chromatography (retention time of 24 min in a gradient of 0 to 60% acetonitrile in 100 mM NH_4OAc at pH 6.5 over 50 min) as described for CCdApPuro (Welch et al., 1995). The product was confirmed by UV spectroscopy (its spectrum was similar to that of CCdApPuro) and electrospray mass spectrometry. Protection of the terminal amino of the linker during coupling was not required in the presence of a large excess of puromycin. LCCdApPuro was observed to be stable under the conditions of experiments and storage used (determined by HPLC analysis).

LCCdApPuro was coupled to *N*-hydroxysuccinimide-activated CH-Sepharose 4B (Pharmacia) for 4 h at 4 °C in 100 mM HEPES (pH 8.0) and 500 mM NaCl. Activated hydroxysuccinimide esters remaining after the coupling step were reacted with ethanolamine (1 M ethanolamine at pH 8.0 for 1 h at 4 °C). The extent of coupling was determined in two ways. (1) The fraction of input LCCdApPuro not retained in the matrix after extensive washing was measured by HPLC analysis (C18) of the washes. (2) The acid-dependent release from the matrix of puromycin (from dissociation of the phosphoramide) and adenine (from depurination of the dA residue) was measured by incubating a known amount of matrix at 37 °C and pH 1 for 3 h and quantitating the released species using analytical reverse-phase HPLC (C18). Both methods suggested nearly quantitative coupling. The latter method was also used to show that the matrix was stable under storage conditions [30 mM HEPES (pH 7.5) 2 mM EDTA, and 30% methanol at −10 °C]. Counterselection matrix was prepared in an identical fashion except that the initial incubation lacked LCCdAp-Puro.

Selection Procedure. Selections were performed at 24 °C on 100 μL columns of LCCdApPuro-linked matrix at a flow rate of ~0.5 bed volume per minute. Selection buffer was 250 mM KCl, 10 mM $MgCl_2$, 30 mM HEPES (pH 7.0), and $ZnCl_2$, $MnCl_2$, and $CaCl_2$ all at 100 μM. The conditions used for selection were chosen because (1) they showed minimal retention of initial random RNA on the column, allowing good enrichment of specifically binding RNAs, (2) they are conditions under which the matrix-linked ligand is expected to be stable, and (3) they are conditions under which ribosomal peptidyl transferase is active. The metals Zn^{2+}, Mn^{2+}, and Ca^{2+} were included because these divalents have proven to increase the variety and potential of RNA sites in previous selections (Illangasekare et al., 1995).

Prior to selection, internally ^{32}P-labeled RNA (1100 pmol in the first selection round, 300−500 pmol in subsequent rounds) was renatured; RNA was heated in distilled H_2O to 65 °C for 5 min. The solution was adjusted to selection conditions without $ZnCl_2$, $MnCl_2$, and $CaCl_2$, and heating was continued for 1 min. $ZnCl_2$, $MnCl_2$, and $CaCl_2$ were then added, and the mixture (total volume of 30 μL) was cooled to 24 °C over 10 min. After loading of the RNA, the column was washed with 1.5−2 mL of selection buffer and then 300 μL of selection buffer containing 15 μM CCdApPuro. RNA eluted by CCdApPuro was ethanol precipitated and converted to cDNA by reverse transcription, and the cDNA was amplified by PCR and transcribed into RNA as described (Ciesiolka et al., 1995).

Counterselections using columns of ethanolamine-blocked resin (see above) were performed prior to and in a manner similar to that of selection columns. The first 67% of the initial peak of RNA was collected, ethanol precipitated, resuspended in water, and renatured for selection as described above.

At the completion of selection, the RNA pool was converted to cDNA by reverse transcription, amplified by PCR, and cloned using the plasmid pUC19 (United States Biochemical) as described by Ciesiolka et al. (1995).

PCR mutagenesis was performed essentially as described by Bartel and Szostak (1993). For the first three rounds of the reselection with mutagenesis, the pool DNA was subjected to error-prone PCR with *Taq* DNA polymerase with 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 0.2 mM dATP and dGTP, 1.0 mM dCTP and dTTP, 0.5 mM $MnCl_2$, and 7 mM $MgCl_2$. The mutation rate of the selection isolate FA-1 amplified under these conditions was measured as approximately 0.06% per position per PCR doubling, similar to that previously reported (Bartel & Szostak, 1993). Each round, DNA subjected to 12−50 doublings (approximately 0.7−3% mutagenesis) was combined and used as transcription template for the next round RNA pool.

For rounds 6Mt−9Mt of the reselection with mutagenisis, specific elution with 15 μM CCdApPuro was preceded by a 400 μL wash containing 1 mM puromycin and 100 μM CCdA. In rounds 10Mt and 11Mt, puromycin (240 and 270 μM in rounds 10Mt and 11Mt, respectively) and CCdA (21 and 40 μM) were included in the first 1.5 mL of column wash before the CCdApPuro elution.

Colony Hybridization Screening. F-type sequences in the reselected pool were identified by screening of colony or streak replicas on a nitrocellulose filter with ^{32}P-labeled 16-nucleotide DNA probes. The two probes used were complementary to either bases 48−63 (probe A) or 31−46 (probe B) of the isolated sequence FA-1. Both probes (Macromolecular Resources) contained inosine at 15% per position to allow hybridization to sequences with limited variation with respect to FA-1 (Takahashi et al., 1985). Filters were prehybridized in 1 M NaCl, 1% SDS, 10% dextran sulfate, and 100 μg/mL sonicated denatured salmon sperm DNA for 60 min at 37 °C. One picomole of 5′-^{32}P-end-labeled probe was added per 10 mL of prehybridization solution, and hybridization was carried out overnight over a descending temperature range of 50−25 °C. The filters were washed five times with 6× SSC at 25 °C and once at 34 °C. Filters to be hybridized with probe B were stripped of probe A by incubation in 0.1× SSC and 0.1% SDS at 85 °C for 15 min. Hybridization was then carried out as above.

Lead Hydrolysis Assay. Pb^{2+} probing was performed by incubating 5′-^{32}P-labeled RNA (60 μg/mL total RNA in 10 μL reaction mixtures; renatured as for selection) with 0, 0.25, or 1 mM $Pb(OAc)_2$ under selection conditions for 35 min.

Reactions were quenched with the addition of 7.5 μL of gel loading buffer containing 20 mM EDTA, 0.5× TBE (pH 8.3), and 7 M urea. The lengths of the hydrolysis products were analyzed by denaturing polyacrylamide gel electrophoresis (10% polyacrylamide, 7 M urea, and 1× TBE at pH 8.3).

DMS Modification Assay. DMS probing experiments were performed as described by Krol and Carbon (1989). After renaturing (as for selection), 0.3 pmol of RNA was incubated with or without ligand (CCdA or CCdApPuro) under selection conditions for 15 min in a total volume of 95 μL. Modification was initiated with the addition of 5 μL of an ethanolic solution of DMS (final concentration was 44 μM) and proceeded for 20 min at 24 °C. Reactions were stopped by addition of 10 μL of 3 M NaOAc at pH 5.2 and immediate chilling to 4 °C. The RNA was twice precipitated with ethanol and resuspended in 10 μL of water.

Modification was assayed by primer extention with AMV reverse transcriptase from a synthetic DNA primer hybridizing to bases 79−95 of the RNAs. Hybridization reaction mixtures contained 0.14 pmol of RNA and 0.07 pmol of 5′-^{32}P-labeled primer in 6 μL of water and were heated to 90 °C for 1 min, cooled at 24 °C for 30 s, and placed on ice. Extensions were initiated with the addition of 4 μL of a buffer/enzyme mixture and continued for 25 min at 42 °C. Final reaction conditions were 50 mM Tris-HCl (pH 8.0 at 42 °C), 7 mM $MgCl_2$, 50 mM KCl, 1.2 units of AMV reverse transcriptase (Life Sciences, Inc.), and dATP, dCTP, dGTP, and dTTP (170 μM each). Reactions were stopped with the addition of 15 μL of 10 M urea and 50 mM EDTA.

Primer extension products were analyzed by polyacrylamide gel electrophoresis (10% polyacrylamide and 8 M urea). Equal counts per minute of reactions were loaded per gel lane. Radioactivity in bands was quantitated using a phosphorimager (Bio-Rad GS-525 Molecular Imager System). K_d's for ligand binding were determined from the relationship of modification band intensity and ligand concentration as described previously (Welch et al., 1995; Ciesiolka & Yarus, 1996) using the equation

$$I = I_{sat} + I_0/(1 + A/K_d^A)$$

where I_{sat} is the intensity at saturating ligand (>0 due to partial inactivity of the RNAs and/or incomplete protection upon ligand binding) and I_0 is the difference in intensity with respect to I_{sat} in the absence of ligand. Radioactivities of bands were normalized to those of lane control bands.

RESULTS

Selection for RNAs That Bind CCdApPuro. The initial RNA pool for selection was generated by published methods (Ciesiolka et al., 1996). Approximately 10^{14} distinct sequences of chemically synthesized single-stranded DNA, comprised of two specific-sequence regions flanking 50 randomized positions, were converted to double-stranded DNA by PCR (Ciesiolka et al., 1996). The resulting DNA contained an optimal promoter for T7 RNA polymerase (Milligan & Uhlenbeck, 1989) and restriction sites for cloning. Transcription yields RNA 95 nucleotides in length. The first 20 and last 25 positions were specified sequences so that synthetic DNA oligonucleotides could prime reverse transcription of the RNA and PCR amplification.

Enrichment of CCdApPuro binding RNAs was accomplished by affinity chromatography on a 100 μL column of beaded agarose attached to CCdApPuro (bed concentration of 15−20 μM) through a linker at the 5′-OH (Figure 1). Selection was comprised of passing the RNA pool over the column, washing with 15−20 bed volumes of selection buffer, and then specifically eluting with 3 bed volumes of 15 μM free CCdApPuro in selection buffer. Specific eluant (CCdApPuro) was used at column ligand concentrations, rather than higher, to hinder selection of RNAs that prefer the column-bound form. From the third round on, selection was preceded by a counterselection on a column lacking CCdApPuro to deplete RNAs that bind nonspecifically. After each selection column, specifically eluted RNA was amplified by previously published methods (Ciesiolka et al., 1996).

Table 1: Evolution of Selection Column Profiles

round	% retained[a]	% specifically eluted[b]
1	0.4	0.03
2	0.6	0.1
3[c]	3	0.2
4[c]	4	0.3
5[c]	5	0.3
6[c]	10	0.9
7[c]	9	1.2
8[c]	23	7.9
9[c]	35	15.8
10[c,d]	36	puro, <0.5 CCdA, 19.1 CCdApPuro, 6.5

[a] Percent retained is that of total applied RNA remaining on the column after 6 bed volumes (600 μL) of selection buffer wash. [b] Percent specifically eluted is that of total RNA recovered in a 300 μL affinity elution with CCdApPuro. [c] Selection was preceded by counterselection with resin lacking CCdApPuro (see the text). [d] Affinity elutions with puromycin (puro), CCdA, and CCdApPuro were performed sequentially (see the text).

The elution volume of an RNA as a function of the K_d for interaction with the column ligand can be approximated (Connell et al., 1993) according to $V_e = V_n([L] + K_d)/K_d$, where V_e is the median elution volume, [L] is the concentration of column ligand, and V_n is the expected median elution volume of RNA that does not interact with the matrix, which was estimated to be about 0.8 bed volume by the elution point of the initial randomized RNA. From the above equation, it is determined that the prewash before specific elution is sufficient to remove RNAs with K_d's greater than approximately 1 μM.

The progress of the selection is summarized in Table 1. In the initial round of selection, 0.03% of the RNA was collected by affinity elution. RNA retained by the column and specifically eluted increased significantly from rounds 6 through 9. At round 9, 35% of the RNA was retained after 6 bed volumes of selection buffer wash and 16% was collected in the specific elution. A comparison of the first and ninth round column profiles is shown in Figure 2. The increased retention of the RNA was absolutely dependent on the CCdApPuro ligand. There was no increase in retention by the control column (see Materials and Methods) over the course of the selection.

RNAs that interact with both the CCdA and puromycin parts of CCdApPuro (both peptidyl transferase substrate representatives) simultaneously were especially desired. To analyze the contributions of the two parts and to enrich the RNA pool for molecules that prefer CCdApPuro to either half alone, the tenth selection-round specific elution was preceded by successive 4 bed volume washes of 450 μM puromycin and 100 μM CCdA (Table 1). No significant

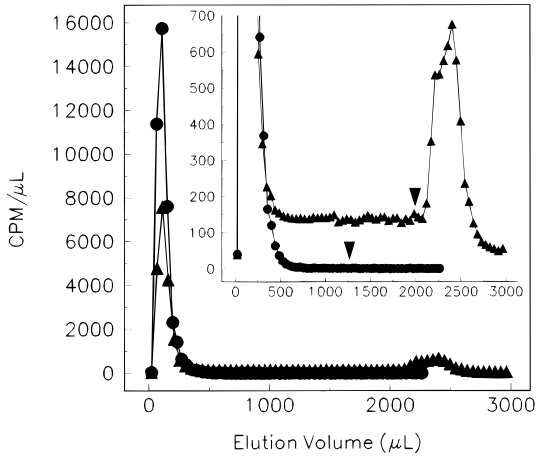

FIGURE 2: Selection column profiles. Comparison of the initial (circles) and ninth-round (triangles) selection column profiles. (Inset) Expansion of the profiles. Arrowheads indicate the start of the elution with CCdApPuro.

response to the puromycin wash was seen. Much of the bound RNA, on the other hand, was eluted by the CCdA wash. Thus, the selected RNA population preferred CCdA over puromycin.

Randomness of the RNA pool through the selection was monitored by gel electrophoresis of the products of limited RNase T1 digestion (Ciesiolka et al., 1995). A distinct pattern in the T1 products (i.e., in the distribution of G's) emerged at round 7 and remained in subsequent rounds (data not shown). The extent of nonrandomness in the T1 pattern at the completion of selection suggested there were only a few predominant sequences in the RNA pool.

Sequence Analysis of the Selected RNAs. The RNAs specifically eluted in the tenth round of selection were converted to cDNA, cloned, and sequenced (Ciesiolka et al., 1996). The set of sequences obtained (Figure 3) confirmed the low level of variation implied by the T1 assay. All of the 47 sequenced clones were classified as belonging to either of two families (FA and FB). The sequence variability within each family was so small that each likely derived from one ancestral RNA molecule in the initial pool. The limited observed familial variation is likely the result of adventitious mutagenesis during the amplification steps. Family FA was dominated by its likely parental sequence, denoted FA-1 (27 of 42 clones).

The two families showed considerable similarity. In particular, they share 21 bases, including 16 contiguous positions, in a stretch of 24 initially randomized nucleotides which are unchanged among all of the clones. The likelihood of selecting two independent clones with such similarity by chance is very low (probability $\approx 1.5 \times 10^{-7}$). It is possible that the extensive contiguous-base stretch in one family is derived from the other through a recombination event during the selection amplification reactions. Whatever the origin of the similarity, its unvaried occurrence in two different selected sequence contexts implies that it contains elements important for selection. This is further supported by the fact that none of the 27 positional variations (including deletions) observed in the 50-base variable region were in the 21-base identity, although it is not known that mutation during the selection was random in distribution; most of the observed variations are at the 3′-end of the randomized region. The occurrence of only two families (i.e., two initial parental RNAs) is consistent with the fact that the size of the region of similarity approaches the maximum size of a specific contiguous sequence expected to be sampled in the initial pool (Ciesiolka et al., 1996).

Reselection with Mutagenesis. Additional support for the requirement of the sequence elements conserved between the two families comes from the results of a second selection for CCdApPuro binding, which incorporated PCR mutagenesis in the amplification steps to increase pool variability. Reselection began with the RNA pool from the fifth round of the initial selection. The fifth-round pool was minimally retained by the selection resin and gave a homogeneous T1 RNase digestion ladder and, therefore, was still diverse. The double-stranded DNA T7 transcription template generated from this RNA was mutagenized by PCR with *Taq* DNA polymerase under conditions that are known to cause increased base substitutions (as well as insertions and deletions) that are nearly random in type and distribution (Cadwell & Joyce, 1992; Bartel & Szostak, 1993). Before each of the first three rounds of the reselection (rounds 6Mt−8Mt), the sequence pool was mutagenized to a degree of 0.7−3% substitution per base. The final three rounds did not include mutagenesis. The selection chromatography was performed essentially as described for the initial selection except that all rounds (six in total) included prewashing with puromycin and CCdA to enrich molecules which prefer complete CCdApPuro over either of its parts.

After round 9, the RNase T1 digestion ladder of the RNA pool became distinctly nonrandom (not shown). In round 11Mt, 37% of the RNA was retained after 6 bed volumes of wash and 11% was specifically eluted by CCdApPuro. The specifically eluted RNAs from this round were converted to cDNA, cloned, and sequenced.

The majority of the sequences obtained from the reselection were distinct from the FA and FB types from the initial selection. This is likely attributable either to the more stringent counterselection with puromycin and CCdA or to the added pool variability due to PCR mutagenesis, or both. Of the 58 clones initially sequenced, 3 were of the FA type (FMA-1, FMA-2, and FMA-3), 1 was of a hybrid FA/FB type (FMA/B-1), likely derived from recombination during PCR or reverse transcription, 48 were of a single new type (family M), and 6 were unique. The family M and unique RNAs are currently being characterized and are outside the focus of the discussion here. For reasons discussed below, we were especially interested in the reselected mutagenized FA- and FB-type sequences.

To increase the collection of the F-type variants for adequate phylogenetic sequence comparison, a colony hybridization method was devised to identify additional clones of this type. The method was designed to effectively identify such clones without significant bias against variant sequences, utilizing oligo−DNA probes partially substituted with inosine to allow hybrization to sequences with limited variation (Takahashi et al., 1985).

Two different 16-nucleotide probes hybridizing to FA bases 48−63 (probe A, corresponding to the FA/FB 16-base identity) or 31−46 (probe B) were used for screening. Both probes contained inosine at 15% per position (a mean of 2.4 inosines per probe). Probe A was designed to identify clones containing variations of the 16-contiguous-base sequence common to both selected F-type families. Of the 750 clones screened with this probe, 29 showed hybridization clearly above the background level; all were F-type, specifically of the FA family (Figure 3B). The frequency of F-type

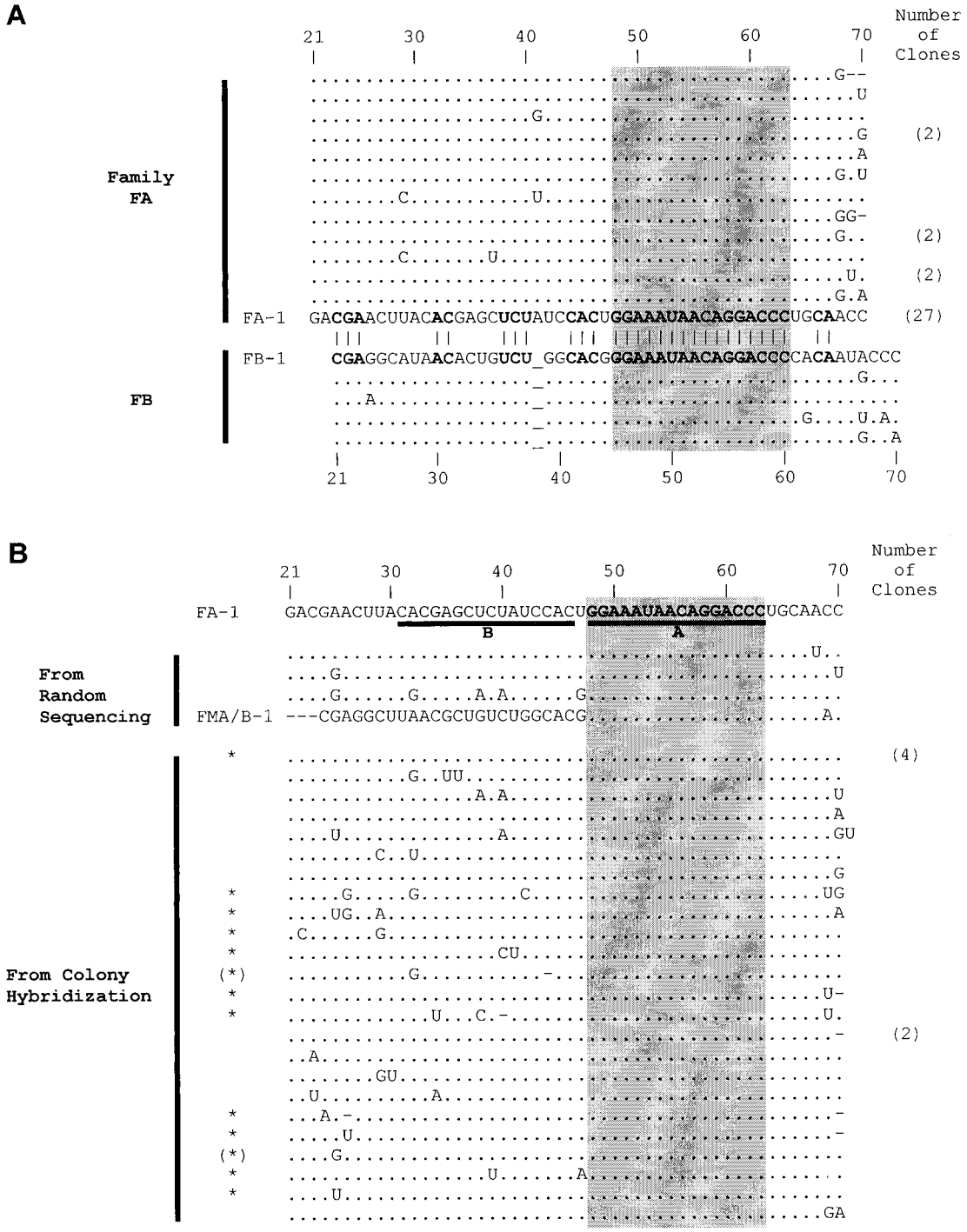

FIGURE 3: Alignment of the selected seqences. (A) Sequences derived from the variable regions obtained from the initial selection. Dots represent positions of identity to the family representative sequence displayed (FA-1 or FB-1). Dashes (−) represent base deletions. Underscores (_) mark gaps in the alignment. Bases of identity between the families are bold and marked by lines. (B) Sequences of the F type obtained from the reselection with mutagenesis compared to sequence FA-1. Clones were identified by either random sequencing or colony hybridization, as indicated (see the text). Lines labeled A and B indicate the sequences complementary to probes A and B, respectively. All sequences from colony hybridization were initially identified by probe A. Sequences from the 280 clones that were screened with probe B are marked by asterisks. Parentheses around asterisks indicate sequences that were not clearly identified in the probe B screening. Other symbols are as in panel A.

clones identified by colony hybridization (28/750 or 3.7%) is statistically consistent with that found in the pool by undirected sequencing (4/58 or 6.9%).

None of the sequences identified by probe A showed any variation in the 16-base identity, suggesting that even single-base changes in this sequence are deleterious for selection. To be sure that the hybridization method truly allowed discovery of sequences which are not perfectly matched to the probe, probe B, which hybridizes to a sequence specifically found in FA RNAs that does not overlap the 16-base identity, was used to screen 280 of the 750 colonies screened by probe A. Probe B clearly identified 13 of the 15 F-type clones identified by probe A among these colonies, and no additional F-type clones were revealed. These sequences had up to three variations in the hybridization region of probe B, and there was no apparent bias in favor of sequences with fewer changes. Thus, the inosine-containing probe A would likely have allowed detection of changes in the 16-base identity. The fact that most of the F-type clones among the 750 colonies screened were likely identified is supported by

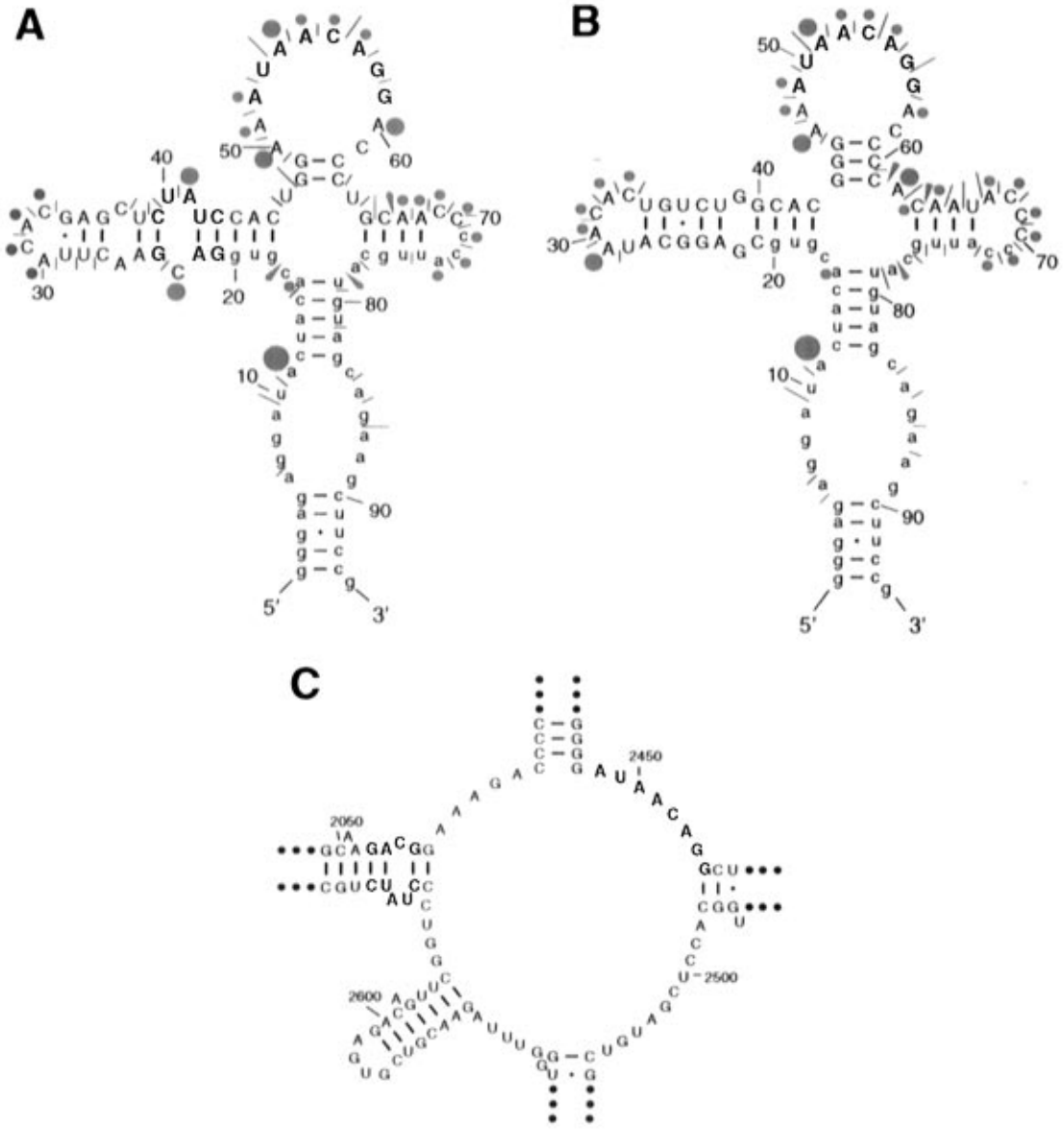

FIGURE 4: Structural probing data superimposed on computer-predicted secondary structures. Optimal secondary structures predicted by *Mfold* (Jaeger et al., 1990) for clones FA-1 (A) and FB-1 (B). Weak, moderate, and strong cleavages by Pb^{2+} are marked by green small dashes, long dashes, and dark arrows, respectively. Weak, moderate, and strong reactivities to DMS (small, medium, and large green circles) are also shown. DMS modification was assayed for bases A4−A73. Bases within observed homology to 23S rRNA are bold and colored red. (C) The peptidyl transferase loop region of 23S rRNA. Bases within homology to the selected clones are bold and colored red.

the fact that each of the two nonoverlapping probes recognized most or all of the clones identified by the other among the 280 colonies screened by both.

In the 32 F-type sequences obtained from the reselection, a total of 61 total positional variations (including deletions and insertions) with respect to FA-1 were observed. This is 3.8% mutagenesis overall, consistent with the performance of the PCR mutagenesis method on virtually unconstrained sequence (see Materials and Methods). None of these changes occurred in a contiguous stretch of 20 bases (positions 48−67) including the 16-base identity, where 21 changes would have been expected if the sequence were under no constraint. This lack of any variation in the these bases is extremely improbable by chance and strongly implicates this region as crucial for CCdApPuro binding.

Structures of the Selected RNAs. The optimal secondary structures predicted by *Mfold* (Jaeger et al., 1990) for representative sequences of families FA and FB are shown in Figure 4. These foldings are supported by a comparative argument; all of the observed variants had optimal (almost all of the RNAs) or near-optimal foldings clearly similar to those shown, and none of the other predicted foldings were common to the two families.

Secondary structures of the representative RNAs were also probed by modification with dimethyl sulfate (DMS) and hydrolysis with Pb^{2+} (Figures 4 and 5). DMS modification at N-1 of A's and N-3 of C's is specifically monitored by primer extension with reverse transcriptase (Krol & Carbon, 1989). Base reactivity indicates that the Watson−Crick face is unpaired and accessible to the modifying agent in solution. Hydrolysis of RNA in the presence of Pb^{2+} has been shown to preferentially occur in single-stranded or flexible regions or near divalent metal ion binding sites (Ciesiolka et al., 1992). Normal A-form helical regions are resistant because a phosphodiester is not aligned for in-line cleavage by the 2′-OH (Ciesiolka et al., 1992).

The DMS modification and Pb^{2+} cleavage patterns for FA-1 and FB-1 support most of the predicted most-stable secondary structure (Figure 4). One helix (bases 65−68 and 74−77 of FA-1 and 63−66 and 73−76 of FB-1) is the

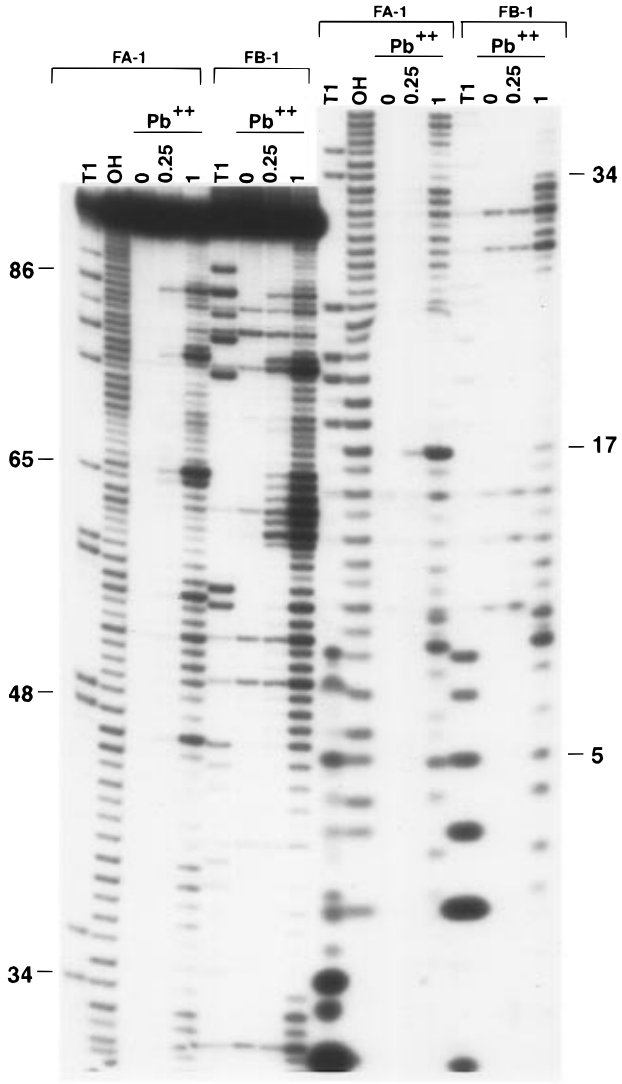

FIGURE 5: Structure probing of selected RNAs by limited hydrolysis with Pb^{2+}. The 5′-^{32}P-labeled RNAs FA-1 and FB-1 were hydrolyzed in the presence of 0, 0.25, or 1 mM Pb^{2+}, as indicated, for 35 min under selection conditions. The number labels at the edges indicate electrophoretic mobilities of the oligonucleotide sizes designated. Other labels are as follows: T1, limited digestion with RNase T1; and OH, limited base hydrolysis.

exception. For both, the 5′-strand of the predicted helix shows significant reactivities to both probes, suggesting that this region is flexible or is not canonical A-form helix. The chemical probing data did not support any of the other computer-predicted foldings.

DMS Footprinting Experiments. DMS footprinting was used to locate nucleotides affected by CCdApPuro binding to RNAs FA-1 and FB-1 (Figures 6 and 7). Altered reactivities toward DMS upon binding were predominantly found in the large conserved loop. Interestingly, A54 (RNA FA-1) was protected strongly by CCdApPuro (25% reactivity at saturating CCdApPuro), but not at all by CCdA (Figures 6 and 7). Therefore, A54 may represent a region of interaction or alteration upon binding of the puromycin moiety of CCdApPuro. Other protections observed were seen with binding of either ligand but tended to be more complete with CCdApPuro binding. The two stimulations observed at A11 and A60 were similar in magnitude for saturating binding of either ligand.

The protection/stimulation pattern observed upon CCdApPuro binding to RNA FB-1 was similar to that shown in Figure 7 for FA-1 (within the between-family similarity and at A11 in the 5′-constant sequence) except that additional protections were seen at A16, C63, and A64 (Figure 6). Thus, the two families appear to interact with CCdApPuro in a similar ways, both involving the conserved 16-base identity.

Dissociation Constants for Binding of CCdA and CCdApPuro. Dissociation constants (K_d's) for ligand binding can be estimated by the dependence of altered chemical reactivities on ligand concentration (Welch et al., 1995; Figure 7B,C). For both FA-1 and FB-1, K_d's were measured for all positions where significant effects were observed. For the particular ligand and RNA used, K_d's determined at all positions were similar, suggesting that all are due to a single binding event. The mean K_d's determined for FA-1 binding to CCdA and CCdApPuro were 170 and 8 nM, respectively (14 nM for CCdApPuro and FB-1). As suggested for the pool by the tenth-round column (see above), there is a clear preference for the CCdA part of the molecule, but both halves of CCdApPuro contribute; the addition of the phosphoropuromycin moiety to CCdA increases affinity approximately 21-fold.

DISCUSSION

Comparison with the Peptidyl Transferase Loop Domain of 23S rRNA. In the predicted structures of the FA and FB RNAs, 12 of the 16 contiguous bases in the between-family similarity constitute a large loop. Remarkably, eight bases of the loop match a highly conserved stretch of bases in the peptidyl transferase loop of 23S rRNA (Figure 4C). Bases 52−59 of FA-1 (FB-1 bases 49−56) are identical to eubacterial 23S rRNA bases 2448−2455 (*Escherichia coli* 23S rRNA numbering). Furthermore, the secondary structure of the eight-base identity is similar in the predicted structures of the selected clones and 23S rRNA. The first seven bases of the homology are predicted to be "single-stranded" in 23S rRNA as well as in the selected RNAs.

The selected eight-base sequence, AUAACAGG, is essentially completely conserved (>99%) among the known 23S rRNA sequences of eubacteria and chloroplasts and very highly conserved among all other phylogenetic domains (Gutell et al., 1993). Some variation in the 23S-like rRNAs of archaebacteria and mitochondria is observed in the eighth base, position 2455, which varies predominantly between A (60%) and G (40%) in archaebacteria or C (66%) and G (30%) in mitochondria (R. R. Gutell, personal communication). In the known rRNA sequences of eukaryotes, the eight-base sequence is completely conserved and identical to that selected except for the sixth base, position 2453, which is always a U. The very high conservation of this selected sequence suggests it has an essential role in peptidyl transferase function.

It should be noted that U2449 of *E. coli* 23S rRNA, within the eight-base homology, was recently found to be a dihydrouridine (Kowalak et al., 1995), a nonplanar modified U which resists base−base stacking (Dalluge et al., 1996). The degree of conservation of this modification of U, however, is not yet clear. It has not been found in any other 23S-like rRNAs analyzed, including that of another eubacterium (J. A. McCloskey, personal communication).

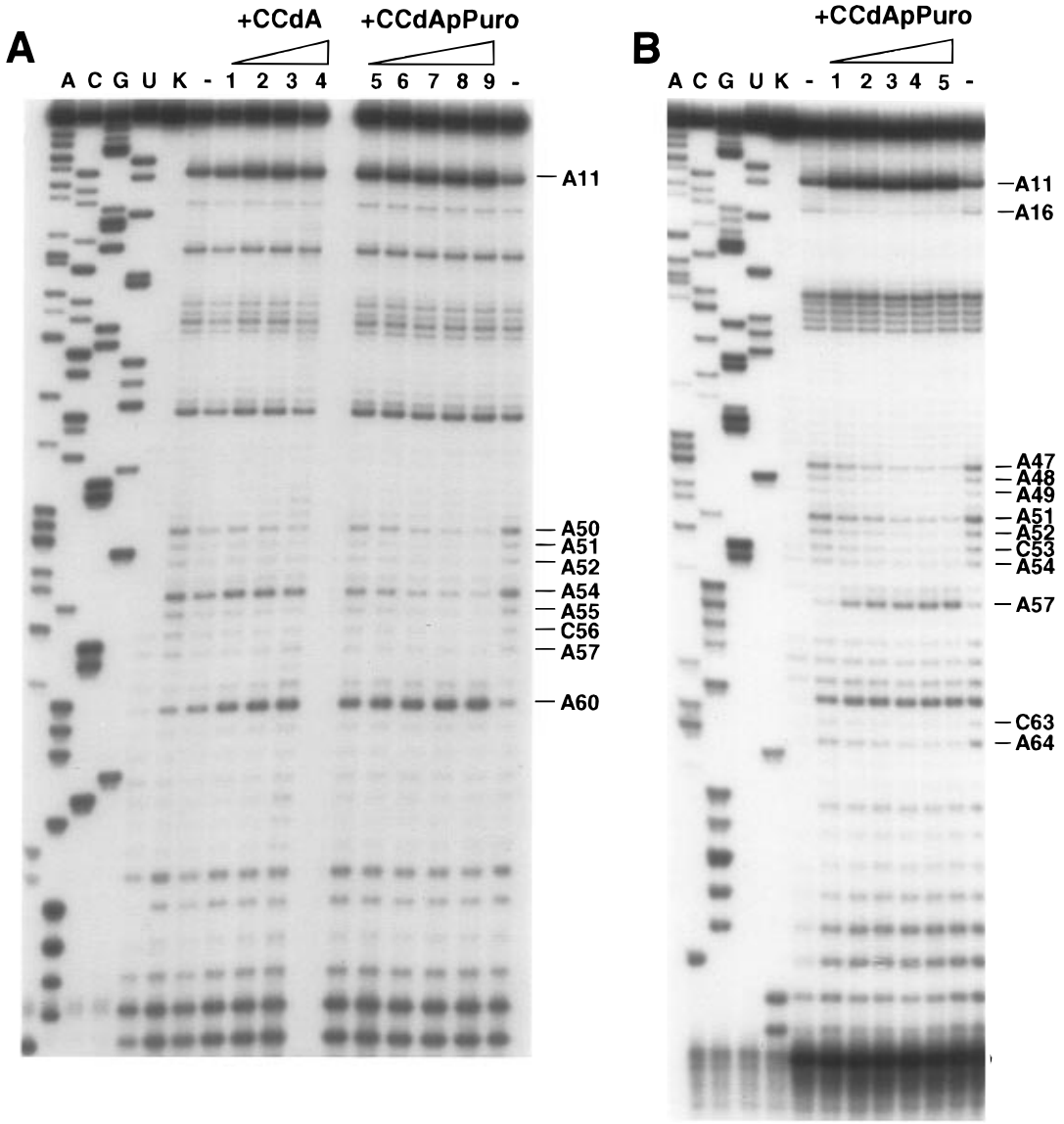

FIGURE 6: DMS modification of selected RNAs and footprinting by ligands. (A) Primer extension analysis of modification of FA-1 in the absence or presence of CCdApPuro or CCdA at varied concentrations. Lanes are labeled as follows. A, C, G, and U are dideoxy sequencing reactions. K marks control reactions lacking DMS. − marks modification reactions in the absence of ligand. Lanes 1−4 are reactions in the presence of 110, 210, 425, and 850 nM CCdA, respectively. Lanes 5−9 are reactions in the presence of 19, 38, 75, 150, and 300 nM CCdApPuro, respectively. Sites of altered reactivity upon ligand binding are indicated by arrows. The trend of protection at the weakly modified A57 with CCdA concentration is not clear due apparently to reverse transcriptase stuttering at the reactivity-stimulated base A60 in the experiment shown; however, we repeatedly observe depression of reactivity at A57 with added CCdA and believe this to be a true protection effect. (B) Modification of FB-1 in the presence or absence of CCdApPuro. Labeling is as above except that lanes 1−5 are reactions in the presence of 19, 38, 75, 150, and 300 nM CCdApPuro, respectively.

FA RNAs show a second homology to 23S rRNA at the peptidyl transferase center (Figure 4). A nine-base structure consisting of bases 21−24 paired with 39−43 is identical in sequence and secondary structure to the *E. coli* 23S rRNA pairing of bases 2053−2056 with 2612−2616. This structure is highly conserved among eubacteria, though more variable between phylogenetic domains (Gutell et al., 1993). However, an internal bulge structure is always present at this position of the 23S rRNA secondary structure. Several of our selected RNAs also show some variation in this structure; overall, changes were observed at seven of the nine positions. Bases G21 and C43 of the FA RNAs, which base pair in the structure, were invariant. These bases correspond in the structure similarity to 23S rRNA bases G2053 and C2616, which are both >95% conserved among all phylogenetic domains (Gutell et al., 1993).

Likelihood of the Selected Homologies to 23S rRNA. To address the statistical significance of observing the eight contiguous-base homology to 23S rRNA shared by both FA and FB RNAs, the probability of the sequence occurring within the 16-base between-family identity (the apparent required sequence for CCdApPuro binding) was calculated as follows. The probability of a specific eight-base sequence occurring in a randomized stretch of eight positions is 4^{-8}. There are nine contiguous eight-base segments in the 16-base between-family similarity; thus, the likelihood of selecting a specific 8-mer in the similarity is $9 \times 4^{-8} = 1.4 \times 10^{-4}$. There are roughly 30 contiguous 8-mers in 23S rRNA that have previously been associated with the peptidyl transferase center (only ten of these are completely within the peptidyl transferase loop itself, as is the eight-base homology). The probability that we would have found any of these is $(1.4 \times 10^{-4}) \times 30 = 4.2 \times 10^{-3}$. Another useful comparison shows that the observed similarity with the peptidyl transferase loop is the only contiguous eight-base identity in a comparison of all eight-base segments of *E.*

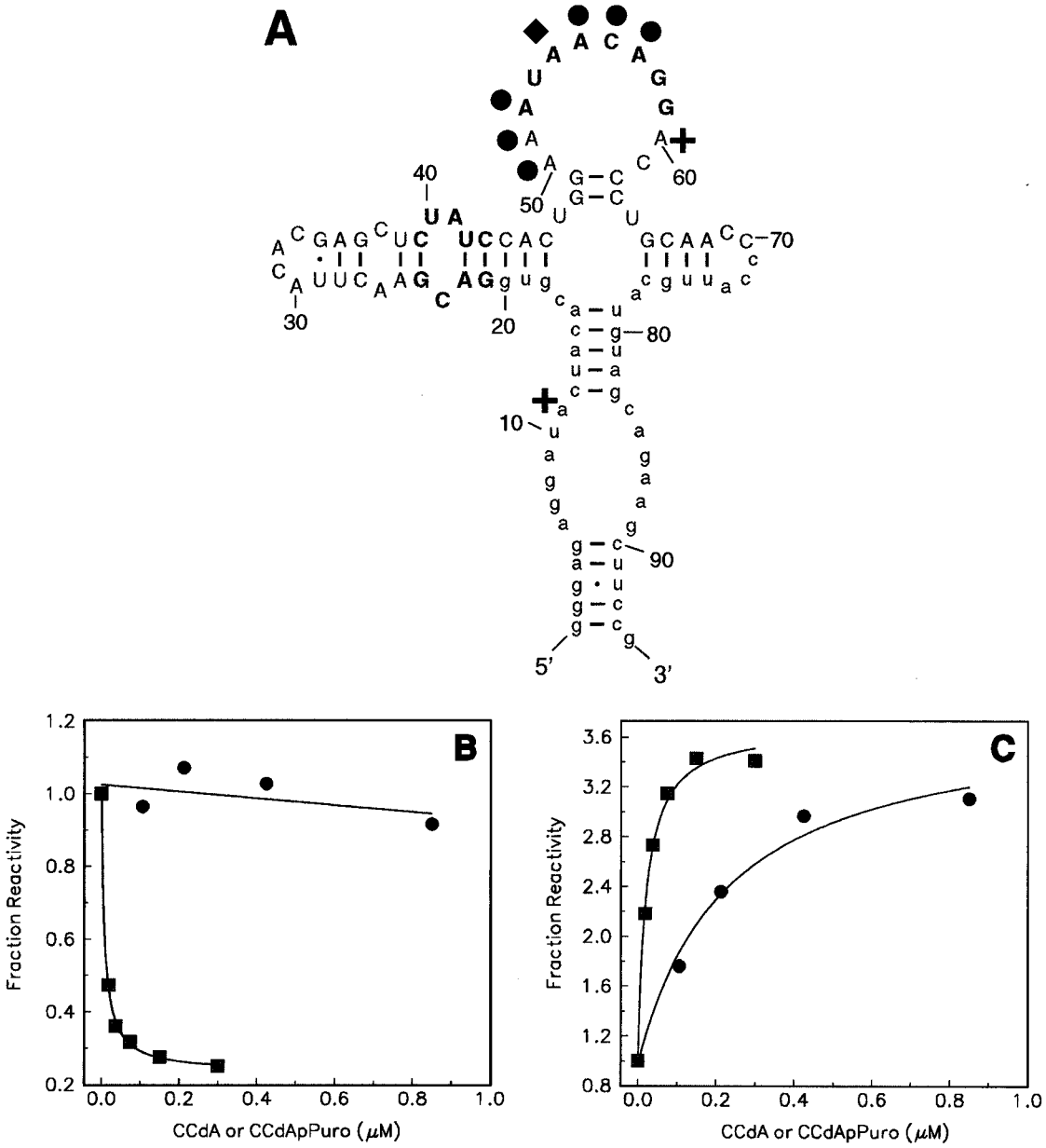

FIGURE 7: Altered reactivities of FA-1 bases toward DMS in the presence of CCdA and CCdApPuro. Results are derived from the data shown in Figure 6. (A) Sites of protection by both CCdA and CCdApPuro (circles) and CCdApPuro only (diamond) are shown. Pluses mark sites of stimulation of reactivity by both CCdA and CCdApPuro. Bases within homologies to 23S rRNA are bold. (B) Dependence of reactivity of bases on ligand concentration. Fraction reactivity at A54 of FA-1 in the presence of CCdA (circles) or CCdApPuro (triangles) relative to that in the absence of ligand. Curves drawn are least-squares fits of the function predicted by simple equilibrium assumptions (see the text). (C) Reactivity at A60 of FA-1. Symbols and curves are as in panel B.

coli 23S rRNA and the consensus sequences of FA or FB (a total of 2×10^5 comparisons). Clearly, such similarity is improbable and suggests structural and functional homology.

The probability of selecting the nine-base internal loop motif observed in family FA RNAs that shows similarity to 23S rRNA is somewhat elusive but can be assumed to be the likelihood of the secondary structure (i.e., three base pairs flanking a three-base asymmetric bulge) occurring and having a particular sequence. Considering only canonical pairing for simplicity, each of the three base pairs has four possibilities (total of 4^3 for the three pairs). There are 36 different bulges with one base opposite two in which Watson−Crick pairing cannot occur. Thus, the probability of the particular sequence is $(4^3 \times 36)^{-1} = 4.3 \times 10^{-4}$. This is taken as an upper estimate of the overall probability of the specific structure since the probability of the three-base pair/three-base bulge motif derived from the randomized region is predicted to be less than 1, on the basis of the observed frequency of occurrence of such a motif in several characterized RNAs from various selections in our lab. Inclusion of noncanonical pairing does not greatly alter the probability estimate, lowering it somewhat.

Implications for 23S rRNA at the Peptidyl Transferase Center. Sequence and structure similarity in the selected RNAs to highly conserved elements of 23S rRNA is intriguing. The eight-base homology in the large loop of the selected molecules is clearly involved in interaction with CCdApPuro. Conservation of these bases between the selected families implies that they are required for selection. Protection of this region from DMS methylation by CCdAp-Puro (all five modified bases of the eight-base similarity to 23S rRNA showed depressed reactivity) can be in part indirect, but the clustering of the effects within the loop suggests that at least some of these conserved nucleotides are directly involved in the binding site.

Several previous experiments have implicated the 23S rRNA bases of the selected similarity in peptidyl transferase function. In the chemical footprinting experiments of Moazed and Noller (1989, 1991), aminoacyl- or peptidyl-tRNA, binding to the A or P sites, respectively, of *E. coli* ribosomes, was shown to protect A2451 of 23S rRNA, within the eight-base selected similarity (A2451 corresponds to A55 of FA-1). This protection was dependent on the acyl group attached to the 3′-terminal CCA. Puromycin is a structural and functional analogue of the 3′-aminoacyl-A of aminoacyl-tRNA. Thus, the region around A2451 may be involved in interaction with puromycin or the 3′-aminoacyl-tRNA terminus. Mutations that confer resistance to and bases that are chemically protected by chloramphenicol and anisomycin are also found in this region (Douthwaite, 1992; Moazed & Noller, 1987; Rodriguez-Fonesca et al., 1995). Both antibiotics compete with puromycin binding at the A site of peptidyl transferase (Pestka et al., 1972; Kalpaxis & Coutsogeorgopoulos, 1989). The dependence of chemical protection at A54 of RNA FA-1 on the puromycin group of CCdApPuro is, therefore, consistent with the eight-base identity being at the puromycin site in 23S rRNA.

Because it is only observed in one of the selected families, the significance of the nine-base bulge structure homology (in FA RNAs) is less certain. Although this structure is apparently not an absolute requirement for binding, it could participate indirectly in the context of the FA RNAs and similarly in 23S rRNA. The primary structure of this motif is not completely conserved among the selected RNAs or all 23S-like RNAs, though the asymmetric bulge and one terminal G-C base pair (G21-C43 in the FA RNAs and G2053-C2616 in 23S rRNA) are widely conserved. Modified constructs of FA sequences lacking the bulge motif, but retaining the 16 conserved nucleotides of the loop, lose the ability to bind CCdApPuro (M. Welch and M. J. Yarus, unpublished), but more refined modifications will be required to assess its role.

Ribosomal peptidyl transferase binds to both CCdA and puromycin moieties of CCdApPuro, and there is evidence that interaction involves both A and P sites of the enzyme (Welch et al., 1995). These selected RNAs also appear to bind both halves of CCdApPuro. Furthermore, K_d's for FA and FB RNAs under the conditions of selection (24 °C) are similar to that observed for binding to the peptidyl transferase center (2−90 nM, depending on preincubation, at 4 °C in fragment assay buffer). Because of the tetrahedral, covalent linkage between the halves of CCdApPuro, the putative selected subsites are necessarily positioned so that the two potential reactants are well-placed for direct nucleophilic attack during peptidyl transfer. Because these sites apparently involve sequences similar to very highly conserved, and therefore ancient, nucleotides at the peptidyl transferase center, our result implies that these 23S rRNA nucleotides may participate in sites that bind and appose peptidyl transferase substrates. Furthermore, such binding can be carried out by a simple RNA domain functioning autonomously in the absence of protein, adding weight to the idea that peptidyl transferase could have arisen in an RNA world (Moore, 1993; Yarus, 1993). In future experiments, we hope to apply the structure and means of interaction of the selected motifs with CCdApPuro toward solution of the structure and function of 23S rRNA at the peptidyl transferase center.

ACKNOWLEDGMENT

We thank members of our laboratory for constructive comments on the data and the manuscript and especially S. Zinnen and J. Connolly for experimental support. We also thank R. Gutell for very helpful suggestions and assistance in construction of figures and the W. M. Keck Foundation for its support of RNA science on our campus.

REFERENCES

Bartel, D. P., & Szostak, J. W. (1993) *Science 261*, 1411.
Ciesiolka, J., & Yarus, M. (1996) *RNA 2*, 785.
Ciesiolka, J., Lorenz, S., & Erdmann, V. A. (1992) *Eur. J. Biochem. 204*, 575.
Ciesiolka, J., Gorski, J., & Yarus, M. (1995) *RNA 1*, 538.
Ciesiolka, J., Illagasekare, M., Majerfeld, I., Nickles, T., Welch, M., Yarus, M., & Zinnen, S. (1996) *Methods Enzymol. 267*, 315.
Connell, G. J., & Yarus, M. (1994) *Science 264*, 1137.
Connell, G. J., Illangasekare, M., & Yarus, M. (1993) *Biochemistry 32*, 5497.
Dalluge, J. J., Hashizume, T., Sopchik, A. E., McCloskey, J. A., & Davis, D. R. (1996) *Nucleic Acids Res. 24*, 1073.
Douthwaite, S. (1992) *J. Bacteriol. 174*, 1333.
Famulok, M. (1994) *J. Am. Chem. Soc. 116*, 1698.
Gutell, R. R., Gray, M. W., & Schnare, M. N. (1993) *Nucleic Acids Res. 21*, 3055.
Illangasekare, M., Sanchez, G., Nickles, T., & Yarus, M. (1995) *Science 267*, 643.
Jaeger, J. A., Turner, D. H., & Zuker, M. (1990) *Methods Enzymol. 183*, 281.
Kalpaxis, D. L., & Coutsogeorgopoulos, C. (1989) *Mol. Pharmacol. 36*, 615.
Kowalak, J. A., Bruenger, E., & McCloskey, J. A. (1995) *J. Biol. Chem. 270*, 17758.
Krol, A., & Carbon, P. (1989) *Methods Enzymol. 180*, 212.
Majerfeld, I., & Yarus, M. (1994) *Nat. Struct. Biol. 1*, 287.
Milligan, J. F., & Uhlenbeck, O. C. (1989) *Methods Enzymol. 180*, 51.
Moazed, D., & Noller, H. F. (1987) *Biochimie 69*, 879.
Moazed, D., & Noller, H. F. (1989) *Cell 57*, 585.
Moazed, D., & Noller, H. F. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 3725.
Moore, P. (1993) in *The RNA world* (Gesteland, R. F., & Atkins, J. F., Eds.) pp 119−135, Cold Spring Harbor Laboratory Press, Plainview, NY.
Noller, H. F. (1991) *Annu. Rev. Biochem. 53*, 119.
Noller, H. F., Hoffarth, V., & Zimniak, L. (1992) *Science 256*, 1416.
Pestka, S., Rosenfeld, H., Harris, R., & Hintikka, H. (1972) *J. Biol. Chem. 247*, 6895.
Rodriguez-Fonesca, C., Amils, R., & Garrett, R. A. (1995) *J. Mol. Biol. 247*, 224.
Samaha, R. R., Green, R., & Noller, H. F. (1995) *Nature 377*, 309.
Sassanfar, M., & Szostak, J. W. (1993) *Nature 364*, 550.
Takahashi, Y., Kato, K., Hayashizaki, Y., Wakabayashi, T., Ohtsuka, E., Matsuki, S., Ikehara, M., & Matsubara, K. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1931.
Tuerk, C., & Gold, L. (1990) *Science 249*, 505.
Welch, M., Chastang, J., & Yarus, M. (1995) *Biochemistry 34*, 385.
Yarus, M. (1993) in *The RNA World* (Gesteland, R. F., & Atkins, J. F., Eds.) pp 205−217, Cold Spring Harbor Laboratory Press, Plainview, NY.

BI963135J