Protein−RNA Sequence Covariation in a Ribosomal Protein−rRNA Complex†

Debraj GuhaThakurta and David E. Draper*

Department of Chemistry, Johns Hopkins University, Baltimore, Maryland 21218

Received November 5, 1998; Revised Manuscript Received January 20, 1999

ABSTRACT: Comparative sequence analysis has successfully predicted secondary structure and tertiary interactions in ribosomal and other RNAs. Experiments presented here ask whether the scope of comparative sequence-based predictions can be extended to specific interactions between proteins and RNA, using as a system the well-characterized C-terminal RNA binding domain of ribosomal protein L11 (L11-C76) and its 58 nucleotide binding region in 23S rRNA. The surface of L11-C76 α-helix 3 is known to contact RNA; position 69 in this helix is conserved as serine in most organisms but varies to asparagine (all plastids) or glutamine (*Mycoplasma*). RNA sequence substitutions unique to these groups of organisms occur at base pairs 1062/1076 or 1058/1080, respectively. The possibility that rRNA base pair substitutions compensate for variants in L11 α-helix 3 has been tested by measuring binding affinities between sets of protein and RNA sequence variants. Stability of the RNA tertiary structure, as measured by UV melting experiments, was unexpectedly affected by a 1062/1076 base pair substitution; additional mutations were required to restore a stably folded structure to this RNA. The results show that the asparagine variant of L11-C76 residue 69 has been compensated by substitution of a 1062/1076 base pair, and plausibly suggest a direct contact between the amino acid and base pair. For some of the protein and RNA mutations studied, changes in binding affinity probably reflect longer-range adjustments of the protein−RNA contact surface.

Searches for compensatory base changes in ribosomal RNAs (rRNAs) have been highly successful at predicting secondary structures, and a number of potential tertiary interactions have also been identified (*1*, *2*). Similarly, covariations between ribosomal proteins and rRNA might also be informative; some conserved ribosomal proteins have evolved with rRNA since early times, and mutations in protein residues contacting RNA may have been compensated by RNA mutations. It has been possible to select for Rev protein variants that compensate for deleterious mutations in the target RRE RNA in vivo, showing that protein−RNA covariation can take place in an artificial situation (*3*).

A good place to look for protein−RNA covariation in the ribosome are those proteins that are conserved across all phylogenetic domains, as they have been under selective pressure for the longest time and are associated with the largest sets of sequences. Here we examine the L11 family of ribosomal proteins, which is conserved in sequence and function among eubacteria, archaebacteria, and eukaryotes. The protein recognizes a small, conserved domain of large subunit rRNA (*4*), and bacterial L11 binds to homologous regions in eukaryotic (*5*) and archaebacterial rRNA (*6*). Replacement of L11 from one bacterial ribosome by that of another restores GTP hydrolysis in the presence of elongation factor G (EF-G) and full functional activities of the ribosome in protein synthesis in vitro (*7*), and the RNA binding domain can be swapped between *E. coli* and yeast ribosomes without any apparent loss of activity in vivo (*8*, *9*). Thus, the basic framework of L11 protein−RNA contacts has remained largely the same since early evolutionary times, though, as discussed below, some details of the recognition strategy used by different groups of organisms vary.

L11 is also one of the best understood rRNA binding proteins in structural terms. The minimum RNA fragment recognizing L11 is ∼58 nt (*10*), though protection studies suggest that only one hairpin of this structure actually contacts the protein (*11*, *12*) (Figure 1A). The rRNA binding domain of L11 is located solely within its 75 amino acid C-terminal domain (*13*), and the structure of this domain from *Bacillus stearothermophilus* (Bst),[1] termed L11-C76, has been determined by NMR (*14*, *15*) (Figure 1B). The domain shows surprising structural homology to the homeodomain class of eukaryotic DNA binding proteins and seems to employ a similar strategy for nucleic acid recognition (*14*). From NMR studies of the protein bound to its RNA target, it was found that the solvent-exposed surface of helix 3 makes extensive contacts with the RNA, and that loops 1 and 2 at either end of helix 3 "clamp" onto the RNA (*16*). Homeodomain proteins position helix 3 in the major groove of the DNA, and many of the homeodomain class of proteins contain extra loops ("wings") that wrap onto the DNA backbone (*17*).

Serine 69 of L11-C76 appears to contribute directly to RNA recognition. Position 69 at the end of helix 3 coincides with the universally conserved asparagine 51 of homeodomains, which is crucial for DNA binding (*14*). NMR studies of the L11-C76·RNA complex show Ser 69 is in close

† This work was supported by NIH Grant GM29048.

* To whom correspondence should be addressed. Email: draper@jhunix.hcf.jhu.edu. Telephone: 410 516 7448. Fax: 410 516 8420.

[1] Abbreviation: Bst, *Bacillus stearothermophilus*.

10.1021/bi9826411 CCC: $18.00
Published on Web 03/02/1999

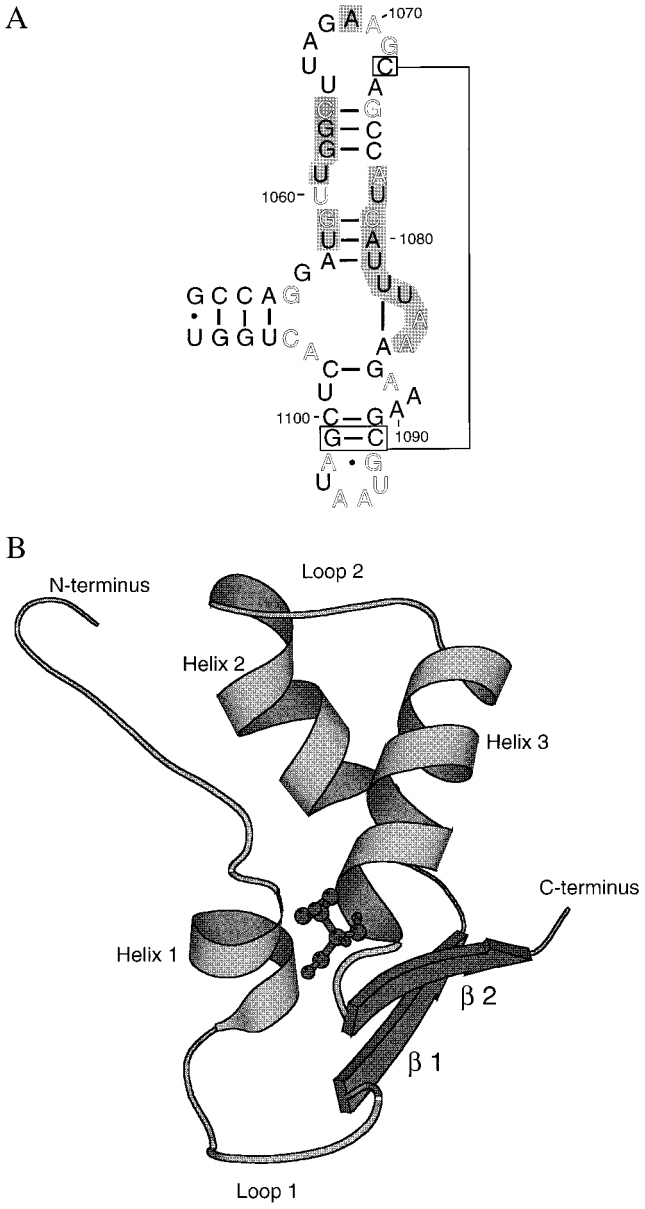

FIGURE 1: Structures of the components of the L11-C76−rRNA complex. (A) The minimal L11 binding site, *E. coli* 23S rRNA nucleotides 1051−1108. Bases in outline type font are nearly invariant among prokaryotic and eukaryotic rRNA sequences, and those which are shaded are protected from hydroxyl radicals by *E. coli* L11 in footprinting experiments (12). Triple base interaction [C1072(C1092/G1099)] is shown (*32*). (B) NMR-derived structure of Bst L11-C76 (14). The side chain of residue S69 is shown. The figure was generated using Molscript.

proximity to RNA (*16*), and its mutation to alanine weakens RNA binding by about an order of magnitude (*14*). It has, however, been substituted by asparagine in all known plastid sequences and by glutamine in a few other instances, notably *Mycoplasma* (see sequence alignments in Supporting Information). Plastid and *Mycoplasma* rRNAs have base substitutions unique to those groups of organisms and located on an irregular helix of 23S rRNA protected by L11 in hydroxyl radical footprinting experiments (Figure 1A) (*12*). These RNA substitutions thus correlate with protein Ser 69 variants and potentially lie within the protein binding site.

In the present work, we use site-directed mutagenesis to ask whether the covariations between L11-C76 residue 69 and rRNA bases in the protein footprint region reflect specific protein−RNA contacts. We find that the amino acid at position 69 does discriminate between different rRNA sequences. Covarying RNA bases may be in contact with this amino acid; however, RNA covariations may also be interpreted as maintaining a specific tertiary structure required for L11 recognition.

MATERIALS AND METHODS

RNA Preparation. The DNA sequence corresponding to the 58 nucleotide L11 binding fragment of 23S rRNA (1051−1108) was previously cloned in a derivative of pUC18 containing a T7 promoter and *Stu*I site (*18*). Mutations were made in this plasmid by the "unique site elimination" protocol (*19*) using a kit available from Pharmacia Biotech. Plasmids were linearized using *Rsa*I, transcribed with T7 RNA polymerase, and purified by denaturing gel electrophoresis followed by electroelution, as described (*18*). RNA fragments produced in this way were 59 nucleotides in length, and contained an extra 5′-G nucleotide.

For filter binding assays, ^{35}S-labeled RNA was prepared by transcription in the presence of [^{35}S]ATP. For some assays, ^{35}S-labeled RNA containing 23S rRNA nucleotides 1029−1126 was transcribed from plasmid pLL1 cleaved with *Bam*HI (*20*). The 1029−1126 RNA fragment containing the AA1089/90 → GU mutation was transcribed from a pLL1 derivative cut with *Eco*RV (*21*). All labeled RNAs were purified by NENsorb (DuPont) reverse-phase columns (*22*), ethanol-precipitated, and checked for purity on 8% denaturing polyacrylamide gels. Autoradiography showed a single major band for each RNA.

Preparation of L11-C76 and L11-C76 Mutants. A pET11a derivative overexpressing L11-C76 has been described (*13*). Mutagenesis of L11-C76 was done in this plasmid by the same "unique site elimination" protocol used to make rRNA mutants. Overexpression plasmids were transformed into *E. coli* BL21(DE3), and L11-C76 and its variants were purified as described (*13*). All proteins except wild-type L11-C76 and mutant R68K/S69Q were purified only through the gel filtration step. Mutant R68K/S69Q was passed through an anion exchange column (Biorad TSK-DEAE-5PW) to remove nucleic acid contaminants. The protein did not adhere to the column (20 mM phosphate buffer, pH 6.0), and the nucleic acids were eluted after the protein with buffer containing 1 M KCl. Wild-type L11-C76 was further purified using a cation exchange column (Biorad TSK-SP-5PW) as described (*13*).

Concentrations of wild-type L11-C76 and R68K/S69Q mutant were measured by absorbance at 230 nm using an extinction coefficient of 24.8×10^3 M^{-1} cm^{-1} (*13*). Concentrations of the other L11-C76 proteins were measured by Bradford assay (*23*), using a Bio-Rad assay kit. Known concentrations of wild-type L11-C76 (determined by 230 nm absorbance) were used to calibrate the Bradford assay.

RNA UV Melting. Unfolding of rRNA fragments was observed as a function of temperature by following hyperchromicity at 260 and 280 nm, using a modified spectrophotometer and procedures as described (*18*, *24*). RNA was renatured at 65 °C for 15 min in 10 mM MOPS (pH 7.0), 5 mM $MgCl_2$, and 175 mM monovalent salt (NaCl, KCl, or NH_4Cl). RNAs were melted from 10 to 90 °C at a heating rate of 0.8 °C per minute. Melting profiles are plots of the

```
      60  Helix 3  69
       |          |
EAAMRMIEGTARS MG   Eubacteria
TAAIKSLLGTARS IG   Archaea
ASVTKEILGTAQS VG   Eukaryote
D---NI------N --   Synechoscystis sp.
-S---I-A---AN --   Spinacia oleracia
SQ--KIVG---KN --   Porphyra purpurea
SS--KIV-----N --   Odontella sinensis
---S-IV-----N --   Cyanophora paradoxa
---LK-VL---KQ --   Mycoplasma pneumonae
---LK-VL---KQ --   Mycoplasma genitalium
```

FIGURE 2: Residues 57−71 of L11-C76 and its homologues; helix 3 extends from 57 to 69. Representative sequences from each phylogenetic class are shown: *Bacillus stearothermophilus* (eubacteria), *Sulfolobus acidocaldarius* (archaea), and *Saccharomyces cerevisiae* (eukaryote). Amino acids in dark shading are conserved in all of the available sequences within a phylogenetic domain, and those in light shading are highly conserved (at least 17 out of 23 available eubacterial sequences, 6 of 8 of eukaryotic sequences, or 6 out of 8 archaeal sequences). L11-C76 sequences of *Synechocystis*, all four chloroplasts, and two *Mycoplasma* are given separately. Amino acids identical to Bst sequence are indicated by dashes.

derivative of absorbance with respect to temperature, averaged over a 4° window and normalized to 10 °C (*18*).

Filter Binding Assays. For filter binding assays, RNA was renatured in the same way as described for melting experiments; 3000−5000 cpm of renatured RNA was used per assay; filtration and scintillation counting were as described (*14*, *22*). Association constants between RNA and thiostrepton were determined in 10 mM MOPS (pH 7.0), 3 mM $MgCl_2$, 175 mM NH_4Cl, and 5% v/v DMSO (to solubilize thiostrepton). Buffer for protein−RNA binding experiments was 10 mM MOPS (pH 7.0), 3 mM $MgCl_2$, and 175 mM KCl. The RNA−thiostrepton or RNA−protein mixtures were incubated for 15 min at room temperature (22−23 °C) prior to filtration through nitrocellulose filter membranes. Plots of RNA retained on filters against thiostrepton or protein concentration were fit to hyperbolic binding isotherms, letting the equilibrium constant, background retention, and maximum retention be variables in a least-squares fit using KaleidaGraph (Synergy Software) (*22*).

RESULTS

Potential rRNA Covariants Associated with Residue 69 of L11-C76. As mentioned in the introduction, serine 69 at the C-terminus of L11-C76 helix 3 appears to make specific contacts with the rRNA target: the serine methylene shows NOEs to the RNA in NMR experiments (*16*), and the binding affinity of the mutant S69A to the rRNA is decreased by about 10-fold as compared to wild-type protein (*14*). Position 69 is highly conserved as a serine among eubacteria (26 out of 29 sequences), archaebacteria (7 out of 8 sequences), and eukaryotes (10 out of 10 sequences). However, in all four chloroplast L11 sequences and in L11 from their close evolutionary relative, the cyanobacterium *Synechocystis* (*25*), asparagine replaces serine at the homologous position. In two other eubacteria (*Mycoplasma genetalium* and *Mycoplasma pneumoniae*) as well as in one archaea (*Acidianus ambivalens*), the residue at position 69 is replaced by a glutamine. A comparison of L11-C76 helix 3 sequences is shown in Figure 2.

In organisms for which L11 sequences are known, the rRNA sequences were searched in the 1057−1086 region, which is most likely to be contacting L11 (Figure 1A), to see if changes that uniquely correlate with protein position 69 variants could be detected (see Supporting Information). In *Synechocystis* and the four plastids (listed in Figure 2), rRNA sequences consistently contain two base pair changes compared to the *E. coli* sequence; viz., base pairs G1062/C1076 and U1058/A1080 are replaced by U1062/A1076 and G1058/C1080 (Figure 3A). (Among all chloroplast 23S rRNA sequences, a U1062·C1076 mismatch is most frequently found.) G1058/C1080 is common in eubacterial and archaebacterial sequences as well and thus a less likely candidate for covariation with Ser 69. A U·G mismatch at position 1058/1080 occurs in *M. genetalium* and *M. pneumoniae* (Figure 3B), the two organisms in which L11 sequences show the unique S69Q change. Thus, S69 → N correlates uniquely with U1062/A1076 (plastids and cyanobacteria), and S69 → Q is found only in conjunction with G1080 (some *Mycoplasma*). Our initial strategy was to prepare these substitutions in the context of *E. coli* rRNA for testing against various position 69 mutations of Bst L11-C76. As detailed below, substitution of G1062/C1076 had unexpected effects on RNA tertiary structure stability, which necessitated preparation of a more extensive set of RNA variants.

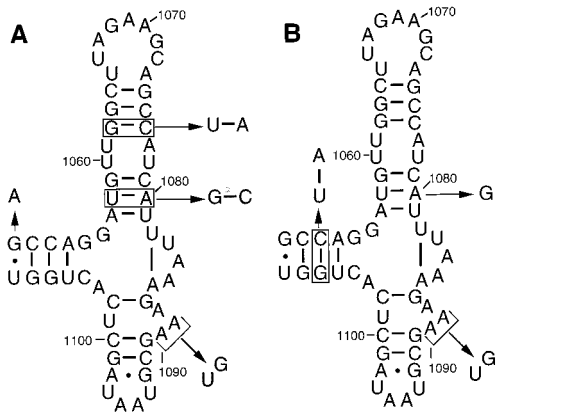

FIGURE 3: Differences between *E. coli* and plastid or *Mycoplasma* L11 binding rRNA domains. (A) Base changes in the 58 nucleotide fragment seen in chloroplasts (*Odontella sinensis*, *Porphyra purpurea*, and *Cyanophora paradoxa*) and *Synechocystis*. Base pair G1062−C1076, common in eubacteria, is replaced by U1062−1076A, and base pair U1058−A1080 is replaced by G1058−C1080. (B) Base changes observed in eubacteria *Mycoplasma genitalium* and *Mycoplasma pneumoniae*. Shaded substitutions also occur frequently in other eubacteria and archaea. In each panel, only those differences from the *E. coli* sequence that are present in all members of the group are shown. The triple base interaction [C1072(C1092/G1099)] is not shown for clarity.

Structural Analysis of Mutant rRNAs by UV Melting Experiments. L11 recognition of the 58 nucleotide rRNA domain requires a specific RNA tertiary structure (*20*). Hence, before studying the binding of L11-C76 to mutant rRNAs, it was important to verify whether these rRNAs had stable tertiary structures. The tertiary structure of the 1051−1108 rRNA fragment can be detected by UV melting experiments. Thermal unfolding of *E. coli* 1051−1108 rRNA monitored by UV hyperchromicity shows a broad unfolding transition at low temperature (*18*, *26*, *27*). Since this low-temperature transition could not be ascribed to unfolding of any of the secondary structural elements of the RNA, it was initially assigned to unfolding of tertiary structure (*18*). This

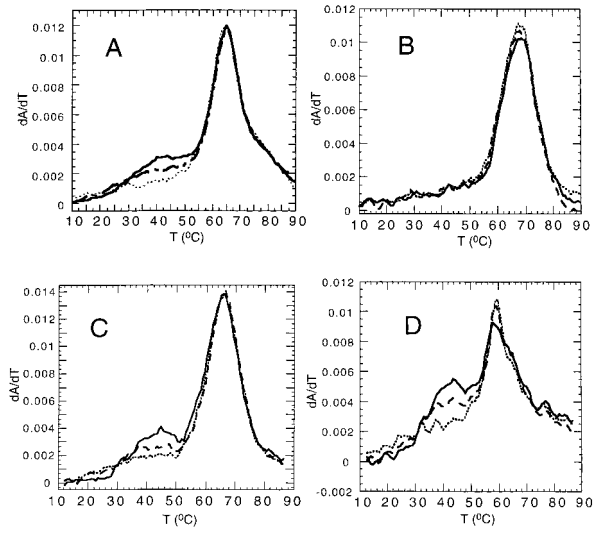

FIGURE 4: Representative UV melting profiles of L11 binding domain RNAs. RNA melting experiments were done in 10 mM MOPS (pH 7.0), 5 mM $MgCl_2$, and 175 mM monovalent salt. Melting profiles in NH_4Cl (—), KCl (---), and NaCl (•••) are shown. (A) 58 nucleotide wild-type *E. coli* rRNA fragment. (B) ChRNA2. (C) ChRNA4. (D) MycRNA1.

identification was later supported by unusual sensitivity of the transition to ions: the tertiary structure is specifically stabilized by Mg^{2+} among divalent ions (*27*, *28*) and by monovalent ions in the order $NH_4^+ > K^+ > Na^+$ (*26*) (Figure 4A). Conveniently, the tertiary unfolding transition shows hyperchromicity only at 260 nm and little or no hyperchromicity at 280 nm (*24*). Thus, comparison of UV melting experiments performed at 260 and 280 nm or in different monovalent salts detects formation of tertiary structure within this rRNA domain.

UV melting studies were performed with the parent *E. coli* RNA and all mutant RNAs (Figure 4 and Table 1). First, a mutant with 1062U/1076A (ChRNA1; Chloroplast-like rRNA Sequence 1), a base pair unique to chloroplasts, did not show any evidence of tertiary structure in melting experiments (data not shown). Additional sets of base changes that are found in all chloroplast and *Synechocystis* sequences, though also common among eubacteria and archaea (Figure 3A), were then added to see if the tertiary structure of this mutant could be rescued. Two additional variants are found outside the putative L11 contact region (Figure 1A). G1051A occurs frequently in eubacteria and is known to increase L11 binding modestly (by about 1.8-fold) (*29*), but is without any effect on the stability of the RNA tertiary structure (*24*); it therefore seemed unlikely to compensate for the destabilization caused by 1062U/1076A. The sequence G1089-U1090 is found in all plastids and most archaea and eubacteria, including *B. stearothermophilus*; mutation of the *E. coli* A1089-A1090 sequence to GU is known to increase the tertiary structure stability by about 20° (*24*). Inclusion of AA1090 → GU in ChRNA1 or ChRNA2 (to give ChRNA3 and ChRNA4, respectively) reversed the destabilizing effect of 1062U/1076A. Like the parent *E. coli* rRNA fragment, ChRNA3 (data not shown) and ChRNA4 (Figure 4C) show a distinct tertiary structure unfolding transition that is observable at 260 nm but not at 280 nm, and is preferentially stabilized by NH_4^+ and K^+ relative to Na^+.

Table 1: Ligand Binding Affinities and Tertiary Structure Stabilities of rRNA Fragments

RNA[a]	Wild Type (E.coli)	1089GU	ChRNA1	ChRNA2	ChRNA3	ChRNA4	MycRNA1	MycRNA2
Sequence	62 G-C 76 U A U U 58 U-A 80 A 89 A	G-C U A U U U-A G U	G-C U A U U U-A G U	U-A U A U U G-C A A	U-A U A U U U-A G U	U-A U A U U G-C G U	G-C U A U U U-G A A	G-C U A U U U-G G U
apparent T_m[b]	41 °C	59 °C	X	X	46 °C	47 °C	44 °C	58 °C
K, μM^{-1} [c]								
Thiostrepton	0.97±0.47	1.13±0.61	< 0.1	0.20±0.07	0.98±0.45	1.12±0.53	0.71±0.41	1.17±0.64
L11-C76	9.4±1.3	32.2±5.7	< 0.1	0.4±.0.1	7.5±1.9	9.6±1.1	2.5±0.4	8.9±0.2
S69N	1.0±0.8	1.6±0.1	< 0.1	1.1±0.7	8.5±1.7	11.1±0.1	< 0.1	1.0±0.5
S69Q	1.7±0.1	6.4±2.9	ND	ND	2.3±0.7	2.5±0.6	0.4±0.2	1.2±0.2
R68K/S69Q	< 0.1	< 0.1	ND	ND	ND	ND	< 0.1	< 0.1
S69A	1.2±0.3	3.0±0.1	ND	ND	1.1±0.4	1.6±0.1	ND	1.1±0.4

[a] Binding constants reported for wild-type and 1089GU RNAs were determined with rRNA fragments containing 23S rRNA sequence 1029−1126; all other determinations were made with the 59mer sequence containing nucleotides 1051−1108. Controls did not detect any significant differences between the two fragments in binding either L11-C76 or thiostrepton: binding of L11-C76 to a U1061A variant was determined to be 28 μM^{-1} and 40 μM^{-1} for the larger and smaller fragments, respectively, and thiostrepton binding to the same two RNAs was identical (0.9 μM^{-1}). [b] Apparent T_ms of 59mer rRNA fragments were determined in 5 mM $MgCl_2$, 175 mM NH_4Cl, and 10 mM MOPS (pH 7.0) as described under Materials and Methods, and are reproducible to ±1 °C. The melting temperature of the 1089GU mutant is taken from Lu and Draper (*24*). X, tertiary unfolding transition not detected. [c] Thiostrepton binding constants were determined by filter binding in 3 mM $MgCl_2$, 175 mM NH_4Cl, 10 mM MOPS (pH 7.0), and 5% v/v DMSO; protein binding constants were determined in the same buffer lacking DMSO and containing 175 mM KCl instead of NH_4Cl. Values are averages of at least two determinations; associated standard deviations are shown. ND, not determined.

MycRNA1 (Mycoplasma-like rRNA sequence 1) has the A1080G mutation unique to *M. genetalium* and *M. pneumoniae* and shows a stable tertiary structure that is preferentially stabilized by NH_4^+ and K^+ (Figure 4D). The part of the helix containing the 1058−1080 base pair is known to unfold first among the secondary structure elements of *E. coli* RNA (*18*). Replacing the U−A base pair at this position with a U•G mismatch further weakens the secondary structure of this region, as seen in the decreased T_m of the second transition relative to the parent sequence (Figure 4A); thus, the first transition appears as a shoulder on the larger secondary structure unfolding transition. As observed with the wild-type and chloroplast-like RNAs, the additional mutation AA1090 → GU in MycRNA1 (which gave MycRNA2) increases the stability of its tertiary structure (Table 1). Another mutation that occurs in both *M. genitalium* and *M. pneumoniae* is 1053C/1106G → 1053A/1106U. This mutation was not made because the sequence change was not located in the protein binding region and both MycRNA1 and MycRNA2 already had stable tertiary structures necessary for protein binding.

The apparent melting temperatures (T_ms) of the tertiary structures of all the RNA variants are shown in Table 1. The tertiary structure T_m is dependent on Mg^{2+} concentration and on the monovalent cation type, NH_4^+ or K^+ (*26*−*28*). Here the apparent T_ms were all estimated in 5 mM $MgCl_2$ and 175 mM NH_4Cl. Note that in rRNAs where melting of secondary and tertiary structures is closely spaced in temperature, the estimated apparent T_m underestimates the actual thermodynamic T_m (see discussion of coupled unfolding transitions in ref *30*).

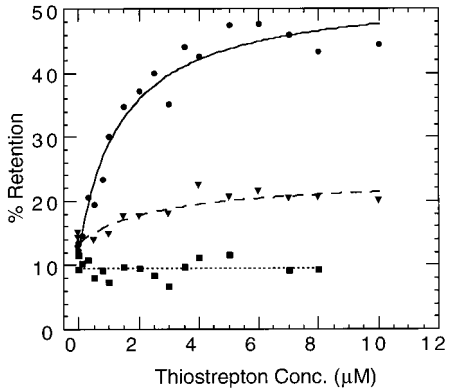

FIGURE 5: Thiostrepton binding to various RNAs, measured by nitrocellulose filter assay. Representative fits to binding curves are shown: ChRNA4 (● and —), ChRNA2 (▼ and ---), and ChRNA1 (■ and ⋯). Binding constants are given in Table 1.

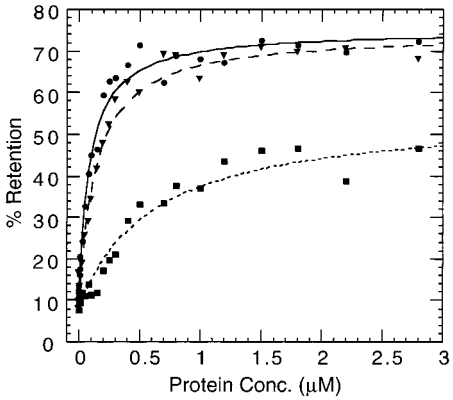

FIGURE 6: Representative L11-C76−RNA binding isotherms, measured by a nitrocellulose filter assay. Binding curves for the S69N protein are shown, with RNA variants GU1089/1090 (■ and ⋯), ChRNA3 (▼ and ---), and ChRNA4 (● and —). Binding constants are given in Table 1.

Thiostrepton Affinity for RNA Variants. The tertiary structure of the 1051−1108 rRNA fragment is specifically recognized by an antibiotic, thiostrepton (*10*, *31*). Thiostrepton protects both hairpin loops in the rRNA fragment from nuclease digestions (*31*), and its binding depends on formation of a triple base interaction involving base pair 1092/1099 and base 1072 (*32*). These observations suggest that the two hairpin loops come very close to each other in the correctly folded RNA molecule and form the binding site for thiostrepton. The rRNA affinity for thiostrepton thus provides a second method to assay for the presence of the tertiary structure in various rRNA mutants.

Nitrocellulose filter binding experiments were used to obtain association constants of various RNAs with thiostrepton. Representative binding isotherms are shown in Figure 5 and the data are summarized in Table 1. There is a good correlation between tertiary structure stability as observed by UV melting and the thiostrepton binding capacity of the RNA variants, in that ChRNA1 and ChRNA2 show undetectable or significantly reduced thiostrepton affinity, respectively. However, it is intriguing that ChRNA2, which shows no detectable tertiary structure in melting experiments, can still bind thiostrepton, though about 5-fold less tightly than RNAs with stable tertiary structures.

Wild-Type and Mutant Protein Binding to Various RNAs. The association constants between protein and RNA variants were measured by filter binding assays. The proteins initially used were the wild-type L11-C76 and its variants S69N (representing the chloroplast and *Synechocystis* substitution at position 69), S69Q (representing the *Mycoplasma* substitution at position 69), and S69A (representing a negative control, because of the inability of alanine to be involved in any hydrogen bonding interaction). Representative binding curves are shown in Figure 6, and results are summarized in Table 1.

Chloroplast-like RNAs with unstable tertiary structures, ChRNA1 and ChRNA2, bind weakly to the S69N protein variant, while chloroplast-like RNAs with stable tertiary structures, ChRNA3 and ChRNA4, show higher affinity for both wild-type L11-C76 and S69N, but bind S69A poorly. A comparison of S69N binding affinities to all RNAs clearly indicates that it recognizes the stably folded chloroplast-like RNAs, ChRNA3 and ChRNA4, in preference to wild-type RNA and all other RNA variants. In contrast, S69A is unable to make such discrimination and binds weakly to all RNAs. Thus, we conclude that an asparagine residue at position 69 makes a positive contribution to RNA binding, but only in the context of chloroplast-like RNA sequences. Surprisingly, serine at the same position, as found in most eubacteria, contributes to binding with both eubacterial and chloroplast-like RNA sequences, though binding to chloroplast-like RNAs is weaker (compare binding affinity of L11-C76 to 1089GU and ChRNA3 or ChRNA4, Table 1).

The *Mycoplasma* substitution, S69Q, however, does not recognize *Mycoplasma*-like RNA sequences well (Table 1). An additional mutation was then added to S69Q to make it more like the *Mycoplasma* L11 sequence. Position 68 is usually an arginine in eubacteria (19 out of 23 sequences), but is substituted by a lysine in the *Mycoplasma* and another eubacterium, *Thermotoga maritima*. This arginine residue contributes to the electrostatic free energy of L11−RNA recognition, as its mutation to alanine reduces the protein−RNA affinity and the salt dependence of the affinity (unpublished observations). However, the double mutant, R68K/S69Q, bound more poorly than S69Q, and did not show detectable affinity to any of the RNAs tested. CD spectra taken of the wild-type Bst L11-C76 and the R68K/S69Q variant in 20 mM phosphate buffer (pH 7.0) were nearly identical in shape (data not shown), the mutant spectrum being slightly more intense. Thus, loss of rRNA binding activity cannot be ascribed to drastic misfolding of protein structure by this mutation.

DISCUSSION

Ribosomal Protein−RNA Covariation. Comparative sequence analysis has been a powerful tool for rRNA structural analysis (*1*, *2*). Because substitution of one Watson−Crick or wobble base pair for another has little structural consequence in most helices, two-position nucleotide covariations have occurred throughout the rRNAs during evolution and make it possible to identify many helical elements in an RNA by analysis of relatively few sequences (*33*). Isosteric substitutions of tertiary interactions may require simultaneous covariation of a number nucleotides [for instance, neighboring base pairs may covary with triple base interactions (*32*, *34*)], making them less frequent and more difficult to detect. Nevertheless, the large number of sequences now available

has made it possible for comparative analysis to suggest a number of intriguing tertiary contacts. For instance, a triple base interaction within the 58 nt rRNA domain [C1072-(C1092/G1099); Figure 1A] was recently predicted and confirmed experimentally (*32*). Since many ribosomal proteins have been an integral part of the ribosome since the last common ancestor, presumably some protein−RNA compensatory mutations have occurred over the intervening eons, in addition to RNA−RNA covariations. Only recently has it become feasible to search for such compensatory variants, as the number of ribosomal protein sequences has expanded.

The analogue of a Watson−Crick covariation in an RNA−protein contact would be the substitution of an interacting base−amino acid pair with a different, isosteric pair yielding the same complex stability; we refer to this as direct compensation. Alternatively, maintenance of complementary protein and RNA surfaces may require an extensive set of amino acid or base changes to accommodate an amino acid or nucleotide mutation, analogous to the sets of rRNA base changes sometimes needed to maintain an rRNA tertiary structure; we call this indirect compensation, since residues some distance apart from each other may covary. In the results presented here, it is clear that an L11 sequence variant occurring in plastids (69N) is selective for plastid-like rRNA sequences versus closely related bacterial rRNAs: comparing 1089GU and ChRNA3, the simple substitution of a U−A base pair for G1062/C1076 is seen to enhance S69N protein binding 5-fold while discriminating against Ser, Ala, or Gln at position 69 by 3−4-fold. Both RNAs have similar (if not identical) tertiary structures at room temperature, as evident from their nearly identical thiostrepton binding affinity (Table 1). (The difference in tertiary structure stability between 1089GU and ChRNA3 can be ignored, since both RNAs are fully folded at the temperature of the binding assays.) Thus, an S69 → N mutation may have been compensated by rRNA mutations at the divergence of *Synechocystis* and plastids from the rest of the eubacteria, and direct compensation of G1062U/C1076A for S69N is suggested.

The free energy differences by which 1062/1076 variants discriminate between L11 sequences are modest, on the order of 0.7−1.0 kcal/mol, and within the range of 0.5−1.8 kcal/mol observed for the gain or loss of a single hydrogen bond in other recognition systems (*35*, *36*). In the few RNA−protein complexes for which atomic resolution structures are available, loss of a single hydrogen bond is frequently accompanied by very small free energy changes. For instance, the side chain of Gln 85 of the U1A protein makes a single hydrogen bond to a cytidine N4, and its mutation to valine reduces the binding affinity by only 2-fold (*37*, *38*). Similarly, Asn 87 of the MS2 coat protein bonds to a cytidine O2, but its mutation to Ala does not significantly affect the binding affinity (*39*). A single hydrogen bond between Ser 69 and base pair 1062/1076 is therefore plausible.

Since this work was completed, a new plastid L11 sequence appeared (*Guillardia theta*) (*40*). This organism has Asn at L11 position 69 and a 23S rRNA C1062/U1076 mismatch which is common among (and unique to) plastids. This suggests that Asn 69 discriminates in favor of both C1062/U1076 and U1062/A1076. The presence of a carbonyl vs an amino group on the minor groove face of 1062 is a potential discriminant that could be detected by Asn vs Ser. Though we consider a Ser 69−G1062/C1076 hydrogen bond a plausible hypothesis, only atomic resolution structure will reveal whether Ser 69 participates directly or indirectly in RNA recognition. In the following sections, we argue that data in Table 1 reveal the presence of indirect effects mediated by both RNA and protein in other aspects of L11−rRNA recognition.

General Effects of RNA Variants on Protein Recognition. The 1051−1108 rRNA domain folds into a compact tertiary structure in which the two distant hairpin loops in the RNA come very close to each other, as indicated by nuclease and chemical protection experiments (*11*) and the formation of a triple base interaction involving nucleotides at 1072 and 1092/1098 (*32*). A number of conserved bases surrounding the central junction are also required for tertiary structure formation and presumably form additional tertiary contacts (*24*); these include A1077, A1085, and A1088 (Figure 1A). L11 binding is strongly linked to formation of the correct tertiary structure: mutations destabilizing the structure also weaken L11 binding, as seen here in the dual effects of 1062U/1076A mutation on RNA stability and protein affinity, and L11 dramatically stabilizes the entire tertiary fold of the rRNA fragment (*13*, *20*). The extensive network of tertiary interactions may thus be able to transmit subtle conformational changes from many parts of the RNA to the protein binding site. This kind of indirect effect may account for the effects of two variants shown in Table 1, 1089GU and 1080G.

The 1089GU substitution is found in nearly all eukaryotic rRNAs and is widespread among prokaryotic rRNA. It appears to enhance tertiary structure stability in any sequence context (compare apparent T_ms in Table 1 for 1089GU vs wild-type RNA, and MycRNA2 vs MycRNA1). Since substitutions at 1089−1090 do not covary with any other part of the rRNA, it is impossible to tell whether these nucleotides are directly involved in tertiary contacts or simply stabilize other nucleotides (e.g., A1088) in a conformation favorable for tertiary folding. In any case, an AA → GU mutation at 1089−1090 should promote L11 binding in an RNA with an otherwise unstable tertiary structure; this accounts for the increases in protein and thiostrepton binding affinity in going from ChRNA1 to ChRNA3 and from ChRNA2 to ChRNA4.

In cases where the RNA fragment already has a stable tertiary structure, the 1089GU mutation still enhances binding of wild-type and variant proteins roughly 3-fold (cf. protein affinities for wild type vs 1089GU and MycRNA1 vs MycRNA2, Table 1). At the temperature of the binding assay, 22 °C, the effects of 1089GU on the stability of the tertiary structure in these RNAs should be negligible, and in fact thiostrepton binding affinities are unaffected; thus, GU1089 must stabilize the protein complex by some other means. A 3-fold enhancement of protein binding affinity could result from a subtle change in RNA structure transmitted to the protein interaction site by RNA tertiary interactions. Direct contacts of 1089GU with the protein are a less likely possibility, as these nucleotides lie outside of the protein "footprint" region (*12*).

The A1080G mutation, which replaces a U−A base pair with a wobble U·G, is similar to the 1089GU variant: RNA tertiary structure remains stable and thiostrepton binding affinity is unaffected, but binding of all protein variants is

reduced 3−5-fold (cf. wild-type RNA vs MycRNA1 and 1089GU vs MycRNA2, Table 1). Since the 1058/1080 base pair is within the presumed protein binding region, subtle changes in helix structure caused by introduction of a wobble pair (*41*, *42*) could be responsible for the small destabilization of protein complexes.

How Do Mycoplasma L11 Homologues Bind rRNA? The attempt to detect compensation between *Mycoplasma* L11 and rRNA sequences was a complete failure. Not only did *Mycoplasma*-like L11-C76 variants S69Q or R68K/S69Q bind the *Mycoplasma*-like RNA variants weakly, they also recognized wild-type and GU1089 RNAs with higher affinity. Conversely, L11-C76 bound the *Mycoplasma*-like RNA variants more strongly than did the *Mycoplasma*-like protein variants. How then does *Mycoplasma* L11 recognize its cognate rRNA site? Aside from the 1080G variant tested (which maintains RNA tertiary structure required for protein binding), *Mycoplasma* L11 binding domain RNA sequences show only variations common among eubacteria. However, *Mycoplasma* L11 sequences are quite different from all other known eubacterial and plastid sequences; presumably variations in other parts of the protein compensate for the inability of 68K/69Q to bind rRNA strongly.

There are several residues within the L11-C76 rRNA binding domain of the two known *Mycoplasma* L11 homologues that are unique or rare among other L11 sequences. The most striking of these differences lies within loop 1. Loop 1 is crucial for RNA binding: NMR experiments show that Ser 24, Glu 26, and Pro 27 make contacts with the RNA (*16*), and Pro 27 → Gly reduces RNA binding by more than 20-fold (unpublished data). Of these loop residues, Ser 24 has become Ala and Pro 27 has been deleted in *Mycoplasma* sequences. At each of these positions, only one other example of any variation is present in bacterial, plastid, and archaeal L11 homologues (see Supporting Information); eukaryotic loop 1 sequences (not shown) are entirely different from those of prokaryotes and presumably binding by a different mechanism. One possibility is that the *Mycoplasma*-like loop 1 sequence orients helix 3 differently on the rRNA surface, so that residues 68 and 69 contact different parts of the RNA structure than in other eubacterial complexes. A larger set of L11 sequences from *Mycoplasma* and more extensive mutagenesis experiments will be needed to find out which loop 1 changes and/or other unique features of *Mycoplasma* L11 sequences are needed for optimum recognition of different rRNAs.

Summary: Protein−RNA Recognition Mechanisms. Sequence comparisons among homologous L11−rRNA complexes have led us to measure binding affinities between a number of protein and RNA variants. These experiments ask, in essence, whether transplanting protein and RNA sequence variations from one organism into the context of a different L11−RNA complex perturbs the stability of the complex. This is analogous to "structural perturbation" experiments used to explore energetics of recognition in other macromolecular systems (*43*). In the present results, we have seen three different kinds of responses to protein and RNA mutations: (i) substitution at protein residue 69 may be compensated by substitution of base pair 1062/1076, possibly because of a direct contact between these components; (ii) the RNA mutation 1089AA → GU generically enhances protein binding affinity, probably because of favorable adjustments to the RNA tertiary structure; and (iii) helix 3 variations found in *Mycoplasma* cannot be transplanted into other L11 proteins, presumably because these mutations require adjustments in other parts of the protein.

Similar kinds of compensatory mutations have been studied in tRNA−synthetase recognition systems. For instance, a base pair substitution (G3/U70 → G3/C70) in the $tRNA^{Ala}$ acceptor stem eliminates charging by the cognate synthetase, but a base pair mutation in the anticodon stem (U27/A43 → U27/U43) revives charging (*44*). These base pairs are some distance apart in the tRNA structure, but both probably interact with the protein; the compensatory effect may be mediated by a protein conformational change. The same G3/U70 → G3/C70 mutation is also compensated by a synthetase mutation in the domain proximal to the acceptor stem (*45*).

It is probably not surprising that protein−RNA complexes, in which both components have specific tertiary structures, respond in both local and global ways to mutations introduced in or near their interaction surfaces. The limited number of pairs of L11 and rRNA sequences now available has already directed us to interesting features of this complex; the growing databases of homologous ribosomal protein and RNA sequences will be a useful guide in designing further directed mutations to probe protein−RNA recognition energetics.

ACKNOWLEDGMENT

We thank Drs. Thomas C. Gluick and K. Asish Xavier for valuable discussions and comments on the manuscript, Dr. Yury V. Bukhman for providing the sequences for L11 and 23S rRNA, and Dr. Doug Barrick for use of the CD apparatus.

SUPPORTING INFORMATION AVAILABLE

Two tables listing aligned sequences for homologues of L11-C76 and 23S rRNA nucleotides 1051−1108 from 25 prokaryotes for which both protein and RNA sequences are available (3 pages). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Gutell, R. R. (1993) *Curr. Opin. Struct. Biol. 3*, 313−322.
2. Gutell, R. R., Larsen, N., and Woese, C. R. (1994) *Microbiol. Rev. 58*, 10−26.
3. Jain, C., and Belasco, J. G. (1996) *Cell 87*, 115−125.
4. Schmidt, F. J., Thomson, J., Lee, K., Dijk, J., and Cundliffe, E. (1981) *J. Biol. Chem. 256*, 12301−12305.
5. El-Baradi, T. T. A. L., Regt, V. H. C. F. d., Einerhand, S. W. C., Teixido, J., Planta, R. J., Ballesta, J. P. G., and Raué, H. (1987) *J. Mol. Biol. 195*, 909−917.
6. Beauclark, A. A. D., Hummel, H., Holmes, D. J., Böck, A., and Cundliffe, E. (1985) *Eur. J. Biochem. 151*, 245−255.
7. Stark, M. J. R., Cundliffe, E., Dijk, J., and Stöffler, G. (1980) *Mol. Gen. Genet. 180*, 11−15.
8. Musters, W., Gonçalves, P. M., Boon, K., Raué, H. A., van Herikhuizen, H., and Planta, R. J. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 1469−1473.
9. Thompson, J., Musters, W., Cundliffe, E., and Dahlberg, A. E. (1993) *EMBO J. 12*, 1499−1504.
10. Ryan, P. C., and Draper, D. E. (1989) *Biochemistry 28*, 9949−9956.
11. Egebjerg, J., Douthwaite, S. R., Liljas, A., and Garrett, R. A. (1990) *J. Mol. Biol. 213*, 275−188.

12. Rosendahl, G., and Douthwaite, S. (1993) *J. Mol. Biol. 234*, 1013−1020.
13. Xing, Y., and Draper, D. E. (1996) *Biochemistry 35*, 1581−1588.
14. Xing, Y., GuhaThakurta, D., and Draper, D. E. (1997) *Nat. Struct. Biol. 4*, 24−27.
15. Markus, M. A., Hinck, A. P., Huang, S., Draper, D. E., and Torchia, D. A. (1997) *Nat. Struct. Biol. 4*, 70−77.
16. Hinck, A. P., Markus, M. A., Huang, S., Grzesiek, S., Kustonovich, I., Draper, D. E., and Torchia, D. A. (1997) *J. Mol. Biol. 274*, 101−113.
17. Brennan, R. G. (1993) *Cell 74*, 773−776.
18. Laing, L., and Draper, D. E. (1994) *J. Mol. Biol. 237*, 560−576.
19. Deng, W. P., and Nickoloff, J. A. (1992) *Anal. Biochem. 200*, 81−88.
20. Xing, Y., and Draper, D. E. (1995) *J. Mol. Biol. 249*, 319−337.
21. Ryan, P., Lu, M., and Draper, D. E. (1991) *J. Mol. Biol. 221*, 1257−1268.
22. Draper, D. E., Deckman, I. C., and Vartikar, J. V. (1988) *Methods Emzymol. 164*, 203−220.
23. Bradford, M. M. (1976) *Anal. Biochem. 72*, 248−254.
24. Lu, M., and Draper, D. E. (1994) *J. Mol. Biol. 244*, 572−585.
25. Sibold, C., and Subramanian, A. R. (1990) *Biochim. Biophys. Acta 1050*, 61−68.
26. Wang, Y., Lu, M., and Draper, D. E. (1993) *Biochemistry 32*, 12279−12282.
27. Bukhman, Y. V., and Draper, D. E. (1997) *J. Mol. Biol. 273*, 1020−1031.
28. Laing, L., Gluick, T. C., and Draper, D. E. (1994) *J. Mol. Biol. 237*, 577−587.
29. Lu, M., and Draper, D. E. (1995) *Nucleic Acids Res. 23*, 3426−3433.
30. Draper, D. E., Xing, Y., and Laing, L. (1995) *J. Mol. Biol. 249*, 231−238.
31. Egebjerg, J., Douthwaite, S. R., and Garrett, R. A. (1989) *EMBO J. 8*, 607−611.
32. Conn, G., Gutell, R. R., and Draper, D. E. (1998) *Biochemistry 37*, 11980−11988.
33. Noller, H. F., Kop, J., Wheaton, V., Brosius, J., Gutell, R. R., Kopylov, A. M., Dohme, F., Herr, W., Stahl, D. A., Gupta, R., and Woese, C. R. (1981) *Nucleic Acids Res. 9*, 6167−6189.
34. Gautheret, D., Damberger, S. H., and Gutell, R. R. (1995) *J. Mol. Biol. 248*, 27−43.
35. Fersht, A. R. (1987) *Trends Biochem. Sci. 12*, 301−304.
36. Lesser, D. R., Kurpiewski, M. R., Waters, T., Connolly, B. A. M., and Jen-Jacobson, L. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 7548−7552.
37. Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H., and Nagai, K. (1994) *Nature 372*, 432−438.
38. Jessen, T. H., Oubridge, C., Teo, C. H., Pritchard, C., and Nagai, K. (1991) *EMBO J. 10*, 3447−56.
39. Johansson, H. E., Dertinger, D., LeCuyer, K. A., Behlen, L. S., Greef, C. H., and Uhlenbeck, O. C. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 9244−9249.
40. Douglas, S. E., and Penny, S. L. (1998) *J. Mol. Evol.* (in press).
41. Mizuro, H., and Sundaralingam, M. (1978) *Nucleic Acids Res. 5*, 4451−4461.
42. Gautheret, D., Konings, D., and Gutell, R. R. (1995) *RNA 1*, 807−814.
43. Ackers, G. K., and Smith, F. R. (1985) *Annu. Rev. Biochem. 54*, 597−629.
44. Hou, Y.-M., and Schimmel, P. (1992) *Biochemistry 31*, 10310−10314.
45. Miller, W. T., Hou, Y.-M., and Schimmel, P. (1991) *Biochemistry 30*, 2635−2641.

BI9826411