

Dissection of the 16S rRNA binding site for ribosomal protein S4

Amalia Sapag, Jailaxmi V. Vartikar and David E. Draper

Department of Chemistry, Johns Hopkins University, Baltimore, MD (U.S.A.)

Key words: Ribosomal protein S4; S4 binding site; rRNA, 16S; Protein-RNA recognition; Deletion, 16S rRNA, Affinity, S4

The ribosomal protein S4 from *Escherichia coli* is essential for initiation of assembly of 30S ribosomal subunits. We have undertaken the identification of specific features required in the 16S rRNA for S4 recognition by synthesizing mutants bearing deletions within a 460 nucleotide region which contains the minimum S4 binding site. We made a set of large nested deletions in a subdomain of the molecule, as well as individual deletions of nine hairpins, and used a nitrocellulose filter binding assay to calculate association constants. Some small hairpins can be eliminated with only minor effects on S4 recognition, while three hairpins scattered throughout the domain (76–90, 376–389 and 456–476) are essential for specific interaction. The loop sequence of hairpin 456–476 is important for S4 binding, and may be directly recognized by the protein. Some of the essential features are in phylogenetically variable regions; consistent with this, *Mycoplasma capricolum* rRNA is only weakly recognized by S4, and no specific binding to *Xenopus laevis* rRNA can be detected.

Introduction

We have undertaken the study of the interaction between *Escherichia coli* ribosomal protein S4 and 16S rRNA. Being one of the first two proteins which bind to 16S rRNA [1], S4 holds a key role in the assembly of 30S subunits. Additionally, as a translational repressor, S4 controls the expression of the ribosomal proteins encoded in the α operon upon binding to the leader region of its polycistronic mRNA [2].

Given the size of 16S rRNA (1542 nucleotides) and our incomplete knowledge of its folding in the small subunit of the ribosome, the characterization of its contribution to complex formation with S4 (204 residues) is a challenging problem. Ribonuclease protection assays (reviewed in Ref. 3) showed early on that non contiguous sections of the molecule, mostly in the 5' domain of 16S rRNA, are involved in the interaction [4–6]. A more refined footprinting analysis has been obtained by chemical modification and enzymatic digestion assays [7]. Changes in reactivity between the naked RNA and the protein-RNA complex confirm the first 560 nucleotides as the binding region for S4; protected sites are found throughout the domain but are mostly clustered in a junction of five helices (27–47/395–556). These assays provide a first approxima-

tion of the limits of the binding site and yield extensive information on conformation and/or possible protein contact sites.

In recent years, in vitro transcription and site directed mutagenesis have made it possible to prepare synthetic RNA. This powerful tool has enabled us to gradually increase the magnification on the S4-16S rRNA complex as we carry out the following strategy: (i) *definition of the limits* of the binding site in terms of primary structure, (ii) *internal dissection* of this minimum domain, and (iii) *identification* of specific sequences or structural features recognized by the protein. We have accomplished the first stage by monitoring the effect of increasing 5' and 3' end deletions in 16S rRNA on the S4 binding capacity [8]. The boundaries of the minimum binding site are now more clearly defined as positions 39 and 500 of the nucleotide sequence.

Here we address the second stage of *internal dissection* of the specific binding site. Methodical pruning of structures from the molecule should allow us to distinguish regions which do not actively contribute to specificity of binding, or which are not required for correct folding, from those that do encompass essential features for recognition by the protein. The secondary structure of the RNA may be used as a guideline to choose subdomains or small structural units whose systematic deletion will be informative. The 16S rRNA molecule is amenable to this type of analysis since its proposed secondary structure is supported by extensive phylogenetic comparisons [9]. In this work, we present

the effects of a number of hairpin deletions in the 5' domain of 16S rRNA on the capacity of the molecule to bind ribosomal protein S4.

Materials and Methods

RNA fragments containing bases 1–559 of the *E. coli* 16S rRNA were prepared by in vitro transcription of plasmid DNA with T7 RNA polymerase, and deletions within this sequence were made by site-directed mutagenesis, as described in Ref. 8. The gene fragment corresponding to nucleotides 1–674 (*E. coli* numbering) of *M. capricolum* 16S rRNA was obtained from pMCH158 [10] by digestion with *Dra*I and *Eco*RI and was cloned into sites *Sma*I/*Eco*RI of pT7/T3-19 (BRL). This construct was linearized with *Sma*I for transcription. The DNA sequence corresponding to the 5' domain of *Xenopus laevis* 18S rRNA was cloned from pXlr12 [11] into pT7/T3-18 using *Pst*I and *Sma*I. The plasmid was linearized with *Pst*I for transcription.

S4-RNA fragment binding constants were measured by nitrocellulose filter binding assays as described in Ref. 8. Binding conditions were 4°C, 0.35 M KCl, 8 mM MgSO₄, 10 mM β-mercaptoethanol, 30 mM Tris (pH 7.6). A least-squares method was used to fit a hyperbolic curve to the experimental data from which an association constant was calculated for each interaction (see Fig. 1). Numerical values were divided by that of the wild-type RNA (14.4 μM⁻¹) [8] to establish relative binding strengths. *M. capricolum* and *X. laevis* RNA fragment binding affinities were also assayed by sedimentation in sucrose gradients with ³H-labeled S4 [8].

Results

The secondary structure of the 5' domain of 16S rRNA falls into two halves defined by the long-range pairing at 52–58/354–359 (Fig. 2). The fact that the

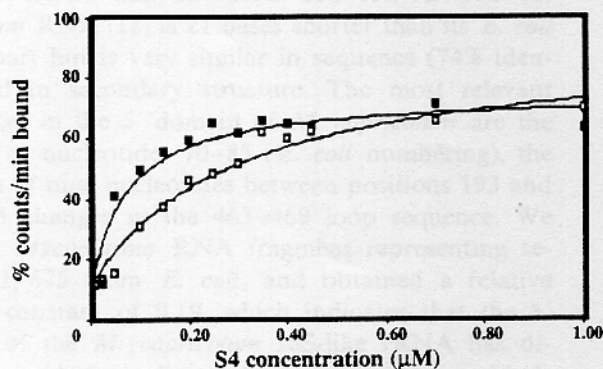


Fig. 1. Typical binding curves of protein S4 to 16S rRNA fragments. The percentage of ³⁵S retained on the filter is plotted. ■, wild-type (1–559); $K = 13.96 \mu\text{M}^{-1}$ and retention efficiency of 75%. □, mutant $\Delta(376-389)$; $K = 3.41 \mu\text{M}^{-1}$ and retention efficiency of 92%.

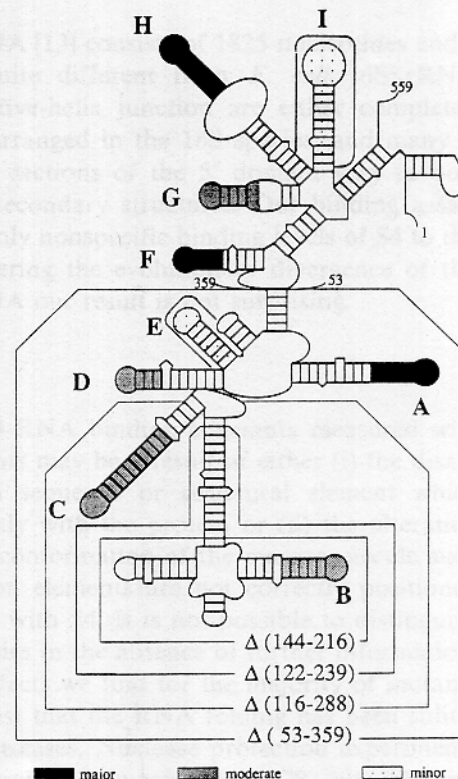


Fig. 2. Deletions in the 5' domain of 16S rRNA. Large deletions are enclosed in simple frames. Single hairpin deletions are indicated by shaded areas; their effect on S4 binding (black, major; grey, moderate; open, minor) is shown.

'lower' area (52–359) contains only one position which presents reduced reactivity towards cleavage reagents in the presence of S4 [7] had previously led us to investigate the possibility that part of it could be expendable. However, complete deletion of the subdomain ($\Delta(53-359)$) resulted in severe loss of S4 binding capacity [8]. We therefore tested a series of smaller nested deletions, namely $\Delta(116-288)$, $\Delta(122-239)$, and $\Delta(144-216)$, depicted in Fig. 2. Wild-type levels of binding were never recovered (relative binding constants are presented in Table I) indicating that these parts of the molecule contain elements required for either correct folding or direct interaction with the protein.

Since smaller deletions are less likely to cause drastic alterations in the conformation of the RNA we proceeded to remove sequences which represent individual hairpins in the secondary structure (Fig. 2). The effects observed are listed in Table I). They range from undetectable to a 10-fold decrease in the association constant relative to the wild-type value of $14.4 \mu\text{M}^{-1}$. In Fig. 2 we have noted these effects as minor (K_s reduced by 1.5-fold or less, essentially within experimental error of wild-type), moderate (reduced by 1.5- to 2.5-fold) and major (effects greater than 2.5-fold). Results reported are averages over three sets of experiments, the reproducibility being $\pm 20\%$. We consider 2.5-fold reductions

TABLE I

Binding for deletions in the 5' domain of 16S rRNA

Values are relative to the wild-type $K = 14.4 \mu\text{M}^{-1}$.

	Deletion	Relative K
Large deletions	$\Delta(53-359)$	0.21 ^a
	$\Delta(116-288)$	0.23
	$\Delta(122-239)$	0.39
	$\Delta(144-216)$	0.48
Single hairpin deletions		
A	$\Delta(76-90)$	0.37
B	$\Delta(153-198)$	0.54
C	$\Delta(247-273)$	0.62
D	$\Delta(294-303)$	0.44
E	$\Delta(317-336)$	0.71
F	$\Delta(376-389)$	0.29
G	$\Delta(415-432)$	0.66
H	$\Delta(456-476)$	0.15
I	$\Delta(501-544)$	1.17 ^a

^a Ref. 8.

as significant losses in binding strength, since other structured RNAs (e.g. tRNA, 23S rRNA) show non-specific binding affinities only 10 to 100 times weaker than the specific 16S rRNA [8].

Hairpin H has the most pronounced effect of the series, so we changed the sequence of the loop (positions 463–469) to its Watson–Crick complement to determine whether the required feature is the loop alone, the helical element or both. The relative binding constant has a value of 0.22, within experimental error of the value obtained for deletion of the full hairpin (0.15). It therefore appears that S4 requires a specific sequence in this hairpin loop (see Discussion).

The use of fragments of 16S-like rRNA from other organisms may serve as a complementary approach to the artificial generation of deletion and sequence mutants. We explored this alternative by evaluating the binding level of *E. coli* S4 to fragments of *M. capricolum* 16S rRNA and *X. laevis* 18S rRNA. The *M. capricolum* RNA [12] is 21 bases shorter than its *E. coli* counterpart but is very similar in sequence (74% identity) and in secondary structure. The most relevant differences in the 5' domain of *M. capricolum* are the deletion of nucleotides 70–85 (*E. coli* numbering), the insertion of nine nucleotides between positions 193 and 194, and changes in the 463–469 loop sequence. We tested a *Mycoplasma* RNA fragment representing sequence 1–675 from *E. coli*, and obtained a relative binding constant of 0.19, which indicates that the 5' domain of the *M. capricolum* 16S-like rRNA has diverged enough from *E. coli* to prevent wild-type binding but retains enough similarity to support intermediate affinity for the *E. coli* protein. Smaller RNA fragments were also evaluated (data not shown) but exhibited only nonspecific S4 binding levels. The *X.*

laevis 18S rRNA [13] consists of 1825 nucleotides and is structurally quite different from *E. coli* 16S²rRNA. Parts of the five-helix junction are either completely missing or rearranged in the 18S species, and many of the remaining sections of the 5' domain vary in both primary and secondary structures. Our binding assays could detect only nonspecific binding levels of S4 to this RNA. Considering the evolutionary divergence of this eukaryotic RNA our result is not surprising.

Discussion

Reduced S4-RNA binding constants measured with deletion mutants may be a result of either (i) the disappearance of a sequence or structural element which interacts directly with the protein or (ii) the alteration of the overall conformation of the macromolecule such that recognition elements are not correctly positioned for interaction with S4. It is not possible to distinguish these possibilities in the absence of further information. The modest effects we find for the majority of mutants analysed suggest that the RNA folding has been subtly altered in most cases. Nuclease protection experiments showing digestion of sequences 246–279, 295–301 and 323–361 [14] are also consistent with hairpin deletions C, D and E (247–273, 294–303 and 317–333, respectively) perturbing only RNA structure and not protein recognition features. The loop of hairpin H presents a contrasting situation. Changes in this sequence substantially reduce the specificity of S4 binding, and S4 binding to the α operon mRNA is reduced by about the same amount when A106, located in an AUA sequence of an exposed loop of the pseudoknot structure, is changed to U [15]. The similar effects of the mutations in the two different RNAs make this sequence a good candidate for a feature directly recognized by S4.

It does not appear possible to prepare a much smaller RNA recognizing S4 by constructing deletions within the 16S rRNA 5' domain. The three hairpins most essential for binding (A, F and H) are scattered throughout the structure, and large deletions in the 52–359 region also have a major effect on binding. S4 may recognize a complex tertiary structure which depends on interactions between elements widely separated in maps of the secondary structure, or S4 itself may be in effect crosslinking different parts of the 5' domain. Further study of the features recognized by S4 may help determine the overall three-dimensional folding of this rRNA domain.

It has sometimes been assumed that protein recognition of ribosomal RNA will take advantage of sequences which are highly conserved along the phylogenetic scale. In the S4-16S rRNA complex, however, both conserved and variable regions are needed for the interaction: two of the three hairpins we find to be most important, A and H, are in the phylogenetically most

variable regions [16]. This is consistent with a previous study which showed that *E. coli* S4 is able to interact with 16S rRNA from very closely related bacteria (species representing *Pseudomonas*, *Proteus*, *Enterobacter* and *Photobacter*) but not from more distant ones such as *Bacillus subtilis* or *Chromatium vinosum* [17]. The intermediate binding levels obtained with *Mycoplasma capricolum* RNA also suggest the importance of variable regions; *Mycoplasma* is thought to be phylogenetically closer to Gram-positive bacteria (e.g., *Bacillus*) than to Gram negatives [12]. The phylogenetic variability of S4 from different bacteria would clearly be of interest (note that *Bacillus subtilis* does contain a gene homologous to *E. coli* S4 [18]). It may be that variable protein-RNA interactions have evolved in this domain to maintain a conserved core of RNA structure in the optimum configuration for protein synthesis.

Acknowledgements

We thank Sarah Morse for the purification of T7 RNA polymerase and for synthesis of oligonucleotides. We are also grateful to Syozo Osawa and Susan Gerbi for plasmids pMCH158 and pXlr12 respectively, and to Eduardo Ostertag for assistance in computer programming. This work was supported by grant GM29048 from the NIH, a Research Career Development Award (CA01081) to D.E.D. and a fellowship from Lever Research Inc. awarded to A.S.

References

- 1 Nowotny, V. and Nierhaus, K.H. (1988) *Biochemistry* 27, 7051-7055.
- 2 Jinks-Robertson, S. and Nomura, M. (1982) *J. Bacteriol.* 151(1), 193-202.
- 3 Zimmermann, R.A. (1980) In *Ribosomes: structure function and genetics* (Chambliss, G., Craven, G.R., Davies, J., Davies, K., Kahan, L. and Nomura, M., eds.), pp. 135-170, University Park Press, Baltimore.
- 4 Zimmermann, R.A., Muto, A. and Mackie, G.A. (1974) *J. Mol. Biol.* 86, 433-450.
- 5 Ungewickell, E., Garrett, R., Ehresmann, C., Stiegler, P. and Fellner, P. (1975) *Eur. J. Biochem.* 51, 165-180.
- 6 Mackie, G.A. and Zimmermann, R.A. (1975) *J. Biol. Chem.* 250(11), 4100-4112.
- 7 Stern, S., Wilson, R.C. and Noller, H.F. (1986) *J. Mol. Biol.* 192, 101-110.
- 8 Vartikar, J.V. and Draper, D.E. (1989) *J. Mol. Biol.* 209, 221-234.
- 9 Noller, H.F. (1984) *Ann. Rev. Biochem.* 53, 119-162.
- 10 Sawada, M., Muto, A., Iwami, M., Yamao, F. and Osawa, S. (1984) *Mol. Gen. Genet.* 196, 311-316.
- 11 Dawid, I.B. and Wellauer, P.K. (1976) *Cell* 8, 443-448.
- 12 Iwami, M., Muto, A., Yamao, F. and Osawa, S. (1984) *Mol. Gen. Genet.* 196, 317-322.
- 13 Salim, M. and Maden, B.E.H. (1981) *Nature* 291, 205-208.
- 14 Ehresmann, C., Stiegler, P., Carbon, P., Ungewickell, E. and Garrett, R.A. (1980) *Eur. J. Biochem.* 103, 439-446.
- 15 Tang, C.K. and Draper, D.E. (1990) *Biochemistry* 29, 4434-4439.
- 16 Woese, C.R., Gutell, R., Gupta, R. and Noller, H.F. (1983) *Microbiol. Rev.* 47(4) 621-669.
- 17 Thurlow, D.L. and Zimmermann, R.A. (1978) *Proc. Natl. Acad. Sci. USA* 75(6), 2859-2863.
- 18 Henkin, T.M. and Chambliss, G.H. (1984) *Mol. Gen. Genet.* 193, 364-369.