

# The crystal structure of ribosomal protein S4 reveals a two-domain molecule with an extensive RNA-binding surface: one domain shows structural homology to the ETS DNA-binding motif

Christopher Davies<sup>1</sup>, Resi B.Gerstner<sup>2</sup>,  
David E.Draper<sup>2,3</sup>, V.Ramakrishnan<sup>3,4</sup> and  
Stephen W.White<sup>1,3,5</sup>

<sup>1</sup>Department of Structural Biology, St Jude Children's Research Hospital, 332 North Lauderdale, Memphis, TN 38105, <sup>2</sup>Department of Chemistry, The Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, <sup>4</sup>Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, UT 84132 and <sup>5</sup>Department of Biochemistry, University of Tennessee, 858 Madison, Suite G01, Memphis, TN 38163, USA

<sup>3</sup>Corresponding authors  
e-mail: stephen.white@stjude.org, v.ramakrishnan@m.cc.utah.edu or draper@jhunix.hcf.jhu.edu

**We report the 1.7 Å crystal structure of ribosomal protein S4 from *Bacillus stearothermophilus*. To facilitate the crystallization, 41 apparently flexible residues at the N-terminus of the protein have been deleted (S4Δ41). S4Δ41 has two domains; domain 1 is completely α-helical and domain 2 comprises a five-stranded antiparallel β-sheet with three α-helices packed on one side. Domain 2 is an insertion within domain 1, and it shows significant structural homology to the ETS domain of eukaryotic transcription factors. A phylogenetic analysis of the S4 primary structure shows that the likely RNA interaction surface is predominantly on one side of the protein. The surface is extensive and highly positively charged, and is centered on a distinctive canyon at the domain interface. The latter feature contains two arginines that are totally conserved in all known species of S4 including eukaryotes, and are probably crucial in binding RNA. As has been shown for other ribosomal proteins, mutations within S4 that affect ribosome function appear to disrupt the RNA-binding sites. The structure provides a framework with which to probe the RNA-binding properties of S4 by site-directed mutagenesis.**

**Keywords:** protein evolution/protein–RNA interactions/ribosome architecture/translational feedback/X-ray crystallography

## Introduction

Despite many years of effort by numerous investigators, our understanding of the structure of the ribosome is still not sufficiently detailed to answer many of the fundamental questions concerning its mechanism. An approach that is now making significant advances is to incorporate high resolution models of isolated ribosomal components into low resolution models of the individual subunits based on images from electron microscopy. This process is made possible by the numerous structural constraints that are being accumulated from the bacterial ribosome (almost

exclusively *Escherichia coli*) using techniques such as RNA footprinting, cross-linking and mutagenesis. Therefore, much effort is being directed towards solving the structures of ribosomal proteins, fragments of rRNA and protein–rRNA complexes, and the past few years have seen a significant increase in the number of protein structures that have been determined. Currently, 15 unique ribosomal protein structures are known (Ramakrishnan and White, 1998), and two of these have been fitted into models of 16S rRNA (Heilek and Noller, 1996; Tanaka *et al.*, 1998).

Although it is now accepted that the fundamental mechanisms of the ribosome are mediated by the RNA component (Dahlberg, 1989), evidence so far indicates that functional ribosomes cannot be formed in the absence of the proteins. Thus, catalytic activity is still possible after the majority, but not all, of the proteins have been removed by proteolysis (Noller *et al.*, 1992). It is likely that the primary function of the proteins is architectural, to help fold the rRNA into the correct three-dimensional structure for biological activity. Antibiotics that are directed against the ribosome generally bind to RNA, and the observation that many of the antibiotic resistance mutations occur in proteins indicates that the proteins can modulate rRNA structure within the intact ribosome.

Ribosomal protein structures have also provided some fascinating insights into the early events of protein evolution and the possible origins of modern protein folds. The 15 known prokaryotic ribosomal protein structures contain a total of 20 independently folded domains, and it is remarkable that only two of these domains are completely unique. All of the rest are clearly recognizable in other proteins, often with related functions such as DNA and RNA binding (Ramakrishnan and White, 1998). These observations are consistent with the view that ribosomal proteins were amongst the earliest proteins to have evolved, and that during the course of evolution their folds have been retained and modified for different functions. It is possible that the set of folds that will eventually be found in all the prokaryotic ribosomal proteins provided the structural templates for many modern proteins. It was noted recently that an efficient strategy will have to be adopted to determine unique structures amongst the vast numbers of proteins that are being discovered by genome sequencing (Pennisi, 1998). Clearly, the ribosomal proteins represent a rich source of structural information that needs to be fully tapped.

S4 is one of the largest bacterial ribosomal proteins, with a mol. wt of 23 kDa and comprising some 200 amino acids. It is also one of the more important proteins, with key roles in the assembly of the 30S subunit and the maintenance of translational fidelity. Its location in the body of the 30S subunit close to proteins S3, S5 and S12 has been established by neutron

**Table I.** Data collection statistics

	Native 1	Native 2	PIP <sup>a</sup>	TET	CNPT	URAC
Molarity (mM)			sat (~5)	0.1	20.0	10.0
Length of soak			8 days	20 h	30 days	16 days
Stabilizing buffer <sup>b</sup>			AS	AS	LS	AS
Temperature (K)	298	100	298	298	298	298
Reflections (unique)	6683	25 148	9776	5639	7799	7627
Reflections (measured)	24 852	139 696	54 551	31 795	46 834	33 263
Completeness (%)	95.3	96.4	94.7	99.6	96.9	98.7
Resolution (Å)	2.25	1.65	2.5	2.8	2.5	2.5
$R_{\text{sym}}^c$	7.8	5.6	7.3	11.7	8.7	5.4

<sup>a</sup>Derivative abbreviations: PIP, di-mu-iodobis-(ethylenediamine) di-platinum nitrate; TET, dipotassium tetrakis (thiocyanato) platinate (II); CNPT, potassium tetracyanoplatinate (II); URAC, uranyl acetate.

<sup>b</sup>AS, ammonium sulfate; LS, lithium sulfate.

<sup>c</sup> $R_{\text{sym}} = \sum |I_i - I_m| / \sum I_m$  where  $I_i$  is the intensity of the measured reflection and  $I_m$  is the mean intensity of all symmetry-related reflections.

scattering (Capel *et al.*, 1987) and immune electron-microscopy (Stöffler and Stöffler-Meilicke, 1984). Based on a variety of footprinting and cross-linking data, the RNA-binding region for S4 in the 30S subunit is well defined. Nuclease protection experiments and hydroxyl radical probing show that S4 contacts a compact structure at the junction of five helical segments formed within a 460 nucleotide region at the 5' end of 16S rRNA (Stern *et al.*, 1986; Powers and Noller, 1995a). This so-called 'S4 junction' site encompasses nucleotides 27–47 and 394–556, and is supported by S4–RNA cross-links to nucleotides 413 (Greuer *et al.*, 1987), 427–500 and 528–565 (Ehresmann *et al.*, 1977).

S4 is one of the six small subunit 'primary' RNA-binding proteins that can interact specifically with 16S rRNA in the absence of other proteins (Held *et al.*, 1974). Moreover, together with S7, it initiates the assembly of the entire 30S subunit and is absolutely essential to the assembly process (Nomura and Held, 1974). S4 initially organizes a subdomain that includes proteins S20, S16, S15, S6 and S18 (Nowotny and Nierhaus, 1988), and this eventually creates the 'body' of the small subunit. Functionally, this region contains the 'accuracy domain' of which S4 is a principal component, together with S5 and S12, and the 530 and 900 loops of 16S rRNA (Powers and Noller, 1991). The other assembly protein, S7, organizes the 3' end of 16S rRNA to form the 'head' of the small subunit.

S4 is also one of the proteins that autogeneously regulates the expression of other ribosomal proteins by binding to polycistronic mRNA (Dean and Nomura, 1980; Thomas *et al.*, 1987). S4 specifically binds to a pseudoknot structure consisting of 110 nucleotides within the  $\alpha$  operon mRNA (Spedding and Draper, 1993). There is no detectable sequence homology between the mRNA- and rRNA-binding sites for S4, and it is not obvious whether the two RNAs have any structural similarity.

Here we present the structure of S4 that has been solved using X-ray crystallography and refined to 1.7 Å resolution. For crystallization purposes, 41 residues at the N-terminus that are apparently flexible have been deleted. This construct was selected on the basis of the trypsin digestion experiments of Changchien and Craven (1976). The resulting protein, referred to as S4Δ41, retains the wild-type protein's ability to interact specifically with rRNA and mRNA, albeit with a slightly lower affinity (R.B.

Gerstner and D.E.Draper, unpublished). The S4Δ41 structure comprises two domains, both of which have recognizable structural homology to other proteins, most notably the ETS DNA-binding domain. The structure also reveals an extensive RNA interaction surface which is consistent with the role of S4 as an important assembly protein.

## Results

### Purification and crystallization

The gene encoding residues 42–200 of ribosomal protein S4 from *Bacillus stearothermophilus* was incorporated into the plasmid pET13 expression vector and transformed into BL21 (DE3) cells. Protein expression was excellent, yielding some 50 mg/l of S4Δ41. Crystals of S4Δ41 were obtained using ammonium sulfate as a precipitant, in the pH range 8.0–9.5. The best crystals grew in 2.15 M ammonium sulfate, 100 mM Tris pH 8.5 at 22°C, using an initial protein concentration of 20 mg/ml. These appeared after 5–6 days and eventually grew to be very large with dimensions 1.0×1.0×1.5 mm<sup>3</sup>. The crystals belong to space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>, with cell dimensions  $a = 50.38$  Å,  $b = 57.23$  Å and  $c = 76.25$  Å, and diffract to 1.65 Å resolution. Assuming an average packing density within the crystal, the dimensions are consistent with one molecule in the asymmetric unit (Matthews, 1968), which eventually proved to be the case.

### Structure determination

The S4Δ41 structure was determined using standard multiple isomorphous replacement (MIR) techniques. A total of nine derivatives were identified using various platinum, uranium, mercury and gold compounds, but some of these heavy atoms occupied identical sites in the crystal lattice, and only the best were used for phasing. These were potassium tetracyanoplatinate (II), di-mu-iodobis-(ethylenediamine) di-platinum nitrate (PIP), uranyl acetate and dipotassium tetrakis (thiocyanato) platinate (II) (Table I). Initial phasing at 2.5 Å produced an overall figure of merit of 0.693, and this increased to 0.828 after solvent flattening. The data collection and phasing statistics are shown in Tables I and II. The resulting MIR electron density map was of excellent quality (Figure 1A) and allowed the fitting of the entire model, including side chains and both termini. This initial model refined to  $R$  and  $R_{\text{free}}$  values of 21.2 and 35.2%,

**Table II.** Phasing statistics

Type of data <sup>b</sup>	PIP <sup>a</sup>		TET		CNPT		URAC	
	iso	ano	iso	ano	iso	ano	iso	ano
Resolution (Å)	2.5	2.5	2.8	2.8	2.5	2.5	2.5	2.5
No. of reflections	7246	5863	5182	4003	7230	5904	7427	6028
$R_{\text{merge}}^c$	20.3		6.5		10.9		6.5	
Phasing power	1.60	1.66	1.25	1.20	2.00	1.69	1.28	1.46
$R_{\text{Cullis}}^d$	0.57		0.67		0.55		0.67	
$R_{\text{Kraut}}^e$	0.33	0.31	0.11	0.21	0.14	0.15	0.10	0.11
Mean figure of merit	0.25	0.23	0.29	0.23	0.34	0.31	0.30	0.25
Overall figure of merit	0.69							

<sup>a</sup>Derivative abbreviations are as in Table I.

<sup>b</sup>iso is isomorphous and ano is anomalous.

<sup>c</sup> $R_{\text{merge}} = \sum |F_{\text{PH}} - F_{\text{p}}| / \sum |F_{\text{PH}}|$ .

<sup>d</sup> $R_{\text{Cullis}} = \sum |(F_{\text{PH}} \pm F_{\text{p}}) - F_{\text{H}}(\text{calc})| / \sum |F_{\text{PH}} - F_{\text{p}}|$ .

<sup>e</sup> $R_{\text{Kraut}} = \sum |F_{\text{PH}}(\text{obs}) - F_{\text{PH}}(\text{calc})| / \sum |F_{\text{PH}}(\text{obs})|$ .

Phasing power =  $F_{\text{H}}/E_{\text{RMS}}$ .

$F_{\text{p}}$ ,  $F_{\text{PH}}$  and  $F_{\text{H}}$  are the protein, derivative and heavy atom structure factors, respectively, and  $E_{\text{RMS}}$  is the residual lack of closure.

respectively, using data to 2.5 Å resolution. This model was refined further by alternating rounds of XPLOR and manual rebuilding using 1.65 Å data collected from cryo-cooled crystals. The final  $R$  and  $R_{\text{free}}$  values are 24.4 and 30.0%, respectively, using all data in the range 15.0–1.7 Å, and the model includes 243 water and six sulfate molecules. A Ramachandran plot shows that 93.6% of the residues are in the most favored region and 5.7% are in the additional allowed region. The statistics of the final model are shown in Table III, and the corresponding  $2F_{\text{o}} - F_{\text{c}}$  map is shown in Figure 1B.

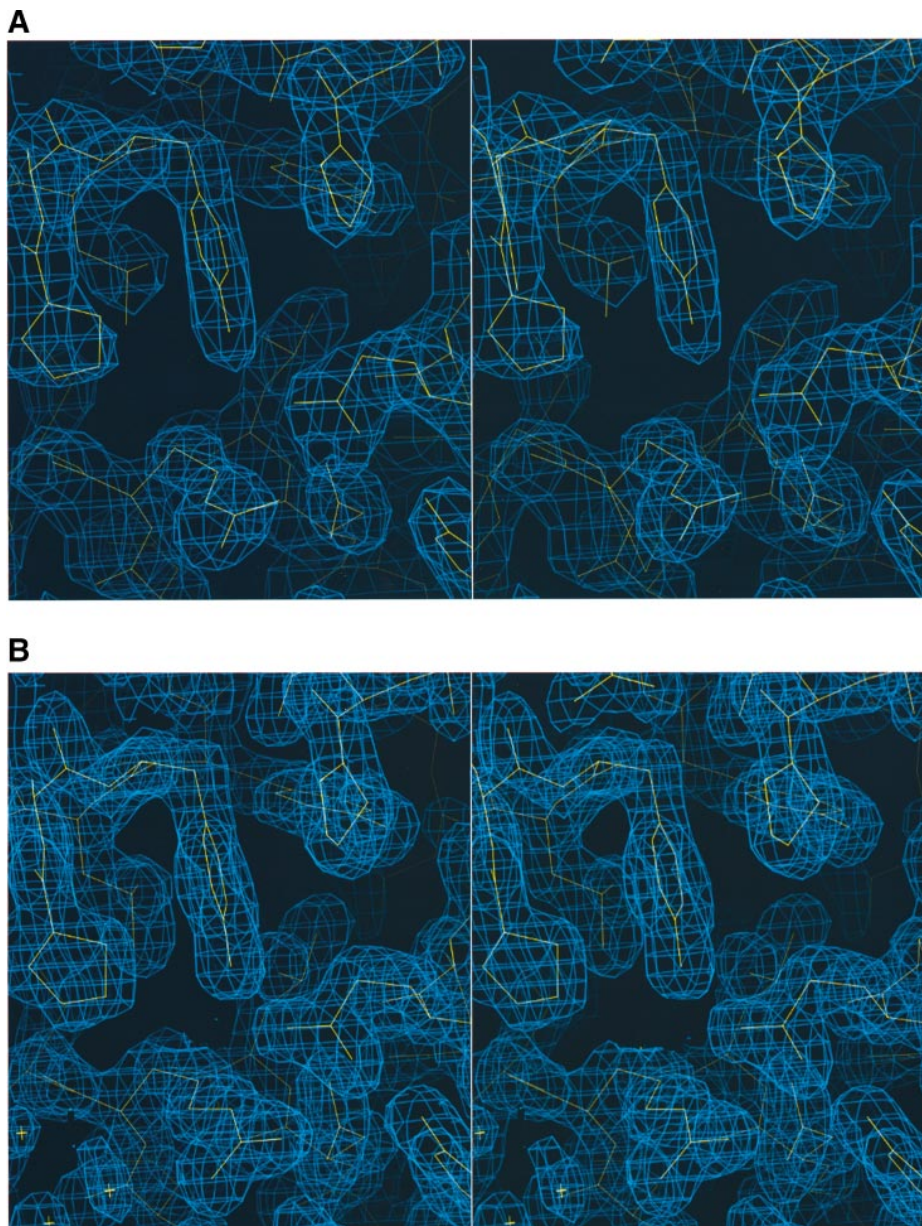
### Structure description

The molecule is numbered from residues 42 to 200 according to the *B.stearothermophilus* sequence (S.E. Gerchmann and V.Ramakrishnan, unpublished), and it starts with a methionine at the N-terminus which replaces an arginine in the wild-type sequence (R.B.Gerstner and D.E.Draper, unpublished). S4Δ41 is a two-domain, slightly elongated molecule with dimensions 40×50×65 Å<sup>3</sup>. The entire structure is well ordered, and the only significant regions of flexibility in the main chain are at the termini and between residues 157 and 158 where the electron densities are weak. With the possible exception of S15 (Clemons *et al.*, 1998), S4, like all the ribosomal proteins, shows no evidence of having an artifactual conformation when isolated from the ribosome. For example, there is a clear and conserved hydrophobic core which extends through the entire molecule, including the domain interface. The sequence of secondary structure elements in the molecule is α1-α2-α3-α4-α5-β1-β2-β3-α6-β4-β5-α7. Domain 1 is entirely α-helical (α1-α2-α3-α7), whereas domain 2 has an α/β structure (α4-α5-β1-β2-β3-α6-β4-β5) and represents an insertion into domain 1. A C<sub>α</sub> trace and a ribbon diagram are shown in Figure 2, and the secondary structure is shown in Figure 3, superimposed on an alignment of S4 sequences from several bacterial and chloroplast species.

Domain 1 comprises a bundle of four α-helices from the non-contiguous residues 42–91 and 180–200. The locations of the α-helices within the sequence are as follows: α1, residues 45–62; α2, residues 64–77; α3, residues 82–92; and α7, residues 192–200. Ser45, Asn64

and Asn190 N-cap helices α1, α2 and α7, respectively. In addition, there is a short region of 3/10 helix between residues 181 and 185. The arrangement of the α-helices within the bundle is rather unusual since α1 and α2 form a 98° angle and interact very little, but α7 is wedged between them. α3 is packed against α2 at a more typical angle, and leads directly into domain 2. A notable feature of domain 1 is a cluster of aromatic residues in the hydrophobic core that interact in a typical edge-to-face orientation (Burley and Petsko, 1985). The residues include phenylalanines 68, 72, 86 and 197 and Tyr198. Of these, Phe197 and Tyr198 are also close to the protein surface, and they probably also have a role in binding RNA (see below). Other hydrophobic core residues include Leu57, Leu90, Val63, Met87, Ile189 and Ile194. The relatively large number of aromatic residues in S4 from *B.stearothermophilus* may contribute to the molecule's thermal stability. In addition to the hydrophobic core, there are two additional types of specific interaction that appear to be important for maintaining the fold of domain 1. The first involves salt bridges that tether potentially flexible regions of the fold. One is between Arg182 and Glu191 that connects the 3/10 helix to α7, and the second is between Lys54 and Glu65 that links helices α1 and α2. It is interesting to note that the 3/10 helix is also held in place by a third salt bridge between Arg178 and Glu184. The conservation of these salt bridges suggests an important role for the 3/10 helix in RNA binding that is specific to prokaryotes (Figure 3). A number of other salt bridges in the domain probably stabilize the helical structures (Marqusee and Baldwin, 1987). The second type of interaction involves polar side chains that provide important structural hydrogen bonds. The Oε of Glu53 is suitably positioned to hydrogen-bond to the main chain nitrogen of Asn190, and the Oδ of Asn85 forms a hydrogen bond to the carboxyl group of Pro79 and stabilizes loop 3 between α2 and α3.

Domain 2 (residues 92–179) is composed of three α-helices and five β-strands organized as an antiparallel sheet in a Greek key motif. The locations of the secondary structure elements within the sequence are as follows: α4, residues 93–101; α5, residues 106–116; β1, residues 119–122; β2, residues 124–125; β3, residues 137–141;



**Fig. 1.** Stereoviews of the electron density map of ribosomal protein S4 $\Delta$ 41. The area shown is in the interdomain hydrophobic core which features the two aromatic residues Tyr164 and Tyr176. Both maps are contoured at  $1.5\sigma$  and displayed using the 'O' program (Jones *et al.*, 1991). (A) The solvent-flattened MIR map calculated at  $2.5\text{ \AA}$  that was used to build the initial model. (B) The  $2F_o - F_c$  calculated phased map generated from the final refined coordinates at  $1.7\text{ \AA}$ .

$\alpha 6$ , residues 148–158;  $\beta 4$ , residues 164–168; and  $\beta 5$ , residues 173–177. The connecting loop between  $\beta 2$  and  $\beta 3$  has an extended conformation and is anchored at the center by a short antiparallel strand–strand interaction involving residues 129–133 with residues 92–95 at the interdomain junction between  $\alpha 3$  and  $\alpha 4$ . This interaction also serves to N-cap  $\alpha 4$ , and Thr106 N-caps  $\alpha 5$ . There is also a 3/10 helical region connecting  $\beta 3$  and  $\alpha 6$  (residues 143–147). This 3/10 helix occupies a potentially flexible region of domain 2, and it is anchored by the side chain of Arg146 that interacts with Glu170 within the type I  $\beta$  turn between  $\beta 4$  and  $\beta 5$ . The guanidinium group forms both a salt bridge with the side chain and a hydrogen bond with the main chain oxygen. The three  $\alpha$ -helices pack against one side of the  $\beta$ -sheet and create the hydrophobic core of the domain. The core residues are

leucines 94, 97, 101, 103 and 155, valines 98, 114, 121, 126, 133 and 141, isoleucines 119, 139 and 151, and Phe167, most of which are highly conserved as hydrophobics. The surface of the sheet is populated by a mixture of charged and polar residues that are not conserved.

The two domains are packed closely together and form a canyon on one side of the molecule (Figure 2). The domain interface is well conserved (Figure 3). The central region is formed by the contiguous hydrophobic cores of each domain and is characterized by five tyrosine residues, 61, 99, 131, 164 and 176, although 99 and 131 are close to the surface and probably also interact with RNA (see below). The periphery of the interface contains three salt bridges, Glu74 to Arg132, Glu91 to Arg100 and Glu184 to Arg178. The 91/100 salt bridge is located at the base of the canyon and is particularly well conserved. Other

**Table III.** Statistics of the final model

Resolution range (Å)	15.0–1.7
Sigma cut-off applied	0.0
No. of reflections used in refinement	21 044
Reflections used in $R_{\text{free}}$ calculation (%)	10.0
Completeness of data in resolution range (%)	98.1
No. of protein atoms	1309
No. of water molecules	243
No. of sulfate molecules	6
$R$ factor (%)	24.4
Free $R$ factor (%)	30.0
r.m.s. deviations from ideal stereochemistry:	
bond lengths (Å)	0.011
bond angles (°)	2.123
dihedrals (°)	21.67
impropers (°)	1.462
Mean $B$ factor (main chain) (Å <sup>2</sup> )	25.67
r.m.s. deviation in main chain $B$ factor (Å <sup>2</sup> )	1.18
Mean $B$ factor (side chains and waters) (Å <sup>2</sup> )	30.18
r.m.s. deviation in side chain $B$ factors (Å <sup>2</sup> )	1.69
Ramachandran plot:	
residues in most favored region (%)	93.6
residues in additionally allowed regions (%)	5.7

features of the interface that may contribute to interdomain stability are an extensive network of ordered water molecules in the canyon, and two proline residues in the loop connecting  $\beta 5$  with  $\alpha 7$ .

### Sites of interaction with RNA

In our previous structural analyses of ribosomal proteins, we have used a phylogenetic approach to deduce the likely regions of contact with RNA. There is a great deal of sequence data available for most of the ribosomal proteins, and these make it possible to look for conserved aromatic, basic or hydrophobic residues on the surfaces of the proteins. The validity of this approach has been vindicated by independent studies in which the RNA-binding regions have been identified by more direct methods such as mutagenesis, protein–RNA cross-linking, hydroxyl radical footprinting and NMR (Urlaub *et al.*, 1995; Adamski *et al.*, 1996; Heilek and Noller, 1996; Hinck *et al.*, 1997). When the sequence conservation pattern of S4 is mapped onto the crystal structure, an extensive potential RNA-binding surface is apparent that covers one side of the protein (Figure 4). This entire surface, which is spread across both domains and which includes the interdomain canyon, is concave and has a distinct positive electrostatic potential (Figure 5). This is in contrast to the opposite surface which is convex and highly electronegative (Figure 5). This putative RNA-binding surface of S4 $\Delta$ 41 can be loosely divided into three regions: two major regions in domain 1 and at the domain interface, and a minor region in domain 2.

The first region covers most of the lower half of domain 1 and includes conserved amino acids from  $\alpha 1$ , the N-terminal end of  $\alpha 2$  and the C-terminal end of  $\alpha 7$ . Each of the residues along the solvent-exposed side of  $\alpha 1$  is well conserved and a candidate for interaction with RNA. Three aromatic residues, Tyr47, Tyr198 and Phe197, dominate the region and are contiguous with the aromatic side chains in the hydrophobic core of the domain. Other conserved residues in the region include Lys43, Lys56, Gln52, Gln55, Arg69, Arg200, Asn190 and Ser199. Tyr47 interacts closely with Leu51 on adjacent, exposed turns

of  $\alpha 1$ , and this pair appears to be particularly crucial to the region since both are almost completely conserved in bacterial-type sequences of S4. The importance of Tyr47 as an RNA-binding residue is also supported by a mutation of this residue to aspartate, in a strain of *E. coli* that is cold-sensitive and which has altered release factor 1-binding properties (A.E.Dahlgren and M.Rydén-Aulin, unpublished).

The most distinctive potential RNA-binding region is centered on the interdomain canyon which covers approximately two-thirds of the circumference of S4. It is possible that either an extended region of RNA binds into the interface and bends around the protein or that several regions of RNA converge on this central portion of S4 from different directions. These scenarios are supported by the surface electrostatic potential of S4 which shows that the highest concentration of basic charge is around the waist of the molecule (Figure 5). Many conserved, putative RNA-binding amino acid side chains project into this canyon from both domains. From domain 1, the residues are Arg58, Arg66, Arg93, Lys70, Lys77, His59, Asn64 and Gln67. From domain 2, they are Tyr99, Tyr131, Arg100, Arg107, Arg108, Arg111, Thr106, Thr115, Gln112, His116, Asn127 and Ser130. Pro79 is immediately adjacent to this region, and in *E. coli* its equivalent is Lys82 which has been cross-linked to rRNA (Wittmann-Liebold *et al.*, 1995). The 3/10 helix (residues 181–185) may also have a role in binding RNA that is specific to prokaryotes. This region is highly conserved only within prokaryotes (Figure 3) and, as mentioned previously, is tightly constrained by two salt bridges. Two residues, Glu181 and Ser183, protrude from this helix and may contact RNA.

Two adjacent arginines, 93 and 111, warrant particular attention since they are conserved in all known species of S4 including eukaryotes and Archaea. The orientation of Arg93 appears to be constrained by an electrostatic interaction with Asp95 which is totally conserved in prokaryotes (Figure 6). Interestingly, these arginines interact with a sulfate ion, and it is tempting to speculate that this mimics a phosphate group on the backbone of bound RNA. A similar situation occurs with Arg100 and Arg107 which are also located in this region. Since the canyon is lined by so many hydrogen bond donors and acceptors, it is not surprisingly that the region also contains a large number of ordered water molecules. These are presumably displaced when RNA is bound.

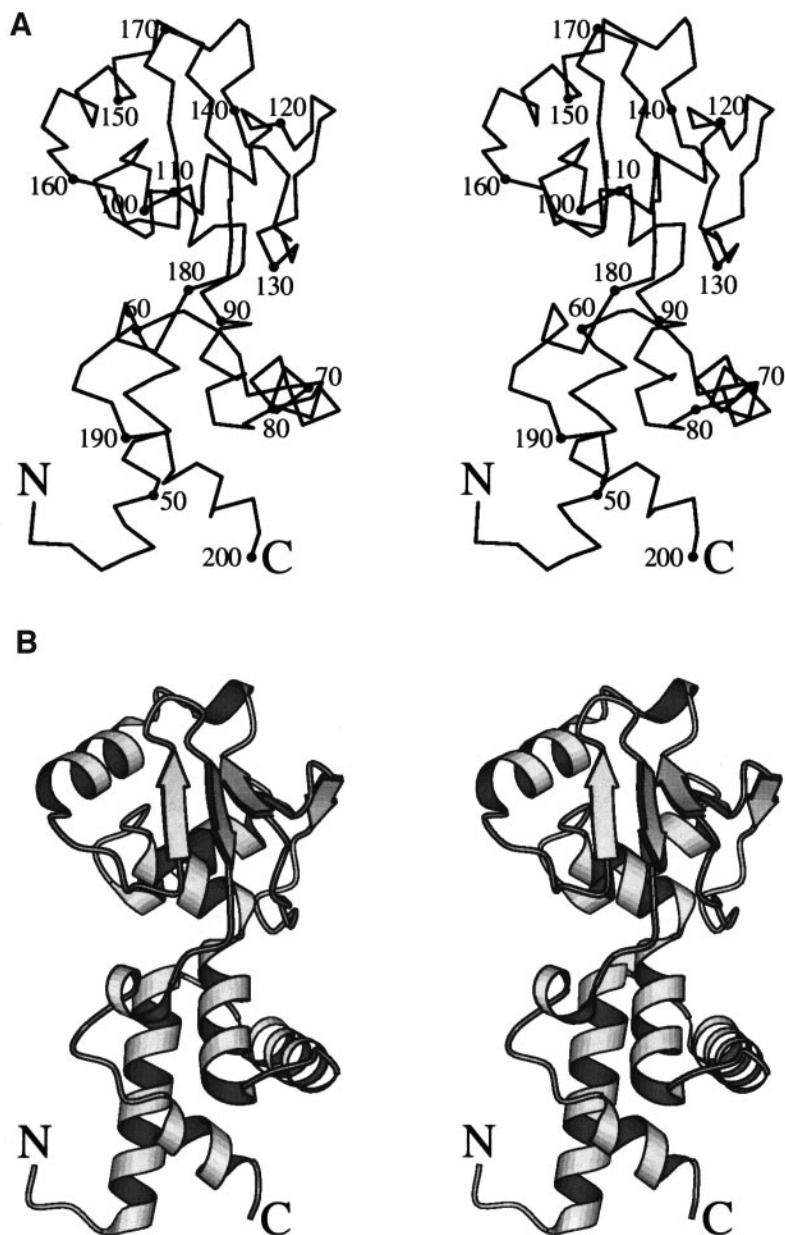
The third region is located on domain 2, at the top of the molecule (Figure 4). It comprises the adjacent His118, and Arg125 and 142, and compared with the other two probably represents a minor site of interaction with RNA.

### Structural homology to other proteins

As noted earlier, the majority of ribosomal protein structures have been found to be homologous to many types of protein families, and they may represent structural prototypes. To search for known protein structures that are homologous to S4, the coordinates were submitted to the Dali server (Holm and Sander, 1995). Two significant matches were obtained, one for each domain.

As shown in Figure 7, helices  $\alpha 2$ ,  $\alpha 3$ , 3/10 and  $\alpha 7$  of domain 1 can be superimposed on helices  $\alpha 1$ – $\alpha 4$  of the N-terminal domain of the Tet repressor (Hinrichs *et al.*,



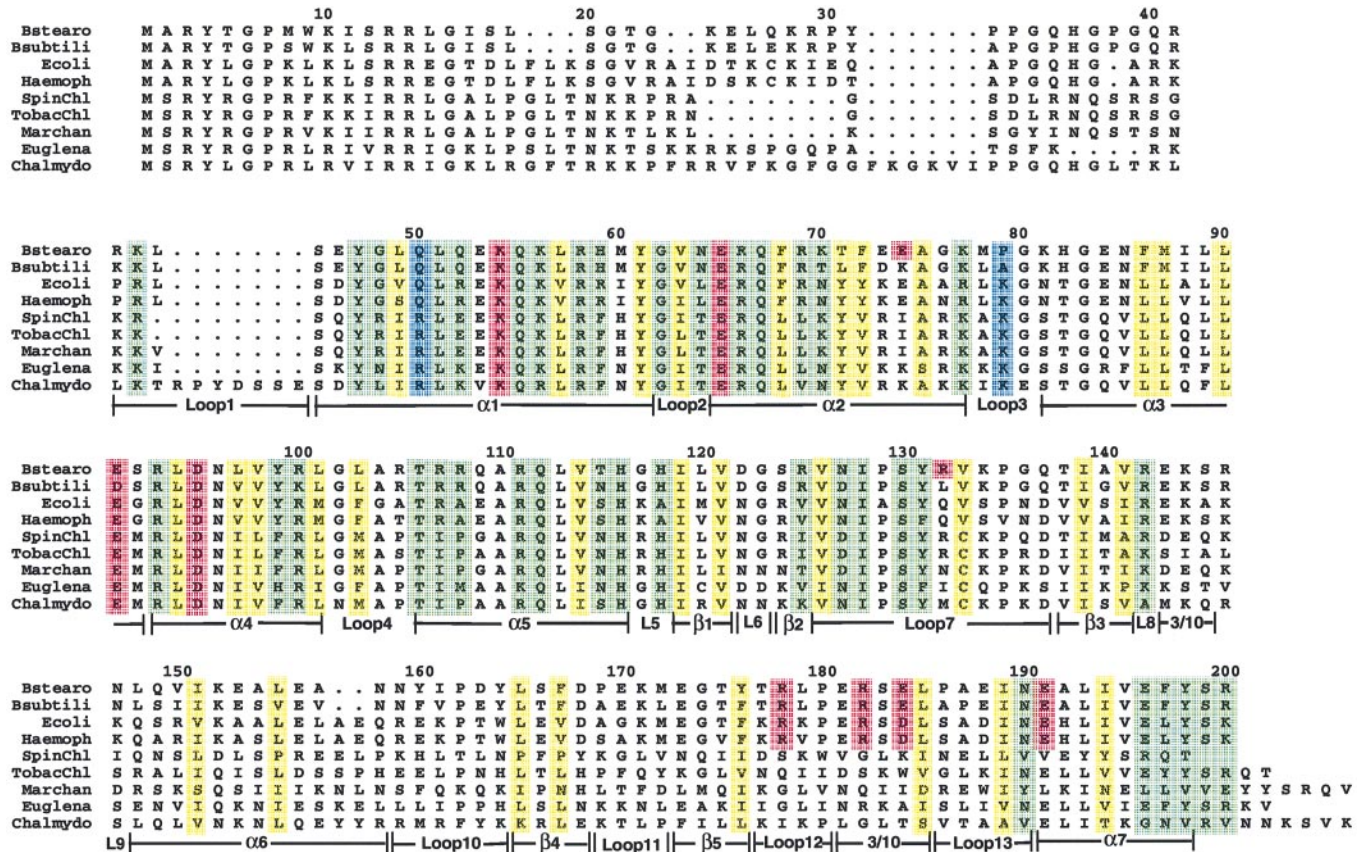


**Fig. 2.** The overall structure of ribosomal protein S4 $\Delta$ 41. (A) A stereo C $\alpha$  trace of the S4 $\Delta$ 41 backbone with every tenth residue labeled and marked with a filled circle. (B) A stereo ribbon representation of S4 $\Delta$ 41 showing the elements of secondary structure determined using PROMOTIF (Hutchinson and Thornton, 1996). The figure was produced using MOLSCRIPT (Kraulis, 1991).

1994). The first three helices of the Tet repressor form the DNA-binding domain which includes a helix–turn–helix (HTH) motif ( $\alpha$ 2–turn– $\alpha$ 3). The major difference between the structures is that the loop connecting the second and third  $\alpha$ -helices of the Tet repressor is replaced by the entire domain 2 of S4. Helical bundles are rather common structural motifs in proteins, and the significance of this homology is difficult to assess. For example, similar arrangements of  $\alpha$ -helices occur in such diverse proteins as the  $\sigma^{70}$  subunit of RNA polymerase (Malhotra *et al.*, 1996), type 1 cytochromes (Dickerson *et al.*, 1971), lysozyme (Blake *et al.*, 1965), the myosin motor domain (Rayment *et al.*, 1993), c-Myb (Ogata *et al.*, 1992) and the finger domain of Klenow (Ollis *et al.*, 1985). In myosin, the equivalent structure (residues 508–553) forms the interaction site with actin (Rayment *et al.*, 1993) and,

in Klenow (residues 751–793) it forms a section of the variable part of the RNA recognition motif (RRM). None of these structural similarities are accompanied by sequence similarities.

Domain 2 of S4 $\Delta$ 41 has an unexpectedly close relationship to the ETS domain, which is the distinguishing characteristic of the ets family of eukaryotic transcription factors. The two domains have a very similar architecture of  $\alpha$ -helices packed onto one surface of an antiparallel  $\beta$ -sheet, and the topologies of the secondary structure elements are identical (Figure 7). When optimally superimposed, the overall r.m.s. deviation in C $\alpha$  positions is 3.3 Å. The correspondence of the two  $\beta$ -sheets is particularly close, with an r.m.s. deviation of 1.6 Å. Despite this similarity, there is no detectable sequence homology between the two domains, and three invariant tryptophans



**Fig. 3.** An alignment of representative sequences of ribosomal protein S4 from bacteria and chloroplasts. The abbreviations for each are as follows: Bstearo, *B.stearothermophilus* (S.E.Gerchmann and V.Ramakrishnan, unpublished); Bsubtili, *B.subtilis* (Grundy and Henkin, 1990); Ecoli, *E.coli* (Thomas *et al.*, 1987); Haemoph, *Haemophilus influenza* (Fleischmann *et al.*, 1995); SpinChl, spinach chloroplast (Ben Tahar *et al.*, 1986); TobacChl, tobacco chloroplast (Shinozaki *et al.*, 1986); Marchan, *Marchantia polymorpha* chloroplast (Ohyama *et al.*, 1986); Euglena, *Euglena chloroplast* (Stevenson *et al.*, 1991); Chalmydo, *Chlamydomonas reinhardtii* chloroplast (Randolph-Anderson *et al.*, 1995). For clarity, 31 residues at the C-terminus of the *C.reinhardtii* chloroplast sequence, which are unique to this species, have been omitted. The numbering corresponds to the *B.stearothermophilus* protein, and the regions of secondary structure within S4Δ41 are indicated. Blocks of amino acids are color coded as follows: yellow, hydrophobic core residues; red, salt bridges; green, putative RNA-binding residues; and blue, residues cross-linked to other ribosomal components and sites of mutation. The alignment was performed using PILEUP [Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI].

that form part of the hydrophobic core in ETS domains (Donaldson *et al.*, 1996) are not observed in S4. The ETS domain is a member of the winged HTH (wHTH) superfamily that interact with DNA via the second  $\alpha$ -helix and the adjacent ‘wing’ or loop that bind at consecutive major and minor grooves, respectively. S4 differs from the ETS domain primarily in the HTH region. The first ‘supporting’  $\alpha$ -helix is replaced by a short region of 3/10 helix which, together with a shorter connecting loop, results in a more direct connection to the second ‘recognition’  $\alpha$  helix ( $\alpha_6$ ). The wing of the ETS domain corresponds to the loop between  $\beta_4$  and  $\beta_5$  in S4.

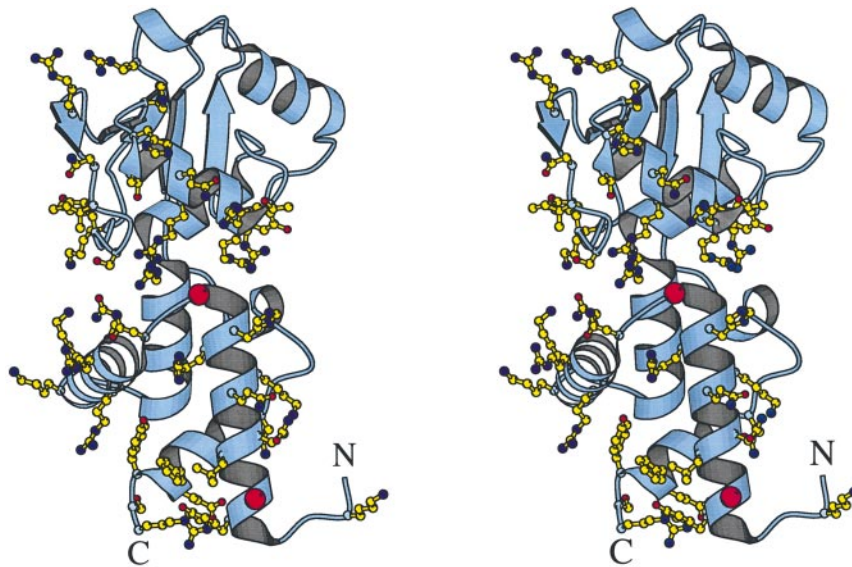
**Discussion**

**Interaction of S4 with RNA**

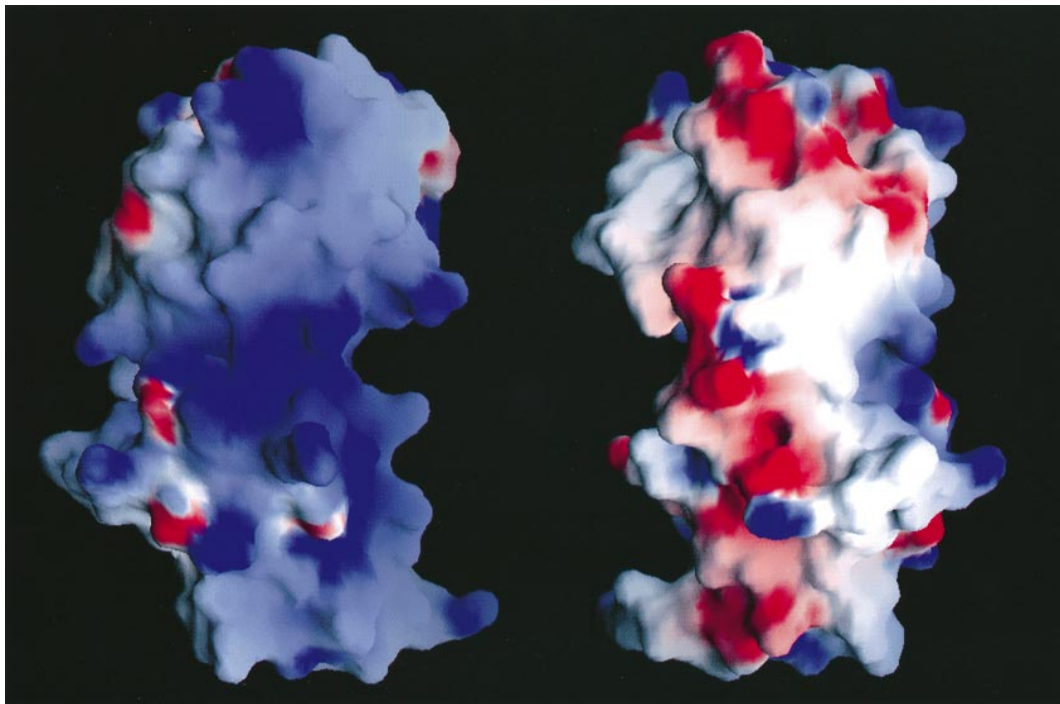
It has been known for some time that S4 organizes and promotes the folding of a substantial fraction of the 16S rRNA molecule (Zimmermann *et al.*, 1975). S4 also has a major role in the translational feedback process that coordinates the synthesis of ribosomal proteins and rRNA (Yates *et al.*, 1980). The S4 gene is located within the  $\alpha$  operon of *E.coli*, and S4 specifically binds to a region of

the polycistronic mRNA that contains the translational initiation site for the first gene (S13). In a series of experiments, Draper and co-workers have shown that the site folds into a complicated double pseudoknot structure that is stabilized by S4 (Deckman and Draper, 1985, 1987a,b; Tang and Draper, 1989, 1990; Gluick *et al.*, 1997). Translational control is thought to occur via an allosteric mechanism in which S4 traps the mRNA in a conformation that is able to bind 30S subunits, but unable to form an initiation complex (Spedding and Draper, 1993). Several other primary ribosomal proteins perform a similar function within their operons, and their rRNA- and mRNA-binding sites are generally very similar (Draper, 1989). This is apparently not the case with S4, although the sites may have common three-dimensional structures that are not obvious from the secondary structures. Clearly, the three-dimensional structures of the S4-rRNA and S4-mRNA complexes are very complicated and impossible to model with any degree of confidence. However, there have been a number of studies of these interactions, and these can now be re-evaluated in the light of the crystal structure.

Craven and co-workers studied a number of *E.coli* S4



**Fig. 4.** A stereoview of ribosomal protein S4Δ41 showing the overall distribution of residues believed to be involved in mediating interactions with rRNA. The residues (in standard coloring) are distributed throughout this entire face of the molecule, but can be grouped into three clusters: domain 1 (bottom), domain 2 (top) and the domain interface. The orientation of the molecule is identical to that shown in Figure 2. The figure was produced using MOLSCRIPT (Kraulis, 1991).

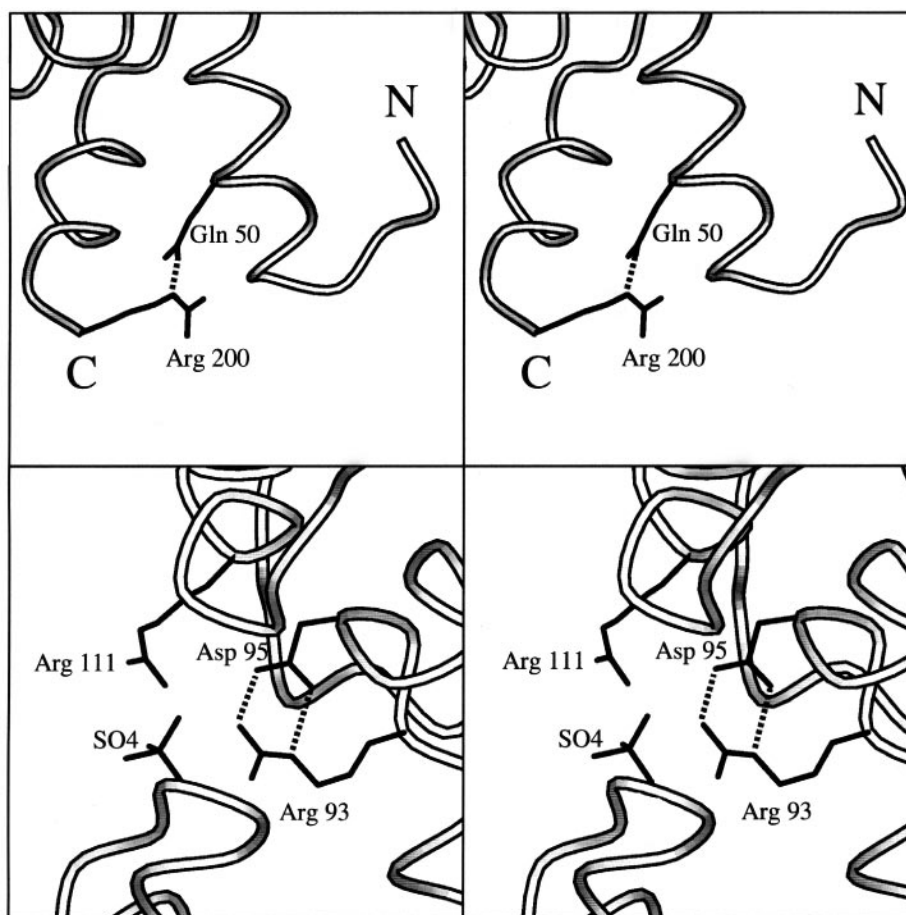


**Fig. 5.** The surface electrostatic potential of ribosomal S4Δ41. The view on the left corresponds to the putative RNA-binding surface shown in Figure 4 and has a high overall positive potential. The view on the right corresponds to the opposite surface and is generally electronegative. The extreme ranges of red (negative) and blue (positive) represent electrostatic potentials of less than  $-9$  to greater than  $+9 k_bT$ , where  $k_b$  is the Boltzmann constant and T is the temperature. The figure was calculated using the GRASP program (Nicholls *et al.*, 1991).

fragments to determine the minimal region required for binding rRNA and promoting 30S assembly. Trypsin digestion of an S4-rRNA complex resulted in loss of the first 46 residues; the remaining fragment was able to bind rRNA and promote normal 30S subunit assembly (Changchien and Craven, 1976). These experiments suggested that the equivalent N-terminal deletion in *B.stearotherophilus* S4, the S4Δ41 of the present study, would yield a stably folded protein, which is indeed the case.

Fragments terminating at residues 171, 119 and 101 (*B.stearotherophilus* numbering) were also made. All retained some affinity for rRNA (Changchien and Craven, 1985, 1986; Conrad and Craven, 1987). However, later studies using circular dichroism showed that any truncation of the molecule prior to residue 171 resulted in a marked destabilization (Baker and Draper, 1995). From the structure, it is now clear that in such fragments domain 2 cannot fold properly. In addition, the absence of the final





**Fig. 6.** Stereoviews of two significant regions of ribosomal protein S4 $\Delta$ 41 discussed in the text. (Top) Gln50 within domain 1 has been identified as the site of a point mutation that results in a *ram* phenotype. It is involved in a hydrogen-bonded interaction with the conserved Arg200, and appears to orient it precisely for interaction with RNA. (Bottom) Arg93 and Arg111 are highly conserved and within the putative RNA-binding site at the domain interface. Conserved Asp95 is mostly buried and appears to orient the adjacent basic residues. A strongly bound sulfate group may reflect the position of a phosphate group when RNA is bound in this region.

helix in domain 1,  $\alpha$ 7, would probably cause a dramatic alteration in the packing arrangement of  $\alpha$ 1 and  $\alpha$ 2, both of which probably contact RNA. It is remarkable that such fragments retain the ability to recognize a specific rRNA site, even though a hydrophobic core is lacking.

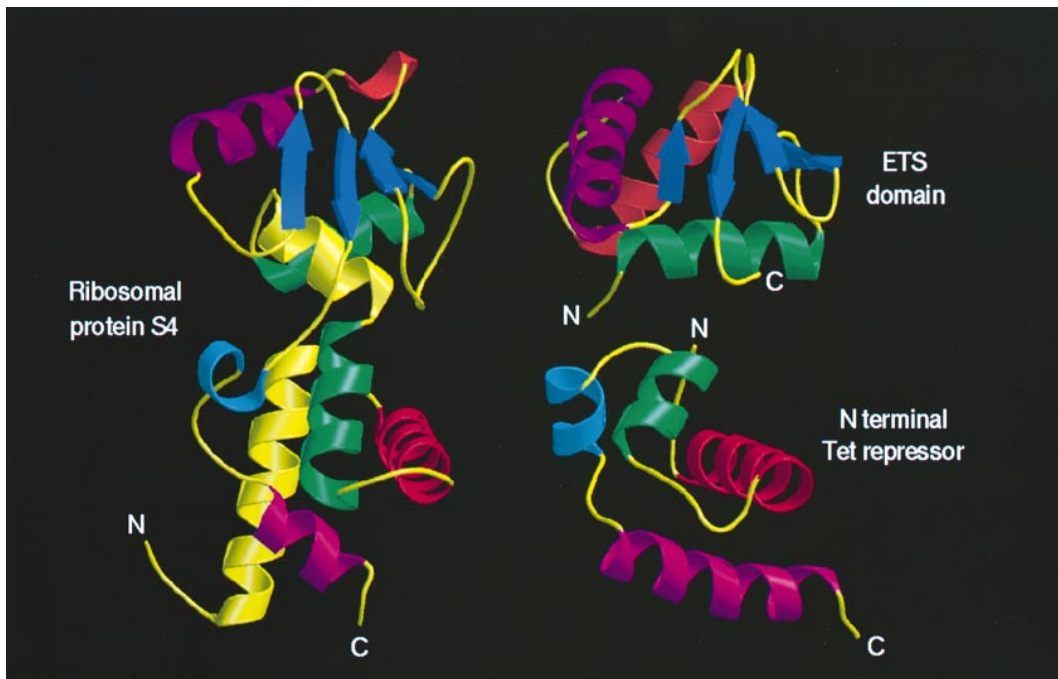
S4 deletions that affect rRNA recognition have parallel effects on mRNA binding (Baker and Draper, 1995), suggesting that the two binding activities utilize the same regions of S4. The concentration of potential RNA-binding residues on one side of the protein certainly supports this idea. The same mRNA binding studies also suggested that residues 43–101 stabilize different mRNA secondary structures than 102–171 (*B. stearothermophilus* numbering). This is consistent with domain 1 and domain 2 recognizing different regions of the mRNA. It is not yet possible to assess in detail whether the mRNA and rRNA structures bind in a similar fashion, but with the structure in hand it will now be possible to address this question directly by site-directed mutagenesis. Efforts are also underway to co-crystallize S4 with cognate RNA fragments.

In terms of its overall architecture and putative interaction with rRNA, S4 most closely resembles ribosomal protein L1 (Nikonov *et al.*, 1996). L1 also comprises two domains that appear to have arisen by gene insertion, and

has a major RNA-binding site at the domain interface. It has been demonstrated that L1 has considerable interdomain flexibility that may allow the protein to adopt open and closed conformations during the RNA-binding process. Unlike L1, where the crystals had a large variation in unit cell parameters related to the flexibility, S4 shows no crystallographic evidence of interdomain movement. However, the packing arrangement in the S4 crystal, where both domains are fixed by symmetry-related molecules, may preclude possible movements. The major interdomain connection in S4 is via the helices  $\alpha$ 3 and  $\alpha$ 4, which have the lowest main chain temperature factors in the entire molecule. In addition, there is also a clear interdomain hydrophobic core as described earlier. These would appear to rule out major movements of the domains. However, a relative twisting of the domains is still possible, and tentative evidence supporting this has been obtained from a parallel NMR analysis of the S4 structure (Markus *et al.*, 1998).

#### **S4 mutations and RNA binding**

We have shown for a number of ribosomal proteins that mutations which produce defined ribosomal phenotypes are invariably within their RNA-binding sites (Ramakrishnan and White, 1998). These mutations appar-



**Fig. 7.** The structural homology of ribosomal protein S4 $\Delta$ 41 to regions of other proteins. Domain 1 of S4 $\Delta$ 41 (bottom) is homologous to the  $\alpha$ -helical DNA-binding domain of the tet repressor, and domain 2 (top) is homologous to the ETS domain. The corresponding elements of secondary structure in both homologies are colored identically. Note that these elements are not in precisely the same orientation, but the topological relationships in both cases are the same. The figures of each protein were produced using MOLSCRIPT (Kraulis, 1991) and Raster3D (Merritt and Bacon, 1997).

ently exert their effects indirectly by perturbing the local rRNA structure. Mutations in *E.coli* S4 were amongst the first to be discovered, and they were found to modulate the accuracy of protein synthesis. They initially were characterized as revertants to non-dependence on streptomycin (Birge and Kurland, 1970) and later identified as *ram* mutants that lower the intrinsic accuracy of translation (van Acken, 1975). S4, together with S5, S12 and the 900 and 530 regions of 16S rRNA are now collectively known as the 'accuracy domain' which is also the site of streptomycin binding (Kurland *et al.*, 1990). Green and Kurland (1971) were the first to demonstrate that S4 mutants have defective rRNA binding, and Noller and co-workers (Allen and Noller, 1989) have shown that S4 *ram* mutants perturb rRNA structure *in situ*. Clearly, these sites of mutations are an excellent indicator of RNA-binding sites.

Most of the S4 *ram* mutants are truncations of the protein that result in defective rRNA binding and poor cell growth (Daya-Grosjean *et al.*, 1972; Funatsu *et al.*, 1972). The truncations occur around residue 175 (A.E. Dahlgren and M.Rydén-Aulin, unpublished), and would allow domain 2 but not domain 1 to fold correctly. This indicates that the putative RNA-binding site in domain 1 is important for interacting with rRNA, and is supported by the observation that *ram* mutants with unaltered lengths show wild-type rRNA binding (Daya-Grosjean *et al.*, 1972). One such mutant has been characterized as a glutamine to leucine change at *E.coli* position 53 (van Acken, 1975). This residue corresponds to Gln50 in the *B.stearothermophilus* molecule, and is highly conserved in prokaryotic sequences. In the structure, the residue forms a hydrogen bond interaction with the highly conserved Arg200 that we have identified as a probable RNA-

binding element, and the interaction appears to orient the guanadinium group precisely (Figure 6). Therefore, the mutation also strongly suggests that the domain 1 RNA-binding site is crucial to the function of S4.

#### **Structural homologies to S4**

Most of the ribosomal protein structures have been found to be structurally homologous to other families of proteins, including those that interact with DNA and RNA (Ramakrishnan and White, 1998). We have suggested that these ancient proteins discovered successful structural prototypes that have been retained throughout evolution. In certain cases, these homologous proteins have proved to be useful indicators of how the ribosomal proteins bind RNA, since the structures of a number of them have been analyzed bound to nucleic acids. None of the putative ribosomal protein-rRNA complexes based on these homologies has been confirmed by detailed structural analysis. However, an NMR analysis of the L11-rRNA complex does support the homology to the homeodomain-DNA complex (Hinck *et al.*, 1997). Each domain of S4 has an intriguing homology to a DNA-binding protein motif, but a closer inspection of the similarities reveals that they are unlikely to represent useful models for how S4 interacts with RNA.

Domain 1 resembles the DNA-binding domain of the Tet repressor which uses a HTH region to interact with the major groove of DNA. However, the corresponding HTH region of S4 is interrupted by the entire domain 2 and the second helix is sterically hindered from binding RNA to any great extent. At first sight, the homology of domain 2 to the ETS domain appears to be a better candidate for modeling. The ETS domain is a member of the 'winged-helix' family of DNA-binding proteins that

also use a HTH motif to interact with the DNA major groove, and one or two loops or wings to bind to the adjacent minor groove (Brennan, 1993). Domain 2 of S4 and the ETS domain both have a single wing, and by superimposing domain 2 onto the ETS domain within the PU.1–DNA co-crystal structure (Kodandapani *et al.*, 1996), it was possible to investigate its putative interaction with a region of double-stranded RNA. The model is generally unconvincing, principally because the RNA site does not match any of the sites suggested by the pattern of sequence conservation. In particular  $\alpha 6$ , which would be the principal ‘recognition’ helix, is in the most weakly conserved region of S4. A closer examination of the S4/ETS homology does reveal an important difference in the HTH motif that also argues against the model. In the ETS–DNA complex, the first helix of the HTH motif acts as a platform for the second ‘recognition’ helix and orients it appropriately with respect to the adjacent wing (Kodandapani *et al.*, 1996). In domain 1 of S4, this precise orientation is absent, and the first helix is a short 3/10 structure that does not act as a platform.

### **S4 and the 30S subunit**

Early studies on the assembly of the 30S subunit demonstrated that S4 has a pivotal role in the early stages of the pathway (Held *et al.*, 1974). S4 is now recognized as a primary RNA-binding protein that specifically binds a region of 16S rRNA and, together with S7, initiates the folding of the molecule into its functional conformation (Nowotny and Nierhaus, 1988). The general location of the S4-binding site on the 16S rRNA molecule has been studied extensively by a number of techniques including base probing (Stern *et al.*, 1986), hydroxyl-radical footprinting (Powers and Noller, 1995a,b), protein–rRNA cross-linking (Greuer *et al.*, 1987) and fragment binding (Vartikar and Draper, 1989; Sapag *et al.*, 1990). The site is within the 5′ domain at the junction of five helical regions encompassing nucleotides 27–47 and 395–556. Using Brimacombe’s numbering scheme (Brimacombe, 1991), these helices are 3, 4, 16, 17 and 18. The latest model of the folded 16S rRNA molecule does reveal a suitable clustering of the S4 footprint and cross-linking sites (Mueller and Brimacombe, 1997), and the S4 structure should now facilitate a more detailed modeling of the S4 environment. High resolution structures of ribosomal proteins together with associated biochemical data have provided important restraints for the evolving models of the ribosome (Mueller and Brimacombe, 1997; Davies *et al.*, 1998; Tanaka *et al.*, 1998).

The complicated pattern of S4–rRNA footprints agrees well with the structure of S4 $\Delta$ 41 which has the general appearance of binding extensively with RNA. The distinctive difference in charge between the two surfaces of S4 suggests that one entire surface of the protein faces the 30S subunit and interacts mainly with 16S rRNA, while the opposite face is exposed to the exterior. Consistent with this, the majority of the putative RNA-binding residues are on the former surface and the latter surface is poorly conserved. This general orientation would suggest a large electrostatic contribution to the S4–rRNA interaction which has been observed (Deckman and Draper, 1985). It also suggests that S4 has minimal interaction with other ribosomal proteins. Hydrophobic patches that

are potential sites of protein–protein interactions have been observed on S5 (Davies *et al.*, 1998), S8 (Davies *et al.*, 1996b) and L14 (Davies *et al.*, 1996a), but are not present in S4 $\Delta$ 41. The role of the N-terminal region of S4 in binding RNA should not be ignored, especially since the construct of S4 used in this study has a somewhat lower affinity for both rRNA and mRNA compared with wild-type S4 (R.B.Gerstner and D.E.Draper, unpublished). Indeed, the first 16 residues of S4 are highly conserved (Figure 3) and may contact RNA.

The positioning of S4 with respect to the 16S rRNA can be refined further from the results of recent, more precise experiments. Directed hydroxyl radical footprinting of 16S rRNA with an Fe-EDTA moiety attached to Cys31 in *E.coli* cleaves at nucleotides 419–432 and 297–303 (Heilek *et al.*, 1995). This experiment suggests that the N-terminal domain of S4 is oriented towards helices 16 and 13. Also, Wittmann-Liebold and co-workers (Urlaub *et al.*, 1995, 1997) have identified S4–cross-links involving *E.coli* residues Lys44 and Lys82 which correspond to *B.stearothermophilus* residues Gln40 and Pro79, respectively. The first location is at the periphery of the interface RNA-binding site, and the second, although not included in our construct, is near the domain 1 RNA-binding site. Thus, as with proteins S8, S17 and L6, these cross-linking data firmly support the RNA-binding sites deduced from the structures. These types of data currently are being used in modeling studies of the 30S subunit which have yet to be completed. Functionally, S4 is located in the ‘accuracy domain’ that also includes S5, S12 and the 530 and 900 regions of 16S rRNA. This domain originally was thought to be distant from the decoding site, but recent data suggest that they may be adjacent or contiguous (reviewed in Davies *et al.*, 1998). This is an important question with fundamental implications for the mechanism of translation which has yet to be resolved satisfactorily. The crystal structure of S5 has had a major impact on this problem (Davies *et al.*, 1998), and has enabled a number of precise RNA footprinting experiments to be performed (Heilek and Noller, 1996). Similar experiments will now be possible using S4 as a site-directed probe.

## **Materials and methods**

### **Cloning and protein purification**

The procedure for cloning and sequencing the full-length S4 gene from *B.stearothermophilus* using the T7 system was the same as that described previously (Davies *et al.*, 1996a). The gene was inserted into the pET13a vector using the standard *Nde*I and *Bam*HI sites at the 5′ and 3′ ends, respectively. To obtain the S4 $\Delta$ 41 fragment, the pET13a vector was cut with *Nde*I which excised a region spanning the initiation codon and nucleotide 177 of the coding region. This was replaced by a synthetic 54 bp fragment which resulted in the deletion of residues 2–42. *Escherichia coli* cells incorporating the T7 expression system [BL21(DE3)] and containing the S4 $\Delta$ 41 pET13a expression vector (Novagen) were grown in Luria broth medium (25 g/l) and 25 mg/l kanamycin. Batches of 2 l were inoculated with 3 ml of an overnight culture and grown until the OD<sub>550</sub> reached 0.6, at which point the cells were induced by adding isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) to a final concentration of 0.4 mM. After 3 h, the cells were harvested by centrifugation at 2000 r.p.m. and resuspended in 40 ml of a buffer containing 25 mM phosphate pH 7.2, 1 mM EDTA, 10 mM  $\beta$ -mercaptoethanol and 50 mM potassium chloride. The sample was frozen at –20°C after adding 2 mg of DNase to promote DNA cleavage, and then subjected to several freeze–thaw cycles to lyse the cells. The

lysed cells were spun at 20 000 r.p.m. for 30 min, and the supernatant was diluted by an equal volume of buffer containing 50 mM Tris pH 8.0, 50 mM NaCl, 0.5 mM EDTA and protease inhibitors. This was applied to an S-Sepharose column (Pharmacia) equilibrated in the same buffer, and the bound protein was eluted by applying a salt gradient of 0–1.0 M NaCl. SDS-PAGE showed that S4Δ41 eluted as a single pure peak at 0.5 M NaCl.

### Crystallization

The extinction coefficient of S4Δ41 was estimated to be 1.0 based on the amino acid sequence. The protein was concentrated to 40 mg/ml using Centricon-10 microconcentrators (Amicon) and subjected to crystallization trials using the hanging drop method (Davies and Segal, 1971). In these trials, 3 μl of the protein solution was mixed with 3 μl of the well solution and the dishes were stored at 22°C. Initial trials used the Crystal Screens from Hampton Research.

### Data collection

Diffraction data were collected using a DIP2030 area detector system (MacScience) mounted on a Nonius FR591 X-ray generator operating at 45 kV and 90 mA and equipped with focusing mirrors (MacScience). Data were collected at room temperature by the standard oscillation method using a crystal-detector distance of 130 mm. For the derivative search, crystals were soaked in heavy atom solutions of varying concentrations and data were collected with an oscillation angle of 2 or 3° (depending on the precise crystal orientation), with an exposure time of 6 min/degree. Crystals typically were rotated through ~100° to collect essentially complete datasets. The crystals generally were stable in the X-ray beam and freezing was not necessary.

For model refinement purposes, a high resolution (1.65 Å) dataset was collected at 100 K using flash-cooled crystals that had been soaked in mother liquor containing 20% glycerol as cryoprotectant. In this case, the crystal-detector distance was 100 mm, with an oscillation angle of 1° and an exposure time of 6 min/degree. To ensure a high redundancy of data, the crystal was rotated through a total of 176°. The space group and cell dimensions were determined, and oscillation data were processed using HKL (Otwinowski, 1993). In the case of derivative data, the Friedel pairs were not merged in order to measure the anomalous signal.

### Phasing

Data were scaled and merged, and potential derivatives were identified by Patterson methods. Several derivatives were identified, many of which had bound to the same position in the crystal lattice. In the case of one derivative, tetracyanoplatinate, the usual soaking buffer was replaced by 2.1 M lithium sulfate, 100 mM Tris pH 8.5. The heavy atom positions were determined by manual inspection of the difference Patterson maps, and the parameters were refined for each individual dataset prior to inclusion in the phasing calculation. Anomalous data for the derivatives were included if significant and consistent peaks were present in the anomalous difference Patterson. Each derivative was confirmed by cross-difference Fourier procedures. The final MIR phasing calculation used a total of four derivatives, all of which included anomalous data. The resulting phases were improved by solvent flattening where the solvent content was assumed to be 50%. All calculations were performed using PHASES (Furey and Swaminathan, 1996).

### Model building and refinement

An MIR electron density map was calculated using CCP4 programs (CCP4, 1994) and displayed using the 'O' program (Jones *et al.*, 1991). This map was of sufficient quality to construct a complete model of S4Δ41, complete with side chains. The model was refined by alternating rounds of XPLOR (Brünger *et al.*, 1987) and manual revision using 'O'. Water and sulfate molecules were included during the latter stages of the refinement. The final round of refinement was performed using REFMAC (CCP4, 1994). The stereochemistry of the final model was examined and evaluated using PROCHECK (Laskowski *et al.*, 1993). The numbering of the final model corresponds to the *B.stearothermophilus* sequence (S.E.Gerchmann and V.Ramakrishnan, unpublished) which differs from the previously published sequence (Arndt *et al.*, 1991) at two positions. Residue 118 is a leucine rather than isoleucine, and there is an extra arginine residue between residues 41 and 42. Accordingly the model is numbered 42–200 beginning with Met42, the N-terminal residue of the S4Δ41 construct (equivalent to Arg42 in the wild-type sequence).

## Acknowledgements

We thank Michelle Markus and Dennis Torchia for providing the coordinates of the NMR structure of S4Δ41 prior to publication and for invaluable discussions. We are also grateful to Monica Rydén-Aulin for providing results in advance of publication and Sue Ellen Gerchman for technical assistance. This work was supported in part by NIH grants to S.W.W. and V.R. (GM44973), and to D.E.D. (GM56968). The authors acknowledge the support of the American Lebanese Syrian Associated Charities (ALSAC).

## References

- Adamski,F.M., Atkins,J.F. and Gesteland,R.F. (1996) Ribosomal protein L9 interactions with 23S rRNA: the use of translational bypass assay to study the effect of amino acid substitutions. *J. Mol. Biol.*, **261**, 357–371.
- Allen,P.N. and Noller,H.F. (1989) Mutations in ribosomal proteins S4 and S12 influence the higher order structure of 16S ribosomal RNA. *J. Mol. Biol.*, **208**, 457–468.
- Arndt,E., Scholzen,T., Kromer,W., Hatakeyama,T. and Kimura,M. (1991) Primary structures from the archaeobacterium *Halobacterium marismortui* and the eubacterium *Bacillus stearothermophilus*. *Biochimie*, **73**, 657–668.
- Baker,A.M. and Draper,D.E. (1995) Messenger RNA recognition by fragments of ribosomal proteins. *J. Biol. Chem.*, **270**, 22939–22945.
- Ben Tahar,S., Bottomley,W. and Whitfield,P.R. (1986) Characterization of the spinach chloroplast genes for the S4 ribosomal protein, transfer RNA-Thr and transfer RNA-Ser. *Plant Mol. Biol.*, **7**, 63–70.
- Birge,E.A. and Kurland,C.G. (1970) Reversion of a streptomycin-dependent strain of *Escherichia coli*. *Mol. Gen. Genet.*, **109**, 356–369.
- Blake,C.C., Koenig,D.F., Mair,G.A., North,A.C., Philips,D.C. and Sarma,V.R. (1965) Structure of hen egg-white lysozyme. A three dimensional Fourier synthesis at 2 Å resolution. *Nature*, **206**, 757–761.
- Brennan,R.G. (1993) The winged-helix DNA binding motif: another helix–turn–helix takeoff. *Cell*, **74**, 773–776.
- Brimacombe,R. (1991) RNA–protein interactions in the *Escherichia coli* ribosome. *Biochimie*, **73**, 927–936.
- Brünger,A.T., Kuriyan,J. and Karplus,M. (1987) Crystallographic *R* factor refinement by molecular dynamics. *J. Mol. Biol.*, **203**, 803–816.
- Burley,S.K. and Petsko,G.A. (1985) Aromatic–aromatic interaction: a mechanism of protein structure stabilization. *Science*, **229**, 23–28.
- Capel,M.S. *et al.* (1987) A complete mapping of the proteins in the small ribosomal subunit of *Escherichia coli*. *Science*, **238**, 1403–1406.
- Changchien,L.-M. and Craven,G.R. (1976) The function of the N-terminal region of ribosomal protein S4. *J. Mol. Biol.*, **108**, 381–401.
- Changchien,L.M. and Craven,G.R. (1985) Specific ribosomal RNA recognition by a fragment of *E.coli* ribosomal protein S4 missing the C-terminal 36 amino acid residues. *Nucleic Acids Res.*, **13**, 6343–6360.
- Changchien,L.M. and Craven,G.R. (1986) The use of hydroxylamine cleavage to produce a fragment of ribosomal protein S4 which retains the capacity to specifically bind 16S ribosomal RNA. *Nucleic Acids Res.*, **14**, 1957–1966.
- Clemons,W.M., Davies,C., White,S.W. and Ramakrishnan,V. (1998) Conformational variability of the N terminal helix in the structure of ribosomal protein S15. *Structure*, **6**, 429–438.
- Collaborative Computation Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.
- Conrad,R.C. and Craven,G.R. (1987) A cyanogen bromide fragment of S4 that specifically rebinds 16S RNA. *Nucleic Acids Res.*, **15**, 10331–10343.
- Dahlberg,A.E. (1989) The functional role of ribosomal RNA in protein synthesis. *Cell*, **57**, 525–529.
- Davies,C., White,S.W. and Ramakrishnan,V. (1996a) The crystal structure of ribosomal protein L14 reveals an important organizational component of the translational apparatus. *Structure*, **4**, 55–66.
- Davies,C., Ramakrishnan,V. and White,S.W. (1996b) Structural evidence for specific S8–RNA and S8–protein interactions within the 30S ribosomal subunit: ribosomal protein S8 from *Bacillus stearothermophilus* at 1.9 Å resolution. *Structure*, **4**, 1093–1104.
- Davies,C., Bussiere,D.E., Golden,B.L., Porter,S.J., Ramakrishnan,V. and White,S.W. (1998) Ribosomal proteins S5 and L6: high resolution crystal structures and roles in protein synthesis and antibiotic resistance. *J. Mol. Biol.*, **279**, 873–888.
- Davies,D.R. and Segal,D.M. (1971) Protein crystallization: micro techniques involving vapor diffusion. *Methods Enzymol.*, **22**, 266–269.



- Daya-Grosjean, L., Garrett, R.A., Pongs, O., Stöffler, G. and Wittmann, H.G. (1972) Properties of the interaction of ribosomal protein S4 and 16S RNA in *Escherichia coli* revertants from streptomycin dependence to independence. *Mol. Gen. Genet.*, **119**, 277–286.
- Dean, D. and Nomura, M. (1980) Feedback regulation of ribosomal protein gene expression in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **77**, 3590–3594.
- Deckman, I.C. and Draper, D.E. (1985) Specific interaction between ribosomal protein S4 and the alpha operon messenger RNA. *Biochemistry*, **24**, 7860–7865.
- Deckman, I.C. and Draper, D.E. (1987a) S4-alpha mRNA translation repression complex. I. Thermodynamics of formation. *J. Mol. Biol.*, **196**, 313–322.
- Deckman, I.C. and Draper, D.E. (1987b) S4-alpha mRNA translation repression complex. II. Secondary structures of the RNA regulatory cycle in the presence and absence of S4. *J. Mol. Biol.*, **196**, 323–332.
- Dickerson, R.E., Takano, T., Eisenberg, D., Kallai, O.B., Samson, L., Cooper, A. and Margoliash, E. (1971) Ferricytochrome c. I. General features of the horse and bonito proteins at 2.8 Å resolution. *J. Biol. Chem.*, **246**, 1511–1535.
- Donaldson, L.W., Peterson, J.M., Graves, B.J. and McIntosh, L.P. (1996) Solution structure of the ETS domain from murine Ets-1: a winged helix–turn–helix DNA binding motif. *EMBO J.*, **15**, 125–134.
- Draper, D.E. (1989) How do proteins recognize specific RNA sites? New clues from autogenously regulated ribosomal protein complexes. *Trends Biochem. Sci.*, **14**, 335–338.
- Ehresmann, C., Stiegler, P., Carbon, P., Ungewickell, E. and Garrett, R.A. (1977) A topographical study of the 5'-region of 16S RNA of *Escherichia coli* in the presence and absence of ribosomal protein S4. *FEBS Lett.*, **81**, 188–192.
- Fleischmann, R.D. et al. (1995) Whole-genome sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, **269**, 496–512.
- Funatsu, G., Puls, W., Schlitz, E., Reinbolt, J. and Wittmann, H.G. (1972) Ribosomal proteins. XXXI. Comparative studies on altered proteins S4 of six *Escherichia coli* revertants from streptomycin dependence. *Mol. Gen. Genet.*, **115**, 131–139.
- Furey, W. and Swaminathan, S. (1996) PHASES-95: a program package for processing and analyzing diffraction data from macromolecules. *Methods Enzymol.*, **277**, 590–620.
- Gluck, T.C., Gerstner, R.B. and Draper, D.E. (1997) Effects of Mg<sup>2+</sup>, K<sup>+</sup>, and H<sup>+</sup> on an equilibrium between alternative conformations of an RNA pseudoknot. *J. Mol. Biol.*, **270**, 451–463.
- Green, M. and Kurland, C.G. (1971) Mutant ribosomal protein with defective RNA binding site. *Nature New Biol.*, **234**, 273–275.
- Greuer, B., Osswald, M., Brimacombe, R. and Stöffler, G. (1987) RNA–protein cross-linking in *Escherichia coli* 30S ribosomal subunits: determination of sites on 16S RNA that are cross-linked to proteins S3, S4, S7, S9, S10, S11, S17, S18 and S21 by treatment with bis-(2-chloroethyl)-methylamine. *Nucleic Acids Res.*, **15**, 3241–3255.
- Grundy, F.J. and Henkin, T.M. (1990) Cloning and analysis of the *Bacillus subtilis rpsD* gene, encoding ribosomal protein S4. *J. Bacteriol.*, **172**, 6372–6379.
- Heilek, G.M. and Noller, H.F. (1996) Site-directed hydroxyl radical probing of the rRNA neighborhood of ribosomal protein S5. *Science*, **272**, 1659–1662.
- Heilek, G.M., Marusak, R., Meares, C.F. and Noller, H.F. (1995) Directed hydroxyl radical mapping of 16S rRNA using Fe(II) tethered to ribosomal protein S4. *Proc. Natl Acad. Sci. USA*, **92**, 1113–1116.
- Held, W.A., Ballou, B., Mizushima, S. and Nomura, M. (1974) Assembly mapping of 30S ribosomal proteins from *Escherichia coli*. Further studies. *J. Biol. Chem.*, **249**, 3103–3111.
- Hinck, A.P., Markus, M.A., Huang, S., Grzesiek, S., Kustanovich, I., Draper, D.E. and Torchia, D.A. (1997) The RNA binding domain of ribosomal protein L11: three-dimensional structure of the RNA-bound form of the protein and its interaction with 23S rRNA. *J. Mol. Biol.*, **274**, 101–113.
- Hinrichs, W., Kisker, C., Duvel, M., Muller, A., Tovar, K., Hillen, W. and Saenger, W. (1994) Structure of the Tet repressor–tetracycline complex and regulation of antibiotic resistance. *Science*, **264**, 418–420.
- Holm, L. and Sander, C. (1995) DALI: a network tool for protein structure comparison. *Trends Biochem. Sci.*, **20**, 478–480.
- Hutchinson, E.G. and Thornton, J.M. (1996) PROMOTIF: a program to identify and analyse secondary structure motifs in proteins. *Protein Sci.*, **5**, 212–220.
- Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in those models. *Acta Crystallogr.*, **A47**, 110–119.
- Kodandapani, R., Pio, F., Ni, C.-Z., Piccialli, G., Klemsz, M., McKercher, S., Maki, R.A. and Ely, K.R. (1996) A new pattern for helix–turn–helix recognition revealed by PU.1 ETS-domain–DNA complex. *Nature*, **380**, 456–460.
- Kraulis, P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
- Kurland, C.G., Jørgensen, F., Richter, A., Ehrenberg, M., Bilgin, N. and Rojas, A.-M. (1990) Through the accuracy window. In Hill, W.E., Dahlberg, A., Garrett, R.A., Moore, P.B., Schlessinger, D. and Warner, J.R. (eds), *The Ribosome: Structure, Function and Evolution*. American Society for Microbiology, Washington DC, pp. 513–526.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
- Malhotra, A., Severinova, E. and Darst, S.A. (1996) Crystal structure of a σ70 subunit fragment from *E. coli* RNA polymerase. *Cell*, **87**, 127–138.
- Markus, M.A., Gerstner, R.B., Draper, D.E. and Torchia, D.A. (1998) The solution structure of ribosomal protein S4Δ41 reveals two subdomains and a positively charged surface that may interact with RNA. *EMBO J.*, **17**, 4559–4571.
- Marqusee, S. and Baldwin, R.L. (1987) Helix stabilization by Glu–...Lys+ salt bridges in short peptides of *de novo* design. *Proc. Natl Acad. Sci. USA*, **84**, 8898–8902.
- Matthews, B. (1968) Solvent content of protein crystals. *J. Mol. Biol.*, **33**, 491–497.
- Merritt, E.A. and Bacon, D.J. (1997) Raster3D: photorealistic molecular graphics. *Methods Enzymol.*, **277**, 505–524.
- Mueller, F. and Brimacombe, R. (1997) A new model for the three-dimensional folding of *Escherichia coli* 16S ribosomal RNA. II. The RNA–protein interaction data. *J. Mol. Biol.*, **271**, 545–565.
- Nicholls, A., Sharp, K.A. and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.
- Nikonov, S. et al. (1996) Crystal structure of the RNA-binding ribosomal protein L1 from *Thermus thermophilus*. *EMBO J.*, **15**, 1350–1359.
- Noller, H.F., Hoffarth, V. and Zimniak, L. (1992) Unusual resistance of peptidyl transferase to protein extraction procedures. *Science*, **256**, 1416–1419.
- Nomura, M. and Held, W.A. (1974) Reconstitution of ribosomes: Studies of ribosome structure, function and assembly. In Nomura, M., Tissieres, A. and Lengyel, P. (eds), *Ribosomes*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 193–224.
- Nowotny, V. and Nierhaus, K.H. (1988) Assembly of the 30S subunit from *Escherichia coli* ribosomes occurs via two assembly domains which are initiated by S4 and S7. *Biochemistry*, **27**, 7051–7055.
- Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., Ishii, S. and Nishimura, Y. (1992) Solution structure of a DNA-binding motif with conserved tryptophans forming a hydrophobic core. *Proc. Natl Acad. Sci. USA*, **89**, 6428–6432.
- Ohyama, K. et al. (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature*, **322**, 572–574.
- Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G. and Steitz, T.A. (1985) Structure of a large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature*, **313**, 762–766.
- Otwinowski, Z. (1993) Data collection and processing. In Sawyer, L., Isaacs, N. and Bailey, S. (eds), *Proceedings of the CCP4 Study Weekend*. SERC Daresbury Laboratory, Warrington, UK, pp. 56–62.
- Pennisi, E. (1998) Taking a structured approach to understanding proteins. *Science*, **279**, 978–979.
- Powers, T. and Noller, H.F. (1991) A functional pseudoknot in 16S ribosomal RNA. *EMBO J.*, **10**, 2203–2214.
- Powers, T. and Noller, H.F. (1995a) Hydroxyl radical footprinting of ribosomal proteins on 16S rRNA. *RNA*, **1**, 194–209.
- Powers, T. and Noller, H.F. (1995b) A temperature-dependent conformational rearrangement in the ribosomal protein S4–16S rRNA complex. *J. Biol. Chem.*, **270**, 1238–1242.
- Randolph-Anderson, B.L., Boynton, J.E., Gillham, N.W., Huang, C. and Liu, X. (1995) The chloroplast gene encoding ribosomal protein S4 in *Chlamydomonas reinhardtii* spans an inverted repeat-unique sequence junction and can be mutated to suppress a streptomycin dependence mutation in ribosomal protein S12. *Mol. Gen. Genet.*, **247**, 295–305.
- Ramakrishnan, V. and White, S.W. (1998) Ribosomal protein structures: insights into the architecture, mechanism and evolution of the ribosome. *Trends Biochem. Sci.*, **23**, 208–212.

- Rayment, I., Rypniewski, W.R., Schmidt-Bäse, K., Smith, R., Tomchick, D.R., Benning, M.M., Winkelmann, D.A., Wesenberg, G. and Holden, H.M. (1993) Three-dimensional structure of myosin sub-fragment-1: a molecular motor. *Science*, **262**, 50–58.
- Sapag, A., Vartikar, J.V. and Draper, D.E. (1990) Dissection of the 16S rRNA binding site for ribosomal protein S4. *Biochim. Biophys. Acta*, **1050**, 34–37.
- Shinozaki, K. *et al.* (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J.*, **5**, 2043–2049.
- Spedding, G. and Draper, D.E. (1993) Allosteric mechanism for translational repression in the *Escherichia coli*  $\alpha$  operon. *Proc. Natl Acad. Sci. USA*, **90**, 4399–4403.
- Stern, S., Wilson, R.C. and Noller, H.F. (1986) Localization of the binding site for protein S4 on 16S ribosomal RNA by chemical and enzymatic probing and primer extension. *J. Mol. Biol.*, **192**, 101–110.
- Stevenson, J.K., Drager, R.G., Copertino, D.W., Christopher, D.A., Jenkins, K.P., Yepiz-Plascencia, G. and Hallick, R.B. (1991) Intercistronic group III introns in polycistronic ribosomal protein operons of chloroplasts. *Mol. Gen. Genet.*, **228**, 183–192.
- Stöffler, G. and Stöffler-Meilicke, M. (1984) Immunoelectron microscopy of ribosomes. *Annu. Rev. Biophys. Bioeng.*, **13**, 303–330.
- Tanaka, I., Nakagawa, A., Hosaka, H., Wakatsuki, S., Mueller, F. and Brimacombe, R. (1998) Matching the crystallographic structure of ribosomal protein S7 to a three-dimensional model of the 16S ribosomal RNA. *RNA*, **4**, 542–550.
- Tang, C.K. and Draper, D.E. (1989) Unusual mRNA pseudoknot structure is recognized by a protein translational repressor. *Cell*, **57**, 531–536.
- Tang, C.K. and Draper, D.E. (1990) Evidence for allosteric coupling between the ribosome and repressor binding sites of a translationally regulated mRNA. *Biochemistry*, **29**, 4434–4439.
- Thomas, M., Bedwell, D.M. and Nomura, M. (1987) Regulation of  $\alpha$  operon gene expression in *Escherichia coli*. *J. Mol. Biol.*, **196**, 333–345.
- Urlaub, H., Kruft, V., Bischof, O., Müller, E.C. and Wittmann-Liebold, B. (1995) Protein-rRNA binding features and their structural and functional implications in ribosomes as determined by cross-linking studies. *EMBO J.*, **14**, 4578–4588.
- Urlaub, H., Thiede, B., Müller, E.C., Brimacombe, R. and Wittmann-Liebold, B. (1997) Identification and sequence analysis of contact sites between ribosomal proteins and rRNA in *Escherichia coli* 30S subunits by a new approach using matrix-assisted laser desorption/ionization-mass spectrometry combined with N-terminal microsequencing. *J. Biol. Chem.*, **272**, 14547–14555.
- van Acken, U. (1975) Protein chemical studies on ribosomal mutants S4 and S12 from *ram* (ribosomal ambiguity) mutants of *Escherichia coli*. *Mol. Gen. Genet.*, **140**, 61–68.
- Vartikar, J.V. and Draper, D.E. (1989) S4–16S ribosomal RNA complex. Binding constant measurements and specific recognition of a 460-nucleotide region. *J. Mol. Biol.*, **29**, 221–234.
- Wittmann-Liebold, B., Uhlein, M., Urlaub, H., Müller, E.-C., Otto, A. and Bischof, O. (1995) Structural and functional implications in the eubacterial ribosome as revealed by protein-rRNA and antibiotic contact sites. *Biochem. Cell Biol.*, **73**, 1187–1197.
- Yates, J.L., Arfsten, A.E. and Nomura, M. (1980) *In vitro* expression of *Escherichia coli* ribosomal protein genes: autogenous inhibition of translation. *Proc. Natl Acad. Sci. USA*, **77**, 1837–1841.
- Zimmermann, R.A., Mackie, G.A., Muto, A., Garrett, R.A., Ungewickell, E., Ehresmann, C., Stiegler, P., Ebel, J.P. and Fellner, P. (1975) Location and characteristics of ribosomal protein binding sites in the 16S RNA of *Escherichia coli*. *Nucleic Acids Res.*, **2**, 279–302.

Received May 5, 1998; revised June 22, 1998; accepted June 23, 1998