The EMBO Journal Vol.17 No.16 pp.4559–4571, 1998

The solution structure of ribosomal protein S4 Δ41 reveals two subdomains and a positively charged surface that may interact with RNA

Michelle A.Markus, Resi B.Gerstner[1], David E.Draper[1] and Dennis A.Torchia[2]

Molecular Structural Biology Unit, National Institute of Dental Research, National Institutes of Health, 30 Convent Drive, Room 132, Bethesda, MD 20892-4320 and [1]Department of Chemistry, Johns Hopkins University, Baltimore, MD 21218, USA

[2]Corresponding author
e-mail: torchia@yoda.nidr.nih.gov

S4 is one of the first proteins to bind to 16S RNA during assembly of the prokaryotic ribosome. Residues 43–200 of S4 from *Bacillus stearothermophilus* (S4 Δ41) bind specifically to both 16S rRNA and to a pseudoknot within the α operon mRNA. As a first step toward understanding how S4 recognizes and organizes RNA, we have solved the structure of S4 Δ41 in solution by multidimensional heteronuclear nuclear magnetic resonance spectroscopy. The fold consists of two globular subdomains, one comprised of four helices and the other comprised of a five-stranded antiparallel β-sheet and three helices. Although cross-linking studies suggest that residues between helices α2 and α3 are close to RNA, the concentration of positive charge along the crevice between the two subdomains suggests that this could be an RNA-binding site. In contrast to the L11 RNA-binding domain studied previously, S4 Δ41 shows no fast local motions, suggesting that it has less capacity for refolding to fit RNA. The independently determined crystal structure of S4 Δ41 shows similar features, although there is small rotation of the subdomains compared with the solution structure. The relative orientation of the subdomains in solution will be verified with further study.

Keywords: modular proteins/multidimensional NMR/protein–nucleic acid interactions/ribosome assembly/ribosomal proteins

Introduction

The protein S4 is an essential component of the ribosome. Prokaryotic S4 binds to the 16S rRNA to nucleate one of the two assembly domains for the small subunit (S7 nucleates the other domain) (Nowotny and Nierhaus, 1988). Mutations in S4 affect the accuracy of translation by the ribosome and influence the higher order structure of the 16S RNA (Allen and Noller, 1989). S4 is strongly conserved among bacteria and chloroplasts, and homologs have been identified in yeast, higher plants and animals based on a well-conserved stretch of ~45 amino acids. Mutations in the yeast homolog affect the accuracy of the eukaryotic ribosome (Liebman *et al*., 1995), suggesting that S4's crucial role in ribosomal function is conserved across phylogenetic domains.

Like other key ribosomal proteins, prokaryotic S4 regulates its own expression by binding to its own mRNA, the α operon mRNA, which encodes not only S4, but also S11, S13 and L17. S4 binds to α operon mRNA and 16S RNA with similar affinity (Deckman and Draper, 1985) despite the fact that these RNA targets have been reduced to quite dissimilar structures: the mRNA target site can be reduced to ~100 nucleotides containing a pseudoknot (Deckman and Draper, 1987; Deckman *et al*., 1987), while the 16S RNA target site can be reduced to 462 nucleotides containing at least three essential hairpins (Vartikar and Draper, 1989; Sapag *et al*., 1990).

The C-terminal fragment of prokaryotic S4 retains almost full binding affinity for both RNA targets (Baker and Draper, 1995). Strikingly, the N-terminal residues 1–46 from the *Escherichia coli* sequence and 1–41 from the *Bacillus stearothermophilus* sequence are rich in proline, glycine, arginine and lysine, suggesting that the N-terminus forms a highly positively charged extension with little regular secondary structure. The high entropy cost of organizing the N-terminal residues may offset a favorable charge interaction with RNA, leaving most of the affinity for RNA in the C-terminal domain. As a first step towards understanding the structural basis for the C-terminal domain's specific recognition of two distinct RNA targets, we have solved the structure of the C-terminal fragment of S4 from *B.stearothermophilus* (S4 Δ41) in solution. S4 Δ41 is comprised of two distinct subdomains, with a concentration of positive side chains lining the crevice between the subdomains. The results are discussed in detail below.

Results

Solution conditions have a major impact on the quality of the spectra for S4 Δ41

The first ^{15}N HSQC spectrum for S4 Δ41, in 10 mM KH_2PO_4, pH 6.0 and 70 mM KCl, at ~1 mM protein concentration, showed only 19 peaks for a protein of 159 residues with a total of 199 peaks expected, including side chains. The observed peaks were mostly in regions typical for side chain NH_2 groups and arginine εs. Dilution of the protein (1:10) dramatically improved the spectrum, which suggests that S4 Δ41 aggregates. Higher temperatures (37°C) and higher salt concentrations (250 mM KCl) improved the solubility sufficiently such that good quality spectra were obtained at protein concentrations near 1 mM, as demonstrated by the ^{15}N HSQC spectrum shown in Figure 1. Despite the high salt and high temperatures, aggregation was still evident at protein concentrations of ~1 mM; for the HCCH-TOCSY experiment, where sensitivity depends critically upon the ^{13}C T_2 relaxation, and thus upon molecular tumbling and aggrega-

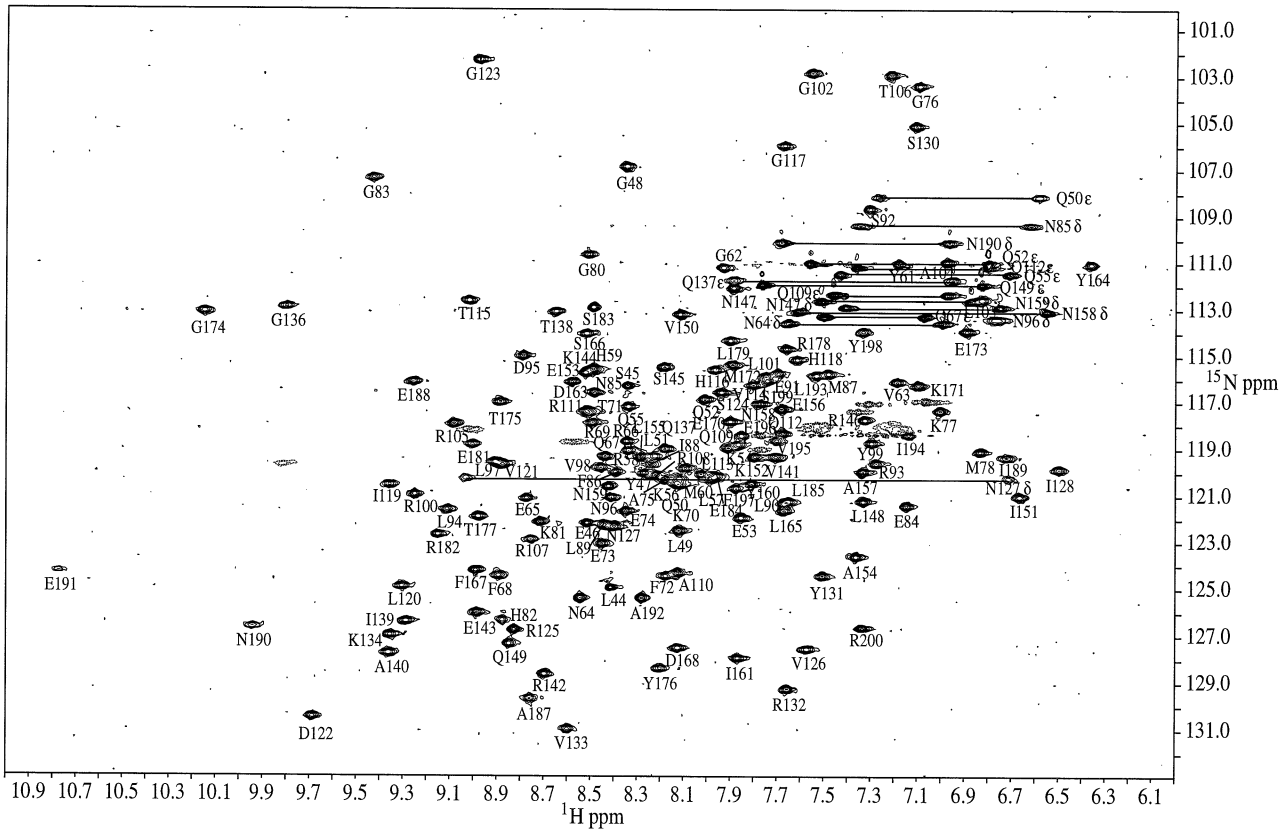

Fig. 1. ^{15}N HSQC for S4 Δ41 in 20 mM deuterated acetate, 250 mM KCl, at 37°C, acquired at 750 MHz. The ^{1}H–^{15}N cross-peaks are labeled with the assigned amino acid type and residue number. Asparagine side chains are additionally labeled δ, and glutamine side chains are labeled ε. Cross-peaks corresponding to the same side chain are connected with a horizontal line. The arginine side chains have been aliased into the middle of the spectrum as negative peaks, represented by dashed lines. They are not labeled.

tion, better signal-to-noise was obtained with a sample at 650 μM than at 920 μM.

Chemical shift assignments for S4 Δ41

Backbone assignments for S4 Δ41 are essentially complete; the only unassigned positions are the ^{15}N chemical shifts for the seven proline residues and both amide proton and nitrogen chemical shifts for the N-terminal residues, Met42 and Lys43. The amide assignments are indicated in the ^{15}N HSQC spectrum in Figure 1. Side chain assignments are complete (with at least one proton assigned per methylene group) for 139 out of 159 residues. Only five of 16 arginine side chains could be completely assigned, due to overlap and to very weak peaks for some residues. The unassigned arginine positions include the εs for residues 58, 66, 69, 108, 125 and 142, the γs for residues 182 and 200, γ and δ for residue 93, δ and ε for residue 107, and γ, δ and ε for residue 178. Side chain carbonyls could not be assigned for the asparagines at positions 85, 96, 127 and 190 nor for the glutamines at positions 50 and 137. The other gaps in assignments occurred in long side chains: the γ position for Lys54, the γ methylene position for Ile119, and the ring ε and ζ positions for Phe167.

The solution structure of S4 Δ41 reveals an elongated molecule with two subdomains

The restraints used in structure calculations for S4 Δ41 in solution are summarized in Table I. The structure was calculated using the hybrid distance geometry/simulated annealing protocol in X-PLOR. Statistics reflecting the merit of the structure calculations are summarized in Table II. Convergence for the structure calculations was quite high: 50 of 50 calculated structures had no restraint violations above 0.5 Å for distance restraints nor above 5° for dihedral angle constraints. The best 26 structures, with no violations above 0.2 Å nor 2°, were selected for analysis.

Table I. Summary of constraints for structure calculations for S4 Δ41 in solution

Total constraints available	2405
Total NOE restraints	2196
intraresidue	698
sequential ($\|i - j\| = 1$)	608
short range ($1 < \|i - j\| \leq 4$)	378
long range ($\|i - j\| > 4$)	512
Hydrogen bond constraints	90
Total dihedral angle constraints	119
ϕ	73
χ_1	46
Stereospecific assignments	
β methylene	28
valine g methyls	8 of 9
leucine δ methyls	18 of 18

Ribbon diagrams depicting the fold for S4 Δ41 are shown in Figure 2. S4 Δ41 is mostly α-helical, with seven segments of α-helix and only five short strands forming a single antiparallel β-sheet. Although the chemical shift

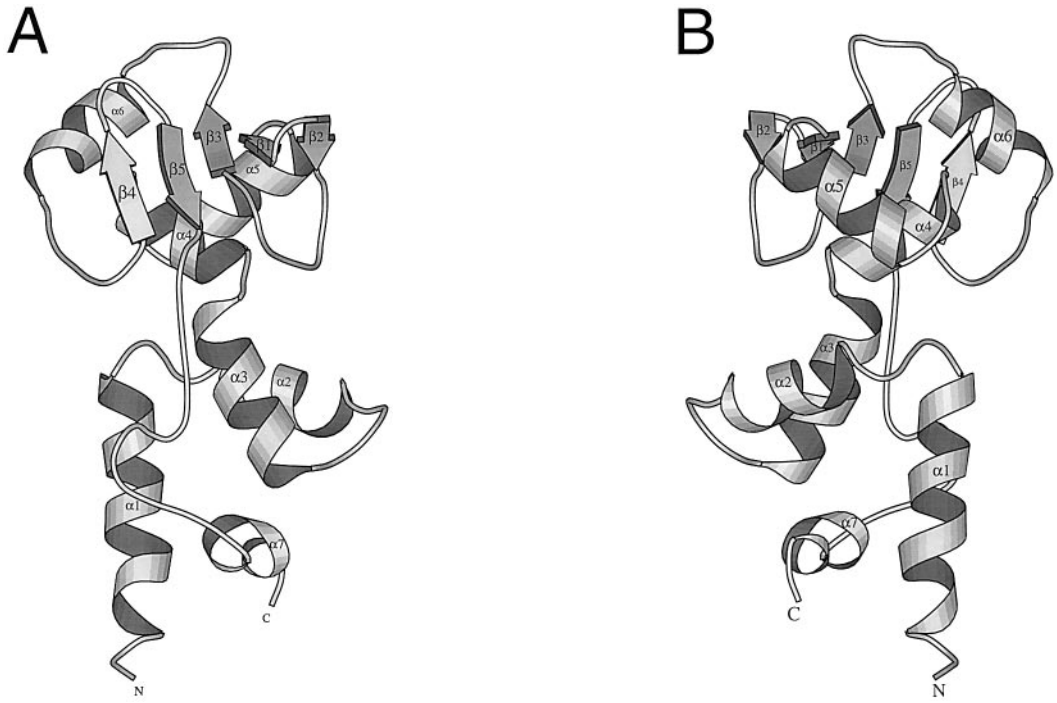

Fig. 2. (**A**) A ribbon diagram depicting the organization of elements of regular secondary structure within the fold of S4 Δ41. (**B**) A ribbon diagram depicting secondary structure for S4 Δ41, with the view rotated by 180° about the vertical axis in the page relative to (A). Elements of secondary structure are labeled, with α denoting α-helices and β denoting extended strands that form the β-sheet. N and C denote the termini of the protein chain. This figure was generated with MOLSCRIPT (Kraulis, 1991).

Table II. Statistics for X-PLOR calculations for S4 Δ41 in solution

Acceptance criteria		
NOE violations above 0.2 Å	0	
dihedral angle violations over 2°	0	
Accepted structures	26 out of 50	
	{SA}[a]	**<SA>$_m$**
Remaining violations		
NOE violations above 0.1 Å	2.58 ± 1.50	0
dihedral angles violations over 1°	0.0769 ± 0.2665	0
R.m.s.d. from idealized covalent geometry		
bonds (Å)	0.00145 ± 0.00008	0.00119
bond angles (°)	0.4899 ± 0.0052	0.4703
improper torsions (°)	0.3574 ± 0.0043	0.3380
R.m.s.d. from experimental constraints		
distances (Å)	0.008310 ± 0.000961	0.005000
dihedral angles (°)	0.05179 ± 0.03618	0.02149
Final energies (kcal/mol)		
distance restraints	7.999 ± 1.892	2.857
dihedral angles	0.0289 ± 0.0379	0.00335
non-bonded (REPEL)	9.720 ± 2.283	3.143

[a]{SA} refers to the ensemble of 26 accepted structures. For the ensemble, the average value plus or minus the standard deviation is quoted. <SA>$_m$ refers to the minimized average structure.

index (Wishart and Sykes, 1994) suggests that helix α3 continues directly into helix α4 to form one long helix, hydrogen exchange and NOESY data show that this is not the case. There is a break in the pattern of slowly exchanging amides at Glu91 and a strong sequential HN–Hα NOE at Arg93 instead of the strong HN–HN sequential cross-peak expected for a helix. Furthermore, our structure appears to be stabilized by hydrogen bonds between residues in the break between the helices and residues in the loop between strands β2 and β3. The break between helix α3 and helix α4 is also the dividing line between the two subdomains of S4 Δ41. Helices α1, α2, α3 and α7 pack together to form one subdomain, while helices α4, α5 and α6 form a layer that packs against the five-stranded antiparallel sheet to form the second subdomain. Most long-range NOEs represent interactions within each subdomain, although there are a few NOEs from the C-terminus of helix α1, the loop, and the N-terminus of helix α2 within the first subdomain to helix α4 of the second subdomain, as well as from helix α3 of the first

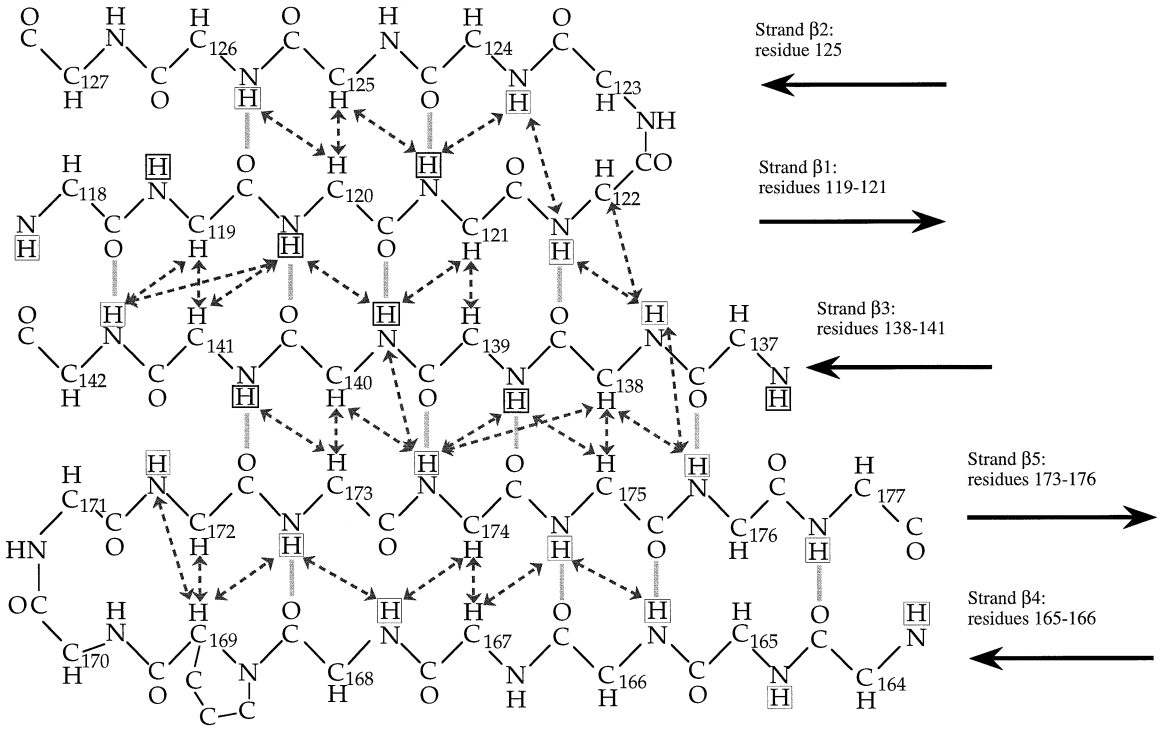

Fig. 3. The chemical structure of the backbone for residues in the five-stranded antiparallel β-sheet found in S4 Δ41. Slowly exchanging amide protons are boxed; amide protons that do not exchange after weeks in D_2O are enclosed in darker boxes, those protected for at least 45 min are enclosed in lighter boxes. Seven of the 21 most slowly exchanging amides are found in this β-sheet. Dashed double-headed arrows connect protons that give rise to NOE cross-peaks. Thick gray lines connect the partners for hydrogen bond restraints used in the structure calculations.

subdomain to the loop between β2 and β3 in the second subdomain which begin to establish the relative positions of the subdomains. Overall, S4 Δ41 is somewhat elongated and approximately axially symmetric; the averages over the 26 accepted structures for the ratios of the principal components of the inertia tensor are 1.00:2.41:2.36.

Although strand β2 is very twisted in the ribbon representation in Figure 2, the pattern of slowly exchanging amides from residues 124 to 126 suggests that these residues hydrogen-bond to strand β1. This is clear in the schematic representation of the sheet, shown in Figure 3. Both Ser124 and Val126 have slowly exchanging amide protons, while amide protons in Arg125 and Asn127 exchange rapidly. Although PROCHECK (Laskowski *et al*., 1993) does not identify these residues as a β-strand, the Protein Design module of QUANTA consistently identifies a strand conformation at position 125, and the pattern of NOEs to strand β1 also suggests pairing. Therefore, we suggest that β2 is functionally the fifth strand of the β-sheet, even though its backbone is somewhat more twisted than typically found in an extended strand.

PROCHECK identifies residues 143–145 as an additional consensus helix in S4 Δ41. Careful examination, structure by structure for the 26 accepted structures, shows that residues 143–145 do not consistently adopt a helical conformation: in some structures, 143–146, 144–146 or 145–147 are identified, and in two structures, no helical residues are identified in this region. Furthermore, a helix should be long enough to include at least one characteristic hydrogen bond, and most of these regions are too short for an α-helix. Therefore, we describe this region as a 'helical turn' rather than an additional helix.

The structure of each subdomain is better defined than the whole

The ensemble of 26 calculated structures is superimposed in Figure 4, and superpositions for different residues are compared in Table III. Superposition of all 26 structures onto an average structure, aligned by the elements of secondary structure identified with PROCHECK (83 residues), gives an r.m.s.d. for the backbone atoms of 0.61 Å. However, for S4 Δ41, the elements of secondary structure are not exclusively the best defined features. Most of the turns within the β-sheet subdomain are very well defined, while the exact boundaries for helices α1, α2 and α7 vary from structure to structure in the ensemble. Therefore, a superposition of 119 residues that leaves out not just the N-terminus but the beginning of helix α1, the turn between helices α2 and α3, the turn between α6 and β4, the extended structure between β5 into α7, and the C-terminal end of α7 to the end of the chain produces a somewhat better alignment. This superposition, shown in Figure 4A, gives an r.m.s.d. for the backbone of 0.54 Å. Each subdomain superimposed separately gives still lower backbone r.s.m.d.s (Table III). The r.m.s.d. for 84 residues in the sheet-containing subdomain is lower than for 34 residues in the helical subdomain (0.35 Å versus 0.40 Å, see Figure 4B and C). The improved definition in the sheet-containing subdomain is probably due to the large number of strong long-range NOEs characteristic of

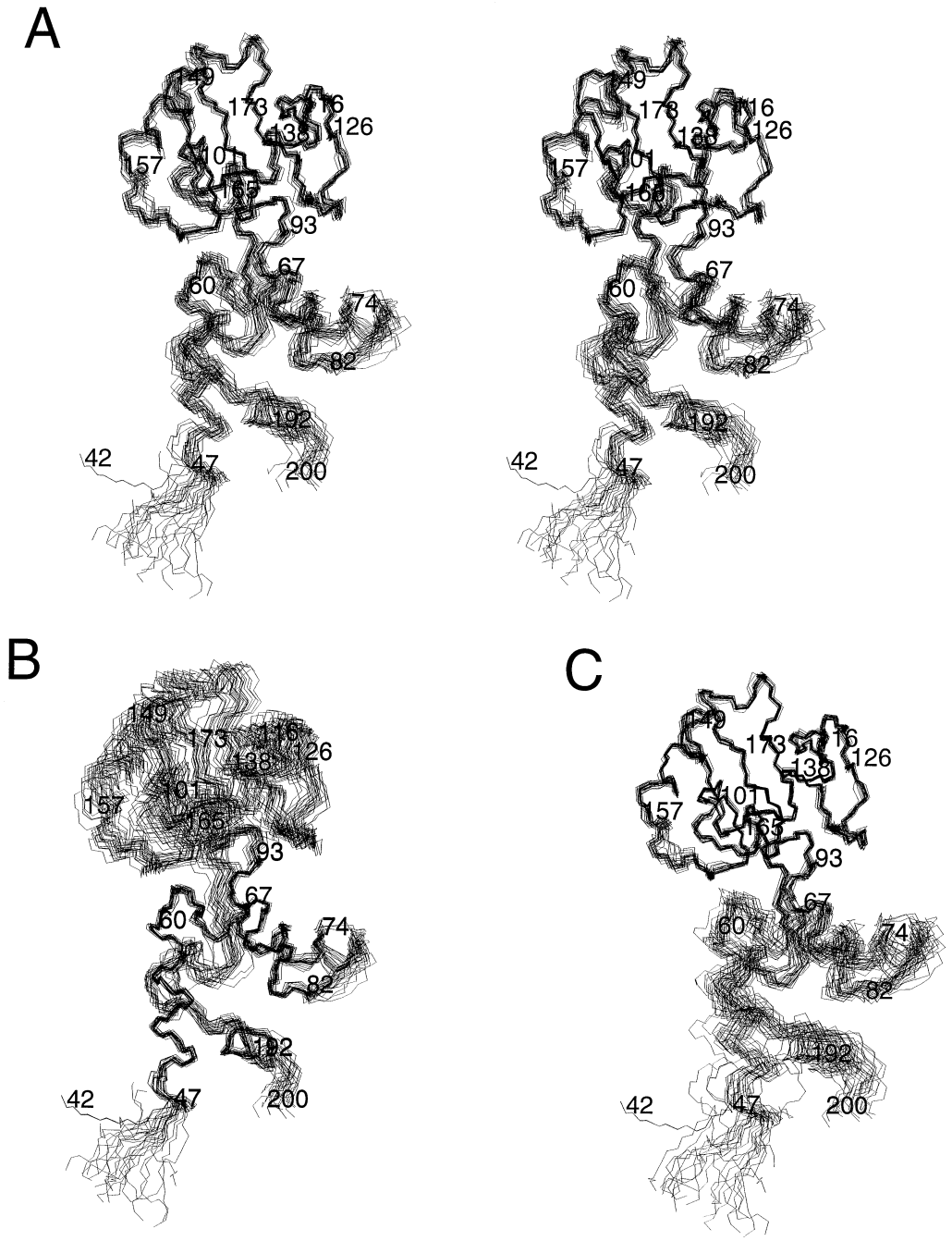

Fig. 4. Views of the ensemble of 26 structures calculated for S4 Δ41. (**A**) Stereoview of the ensemble of structures, superimposed by the backbone throughout the molecule, omitting three loops (i.e. superimposed by residues 50–72, 85–157, 161–180 and 193–195). (**B**) Single image of the ensemble, superimposed by the well-defined regions in the helical (lower) subdomain (residues 50–72, 85–92 and 193–195). Note that alignment within the helical subdomain is better, but within the sheet-containing (upper) subdomain is worse, than in (A). (**C**) Single image of the ensemble, superimposed by the well-defined regions in the sheet-containing subdomain (residues 94–157 and 161–180). In this case, the alignment within the sheet-containing subdomain is better, but that within the helical subdomain is worse, than in (A). Some of the endpoints of regular secondary structure and the chain termini are labeled with the residue number for reference. This figure was generated with QUANTA (Molecular Simulations, Inc., San Diego, CA).

β-sheets and to the improved resolution of the chemical shifts in the sheet-containing subdomain.

Locally high r.m.s.d. correlate with fast time scale mobility in one loop

The r.m.s.d. per residue for S4 Δ41 is shown in Figure 5A. The most disordered regions include the chain termini, the loop between helices α2 and α3, the loop between helix α6 and strand β4, and the extended structure between strand β5 and helix α7. In general, the connections between the other elements of regular secondary structure are well defined, though there are smaller increases in r.m.s.d. between helices α1 and α2, α4 and α5, helix α5 and strand β1, and between strands β4 and β5. Typically, a local increase in r.m.s.d. correlates with a local decrease in the number of NOE restraints (Figure 5B), although the number of NOE restraints per residue also reflects the number of protons in that residue. The most noticeable

Table III. R.m.s.d.[a] for different superpositions of the calculated S4 Δ41 structures

Description	Residues	Total[b]	{SA}[c] versus <SA>		<SA> versus $<SA>_m$	
			bb[d]	heavy	bb	heavy
Structured region	47–199	153	0.76	1.15	0.32	0.59
Consensus regular 2° structure[e]	47–60, 67–74, 82–101, 107–116,119–121, 138–141, 149–157, 165–168, 173–176, 192–198	83	0.61	1.04	0.22	0.53
Omit termini and three loops	50–72, 85–157, 161–180, 193–195	119	0.54	0.98	0.22	0.51
Helical subdomain	50–72, 85–92, 193–195	34	0.40	0.87	0.21	0.52
β-Sheet subdomain	94–157, 161–180	84	0.35	0.86	0.22	0.50

[a]R.m.s.ds are quoted in angstroms.
[b]Total number of residues included in the superposition.
[c]{SA} is the ensemble of 26 accepted structures. <SA> is the average structure, calculated by superimposing the consensus regular secondary structure and then averaging the coordinates. $<SA>_m$ is the average structure after restrained energy minimization.
[d]Backbone atoms.
[e]The consensus secondary structure is based on the secondary structure identified in the ensemble of structures, calculated with PROCHECK (Laskowski *et al*., 1993).

local decrease in the number of NOE restraints occurs between strand β5 and helix α7, in the extended bridge of residues that runs from the sheet-containing subdomain back to the helical subdomain.

While it is tempting to suggest that disordered regions in the superimposed ensemble of structures correspond to mobile regions in solution, relaxation measurements offer a more direct probe of mobility. Figure 5C shows the ^{15}N T_2 time constants for backbone amide in S4 Δ41. Elevated T_2 time constants suggest motions on the picosecond to nanosecond (fast) time scale. Disorder at the chain termini (residues 42–46 and 200) and in the loop between helix α6 and strand β4 (residues 158–160) correlates well with fast time scale motions, as suggested by the decidedly elevated T_2 time constants. The disorder in the loop between helices α2 and α3 (residues 78–81) correlates with slightly elevated ^{15}N T_2s. Thus, disorder between α2 and α3 could be associated with fast time scale motions, but it could also reflect low proton density due to the geometry of the chain (this loop sticks out somewhat, away from the rest of the molecule). Reduced T_2 time constants typically reflect slower motions, on the chemical shift (millisecond) time scale. Strikingly low T_2s are observed for residues Thr71 and Ala187, although the measurement at Thr71 is associated with a large error due to substantial overlap with Arg111. The low value at Ala187 suggests that there could be slow time scale motions in the long bridge of residues connecting the two subdomains, which correlates with the local increase in the r.m.s.d. between helix α6 and strand β4. In summary, for S4 Δ41, the regions with the largest local r.m.s.ds correspond qualitatively to regions with elevated or reduced ^{15}N T_2s, suggesting that mobility, at least in part, contributes to the disorder in the ensemble of structures.

In our previous work on the NMR structure of the RNA-binding domain of ribosomal protein L11 (L11-C76, Markus *et al*., 1997), we identified 15 residues within the domain that form an extended loop undergoing fast motions. This flexible loop becomes ordered upon binding RNA (Hinck *et al*., 1997), suggesting that the protein adapts to fit the RNA upon binding. In contrast, the structure of S4 Δ41 is well defined all along the chain, with only a very short loop (residues 158–160) that shows evidence of fast time scale motions. This suggests that S4 lacks L11's ability to adapt to fit RNA using picosecond time scale motions.

Discussion

S4 Δ41 is not as elongated as suggested by previous biophysical measurements in solution

Measurements of the sedimentation coefficient and diffusion coefficient for full-length S4 protein from *E.coli* imply an axial ratio of ~5:1 for the protein, modeled as a prolate ellipsoid, and a radius of gyration of 29–30 Å (Dodd and Hill, 1987), assuming internal hydration of ~0.3 g H_2O per g of protein. Since the full-length *E.coli* sequence has ~50% identity with the *B.stearothermophilus* sequence, we would expect the three-dimensional structures of the two proteins to be very similar. Based on the ratios of the principle moments of the inertia tensor and modeling in QUANTA, S4 Δ41 can be modeled as an axially symmetric prolate ellipsoid with an axial ratio of ~1.5:1, a major axis of ~58 Å, and a minor axis near 40 Å. Its radius of gyration, averaged over the 26 accepted structures, is 16.56 ± 0.12 Å. These results are strikingly different from measurements on the full-length protein, despite the fact that in both studies the protein was kept in high salt (250 mM KCl for NMR, 350 mM KCl for sedimentation and diffusion) to avoid aggregation. Since S4 Δ41 lacks 41 of the N-terminal residues, one possible explanation is that these residues form an extension that greatly affects the hydrodynamic measurements. However, the N-terminal region comprises only ~20% of the entire protein, and the high number of prolines and glycines among these residues (five and six, respectively, for the *B.stearothermophilus* sequence, about twice as many as expected by the occurrence of the amino acid types in proteins) suggests a lack of regular secondary structure. Low chemical shift dispersion for additional glycine cross-peaks in a preliminary ^{15}N HSQC spectrum of the full-length S4 (spectrum not shown) also implies little regular

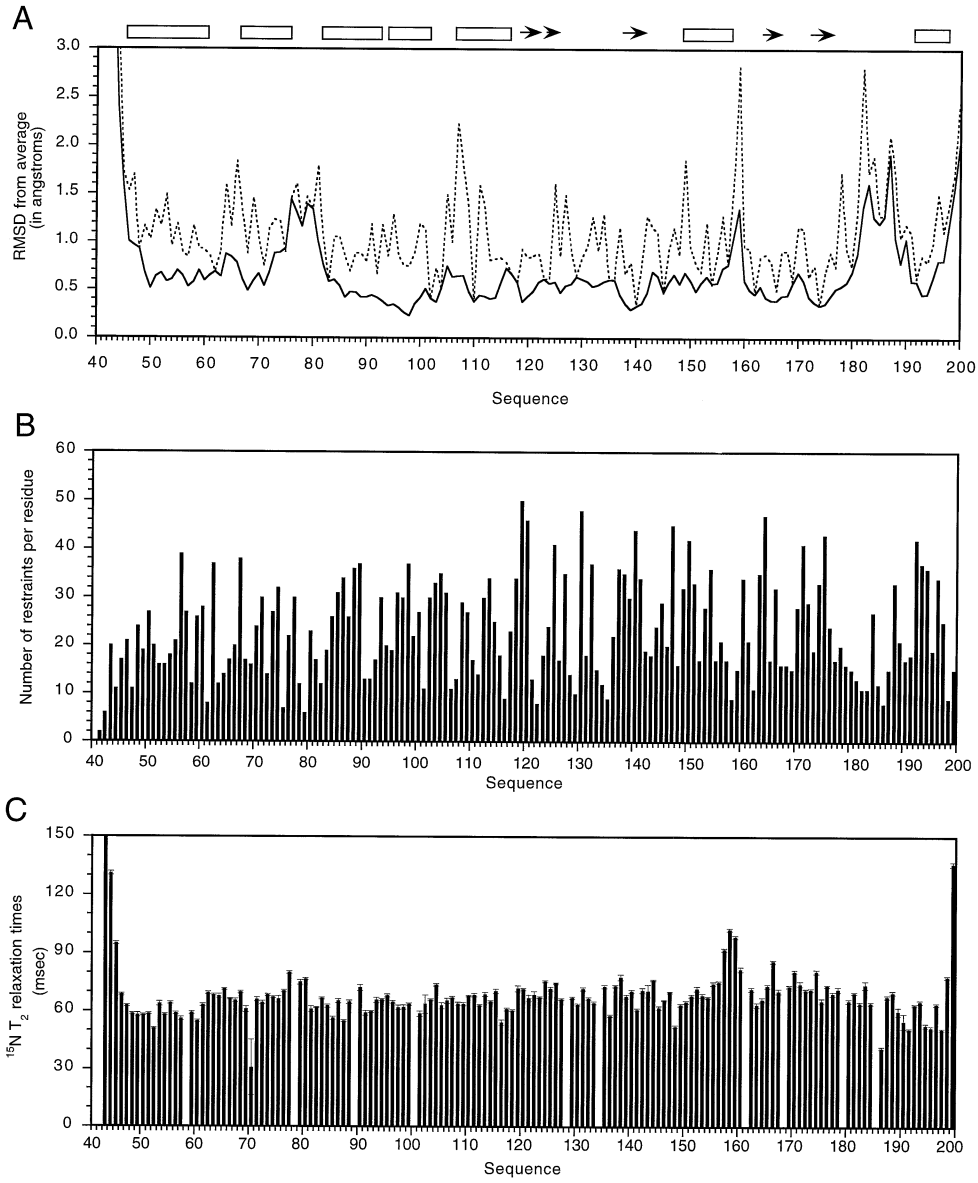

Fig. 5. Correlating the r.m.s.d. per residue with the number of restraints and with fast time scale mobility. (**A**) The average r.m.s.d. per residue for the ensemble of 26 structures for S4 Δ41, both for the backbone atoms (solid line) and all of the heavy atoms (dashed line). (**B**) The number of NOE restraints per residue. Intraresidue NOEs are counted once, while interresidue NOE restraints are counted twice, once for each residue contributing to the restraint. Note that a higher local r.m.s.d. correlates with locally fewer NOE restraints. (**C**) ^{15}N T_2 relaxation times for backbone amide nitrogens in S4 Δ41, based on experiments recorded at 750 MHz. Long T_2s suggest motion on the picosecond to nanosecond time scale. Fast motion is evident at the chain termini and in the loop between helix α6 and strand β4. The elements of secondary structure are indicated above (A), with open rectangles representing α-helices and arrows representing β-strands.

secondary structure in the N-terminal region. With little regular secondary structure, it seems less likely that the N-terminal residues form an extended structure that could greatly affect the hydrodynamic properties of the molecule and more likely that the assumptions required to deduce a simple shape for a macromolecule of unknown structure from measurements of sedimentation and diffusion may have broken down for S4.

Each subdomain of S4 Δ41 has structural homology to helix–turn–helix DNA-binding domains

To check the uniqueness of the S4 Δ41 fold, the coordinates for all 26 structures of the ensemble were compared against a database of known structures using DALI (Holm and Sander, 1993). S4 Δ41 is like no other structure in the database. However, each subdomain shows similarities to known structures. For the helical subdomain, the DNA-binding domain of the tetracycline repressor (Kisker *et al*., 1995, pdb code 2tct) matches 15 of 26 structures with *Z* scores over 2.0. The tetracycline repressor can be aligned with helices α2, α3 and α7 of the helical subdomain with a maximum *Z* score of 2.7 and r.m.s.d. for equivalent C_α positions ranging from 2.6 to 3.3 Å. The recognition helix in the helix–turn–helix motif of the tetracycline repressor aligns with the extended region from Glu184 to Glu188 in the S4 Δ41 structure (Figure 6A). Eleven structures in the ensemble also match a poorly conserved region of the

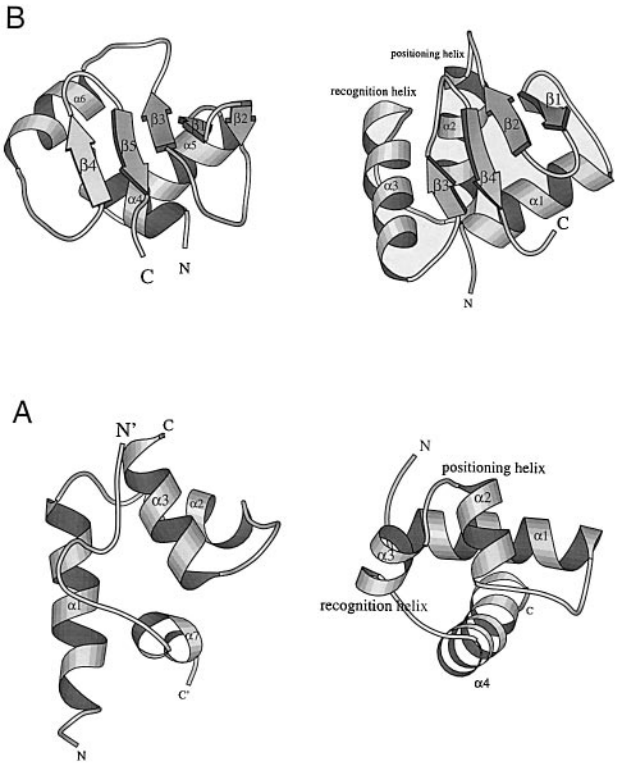

Fig. 6. Comparing the folds of the subdomains of S4 Δ41 with previously discovered folds. (**A**) The helical subdomain of S4 Δ41 shows some structural similarity to the DNA-binding domain of the tetracycline repressor. (**B**) The sheet-containing subdomain of S4 Δ41 shows some structural similarity to the PU.1 ETS DNA-binding domain. S4 Δ41 is shown on the left, while the tetracycline repressor and the PU.1 ETS domain are shown on the right in (A) and (B), respectively. Elements of secondary structure are labeled sequentially. The helices from the helix–turn–helix motif for both the tetracycline repressor and the PU.1 ETS domain are labeled, as 'positioning helix' and 'recognition helix'. The chain termini are also indicated. Note that for the helical subdomain of S4 Δ41, there are actually two protein chains, with endpoints marked as N and C, primed and unprimed for the separate chains. This figure was generated with MOLSCRIPT (Kraulis, 1991).

σ^{70} subunit from *E.coli* RNA polymerase (Malhotra *et al*., 1996, pdb code 1sig), with a maximum *Z* score of 2.4 and r.m.s.d. ranging from 3.4 to 5.4 Å over helices α1, α2, α3 and α7. Additional matches occur for only a few structures in the ensemble, which probably reflect less on possible structure–function relationships for S4 Δ41 and more on the high occurrence of packed α-helices in known protein structures.

The sheet-containing subdomain matches one known structure 12 out of 26 times: the PU.1 ETS domain, from its co-crystal with DNA (Kodandapani *et al*., 1996; pdb code 1pue, chain E). S4 Δ41 strands β1, β3, β4 and β5 and helix α6 as well as parts of helices α4 and α5 can be aligned with the PU.1 ETS domain with a maximum *Z* score of 2.7 and C_α r.m.s.d. ranging from 3.4 to 4.5 Å. The recognition helix in the ETS domain aligns with α6 in the S4 Δ41 structure (Figure 6B), although a candidate for the positioning helix of the helix–turn–helix motif is hard to identify in S4. In summary, although any functional analogies remain to be proven, it is striking that both subdomains of S4 Δ41 show similarity to DNA-binding domains based on the helix–turn–helix motif.

The strongly conserved region includes two helices and two strands

The amino acid sequence of *B.stearothermophilus* S4 Δ41 was compared against sequences in several databases using the BLAST network service at the National Center for Biotechnology Information (Altschul *et al*., 1990). The prokaryotic sequences show strong conservation throughout the domain, with 82% identity for *Bacillus subtilis* and 54% identity for *E.coli*, for example. For the eukaryotic homolog, S9, sequence identity is also very high (42–51%) for sequences in the databases, including yeast (*Saccharomyces cerevisiae*), slime mold (*Dictyostelium discoideum*), worm (*Caenorhabditis elegans*), rat (*Rattus norvegicus*) and humans (*Homo sapiens*), but the similarity is confined to a region of 44 amino acids, corresponding to residues 90–134 in the S4 Δ41 sequence (Figure 7). In the structure, these residues form the beginning of the β-sheet-containing subdomain and include helices α4 and α5, strands β1 and β2, and much of the loop from β2 to β3 (refer to Figure 2). Helix α4 is mostly buried, but much of helix α5, strands β1 and β2, and the loop from β2 to β3 is accessible from the surface. Notice that although it was not identified through structural homology with DALI, the region from helix α4 to helix α5 could also be described as a helix–turn–helix unit. Because most of the conserved residues map to the surface of S4 Δ41, we can speculate that they are important for intermolecular interactions. However, since S4 binds to two distinct RNA molecules and facilitates interactions with other proteins in the context of ribosomal assembly, it is likely that much of its surface is used for intermolecular interactions and it is not readily apparent which target interacts with the conserved region.

A further search was made based on the consensus from the strongly conserved S4 proteins from eubacteria and the S9 proteins from eukaryotes using PSI-BLAST (Altschul *et al*., 1997). This search reveals sequence similarities to additional proteins, including a segment of eubacterial tyrosyl-tRNA synthetases, reverse transcriptase from a viral polyprotein and an *E.coli* protein of unknown function (Figure 7). Although the similarities are confined to a short stretch of amino acids, it is possible that S4 is the prototype for a novel family of RNA-binding proteins.

Cross-linking studies suggest an RNA-binding site

Protein regions important for interaction with the 16S rRNA have been identified through cross-linking studies (Urlaub *et al*., 1995). RNA was cross-linked to proteins in the context of the assembled ribosomal subunits, using subunits from both *E.coli* and *B.stearothermophilus*. The proteins were digested, and the peptides containing cross-links were sequenced. Although no peptides corresponding to S4 were identified from the *B.stearothermophilus* subunits, Lys82 from *E.coli* S4 was cross-linked to the 16S RNA. In the *B.stearothermophilus* sequence, this residue corresponds to Pro79, found in the turn between helices α2 and α3 in the first subdomain. This residue is on the surface of the domain, in a good position for cross-linking if the correct functional group were available. Nearby lysines at positions 77 and 81 in the *B.stearothermophilus* sequence are apparently not close enough to the RNA to form cross-links, possibly due to interactions with side

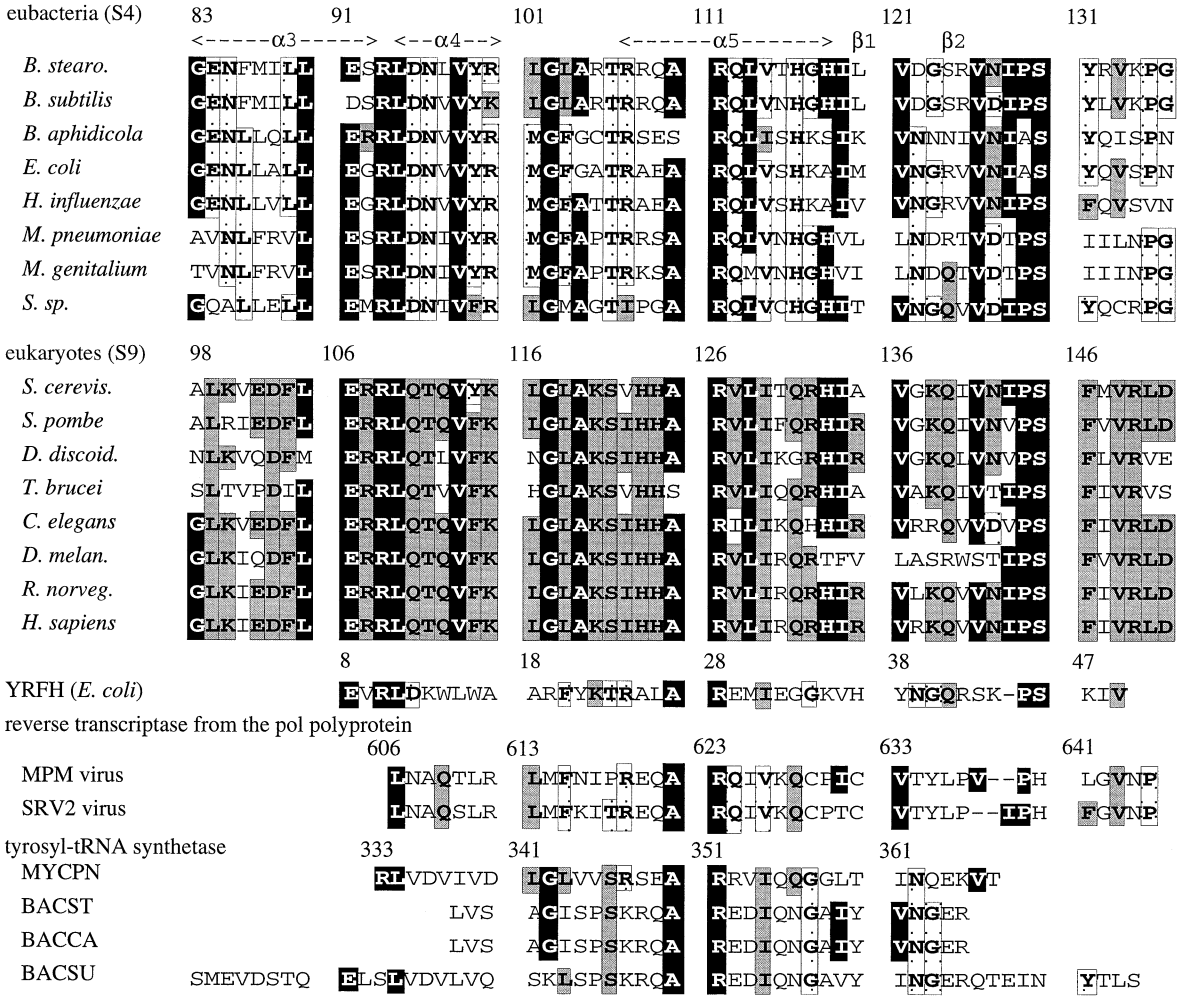

Fig. 7. Comparing the amino acid sequence of S4 Δ41 with closely related eubacterial S4 proteins, related eukaryotic S9 proteins and possible members of a family of RNA-binding proteins. Amino acid types that are strongly conserved among eubacteria are shown in bold outlined with light boxes; amino acids that are strongly conserved among eukaryotes are shown in bold in gray boxes, and residues that are conserved across phylogenetic domains are shown in white on black boxes. Only the strongly conserved region is shown. Numbering in the first line for each class corresponds to the first protein listed. The secondary structure of S4 Δ41 from *B.stearothermophilus* is shown above the eubacterial sequences for reference.

chains Glu74 and Glu84, respectively, in the nearby helices.

The distribution of positive charges on the protein surface gives another clue as to where RNA might bind to S4 Δ41

Another rationale for suggesting a nucleic acid-binding site is to look for excess positive charge on the surface of the protein, appropriate for interaction with the negatively charged phosphate backbone. The map of electrostatic potential onto the surface of S4 Δ41 produces a striking result—one face of the molecule is generally negative while the opposite face has a strong concentration of positive charge (Figure 8). This is entirely consistent with our observation that protein aggregation is reduced by increasing the salt concentration; the high salt shields the charges and reduces electrostatic attraction and aggregation. On one face of the protein, negative side chains extending toward the surface include Glu181, Glu184 and Glu188 in the sequence linking strand β5 to helix α7, and Asp168, Glu170 and Glu173 near the turn between strands β4 and β5. On the opposite side of the molecule, a concentration of positive charge is created when the two subdomains come together: Arg66, Arg69 and Lys70 create a positive surface on helix α2 from the helical subdomain, while Arg107 and Arg108, with Arg111 one turn later, create a positive surface on helix α5 from the sheet-containing subdomain. Of these positively charged side chains, Arg111 is conserved across phylogenetic domains (Figure 7). Note that if this positive surface is part of the RNA-binding site, the cleft may need to adjust to accommodate the RNA, changing the relative orientation of the subdomains. Such a change in the orientation of the subdomains may be facilitated by the millisecond timescale motions already mentioned at Ala187, a residue in the bridge between strand β5 and helix α7 that connects the subdomains.

Comparison with the X-ray structure

Comparison of the solution structure with the independently solved crystal structure (C.Davies and S.White, personal communication) reveals that each subdomain superimposes very well: the pairwise r.m.s.d. for the backbone in the helical subdomain (residues 50–72,

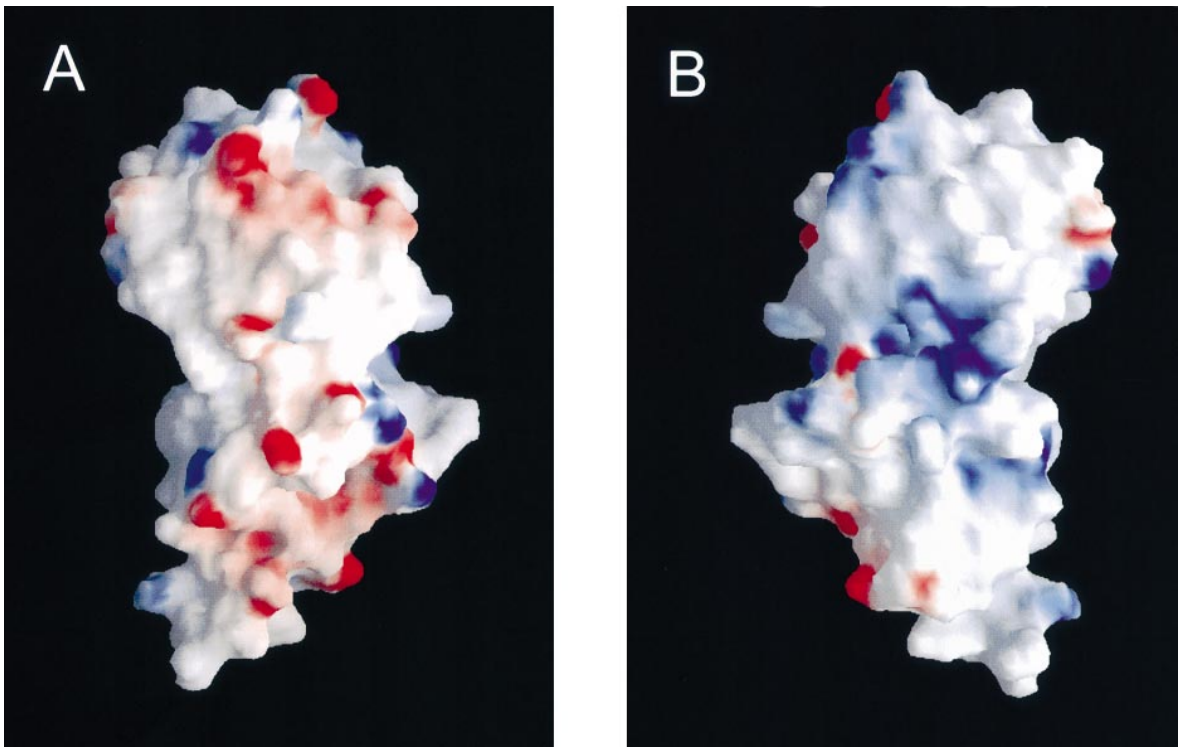

Fig. 8. The electrostatic potential at the surface of S4 Δ41. (**A**) Surface of S4 Δ41, with the molecule rotated +30° about the vertical axis in the page compared with the view in Figure 2A. (**B**) The opposite side of S4 Δ41. The molecule is rotated 180° compared with (A). The surfaces are color coded, with red indicating net negative electrostatic potential and blue indicating positive potential. Regions with a high concentration of positive charge are good candidates for interaction with the negatively charged phosphate backbone of RNA. This figure was generated with GRASP (Nicholls *et al*., 1991).

85–92 and 193–195) is 0.92 Å and for the backbone in the sheet-containing subdomain (residues 94–157 and 161–180) is 1.23 Å. However, the r.m.s.d. for backbone atoms in the full protein is rather high: 3.31 Å for residues 50–72, 85–157, 161–180 and 193–195. Careful examination of the structures reveals that the helical subdomain is rotated by ~45° about the long axis of the molecule and shifted down ~2 Å and to the right ~2 Å in the crystal structure compared with the solution structure, from the orientation of Figure 2A.

Efforts to understand which NMR restraints determine the relative orientation of the subdomains produced a surprise. Omitting 173 NOEs violated in the crystal structure by >0.5 Å from the calculations shifted the ensemble of structures toward the crystal structure (with an average pairwise r.m.s.d. for backbone atoms in residues 47–199 of 1.89 Å to the crystal structure and 3.03 Å to the original NMR structure) without significantly increasing the r.m.s.d. within the ensemble of structures. Reducing the number of restraints in general would be expected to increase the r.m.s.d. within the bundle of structures. Additional calculations revealed that the r.m.s.d. for residues 47–199 could be increased by more than a factor of 2 even as the subdomains themselves remained well defined and the overall energies increased only 10%, simply by modifying the simulated annealing protocol (with all of the original NOE restraints included). We conclude that the orientation of the subdomains is not well specified by the data analyzed so far. To examine the relative orientation of the subdomains in solution further, we plan additional analysis of NOESY data and measurements of dipolar splittings of samples in liquid crystals (Bax and Tjandra, 1997; Tjandra and Bax, 1997). To determine if the intersubdomain orientation changes on a slow time scale, we plan to measure ^{15}N $T_{1\rho}$ relaxation time constants as a function of spin locking field strength.

Although further study is required to decide whether the orientation of the subdomains in solution is actually different from that in the crystal, the general conclusions regarding the homology of the individual subdomains and the distribution of surface charge are not affected by the present uncertainty in the subdomain orientation in solution. Better experimental evidence for multiple orientations of the subdomains is required before we can speculate on the significance of reorientation for RNA binding.

Conclusion

The structure of S4 Δ41 in solution reveals a novel fold, featuring two globular subdomains that form an elongated molecule. Many surfaces with interesting conformations and electrostatic characteristics are available for interactions with RNA and possibly proteins, consistent with S4's known functions in binding 16S rRNA, its own mRNA, and facilitating assembly of the small ribosomal subunit. Our results point to additional studies probing both the relative orientation of these surfaces in solution and ligand interactions with these surfaces that should provide deeper insight into how S4 Δ41 facilitates ribosome assembly and regulates its own translation.

Materials and methods

Sample preparation

Residues 43–200 of the *B.stearothermophilus* sequence for ribosomal protein S4, plus an initiator methionine, were cloned into the pET13a expression vector (Studier *et al*., 1990) and expressed in *E.coli* strain BL21 (DE3). The numbering of the full-length protein (based on the unpublished sequence of S.E.Gerchman and V.Ramakrishnan) has been retained, with 42 as the initiator methionine. Note that resonances for the N-terminal methionine have been observed and assigned, indicating that this methionine was not cleaved during expression. To produce labeled samples, the bacteria were expressed in M9 minimal medium containing $^{15}NH_4Cl$ (Cambridge Isotope Laboratories, Cambridge, MA) and D-[$^{13}C_6$]glucose (Isotec Inc., Miamisburg, OH) as appropriate. Details of the purification are similar to those described by Baker and Draper (1995). Briefly, cells were lysed with a French pressure cell. The expressed protein was found in the supernatant after spinning down the cell debris. The protein was purified by high-performance liquid chromatography on a TSK SP-5-PW column (Bio-Rad Laboratories).

To determine the best solution conditions for NMR spectroscopy, the temperature, pH, salt concentration and protein concentration for ^{15}N-labeled S4 Δ41 were varied systematically (sample conditioning, Wagner, 1993). The results were judged based on the appearance of the ^{15}N HSQC spectrum and the measured values of the ^{15}N T_2 relaxation time constants. Sample conditioning suggested that relatively high concentrations of salt and low concentrations of protein were desirable. For NMR spectroscopy, S4 Δ41 was dissolved in 20 mM deuterated acetate at pH 5.4 and 250 mM KCl. Sodium azide was added to 0.1 mM to inhibit bacterial growth, and 0.1 mg/ml phenylmethylsulfonyl fluoride (PMSF) was included in some samples to prevent proteolysis. Protein concentrations, based on an extinction coefficient of 10 240/M/cm at 280 nm (Gill and von Hippel, 1989), ranged from 450 μM to 1.14 mM. H_2O samples included D_2O to 6% to provide a lock signal. Samples were generally prepared by ultrafiltration in Centricon 10 concentrators (Amicon, Inc., Beverly, MA) and examined in NMR microtubes (Shigemi, Inc., Allison Park, PA).

NMR spectroscopy

NMR spectra were acquired on DMX 500 and DMX 750 spectrometers equipped with pulsed-field gradients (Bruker Instruments, Billerica, MA). Spectra were recorded at 310 K (37°C). Data were processed with NMRPipe and NMRDraw (Delaglio *et al*., 1995) and analyzed with PIPP (Garrett *et al*., 1991) on SPARC and UltraSPARC computer workstations (Sun Microsystems, Mountain View, CA).

Chemical shift assignments

Sequential backbone assignments were based on scalar couplings through the backbone (Kay *et al*., 1990), specifically the HNCACB (Wittekind and Mueller, 1993), CBCA(CO)NH (Grzesiek and Bax, 1992a) and HNCA (Grzesiek and Bax, 1992b) experiments. Backbone assignments were checked using carbonyl chemical shift correlations, from the HCACO (Powers *et al*., 1991) and HNCO (Grzesiek and Bax, 1992b) experiments, and further confirmed by searching for characteristic sequential cross-peaks from the backbone amide proton to the preceding amide or α proton (Wüthrich, 1986) in the ^{15}N-separated NOESY experiments.

The HNCACB/CBCA(CO)NH experiments, used for sequential assignments, also provide information about the $^{13}C_\beta$ chemical shift. Assignments were extended to the $^1H_\beta$ position with the HBHA(CBCACO)NH (Grzesiek and Bax, 1993) and HNHB (Archer *et al*., 1991) experiments. Carbon and proton assignments for positions further down the side chain were obtained with the C(CO)NH and H(CCO)NH experiments (Grzesiek *et al*., 1993), respectively. To resolve any ambiguities from overlap in the backbone amide signals, side chain spin system connectivities were also examined in the HCCH-TOCSY spectrum (Bax *et al*., 1990). In rare cases, ^{1}H assignments within the main amino acid spin system were based on a consistent pattern observed in the NOESY spectra. Spin systems for the aromatic rings, as well as arginine ε protons, asparagine and glutamine amide groups, and methionine methyl groups, were connected to the main spin systems via NOESY cross-peaks. Connections for the asparagine and glutamine side chain amides were also confirmed in the CBCA(CO)NH experiment, and carbonyl assignments for some of those side chains were obtained from the HNCO and HCACO experiments.

Stereospecific assignments and determination of χ_1

For the β-methylene protons, stereospecific assignments were determined from the strengths of the $^3J_{NH\beta}$ and $^3J_{H\alpha H\beta}$ coupling constants, measured with the HNHB (Archer *et al*., 1991) and HACAHB-COSY (Grzesiek *et al*., 1995) experiments, respectively. The pattern of stronger and weaker NOE cross-peaks from the amide and α protons to the β protons was also examined to confirm conclusions based on the coupling constants. Wherever β-methylene protons could be assigned stereospecifically, the χ_1 angle was constrained to one of the staggered conformations (+60°, +180°, –60°) ± 40°. χ_1 was also constrained for some valine, isoleucine and threonine side chains, based on the $^3J_{NH\beta}$ and $^3J_{H\alpha H\beta}$ coupling constants.

Stereospecific assignments for the prochiral methyl groups of valine and leucine were determined readily from the splitting patterns of methyl cross-peaks in the ^{13}C HSQC spectrum of a 10% ^{13}C-labeled sample (Neri *et al*., 1989; Senn *et al*., 1989). Due to the high resolution of the spectrum, which was acquired at 750 MHz, both methyl groups for 17 out of 18 leucines and eight out of nine valines could be identified and assigned without ambiguity. Although only one methyl cross-peak could be resolved in the ^{13}C HSQC spectrum for Leu94, it provides stereospecific assignments for both methyl groups. Only the methyl groups of Val150, with nearly completely degenerate proton and carbon chemical shifts, could not be assigned stereospecifically.

Restraints on the ϕ backbone dihedral angle

The $^3J_{HNH\alpha}$ coupling constants were determined from the HNHA (Vuister and Bax, 1993; Kuboniwa *et al*., 1994) experiment. The backbone dihedral angle φ was restrained to –65 ± 25° for $^3J_{HNH\alpha}$ <5.5 Hz and to –120 ± 40° for $^3J_{HNH\alpha}$ >8.0 Hz. No restraints were included for Ala187 and Glu188, despite coupling constants of 5.1 and 5.3 Hz, respectively, because both positive and negative φ angles, each predicting coupling constants around 5.0 Hz, are well tolerated at either position. These residues follow Pro186 as part of an apparently irregular structure.

NOE information

Four three-dimensional ^{15}N-separated NOESY spectra were acquired for S4 Δ41: one at 500 MHz, one at 750 MHz, one with ^{13}C evolution in the indirect dimension and one with ^{15}N evolution in the indirect dimension. By cross-checking possible assignments for the chemical shift of both the proton and the attached heteronucleus (^{13}C or ^{15}N), using the proton readout spectra together with the ^{13}C and ^{15}N readout spectra, a set of virtually unambiguous initial assignments was established. Comparing the proton readout spectra, the spectrum at 750 MHz has higher resolution, but the spectrum at 500 MHz has some stronger signals due to the longer acquisition and longer mixing time (75 ms versus 60 ms). Two ^{13}C-separated NOESY spectra, acquired in D_2O, were analyzed in detail for S4 Δ41: a 3D spectrum acquired at 750 MHz and a 4D spectrum acquired at 500 MHz. The 4D experiment proved essential for providing nearly unambiguous assignments at the early stages of structure determination.

NOE distance restraints were derived from NOE cross-peak intensities, using reference distances from elements of regular secondary structure and an inverse 4th power relationship between distance and intensity, instead of the theoretical inverse 6th power, to account for spin diffusion in a general way (Suri and Levy, 1993). Cross-peaks were classified as strong, medium, weak and very weak, corresponding to distances of 2.5, 3.0, 4.0 and 5.0 Å with upper limits extended by 0.2, 0.3, 1.0 and 1.0 Å, respectively. Pseudoatom corrections were added for methyl groups, aromatic rings and non-stereoselectively assigned methylene protons (Wüthrich, 1986). After the initial rounds of structure calculations, based on NOE cross-peaks between the most unique chemical shifts, a set of calculated structures with low restraint violations was used to discriminate between possible cross-peak assignments. Potential assignments that gave distances >10 Å were not considered further. As the amount of information in the calculated structures increased, this cutoff distance was reduced to 6 Å.

Amide proton exchange

A lyophilized sample of ^{15}N-labeled S4 Δ41 was dissolved in D_2O and monitored by ^{15}N HSQC experiments, recorded with gradient water suppression and water flip-back in 20 min. Amide protons were considered to exchange slowly if they were still visible in the spectrum recorded 45–65 min after adding D_2O. Twenty one backbone amide protons (in residues Leu94, Asn96, Leu97, Val98, Tyr99, Arg100, Leu101, Gly102, Leu103, Ala104, Ala110, Val114, Ile119, Leu120, Val121, Val133, Gln137, Ile139, Ala140, Val141 and Leu165) exchange very slowly, with lifetimes of the order of months. For each slowly exchanging amide proton, two distance restraints representing a hydrogen bond (N to O and NH to O) were included if only a single hydrogen bonding partner was identified in preliminary structures and if that hydrogen bond corresponded to one expected in regular secondary structure.

Structure calculations for S4 Δ41

Structures were calculated with X-PLOR 3.8 (Brünger, 1992) using the hybrid distance geometry/simulated annealing protocol. The potential energy function featured terms for NOE restraints and dihedral angle restraints in addition to chemically derived terms for bond lengths, bond angles, improper torsion angles and non-bonded interactions. The number of steps and the initial temperatures were increased compared with the default values for all simulated annealing steps in the protocol to improve convergence. Of 50 structures calculated, all 50 converged, with no restraint violations above 0.5 Å nor 5°. The best 26 structures, with no violations above 0.2 Å nor 2°, have been analyzed in detail.

Additional information

Supplementary tables containing the chemical shift assignments and the details of the NMR experiments are available from the authors.

Acknowledgements

We thank Frank Delaglio and Dan Garrett for the computer programs used to process and analyze NMR data, York Tomita for computer support, Shengrong Huang for enthusiastically encouraging the structural studies of S4 Δ41 and for many interesting discussions, and Chris Davies and Stephen W.White for sharing the coordinates for the X-ray structure of S4 Δ41 prior to publication. This work was supported in part by the AIDS Targeted Anti-Viral Program of the Office of the Director of the National Institutes of Health.

References

Allen,P.N. and Noller,H.F. (1989) Mutations in ribosomal proteins S4 and S12 influence the higher order structure of 16 S ribosomal RNA. *J. Mol. Biol*., **208**, 457–468.

Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Archer,S.J., Ikura,M., Torchia,D.A. and Bax,A. (1991) An alternative 3D NMR technique for correlating backbone ^{15}N with side chain Hβ resonances in larger proteins. *J. Magn. Resonance*, **95**, 636–641.

Baker,A.-M. and Draper,D.E. (1995) Messenger RNA recognition by fragments of ribosomal protein S4. *J. Biol. Chem.*, **270**, 22939–22945.

Bax,A. and Tjandra,N. (1997) High-resolution heteronuclear NMR of human ubiquitin in an aqueous liquid crystalline medium. *J. Biomol. NMR*, **10**, 289–292.

Bax,A., Clore,G.M. and Gronenborn,A.M. (1990) ^{1}H–^{1}H correlation via isotropic mixing of ^{13}C magnetization, a new three-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins. *J. Magn. Resonance*, **88**, 425–431.

Brünger,A.T. (1992) *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.

Deckman,I.C. and Draper,D.E. (1985) Specific interaction between ribosomal protein S4 and the α operon messenger RNA. *Biochemistry*, **24**, 7860–7865.

Deckman,I.C. and Draper,D.E. (1987) S4-α mRNA translation regulation complex. II. Secondary structures of the RNA regulatory site in the presence and absence of S4. *J. Mol. Biol*., **196**, 323–332.

Deckman,I.C., Thomas,M.S. and Draper,D.E. (1987) S4-α mRNA translation repression complex. I. Thermodynamics of formation. *J. Mol. Biol*., **196**, 313–322.

Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.

Dodd,J. and Hill,W.E. (1987) Physical characteristics of ribosomal protein S4 from *Escherichia coli*. *J. Biol. Chem*., **262**, 2478–2484.

Garrett,D.S., Powers,R., Gronenborn,A.M. and Clore,G.M. (1991) A common sense approach to peak picking in two-, three- and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Resonance*, **95**, 214–220.

Gill,S.C. and von Hippel,P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.*, **182**, 319–326.

Grzesiek,S. and Bax,A. (1992a) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.*, **114**, 6291–6293.

Grzesiek,S. and Bax,A. (1992b) Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J. Magn. Resonance*, **96**, 432–440.

Grzesiek,S. and Bax,A. (1993) Amino acid type determination in the sequential assignment procedure of uniformly ^{13}C/^{15}N-enriched proteins. *J. Biomol. NMR*, **3**, 185–204.

Grzesiek,S., Anglister,J. and Bax,A. (1993) Correlation of backbone amide and aliphatic side-chain resonances in ^{13}C/^{15}N-enriched proteins by isotropic mixing of ^{13}C magnetization. *J. Magn. Resonance*, **B101**, 114–119.

Grzesiek,S., Kuboniwa,H., Hinck,A.P. and Bax,A. (1995) Multiple-quantum line narrowing for measurement of H^{α}–H^{β} *J* coupling in isotopically enriched proteins. *J. Am. Chem. Soc.*, **117**, 5312–5315.

Hinck,A.P., Markus,M.A., Huang,S., Grzesiek,S., Kustnovich,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 23 S rRNA. *J. Mol. Biol.*, **274**, 101–113.

Holm,L. and Sander,C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.

Kay,L.E., Ikura,M., Tschudin,R. and Bax,A. (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Resonance*, **89**, 496–514.

Kisker,C., Hinrichs,W., Tovar,K., Hillen,W. and Saenger,W. (1995) The complex formed between Tet repressor and tetracycline-Mg^{2+} reveals mechanism of antibiotic resistance. *J. Mol. Biol.*, **247**, 260–280.

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 the 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.

Kuboniwa,H., Grzesiek,S., Delaglio,F. and Bax,A. (1994) Measurement of H^{N}–H^{α} J couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods. *J. Biomol. NMR*, **4**, 871–878.

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.

Liebman,S.W., Chernoff,Y.O. and Liu,R. (1995) The accuracy center of a eukaryotic ribosome. *Biochem. Cell Biol*., **73**, 1141–1149.

Malhotra,A., Severinova,E. and Darst,S.A. (1996) Crystal structure of a σ^{70} subunit fragment from *E.coli* RNA polymerase. *Cell*, **87**, 127–136.

Markus,M.A., Hinck,A.P., Huang,S., Draper,D.E. and Torchia,D.A. (1997) High resolution solution structure of ribosomal protein L11-C76, a helical protein with a flexible loop that becomes structured upon binding to RNA. *Nature Struct. Biol.*, **4**, 70–77.

Neri,D., Szyperski,T., Otting,G., Senn,H. and Wüthrich,K. (1989) Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling. *Biochemistry*, **28**, 7510–7516.

Nicholls,A., Sharp,K.A. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct., Function Genet.*, **11**, 281–296.

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.

Powers,R., Gronenborn,A.M., Clore,G.M. and Bax,A. (1991) Three-dimensional triple-resonance NMR of ^{13}C/^{15}N-enriched proteins using constant time evolution. *J. Magn. Resonance*, **94**, 209–213.

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.

Senn,H., Werner,B., Messerle,B.A., Weber,C., Traber,R. and Wüthrich,K. (1989) Stereospecific assignments of the methyl ^{1}H NMR lines of valine and leucine in polypeptides by nonrandom ^{13}C labelling. *FEBS Lett*., **249**, 113–118.

Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.*, **185**, 60–89.

Suri,A.K. and Levy,R.M. (1993) Estimation of interatomic distances in proteins from NOEe spectra at longer mixing times using an empirical two-spin equation. *J. Magn. Resonance*, **B101**, 320–324.

Tjandra,N. and Bax,A. (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science*, **278**, 1111–1114.

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.

Vartikar,J. and Draper,D.E. (1989) S4–16 S ribosomal RNA complex. Binding constant measurements and specific recognition of a 460-nucleotide region. *J. Mol. Biol.*, **209**, 221–234.

Vuister,G.W. and Bax,A. (1993) Quantitative *J* correlation: a new approach for measuring homonuclear three-bond $J(H^N–H^\alpha)$ coupling constants in ^{15}N-enriched proteins. *J. Am. Chem. Soc.*, **115**, 7772–7777.

Wagner,G. (1993) Prospects for NMR of large proteins. *J. Biomol. NMR*, **3**, 375–385.

Wishart,D.S. and Sykes,B.D. (1994) The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR*, **4**, 171–180.

Wittekind,M. and Mueller,L. (1993) HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *J. Magn. Resonance.*, **B101**, 201–205.

Wüthrich,K. (1986) *NMR of Proteins and Nucleic Acids*. John Wiley & Sons, New York.

Received May 5, 1998; revised May 27, 1998; accepted June 2, 1998