# Solution Structure of Prokaryotic Ribosomal Protein S17 by High-Resolution NMR Spectroscopy†

T. N. Jaishree,‡,§ V. Ramakrishnan,*,||,⊥ and Stephen W. White*,‡

*Department of Microbiology, Box 3020, Duke University Medical Center, Durham, North Carolina 27710, and Department of Biology, Brookhaven National Laboratory, Upton, New York 11973*

*Received May 10, 1995; Revised Manuscript Received January 3, 1996*⊗

ABSTRACT: The solution structure of a primary 16S rRNA-binding ribosomal protein, S17, was investigated by two- and three-dimensional homonuclear and heteronuclear magnetic resonance spectroscopy. Almost complete chemical shift assignments for the $^{1}$H, $^{15}$N, and $^{13}$C resonances have been obtained. The NMR data have been rigorously analyzed using a combination of distance geometry, back-calculation, and simulated annealing refinement techniques, and a high-resolution three-dimensional structure has been deduced. The protein consists of a single twisted antiparallel $\beta$-pleated sheet with Greek-key topology. The five $\beta$-strands are connected by extended loops that are flexible compared to the $\beta$-sheet core structure and appear not to adopt one definite conformation in solution. Two of these loops contain many of the residues that have been implicated in binding ribosomal RNA. The location and distribution of these residues and other positively charged side chains on the protein surface suggest an interaction with two distinct regions of ribosomal RNA.

Structural studies on individual ribosomal proteins have been important for understanding the mechanism of the translational apparatus. The detailed information from these studies can be used to supplement and confirm our understanding of the structure/function aspects of the ribosome obtained by other techniques such as chemical cross-linking (Lambert et al., 1983; Brimacombe et al., 1990), chemical probing (Stern et al., 1989), electron microscopy (Stöffler-Meilicke & Stöffler, 1990; Frank et al., 1995; Stark et al., 1995), neutron scattering (Capel et al., 1987), and X-ray crystallography (Yonath, 1992). The structures of several ribosomal proteins have now been determined, and these have provided invaluable insights into rRNA[1] −protein interactions (Ramakrishnan & White 1992; Golden et al., 1993a; Hoffman et al., 1994; Lindahl et al., 1994), antibiotic resistance (Ramakrishnan & White 1992; Golden et al., 1993a), and ribosome evolution (Leijonmarck et al., 1980, 1988; Wilson et al., 1986; Golden et al., 1993a; Hoffman et al., 1994; Lindahl et al., 1994). The eventual goal of this work is to incorporate the component structures into the constantly improving models of the ribosome and to understand this ribonucleoprotein complex at the molecular level.

During the last 10−15 years, it has become increasingly apparent that the ribosomal RNA is the important functional component of the ribosome (Noller, 1991; Noller et al., 1992; Purohit & Stern, 1994) and that one of the principal roles of the proteins is to ensure that the RNA folds into its correct three-dimensional structure (Stern et al., 1988a,b). On the basis of this criterion, the proteins vary with respect to their importance to the assembly and integrity of the ribosome. Most important are the primary rRNA-binding proteins that recognize specific sequences on the rRNA and initiate its folding (Stern et al., 1989). These proteins are generally difficult to remove from the ribosome in a nondenatured state and, until quite recently, were unavailable for structural studies. However, due to our development of cloning techniques for ribosomal proteins (Ramakrishnan & Gerchman, 1991), this is no longer a practical limitation. Here we present a high-resolution structure of a *bona-fide* primary rRNA-binding ribosomal protein, S17 from the small (30S) subunit. As with our previous structural studies, the protein is derived from the thermophilic bacterium *Bacillus stearothermophilus*. This work represents a continuation of our previously reported studies on S17 at lower resolution (Golden et al., 1993b).

With a molecular mass of 10 000 Da, S17 is one of the smaller ribosomal proteins, and its fundamental importance to the ribosome is reflected in its high conservation across all life forms (Golden et al., 1993b). It is located toward the "bottom" of the small subunit in a region that is relatively free of other protein components (Brimacombe et al., 1988; Stern et al., 1988b), and point mutations in the protein that affect subunit assembly (Herzog et al., 1979) and translational fidelity (Yaguchi et al., 1976) support a crucial organizational role. Clear evidence that S17 is a primary rRNA-binding

† This work was supported by Grant GM44973 from the National Institutes of Health and a generous award from the Rippel Foundation.

* To whom correspondence should be addressed.

‡ Duke University Medical Center.

§ Present address: Cellular Biochemistry & Biophysics Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10021.

|| Brookhaven National Laboratory.

⊥ Present address: Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, UT 84132.

⊗ Abstract published in *Advance ACS Abstracts,* February 15, 1996.

[1] Abbreviations: DDS, sodium 4,4-dimethyl-4-silapentane-1-sulfonate; DQF-COSY, double-quantum-filtered correlated spectroscopy; 2D, two dimensional; 3D, three dimensional; DNA, deoxyribonucleic acid; HMQC-J, heteronuclear multiple-quantum coherence *J*-correlated spectroscopy; HMQC, heteronuclear multiple-quantum coherence spectroscopy; HSMQC, heteronuclear single- and multiple-quantum coherence spectroscopy; HSQC, heteronuclear single-quantum coherence spectroscopy; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy; PCR, polymerase chain reaction; rmsd, root mean square deviation; rRNA, ribosomal RNA; TOCSY, total correlated spectroscopy; TSP, 3-(trimethylsilyl)propanesulfonic acid, sodium salt.

0006-2960/96/0435-2845$12.00/0 

protein came from studies on ribosome assembly (Held et al., 1974; Zimmermann et al., 1974; Mackie & Zimmermann, 1978), and subsequent studies have localized the probable RNA-binding site to a small stem−loop structure (Greuer et al., 1987; Stern et al., 1988a; Weitzmann et al., 1993; Powers & Noller, 1995). Indeed, it was the compactness of the cognate RNA site that made S17 such an attractive molecule for investigating the RNA−protein interactions within the ribosome.

## EXPERIMENTAL PROCEDURES

*Cloning, Expression, and Purification of S17.* The S17 gene was cloned from *B. stearothermophilus* genomic DNA by PCR methods using the procedure described earlier (Ramakrishnan & Gerchman, 1991; Golden et al., 1993b). For protein expression, cells were grown in different media depending on the labeling conditions required for the various NMR experiments. In the case of unlabeled protein for use in $^1$H NMR experiments, the cells were grown in Luria broth. For uniformly $^{15}$N-labeled protein samples, the cells were grown in a defined medium containing $^{15}NH_4Cl$ (5 g/L) as the sole nitrogen source. Finally, for uniformly $^{15}$N/$^{13}$C-labeled protein, the cells were grown in a defined medium containing $^{15}NH_4Cl$ (5 g/L) as the sole nitrogen source and [$^{13}$C]acetate (4 g/L) as the sole carbon source using procedures described previously (Venters et al., 1991; Golden et al., 1993b). The procedures for purifying the protein and preparing the NMR samples are as described earlier (Golden et al., 1993b).

*NMR Spectroscopy.* For the present structural analysis, it was not necessary to collect any 2D NMR spectra other than those already acquired. A full description of these data acquisitions has already been reported (Golden et al., 1993b). However, it should be noted that 1024 complex points were collected in the $t_2$ dimension and 512 in $t_1$ and that a mixing time of 200 ms was used for the NOESY experiments. Also, the 2D data were processed using FELIX 1.1 (Hare Research Inc., Woodenville, WA).

The 3D NMR spectra were acquired on a Varian unity 600 MHz spectrometer. This instrument enabled gradient-enhanced versions (Bax & Pochapsky, 1992) of the following spectra to be collected in water: 3D CBCA(CO)NH (Grzesiek & Bax, 1992a, 1993), HNCACB (Grzesiek & Bax, 1992b; Wittekind & Mueller, 1993; Yamazaki et al., 1994; Muhandiram & Kay, 1994), HCCH-TOCSY (Kay et al., 1993), $^{15}$N-edited HMQC-NOESY (Fesik & Zuiderweg, 1988), and $^{13}$C-edited HSQC-NOESY (Majumdar & Zuiderweg, 1993). Sweep widths of 7000.4, 9000.9, and 2200 Hz were used for the $^1$H, $^{13}$C, and $^{15}$N dimensions, respectively. Referencing was done with respect to DDS for $^1$H, 2.9 M $^{15}NH_4Cl$ in 1 M HCl (24.93 ppm) for $^{15}$N, and TSP for $^{13}$C. For the CBCA(CO)NH and HNCACB experiments, 40 increments were collected in the carbon and nitrogen dimensions. The 3D HCCH-TOCSY was acquired using 57 increments in the carbon dimension and a spin lock time of 18.8 ms. For the $^{15}$N- and $^{13}$C-edited 3D NOESY experiments, 32 and 40 increments were collected in the nitrogen and carbon dimensions, respectively, with mixing times of 100 ms. All spectra were collected at 25 °C. The data were processed using a Gaussian function for apodization in all three dimensions on a SPARC station using VNMR 4.0 (Varian Nuclear Magnetic Resonance Instruments, Palo Alto, CA). Linear prediction was used in the carbon and nitrogen dimensions where applicable.

*Structure Calculation.* An initial distance geometry model was built by assembling sets of distance, hydrogen-bonding, and $\phi$-angle constraints. The distance constraints were generated by searching the $^1$H NOESY spectra for unambiguous cross-peaks that show no overlap and then confirming them with the $^{15}$N- and $^{13}$C-edited NOESY spectra. This procedure generated 250 distance constraints that were classified into small (2−3 Å), medium (2−4 Å), and large (2−5 Å) depending on the intensity of the corresponding integrated NOE peaks. The H-bond constraints were applied to the amide protons that were judged nonexchangeable on the basis of $D_2O$ exchange experiments, and the dihedral $\phi$-angle constraints were derived from an HMQC-J spectrum. Apart from a few additions and deletions, these constraints were very similar to those used for the preliminary analysis of the S17 structure (Golden et al., 1993b). The structures obtained from distance geometry were subjected to simulated annealing in X-PLOR (Brünger, 1992) using the NOE constraints applied as a soft-square NOE potential with a scale factor of 50. In preparation for the next stage, an average set of coordinates was calculated from these structures and subjected to additional simulated annealing. This produced a starting model for determining the maximum number of accurate NOE constraints using a back-calculation refinement protocol. The following procedure, which uses the well-accepted method of complete relaxation matrix analysis (Macura & Ernst, 1980; Post et al., 1990), has been applied many times to structural calculations on proteins and nucleic acids (Bonvin et al., 1993, 1994; Hoffmann et al., 1993; Knegtel et al., 1993; Santoro et al., 1993; Costa et al., 1994; Frechet et al., 1994; Robinson et al., 1994).

The first step was to extract the integrated peaks from the 2D NOE data set. This was done using the known shapes of each spin in the $\omega_1$ and $\omega_2$ dimensions as determined by the spectral analysis program MYLOR (Robinson & Wang, 1992). The NOE constraints were obtained by the sequence of iterative steps that comprise the SPEDREF package (Robinson & Wang, 1992). The first step involves the calculation of the NOEs from the model by the MORASS program using a complete relaxation matrix analysis (Macura & Ernst, 1980; Post et al., 1990). The isotropic rotational correlation time was obtained by an empirical comparison between the experimental and calculated NOE volumes at different correlation times and choosing the value that gave the minimum errors. Similar procedures have been used by other groups (Robinson & Wang, 1992; Schultze et al., 1994), and the value eventually chosen was 1.2 ns. During the second step, the simulated NOE volumes are used to deconvolute overlapping regions of the experimental spectra. This is a model-dependent deconvolution procedure that works well since the starting model was generated by distance geometry. The new set of observed NOE volumes and associated NOE-derived force springs is then used as inputs to conjugate gradient minimization performed in X-PLOR. This third step uses the biharmonic potential option (Clore et al., 1985, 1986) with an NOE scale factor of 30 and results in a new model that can be recycled through the whole procedure until the model no longer changes and the residual error factor is minimized.

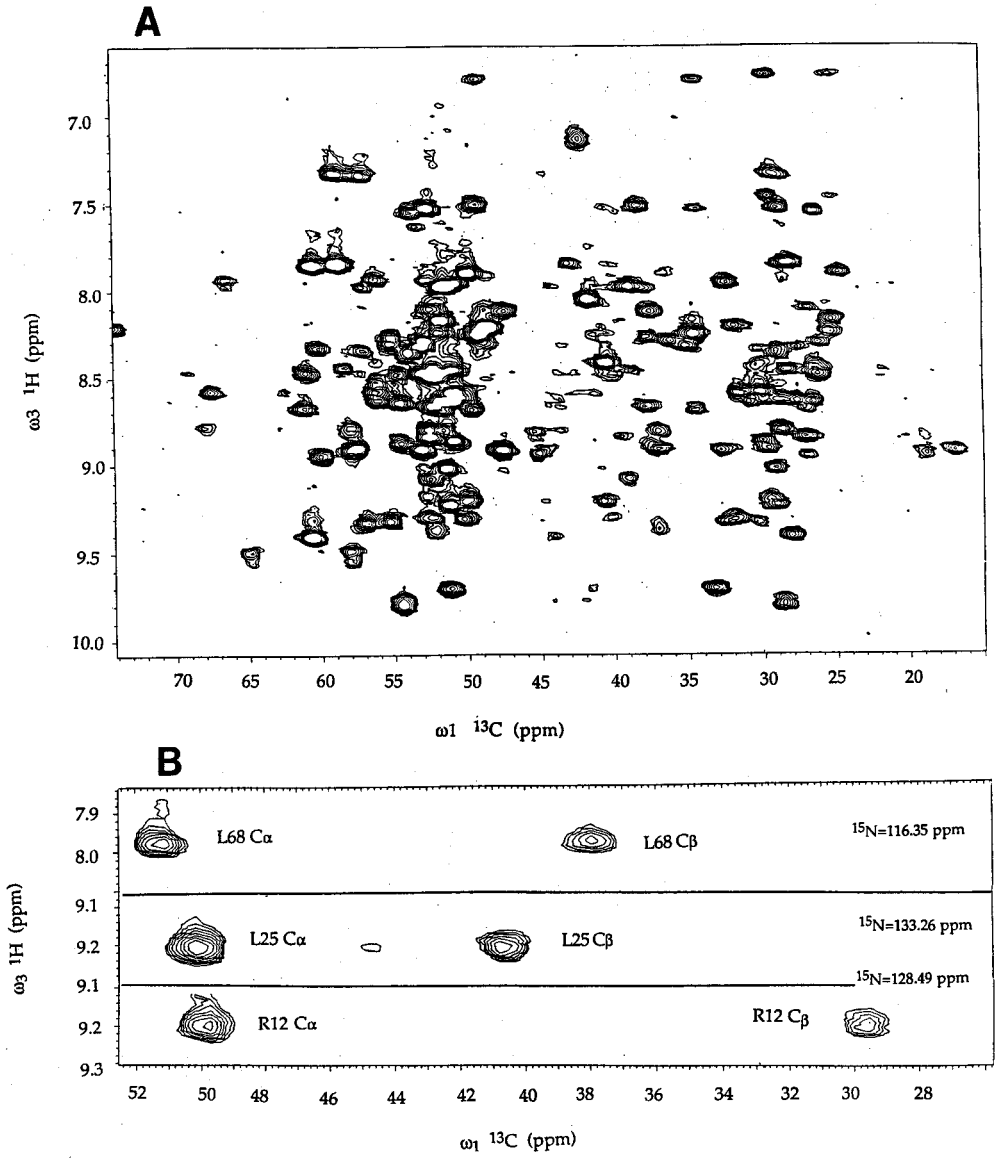


FIGURE 1: The CBCA(CO)NH NMR spectrum of ribosomal protein S17. (A) Compressed projection of all 3D slices along amide nitrogen chemical shifts showing cross-peaks between NH($i$) and $C_\beta(i-1)$ and $C_\alpha(i-1)$. Sweep widths of 7000.4, 9000.9, and 2200 Hz were used for $^1$H, $^{13}$C, and $^{15}$N, respectively. A total of 40 increments were collected in the carbon and nitrogen dimensions. Linear prediction to 80 was used in both these dimensions. (B) Representative 2D slices through the spectrum. The slices are taken at the $^{15}$N chemical shifts which are shown on the upper right-hand corner of each slice. The $C_\alpha$ and $C_\beta$ assignments are shown on the slice.

The SPEDREF procedure eventually extracted a total of 853 NOE constraints, 249 intraresidue ($i = j$), 294 sequential ($i - j = 1$), 85 medium range ($i - j < 5$), and 225 long range ($i - j > 5$). These were applied in two stages to the initial structures obtained by distance geometry to generate a final family of structures. In the first stage, each structure was subjected to the simulated annealing procedure in X-PLOR (Nilges et al., 1988) by heating the molecules to 1000 K, running molecular dynamics for 40 ps, and finally slowly cooling to 300 K in the presence of the NOE constraints represented as biharmonic potentials to obtain 20 structures. During the second stage, each of the models was subjected to multiple iterations of the simulated annealing procedure in X-PLOR using the NOE constraints from SPEDREF as a soft-square potential in order to overcome the problem of using tight restraints and to minimize the number of NOE violations. The final structures were energy minimized using conjugate gradient minimization in X-PLOR.

## RESULTS AND DISCUSSION

*Structure Determination.* The $^1$H assignments were made by conventional methods using the information from $^1$H NOESY, DQF-COSY (Wüthrich, 1986), HNCACB, CBCA-(CO)NH, and HCCH-TOCSY experiments. The $^{15}$N assignments were made using the information from the HSMQC (Zuiderweg, 1990), 2D HMQC-NOESY, and 2D HMQC-TOCSY experiments performed on a uniformly $^{15}$N-labeled sample. It was possible to confirm many of these $^{15}$N assignments using the previously acquired data from the $^{15}$N type-labeled samples (Golden et al., 1993b). The $J_{\alpha NH}$ coupling constants were extracted using HMQC-J spectra as previously reported (Golden et al., 1993b). The $^{13}$C assignments were made using the 3D HNCACB, CBCA-CONH, and HCCH-TOCSY spectra. All assignments are shown in Tables 1 and 2 (supporting information). Figure 1A shows a projection through all the slices of the CBCA-(CO)NH spectrum. It can be seen from this spectrum that the sample is uniformly labeled with $^{13}$C and that the α and

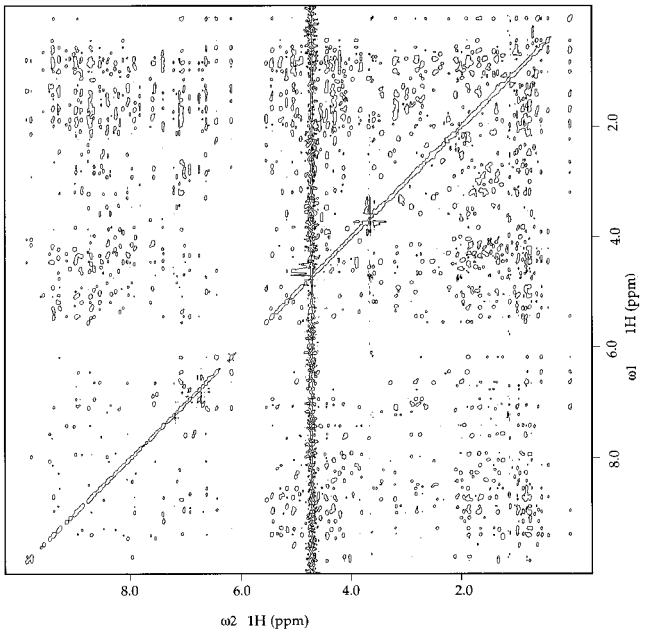


FIGURE 2: The $^1$H 2D NOESY spectrum of ribosomal protein S17 in 90% $H_2O$/10% $D_2O$. The data were acquired at 25 °C and pH 6.5 using the TPPI method (Marion & Wüthrich, 1983). Low-level presaturation of the water signal followed by a SCUBA pulse sequence with a 28 $\mu$s SCUBA delay was used (Brown et al., 1988). The chemical shift referencing was done relative to DDS.

$\beta$ carbons show the usual dispersion. Examples of some slices taken through this spectrum at different $^{15}$N chemical shifts are shown in Figure 1B. Figure 2 shows the complete $^1$H 2D NOESY spectrum in 90% $H_2O$ /10% $D_2O$, and panels A and B of Figure 3 show the amide−α and α−α regions, respectively, of this spectrum. Some stereospecific assignments were made using the initial structures, and the pseudoatom approximations were used in X-PLOR (Brünger, 1992) where appropriate.

The decision to use the 2D NOESY data in the structural analysis was governed by its overall superior quality when compared to the $^{15}$N- and $^{13}$C-edited 3D NOESY data. Also taken into account was the fact that S17 spectra generally display excellent chemical shift dispersion (Figure 2). The only crowded region corresponds to the $\gamma$- and $\delta$-proton resonances, but it was clear from the HCCH-TOCSY that the $\gamma$ and $\delta$ *carbons* were also poorly dispersed and that peak overlap was still a problem in this region of the $^{13}$C-edited NOESY spectrum. Our solution for dealing with the overlapping cross-peaks involving these protons was based on inference from initial models, and the described procedure does this rather efficiently. The combined use of the line-shape fitting program MYLOR, the model-dependent deconvolution procedure, and the fact that the initial model was derived from distance geometry deals with overlaps rather well. The $^{13}$C-edited NOESY data *were* used to eventually verify the findings from the relaxation matrix analysis. It was possible to evaluate many distance restraints from the 3D spectra, and these generally agreed well with the values from the analysis.

An unusual finding from the analysis was the rather low value for the optimal correlation time of 1.2 ns. Typical proteins of the size of S17 normally have a correlation time of around 3 ns, but using this value in our simulations produced large errors. The correlation time is closely linked to the molecular shape, and we believe that the lower value for S17 is consistent with its two extended and perhaps disordered loop regions (see below).

A final total of 853 constraints were eventually used to define the S17 three-dimensional structure, and their distribution between the residues is shown in Figure 4. The family of 20 accepted structures has no NOE violations greater than 0.7 Å, an rmsd on bonds of less than 0.017 Å, and an rmsd on angles of less than 1.7°. These values are comparable to typical crystal structures and also to other NMR structures in the literature (Andersen et al., 1992; Constantine et al., 1992; Bonvin et al., 1994). A superposition of 15 of the structures calculated from NMR data is shown in Figure 5. A set of average coordinates was calculated from this family, and the average rmsd from the mean is 0.8 Å for the core backbone atoms and 1.8 Å for all the backbone atoms. The average of 11 constraints per residue qualifies the complete molecule as a second to third generation structure as defined

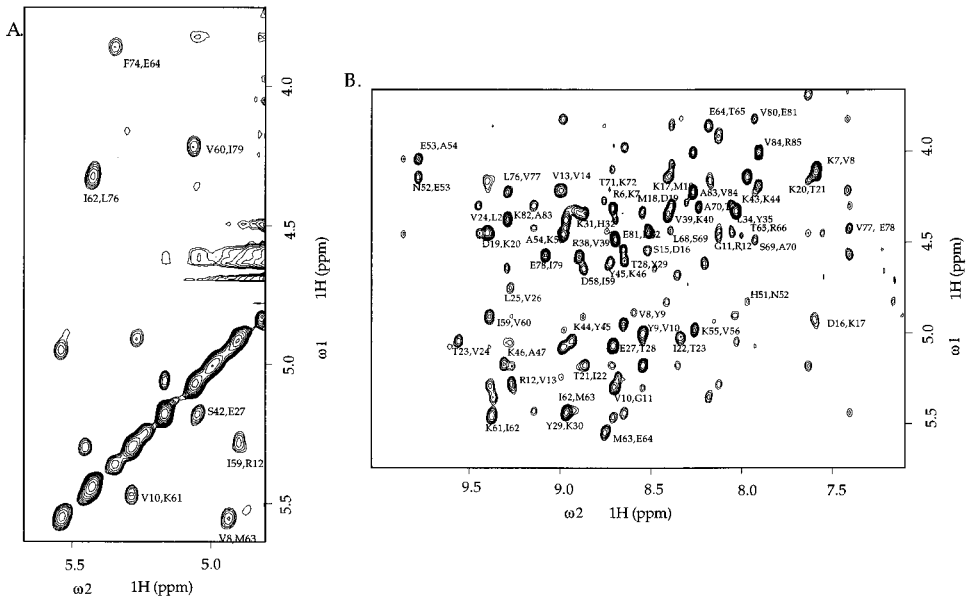


FIGURE 3: Expanded regions of the $^1$H 2D NOESY spectrum of ribosomal protein S17 in 90% $H_2O$/10% $D_2O$. (A) The α−α region showing some crucial α−α NOEs. (B) The α−NH region showing some crucial α−NH NOEs.

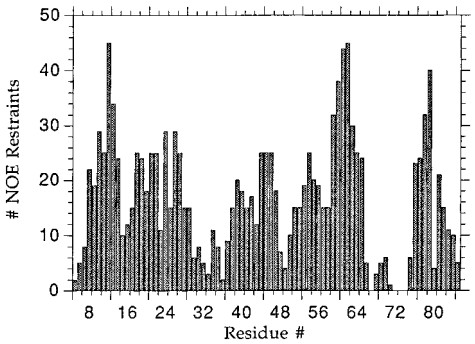


FIGURE 4: Plot showing the distribution of NOE constraints per residue for ribosomal protein S17. The number of NOE constraints are higher for the residues in the barrel-like core region of the protein than for the residues in the loop regions. This suggests an inherent flexibility of the loops (see text for details).

Table 1: NMR and Structural Statistics for S17

| quantity considered | values |
|---|---|
| no. of NOE restraints | |
| intraresidue | 249 |
| sequential | 294 |
| medium range ($i - j < 5$) | 85 |
| long range ($i - j > 5$) | 225 |
| no. of NOE violations | |
| >0.7 Å | 0 |
| >0.5 Å | <10 |
| rmsd NOE violations (Å) | <0.3 |
| rmsd bond lengths from ideal (Å) | <0.017 |
| rmsd bond angles from ideal (deg) | <1.7 |
| rmsd from mean for backbone (Å) | |
| for $\beta$-sheet core | 0.8 |
| for whole protein | 1.8 |
| $E_{elec} + E_{vdw}$ (X-PLOR) (kcal/mol) | $-115 \leftrightarrow -27$ |
| PROCHECK criteria | |
| overall goodness factor | $-0.60 \leftrightarrow -0.50$ |
| % residues in allowed regions of Ramachandran plot | $77.7 \leftrightarrow 84.1$ |

by Clore and Gronenborn (1991). Table 1 shows the NMR and structural statistics for the calculations.

*Description of the Structure.* Figure 6 shows one of the structures in detail. The general fold of the protein agrees well with the structure previously reported (Golden et al., 1993b) and will be only briefly described. S17 consists of a five-stranded antiparallel $\beta$-pleated sheet in which the strands are arranged in a Greek-key topology (Richardson, 1977). Strand $\beta$1 comprises residues 9−15, $\beta$2 residues 21−27, $\beta$3 residues 42−49, $\beta$4 residues 56−67, and $\beta$5 residues 72−80. The $\beta$-sheet is highly twisted in the usual left-handed sense and forms a barrel-like structure with a gap between $\beta$3 and $\beta$5 which a sixth $\beta$-strand could occupy if present. There are three extended connecting loops which contain 15, 8, and 6 residues, respectively, loop 1 residues 27−42, loop 2 residues 49−56, and loop 3 residues 67−72. The higher resolution structure confirms our earlier prediction that $\beta$5 contains a $\beta$-bulge (Golden et al., 1993b). In the $\beta$-bulge, the amide protons of valine 77 and glutamate 78 are slow exchangers and are probably involved in H-bonding interactions (Golden et al., 1993b).

An alignment of all the available S17 amino acid sequences clearly showed that a subset of the hydrophobic residues are highly conserved and were presumed to occupy the core of the molecule (Golden et al., 1993b). The present structure shows that this is indeed correct and provides a convenient internal check of the validity of the structure. Particularly well conserved are valines 13, 24, and 26, isoleucines 22, 59, and 62, and alanine 47, which are at the very center of the molecule. It was also suggested that the totally conserved glycine 11 might occupy a highly crowded region of the core that is typically seen in these barrel-like structures (Murzin, 1993), and this is also the case. With regard to the charged and polar amino acids, these populate the surface of the molecule as would be expected for a typical soluble globular protein.

It can be seen from Figure 4 that there are very few NOE constraints for the loop regions compared to the barrel-like core of the molecule, and it is not possible to assign to them a single conformation. This could be due either to the inherent flexibility of the loop regions or to the absence of distance constraints less than 0.5 Å. An indication of the loops' poorly defined structure is that when excluded, the rmsd of the molecule's backbone drops from 1.8 to 0.8 Å. With more than 15 constraints per residue, the core of the protein qualifies as a third to fourth generation structure (Clore & Gronenborn, 1991).

In the earlier structural analysis of S17, it was pointed out that there may be two alternate conformations of the molecule toward one end of the $\beta$-barrel. This was suggested on the basis of a doubling of a subset of the cross-peaks in the NMR spectra involving amino acids in this region. We did not see this phenomenon in the data sets collected for this analysis, and this may be due to some subtle differences in sample preparations and/or data collection conditions.

*Distribution of Charged and Aromatic Amino Acids.* S17 contains a large number of positively charged amino acids.

FIGURE 5: A stereoview of the superposition of 15 of the structures for ribosomal protein S17 calculated on the basis of the NMR data. The superposition was performed using only the core atoms by the program Insight II [Version 2.2.0 (1993) Biosym Technologies, San Diego, CA]. An average structure was calculated, and the rmsd relative to the backbone of the mean coordinates was 0.8 Å without the loop regions and 1.8 Å with the loop regions.

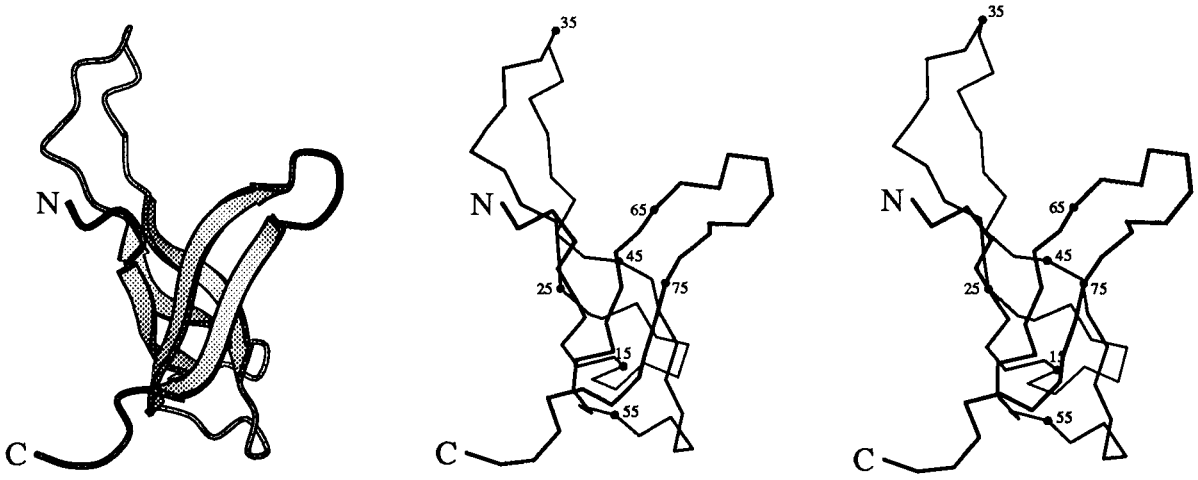


FIGURE 6: One of the final structures of ribosomal protein S17 calculated on the basis of the NMR data. Shown are a ribbon diagram and a stereoview of the α-carbon backbone with every 10 atoms numbered. Loop 1 and loop 3 which are proposed to interact with ribosomal RNA are on the left and right, respectively. Both figures were produced using the MOLSCRIPT program (Kraulis, 1991).

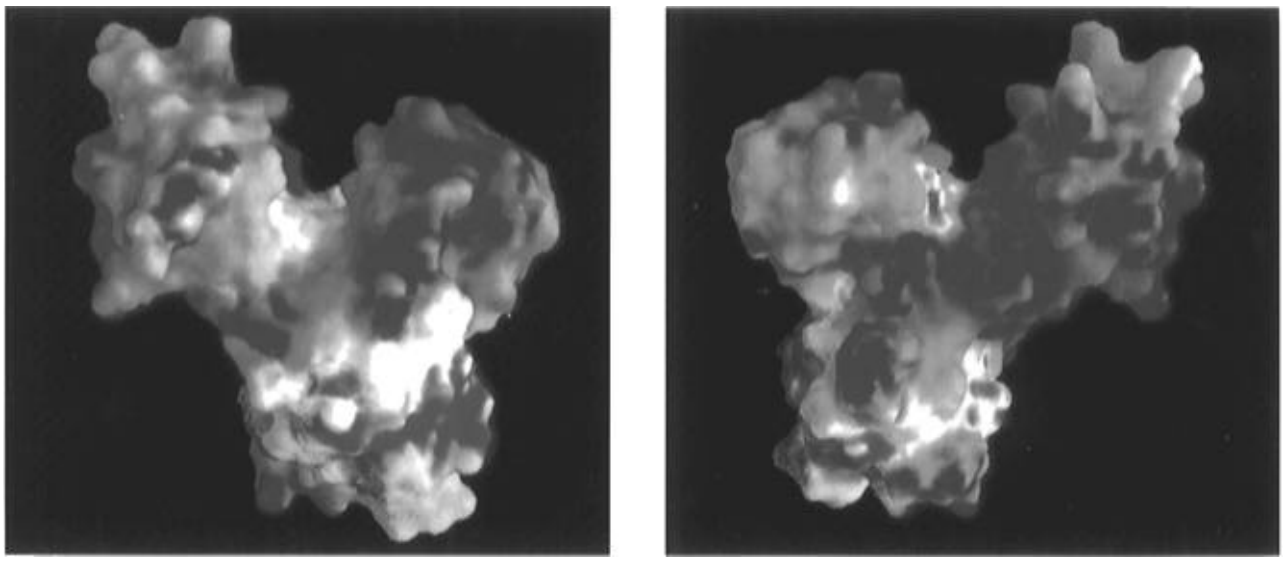

FIGURE 7: Two views of the electrostatic surface of ribosomal protein S17. The view on the left is in the same orientation as Figure 6. The calculations were performed using the program GRASP (Nicholls & Honig, 1993) on the average coordinates. Blue and red indicate positively charged and negatively charged surfaces, respectively. The extremes represent a range of electrostatic potentials from $<-12$ to $>+12$ $kT$, where $k$ is the Boltzmann constant and $T$ the temperature. The calculation was done at pH 8.0, and the dielectric constants were 80 for water and 2 for the protein. The entire upper part of the protein (toward loop 1 and loop 3) is more positively charged than the lower half, and two highly positively charged surfaces centered on loop 1 and loop 3 can be identified, pointing in opposite directions.

They are distributed fairly equally throughout the molecular surface, but those in the lower half (relative to the standard orientation shown in Figure 6) are intermixed with acidic residues to create a net neutral but highly polar surface. However, the upper half of the molecule is almost totally devoid of aspartates and glutamates, and this entire surface is predominantly basic. The most basic regions of S17 occur on loop 1 and loop 3, and two particularly positively charged patches are directed toward opposite sides of the molecule. The arginine and lysine residues responsible for these features are well conserved. Figure 7 shows the distinctive electrostatic surfaces of S17 which were calculated from the average coordinates using the program GRASP (Nicholls & Honig, 1993). The orientation of the positively charged surfaces involving loop 1 and loop 3 is clearly seen.

The aromatic amino acids are distributed in two distinct areas of the protein. The first is in the structured core of the molecule in positions that are conserved for either aromatic or hydrophobic residues. The second is in loop 1 which contains two of the most conserved aromatic residues, histidine 32 and tyrosine 35, together with tyrosine 41 and tyrosine 29. There are no aromatic amino acids in loop 3 *per se* although phenylalanine 74, which appears to be partially exposed, is located at the junction of loop 3 and strand $\beta$5.

*Putative RNA-Binding Surfaces.* Although the high-resolution structure of a ribosomal protein−RNA complex has yet to be determined, we have been able to identify the probable RNA-binding locations on the ribosomal proteins by features that are common to these types of proteins (Ramakrishnan & White, 1992; Golden et al., 1993a; Hoffman et al., 1994). They often have localized and flexible regions that are populated with conserved positively charged and aromatic residues. The structural basis of these features has recently been clarified by the crystal structure of the U1A−RNA complex (Oubridge et al., 1994). Conserved lysine and arginine residues define the path of the sugar−phosphate backbone with which they form electrostatic

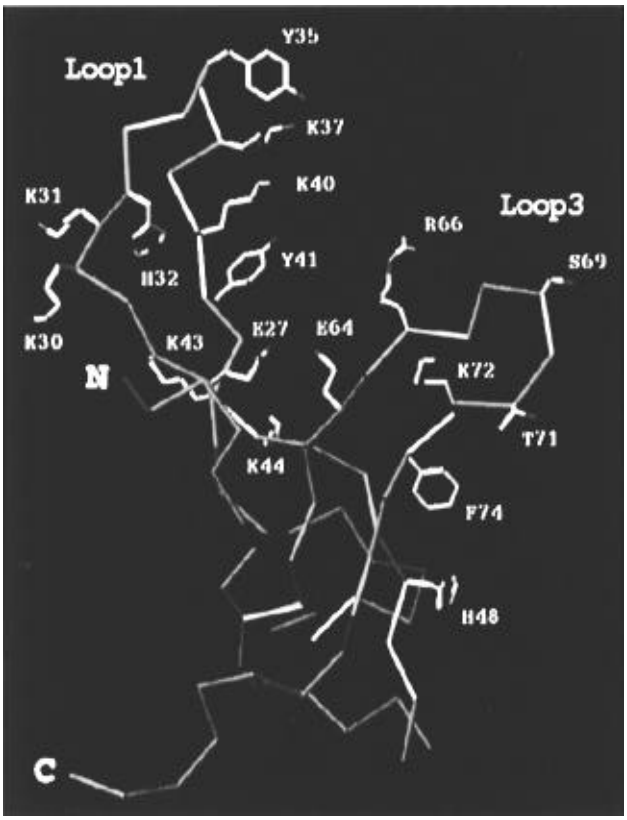


FIGURE 8: Detailed view of the proposed RNA-binding region of ribosomal protein S17. Loop 1 and loop 3 are indicated on the left and right, respectively. Histidine 32 is the site of a point mutation that renders the ribosome resistant to the antibiotic neamine. Serine 69 is the site of a second mutation that results in defective 30S assembly at restrictive temperatures. Lysine 31 has been cross-linked to ribosomal RNA. A pair of glutamic acids (27 and 64) at the base of loop 1 and loop 3 are highly conserved and appear to fix the relative orientation of the loops. The figure was produced using the O program (Jones & Kjeldgaard, 1993).

interactions, and the aromatic rings stack with the exposed bases of single-stranded regions of RNA. It has also been noted that the putative RNA-binding surfaces of S5 (Ramakrishnan & White, 1992), L6 (unpublished results), and L9 (Herbst et al., 1994) are the sites of point mutations that produce defined ribosome phenotypes.

On the basis of these features, we suggested from the early model of S17 that the protein might recognize and bind two regions of ribosomal RNA using elements in loop 1 and loop 3 (Golden et al., 1993b). This is fully supported by the present structure which shows these two regions in much greater detail despite the apparent flexibility of the loops. A detailed view of the region is shown in Figure 8. A feature of the two sites which is also reflected in the electrostatic surface of S17 (Figure 7) is that they point in opposite directions. Unless the RNA wraps around the molecule, this architecture clearly supports our previous suggestion that S17 binds two separate RNA moieties. An unusual region of the molecule that may be related to its RNA-binding function is a pair of acidic residues, glutamic acids 27 and 64, that are directly juxtaposed at the bases of loop 1 and loop 3 (Figure 8). These amino acids are highly conserved and are positionally well determined in the basic top half of the molecule. Their possible role may be to keep loop 1 and loop 3 apart and/or to form an electronegative barrier between the two bound RNA moieties.

Recent experiments by Noller and co-workers (personal communication) support this general RNA-binding model for S17. They have covalently attached $Fe^{2+}$−EDTA complexes to S17 at various positions and then incorporated the derivatized protein into a 16S RNP complex and complete 30S particles. The positions of the $Fe^{2+}$−EDTA complexes and their ability to produce hydroxy radical footprints on the 16S rRNA correlate well with our putative RNA-binding sites on S17. In addition, the RNP complex produces footprints at one general location of the 16S rRNA, whereas the 30S subunit produces footprints at two locations. The second footprint appears to represent a secondary S17−RNA interaction that depends on RNA folding. There are several reasons to suggest that loop 1 is the primary site. First, it is more extensive and electropositive than the loop 3 site (Figures 7 and 8). Second, a mutation at serine 69 in loop 3 results in defective 30S assembly at restrictive temperatures (Herzog et al., 1979) which is consistent with an incompletely folded 16S rRNA. Finally, recent protein−RNA cross-linking data from Wittmann-Liebold and co-workers (Urlaub et al., 1995) confirm that lysine 31 in loop 1 is adjacent to 16S rRNA in the ribosome, but a cross-link to loop 3 was not found, suggesting a less intimate interaction.

Noller's results are important for confirming and further understanding the general role of the ribosomal proteins which is to help organize the rRNA into its functional three-dimensional structure. The stepwise binding of RNA to S17 may be a general feature of many ribosomal proteins. Two potential RNA-binding sites have also been located on S5 (Ramakrishnan & White, 1992), L9 (Hoffman et al., 1994), and L14 (Davies et al., 1996), and a stepwise binding of S5 is supported by other footprinting data from Noller's group (Noller et al., 1995). These types of experiments also show the power of having high-resolution protein structures with which to probe the ribosome's architecture. Further work is underway in our laboratory to identify the precise RNA sequence that primarily binds to S17 using *in vitro* selection techniques and binding studies. Once identified, we intend to determine the structure of the S17−RNA complex.

*Structural Similarity to the OB/OG Protein Family*. Several of the ribosomal proteins whose structures have been determined contain a common motif comprising an antiparallel $\beta$-pleated sheet and an $\alpha$-helix (Leijonmarck et al., 1980, 1988; Wilson et al., 1986; Golden et al., 1993a; Lindahl et al., 1994; Hoffman et al., 1994). This motif is also present in the RNA recognition motif or RRM and is intimately involved in binding RNA (Nagai et al., 1990; Hoffman et al., 1991; Oubridge et al., 1994). S17 does not seem to be related to this structural family but appears instead to belong to a separate class of proteins that contain a motif called the oligonucleotide/oligosaccharide binding fold (OB fold) (Murzin, 1993). The common fold of these proteins comprises a five-stranded $\beta$-sheet that is twisted to form a $\beta$-barrel which is capped by an $\alpha$-helix located between the third and fourth strands.

S17 deviates from this basic structure in two ways. A loop (loop 2) replaces the $\alpha$-helix, and the $\beta$-strands do not close into a continuous $\beta$-barrel but instead leave a gap between strands 3 and 5. Many different types of proteins belong to this class, but two are of particular interest since they bind single-stranded nucleic acids; these are the gene 5 protein from bacteriophage fd (Skinner et al., 1994) and the major cold-shock protein of *Escherichia coli* (Schindelin et

al., 1994; Newkirk et al., 1994). Murzin has noted that the oligomer-binding site in these proteins is always in the same location, on one side of the $\beta$-barrel between the loops that connect strands 1 and 2 and strands 4 and 5. This location is consistent with the putative primary RNA-binding site of S17 that is centered on loop 1. This further supports our suggestion that the second putative RNA-binding site at loop 3 is a secondary site.

## ACKNOWLEDGMENT

We thank Dr. Howard Robinson and Dr. Andrew H.-J. Wang for their generous help with the SPEDREF package and Drs. Leonard Spicer and Ronald Venters for assistance with the NMR data collection and processing. We are particularly grateful to Dr. Harry Noller for providing us with unpublished data. Finally, we thank Mike Finnin, Christopher Davies, Dirk Bussiere, and Stephanie Porter for helpful discussions.

## SUPPORTING INFORMATION AVAILABLE

Two tables listing the $^{1}H$ chemical shifts and the $^{13}C$ and $^{15}N$ chemical shifts (2 pages). Ordering information is given on any current masthead page.

## REFERENCES

Andersen, N. H., Chen, C. P., Marchner, T. M., Krystek, S. R., & Bassolono, D. A. (1992) *Biochemistry 31*, 1280−1295.
Bax, A., & Pochapsky, S. (1992) *J. Magn. Reson. 99*, 638−643.
Bonvin, A. M., Rullmann, J. A., Lamerichs, R. M., Boelens, R., & Kaptien, R. (1993) *Proteins 15*, 385−400.
Bonvin, A. M., Vis, H., Breg, J. N., Burgering, M. J., Boelens, R., & Kaptien, R. (1994) *J. Mol. Biol. 236*, 328−341.
Brimacombe, R., Atmadja, J., Stiege, W., & Schüler, D. (1988) *J. Mol. Biol. 199*, 115−136.
Brimacombe, R., Greuer, B., Mitchell, P., Osswald, M., Rinke-Appel, J., Schüler, D., & Stade, K. (1990) in *The Ribosome: Structure, Function and Evolution* (Hill, W. E., Dahlberg, A., Garrett, R. A., Moore, P. B., Schlessinger, D., & Warner, J. R., Eds.) pp 93−106, American Society for Microbiology, Washington, DC.
Brown, S. C., Weber, P. L., & Mueller, L. (1988) *J. Magn. Reson. 77*, 166−169.
Brünger, A. T. (1992) *X-PLOR, Version 3.1*, The Howard Hughes Medical Institute and Yale University, New Haven, CT.
Capel, M. S., Engelman, D. M., Freeborn, B. R., Kjeldgaard, M., Langer, J. A., Ramakrishnan, V., Schindler, D. G., Schneider, D. K., Schoenborn, B. P., Sillers, I.-Y., Yabuki, S., & Moore, P. B. (1987) *Science 238*, 1403−1406.
Clore, G. M., & Gronenborn, A. M. (1991) *Science 252*, 1390−1399.
Clore, G. M., Gronenborn, A. M., Brünger, A. T., & Karplus, M. (1985) *J. Mol. Biol. 185*, 435−455.
Clore, G. M., Nilges, M., Sukumaran, D. K., Brünger, A. T., Karplus, M., & Gronenborn, A. M. (1986) *EMBO J. 12*, 2729−2735.
Constantine, K. L., Madrid, M., Banyai, L., Trexler, M., Patthy, L., & Llinas, M. (1992) *J. Mol. Biol. 223*, 281−298.
Costa, H. S., Santos, H., & Turner, D. L. (1994) *Eur. J. Biochem. 223*, 783−789.
Davies. C., White, S. W., & Ramakrishnan, V. (1996) *Structure 4*, 55−66.
Fesik, S. W., & Zuiderweg, E. R. P. (1988) *J. Magn. Reson. 78*, 588−593.
Frank, J., Zhu, J., Penczek, P., Li, Y., Srivastava, S., Verschoor, A., Radermacher, M., Grassucci, R., Lata, R. K., & Agrawal, R. K. (1995) *Nature 376*, 441−444.
Frechet, D., Guitton, J. D., Herman, F., Faucher, D., Helynck, G., Monegier du Sorbier, B., Ridoux, J. P., James-Surcouf, E., & Vuilhorgne, M. (1994) *Biochemistry 33*, 42−50.
Golden, B. L., Ramakrishnan, V., & White, S. W. (1993a) *EMBO J. 12*, 4901−4908.
Golden, B. L., Hoffman, D. W., Ramakrishnan, V., & White, S. W. (1993b) *Biochemistry 32*, 12812−12820.
Greuer, B., Osswald, M., Brimacombe, R., & Stöffler, G. (1987) *Nucleic Acids Res. 15*, 3241−3255.
Grzesiek, S., & Bax, A. (1992a) *J. Am. Chem. Soc. 114*, 6291−6293.
Grzesiek, S., & Bax, A. (1992b) *J. Magn. Reson. 99*, 201−207.
Grzesiek, S., & Bax, A. (1993) *J. Biomol. NMR 3*, 185−204.
Held, W. A., Ballou, B., Mizushima, S., & Nomura, M. (1974) *J. Biol. Chem. 249*, 3103−3111.
Herbst, K. L., Nichols, L. M., Gesteland, R. F., & Weiss, R. B. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 12525−12529.
Herzog, A., Yaguchi, M., Cabezón, T., Corchuelo, M. C., Petre, J., & Bollen, A. (1979) *Mol. Gen. Genet. 171*, 15−22.
Hoffman, D. W., Query, C. C., Golden, B. L., White, S. W., & Keene, J. D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2495−2499.
Hoffman, D. W., Davies, C., Gerchman, S. E., Kycia, J. H., Porter, S., White, S. W., & Ramakrishnan, V. (1994) *EMBO J. 13*, 205−221.
Hoffman, R. C., Horvath, S. J., & Klevit, R. E. (1993) *Protein Sci. 2*, 951−65.
Jones, T. A., & Kjeldgaard, M. (1993) *O Version 5.9*, Uppsala University, Sweden, and Aarhus University, Denmark.
Kay, L. E., Xu, G.-Y., Singer, A. U., Muhandiram, D. R., & Kay, J. D. F. (1993) *J. Magn. Reson. 101*, 333−337.
Knegtel, R. M., Katahira, M., Schilthuis, J. G., Bonvin, A. M., Boelens, R., Eib, D., van der Saag, P. T., & Kaptien, R. (1993) *J. Biomol. NMR 1*, 1−17.
Kraulis, P. (1991) *J. Appl. Crystallogr. 24*, 946−950.
Lambert, J. M., Boileau, G., Cover, J. A., & Traut, R. R. (1983) *Biochemistry 22*, 3913−3920.
Leijonmarck, M., Ericksson, S., & Liljas, A. (1980) *Nature 286*, 824−826.
Leijonmarck, M., Appelt, K., Badger, J., Liljas, A., Wilson, K. S., & White, S. W. (1988) *Proteins: Struct., Funct., Genet. 3*, 243−251.
Lindahl, M., Svensson, A., Liljas, A. Sedelnikova, S. E., Eliseikina, A., Fomenkova, N. P., Nevskaya, N., Nikonov, S. V., Garber, M. B., Muranova, T. A., Rykonova, A. I., & Amons, R. (1994) *EMBO J. 13*, 1249−1254.
Mackie, G. A., & Zimmermann, R. A. (1978) *J. Mol. Biol. 121*, 17−39.
Macura, S., & Ernst, R. R. (1980) *Mol. Phys. 41*, 95−117.
Majumdar, A., & Zuiderweg, E. R. P. (1993) *J. Magn. Reson. 102*, 242−244.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
Muhandiram, D. R., & Kay, L. E. (1994) *J. Magn. Reson. 103*, 203−216.
Murzin, A. G. (1993) *EMBO J. 12*, 861−867.
Nagai, K., Oubridge, C., Jessen, T. H., Li, J., & Evans, P. R. (1990) *Nature 348*, 515−520.
Newkirk, K., Feng, W., Jiang, W., Tejero, R., Emerson, S. D., Inouye, M., & Montelione, G. T. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 5114−5118.
Nicholls, A., & Honig, B. (1993) *GRASP*, Columbia University, New York.
Nilges, M., Clore, G. M., & Gronenborn, A. M. (1988) *FEBS Lett. 229*, 317−324.
Noller, H. F. (1991) *Annu. Rev. Biochem. 60*, 191−227.
Noller, H. F., Hoffarth, V., & Zimniak, L. (1992) *Science 256*, 1416−1419.
Noller, H. F., Green, R., Heilek, G., Hoffarth, V., Huttenhofer, A., Joseph, S., Lee, I., Lieberman, K., Mankin, A., Merriman, C., Powers, T., Samaha, R. R., & Weiser, B. (1995) *Biochem. Cell Biol.* (in press).
Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H., & Nagai, K. (1994) *Nature 372*, 432−438.
Post, C. B., Meadows, R. P., & Gorenstein D. G. (1990) *J. Am. Chem. Soc. 112*, 6796−6803.
Powers, T., & Noller, H. F. (1995) *RNA 1*, 194−209.
Purohit, P., & Stern, S. (1994) *Nature 370*, 659−662.

Ramakrishnan, V., & Gerchman, S. E. (1991) *J. Biol. Chem. 266*, 880−885.
Ramakrishnan, V., & White, S. W. (1992) *Nature 358*, 768−771.
Richardson, J. S. (1977) *Nature 268*, 495−500.
Robinson, H., & Wang, A. H.-J. (1992) *Biochemistry 31*, 3524−3533.
Robinson, H., Yang, D., & Wang, A. H.-J. (1994) *Gene 149*, 179−188.
Santoro, J., Gonzalez, C., Bruix, M., Neira, J. L., Nieto, J. L., Herranz, J., & Rico, M. (1993) *J. Mol. Biol. 229*, 722−734.
Schindelin, H., Jiang, W., Inouye, M., & Heinemann, U. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 5119−5123.
Schultze, P., Macaya, R. F., & Feigon, J. (1994) *J. Mol. Biol. 235*, 1532−1547.
Skinner, M. M., Zhang, H., Leschnitzer, D. H., Guan, Y., Bellamy, H., Sweet, R. M., Gray, C. W., Konnings, R. N., Wang, A. H.-J., & Terwilliger, T. C. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 2071−2075.
Stark, H., Mueller, F., Orlova, E. V., Schatz, M., Dube, P., Erdemir, T., Zemlin, F., Brimacombe, R., & van Heel, M. (1995) *Structure 3*, 815−821.
Stern, S., Changchien, L.-M., Craven, G. R., & Noller, H. F. (1988a) *J. Mol. Biol. 200*, 291−299.
Stern, S., Weiser, B., & Noller, H. F. (1988b) *J. Mol. Biol. 204*, 447−481.
Stern, S., Powers, T., Changchien, L.-M., & Noller, H. F. (1989) *Science 244*, 783−790.
Stöffler-Meilicke, M., & Stöffler, G. (1990) in *The Ribosome: Structure, Function and Evolution* (Hill, W. E., Dahlberg, A. D., Garrett, R. A., Moore, P. B., Schleissinger, D., & Warner, J. R., Eds.) pp 123−133, American Society for Microbiology, Washington, DC.
Urlaub, H., Kruft, V., Bischof, O., Mueller, E. C., & Wittmann-Liebold, B. (1995) *EMBO J. 14*, 4578−4588.
Venters, R. A., Calderone, T. L., Spicer, L. D., & Fierke, C. A. (1991) *Biochemistry 30*, 4491−4494.
Weitzmann, C. J., Cunningham, P. R., Nurse, K., & Ofengand, J. (1993) *FASEB J. 7*, 177−180.
Wilson, K. S., Appelt, K., Badger, J., Tanaka, I., & White, S. W. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7251−7255.
Wittekind, M., & Mueller, L. (1993) *J. Magn. Reson. B 101*, 201.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Yaguchi, M., Wittmann, H. G., Cabezón, T., DeWilde, M., Villaroel, R., Herzog, A. & Bollen, A. (1976) *J. Mol. Biol. 104*, 617−620.
Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R., & Kay, L. E. (1994) *J. Am. Chem Soc. 116*, 11655−11666.
Yonath, A. (1992) *Annu. Rev. Biophys. Biomol. Struct. 21*, 77−93.
Zimmermann, R. A., Muto, A., & Mackie, G. A. (1974) *J. Mol. Biol. 86*, 433−450.
Zuiderweg, E. R. P. (1990) *J. Magn. Reson. 86*, 346−357.

BI951062I