Solution Structure of HI0257, a Bacterial Ribosome Binding Protein[†,‡]

Lisa Parsons, Edward Eisenstein, and John Orban*

Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, 9600 Gudelsky Drive, Rockville Maryland 20850

Received May 25, 2001; Revised Manuscript Received July 25, 2001

ABSTRACT: A novel bacterial ribosome binding protein, protein Y (also known as YfiA), was recently shown to reside at the 30S/50S subunit interface and to stabilize the ribosomal 70S complex against dissociation at low magnesium ion concentrations. We report here the three-dimensional NMR structure in solution of a homologue from *Haemophilus influenzae*, HI0257, that has 64% sequence identity to protein Y. The 107 residue protein has a β-α-β-β-β-α folding topology with two parallel α-helices packed against the same side of a four-stranded β-sheet. The closest structural relatives are proteins with the double-stranded RNA-binding domain (dsRBD) motif although there is little (<10%) sequence homology. The most immediate differences between the dsRBD and HI0257 structures are that (1) HI0257 has a larger β-sheet motif with an extra β-strand at the N-terminus, (2) the helices are parallel in HI0257 but at an angle of about 30° to each other in the dsRBD, and (3) HI0257 lacks the extended loop commonly seen between the first and second β-strands of the dsRBD. Further, an analysis of the surface electrostatic potential in HI0257 and the dsRBD family reveals significant differences in the location of contiguous positively (and negatively) charged regions. The structural data, in combination with sequence analysis of HI0257 and its homologues, suggest that the most likely mode of RNA recognition for HI0257 may be distinct from that of the dsRBD family of proteins.

HI0257 is a 107 residue protein from *Haemophilus influenza* that had no known function when the genome was sequenced (*1*). Two recent papers report that the *Escherichia coli* homologue, YfiA, also known as protein Y, has a role in stabilizing the 70S ribosome. From a BLAST[1] analysis (*2*), YfiA is 64% identical to HI0257. YfiA binds to the 30S subunit and becomes solvent inaccessible when the 50S subunit is added, suggesting that the protein is located at the subunit interface (*3*, *4*). In the presence of YfiA, dissociation of the subunits by low magnesium ion concentrations is inhibited. Maki et al. (*5*) showed that YfiA is associated with 70S and 100S ribosomes in the resting state, but not after cell growth is reinitiated [100S ribosomes are dimeric 70S ribosomes in the resting state (*6*)]. Another *Escherichia coli* protein called YhbH is 35% identical to HI0257 and was also found to associate with 100S ribosomes in a manner similar to YfiA (*5*).

A distant homologue to HI0257 exists in spinach chloroplast. This protein, known as S22 (formerly known as Psrp-1 or CS−S5) has also been shown to associate with the 30S ribosomal subunits of both its native 70S ribosome and that of *E. coli* (*7*, *8*). The mature S22 protein contains 236 residues and is only 21% identical to HI0257. However, the region of similarity is important for ribosome binding (*8*) and is recognized as having the same Ribosomal S30 domain [as defined by Pfam (*9*, *10*)] as HI0257.

We describe here the three-dimensional solution structure of HI0257 as determined by NMR spectroscopy. To our knowledge, this is the first structure to be reported from this family of proteins. The fold contains the α-β-β-β-α topology of the double-stranded RNA-binding domain (dsRBD) but has an additional β-strand at the N-terminus and a C-terminal unstructured tail. HI0257 has little sequence homology to dsRBD proteins (<10%), and the conserved residues which are most likely involved in binding to RNA are located in a different region of the protein. The most striking feature of HI0257 is the arrangement of charged residues along its helices which is dissimilar to the dsRBD protein family. This suggests that HI0257 and its homologues may have a different mode of interaction with RNA as compared to other dsRBD proteins.

MATERIALS AND METHODS

Sample Preparation. The gene for HI0257 was amplified from *Haemophilus influenzae* genomic DNA and was inserted into the pET-15b vector (Novagen) using *Nde*I/*Bam*HI restriction sites. The sequence was verified by DNA sequencing. BL21(DE3) *E. coli* cells were grown in LB media to an A_{600} of 1.0 and induced with 1 mM IPTG. After 3 h, the cells were harvested, resuspended in lysis buffer

[†] Supported by NIH Grant GM57890. L.P. is a member of the Molecular and Cell Biology Graduate Program at University of Maryland College Park and is the recipient of a Life Technologies Fellowship.

[‡] The coordinates and restraint file have been deposited in the Protein Data Bank (accession code 1IMU). Chemical shift assignments have been deposited in BioMagResBank (entry 5055).

* To whom correspondence should be addressed. Phone: (301) 738-6221. Fax: (301) 738-6255. E-mail: john@magpie.carb.nist.gov.

[1] Abbreviations: ARIA, ambiguous restraints for iterative assignment; BLAST, basic local alignment search tool; CATH, class, architecture, topology, homologous superfamily; CD, circular dichroism; CNS, crystallography and NMR system; dsRBD, double-stranded RNA binding domain; dsRNA, double-stranded RNA; ExPASy, expert protein analysis system; HSQC, heteronuclear single quantum coherence; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser effect spectroscopy; Pfam, protein families database of alignments and HMMs; RNP, ribonucleoprotein; SCOP, structure classification of proteins; TOCSY, total correlation spectroscopy.

10.1021/bi011077i CCC: $20.00
Published on Web 08/23/2001

(10 mM imidazole, 0.3 M NaCl, and 50 mM sodium phosphate, pH 8.0), and sonicated. Lysed cells were centrifuged at 32000*g* for 25 min and the supernatant was passed through a 0.45 μm membrane. The clear filtrate was loaded onto a Ni-NTA column (Qiagen) and washed with 6 column volumes of lysis buffer followed by 6 column volumes of lysis buffer with 20 mM imidazole. The desired protein was eluted with 5 column volumes of lysis buffer containing 250 mM imidazole and fractions were tested for purity using SDS−PAGE under denaturing conditions. The eluent was dialyzed against lysis buffer without imidazole at 4 °C overnight. The resulting protein solution was concentrated to 0.1−0.3 mM and treated with 60 μg of human α-thrombin (Haematologic Technologies) at 25 °C for 4 h. The extent of the reaction was monitored by SDS−PAGE. Upon completion of the reaction, the mixture was loaded onto a Ni-NTA column and washed with lysis buffer plus 5 mM imidazole. This bound the cleaved N-terminal $(His)_6$-tag to the column. The eluent containing HI0257 flowed directly into a connected column of *p*-aminobenzamidine on cross-linked 6% beaded agarose (Sigma) to remove thrombin. The final eluent was concentrated to 1.5 mM using Centriprep and Centricon 3 kDa cutoff membrane filters (Amicon) and then dialyzed against NMR buffer (see below). Typical yields of pure HI0257 were 20−30 mg/L of fermentation culture. The thrombin cleavage left a permanent Gly-Ser-His tail attached to the N-terminus of the protein. ^{15}N-Labeled and ^{13}C/^{15}N-labeled proteins were purified in the same manner, except that the cells were grown in minimal media with ^{15}N-ammonium chloride and ^{13}C-labeled glucose as the ^{15}N- and ^{13}C-sources, respectively. Unless otherwise stated, the final buffer conditions were 50 mM sodium phosphate, 100 mM sodium chloride, 3 mM DTT, and pH 7.0. ^{13}C/^{15}N-Labeled samples were at a concentration of 1.3 mM. The samples used for the 2D NOESY, 2D TOCSY, and 3D ^{13}C NOESY-HSQC experiments were lyophilized and dissolved in 99.9% D_2O. All other samples were in 10% D_2O and 90% H_2O.

Protein Characterization. UV−vis spectra were recorded on a Cary4 UV−vis spectrophotometer between 250 and 400 nm. Circular dichroism measurements were performed using a Jasco spectropolarimeter, model J-720 and a water-jacketed quartz cell with a path length of 0.01 cm. The protein concentration was 3 mg/mL in 50 mM sodium phosphate and 100 mM sodium chloride, pH 7.0. Temperature control was maintained by a Neslab RTE−110 circulating water bath connected to a MTP-6 temperature programmer. The absorbance of the protein at 222 nm was recorded as the temperature was increased 1 °C/min from 35 to 90 °C. Reversibility was confirmed by comparing the spectra between 200 and 250 nm at 35 °C.

Mass spectra were collected on a MALDI-TOF Voyager-DE Biospectrometry Workstation (PerSeptive Biosystems) equipped with a nitrogen laser (337 nm, 3 ns pulse). Unlabeled protein at a concentration of 1 mM in the above buffer was diluted 1:10 with water and mixed with equal amounts of sinapinic acid prepared using the manufacturers instructions. With the laser intensity set at 1880, accelerating voltage at 25 kV, grid voltage at 93%, and a delay time of 700 ns, 256 scans were averaged in the positive mode.

NMR Spectroscopy. All spectra were collected at 25 °C on Bruker DRX-600 and DRX-500 NMR spectrometers equipped with triple resonance 3-axis gradient probes. Data were processed on a Dell Precision 410 Linux workstation using nmrPipe (*11*) and analyzed with Sparky (*12*). Unless otherwise stated a 1 s relaxation delay and 65.8 ms acquisition time were employed with 16 scans/increment. Most spectra were recorded in States-TPPI mode using a gradient water flip-back scheme for solvent suppression (*13*, *14*).

Two-dimensional NOESY (*15*, *16*) and TOCSY (*17*) spectra were recorded on the ^{15}N-labeled sample in D_2O. The spectral widths for both dimensions were set to 11.0 ppm in the two experiments with $4096t_2 \times 1024t_1$ complex points in the NOESY experiment and $4096t_2 \times 800t_1$ complex points in the TOCSY. The NOESY spectrum had a 150 ms mixing time and used 48 scans/increment. The TOCSY spectrum had a 60 ms mixing time and used 80 scans/increment. Both experiments employed a 1.5 s relaxation delay.

The three-dimensional HNCACB (*18*), CBCACONH (*19*), HNCO (*20*), and HBHA(CBCACO)NH (*19*) spectra were all collected with ^{1}H and ^{15}N spectral widths of 13.0 and 26.0 ppm, respectively, and $1024t_3 \times 64t_2 \times 128t_1$ complex points. The ^{13}C spectral widths were 65.0, 65.0, and 20.0 ppm for the first three experiments, respectively. The second ^{1}H spectral width (f_1) was 8.3 ppm for the HBHA(CBCACO)NH experiment. Only 8 scans/increment were used to record the HNCO spectrum. The HNCACB experiment had a 1.4 s relaxation delay. The 3D ^{15}N TOCSY-HSQC and ^{15}N NOESY-HSQC spectra (*21*, *22*) were recorded with 32 and 16 scans with mixing times of 70 and 150 ms, respectively. Spectral widths of 13.0, 26.0, and 13.0 ppm (f_3, f_2, f_1) were used for both, but the complex points utilized differed; $1024t_3 \times 40t_2 \times 128t_1$ were used for the TOCSY while $1024t_3 \times 64t_2 \times 196t_1$ were used for the NOESY. A 3D HCCH-TOCSY experiment (*23*) was collected with an acquisition time of 63.9 ms, spectral widths of 16.0, 30.0, and 12.0 ppm (f_3, f_2, f_1), and $1024t_3 \times 64t_2 \times 200t_1$ complex points. A 3D HNHA experiment (*24*) was collected with spectral widths of 16.0, 60.0, and 16.0 ppm (f_3, f_2, f_1), and $1024t_3 \times 64t_2 \times 128t_1$ complex points. Three-dimensional ^{13}C NOESY-HSQC (*25*, *26*) experiments were collected both in H_2O and D_2O with spectral widths of 13.0, 40.0, and 13.0 ppm (f_3, f_2, f_1). Acquisition matrixes were $2048t_3 \times 60t_2 \times 194t_1$ and $1024t_3 \times 64t_2 \times 164t_1$ complex points, respectively. The mixing time for the H_2O experiment was 100 ms, while that for the D_2O experiment was 118 ms.

For the hydrogen exchange experiment at 25 °C, the protein was lyophilized and dissolved in 99.9% D_2O. The ^{15}N HSQC experiment was started after 30 min had elapsed using 16 scans/increment, $1024t_2 \times 64t_1$ complex points, and spectral widths of 13.0 (^{1}H) and 33.0 (^{15}N) ppm.

Structure Calculations. Backbone dihedral angle restraints (ϕ) were identified from secondary ^{1}H and ^{13}C chemical shifts (*27*, *28*), characteristic NOE restraints, and from $^3J_{HNH\alpha}$ coupling constants derived from the 3D HNHA experiment. In well-defined regions, ϕ was set to $-60 \pm 20°$ for α-helix residues and $-120 \pm 30°$ for β-strand residues. Inter-proton distance restraints obtained from the 2D and 3D NOESY experiments were classified based on peak intensities of strong (1.8−2.7 Å), medium-strong (1.8−3.1 Å), medium (1.8−3.5 Å), medium-weak (2.3−4.2 Å), weak (2.8−5.0 Å),

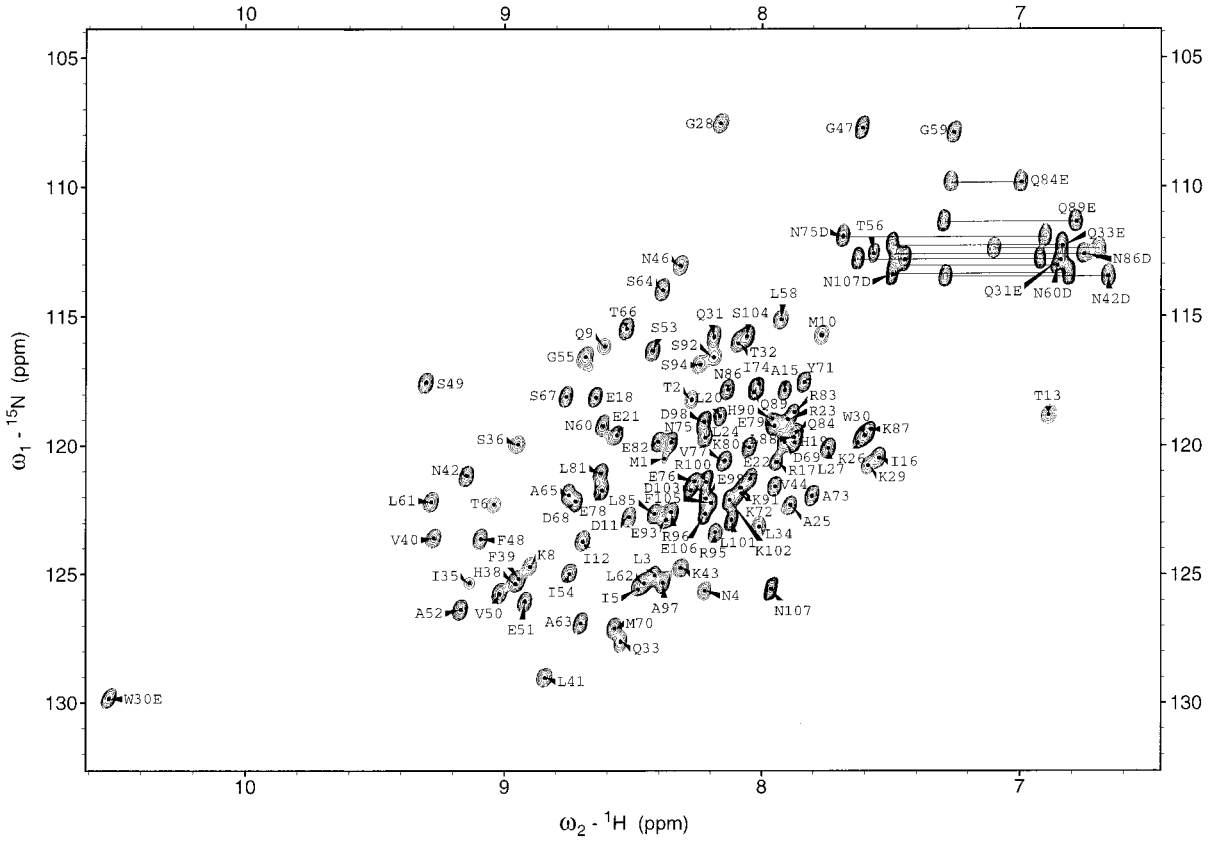

FIGURE 1: ^{15}N HSQC spectrum of HI0257 with assignments. Horizontal lines connect cross-peaks due to side chain Asn/Gln amide protons. The spectrum was recorded at 25 °C, pH 7.0.

and very weak (2.8−6.0 Å). Very weak peaks were used only when reciprocal NOE correlations were observed. Two restraints, 2.3−3.2 Å for r_{N-O} and 1.5−3.0 Å for r_{HN-O}, were used for each hydrogen bond. These were only included in the final stages of refinement when well defined candidate acceptor/donor heavy atoms were consistently within 3.2 Å of each other and had favorable geometries (HN···O angle > 120°) in all the calculated structures.

Initial structures were calculated with the CNS 1.0 program (*29*) starting from an extended polypeptide chain. Standard simulated annealing protocols with torsion angle dynamics were utilized. Prochiral groups were given floating stereospecific assignments. Nonbonded contacts were represented by a quartic van der Waals repulsion term. The final values for the force constants employed were as follows: 1,000 kcal mol^{-1} Å $^{-2}$ for bond lengths, 500 kcal mol^{-1} rad^{-2} for angles and improper torsions, 40 kcal mol^{-1} Å $^{-2}$ for experimental distance restraints (interproton distances and hydrogen bonds), 200 kcal mol^{-1} rad^{-2} for dihedral angle restraints, and 4.0 kcal mol^{-1} Å $^{-4}$ for the van der Waals repulsion term. Refinement from intermediate to final structures was carried out in an automated fashion using ARIA 1.0 (*30*). Inputs for ARIA were 10 of the lowest energy structures, 1450 fully assigned NOE restraints, about 1500 partially assigned/ambiguous NOE restraints, and the chemical shift assignment list. For atoms where chemical shift assignments were not made, standard values with large uncertainties (*31*) were used (cf. ref *32*). The NOE restraint list from ARIA after eight iterations was compared to the previously assigned NOE list resulting in about 200 new assignments. A total of 1685 NOE distance restraints were used in the final structure calculations; 420 intraresidue, 449 sequential, 324 medium

Table 1: Experimental Restraints and Structural Statistics[a]

	⟨20⟩	best
rmsds from exptl distance restraints (Å)[b]		
all (1775)	0.043 ± 0.001	0.042
intraresidue (420)	0.033 ± 0.001	0.031
sequential ($\|i-j\|=1$) (449)	0.051 ± 0.002	0.047
medium-range ($1<\|i-j\|\le 5$) (324)	0.042 ± 0.002	0.039
long-range ($\|i-j\|>5$) (492)	0.043 ± 0.003	0.035
hydrogen bonds (90)	0.050 ± 0.003	0.047
rmsds from exptl dihedral angle restraints[b] (deg) (83)	0.284 ± 0.056	0.175
rmsds from idealized covalent geometry[b]		
bonds (Å)	0.003 ± 0.000	0.003
angles (deg)	0.493 ± 0.008	0.477
impropers (deg)	0.429 ± 0.012	0.401
avg rmsds to mean structure (Å)[b]		
N,C_α,CO of residues 1−91	0.36 ± 0.07	0.24
all heavy atoms of residues 1−91	0.91 ± 0.09	0.68
measures of structure quality (PROCHECK)[c]		
% of residues 1−91 in most favorable region of Ramachandran plot	77.7 ± 2.1	82.7
no. of bad contacts per 100 residues	8.4 ± 1.5	5.0
overall dihedral *G* factor	0.002 ± 0.016	0.04
CNS potential energies (kcal mol^{-1})		
E_{tot}	335.75 ± 6.38	312.60
E_{bond}	14.17 ± 0.55	12.81
E_{ang}	99.57 ± 3.05	93.96
E_{imp}	20.49 ± 1.15	17.85
E_{noe}	131.05 ± 4.19	122.58
E_{cdih}	0.24 ± 0.10	0.09

[a] Statistics for the 20 best structures ⟨20⟩ and the best value out of the 20 for each statistic. The values under the heading ⟨20⟩ are the mean±standard deviation. All values are from structures calculated without residues 92−107. [b] The final force constants for these functions are noted in the text. [c] Structures were chosen based on the following parameters from PROCHECK (*43*): fewer than 10 bad contacts/100 residues, a dihedral *G* factor greater than −0.5, and no residues in disallowed regions of the Ramachandran plot.

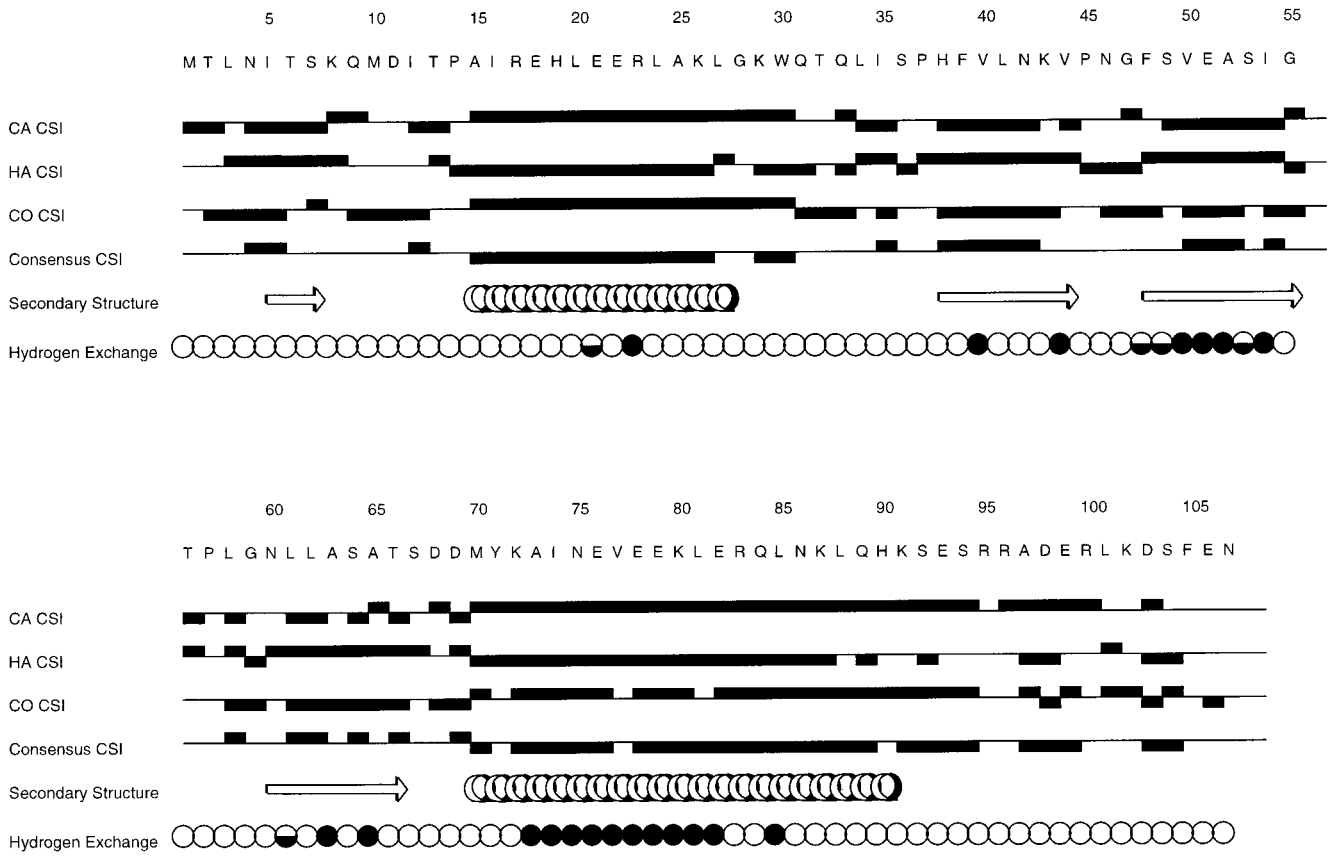

FIGURE 2: Deviations from random coil for C_α, H_α, and CO chemical shifts. The consensus chemical shift index (CSI) is shown along with the secondary structure as determined from the final structure of HI0257. For the hydrogen exchange experiment, filled circles represent slowly exchanging amide protons, half-filled circles represent amide protons with intermediate exchange rates, and open circles represent fast exchanging amides. The figure was made using VINCE (http://www.rowland.org/rnmrtk/vince.html).

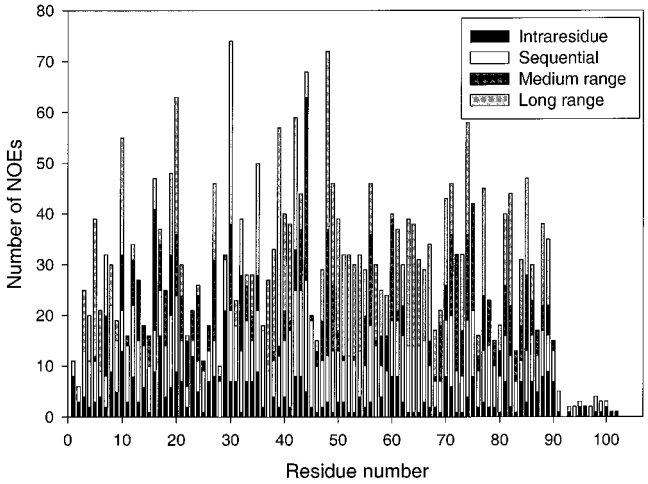

FIGURE 3: NOE distribution per residue for HI0257. Intraresidue NOEs are only represented once. All other NOEs are represented twice. For example, if a sequential NOE was found between residue A and residue B, the sequential NOE count is incremented for both residues. The different categories refer to the number of amino acids in the protein sequence that separate residues (i, j) with an NOE contact. NOE classes are defined as sequential, $|i - j| = 1$; medium range $1 < |i - j| \leq 5$; long range, $|i - j| > 5$.

range, and 492 long range. Ninety hydrogen bond restraints and 83 dihedral restraints were also included. Structures were chosen based on a low total energy, no NOE distance violations greater than 0.32 Å, no dihedral restraint violations greater than 5°, and standard indicators of structure quality as shown in Table 1.

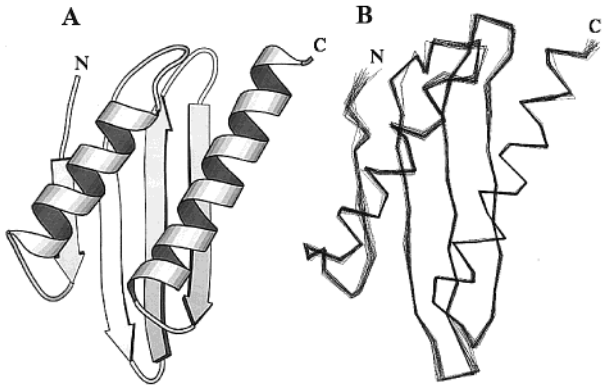

FIGURE 4: (A) Molscript (*44*) figure of the average structure of HI0257. (B) Backbone atom superposition of the twenty best structures chosen for the reasons discussed in the text. The unstructured tail was not included in the structure calculations and is not displayed in these representations. The N-terminal Gly-Ser-His sequence is also not shown.

RESULTS

Protein Characterization. Circular dichroism measurements in the far-UV region were used to determine the melting temperature (T_m) of the protein in the buffer used for NMR experiments (see Materials and Methods). The absorbance at 222 nm was monitored as a function of temperature yielding a sigmoidal curve with a median point around 77 °C. Denaturation of HI0257 was reversible from comparison of the CD spectra before and after heating although a small percentage (<5%) did not refold. Mass

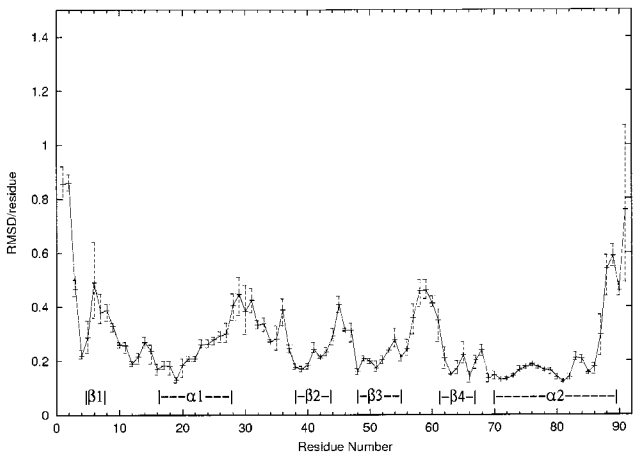

FIGURE 5: RMSD per residue of the backbone atoms (N, C_α, and CO) for residues 1−91. The secondary structure is represented beneath the graph. Error bars show ± 1 standard deviation.

spectrometry analysis of an unlabeled sample gave a molecular weight equal to 12 473.3 Da, which is in good agreement with the calculated value of 12 472.8 Da (from ExPASy, http://ca.expasy.org/).

Resonance Assignments. The N, H_N, H_α, C_α, H_β, and C_β chemical shifts for all 107 residues were assigned except for the backbone N and H_N resonances of Ser7. This was done using a combination of HNCACB, CBCA(CO)NH, HNCO, and HBHA(CBCACO)NH experiments. Due to overlap, especially in the region where γ protons are found, not all of the side-chain proton resonances could be assigned. This was especially true for residues 92−107, which give the most intense peaks in the backbone amide region of the HSQC spectrum (Figure 1). Eighty-eight percent of the H_γ, H_δ, and H_ϵ protons attached to carbons in nonaromatic amino acids were assigned for residues 1−91 using HCCH-TOCSY and ^{13}C NOESY-HSQC experiments. Half of the unassigned resonances were due to H_γ protons. For residues 1−91, 100% of the aromatic proton chemical shifts were assigned using the 2D TOCSY and NOESY spectra recorded in D_2O. The side chain amides were assigned for eight of the eleven Asn/Gln residues from the ^{15}N NOESY-HSQC spectrum. Finally, the hydroxyl proton of Thr56 had slowed exchange and was detected in the NOESY spectrum. Analysis of the final structure indicates that this proton is within hydrogen bonding distance of the Gly59 carbonyl oxygen atom. NOE data also indicated that the $O_{\gamma 1}$ atom of Thr56 is likely hydrogen bonded to the backbone amide of Leu58 and this was included as a restraint in the late stages of refinement.

Structure Determination and Description. The secondary structure elements predicted on the basis of chemical shifts (*28*) were in good agreement with those calculated from the final structure (Figure 2). An average of about 18 restraints per residue were used in the structure calculations and the class distribution of NOEs per residue is shown in Figure 3. Residues with a low number of NOEs are generally more solvent exposed, as in loop regions, or could not be fully assigned due to overlap. The C-terminal residues 92−107 had random coil-type chemical shifts, intense HSQC peaks, and no medium or long-range NOEs, indicating that this region is unstructured. Resonances due to the extra N-terminal Gly-Ser-His sequence remaining after thrombin cleavage of the $(His)_6$ tag were not detected, presumably due to exchange broadening, and were not included in the structure calculation.

HI0257 has a β-α-β-β-β-α topology with two parallel helices on one side of a four-stranded β-sheet (Figure 4, panels A and B). The first two β-strands are parallel while strands two, three, and four are antiparallel. Figure 4B shows an overlay of the 20 lowest energy structures and Table 1 lists the structure statistics. The RMSD per residue for backbone atoms is shown in Figure 5. The secondary structure elements are generally well defined with low RMSDs around 0.2 Å although the first β-strand is less precisely determined. Residues 70−86 in the highly conserved second helix have the lowest RMSDs on average.

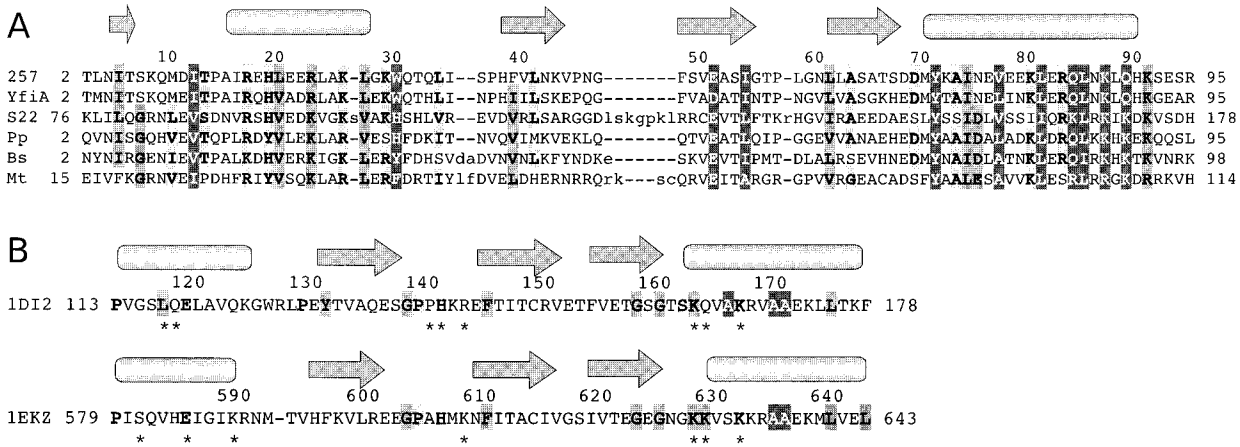

FIGURE 6: (A) Pfam alignment for HI0257 (257). Only 6 of 28 sequences are shown. The dark shade shows residues conserved in >96% of the 28 sequences, the medium shade in >89%, and the light shade in >78%. Because the similarities between the sequences were difficult to see when comparing like residues alone, conservation is based on like charges, hydrophobicity, and size of the amino acids. Sequences shown are from *Escherichia coli* (YfiA), *Spinach oleracea* (S22), *Pseudomonas putida* (Pp), *Bacillus subtilus* (Bs), and *Mycobacterium tuberculosis* (Mt). The secondary structure of HI0257 is shown above the alignment with β-strands as arrows and α-helices as cylinders. The sequence is numbered so that the 0's are above the residue being counted. (B) Sequences of the dsRNA binding protein A (1DI2) and staufen (1EKZ). Conserved residues from their respective Pfam alignments are highlighted such that the dark shade is conserved in >80% of the 88 sequences, the medium shade in >60%, and the light shade in >40%. Conservation reflects only identical residues. Secondary structures for each protein are above their sequences. Residues involved in dsRNA binding are marked with a star below each sequence.

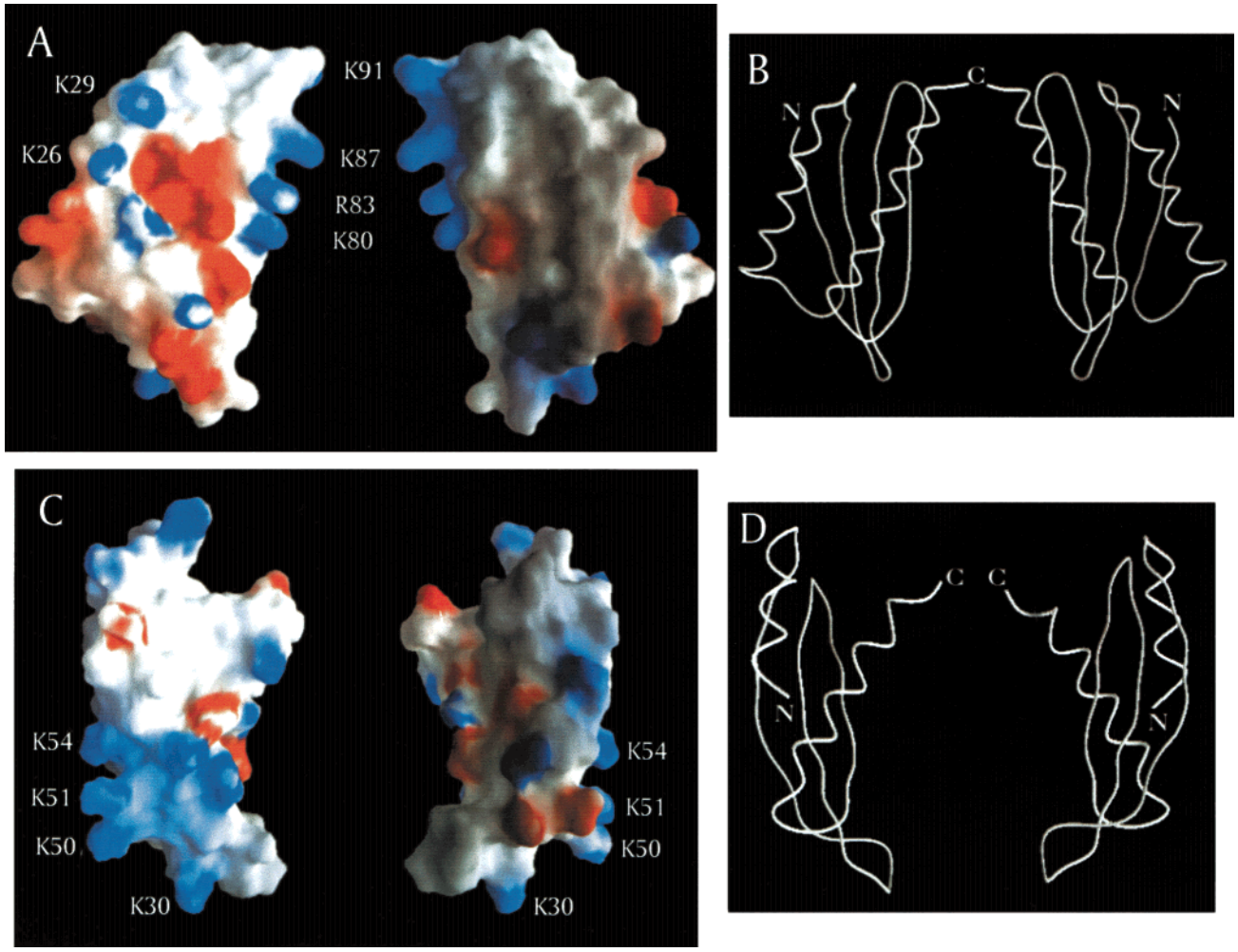

FIGURE 7: (A) GRASP (*45*) representation of the surface potential of HI0257. The picture on the left is the view from the helical side of the molecule, while that on the right is from the β-sheet side. Negative charge is colored red and positive charge is shown in blue. (B) Worm representation of HI0257 reflecting the orientation of the protein in panel A. (C) GRASP representation of the surface potential of the third dsRBD from the *Drosophila* staufen protein. The picture on the left is the helix view while that on the right is the sheet view. (D) Worm figure showing the orientation of the staufen protein as seen in panel C. Both proteins are arranged so that the C-terminus of the second helix is seen at the top of the representation.

The loop/turn regions at residues 29−37, 44−46, and 56−60 are more disordered with RMSDs of 0.3−0.5 Å whereas the short β4-α2 loop at residues 67−69 is well determined. The N-terminus of the β1-strand and the C-terminus of the α2-helix both have high RMSDs consistent with poor structure definition.

DISCUSSION

Sequence Analysis. The HI0257 sequence is conserved among many bacterial species including those in the genuses of Klebsiella, Azotobacter, Pseudomonas, Bacillus, Vibrio, and Salmonella. It also has similarity to proteins in at least two plant species, *Arabidopsis thelia* and *Spinach oleracea*. Homologues from *Pasteruella multocida* and *E. coli* are the most similar to HI0257 with 84 and 64% identical residues, respectively. The HI0257 homologues are annotated as putative sigma 54 modulation factors, 30S ribosomal binding proteins, or light repressed proteins. A search of the *Haemophilus influenzae* genome did not reveal any proteins with homology to the *E. coli* sigma 54 protein, making it unlikely that HI0257 is involved in modulation of sigma 54. The sequence of HI0257 has little (19% positive over 31 residues) or no homology to the structurally similar dsRBD proteins and less than 10% homology to other RNA binding proteins. Alignment of HI0257 with its homologues shows that the most conserved residues are in helix 2 (Figure 6A).

Structural Analysis and Functional Implications. An extensive search of all known structures in the Protein Data Bank using CATH (*33*), SCOP (*34*, *35*), and Dali (*36*) did not result in any structures with the same secondary and tertiary arrangement as HI0257. Of those proteins with a tertiary structure similar to HI0257, the secondary structure was frequently arranged in a β-α-β-α-β-β motif as seen in three domains of the RNA 3′-terminal phosphate cyclase (*37*). The fourth domain of the cyclase has the same secondary structure as HI0257 (β-α-β-β-β-α) but the first two strands are exchanged in the tertiary structure so that the second strand is on the outside.

The protein most structurally similar to HI0257 from a Dali search is the double stranded RNA binding protein A from *Xenopus laevis* (*38*) which has a Z-score of 5.0 and an RMSD (C_α atoms) of 2.6 Å. This protein has the characteristic α-β-β-β-α motif of the dsRBD family and similar structures have been determined for other dsRBD proteins including human protein kinase PKR (*39*) and the maternal effect protein staufen from *Drosophila melanogaster* (*40*,

41). The most immediate differences between the dsRBD and HI0257 structures are that (1) HI0257 has a larger β-sheet motif with an extra β-strand at the N-terminus, (2) the helices are parallel in HI0257 whereas, in the dsRBD structures determined to date, the helices are at an angle of about 30° to each other (compare Figure 7, panels B and D), and (3) HI0257 lacks the extended loop commonly seen between the first and second β-strands of the dsRBD.

Further comparison of surface charge distribution reveals some considerable differences between HI0257 and the dsRBD family. In HI0257, the α-helical elements are considerably more charged than the β-sheet motif (Figure 7A). Helix 1 (residues 16−28) contains contiguous negatively charged surface residues at E18, E21, and E22, and proximal positively charged residues in its C-terminal half at residues R23, K26, and K29. Helix 2 (residues 70−91) has negatively charged residues at E76, E78, E79, and E82, and positively charged residues at K72, K80, R83, K87, and K91. The basic residues K80, R83, K87, and K91 form a continuous surface of positive charge in the C-terminal half of helix 2 while the acidic residues E78, E79, and E82 form a negatively charged region between the rows of positive charge in the two helices. Another negatively charged surface also exists near the N-terminus of helix 2, formed by residues D68 and D69 in the β4-α2 loop. The β-sheet surface of the molecule has only three charged groups at residues K8, K43, and E51. In contrast to HI0257, the dsRBD proteins have an array of basic residues near the N-terminus of helix 2 and in the loop between β-strands 1 and 2 which provide an extensive contiguous surface of positive charge (Figure 7C). These residues have been shown to be essential for RNA recognition in *Drosophila* Staufen (Figure 6B) and mutations at these positions abolish binding to dsRNA (*41*).

Many of the proteins that are structurally similar to HI0257 bind to dsRNA, perhaps indicating that the region where HI0257 interacts with the 30S ribosome contains dsRNA. If this is correct, then the differences highlighted in Figure 7 would suggest that the mode of RNA recognition for HI0257 is distinct from that of the dsRBD family of proteins. HI0257 is most likely to bind ribosomal RNA using, at least in part, the C-terminal half of helix 2 based on the arrangement of charged residues and the high degree of sequence identity in the second helix (Figure 6A). The positively charged residues at positions 23, 26, and 29 in helix 1 are also conserved, suggesting that they too may play some part in RNA recognition. The negatively charged residues in helix 1 and 2 do not appear to be as strongly conserved, however (Figure 6A). Another possibility is that the C-terminal tail (residues 92−107), which contains a number of charged residues (5 acidic, 4 basic) and is not conserved in HI0257 homologues, is involved in RNA binding. This is known to occur in the RNP family, for example, where the C-terminal tail becomes structured upon RNA binding and provides specificity for the binding interaction (*42*).

The structure determined here lays the groundwork for further studies to determine which residues are involved in ribosomal RNA recognition. Preliminary HSQC screening with an 18 nucleotide double-stranded RNA revealed a limited number of shift perturbations localized in the C-terminal half of helix 2, although the changes in ^{1}H chemical shifts were very small (0.03−0.5 ppm). These results are consistent with weak nonspecific binding and have yet to be optimized for specific 30S ribosomal RNA sequences. We are currently working to locate the exact sites of interaction on both the protein and the ribosome.

ACKNOWLEDGMENT

We thank Phil Bryan and John Toedt for assistance with CD measurements, and Jim Parsons for his advice on cloning and expression.

REFERENCES

1. Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J.-F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O., and Venter, J. C. (1995) *Science 269*, 496−512.
2. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res. 25*, 3389−402.
3. Agafonov, D. E., Kolb, V. A., and Spirin, A. S. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 12892−12897.
4. Agafonov, D. E., Kolb, V. A., Nazimov, I. V., and Spirin, A. S. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 12345−12349.
5. Maki, Y., Yoshida, H., and Wada, A. (2000) *Genes Cells 5*, 965−974.
6. Wada, A. (1998) *Genes Cells 3*, 203−208.
7. Johnson, C. H., Kruft, V., and Subramanian, A. R. (1990) *J. Biol. Chem. 265*, 12790−12795.
8. Bubunenko, M. G., and Subramanian, A. R. (1994) *J. Biol. Chem. 269*, 18223−18231.
9. Sonnhammer, E. L., Eddy, S. R., Birney, E., Bateman, A., and Durbin, R. (1998) *Nucleic Acids Res. 26*, 320−322.
10. Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Finn, R. D., and Sonnhammer, E. L. (1999) *Nucleic Acids Res. 27*, 260−262.
11. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 6*, 277−293.
12. Goddard, T. D., and Kneller, D. G. University of California San Francisco.
13. Grzesiek, S., and Bax, A. (1993) *J. Am. Chem. Soc. 115*, 12593−12594.
14. Mori, S., Abeygunawardana, C., Johnson, M. O., and van Zijl, P. C. (1995) *J. Magn. Reson. B 108*, 94−98.
15. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546−4553.
16. Kumar, A., Ernst, R. R., and Wuthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1−6.
17. Braunschweiler, L., and Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521−528.
18. Wittekind, M., and Mueller, L. (1993) *J. Magn. Reson., Ser. B 101*, 201−205.
19. Grzesiek, S., and Bax, A. (1992) *J. Am. Chem. Soc. 114*, 6291−6293.
20. Grzesiek, S., and Bax, A. (1992) *J. Magn. Reson. 96*, 432−440.
21. Fesik, S. W., and Zuiderweg, E. R. P. (1988) *J. Magn. Reson. 78*, 588−593.
22. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) *Biochemistry 29*, 6150−6156.
23. Sattler, M., Schwendinger, M. G., Schleucher, J., and Griesinger, C. (1995) *J. Biomol. NMR 6*, 11−22.
24. Vuister, G. W., and Bax, A. (1993) *J. Am. Chem. Soc. 115*, 7772−7777.

25. Ikura, M., Kay, L. E., Tschudin, R., and Bax, A. (1990) *J. Magn. Reson. 86*, 204−209.
26. Zuiderweg, E. R. P., McIntosh, L. P., Dahlquist, F. W., and Fesik, S. W. (1990) *J. Magn. Reson. 86*, 210−216.
27. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) *Biochemistry 31*, 1647−1651.
28. Wishart, D. S., and Sykes, B. D. (1994) *J. Biomol. NMR 4*, 171−180.
29. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse, K. R., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr., Sect. D 54*, 905−921.
30. Nilges, M., Macias, M. J., O'Donoghue, S. I., and Oschkinat, H. (1997) *J. Mol. Biol. 269*, 408−422.
31. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., and Skelton, N. J. (1996) *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, San Diego.
32. Hare, B. J., and Wagner, G. (1999) *J. Biomol. NMR 15*, 103−113.
33. Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B., and Thornton, J. M. (1997) *Structure 5*, 1093−108.
34. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) *J. Mol. Biol. 247*, 536−540.
35. Lo Conte, L., Ailey, B., Hubbard, T. J., Brenner, S. E., Murzin, A. G., and Chothia, C. (2000) *Nucleic Acids Res. 28*, 257−259.
36. Holm, L., and Sander, C. (1995) *Trends Biochem. Sci. 20*, 478−480.
37. Palm, G. J., Billy, E., Filipowicz, W., and Wlodawer, A. (2000) *Structure Folding Des. 8*, 13−23.
38. Ryter, J. M., and Schultz, S. C. (1998) *EMBO J. 17*, 7505−7513.
39. Nanduri, S., Carpick, B. W., Yang, Y., Williams, B. R., and Qin, J. (1998) *EMBO J. 17*, 5458−5465.
40. Bycroft, M., Grunert, S., Murzin, A. G., Proctor, M., and St. Johnston, D. (1995) *EMBO J. 14*, 3563−3571.
41. Ramos, A., Grunert, S., Adams, J., Micklem, D. R., Proctor, M. R., Freund, S., Bycroft, M., St Johnston, D., and Varani, G. (2000) *EMBO J. 19*, 997−1009.
42. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. (1994) *Nature 372*, 432−438.
43. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR 8*, 477−486.
44. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
45. Nicholls, A. (1992) Columbia University, New York.

BI011077I