Structural Preordering in the N-Terminal Region of Ribosomal Protein S4 Revealed by Heteronuclear NMR Spectroscopy†

Eric W. Sayers,‡ Resi B. Gerstner,§,|| David E. Draper,§ and Dennis A. Torchia*,‡

Department of Chemistry, Johns Hopkins University, Baltimore, MD 21218, Molecular Structural Biology Unit, National Institute of Dental and Craniofacial Research, National Institutes of Health, 30 Convent Drive, MSC 4307, Bethesda, Maryland 20892-4307

Received June 12, 2000; Revised Manuscript Received August 22, 2000

ABSTRACT: Protein S4, a component of the 30S subunit of the prokaryotic ribosome, is one of the first proteins to interact with rRNA in the process of ribosome assembly and is known to be involved in the regulation of this process. While the structure of the C-terminal 158 residues of *Bacillus stearothermophilus* S4 has been solved by both X-ray crystallography and NMR, that of the N-terminal 41 residues is unknown. Evidence suggests that the N-terminus is necessary both for the assembly of functional ribosomes and for full binding to 16S RNA, and so we present NMR data collected on the full-length protein (200 aa). Our data indicate that the addition of the N-terminal residues does not significantly change the structure of the C-terminal 158 residues. The data further indicate that the N-terminus is highly flexible in solution, without discernible secondary structure. Nevertheless, structure calculations based on nuclear Overhauser effect spectroscopic data combined with ^{15}N relaxation data revealed that two short segments in the N-terminus, $S_{12}RRL_{15}$ and $P_{30}YPP_{33}$, adopt transiently ordered states in solution. The major conformation of S_{12}-RRL_{15} appears to orient the arginine side chains outward toward the solvent in a parallel fashion, while that of $P_{30}YPP_{33}$ forms a nascent turn of a polyproline II helix. These segments contain residues that are highly conserved across many prokaryotic species, and thus they are reasonable candidates respectively for sites of interaction with RNA and other ribosomal proteins within the intact ribosome.

Protein S4, a component of the 30S subunit of the prokaryotic ribosome, has been shown to be intimately involved in both ribosome regulation and function. S4 binds to 16S RNA (*1*−*3*) and serves as one of two nucleation sites during the assembly of the small subunit (*4*). In addition, S4 also regulates its own synthesis by binding to the α operon mRNA, which encodes S11, S13, and L17 in addition to S4 (*5*). Mutations in S4 have been shown to alter the translational accuracy of the ribosome (*6*), the conformation of the 16S RNA (*7*), and the sensitivity of the ribosome to the antibiotic streptomycin (*8*). In an effort to begin to understand the structural basis of the many roles that S4 plays in the intact ribosome, the three-dimensional structure of a fragment consisting of the 158 C-terminal residues of S4 from the thermophile *Bacillus stearothermophilus* was recently solved by both NMR spectroscopy (*9, 10*) and X-ray crystallography (*11*). This protein fragment, labeled S4Δ41, lacked the N-terminal 41 residues of wild-type S4 but retained those regions previously identified as being required for binding to both 16S RNA and the α operon mRNA (*12, 13*). The results of these studies showed that S4Δ41 adopts a somewhat elongated structure consisting of two subdomains, and on the basis of the distribution of surface positive charges, a putative RNA-binding surface was identified.

The role of the remainder of the S4 sequence, the N-terminal 41 residues (47 in *Escherichia coli*), has also been the subject of considerable interest. Early work showed that *E. coli* S4 fragments lacking the N-terminal 44−47 residues via proteolytic cleavage competed with intact S4 for binding to 16S RNA but, when included in ribosomal assembly reactions, produced incomplete 30S particles missing five proteins (*14*). Additional work on an S4 fragment containing residues 32−206[1] showed that this construct could bind to 16S RNA and allowed complete ribosomes to assemble, but the resulting particles could not bind tRNA (*15*). Thus, these authors concluded that residues 1−32 (1−27 for *B. stearothermophilus*) were important for ribosomal function, particularly tRNA binding, while residues 33−47 (28−43 for *B. stearothermophilus*) were required for proper ribosome assembly (*16*). Further work on other S4 fragments narrowed the RNA binding region of S4 to residues 47−177 (44−174 for *B. stearothermophilus*) (*12, 17*). Later studies of the binding of the α operon mRNA to S4 fragments prepared by modifying the *E. coli* S4 gene and overexpressing the resulting proteins in *E. coli* suggested that the same domain of S4 is involved in binding to both the mRNA and 16S RNA (*13*). Interestingly, more recent experiments with *B. stearothermophilus* S4 and S4Δ41 suggest that the N-terminal residues make possible a high-affinity 16S RNA binding mode that was not detected with the *E. coli* protein,

† 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, along with a Pharmacology Research Associate Award from the National Institute of General Medical Sciences to E.W.S.

* To whom correspondence should be addressed: phone (301) 496-5750; fax (301) 402-5321; e-mail dtorchia@dir.nidcr.nih.gov.

‡ National Institute of Dental and Craniofacial Research.

§ Department of Chemistry, Johns Hopkins University.

|| Present address: Department of Molecular Oncology, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080.

[1] The numbering for *E. coli* S4 is as described in ref *13*.

10.1021/bi0013391 CCC: $19.00
Published on Web 10/11/2000

perhaps because of the marginal stability of that protein (R. B. Gerstner and D. E. Draper, unpublished experiments).

The 16S rRNA and mRNA binding behavior observed for *E. coli* and *B. stearothermophilus* S4 proteins raises the question of whether these proteins are functionally equivalent. The sequence identity between S4Δ41 of the two species is 54% (*9*), while that for the N-terminus is 38% (see below). Previous work in which 16S RNA from either *E. coli* or *B. stearothermophilus* was combined with the ribosomal proteins from one of these species showed that active ribosomes could be assembled from each of the four resulting combinations (*18*). A later study further showed that a ribosome containing all *E. coli* components except for S4, which was from *B. stearothermophilus*, retained 92% of the activity of a fully *E. coli* ribosome (*19*). Therefore, the data suggest that S4 molecules from these two species are well conserved and functionally interchangeable. We are therefore left with the task of untangling the functional roles of the S4 N-terminus in the intact ribosome. As a step toward this goal, we present here a detailed analysis of full-length S4 from *B. stearothermophilus* by heteronuclear NMR spectroscopy. We report that the residues corresponding to S4Δ41 maintain their previously determined fold, while the N-terminus is highly flexible in solution. Nevertheless, two short, conserved segments in the N-terminus were found to have transient structural order, and possible functional roles of these segments will be explored.

EXPERIMENTAL PROCEDURES

Sample Preparation. The gene encoding wild-type S4 from *B. stearothermophilus* was cloned into the pET-13a expression vector (*20*), and the protein was overexpressed in *E. coli* strain BL21(DE3). Wild-type S4 was purified by a modified version of the previously reported procedure (*13*). Cultures were grown from single colonies in M9 minimal medium containing $^{15}NH_4Cl$ and D-[$^{13}C_6$]glucose as desired for isotopic labeling. Cultures were induced with 0.5 mM IPTG[2] at an OD_{600} of 0.6−0.8 and were grown for an additional 3 h at 37 °C. The cells were then harvested and resuspended in 20−30 mL of cold buffer A (20 mM MES, pH 5.5, containing 500 mM KCl, 2 M urea, 1 mM EDTA, and 0.1 mM PMSF) supplemented with 1 mM TLCK. The cells were lysed in a French press, and the cellular debris was removed by centrifugation at 14000*g*. The supernatant was then passed successively through 0.45 and 0.22 μm filters and then loaded at 2 mL/min onto a HiLoad 16/10 SP-Sepharose cation-exchange column (Pharmacia Biotech, Piscataway, NJ) equilibrated in buffer A. S4 was recovered with a 0.5−1.0 M KCl gradient over 100 min at a flow rate of 2 mL/min, and the protein eluted as a single, sharp peak at 0.75 M KCl, yielding approximately 20 mg of pure protein per liter of cell culture. Fractions containing pure protein were pooled and dialyzed against buffer B (20 mM sodium acetate, pH 5.4, containing 250 mM KCl and 1 mM EDTA). The samples were concentrated and exchanged into buffer B-d_3 [buffer B containing sodium acetate-d_3 (98%)] by use of Centricon-10 concentrators (Amicon, Beverly, MA). Final NMR samples contained either 0.3 or 0.7 mM S4 in 300 μL of buffer B-d_3 (90% H_2O/10% D_2O or 100% D_2O) in a Shigemi microcell.

NMR Spectroscopy. NMR spectra were acquired at 310 K on either Bruker DMX spectrometers operating at proton resonance frequencies of 499.53, 500.14, or 749.74 MHz or a Bruker DRX spectrometer operating at a proton resonance frequency of 800.13 MHz. All spectra were acquired with a triple-resonance probe equipped with pulsed field gradients. NMR data were processed with NMRPipe software (*21*) and analyzed with PIPP (*22*). The experimental parameters used for the various spectra are collected in the Supporting Information.

Chemical Shift Assignments. The numbering system for full-length S4 used in the present study is consistent with that used in structural studies of S4Δ41 (*9*), with residues 43−200 being identical between the two proteins. In this system, residue 1 of S4 corresponds to the initiator methionine and residue 200 to the C-terminal arginine. We note that resonances for M1 were observed and assigned, justifying its inclusion as residue 1. In S4Δ41, residue 42 is the initiator methionine, while residue 42 in S4 is an arginine. The backbone atoms, along with C^β and H^β, were initially assigned from the results of HSQC (*23*−*25*), HNCACB (*26*), CBCA(CO)NH (*27*), HBHA(CO)NH (*28*), and HNCO (*29*) experiments. Instances of simultaneous degeneracy at the amide ^{15}N and ^1H shifts were resolved by a CBCACO(N)H experiment, a variant of the CBCA(CO)NH experiment in which the carbonyl carbon chemical shift rather than that of the amide ^{15}N evolves in t_2. Confirmation of proline assignments, along with the assignment of proline amide ^{15}N resonances, was facilitated by an H(CA)CON experiment, a similar modification of the HCA(CO)N experiment (*30*) in which the carbonyl chemical shift rather than that of the C^α evolves in t_1. Experimental details of the CBCACO(N)H and H(CA)CON experiments will be published elsewhere. These assignments were checked and extended along the side chains by HCACO (*31*), HCCH-TOCSY (*32*), and C(CO)NH and HC(CO)NH (*33*) experiments. Amide side-chain assignments were achieved with CBCA(CO)NH and NOESY experiments, while other side-chain assignments, such as those in aromatic rings and the stereospecific assignment of proline ring protons, were achieved from NOESY connectivities.

^{15}N Relaxation Experiments. We measured ^{15}N T_1, T_2, and $\{^1H\}-^{15}N$ NOE values at 500 MHz by following published experiments (*25*, *34*). For T_1 measurements, delay values were 8, 96, 256, 512, 800, 960, and 1200 ms, while for T_2 measurements, delay values were 9.6, 19.2, 38.4, 67.2, 115.2, 153.6, and 192.0 ms. For each experiment we recorded 32 scans for each FID and collected a total of 128 complex t_1 points. We derived T_1 and T_2 values for 132 residues with well-resolved signals by fitting measured peak intensities to a single-exponential function (*35*). $\{^1H\}-^{15}N$ NOE values were calculated as intensity ratios from data sets acquired with and without saturating the protons. The results from two independent measurements were averaged. For each measurement, we recorded 64 scans for each FID and collected a total of 256 complex t_1 points.

Acquisition of Structural Restraints. Three 3D NOESY spectra (*36*−*38*) were acquired: (1) an ^{15}N-separated experi-

[2] Abbreviations: IPTG, isopropyl β-thiogalactoside; MES, 2-(*N*-morpholino)ethanesulfonic acid; EDTA, ethylenediaminetetraacetic acid; PMSF, phenylmethanesulfonyl fluoride; TLCK, 1-chloro-3-tosylamido-7-amino-2-heptanone hydrochloride; NOESY, nuclear Overhauser enhancement spectroscopy; FID, free induction decay; RMSD, root-mean-square deviation.

ment at 750 MHz with a mixing time of 75 ms; (2) a ^{13}C-separated experiment in D_2O at 500 MHz with a mixing time of 75 ms; and (3) a ^{13}C-separated experiment in D_2O at 800 MHz with a mixing time of 200 ms. Distance restraints were derived from cross-peak intensities by use of the standard inverse sixth power relationship. The mobility of the N-terminus prevented us from deriving reference distances from secondary structure elements; instead, reference distances were derived from a combination of the following observed NOESY connectivities: intraresidue $H^{\beta 1}-H^{\beta 2}$ in proline side chains, intraresidue $H^{\alpha}-H^{\beta}$ in alanines, and the observed range of intraresidue $H^{\alpha}-H^{N}$ (*39*). Cross-peaks were classified as strong, medium, or weak. The lower limit for each restraint was set to 1.8 Å, while the upper limits for strong, medium, and weak restraints were 2.9, 3.5, and 5.0 Å, respectively. For torsional restraints on the ϕ dihedral angle, $^3J_{HNH\alpha}$ values were measured in the HNHA experiment (*40*, *41*). For values of $^3J_{HNH\alpha} > 8.0$ Hz, ϕ was restrained to $-120 \pm 40°$.

Structure Calculations. Model structures were calculated with X-PLOR 3.8 (*42*), beginning with a molecular template consisting of residues 1−45. Calculations were initiated from random coordinates by use of the *random* protocol. Following this, the resulting structures were regularized and subsequently refined with the *dgsa* and *refine* protocols. Those structures that had no violations greater than 0.5 Å or 5° were accepted and analyzed.

RESULTS

Purification of Full-Length S4. Initial attempts to purify full-length S4 followed the previously published protocol (*13*). While this procedure yielded protein that was >95% pure, a close examination of $^1H-^{15}N$ HSQC spectra taken over a span of a few days revealed that several peaks either shifted or disappeared, while other sharp, weak signals became visible. These observations prompted us to analyze our samples more carefully. Mass spectrometry revealed that >95% of the protein in these ^{15}N-labeled samples was a single species with molecular weight 18691, compared to 23305 expected for full-length S4 and 18676 for S4Δ41 (assuming 100% ^{15}N content). N-Terminal sequencing indicated that multiple peptide species were present, consistent with multiple cleavages of X−Lys bonds in the N-terminus, specifically at K10, K28, and K43. These results convinced us that the purification procedure did not reduce contaminating proteases to a level that would permit acquisition of multidimensional NMR spectra without substantial proteolysis. This led us to develop the revised procedure described under Experimental Procedures. With this revised procedure, significantly sharper elution peaks were observed for S4, and the total S4 yield per liter of cell culture increased by as much as a factor of 10. However, very slow proteolysis was still evident in $^1H-^{15}N$ HSQC spectra of the resulting protein. We subsequently found that this residual proteolysis could be eliminated by adding 1 mM EDTA to the NMR sample, rendering such samples stable for at least 6 months.

Resonance Assignment of S4. Our initial assignment strategy for S4 was two-fold: first, to assign the backbone resonances of residues 43−200 and compare them to those of S4Δ41; and second, to perform a complete assignment of residues 1−42. With the exception of the ^{15}N resonances of the seven proline residues C-terminal to K43, which were not assigned in S4Δ41, the backbone assignment for residues 43−200 is complete. The differences between the backbone assignments for these residues in S4 determined during this work and those previously determined for S4Δ41 (*9*) are displayed in Figure 1. Except for K43 and L44, which were immediately C-terminal to the N-terminal methionine in S4Δ41, the differences between the backbone shifts are mostly small and within the experimental errors. The few resonances that have somewhat larger differences are in residues either at the end of a secondary structure element (R107 H^{α}, L165 ^{13}CO) or in a loop region (I161 ^{13}CO, N190 H^N, E191 H^N). Moreover, the chemical shifts of the amide protons of N190 and E191 are known to be highly pH-sensitive. Therefore, we found no significant differences in backbone chemical shifts for residues 45−200, demonstrating that the secondary structure elements and the overall fold of these residues are essentially the same in the two proteins.

With the exception of the amide proton and ^{15}N resonances of the two N-terminal residues, M1 and A2, the backbone assignment of residues 1−42 is complete. In addition, we were also able to assign the side chains of 38 of these residues. These assignments are tabulated in the Supporting Information. The side chain of M1 could not be assigned due to resonance overlap and weak signals. The remaining unassigned resonances are the N^{ϵ} positions in the six arginines, which could not be definitively linked to specific arginine side chains because of a lack of sufficient NOE connectivities and resonance overlap. After completion of these assignments, it was apparent that several weaker peaks in many of the spectra remained to be assigned. We soon realized that many of these peaks could be sequentially connected in the usual manner and that they corresponded to short segments of amino acids surrounding proline residues in the N-terminus. The identity of these peaks as minor conformations caused by proline *cis*−*trans* isomerism was confirmed by the characteristic C^{β} and C^{γ} shifts of the proline side chains (*43*). These minor conformations are summarized in Figure 2, and the available assignments are collected in the Supporting Information. Notably, the major conformation of all prolines in the N-terminus was found to be *trans*.

Initial Characterization of the Dynamics of S4. Simply on the basis of $^1H-^{15}N$ HSQC spectra, it seemed clear that residues 1−44 exhibited significantly different dynamic behavior compared to residues 45−200. A portion of such a spectrum is displayed in Figure 3 with assignments indicated. As shown, peaks corresponding to residues 1−44 are generally both narrower and markedly more intense than those of the C-terminal domain equivalent to S4Δ41. Moreover, the majority of these peaks cluster near frequency values commonly found in random coil regions of proteins. We also analyzed the chemical shifts of the H^{α}, C^{α}, CO, and H^{β} spins using the CSI program (*44*), which indicated that residues 1−45 form a random coil. All of these data suggested that the N-terminus of S4 is significantly more dynamic than the C-terminal domain, and we therefore measured ^{15}N T_1 and T_2 relaxation times along with the $\{^1H\}-^{15}N$ NOE at 500 MHz to test this hypothesis.

The resulting data, shown in Figure 4, confirm that residues 1−44 indeed are significantly more dynamic than residues 45−200. The average NOE values for these two sets of residues are -0.59 ± 0.27 and 0.79 ± 0.13, respectively, indicating that residues 1−44 are highly

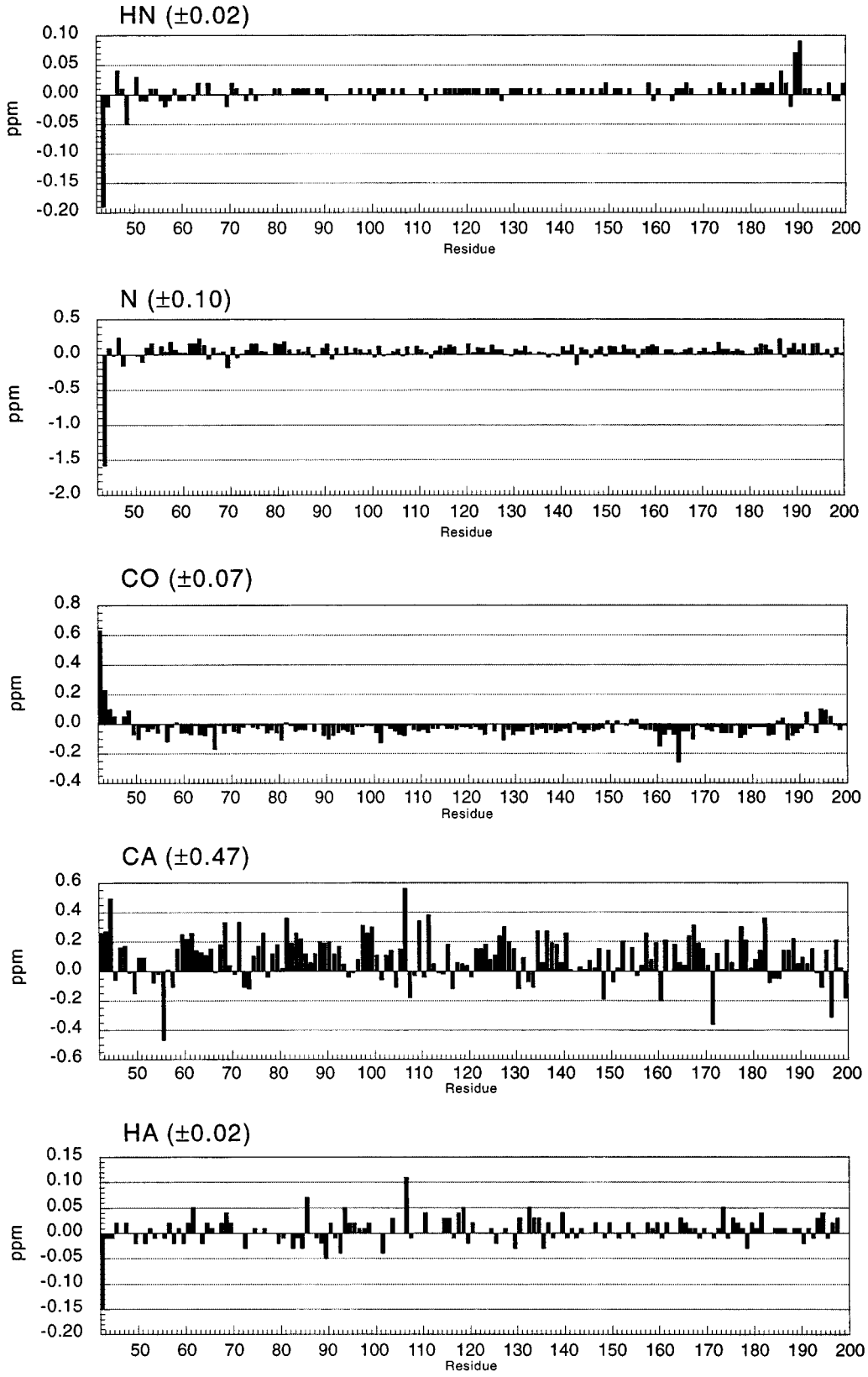

FIGURE 1: Comparison of backbone chemical shifts of S4 and S4Δ41. Data are plotted as $\Delta\delta = \delta(\text{S4}) - \delta(\text{S4}\Delta41)$, where δ is the chemical shift in parts per million (ppm). Shown in parentheses are the estimated errors in $\Delta\delta$ derived from the spectral resolutions of appropriate data sets. The large error in the C^{α} chemical shifts is due to the limited digital resolution of the HNCACB and CBCA(CO)NH experiments.

dynamic on the picosecond to nanosecond time scale while residues 45−200 form a stable tertiary structure, as was observed for S4Δ41. The T_2 data mirror the NOE values, with respective average values of 387 ± 101 ms and 68.0 ± 14.4 ms for the two sets of residues. The long T_2 values for residues 1−44 are typical of disordered polypeptides, while those for residues 45−200 are typical of a well-ordered protein with an overall correlation time (τ_c) of ∼12 ns.

M1	A2	R3	Y4 b	T5 b	G6 b	P7t b(cis)	M8 b	W9 b	K10 b
I11 b	S12 b	R13	R14	L15	G16	I17	S18	L19	S20
G21	T22	G23	K24	E25	L26	Q27	K28 b c d	R29 b c d x	P30t b(trans) c(cis) d(cis) e(trans)
Y31 b c d e	P32t x(?) x(?)	P33t b(trans) c(cis)	G34 b c	Q35 b	H36 b	G37 b c d	P38t	G39 b	Q40 b
R41 b	R42 b	K43	L44	S45	E46	Y47	G48	L49	Q50

FIGURE 2: Minor conformations in the S4 N-terminus. Conformations labeled b−e indicate that sequential connectivities were identified between residues in that conformation. If no such connectivities could be definitively determined, the conformation is labeled x. The configuration of the X−Pro peptide bonds (*cis* or *trans*) for each minor conformation is listed if known. All X−Pro bonds in the major conformation were found to be *trans* and are so indicated by the suffix t.

Interestingly, the T_2 and NOE values for residues 1−44 do not appear to be uniform; two regions centered near residues 10 and 30 have both reduced T_2 and increased NOE values relative to flanking regions near residues 20 and 35. In contrast to the T_2 and NOE data, the T_1 values for the two sets of residues (1−44 and 45−200) are essentially identical, with respective average values of 707 ± 60 ms and 702 ± 83 ms. This result is easily explained by the fact that essentially every unique value of T_1 is consistent with two values of the overall correlation time (*45*). Theoretical calculations of relaxation times for dipolar relaxation of the H−^{15}N system indicated that the T_1 data could be satisfied by τ_c values of ∼0.5 ns or ∼12 ns. Therefore, taken together, the relaxation data strongly suggest that residues 1−44 are significantly more dynamic than residues 45−200. These results also correlate well with relaxation data previously measured for S4Δ41, which suggested an overall correlation time of ∼9 ns (M. A. Markus and D. A. Torchia, unpublished experiments). A more detailed analysis of the relaxation data for these proteins will be published elsewhere.

Collection of Structural Restraints. Our strategy for collecting distance restraints was analogous to that used for resonance assignment. After collecting both ^{15}N- and ^{13}C-separated NOESY data sets with mixing times of 75 ms, we first compared the data for residues 44−200 to the corresponding data previously collected on S4Δ41 (*9*). We note that M42 in S4Δ41 is replaced by R42 in wild-type S4, rendering any comparison of this residue meaningless, along with sequential contacts between residue 42 and K43. Therefore, we excluded residues 42 and 43 from the comparison. In addition, the first helix of S4Δ41 begins with S45, and thus the data for L44 have little structural relevance. A careful comparison revealed no significant differences in the NOESY cross-peaks of the two sets of spectra, thus strongly confirming the conclusion obtained from the chemical shift comparison, that the structures of these residues in S4 and S4Δ41 are the same. We next turned to residues 1−41 and first asked if we should expect to observe efficient cross-relaxation for them. Theoretical calculations indicated that, at both 500 and 750 MHz, we should be able to observe negative NOEs for spins with correlation times longer than 500 ps. Since our relaxation data suggested that the correlation time for these residues is on the order of 1 ns, we proceeded with a standard analysis of the NOESY data on this basis. These predictions were confirmed by the data, and we were able to derive a total of 109 distance restraints from the NOESY data sets. These included 49 intraresidue and 59 sequential restraints, along with a single short-range restraint. These restraints are summarized in Figure 5. Even though we were able to observe negative NOEs, the predicted cross-relaxation rates for protons in these residues are small (<1 s^{-1}), and we therefore collected an additional ^{13}C-separated NOESY experiment at 800 MHz with a longer mixing time of 200 ms. While numerous NOE cross-peaks were confirmed by this experiment and some additional intraresidue NOEs were observed, no new distance restraints could be definitively derived from these data.

In addition to the cross-peaks from protein protons described above, we found exchange cross-peaks from water protons to all observable backbone amide protons of the N-terminus in NOESY spectra recorded with a 75 ms mixing time. We quantitated the intensities of the water cross-peaks and their corresponding amide diagonal peaks, and for each residue calculated an estimate of the exchange rate constant k from the ratio of these intensities:

$$\frac{I_c}{I_d} = \frac{1 - \exp(-2k\tau)}{1 + \exp(-2k\tau)} \quad (1)$$

where I_c and I_d are the intensities of the cross-peak and diagonal peak, respectively, and τ is the mixing time (*46*). We then computed protection factors $P = k_{calc}/k$, where k_{calc} is the random coil exchange rate constant for that residue predicted from neighboring group effects (*47*). This analysis was performed for all well-resolved and observable amide resonances of residues 1−45 except that of H36, for which $I_c \approx I_d$ at $\tau = 75$ ms. Interestingly, the predicted exchange rate constant for H36 is approximately 10 times that of the remaining residues, and its amide resonance is strongly broadened in HSQC spectra, presumably because of its rapid exchange with solvent water. The mean value of P for the 26 residues analyzed was 0.63 ± 0.57.

In addition to the NOESY experiments, we also collected an HNHA data set in order to derive information regarding the ϕ dihedral angle. We were able to measure $^3J_{HNH\alpha}$ for 22 residues in the N-terminus, and of these, all but two fell within the range of 6.0−8.0 Hz and thus were not used as structural restraints. The two exceptions were T5 and M8, which yielded $^3J_{HNH\alpha}$ values of 8.3 and 5.3 Hz, respectively. We found that when these restraints were not included, all of the resulting structures had values of ϕ that were consistent with $^3J_{HNH\alpha}$ for M8 but not T5. Therefore, we only included the restraint for T5 in the final calculations.

Structure Calculations. Given that the structure of residues 44−200 appeared to be essentially equivalent to that determined previously for S4Δ41, we decided to focus our calculations solely on the N-terminus. We thus constructed a molecular template consisting of residues 1−45, corresponding to the N-terminus plus residues 42−45, which were mostly unstructured in S4Δ41. We initially calculated families of 25 structures beginning with random coordinates, followed by regularization and refinement by simulated annealing. After repeating the refinement protocol six successive times,

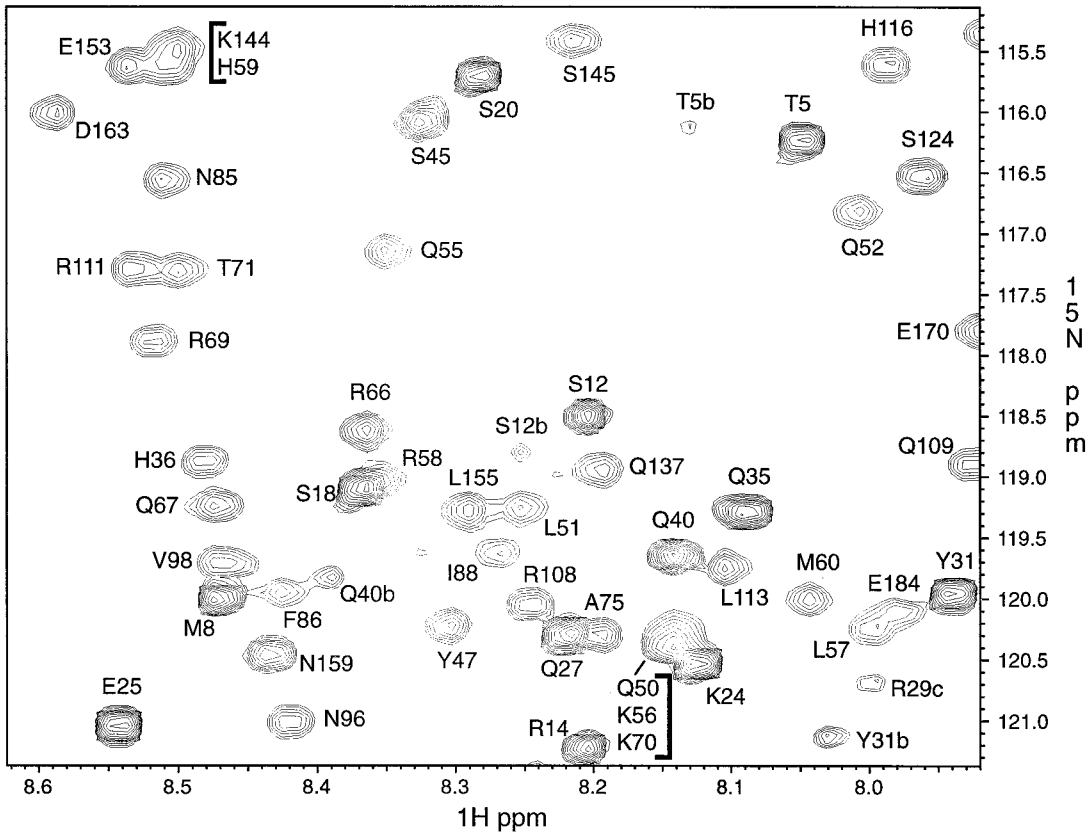

FIGURE 3: Portion of a $^{1}H-^{15}N$ HSQC spectrum of S4 acquired at 750 MHz and 37 °C. The sample conditions were 0.3 mM [*U*-^{15}N]S4 in 20 mM sodium acetate-d_3, pH 5.3, and 250 mM KCl.

we could not detect any improvement measured either by individual structure energies or by the frequency and number of restraints satisfied. In general, 20−23 of the 25 structures contained no restraint violations, and these accepted structures were analyzed further. As expected from the small number of restraints, the resulting structures were largely disordered and contained no stable elements of secondary structure. However, preliminary calculations of angular order parameters, S^{angle}, for backbone ϕ and ψ (*48*), along with pairwise RMSD values for backbone atoms of short residue segments, suggested that residues 12−15 and 30−33 exhibited a higher degree of order compared to the remainder of the peptide. To investigate this finding further, we calculated two sets of 250 structures each, one with and one without the experimental restraints. In each case, the same protocols were used, except that the structures were only subjected to one round of refinement. Of the set of 250 structures calculated with restraints, 217 contained no restraint violations greater than 0.5 Å or 5° (the remainder violated no more than three restraints), and these structures were accepted and analyzed. We again calculated $S^{angle}(\phi)$ and $S^{angle}(\psi)$, along with a series of pairwise backbone RMSD values for segments of four residues. The results of these calculations are displayed in Figures 6 and 7.

Of significance are residues that have higher values of S^{angle} when restraints are included in the calculation. Beginning with ϕ, we note two trivial cases: first, the values for the five prolines are near unity given the fixed geometry of the pyrrolidine ring, and second, that of T5 is high as a direct result of the torsional restraint on this residue. Disregarding these cases, the residues that show an increase when restraints are included are Y4, R13, R14, L15, and I17. The most notable of these are R13, L15, and I17, each of which has an S^{angle} value near 0.7 when restraints are included. Turning to ψ, we note that several ranges of residues show substantial increases in S^{angle}, particularly 10−15, 26−33, and 40−44. However, only R13, R14, L15, and I17 show increases for both dihedral angles. In addition, we note that P30, P32, and P33 show a substantial increase in order at ψ, which when combined with their inherent order at ϕ implies that these residues are particularly well-ordered when restraints are included.

Considering the pairwise RMSD calculations of backbone atoms of four-residue segments shown in Figure 7, we find a pattern similar to that shown by the angular order parameters. While most of the residues show similar pairwise RMSDs both when restraints are present and when they are not, two segments have sharply reduced RMSDs when restraints are included. These segments are $S_{12}RRLG_{16}$ and $P_{30}YPPG_{34}$, which each have RMSDs near 0.8 Å or less, indicating that the backbone atoms of these residues are reasonably well-ordered in solution.

In an attempt to understand the nature of the structures formed by segments $S_{12}RRLG_{16}$ and $P_{30}YPPG_{34}$, we analyzed the (ϕ, ψ) distributions of the individual residues in the 30 accepted structures having the lowest energies. Because the data poorly defined these torsions for residues S12, G16, and G34, we selected those structures that placed ϕ and ψ for the remaining residues in the most favored or additionally allowed regions of the Ramachandran plot (*49, 50*). This resulted in sets of 21 structures each for $S_{12}RRL_{15}$ and $P_{30}YPP_{33}$. Analysis of the structures for $S_{12}RRL_{15}$ revealed that dihedral angles R14 ϕ and L15 ψ each adopted two distinct sets of values. The latter angle has little influence on the

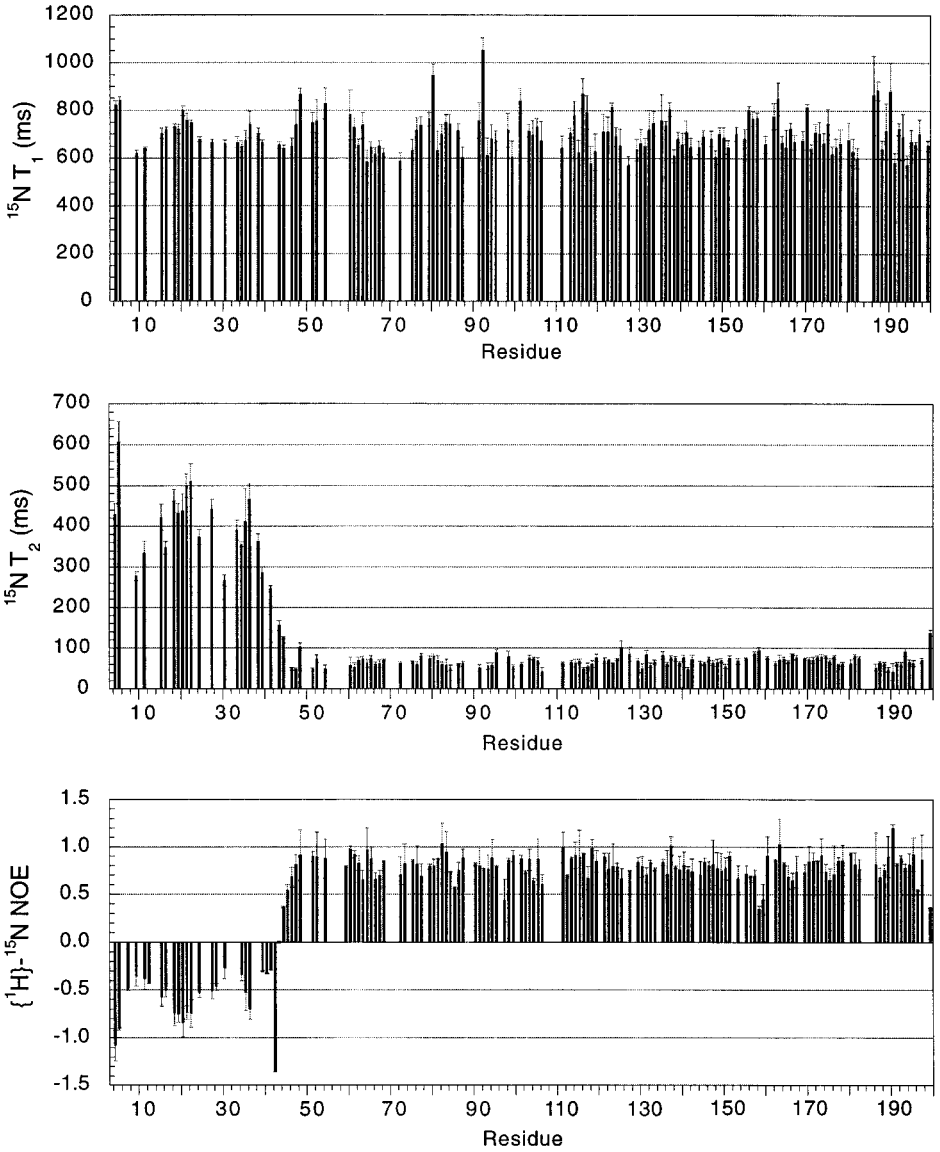

FIGURE 4: ^{15}N relaxation data measured at 500 MHz and 37 °C. Sample conditions were as indicated in the caption of Figure 3. Residues without data are either prolines or had severely overlapped resonances.

conformation of $S_{12}RRL_{15}$; however, R14 ϕ lies in the middle of this segment and significantly influences its conformation. Of the 21 structures, 16 placed R14 ϕ and ψ at 50.6 ± 8.0° and 61.2 ± 10.6°, respectively, while the remaining five structures placed these angles at −63.9 ± 17.4° and 70.2 ± 25.8°. Thus, the major conformation ensemble for R14 lies near the favorable region of the Ramachandran plot corresponding to a left-handed α helix, while the minor ensemble lies near the favorable region corresponding to a β-strand. In addition to ϕ and ψ, we calculated χ^1 and $\chi^{2,1}$ dihedral angles for the L15 side chain, which was restrained by six NOEs. These calculations showed that 15 of the 16 structures of the major ensemble placed both χ^1 and $\chi^{2,1}$ in a single, well-defined conformation. These 15 structures are summarized in Table 1 and displayed in Figure 8, along with the 21 structures of $P_{30}YPP_{33}$, all of which placed ϕ and ψ for these residues in a single conformational ensemble.

Sequence Alignment. We conducted a BLAST search (*51*) for residues 1−40 to identify the portions of the S4 N-terminus that are most conserved across various prokaryotes. A simple gapped BLAST search produced 27 unique

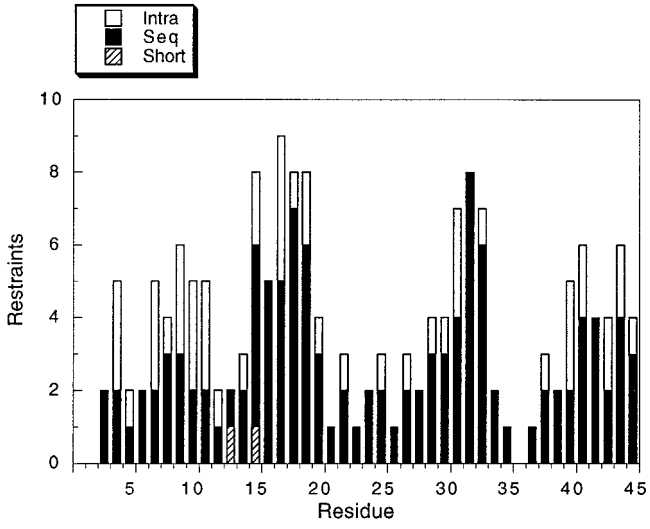

FIGURE 5: Summary of NOESY-derived distance restraints used in structure calculations. Intraresidue, sequential, and short-range restraints are indicated by open, filled, and hatched bars, respectively.

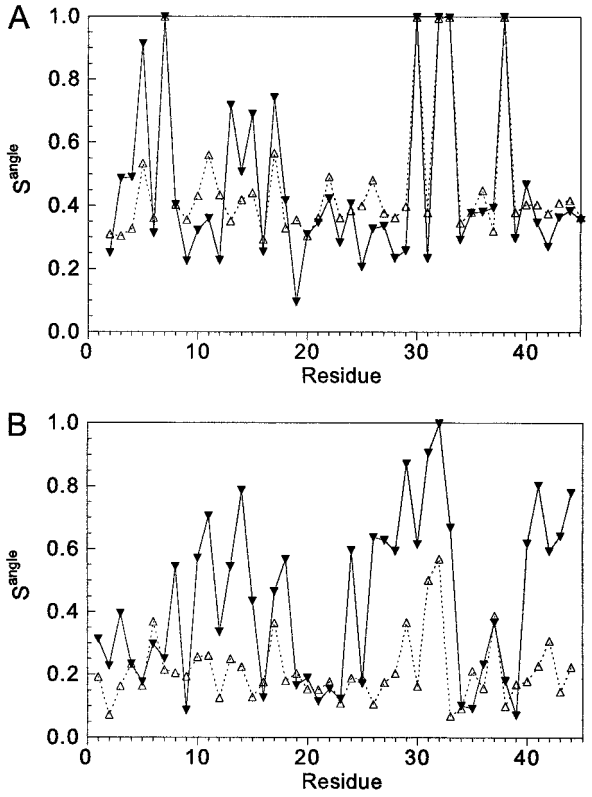

FIGURE 6: Angular order parameters (S^{angle}) measured for (A) ϕ and (B) ψ dihedral angles. (▲) Values derived from the family of 217 structures calculated with restraints that contained no restraint violations; (Δ) values derived from the family of 250 structures calculated without structural restraints.

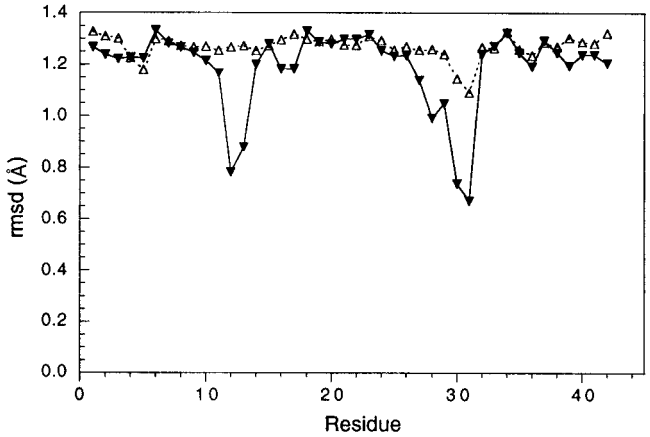

FIGURE 7: Pairwise backbone RMSD values calculated for four-residue segments, and plotted according to the first residue in each segment. (▲) Values derived from the family of 217 structures containing no restraint violations; (Δ) values derived from the family of 250 structures calculated without structural restraints.

sequences, all encoding protein S4 from either a eubacterial or a chloroplast ribosome. These sequences are summarized in Figure 9, along with the ribosomal S4 sequences for *E. coli* and *Salmonella typhimurium*, which were two of 162 sequences produced by a PSI-BLAST search. We note that there appear to be three regions in the N-terminus that contain highly conserved residues: residues 2−7, 12−16, and 28−37. In particular, the latter two regions contain the five residues that are identically conserved in all listed species (R13, R14, P33, G34, and H36) while also containing additional residues that are identically conserved in at least 75% of the species listed (G16, Q35, and G37). These two conserved regions are separated by a region whose length and composition are highly variable among the listed species.

Table 1: Mean Dihedral Angles and Standard Deviations of Major Conformations of $S_{12}RRL_{15}$ and $P_{30}YPP_{33}$

residue	ϕ	ψ
R13	−150.4 ± 18.9	102.0 ± 34.2
R14	51.4 ± 7.4	61.8 ± 10.7
L15 I[a]	−102.9 ± 26.4	−16.8 ± 11.4
L15 II[b]	−114.0 ± 17.7	−156.1 ± 12.4
P30	−73.7 ± 2.3	149.8 ± 42.7
Y31	−107.2 ± 47.2	131.7 ± 19.4
P32	−72.8 ± 2.4	171.8 ± 3.0
P33	−78.5 ± 2.7	151.3 ± 45.7
	χ^1	$\chi^{2,1}$
L15	64.2 ± 5.7	124.4 ± 18.2

[a] Representing 8 of 15 structures. [b] Representing 7 of 15 structures.

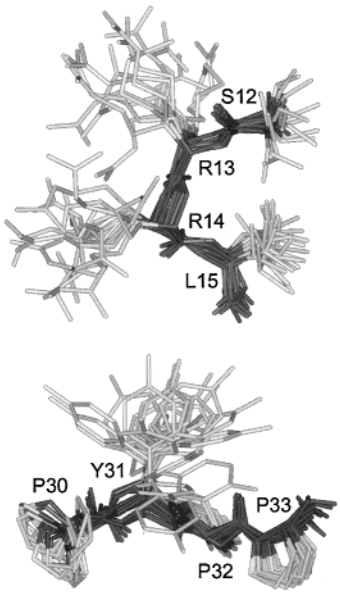

FIGURE 8: Structures taken from a set containing the 30 lowest energy members of the family of 217 structures calculated with restraints. Backbone atoms are shown in black, and non-hydrogen side chain atoms are shown in gray. Residues are labeled near the C^{α} positions. The structures shown place both ϕ and ψ in the most favored or additionally allowed regions of the Ramachandran plot (*50*), except those for S12, which were poorly defined by the data.

DISCUSSION

The Fold of Residues 45−200 Is Unaffected by the Presence of the N-Terminus. One of our initial concerns in this study was to determine whether the previously determined structure of S4Δ41 was valid for the corresponding residues (45−200) of the full-length protein. The first evidence that this was the case was provided by the backbone chemical shifts, which are known to be sensitive to local structural changes. Despite different purification protocols, the assignments for these spins are remarkably similar between S4 and S4Δ41. Even stronger evidence was provided by the ^{15}N- and ^{13}C-separated NOESY data, which also were indistinguishable between the two proteins. The only significant difference detected between these residues in S4 and S4Δ41 was revealed by the ^{15}N relaxation data. While in general the trends in the relaxation data are very

				5					10						15										20							
B. stearo.	**A**	**R**	**Y**	**T**	**G**	**P**	**M**	**W**	**K**	**I**	**-**	**S**	**R**	**R**	**L**	**G**	**-**	**-**	**I**	**S**	**L**	**-**	**-**	**-**	**S**	**G**	**-**	**-**	**-**	**T**	**G**	**-**
Eubacteria																																
B. subtilis	o	o	o	o	o	o		o	o	+	-	o	o	o	o	o	-	-	o	o	o	-	-	-	o	o	-	-	-	o	o	-
M. leprae	o	o	o	o	o	o	+		+		-	o	o	o	o		-	-	o		o	-	-	-		o	-	-	-			-
M. bovis	o	o	o	o	o	o	+		+		-	o	o	o	o		-	-			o	-	-	-		o	-	-	-			-
M. tuberculosis	o	o	o	o	o	o	+		+		-	o	o	o	o		-	-			o	-	-	-		o	-	-	-			-
S. sp.	+	o	o		o	o			+	o	-		o	o	o	o	-	-	-		o	-	-	-		o	-	-	-			-
C. trachomatis	o	o	o		o	o			+	o	-	+	o	o		o	-	-		+	+	-	-	-		o	-	-	-		o	-
C. pneumoniae	o	o	o		o	o			+	+	-	+	o	o		o	-	-		+	+	-	-	-		o	-	-	-			-
A. aeolicus	-	o	o		o	o			o	+	-		o	o		o	-	-	+		+	-	-	-	o	o						-
T. thermophilus	-	o	o		o	o	+		+	+	-		o	o		o	-	-	+		o	-	-	-			-	-	-		o	-
T. maritima	o	o	o	o	o	o	+		o	+	-		o	o		o	-	-	+		o	-	-	-			-	-	-		o	-
M. pneumoniae	-	+	o	o	o		+	+	o		-	o	o	o	o	o	-	-		o	o	-	-	-			-	-	-			-
M. genitalium	-	+	o	o	o		+	+	o		-	o	o	o	o	o	-	-		o	o	-	-	-			-	-	-			-
H. pylori	o	o	o		o		+		+	+	-		o	o		o	-	-	+	o	o	-	-	-	+		-	-	-		o	-
R. prowazekii	-	-	-	-	-	-	-	+	o		-	o	o	o	o	o	-	-	+	o	o	-	-	-		o	-	-	-			-
B. aphidicola	o	o	o		o	o			o	+	-	o	o	o		o	-	-			o				o	o			-	+		-
Z. mobilis	-	-	-	-	-	-	-	+	o	o	-		o	o	+	o	-	-		+	+	-	-	-		o	-	-	-			-
B. burgdorferi	-	-	-	-	-	-	-	-	o	+	-		o	o		o	-	-	o	+	+	-	-	-			-	-	-			-
X. campestris	o	o	o		o	o			o	+	-	+	o	o		o			+	o	o	-	-	-			-	-	-			-
E. coli	o	o	o		o	o			o	+	-	o	o	o		o	-	-			o	-	-	-			-	-	-			
S. typhimurium	o	o	o		o	o			o	+	-	o	o	o		o	-	-			o	-	-	-			-	-	-			
Chloroplast																																
P. purpurea	+	o	o		o	o			+	o	-	o	o	o	o	o	-	-	-		o	-	-	-		o	-	-	-			-
C. paradoxa	+	o	o		o	o			+	o	-		o	o	o	o	-	-				-	-	-			-	-	-	o		-
C. ellipsoidea	o	o	o		o	o			+	o	-		o	o	o	o	-	-	-		o	-	-	-		o	-	-	-			-
P. wickerhamii	+	o	o		o	o				o	-		o	o	o	o	-	-	-		o	-	-	-		o	-	-	-			-
O. sinensis	+	o	o		o	o			+	o	-	+	o	o	o	o	-	-	-	+	o	-	-	-		o	-	-	-			-
C. caldarium	+	o	o		o	o			+	+	-		o	o	o	o	-	-				-	-	-	+		-	-	-	o		-
C. sp.	+	o	o		o		+		o	o	-		o	o	o	o	-	-	-		o	-	-	-		o	-	-	-			-
C. merolae	+	o	o		o	o			+	o			o	o	o	o	-	-	o			-	-	-	+		-	-	-	o		-

					25															30					35					40	
B. stearo.	**-**	**-**	**-**	**-**	**K**	**E**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**L**	**-**	**-**	**-**	**Q**	**K**	**R**	**P**	**Y**	**P**	**P**	**G**	**Q**	**H**	**G**	**P**	**G**	**Q**	**R**
Eubacteria																																
B. subtilis	-	-	-	-	o	o	-	-	-	-	-	-	-	o	-	-	-	+	o	o	o	o		o	o		o	o	o	o	o	o
M. leprae	-	-	-	-	+		-	-	-	-	-	-	-		-	-	-	+	o	o	o	o	o	o	o	o	o	o	-	-	-	-
M. bovis	-	-	-	-	+		-	-	-	-	-	-	-		-	-	-	+	o	o	o	o	o	o	o	o	o	o	-	-	-	-
M. tuberculosis	-	-	-	-	+		-	-	-	-	-	-	-		-	-	-	+	o	o	o	o	o	o	o	o	o	o	-	-	-	-
S. sp.	-	-	-	-	+	+	-	-	-	-	-	-	-		-	-	-		+	o		o	o	o	o	o	o	o			+	o
C. trachomatis	-	-	-	-	+			-	-	-	-	-	-	o	-	-	-		+	+	o		o	o	o	o	o	o	-	-	-	-
C. pneumoniae	-	-	-	-	+			-	-	-	-	-	-	o	-	-	-		o	+	o	+	o	o	o	o	o	o	-	-	-	-
A. aeolicus	-	-	-	-	o		-	-	-	-	-	-	-	o	-	-	-		+	o		o	o	o	o	o	o	o	-	-	-	-
T. thermophilus	-	-	-	-	+									+	-	-	-	+	+	o	o	o	o	o	o	o	o	o	-	-	-	-
T. maritima	-	-	-	-	+										-	-	-	+	+	o	o	o		o	o	o	o	o	-	-	-	-
M. pneumoniae	-	-	-	-	o	o	-	-	-	-	-	-	-					+	o	o				o	o	o	o	o	-	-	-	-
M. genitalium	-	-	-	-	o	o	-	-	-	-	-	-	-					+	o	o				o	o	o	o	o	-	-	-	-
H. pylori	-	-	-	-	+									o	-	-	-		o	o		o		o	o	o	o	o	-	-	-	-
R. prowazekii	-	-	-	-	o	+		-	-	-	-	-	-		-	-	-			o		o		o	o	o	o	o	-	-	-	-
B. aphidicola	-	-	-	-			-	-	-	-	-	-	-		-	-	-		o			+	o	o	o	o	o	o	-	-	-	-
Z. mobilis	-	-	-	-	o			-	-	-	-	-	-	+	-	-	-		+	o		o		o	o	o	o	o			+	o
B. burgdorferi	-	-	-	-	o				-	-	-	-	-	o	-	-	-	+	o	+	o	+	o	o	o		o	o			+	+
X. campestris	-	-	-	-	+		-	-	-	-	-	-	-	o					o					o	o	o	o	o			+	+
E. coli							-	-	-	-	-	-	-		-	-	-		o					o	o	o	o	o			+	
S. typhimurium							-	-	-	-	-	-	-		-	-	-		o					o	o	o	o	o			+	
Chloroplast																																
P. purpurea	-	-	-	-	+	+	-	-	-	-	-	-	-		-	-	-		o	o	o	o	o	o	o	+	o	o				o
C. paradoxa	-	-	-	-	o		-	-	-	-	-	-	-	+	-	-	-	+	o	o		o	o	o	o	o	o	o				o
C. ellipsoidea	-	-	-	-	+	+	-	-	-	-	-	-	-		-	-	-			o		+	o	o	o	o	o	o	o		+	+
P. wickerhamii	-	-	-	-	o	+	-	-	-	-	-	-	-	+	-	-	-		+				o	o	o	o	o	o	-	-	-	-
O. sinensis	-	-	-	-	+	+	-	-	-	-	-	-	-		-	-	-		o	+				o	o	o	o	o	-	-	-	-
C. caldarium	-	-	-	-	o		-	-	-	-	-	-	-		-	-	-	+	o	o				o	o	+	o	-	-	-	-	-
C. sp.	-	-	-	-	+	+	-	-	-	-	-	-	-		-	-	-			o				o	o	o	o	o				o
C. merolae	-	-	-	-	o		-	-	-	-	-	-	-		-	-	-		o	o				o	o	+	o	-	-	-	-	-

FIGURE 9: Sequence alignment of S4 N-termini from a variety of eubacterial and chloroplast ribosomes. These sequences are compared to that of *B. stearothermophilus* (displayed in boldface type). Symbols have the following meanings: (○) identical residue; (+) similar residue; (−) gap; a blank space indicates a dissimilar residue.

similar for these residues in the two proteins, the average T_2 at 500 MHz for residues 45−200 in S4Δ41 was 104 ± 17 ms (M. A. Markus and D. A. Torchia, unpublished experiments), while that for S4 was 68.0 ± 14.4 ms as mentioned above. This is consistent with S4 having a longer overall correlation time than S4Δ41, as expected given the 25% increase in molecular weight. Therefore, the data suggested that the only significant effect of the N-terminus on residues 45−200 was simply to reduce their overall rate of tumbling.

The N-Terminus Is Mobile on Multiple Time Scales. While the relaxation data suggest that residues 45−200 tumble in solution as a well-structured unit, these data paint a radically different picture of the behavior of residues 1−44. The mere fact that the $\{^1H\}-^{15}N$ NOE values for these residues are negative indicates that they are mobile on very fast time scales (picoseconds to nanoseconds); on the other hand, the fact that we observed separate peaks for the minor conformations produced by proline *cis*−*trans* isomerization establishes that there is slow rotation (time scales milliseconds to seconds) around four peptide bonds. Against the background of these two motional regimes, the data suggest that there are additional subtleties to the dynamics of these residues. If one proceeds along the backbone from Q50 toward M1, one finds that the T_2 values smoothly increase to a maximum at G37 and then decrease to a minimum near Y31, followed by another maximum near G21 and a minimum near I11, and then a final increase toward M1. Notably, these trends are mirrored almost precisely in the $\{^1H\}-^{15}N$ NOE data, with T_2 values decreasing as NOEs increase and vice versa. These trends are therefore complementary and, given the changes in the NOE values, show that at least the fast time scale motions are not uniform. Interestingly, the two T_2 minima correspond very well with $S_{12}RRLG_{16}$ and $P_{30}YPPG_{34}$, the two segments identified by structural calculations as having increased order in solution, lending support to a model in which the overall motion of the peptide backbone is reduced at these segments.

In principle, further information on backbone dynamics and hydrogen bonding can be inferred from the proton exchange rates of the backbone amides. Given that we observed cross-peaks from water protons to all resolved backbone amide protons in the N-terminus, we calculated estimated protection factors from these data. We note that these protection factors are based on estimated rate constants that may also contain contributions from both direct and exchange-mediated cross-relaxation; however, given the highly flexible nature of these residues, cross-relaxation rates are small relative to the rate of direct exchange with solvent protons. As mentioned above, all of the observable amide protons in the N-terminus have protection factors that both are near unity and are relatively invariant throughout the N-terminus, suggesting that the N-terminus behaves largely as a flexible peptide. While the protection factors for residues 12−18 are slightly higher than those of surrounding residues, and the protection factor for Y31 is the second highest in the N-terminus, these values are no more than twice that of the mean value of *P*. We therefore find little evidence of transient hydrogen bonding in the N-terminus. On the basis of all these data, we propose a model in which the N-terminus is, in general, flexible and disordered but which nevertheless contains two segments, $S_{12}RRLG_{16}$ and $P_{30}YPPG_{34}$, that have reduced flexibility and that may be represented by a small number of solution conformations.

A Backbone Bend Orients Two Conserved Arginines. Residues 12−16 form one of two segments identified by the data as having increased structural order relative to the N-terminus as a whole. The only NOE connecting nonsequential residues in the N-terminus was observed in this segment, and this NOE along with nine others significantly reduced the backbone RMSD values for these residues and also increased the angular order parameters for ϕ and ψ for R13, R14, and L15. Careful inspection of the family of selected low-energy structures described above confirmed that while the peptide backbone does bend at residues 13−15, this bend does not form a typical tight turn. The ϕ and ψ values for R13 and R14 (putative residues 2 and 3 of the turn) conform most closely to those of a type II turn (*52*); however, the "outer" two dihedral angles (R13 ϕ and R14 ψ) differ by ∼50° from their canonical values. These values position the amide proton of L15 away from residues 12−14 so that the characteristic 1−4 hydrogen bond between this proton and the carbonyl oxygen of S12 cannot be formed. Rather, it is the side chain of L15 that is directed back toward S12 by the NOE between L15 H^{β} and R13 H^N and by five additional L15 side-chain NOEs, two of which were strong intraresidue contacts to the L15 H^N. These restraints were sufficient to fix both χ^1 and $\chi^{2,1}$ of the L15 side chain in one predominant conformation (Table 1). The bend was further determined by strong or medium restraints on $d_{\alpha N}$ for R13, R14, and L15. An additional NOE is predicted by the model between the amide protons of R14 and L15; however, the proton resonance frequencies of these two spins differ by only 0.04 ppm, preventing us from observing this NOE. Perhaps the most striking feature of this bend is the resulting parallel positioning of the side chains of R13 and R14, which are both directed away from the bend out into solvent. Given that both of these arginines are absolutely conserved in a wide variety of eubacteria and chloroplast ribosomes, and given the known propensity for arginines to be involved in RNA binding (*53*, *54*), the segment containing these two residues is a reasonable candidate for a preordered motif that may be involved in interactions between S4 and the 16S RNA.

A Proline-Rich Segment Forms a Nascent Turn of a Polyproline II Helix. The second segment of the N-terminus identified as having increased structural order is the proline-rich segment $P_{30}YPPG_{34}$. Proline has long been known to restrict the conformational flexibility of the polypeptide backbone (*55*), and this is reflected in the slightly decreased backbone RMSD values for this segment when no experimental restraints are included in the structure calculations (Figure 7). A similar reduction in RMSD is also evident near P7. However, unlike the region near P7, the RMSD values for $P_{30}YPPG_{34}$ decreased markedly when the experimental restraints were included in the calculations. This segment was restrained by 15 NOEs, including four backbone to side-chain contacts and five side-chain to side-chain contacts between sequential residues. These restraints also increased the $S^{angle}(\psi)$ values of residues 30−33. Turning to the ϕ torsions, no change in S^{angle} was observed for the proline residues, as expected given the relatively fixed geometry of their pyrrolidine rings; however, no change was observed for Y31 ϕ either, suggesting that this torsion is ill-defined

by the data. Inspection of the (ϕ, ψ) distribution of Y31 showed that ϕ populated two states: one in the β strand region and one in an unfavored region, (ϕ, ψ) = (~60°, 120−180°), of the Ramachandran plot. When we limited the set of lowest energy structures to those that placed ϕ and ψ within the most favored or additionally allowed regions (*50*), the situation naturally improved, but Y31 ϕ retained significant variability. Similarly, the χ^1 and χ^2 dihedral angles of Y31 are poorly defined by the data, and as shown in Figure 9, the aromatic side chain sampled a variety of conformations.

The conformations of proline-rich regions in proteins have been increasingly studied in recent years, particularly because such segments have been found to be involved in numerous protein−protein interactions (*56*). While increasing the proline content of a peptide segment has been shown to reduce the overall backbone dynamics in solution (*57*), it has also been found to predispose such segments to adopt a secondary structure known as the polyproline II helix (*58*). The polyproline II helix is typically observed for sequences of four or more adjacent prolines; however, these helices have been found in segments containing no prolines (*59*), and a survey of XPPX sequences in solved structures revealed a tendency of all four residues to adopt this secondary structure (*55*). While the ϕ torsion in prolines tends to be near −65°, in the polyproline II helix this torsion shifts to near −78°, while ψ is positioned near +146° (*58*). Interestingly, the mean values of ϕ and ψ for residues 30−33 agree well with these canonical values for the polyproline II helix. Moreover, the resulting structures position P30 and P33 in very similar orientations, as expected given the three-residue length of a polyproline II helix turn. Thus, the data suggest that the $P_{30}YPP_{33}$ segment forms a single, nascent turn of a polyproline II helix. We again note that this region of the N-terminus, spanning residues 28−37, is well conserved in eubacteria and chloroplast ribosomes, particularly residues 33−37. Given that S4 is known to be involved in the binding of other ribosomal proteins in the assembling ribosome and that the equivalent residues in the *E. coli* protein (*33*−*47*) were identified as being necessary for proper ribosome assembly, it is reasonable to suggest that this proline-rich region may be a putative site for these important protein−protein interactions.

ACKNOWLEDGMENT

We thank Frank Delaglio for NMRPipe and NMRDraw and Dan Garrett for PIPP. We thank Nga Nguyen for performing N-terminal peptide sequencing and Lewis Pannell for protein mass spectrometry. We also thank Michelle Markus, Rieko Ishima, and Yun-Xing Wang for many helpful discussions.

SUPPORTING INFORMATION AVAILABLE

Four tables containing the chemical shift assignments determined for S4 along with the experimental parameters for the NMR spectra described in this work. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Schuap, H. W., Sogin, M., Woese, C., and Kurland, C. G. (1971) *Mol. Gen. Genet. 114*, 1−8.
2. Stern, S., Wilson, R. C., and Noller, H. F. (1986) *J. Mol. Biol. 192*, 101−110.
3. Vartikar, J. V., and Draper, D. E. (1989) *J. Mol. Biol. 209*, 221−234.
4. Nowotny, V., and Nierhaus, K. H. (1988) *Biochemistry 27*, 7051−7055.
5. Deckman, I. C., and Draper, D. E. (1985) *Biochemistry 24*, 7860−7865.
6. Rosset, R., and Gorini, L. (1969) *J. Mol. Biol. 39*, 95−112.
7. Allen, P. N., and Noller, H. F. (1989) *J. Mol. Biol. 208*, 457−68.
8. Gorini, L. (1974) in *Ribosomes* (Nomura, M., Tissières, A., and Lengyel, P., Eds.) pp 791−803, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
9. Markus, M. A., Gerstner, R. B., Draper, D. E., and Torchia, D. A. (1998) *EMBO J. 17*, 4559−4571.
10. Markus, M. A., Gerstner, R. B., Draper, D. E., and Torchia, D. A. (1999) *J. Mol. Biol. 292*, 375−387.
11. Davies, C., Gerstner, R. B., Draper, D. E., Ramakrishnan, V., and White, S. W. (1998) *EMBO J. 17*, 4545−4558.
12. Changchien, L.-M., and Craven, G. R. (1986) *Nucleic Acids Res. 14*, 1957−1966.
13. Baker, A.-M., and Draper, D. E. (1995) *J. Biol. Chem. 270*, 22939−22945.
14. Changchien, L.-M., and Craven, G. R. (1976) *J. Mol. Biol. 108*, 381−401.
15. Changchien, L.-M., Schwarzbauer, J., Cantrell, M., and Craven, G. R. (1978) *Nucleic Acids Res. 5*, 2789−2799.
16. Changchien, L.-M., and Craven, G. R. (1978) *J. Mol. Biol. 125*, 43−56.
17. Changchien, L.-M., Conrad, R. C., and Craven, G. R. (1986) *Nucleic Acids Res. 14*, 6929−6944.
18. Nomura, M., Traub, P., and Bechmann, H. (1968) *Nature 219*, 793−799.
19. Higo, K., Held, W., Kahan, L., and Nomura, M. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 944−948.
20. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60−89.
21. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 6*, 277−293.
22. Garrett, D. S., Powers, R., Gronenborn, A. M., and Clore, G. M. (1991) *J. Magn. Reson. 95*, 214−220.
23. Santoro, J., and King, G. C. (1992) *J. Magn. Reson. 97*, 202−207.
24. Vuister, G. W., and Bax, A. (1992) *J. Magn. Reson. 98*, 428−435.
25. Grzesiek, S., and Bax, A. (1993) *J. Am. Chem. Soc. 115*, 12593−12594.
26. Wittekind, M., and Mueller, L. (1993) *J. Magn. Reson., Ser. B 101*, 201−205.
27. Grzesiek, S., and Bax, A. (1992) *J. Am. Chem. Soc. 114*, 6291−6293.
28. Grzesiek, S., and Bax, A. (1993) *J. Biomol. NMR 3*, 185−204.
29. Grzesiek, S., and Bax, A. (1992) *J. Magn. Reson. 96*, 432−440.
30. Kay, L. E., Ikura, M., Tschudin, R., and Bax, A. (1990) *J. Magn. Reson. 89*, 496−514.
31. Powers, R., Gronenborn, A. M., Clore, G. M., and Bax, A. (1991) *J. Magn. Reson. 94*, 209−213.
32. Bax, A., Clore, G. M., and Gronenborn, A. M. (1990) *J. Magn. Reson. 88*, 425−431.
33. Grzesiek, S., Anglister, J., and Bax, A. (1993) *J. Magn. Reson., Ser. B 101*, 114−119.
34. Kay, L. E., Nicholson, L. K., Delaglio, F., Bax, A., and Torchia, D. A. (1992) *J. Magn. Reson. 97*, 359−375.
35. Freedberg, D. I., Wang, Y.-X., Stahl, S. J., Kaufman, J. D., Wingfield, P. T., Kiso, Y., and Torchia, D. A. (1998) *J. Am. Chem. Soc. 120*, 7916−7923.
36. Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., and Bax, A. (1989) *J. Am. Chem. Soc. 111*, 1515−1517.
37. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) *Biochemistry 28*, 6150−6156.

38. Ikura, M., Kay, L. E., Tschudin, R., and Bax, A. (1990) *J. Magn. Reson. 86*, 204−209.
39. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
40. Vuister, G. W., and Bax, A. (1993) *J. Am. Chem. Soc. 115*, 7772−7777.
41. Kuboniwa, H., Grzesiek, S., Delaglio, F., and Bax, A. (1994) *J. Biomol. NMR 4*, 871−878.
42. Brünger, A. T. (1992) *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.
43. Torchia, D. A., Lyerla, J. R., Jr., and Deber, C. M. (1974) *J. Am. Chem. Soc. 96*, 5009−5011.
44. Wishart, D. S., and Sykes, B. D. (1994) *J. Biomol. NMR 4*, 171−180.
45. Goldman, M. (1992) *Quantum Description of High-Resolution NMR in Liquids*, Clarendon Press, Oxford, U.K.
46. Ernst, R. R., Bodenhausen, G., and Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford, U.K.
47. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) *Proteins: Struct., Funct., Genet. 17*, 75−86.
48. Hyberts, S. G., Goldberg, M. S., Havel, T. F., and Wagner, G. (1992) *Protein Sci. 1*, 736−751.
49. Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963) *J. Mol. Biol. 7*, 95−99.
50. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR 8*, 477−486.
51. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res. 25*, 3389−3402.
52. Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167−339.
53. Weiss, M. A., and Narayana, N. (1998) *Biopolymers 48*, 167−180.
54. Patel, D. J. (1999) *Curr. Opin. Struct. Biol. 9*, 74−87.
55. MacArthur, M. W., and Thornton, J. M. (1991) *J. Mol. Biol. 218*, 397−412.
56. Kay, B. K., Williamson, M. P., and Sudol, M. (2000) *FASEB J. 14*, 231−241.
57. Torchia, D. A., and Lyerla, J. R., Jr. (1974) *Biopolymers 13*, 97−114.
58. Williamson, M. P. (1994) *Biochem. J. 297*, 249−260.
59. Adzhubei, A. A., and Sternberg, M. J. E. (1993) *J. Mol. Biol. 229*, 472−493.

BI0013391