Solution Structure of the DNA-Binding Protein Sac7d from the Hyperthermophile *Sulfolobus acidocaldarius*[†,‡]

Stephen P. Edmondson,* Lingshi Qiu, and John W. Shriver*

Department of Medical Biochemistry, School of Medicine, Southern Illinois University, Carbondale, Illinois 62901

Received April 13, 1995; Revised Manuscript Received July 12, 1995[⊗]

ABSTRACT: The Sac7 proteins from the hyperthermophile *Sulfolobus acidocaldarius* are a heterogeneous mixture of small, thermostable, nonspecific DNA-binding proteins. One of these proteins, Sac7d, has been overexpressed in *Escherichia coli* to provide a homogeneous preparation for structure, stability, and function studies. We present here essentially complete sequence-specific ^{1}H NMR assignments for Sac7d, a delineation of secondary structural elements, and the high-resolution solution structure obtained from a full relaxation matrix refinement. The final structure provides an excellent fit to the NMR data with an NOE *R*-factor of 0.27 for backbone NOEs. The structure has a compact globular fold with 82% of the sequence involved in regular secondary structure: an antiparallel two-stranded β-ribbon with a tight turn, followed by a short 3_{10} helix, an antiparallel three-stranded β-sheet, another short 3_{10} helix, and finally four turns of α-helix. The amphipathic α-helix packs across the hydrophobic face of the three-stranded β-sheet in an open-faced sandwich arrangement with at least one turn of the helix exposed beyond the sheet. The hydrophobic face of the β-ribbon packs against a corner of the twisted β-sheet. The single tryptophan responsible for the 88% fluorescence quenching upon DNA binding is exposed on the surface of the three-stranded β-sheet. Lysines 5 and 7, whose monomethylation may be associated with enhanced thermostability, are highly solvent exposed along the inner edge of the two-stranded ribbon. The structure of Sac7d differs in many respects from that reported for the homologous native Sso7d [Baumann et al. (1994) *Nature Struct. Biol. 1*, 808] with a backbone RMSD greater than 3.0 Å, largely due to the packing and length of the C-terminal α-helix which may be important in Sac7d DNA binding.

Extremely thermostable proteins from the hyperthermophiles are expected to provide useful insights into the rules governing protein folding and stability (Jaenicke & Zavodszky, 1990). Hyperthermophiles largely occur within the archaebacteria, or archaea (Woese et al., 1990), and grow at temperatures above 80 °C, often with no growth below 60 °C (Stetter et al., 1990). Examples from the archaea include *Sulfolobus*, *Acidianus*, *Pyrococcus*, *Methanothermus*, and *Thermococcus* (Stetter et al., 1990). Only a few eubacteria are hyperthermophiles, e.g., *Thermotoga*. A large number of proteins have been isolated and characterized recently from the various hyperthermophiles. For example, alcohol dehydrogenase (Ammendola et al., 1992), adenylate kinase (Lacher & Schafer, 1993), chaperonin (Guagliardi et al., 1994; Knapp et al., 1994), cytochrome aa_3 (Anemuller & Schafer, 1990), ribonuclease P (Darr et al., 1990), DNA polymerase (Elie et al., 1989), reverse gyrase (Nadal et al., 1988), and the protease thermopsin (Fusek et al., 1990) have been isolated from *Sulfolobus*. Many of these are large oligomeric proteins, often containing metal cofactors, and provide intriguing systems for structure/stability studies. To make the analysis of folding and stability as tractable as possible, we have taken a reductionist approach and limited our focus to small, monomeric systems lacking metal binding sites and disulfide cross-bridges. Unfortunately, most small proteins are rich in disulfides (e.g., bovine pancreatic trypsin inhibitor, metallothionein), contain metal cofactors (e.g., cytochrome *c*, rubredoxin, parvalbumin), or contain relatively small amounts of secondary structure (Richardson & Richardson, 1989).

Sulfolobus expresses large amounts of several small DNA-binding proteins with molecular weights ranging from 7000 to 10 000 (Dijk & Reinhardt, 1986; Grote et al., 1986; Choli et al., 1988b; Reddy & Suryanarayana, 1988). In *Sulfolobus acidocaldarius*, the 7000 Da class is heterogeneous and is composed of a number of isoforms referred to individually as Sac7a,[1] Sac7b, Sac7c, Sac7d, and Sac7e, in order of increasing basicity (Choli et al., 1988b). Only one member of the 7 kDa class, Sso7d, has been isolated and characterized from *Sulfolobus solfataricus* (Choli et al., 1988a), but additional isoforms are most likely present. [The native Sso7 proteins elute as two distinct species during ion-exchange chromatography (Qiu and Shriver, unpublished results).] None of the Sac7 proteins contain cysteine, and initial studies indicate that they lack metal-binding sites and are monomeric (McAfee et al., 1995).

[†] This work was supported by the Biotechnology Research and Development Corp. (J.W.S. and S.P.E.) and the National Institutes of Health (GM 49686) (J.W.S. and S.P.E.).

[‡] Coordinates of the Sac7d structure (identification code 1SAP) have been deposited in the Protein Data Bank at the Brookhaven National Laboratory.

* Authors to whom correspondence should be sent [telephone, (618) 453-6479; Fax, (618) 453-6408; email, jshriver@som.siu.edu or sedmondson@som.siu.edu].

[⊗] Abstract published in *Advance ACS Abstracts,* September 15, 1995.

[1] Abbreviations: RMSD, root mean square deviation; rMD, restrained molecular dynamics; NOE, nuclear Overhauser enhancement; NMR, nuclear magnetic resonance; Sac7, a group of 7 kDa DNA-binding proteins from *Sulfolobus acidocaldarius,* individually referred to as Sac7a, Sac7b, Sac7c, Sac7d, and Sac7e, in order of increasing basicity; Sso7, a group of 7 kDa DNA-binding proteins from *Sulfolobus solfataricus.*

0006-2960/95/0434-13289$09.00/0

The function of the Sac7 and Sso7 proteins *in vivo* is unknown. They are believed to be "chromatin" proteins (Dijk & Reinhardt, 1986), but they share no sequence homology to any histone. The proteins bind DNA to form nucleoprotein complexes, as visualized in electron micrographs (Dijk & Reinhardt, 1986), and probably play a role in DNA packaging and duplex stabilization in these archaebacteria at the high growth temperatures, viz. 75−85 °C in *S. acidocaldarius* and slightly higher in *S. solfataricus*. In a search for ribonucleases in extreme thermophiles, Fusi et al. (1993) isolated a protein (p2) from *S. solfataricus* with the same amino acid sequence as Sso7d. However, the nuclease activity was extremely low and may be derived from contaminating RNases. Whether or not the 7 kDa DNA-binding proteins are indeed RNases remains to be confirmed, preferably with recombinant protein. In addition, Gualiardi et al. (1992) claimed to have isolated a 7 kDa thermostable enzyme from *S. solfataricus* capable of catalyzing disulfide bond formation, and partial amino-terminal sequencing showed it to coincide with Sso7d. However, further purification by ion-exchange chromatography led to total loss of activity.

Genes for the Sac7d and Sac7e proteins recently have been cloned and sequenced (McAfee et al., 1995). The *sac7d* gene encodes a 66 amino acid protein with a molecular weight of 7600. The gene has been overexpressed in *Escherichia coli*, and various physical properties have been compared to the native Sac7 mixture. Nearly identical UV absorbance, fluorescence excitation and emission, circular dichroism, and DQF-COSY spectra are all indicative of similar global folds for the native and recombinant proteins. In addition, both the recombinant and native proteins bind double-stranded DNA with similar affinities and site sizes, and both stabilize duplex DNA by over 30 °C.

The native Sac7 proteins are remarkably thermostable with an unfolding midpoint temperature at pH 7 near 100 °C (McAfee et al., 1995). These are the smallest, most stable proteins known which unfold *reversibly* and lack disulfide linkages and cofactors. The recombinant Sac7d protein is also thermostable but unfolds approximately 7 °C lower than the native protein under similar conditions (McAfee et al., 1995). The only known difference between the native and recombinant proteins is methylation of the native protein (in addition to the initiating methionine in recombinant Sac7d). Native Sac7 and Sso7 proteins are monomethylated to various degrees at specific lysine residues. Lysine 5[2] is 70%, 25%, and 20% ϵ-aminomonomethylated in native Sac7a, Sac7b, and Sac7d, respectively, while that for lysine 7 is 50%, 40%, and 50%, respectively (Choli et al., 1988b). Monomethylation has also been observed in other *Sulfolobus* proteins (Minami et al., 1985; Maras et al., 1992), and a few authors have speculated that methylation may be associated with heat tolerance (Maras et al., 1992; Baumann et al., 1994).

Few structures of extreme thermophile proteins have been deposited in the Brookhaven Protein Data Bank. The crystal and NMR structures of rubredoxin from the hyperthermophile *Pyrococcus furiosus* have been reported (Blake et al., 1992), as well as 3-isopropylmalate dehydrogenase from *Thermus thermophilus* (Imada et al., 1991) and glyceraldehyde-3-phosphate dehydrogenase from *Thermotoga maritima*. Structures have been deposited for a few proteins from the eubacteria thermophiles *Bacillus stearothermophilus* and *Bacillus thermoproteolyticus*, which grow optimally at approximately 50 °C. Models of structures of thermophile proteins have also been deduced from the X-ray crystal structures of homologous proteins from mesophiles (Menendez-Arias & Argos, 1989).

In this work, we present the three-dimensional solution structure of the recombinant DNA-binding protein Sac7d from the hyperthermophile *S. acidocaldarius* using a direct fitting of the NMR data with a full relaxation matrix refinement. While this work was being completed, the structure of the homologous native Sso7 protein from *S. solfataricus* was reported on the basis of simulated annealing with qualitative restraints (Baumann et al., 1994). The two structures have similar overall global folds but differ with a backbone RMSD greater than 3 Å, primarily due to a C-terminal extended α-helix in Sac7d which is apparently truncated in Sso7d. This helix may be important in Sac7d DNA binding.

MATERIALS AND METHODS

Materials. Recombinant Sac7d protein was overexpressed in *E. coli* and purified as described elsewhere (McAfee et al., 1995). Protein to be used for NMR was dialyzed extensively against distilled water, lyophilized, and then dissolved in 90% H_2O/10% D_2O. A small amount of undissolved material was removed by centrifugation. The pH was measured with a micro-glass electrode and adjusted to pH 4.3 using DCl. The pH was not corrected for the isotope effect. The final protein concentration was approximately 5 mM.

NMR Spectra. All NMR data were collected on a Varian 500 MHz NMR spectrometer and processed using Varian software or transferred to a Silicon Graphics IRIS workstation for processing using FELIX (Biosym Technologies). Two-dimensional spectra (DQF-COSY, NOESY, and HOHAHA) were recorded in the phase-sensitive mode using the hypercomplex method (States et al., 1982). Double-quantum-filtered (DQF) COSY (Piatini et al., 1982; Rance et al., 1983), HOHAHA (Braunschweiler & Ernst, 1983; Bax & Davis, 1985; Bax, 1989), and NOESY (Jeener et al., 1979; Macura et al., 1981) pulse sequences were implemented using standard procedures except that the phase cycling in the HOHAHA experiment was reduced by two to facilitate implementation on our system. The spectrometer transmitter frequency was set on the water line (Zuiderweg et al., 1986), and water suppression was achieved with low-power presaturation during the relaxation delay (2 s) between cycles and also during the mixing period in the NOESY experiments. HOHAHA experiments were performed with a spin-lock period of 100 ms (Bax & Davis, 1985; Bax, 1989). NOESY spectra were collected with mixing times of 100, 200, 300, 400, and 600 ms. Data were collected at 5, 35, and 55 °C in most experiments to help remove ambiguities due to spectral overlap, especially with the water line.

Two-dimensional data sets typically consisted of 512 t_1 increments, with 32 FIDs signal-averaged per increment and 1024 complex data points in the t_2 acquisition time domain.

[2] The sequence numbering scheme used here for Sac7d starts with the initiating methionine as residue 1 and differs from the native Sac7d which lacks the methionine and starts with valine as residue 1.

A spectral width of 6000 Hz was used in both dimensions. The data were zero-filled to 2048 points in both dimensions and apodized with a Lorentzian–Gaussian window function. Baseline corrections were applied using FELIX routines. $J_{NH\alpha}$ coupling constants were determined from DQF-COSY spectra with resolution increased by folding in D_2 using 256 t_1 increments with half the usual spectral width and acquiring 4096 complex data points in t_2 with zero-filling to 16 384 points. $J_{\alpha\beta}$ coupling constants were determined from PE-COSY spectra (Mueller, 1987) using a 35° β mixing pulse, 512 t_1 increments, 96 FIDs per increment, and 2048 complex data points in t_2 with spectral widths of 5000 Hz.

Chemical shifts are referenced relative to the appropriate shift of the water line which had been previously measured at 5, 35, and 55 °C relative to TSP in separate experiments without protein.

Spectral Analysis and Structure Determination. The NOESY spectrum collected with a 300 ms mixing time was chosen for structure determination and refinement to obtain the highest signal-to-noise ratio. Simulations using a rotational correlation time of 2.5 ns demonstrated that the NOE cross-peak intensities should reach a maximum at approximately 300 ms mixing time, and this was confirmed by experiment. Errors due to spin diffusion were removed in the final phase of the structural refinement using a full relaxation matrix analysis. The effects of spin diffusion were checked using back-calculations from the final structure (see Discussion). The volumes of all assigned NOE cross-peaks from the NOESY spectra collected at 35 °C were measured using the FELIX integration routines with a footprint width factor of 1.7.

A cyclical assignment/structural calculation procedure (Kraulis et al., 1989) was adopted in which NOESY cross-peaks with potential ambiguity due to degenerate resonances were resolved during the structural calculation process. Initial interproton distance restraints were estimated from 578 well-resolved and unambiguously assigned NOESY intensities using a two-spin approximation and the Tyr and Phe ring side-chain protons for distance calibration. Distance restraints involving methylene, propyl methyl groups, and aromatic protons were calculated with the variable pseudo-atom approach (Güntert et al., 1991a,b) using a program written by us for that purpose. A correction factor of ±0.4 Å was added to all distances involving methyl protons to account for the hindered rotation of methyl groups (Edmondson, 1993). An additional 0.5 Å was added to the upper distance restraints to account for error, flexibility, and spin diffusion. The lower bounds were set to 1.5 Å. The structure calculation program DIANA was used to generate 25 preliminary structures using the standard refinement strategy in DIANA (Güntert et al., 1991a,b). The structure with the lowest target function was examined and used to help assign additional cross-peaks, followed by another round of DIANA calculations and cross-peak assignments. β-Structure was clearly evident at this stage, and intraresidue NH–αCH cross-peaks of residues 5, 12, 14, 23, 24, 30, 32, and 44 involved in putative β-structure were used to define a reference volume associated with an average distance of 2.95 Å. New upper distance restraints were defined, and the standard DIANA strategy was modified to use 300 iterations per step instead of the default 150. After 100 DIANA structures were generated, the five structures with the lowest target function were examined for another assignment stage, and the procedure was repeated again.

Table 1: NOEs and Distance Restraints Used during Successive Stages of the Assignment/DIANA Structural Calculation Process[a]

stage	total NOEs	short range[b]	medium range	long range		distance restraints[d]
				backbone[c]	others	
1	578	264	32	29	253	394
2	634	278	33	32	291	449
3	696	319	33	33	311	648
4	748	348	33	33	334	740
final[e]	775	360	34	35	346	736

[a] Number of different types of NOEs assigned at different stages of the assignment/structure calculation procedure (see Materials and Methods). [b] Intraresidue and sequential NOEs. [c] NOEs involving backbone–backbone interactions. [d] Number of distance restraints considered useful by DIANA; includes pseudoatoms but not hydrogen bond restraints. [e] Final data set includes stereospecific assignments.

One hundred DIANA structures were then calculated using a set of restraints defined using an error of 0.25 Å (instead of 0.5 Å), 18 well-defined hydrogen bond restraints (from amide H-exchange data), 72 torsional restraints generated from $J_{NH\alpha}$ and $J_{\alpha\beta}$ coupling constants [from double-quantum-filtered COSY and PECOSY spectra using Karplus-type relations as determined by the program HABAS (Güntert et al., 1989)], and stereospecific assignments from intraresidue and sequential NOEs using HABAS. The 10 structures with the lowest target function were used to make additional stereospecific assignments using the program GLOMSA (Güntert et al., 1991a). These were then incorporated in a subsequent DIANA calculation of 200 structures, and the 10 structures with the lowest target function were selected for further study. To calculate an "average" DIANA structure, the 10 structures were superimposed and averaged, and the result was used as the starting structure for a final DIANA calculation. The resulting structure was then energy minimized with the all-atom AMBER force field (Weiner & Kollman, 1981; Weiner et al., 1984, 1986). The NOEs and distance restraints used during the sequential structural calculation/assignment process are summarized in Table 1.

The structure of Sac7d was refined by using an iterative relaxation matrix procedure using the program FIRM (Edmondson et al., 1991; Edmondson, 1992). FIRM refinement started with the final average DIANA structure. An isotropic correlation time of 2.8 ns was used for the FIRM calculations, as determined by fluorescence depolarization studies on Sac7d (Maurice Eftink, personal communication). A three-site jump model was used for methyl groups with a correlation time for methyl spinning of 50 fs (Edmondson, 1993). Effects of presaturation of the solvent via chemical exchange with amide protons were not included in the relaxation matrix due to the lack of quantitative measures of the individual amide exchange rates at this time. The resultant error is reflected in the final R-factor and is expected to be small due to the large number of amide protons involved in the hydrogen-bonded secondary structure. The experimental NOEs were scaled to the calculated values using backbone protons within well-defined regions (24 NOEs in the three-stranded β-sheet and α-helix). An error of ±5% (±2.5% in the later stages) was added to the back-calculated distances after the experimental and calculated NOEs were merged with FIRM. These were used as upper and lower distance restraints for DIANA. Lower distance

restraints were not allowed to exceed 4.5 Å. Hydrogen bonds were incorporated as restraints of 2 Å for NH−O and 3 Å for N−O distances, respectively, and torsional restraints were included as described in the DIANA manual. The structure with the lowest target function (out of 100 generated structures) after each cycle of FIRM/DIANA was subjected to 200 cycles of energy minimization with an NMR restraint weight of 1, followed by 200 cycles with a restraint weight of 10 using SANDER (see below).

The structures generated during the FIRM/DIANA refinement were checked for additional stereospecific assignments with GLOMSA (Güntert et al., 1991a), and four additional assignments were identified. The final number of stereospecifically assigned diasteric groups was 19 out of a total of 57 resolvable methylene groups, and 6 of the 8 resolvable propyl methyl groups were stereospecifically assigned. Several additional hydrogen bonds were also identified that had been observed as nonexchangable in D_2O but had not been assigned, bringing the number of identified hydrogen bonds to 26. (The number of observed backbone hydrogen bonds that were assigned in the final structure totaled 31, but 5 were not used during structure refinement because of possible ambiguities.)

The final structure from the FIRM/DIANA was refined with rMD (see below) and then with four cycles of FIRM/rMD. The total number of restraints (including pseudoatom distances, dihedral angles, and hydrogen bonds) in the final stages of refinement was 993. Structures obtained from rMD were energy minimized for 200 cycles with an NMR restraint weight of 2, followed by 200 cycles with a restraint weight of 10.

Restrained MD and energy minimization were performed with the SANDER module of AMBER 4.0 (Pearlman et al., 1991) with the NMR atomic parameter database containing reduced partial charges on ionizable side chains, a distance-dependent dielectric constant, a nonbonded cutoff of 8 Å, and the one to four interactions scaled down by a factor of 2. MD simulations were performed with SHAKE using a step size of 2 fs and a temperature scaling limit of 50° (ntt = 4). The limits of the flat potential wells used for distance restraints in SANDER were obtained by adding ±2.5% error (and any pseudoatom correction factor where appropriate) onto the FIRM back-calculated distances. The pseudoatoms were similar to those employed in DIANA and were implemented by coordinate-averaging the proton positions within SANDER. Parabolic sides of the flat wells extended for 1 Å with a force constant (RK) of 1 kcal/(mol·Å^2), beyond which they were linearly extrapolated. A well width of 0.0 Å was used for hydrogen bond restraints. The force constant (RK) for torsional restraints was 2 kcal/(mol·rad^2). Lower distance restraints were not allowed to exceed 4.5 Å, as for the DIANA calculations. The rMD protocol consisted of 200 cycles of restrained minimization followed by 5 ps rMD at 500 K. The NMR restraint weighting factor was increased from 1.0 to 2.0, and rMD continued for 2 ps at 500 K. Electrostatic contributions were reduced by a factor of 2 for the first 6 ps. The temperature was decreased to 100 K over the next 1 ps, and the resulting structure was subjected to 200 cycles of energy minimization with an NMR restraint weight of 2, followed by another 200 cycles with a weight of 10.

An *R*-factor analogous to that used in X-ray crystallography was defined as (Baleja et al., 1990; Nikonowicz et al., 1990; Gonzalez et al., 1991; Edmondson, 1992)

$$R = \frac{\sum_{\text{NOEs}} |\text{NOE}_{\text{obs}} - \text{NOE}_{\text{cal}}|}{\sum_{\text{NOEs}} |\text{NOE}_{\text{obs}}|} \quad (1)$$

where NOE_{obs} is the experimental NOE and NOE_{cal} is the corresponding back-calculated NOE intensity. *R*-factors were based on all measured cross-peaks used in the structure determination, or a subset consisting of all backbone (NH, CαH) cross-peaks, as specified in the text. Diagonal peaks were not included in the *R*-factor calculations. Random structures give an *R*-factor near 1, and a typical best fit structure has an NOE *R*-factor on the order of 0.2−0.4. Although an *R*-factor of 0.2 is respectable for X-ray-determined structures, the correspondence is not straightforward, and a quantitative comparison of *R*-factors should not be used to compare X-ray and NMR structures.

Chemical shifts were calculated with the DSHIFTS program (obtained from D. A. Case at http://www.scripps.edu/) as described by Osapay and Case (1991, 1994). The calculated shifts include contributions from ring current effects, electrostatic interactions, and magnetic anisotropy of the peptide group. Chemical shifts were referenced to the random coil values tabulated by Bundi and Wüthrich (1979) for tetrapeptides GGXA.

Structures were displayed and manipulated on a Silicon Graphics IRIS workstation using MidasPlus 2.0 software from the Computer Graphics Laboratory, University of California, San Francisco (Ferrin et al., 1988; Huang et al., 1992). Solvent-accessible surfaces, as defined by Richards (1977), were calculated using a 1.4 Å radius probe with the dms program distributed with MidasPlus 2.0.

RESULTS

Sequential Assignments. High-quality NMR spectra were obtained at relatively high protein concentrations and demonstrated a high signal-to-noise ratio with exceptionally wide chemical shift dispersion and resolution (Figure 1). A large number of αCH proton resonances shifted downfield of the water line at 4.7 ppm were immediately obvious and indicated extensive β-structure (Wishart et al., 1991). Data at 5, 35, and 55 °C were used to resolve ambiguities due to cross-peak overlap and also to locate resonances obscured by the water line.

^{1}H NMR resonance assignments were obtained using standard two-dimensional NMR procedures (Wüthrich, 1986). Amino acid spin systems were identified primarily from HOHAHA cross-peak patterns (Figure 2). The five Gly residues were readily identified from the multiplet patterns in HOHAHA and DQF-COSY spectra. One of the diastereotopic αCH_2 protons of two of the Gly residues (later shown to be Gly 41 and Gly 43) displayed unusual upfield chemical shifts, indicating significant ring current interactions (see below). The four Ala residues were easily located from their strong methyl HOHAHA cross-peaks. The single Ile, three Leu, and most of the seven Val residues were apparent from the double methyl resonances in the HOHAHA spectra. One of the three Thr residues was apparent from the methyl chemical shift, and another became obvious during sequential assignments. Thr 17 could not be distinguished until the

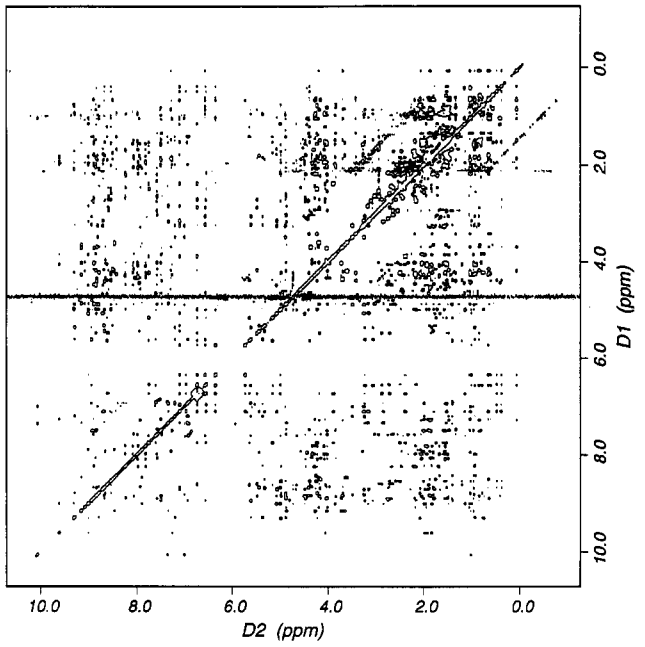

FIGURE 1: Two-dimensional 500 MHz NMR NOESY spectrum of recombinant Sac7d (5 mM Sac7d, pH 4.3, 35 °C, 90% H_2O/10% D_2O, 300 ms mixing time).

sequential assignments were nearly complete since its αCH was shifted upfield of the βH and the NH−γCH_3 cross-peak was missing in the HOHAHA spectrum. Pro 51 was located by the conspicuous absence of an NH cross-peak pattern in the HOHAHA spectra. The four Arg residues were found by ϵNH cross-peak patterns that mirrored the amide NH patterns in HOHAHA spectra. Most of the 14 Lys residues could be identified from the HOHAHA spectrum by their extended side chains and NH−ϵCH_2 cross-peaks between 2.9 and 3.2 ppm. The aromatic side chains (two Phe, two Tyr, one Trp) were readily assigned from the HOHAHA and DQF-COSY spectra and were correlated with their backbone and β-protons via strong NOEs. Except for Ser 46 and Ser 18 (which had characteristic low βCH_2 chemical shifts) most of the AMX residues (one Ser, one Asn, five Asp) and the seven Glu and three Met residues could not be distinguished until NOESY connectivities were established. Side-chain protons identified from the HOHAHA spectra were assigned primarily from DQF-COSY spectra. Many of the methylene groups had nonequivalent protons. The ϵCH_3 resonances of two methionines (Met 29 and Met 57) were identified by intraresidue NOEs. Met 1 ϵCH_3 was not observed.

Sequential connectivities were determined initially using the NOESY α-amide proton fingerprint region (Figure 3) in conjunction with the protein sequence (Choli et al., 1988a; McAfee et al., 1995), with the unique residues providing obvious starting points (e.g., Trp 24 and Pro 51). Some NOEs involving sequential NH-NH interactions were also apparent, especially for residues 52−64 which later proved to be α-helical. Sequential assignments were verified in most cases by examining NH−βCH_2 cross-peaks. A table of the assignments and chemical shifts is available as supporting information.

Secondary Structure. Extensive β secondary structure in Sac7d is indicated by strong sequential αCH−NH NOEs from residues 1 to 50 (Figure 4). Many of these α-protons also display downfield shifts characteristic of β-sheet (Wishart et al., 1991). Further support for β-structure is provided by antidiagonal patterns in an interresidue map of backbone NOEs (data not shown). The β-structures deduced from the interconnecting network of medium- and long-range NOEs consist of two independent β-regions (Figure 5). One is a two-stranded, antiparallel ribbon with an intervening tight hairpin turn composed of residues 8 through 11. The other consists of a three-stranded, antiparallel sheet with a tight hairpin turn (residues 26 through 29) between the first and second strands and a loop of six residues (residues 35 through 40) between the second and third strands. The majority of the amide protons involved in hydrogen bonds in the two β-structures exhibit slow exchange with water and are consistent with the hydrogen-bonded structure shown in Figure 5. The amide protons in the hairpin turn and the loop of the three-stranded β-sheet are not observed in the single long, exchange experiment used here, suggesting distortions and/or flexibility in these regions. More extensive proton exchange results will be presented elsewhere.

The two-stranded β-ribbon is amphipathic (Figure 6). One face of the β-ribbon is composed solely of charged residues, and the other face is significantly hydrophobic. There is also an asymmetry in the distribution of amino acid residues in the three-stranded β-sheet, but the amphipathic nature is not as pronounced.

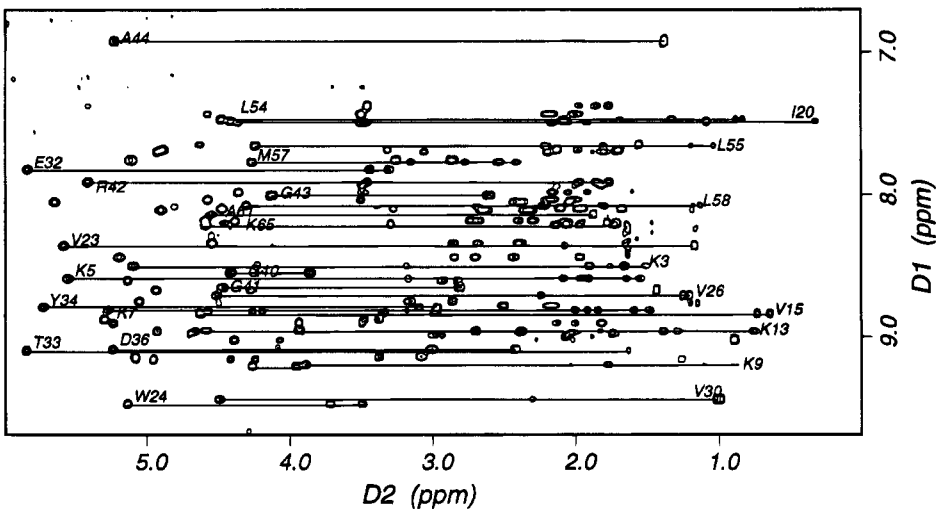

FIGURE 2: NH to αCH and side-chain proton correlations in the HOHAHA spectrum of recombinant Sac7d (conditions as in Figure 1, except at 55 °C). Through bond connectivities for selected residues are indicated by horizontal lines.

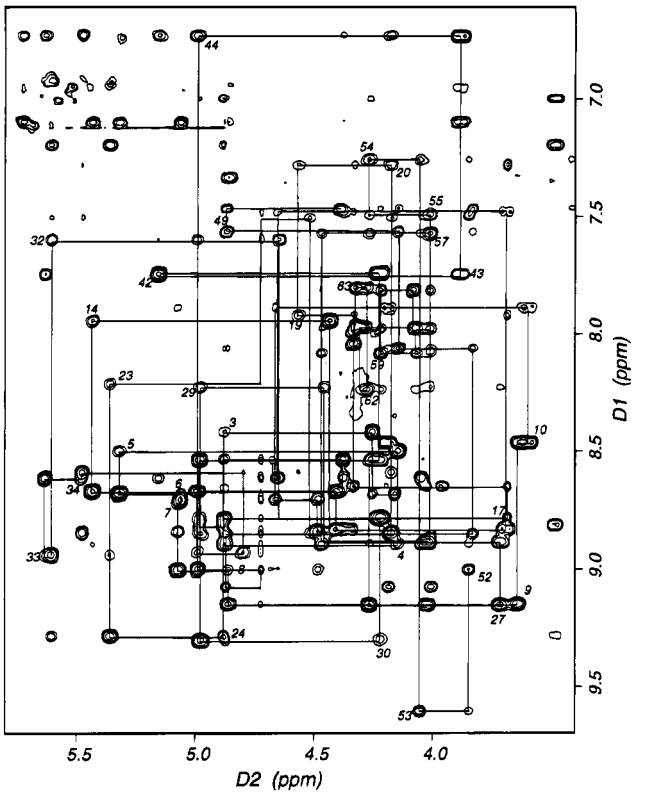

FIGURE 3: Sequential connectivities within the two-dimensional NOESY NH to αCH region of Sac7d (conditions as in Figure 1). Selected intraresidue cross-peaks are labeled.

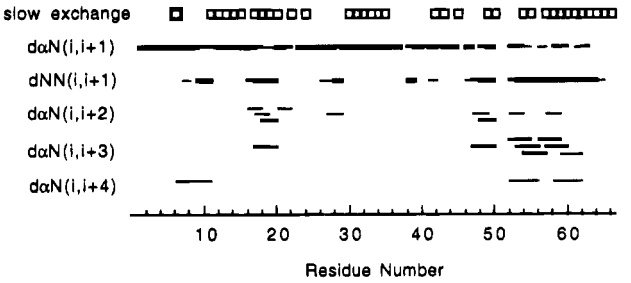

FIGURE 4: Summary of short- and medium-range NOEs and slowly exchanging amide protons. The relative strength of the NOE observed at 35 °C is represented by three line thicknesses (strong, medium, and weak). Slowly exchanging amide protons observed after 16 h in D_2O at 20 °C are indicated by a □.

The C-terminus of Sac7d exhibits weak sequential αCH−NH NOEs and strong sequential NH−NH interactions (Figure 4). Furthermore, there are a number of αCH−NH $i,i+3$ and $i,i+4$ NOEs indicating α-helical secondary structure. There are also many nonexchangable amide protons in this region. Folding the C-terminus into an α-helix results in an amphipathic helix with distinct hydrophobic and hydrophilic faces (Figure 6).

Starting Structures. Preliminary global folding patterns consistent with qualitative distance restraints derived from the observed NOEs were obtained by minimization of multiple random starting structures in torsional angle space using the structure calculation program DIANA (see Materials and Methods). The family of initial structures confirmed the presence of the three secondary structural elements. The root mean square deviations (RMSDs) of the backbone atoms for the 10 structures with the lowest target function were about 0.7 Å over much of the molecule (Figure 7A). The

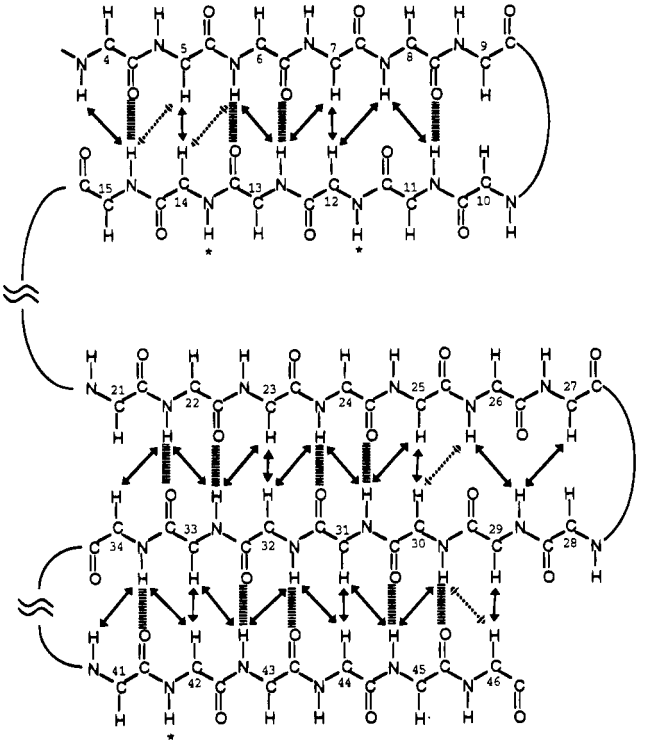

FIGURE 5: Schematic diagram of the β-structures in Sac7d showing interstrand NOEs and slowly exchanging protons. Solid arrows indicate observed NOEs, dotted arrows indicate possible NOEs that are obscured by overlap, and thick hashed lines represent hydrogen bonds inferred from slowly exchanging amide protons. Stars (*) indicate slowly exchanging amide protons that were not included as restraints in the structural calculations.

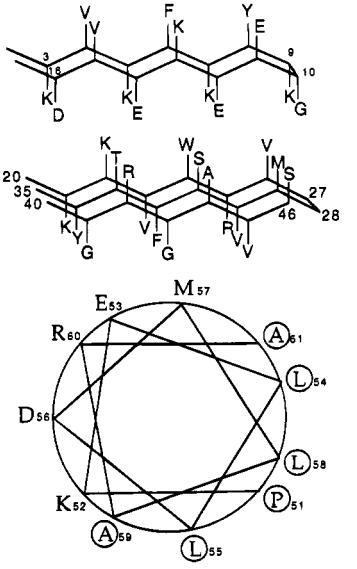

FIGURE 6: Schematic diagram of the secondary structural elements in Sac7d showing the amphipathic distribution of amino acid side chains in the antiparallel two-stranded β-ribbon, three-stranded β-sheet, and α-helix. The hydrophobic amino acids are circled in the helix wheel. Residues 62 through 66 are not shown since they define a hydrophilic turn extending beyond the amphipathic domain of the helix.

turns between strands of the β-structures were not as well-defined, possibly due to flexibility, with RMSDs approaching 3 Å. This was especially apparent in the hairpin of the two-stranded ribbon (residues 8−11) where there appeared to be considerable lack of definition extending into the adjacent strands. The turn between the first and second strands of

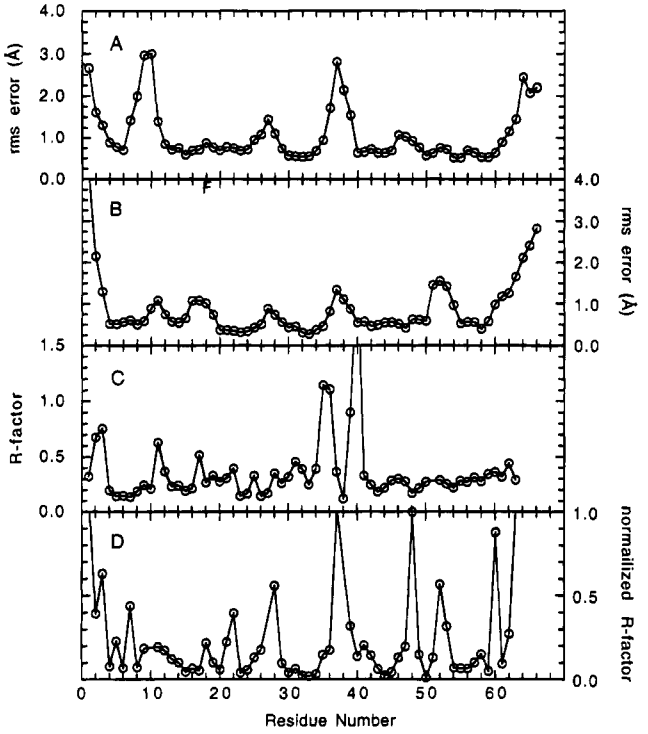

FIGURE 7: Sequence dependence of (A) RMSDs of backbone coordinates for the 10 best DIANA structures, (B) RMSDs of backbone coordinates for eight rMD structures using the restraints from the last FIRM refinement, (C) NOE R-factors of backbone protons for the final FIRM structure, and (D) normalized R-factors of side chains given as R-factor $\times$ n/N, where n is the number of observed resonances and N is the number of measured NOEs.

Table 2: Structural Parameters at Successive Stages of the Iterative Full Relaxation Matrix Refinement

refinement step	NOE R-factor[a]		energy[b]	RMSD[c]
	all atoms	backbone		
DIANA[d]	0.50	0.37	−480	1.89
FIRM/DIANA	0.46	0.30	−163	1.95
FIRM/DIANA	0.43	0.29	−176	1.86
FIRM/DIANA	0.43	0.32	−169	1.78
rMD	0.43	0.29	−510	1.51
FIRM/rMD	0.42	0.29	−484	1.18
FIRM/rMD	0.41	0.27	−493	1.77
FIRM/rMD	0.42	0.27	−516	0.75
FIRM/rMD	0.40	0.25	−537	0.74
rMD[e]	0.40 ± 0.00	0.27 ± 0.02	−550 ± 20	0.75 ± 0.14

[a] NOE R-factors for all experimentally observed protons (all atoms) and only those among αCH, NH backbone protons (backbone). [b] Potential energy in kcal/mol calculated without NMR restraints using AMBER 4.0 with the NMR atomic parameter database. [c] RMSDs of backbone coordinates over residues 4−62 relative to the average of eight rMD structures following the last FIRM refinement. [d] Average DIANA structure before the FIRM refinement procedure was started. [e] Mean ±1 SD of eight rMD calculations using restraints from the last FIRM cycle.

the three-stranded sheet (residues 26−29) showed less variation with an RMSD of about 1.2. The segments linking the secondary structural elements were surprisingly ordered with RMSDs similar to those found within the β-structure and α-helical regions.

The NOESY spectrum back-calculated from the average DIANA structure was in surprisingly good agreement with the experimental NOESY spectrum, giving an NOE R-factor of 0.50 for all protons and 0.37 for backbone protons (Table 2). For comparison, random structures give an NOE R-factor of approximately 1.0, and accurate structures can be expected to have R-factors of 0.3 to 0.4 (Edmondson, 1992).

Full Relaxation Matrix Refinement. The fit between experimental and calculated NOESY spectra was improved with an iterative relaxation matrix refinement using FIRM (see Materials and Methods). After seven FIRM cycles, the overall NOE R-factor decreased to a mean value of 0.40, and the backbone NOE R-factor decreased from 0.37 to 0.27 (Table 2). The backbone RMSDs between the last two FIRM cycles (0.75 Å) were identical to those for multiple rMD simulations, indicating convergence. There was no further improvement in RMSDs or R-factors with additional FIRM cycles. The potential energies of some of the FIRM structures tended to be slightly greater than that of the starting DIANA structure due to the relatively strong emphasis [NMR restraint weight of 10 kcal/(mol·Å^2)] placed on the experimental data, mostly at the expense of bond angle distortions.

A common method of estimating error in NMR structure determination is to examine the RMSDs between various structures obtained by applying the same restraints to different starting structures. This method does not define the accuracy or precision of the structure but examines the uniqueness of the structures and the limitations of the data (Shriver & Edmondson, 1993, 1994). We have repeated the rMD calculation following the last FIRM calculation eight times with different random number seeds and examined the spread among structures (see Figure 8). R-factors ranged from 0.25 to 0.30 using backbone NOEs and from 0.40 to 0.41 using all NOEs. The average RMSD for backbone atoms was 0.75 Å and for all heavy atoms was 1.1 Å. Over much of the sequence, the backbone atom RMSDs were on the order of 0.5 Å (see Figure 7B). As found for the two-spin DIANA calculations, RMSDs were greater in the hairpin turns and loop. However, these regions appeared much better determined after FIRM refinement compared to the DIANA structures. In contrast, the FIRM structures had greater RMSDs in the connecting element between the β-ribbon and sheet structures (residues 16−19) and around the N-terminus of the α-helix (residues 51−54). A quantitative analysis of the error in the Sac7d structure using the Monte Carlo method is in progress and will be presented elsewhere. The eight FIRM/rMD structures were averaged and energy-minimized, and this average structure is presented in this study as the final FIRM-refined structure and best representation of the solution structure. The NOE R-factor for this energy-minimized average structure was 0.28 for the backbone and 0.41 for all NOEs. For comparison with the energies listed in Table 2, the potential energy of the average structure, using the same force field parameters, was −696 kcal/mol.

The R-factors for backbone protons (NH/αCH) were about 0.25 in the α and β secondary structures and in the connecting elements between secondary structures (see Figure 7C). The backbone regions with the greatest NOE deviations were those around the N-terminus and the loop between the second and third strands of the three-stranded sheet. Large R-factors in the vicinity of the loop, similar in magnitude to those expected for random structures, indicate a poorly fit structure in this region even though the RMSDs are about only 1.0 Å (Figure 7B). This demonstrates the fallacy of using structural variations among multiple fits of the same data set to evaluate the precision of structure determination. A significant deviation between observed and

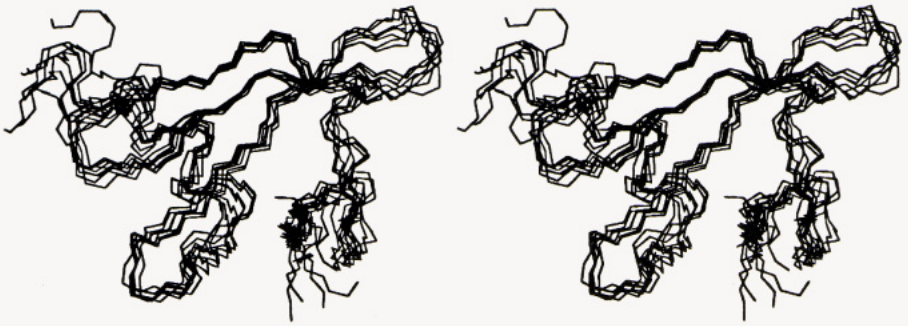

FIGURE 8: Stereoview showing a superposition of the backbone atoms of residues 4 through 62 of structures obtained from eight rMD simulations using the restraints from the last FIRM cycle. The C-terminus is shown at the upper left.

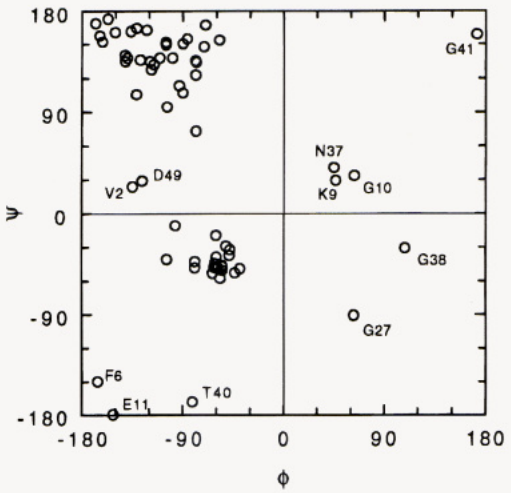

FIGURE 9: Ramachandran ϕ,ψ plot for the final FIRM structure. Selected points are labeled with the corresponding residue.

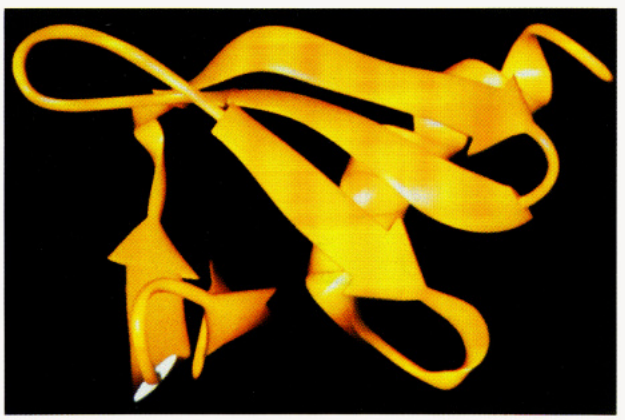

FIGURE 10: Ribbon representation of the final FIRM-refined solution structure of Sac7d illustrating the secondary structural elements and their packing. The view is from the side opposite that in Figure 8 with the C-terminus at the upper right. The view demonstrates the packing of the β-ribbon against a corner of the twisted β-sheet composed of the C-terminus of strand 2 and the N-terminus of strand 3.

calculated NOE intensities was also seen for Glu 11 in the first hairpin.

Not surprisingly, *R*-factors for the side chains (average of about 0.4) were greater than those for the backbone and exhibited significant scatter. Nonetheless, many side-chain protons contributed numerous NOEs, and some of their positions appear fairly well-defined. To qualitatively describe these side-chain interactions, we have computed a "normalized *R*-factor" as a product of the NOE *R*-factor and the ratio n/N, where n is the number of resolved side-chain resonances and N is the number of NOEs involving these side chains. As seen in Figure 7D, *side chains* whose positions are poorly determined by the experimental data include the three N-terminal residues, side chains around each of the β-turns (residues 9−12, 26−28, and 36−39), most charged side chains (e.g., Lys 5, Lys 7, Lys 21, Lys 22, Lys 48, Lys 52, Arg 60), and the four C-terminal residues. Side chains that appear to be well-defined by the NMR data (that is, those with low *R*-factors and many NOEs) are the hydrophobic residues in the β-ribbon (Val 4, Phe 6, Tyr 8, Val 15), residues 16, 17, 20 in the first connecting element, numerous side chains of the β-sheet (including some on the outside surface), and several residues around the α-helix (residues 50, 54−59). Note that most of the ill-defined side chains are polar or charged and that most of the hydrophobics appear well-defined.

All of the ϕ and ψ angles for the final FIRM structure are within allowed ranges, but several residues have notable torsional angles (Figure 9). These include Thr 40 and Gly 41 adjacent to the loop, Asn 37 which is in a left-handed helical conformation in the loop (not uncommon for Asn), and residues around the β-hairpin of the two-stranded ribbon (Phe 6, Lys 9, Gly 10, Glu 11). The distortion of backbone torsional angles around this hairpin reflects the twisted and curled nature of the β-ribbon (see below).

Description of Final Structure. A cartoon of the final refined structure is shown in ribbon form in Figure 10, and a detailed picture including side chains is shown in Figure 11. The molecule is globular with dimensions of approximately 20 × 20 × 30 Å and a surface area of 4050 Å^2. Considering the size of the protein, an unusually large portion of the sequence is involved in repetitive secondary structure (82%): 24 residues (36%) in β-sheet, 8 residues (12%) in β-turns, 15 residues (23%) in α-helix, and 7 residues (11%) in 3_{10} helix. The α-helix and two β-structures are oriented approximately orthogonally to each other and wrap around a cylindrical core of hydrophobic side chains. The global fold is dominated by the packing of the amphipathic helix onto the hydrophobic face of the three-stranded β-sheet, with the two-stranded ribbon positioned over a corner of the β-sheet. The final turn of the α-helix extends beyond the face of the three-stranded sheet and becomes significantly exposed at the carboxy terminus.

The β-ribbon structure in the amino terminus has four backbone hydrogen bonds observed by hydrogen exchange, including one in the hairpin between residues $n,n+4$ (see Figure 12). The hairpin, which has Gly at position 3, is similar to a left-handed 3_{10} helix and is probably best classified as a type III′ turn. One face of the ribbon consists mainly of hydrophobic side chains, resulting in a strongly

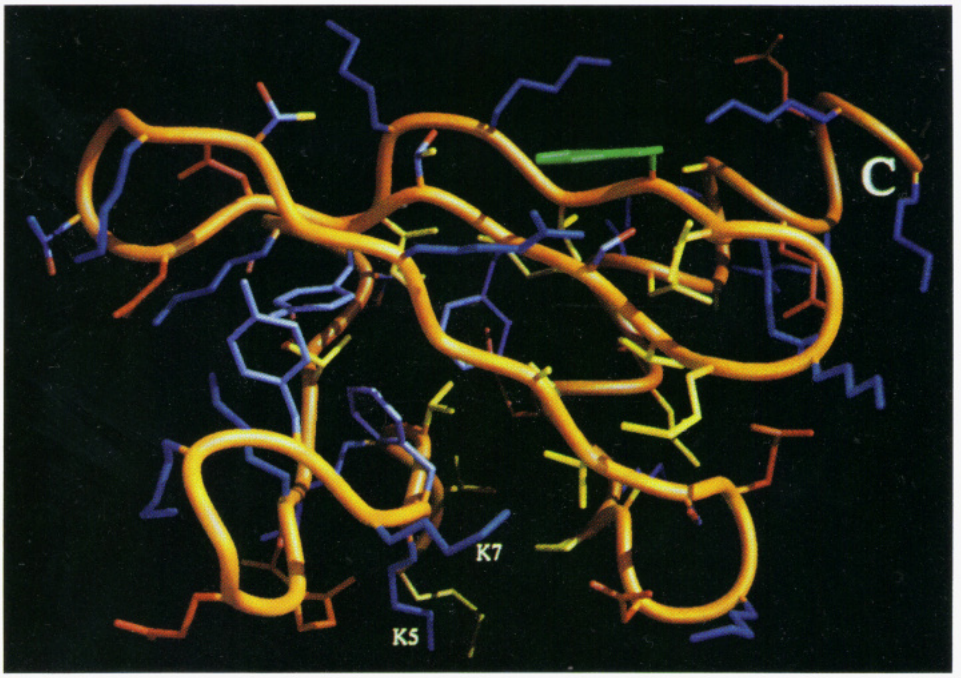

FIGURE 11: Final FIRM-refined solution structure of Sac7d showing all heavy atoms. The view is similar to that in Figure 10. The two lysines that are monomethylated in the native protein are indicated (K5 and K7). The side chains are color coded as follows: blue, Lys and Arg; red, Asp and Glu; yellow, aliphatic; blue-gray, Phe and Tyr; green, Trp.

curled and twisted β-ribbon. There are numerous NOEs to these hydrophobic side chains, as well as to Lys 13 which is also on the inner face of the ribbon. The charged residues on the outer surface of the ribbon are probably rather disordered, as indicated by their larger normalized *R*-factors (Figure 7D).

A short, single turn of 3_{10} helix extending from Asp 16 to Lys 19 connects the two β-structures. The 3_{10} helix assignment is defined by the algorithm of Kabsch and Sanders (1983), ϕ/ψ angles, and inspection of the positions of the Cα atoms (Richardson & Richardson, 1989). The structure is stabilized by two $i,i+3$ backbone hydrogen bonds from Thr 17 to Ile 20 and Asp 16 to Lys 19. The carboxyl side chain of Asp 16, which appears well determined, forms an additional hydrogen bond with the backbone NH of Ser 18, further stabilizing the 3_{10} helix. A hydrogen bond is also observed between the carbonyl of Lys 19 and the NH of Asp 35, linking the 3_{10} helix with the β-sheet (see Figure 12).

The three-stranded β-sheet has nine observable interstrand hydrogen bonds (see Figure 12). The first hairpin of this sheet is a tight inverse glycine (type II′) turn (Richardson & Richardson, 1989) with Gly at position 2, but no $n,n+3$ hydrogen bond was observed in this hairpin by NMR after 16 h in D_2O. Nonetheless, the hairpin is fit well by the experimental NOE data (Figure 7). In contrast, the loop between the second and third strands is fit poorly by the experimental NOE data and may be flexible in solution, although the structural variation in multiple FIRM/rMD calculations was not large. The side chains of the β-sheet are fairly well positioned, including those on the outer surface, such as Trp 24 which is anchored by numerous NOEs. Independent fluorescence depolarization studies also indicate that the Trp is exposed and rigid with respect to the rest of the molecule (Maurice Eftink, personal communication).

The connection between the three-stranded β-sheet and the α-helix is comprised of another single turn of 3_{10} helix extending from Glu 47 to Asp 49 (defined using the same criteria as listed above). This 3_{10} helix is stabilized by two $i,i+3$ hydrogen bonds between Ser 46−Asp 49 and Glu 47−Ala 50 (Figure 12).

The carboxy-terminal α-helix starts after Pro 51, which contains a *trans* peptide bond. The carbonyl of Pro 51 forms a bifurcated hydrogen bond with the NH of residues $i+3$ and $i+4$ (Leu 54 and Leu 55, respectively). Amide exchange data indicates 11 hydrogen bonds in the α-helix extending out to the NH of terminal Lys 66 (see Figure 12). The helix is slightly bent in the direction of the molecule's long axis. Most of the charged side chains of the α-helix are solvent exposed and not well-defined. In contrast, the aliphatic side chains of the α-helix are very well defined within the hydrophobic core.

The final turn of the α-helix (Glu 62, Arg 63, Glu 64, Lys 65, Lys 66) is not amphipathic and extends beyond the hydrophobic face of the three-stranded sheet. Glu 62, the first nonhydrophobic residue on the inside face of the helix, is positioned to form a potential salt bridge with Arg 25 in the β-sheet, further stabilizing the helix and its packing against the β-sheet. Although few NOEs are resolved for the four terminal residues, the hydrogen bonds inferred from the H-exchange experiment indicate that the helical structure persists to the terminus, possibly stabilized by the positively charged residues interacting with the helix dipole.

The relative positioning of the main secondary structures (two-stranded β-ribbon, three-stranded β-sheet, and four-turn α-helix) is determined by numerous NOEs both between the β-sheet and α-helix and between the β-sheet and β-ribbon. Helix−sheet interactions are dominated by NOEs involving the aliphatic side-chain protons of Leu 54, Met 57, Leu 58, and Ala 61 of the α-helix and core residues Val 23, Val 30, and Phe 32 of the β-sheet. There are also extensive NOEs

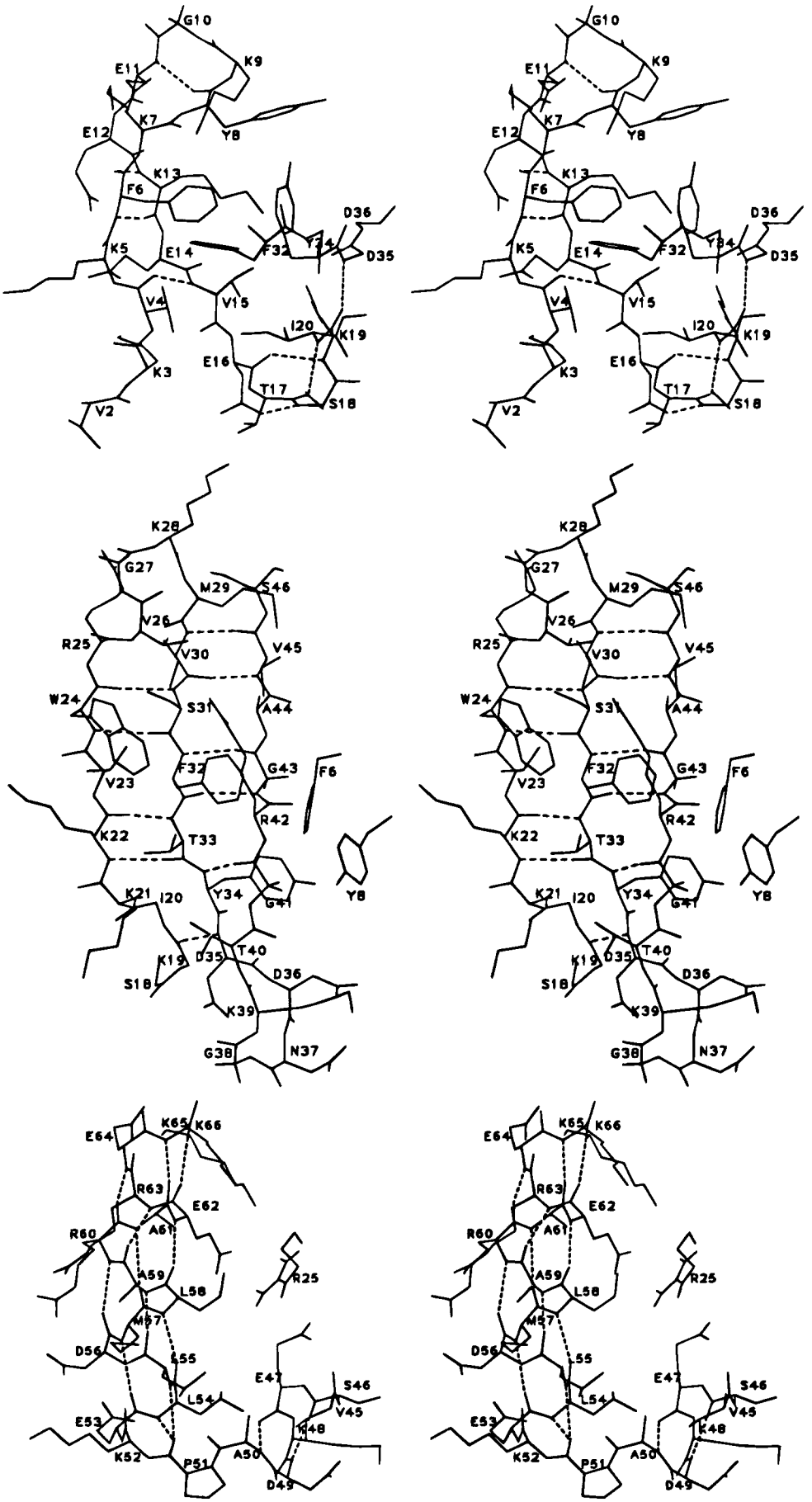

FIGURE 12: Stereo drawings of heavy atom positions for various regions of the final solution structure of Sac7d: (top) N-terminal β-ribbon and adjacent 3_{10} helix; (middle) three-stranded β-sheet; (bottom) 3_{10} helix and C-terminal α-helix. Hydrogen bonds with slowly exchanging amide protons are indicated with dashed lines.

Table 3: Secondary Structural Elements and Solvent Accessibilities[a]

residue	accessibility	residue	accessibility
nonrepetitive		loop	
Met 1	97	Asp 35	61
Val 2	80	Asp 36	62
Lys 3	67	Asn 37	80
β-ribbon strand 1		Gly 38	100
Val 4	**10**[b]	Lys 39	78
Lys 5	98	Thr 40	81
Phe 6	**2**	β-sheet strand 3	
Lys 7	100	Gly 41	80
β-turn		Arg 42	80
Tyr 8	61	Gly 43	50
Lys 9	62	Ala 44	49
Gly 10	100	Val 45	24
Glu 11	83	Ser 46	69
β-ribbon strand 2		connecting element	
Glu 12	76	with 3_{10} helix	
Lys 13	36	Glu 47	68
Glu 14	62	Lys 48	81
Val 15	**0**	Asp 49	68
connecting element		nonrepetitive	
with 3_{10} helix		**Ala 50**	**0**
Asp 16	78	Pro 51	97
Thr 17	85	α-helix	
Ser 18	78	Lys 52	100
Lys 19	63	Glu 53	49
nonrepetitive		**Leu 54**	**5**[b]
Ile 20	**14**[b]	Leu 55	70
β-sheet strand 1		Asp 56	73
Lys 21	76	Met 57	53[b]
Lys 22	78	**Leu 58**	**9**
Val 23	22[b]	Ala 59	54
Trp 24	52	Arg 60	76
Arg 25	53	Ala 61	46
β-turn		Glu 62	50
Val 26	64	Arg 63	80
Gly 27	100	Glu 64	84
Lys 28	100	Lys 65	71
Met 29	68	Lys 66	83
β-sheet strand 2			
Val 30	**0**		
Ser 31	0		
Phe 32	**0**[b]		
Thr 33	47		
Tyr 34	**6**		

[a] Percent solvent accessibility of side chains relative to the free amino acids. Core residues with solvent accessibilities less than 15% are indicated in bold type. [b] Does not include surface area of the internal cavity.

between the β-ribbon and strands 2 and 3 of the β-sheet (there are none between the ribbon and the first strand of the sheet). There were few unambiguously resolved NOEs between the α-helix and β-ribbon, and only those involving Phe 6 and Leu 54 were used for structure calculations. Residues flanking the β-ribbon (Val 2, Thr 17, Ile 20) have NOEs to the N-terminal region of the helix, and Val 4 has NOEs to Pro 51 preceding the helix.

All polar, un-ionized residues are located on the surface of the protein, except for Ser 31 which is shielded by the side chains of Val 26, Trp 24, and Arg 42 on the outer surface of the β-sheet. The two lysine residues that are partially monomethylated in native Sac7d (Lys 5 and Lys 7) are located along the inner edge of the two-stranded β-ribbon (see Figure 11). There are no internal salt bridges. All but one of the ionizable residues, including 13 lysines, 4 arginines, 5 aspartates, and 7 glutamates, are also solvent exposed on the surface (Table 3). Lys 13 is the exception with only 36% solvent accessibility.

The core of Sac7d is composed of 10 residues with less than 15% solvent accessibility (see Table 3). As seen in Figures 11 and 13, the hydrophobic core consists of aliphatic and aromatic regions. The aliphatic side chains are mainly packed between the α-helix and β-structures, with numerous NOEs defining the placement of the helix relative to the three-stranded sheet (e.g., from Leu 54 to Ile 20, Val 23, and Val 30; Met 57 to Thr 17, Ile 20, and Val 23; Leu 58 to Trp 24, Arg 25, and Val 30; Ala 61 to Val 23). The aromatic side chains are stacked in a herringbone pattern between the two β-structures. One face of Tyr 8 is solvent exposed, giving a 61% solvent accessibility for the side chain.

A number of exposed hydrophobic residues are observed. Particular notice should be paid to Val 2 which has a solvent accessibility of 80%, although there are few NOEs to the amino terminus. The hydrophobicity of this residue appears to be essential: it is either a valine or an alanine in all of the Sac7 and Sso7 proteins sequenced. Other relatively exposed hydrophobes include Pro 51 (97% accessibility), Met 29 (68%), Val 26 (64%), and Ala 44 (49%). Leu 55 (70%), Met 57 (53%), and Ala 59 (54%) are somewhat exposed in the helix; they also appear well determined and are probably fairly rigid in the solution structure. The single tryptophan (Trp 24) is located on the surface of the molecule and, as described above, appears to be rigidly placed.

There is a well-defined 5.1 Å diameter cavity between the β-sheet and the α-helix surrounded by side chains of Val 4, Ile 20, Val 23, Phe 32, Leu 54, and Met 57. The

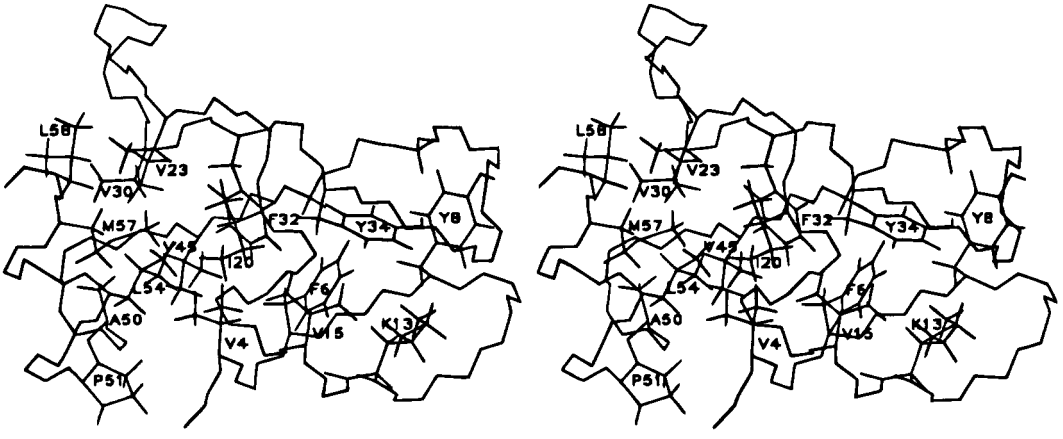

FIGURE 13: Stereoview of the side-chain packing within the hydrophobic core of Sac7d showing the packing of the α-helix against the three-stranded sheet and also the packing of the two-stranded ribbon against the second and third strand of the β-sheet.

cavity could be accessible to solvent by displacement of the Met 57 side chain, but none of the NMR structures appeared to allow such motions. There are no polar groups in the vicinity of this cavity other than Met 57 whose lone-pair electrons are oriented toward the solvent.

Two of the polar side chains had numerous NOEs and appeared to be relatively well-defined. The aliphatic portion of the side chain of Lys 13, which lies on the hydrophobic side of the β-ribbon, is solvent inaccessible and contributes to the core of the aromatic subdomain. The terminal amino group of the Lys 13 side chain appears to be exposed on the surface. The guanidino group of Arg 25 is positioned to form a potential double salt bridge with the side chains of Glu 47 and Glu 62 and to form potential hydrogen bonds with the backbone carbonyl of either Gly 27 or Lys 28 (see Figure 12). Due to its central location, Arg 25 may help stabilize interactions between the β-sheet, α-helix, and 3_{10} helix.

Backbone NH protons of Glu 12, Glu 14, Thr 17, and Arg 42 exhibit weak NMR signals after 16 h in D_2O. All of these residues, except for Thr 17, are in the β-structure and have their NH pointing to the outside (Figure 5). The exchange behavior of Glu 12 and Glu 14 amide protons may be explained by hydrogen bonds with side-chain carboxyls. The NH of Arg 42 is situated between Tyr 8 and Tyr 34 and may be shielded from interaction with solvent by these ring systems (see Figure 12). Thr 17 flanks the β-ribbon and could, with some minor rearrangements of the amino-terminal (Val 2, Lys 3) backbone (which is not very well determined), form an hydrogen bond with the carbonyl of Val 2. Interestingly, this would extend the β-ribbon (and its hydrophobic face) to the N-terminus, which might explain the strict conservation for a hydrophobe in this position for all Sac7 and Sso7 proteins.

Several protons exhibit unusual chemical shifts that can be explained primarily by ring current interactions and provide an opportunity to independently assess the reliability of the final structure. One of the αCH's of both Gly 41 and Gly 43 is adjacent to the aromatic rings of Tyr 34 and Phe 32, respectively, resulting in dramatic upfield shifts for these protons. Chemical shift calculations predict a large upfield shift for one of the αCH's of Gly 43, but a small chemical shift is calculated for Gly 41 (Table 4). The close proximity of Gly 41 to Tyr 34 suggests that minor refinement might explain the observed upfield shift. The βCH_2 protons of Ser 31 are adjacent to Trp 24, and the observed large upfield shifts are predicted by the chemical shift calculations. The side chains of Lys 9 and Lys 13 are in close proximity to the aromatic rings of Tyr 8 and Phe 6, respectively, and the chemical shifts calculated from the NMR structures are in good agreement with the observed shifts. Protons with unusual chemical shifts which are not predicted by the calculations include the NH of Ala 44 (shifted 1.5 ppm upfield) and the NH of Glu 53 (shifted 1.2 ppm downfield).

Table 4: Comparison of Observed and Predicted Chemical Shifts[a]

		Δ chemical shift (ppm)		
residue	proton	obsd	calcd[b]	errors[c]
Lys 9	αH	−0.72	−0.66 ● 0.08	0.30
	βH_2	−0.14	0.03 ± 0.13	0.20
	γH_2	−0.89	−0.31 ± 0.05	0.20
	δH_2	−0.87	−0.09 ± 0.03	0.20
Lys 13	αH	0.07	−0.20 ± 0.14	0.30
	βH_2	−1.38	−1.13 ± 0.16	0.20
	γH_2	−0.42	−0.31 ± 0.10	0.20
	δH_2	−0.55	−0.69 ± 0.64	0.20
Ile 20	δCH_3	−0.82	−0.47 ± 0.18	0.14
Ser 31	βH_2	−1.62	−1.28 ± 0.22	0.20
Gly 41	αH	0.26	−0.17 ± 0.20	0.30
	αH′	−1.48	−0.53 ± 0.61	0.30
Gly 43	αH	−0.08	−0.14 ± 0.40	0.30
	αH′	−1.59	−2.16 ● 0.48	0.30

[a] Change in chemical shifts relative to the random coil values. Except for glycines, the chemical shifts for diasteric protons of methylene groups were averaged. [b] Calculated values are the mean of results from the final eight rMD structures (±1 SD). [c] Estimated rms errors for chemical shift calculations [from Table 3 in Osapay and Case (1991)].

DISCUSSION

Overall Structure Quality. The structure of homogeneous, recombinant Sac7d protein has been determined from initial structures obtained from qualitative distance restraints followed by a direct fit of the NMR data using a full relaxation matrix refinement. The solution structure is well-defined by the NMR data, including 775 NOEs (i.e., 12 per residue), 72 dihedral angle restraints, and 26 hydrogen bonds inferred from H-exchange experiments. The precision with which the Sac7d structure is defined is currently being addressed by Monte Carlo techniques and will be presented elsewhere. Except for the loop between the second and third strands of the three-stranded β-sheet, back-calculated backbone NOEs are in excellent agreement with experimental data and show an NMR *R*-factor of 0.27. The backbone regions with the greatest NOE deviations were those around the N-terminus and the loop between the second and third strands of the three-stranded sheet. The "imprecision" in the loop was not apparent in overlays of multiple refined structures. Overlays of multiple rMD structures using the full relaxation matrix refinement restraints suggest that the positions of the backbone atoms are known to within 0.5 Å over much of the structure.

The global fold of the protein is largely dominated by the packing of the amphipathic helix onto the hydrophobic face of the three-stranded sheet in an open-face sandwich arrangement (Figure 10). The positioning of the α-helix across the three-stranded sheet is well-defined, with numerous NOEs extending across the hydrophobic interface. In contrast, few NOEs are observed between the α-helix and β-ribbon; rather, the orientation of the ribbon is determined almost exclusively by the many NOEs with strands 2 and 3 of the β-sheet.

The positioning of the two-stranded β-ribbon relative to the three-stranded sheet is interesting and seems to be largely dictated by an attempt to accommodate the packing of the amphipathic helix and still reduce the exposure of the two phenylalanines and two tyrosines (Figure 13). The ribbon is packed against a corner of the sheet involving a portion of strands 2 and 3 only, with the interface predominantly between Phe 6, Tyr 8, Lys 13, and Val 15 on one side of the β-ribbon and between Phe 32, Tyr 34, and Gly 43 on one face of the β-sheet. The length of the connecting element between the two β-structures displaces the β-ribbon toward an edge of the β-sheet. There is significant curling of the ribbon around the outer edge of the third strand of the sheet. The interaction of the ribbon over the third strand is permitted by packing the Phe 32 and Tyr 34 side chains

from the center strand of the sheet over the glycines at positions 41 and 43. Such packing explains the large upfield shifts observed and calculated for the αCH protons of these two glycines (Table 4).

Charged and polar side chains located in connecting elements or on the exterior surfaces of the secondary structural elements are not as well-defined as the core side chains. However, Lys 13, Trp 24, and Arg 25 contribute numerous NOEs and appear to be relatively well-defined. Lys 13 is placed to utilize the hydrophobic character of the methylenes. Arg 25 appears to be positioned to link the β-sheet, α-helix, and interconnecting 3_{10} helix through electrostatic interactions and hydrogen bonding. Although Trp 24 is exposed on the surface of the three-stranded sheet, it appears to be largely immobilized, consistent with fluorescence depolarization measurements of the rotational correlation time of the protein.

Spin diffusion was accounted for in the final stages of refinement using a full relaxation matrix analysis. Back-calculations using the final structure can be used to investigate the magnitude of this effect. Selected NOE intensities as a function of mixing time are shown in Figure 14 and compared to that expected for two isolated spins at various distances from 2.5 to 7.5 Å. Spin diffusion in the protein leads to a relative loss of intensity (compared to that predicted for two isolated spins) for short- and medium-range interactions with increasing mixing time and no unexpected buildup of long-range interactions. For example, the NH protons of residues 54 and 55 are separated by 2.40 Å in the final structure, but the NOE intensity at 300 ms would indicate a distance of approximately 2.9 Å if a full relaxation matrix analysis were not performed and the analysis were limited to qualitative restraints derived from a two-spin approximation. In comparison, the αCH of residue 31 is closer to the NH of residue 32 (2.34 Å), although the buildup rate is significantly less than that observed for the NH 54 to NH 55 interaction. The resulting distance would be in error even at short mixing times if only a two-spin approximation were used. Very long range interactions [e.g., between a β-proton of Leu 32 in the β-sheet and a methyl proton in Met 57 in the α-helix (6.48 Å) or between the Hα of Leu 32 and the Hα of Met 57 (10.51 Å)] remain unobservable even at 500 ms mixing times, although intervening protons lead to considerable spin diffusion.

Comparison to Sso7d. The structure of the recombinant Sac7d protein presented here differs from the recently published NMR structure of native Sso7d (PDB entry 1SSO; Baumann et al., 1994) with an overall Cα RMSD of 3.2 Å (overlaying Cα's of residues 4−34 and 41−62 in Sac7d with residues 3−33 and 41−62 in Sso7d). Sso7d has been described as a β-sandwich with the dominant tertiary packing motif being the packing of the two-stranded β-ribbon onto the three-stranded β-sheet. The hydrophobic core is defined by the residues at the interface of the two β-structures. In addition, the 3_{10} helices described here for Sac7d are single turns of α-helix in the Sso7 structure. The most pronounced difference is the packing and length of the extended C-terminal α-helix in Sac7d compared to the largely unstructured C-terminus of Sso7d. No slowly exchanging amide protons were observed outside of the β-structures in Sso7d. Therefore, the C-terminus was not well-defined and received little discussion in Baumann et al. (1994). Residues 52 through 59 appeared to form two turns of an α-helix which packed onto the edge of the hydrophobic core of the β-sheet sandwich.

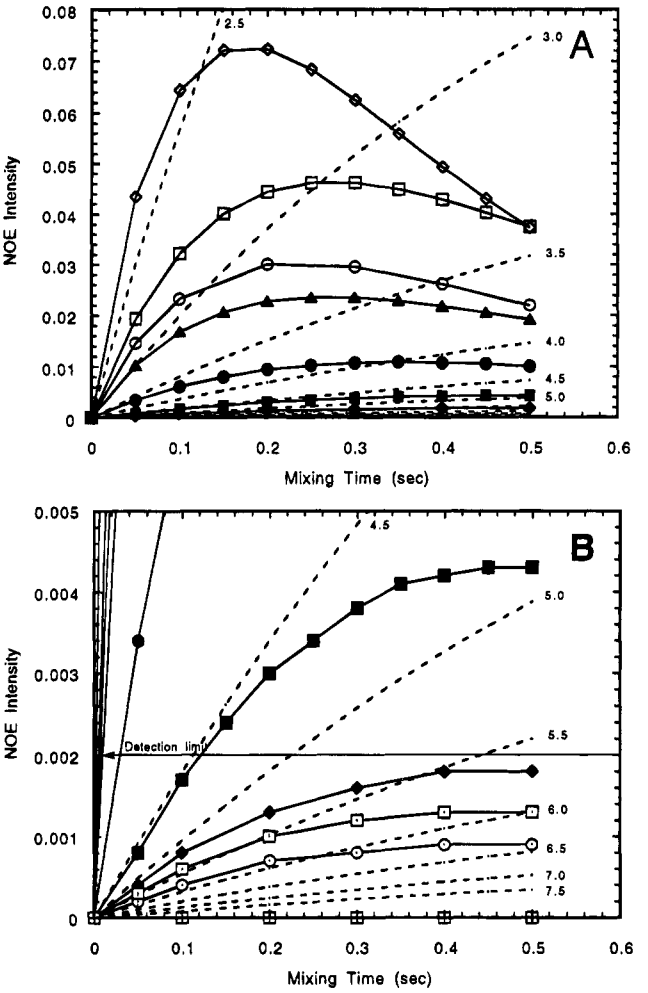

FIGURE 14: Back-calculated NOE intensities as a function of mixing time demonstrating the effects of spin diffusion. Back-calculations were performed with FIRM using the final structure of Sac7d reported here with an overall isotropic rotational correlation time of 2.8×10^{-9} s, a methyl rotor correlation time of 5×10^{-11} s with a three-site jumping model, and a 0.5 s additional leakage rate. (A) Short- and medium-range interactions: NH 54 to NH 55 (2.40 Å in the final structure, open diamonds), Hα 31 to NH 32 (2.34 Å, open squares), NH 57 to NH 58 (2.79 Å, open circles), Hα 57 to NH 57 (2.89 Å, closed diamonds), Hα 54 to NH 55 (3.52 Å, closed circles), and NH 31 to NH 32 (4.45 Å, closed squares). (B) Long-range interactions: NH 31 to Hα 32 (4.97 Å, closed diamonds), Hβ 20 to Hβ 23 (5.73 Å, open squares with internal dot), Hβ 32 to Hε 57 (6.48 Å, open circles with internal dot), Hα 32 to Hα 57 (10.51 Å, open squares with internal +). For comparison, the NOE intensities for two isolated spins at various indicated distances are shown with an overlaid grid of dotted lines. The detection limit is indicated at 0.002 in panel B.

Baumann et al. (1994) suggested that the packing of the β-sheets in Sso7d was similar to that of the eukaryotic SH3 domain, a β-sandwich consisting of orthogonal packing of two three-stranded β-sheets and no helix. The RMSD for the Cα atoms involved in the β-structure was reported to be 2 Å. We find Cα RMSDs for Sac7d similarly aligned to the SH3 domain of chicken brain α-spectrin (PDB entry 1SHG) of about 3.7 Å if we ignore the presence of the α-helix. This alignment does not justify a strong comparison of Sac7d with the SH3 domain. Rather, if we neglect the ordering of the secondary structural elements, the gross topology of Sac7d is reminiscent of ubiquitin (Vijay-Kumar et al., 1987), another very stable small protein.

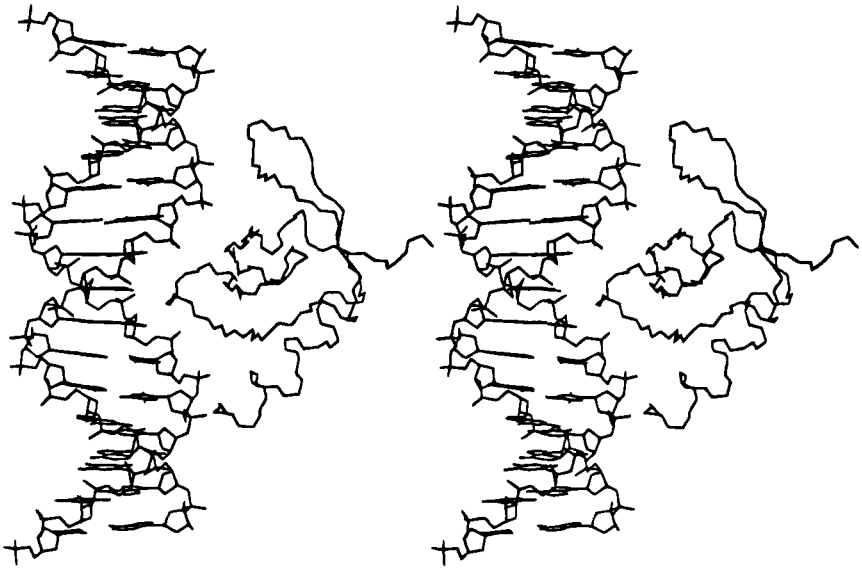

FIGURE 15: Stereoview of a model of the Sac7d−DNA complex extrapolated from that for Sso7d (Baumann et al., 1995) showing the potential contact of the extended C-terminal α-helix in Sac7d with the minor groove of an arbitratry B-DNA sequence. The placement of the three-stranded sheet in the major groove and the turn of the β-ribbon in the minor groove are as in the Sso7d complex. The N-terminus is on the right, away from the protein−DNA interface.

Given the almost identical amino acid sequences, significant structural differences between Sac7d and Sso7d seem unlikely. Structural differences may result from methylation of the native protein, but these are not expected to be large (McAfee et al., 1995). The structure determination of native Sso7d was complicated by the use of a heterogeneous sample composed of different isoforms with incomplete methylation of five different lysines. The effect of lysine monomethylation on the Sac7 structure remains to be determined.

DNA Binding. The solvent exposure of Trp 24 (48%) is consistent with the fluorescence excitation and emission spectra of Sac7d (McAfee et al., 1995). The fluorescence of Trp 24 is dramatically quenched by nearly 90% upon protein binding to DNA, indicating that Trp 24 helps to define the DNA-binding site, although this is not necessarily the case, especially given the small size of the protein. Interactions of adjacent Sac7 molecules on the DNA duplex might lead to fluorescence quenching; however, initial studies indicate a single binding mode with no cooperative interactions (McAfee, 1993).

Baumann et al. (1995) have shown that amide resonances in the three-stranded β-sheet of Sso7d are perturbed in the presence of DNA under conditions of fast exchange. They have presented a model of the DNA complex in which the sheet is associated with the major groove, while the residues in the turn between the two strands of the β-ribbon interact with the minor groove. The structure of Sac7d presented here is consistent with such a model. We note that the extended helix in Sac7d (which is lacking in Sso7d) is positioned to also interact with the minor groove (Figure 15), possibly enhancing the affinity of Sac7d relative to Sso7d. The terminal lysines 65 and 66 are positioned to make electrostatic interactions with the phosphate backbone. This would also explain why these lysines are not methylated in the native protein, which constrasts with the C-terminal lysines of Sso7d which are solvent exposed in the complex and available for methylation (Baumann et al., 1995). The model in Figure 15 indicates that Sac7d would cover approximately 9 base pairs in the complex, which is significantly larger than the 3.5 determined by reverse titrations (McAfee et al., 1995). However, molecular graphics docking experiments indicate that effective packing of monomers around the coil of the helix can lead to an effective site size of approximately 3.5 (Edmondson and Shriver, unpublished results).

Origin of Thermostability. Protein stability can be dramatically affected by only a few interactions. For example, disulfide bonds can make substantial contributions to protein stability and are preferentially found in smaller proteins (White, 1992). While it has proven difficult to dramatically increase the stability of a protein by engineering disulfide cross-links into their structures, their effect can be most easily seen by noting that formation of the three disulfides in bovine pancreatic trypsin inhibitor increases the T_m from approximately 25 °C to over 100 °C (Moses & Hinz, 1983). Disulfides appear to be rare in thermophile proteins due to their thermolability (Volkin & Klibanov, 1987), demonstrating that they are not essential for high stability. Metal cofactors can also have dramatic effects on the stability of proteins; e.g., the binding of calcium to parvalbumin increases the T_m from approximately 35 to 90 °C at pH 7.0 (Filimonov et al., 1978).

The structure of Sac7d does not show any obvious factors which might explain its extreme thermostability. There are no metal binding sites, cysteines, or internal salt bridges. The protein is globular with a surface area of 4050 Å^2, in good agreement with the expected value of 4289 Å^2 using the relationship derived by Miller et al. (1987) from a study of 46 monomeric proteins with molecular weights from 4000 to 35 000 (i.e., $A = 6.3M^{0.73}$, where M is the molecular weight). In fact, the agreement is better than indicated, since the surface areas calculated by Miller et al. were one probe radius (1.4 Å) away from the protein surface, while the routine used here calculates the actual molecular surface area. It would therefore appear that the packing density of the Sac7d protein is not exceptionally different from that of a mesophile protein. Indeed, the sizable cavity in the structure of Sac7d indicates that packing is not very efficient.

A study of the sizes and distributions of cavities in 12 proteins has indicated that cavities are quite common and

sometimes large (Rashin et al., 1986). The energy contribution of a cavity has been estimated to be on the order of 25−60 cal/(mol·Å^3) (Rashin et al., 1986; Eriksson et al., 1993). The volume of the cavity in Sac7d is approximately 70 Å^3, resulting in a destabilization of 1.8−4.2 kcal/mol.

It is interesting to compare the global fold of Sac7d to that of the comparably highly stable immunoglobulin binding domain B1 of streptococcal protein G (Gronenborn et al., 1991). Gronenborn et al. argued that an important factor in conferring stability on this protein was the novel (−1, +3x, −1) topology which prevented fraying of the amino and carboxy termini to initiate unfolding. This now seems unlikely given the presentation here of the structure of a protein which is even more stable but which has two highly exposed termini quite susceptible to fraying.

The most straightforward explanation for the extreme thermostability of the Sac7 and Sso7 proteins (as well as the B1 domain of protein G) is the expected relatively low molar ΔC_p for unfolding of small proteins due to small numbers of hydrophobic residues (Alexander et al., 1992). Decreasing the molar ΔC_p leads to a flattening of the stability curve (Schellman et al., 1981) and therefore an increase in T_m, but not an increase in stability except at high and low temperatures. A determination of the thermodynamic state functions using differential scanning calorimetry is currently in progress to determine if this is applicable to Sac7.

There appears to be a novel mechanism for stabilizing proteins utilized by the hyperthermophiles which has not been described for any mesophile protein. The native Sac7 proteins are approximately 7° more thermostable than the recombinant Sac7d protein (McAfee et al., 1995); e.g., native Sac7 unfolds reversibly with a midpoint temperature of approximately 100 °C at pH 6, while Sac7d unfolds with a T_m of 92.7 °C (McAfee et al., 1995). The only apparently significant difference is the specific monomethylation of lysines 5 and 7 in the native protein. The effect is not electrostatic in origin since decreased thermostability is independent of pH from pH 3 to pH 10 (Shriver, unpublished results) [monomethylation of lysine shifts the pK of the ϵ-amino group from 9.25 to 10.63 (Paik & Kim, 1980)].

With the structure of Sac7d in hand, it is tempting to speculate on the structural implications of lysine methylation. Increased hydrophobicity and decreased flexibility have been associated with increased stability (Menendez-Arias & Argos, 1989). Lysines 5 and 7 are highly solvent exposed but lie along the inner edge of the two-stranded ribbon. Methylation may lead to insertion of the lysine side chains into the hydrophobic core and adjustment of the size of the cavity, with increased stability resulting from filling of the cavity and enhanced van der Waals interactions. However, increased burying of hydrophobic residues will increase the ΔC_p of unfolding and therefore reduce the flattening of the stability curve described above.

It should also be noted that Sac7e and Sso7d (but not Sac7d) are monomethylated at specific lysines in the carboxy terminus as well as in the amino terminus. Such methylation should serve to extend the amphipathic face of the helix and increase the number of hydrophobic interactions. The helix in Sso7d is three residues shorter than in Sac7d, and therefore the carboxy terminus is not expected to be as exposed, whereas some shifting of the helix in Sac7e may be required to take full advantage of the increased hydrophobic surface area.

SUPPORTING INFORMATION AVAILABLE

One table containing ^{1}H NMR assignments at 35 °C and secondary structural elements (4 pages). Ordering information is given on any current masthead page.

REFERENCES

Alexander, P., Fahnestock, S., Lee, T., Orban, J., & Bryan, P. (1992) *Biochemistry 31*, 3597−3603.

Ammendola, S., Raia, C., Caruso, C., Camardella, L., D'Auria, S., De Rosa, M., & Rossi, M. (1992) *Biochemistry 31*, 12514−12523.

Anemuller, S., & Schafer, G. (1990) *Eur. J. Biochem. 191*, 297−305.

Baleja, J., Moult, J., & Sykes, B. (1990) *J. Magn. Reson. 87*, 375−384.

Baumann, H., Knapp, S., Lundbach, T., Ladenstein, R., & Härd, T. (1994) *Nature Struct. Biol. 1*, 808−819.

Baumann, H., Knapp, S., Karshikoff, A., Ladanstein, R.,, & Härd, T. (1995) *J. Mol. Biol. 247*, 840−846.

Bax, A. (1989) *Methods Enzymol. 176*, 151−168.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355−366.

Blake, P., Day, M., Hsu, B., Joshua-Tor, L., Park, J.-B., Hare, D., Adams, M., Rees, D., & Summers, M. (1992) *Proteins 1*, 1522.

Braunschweiler, L., & Ernst, R. (1983) *J. Magn. Reson. 53*, 521−558.

Bundi, A., & Wüthrich, K. (1979) *Biopolymers 18*, 285−297.

Choli, T., Henning, P., Wittmann-Liebold, B., & Reinhardt, R. (1988a) *Biochim. Biophys. Acta 950*, 193−203.

Choli, T., Wittmann-Liebold, B., & Reinhardt, R. (1988b) *J. Biol. Chem. 263*, 7087−7093.

Darr, S., Pace, B., & Pace, N. (1990) *J. Biol. Chem. 265*, 12927−12932.

Dijk, J., & Reinhardt, R. (1986) in *Bacterial Chromatin* (Gualerzi, C., & Pon, C., Eds.) Springer-Verlag, Berlin.

Edmondson, S. (1992) *J. Magn. Reson. 98*, 283−298.

Edmondson, S. (1993) *J. Magn. Reson. 103*, 222−233.

Edmondson, S., Khan, N., Shriver, J., Zdunek, J., & Gräslund, A. (1991) *Biochemistry 30*, 11271−11279.

Elie, C., DeRecondo, A.-M., & Forterre, P. (1989) *Eur. J. Biochem. 178*, 619−626.

Eriksson, A., Baase, W., Zhang, X.-J., Heinz, D., Blaber, M., Baldwin, E., & Matthews, B. (1993) *Science 255*, 178−183.

Ferrin, T., Huang, C., Jarvis, L., & Langridge, R. (1988) *J. Mol. Graphics 6*, 13−27.

Filimonov, V., Pfeil, W., Tsalkova, T., & Privalov, P. (1978) *Biophys. Chem. 8*, 117−122.

Fusek, M., Lin, X., & Tang, J. (1990) *J. Biol. Chem. 265*, 1496−1501.

Fusi, P., Tedeschi, G., ALiverti, A., Ronchi, S., Tortora, P., & Guerritore, A. (1993) *Eur. J. Biochem. 211*, 305−310.

Gonzalez, D., Rullmann, J., Bonvin, A., Boelens, R., & Kaptein, R. (1991) *J. Magn. Reson. 91*, 659.

Gronenborn, A., Filpula, D., Essig, N., Achari, A., Whitlow, M., Wingfield, P., & Clore, G. M. (1991) *Science 253*, 657−661.

Grote, M., Dijk, J., & Reinhardt, R. (1986) *Biochim. Biophys. Acta 873*, 405−413.

Guagliardi, A., Cerchia, L., DeRosa, M., Rossi, M., & Bartolucci, S. (1992) *FEBS Lett. 303*, 27−30.

Guagliardi, A., Cerchia, L., Bartolucci, S., & Rossi, M. (1994) *Protein Sci. 3*, 1436−1443.

Güntert, P., Braum, W., Billeter, M., & Wüthrich, K. (1989) *J. Am. Chem. Soc. 111*, 3997−4004.

Güntert, P., Braun, W., & Wüthrich, K. (1991a) *J. Mol. Biol. 217*, 517−530.

Güntert, P., Qian, Y., Otting, G., Müller, M., Gehring, W., & Wüthrich, K. (1991b) *J. Mol. Biol. 217*, 531−540.

Huang, C. C., Pettersen, E. F., Klein, T. E., Ferrin, T. E., & Langridge, R. (1992) *J. Mol. Graphics 9*, 230−236.

Imada, K., Sato, M., Tanaka, N., Katsube, Y., Matuura, Y., & Oshima, T. (1991) *J. Mol. Biol. 222*, 725−738.

Jaenicke, R., & Zavodszky, P. (1990) *FEBS Lett. 268*, 344−349.

Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. (1979) *J. Chem. Phys. 71*, 4546−4553.

Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2577−2637.
Knapp, S., Schmidt-Krey, I., Hebert, H., Bergman, T., Jornvall, H., & Ladenstein, R. (1994) *J. Mol. Biol. 242*, 397−407.
Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Petterson, G., Knowles, J., & Gronenborn, A. M. (1989) *Biochemistry 28*, 7241−7257.
Lacher, K., & Schafer, G. (1993) *Arch. Biochem. Biophys. 302*, 391−397.
Macura, S., Huang, Y., Suter, D., & Ernst, R. (1981) *J. Magn. Reson. 43*, 259.
Maras, B., Consalvi, V., Chiaraluce, R., Politi, L., De Rosa, M., Bossa, R., Scandurra, R., & Barra, D. (1992) *Eur. J. Biochem. 203*, 81−87.
McAfee, J. (1993) Genetic and DNA Binding Studies on a 7 kDa Protein from *Sulfolobus acidocaldarius*, Ph.D. Dissertation, Southern Illinois University, Carbondale.
McAfee, J., Edmondson, S., Datta, P., Shriver, J., & Gupta, R. (1995) *Biochemistry 34*, 10063−10077.
Menendez-Arias, L., & Argos, P. (1989) *J. Mol. Biol. 206*, 397−406.
Miller, S., Janin, J., Lesk, A., & Chothia, C. (1987) *J. Mol. Biol. 196*, 641−656.
Minami, Y., Wakabayashi, S., Wada, K., Matsubara, H., Kerscher, L., & Oesterhelt, D. (1985) *J. Biochem. 97*, 745−753.
Moses, E., & Hinz, H.-J. (1983) *J. Mol. Biol. 170*, 765−776.
Mueller, L. (1987) *J. Magn. Reson. 72*, 191−196.
Nadal, M., Jaxel, C., Portemer, C., Forterre, P., Mirambeau, G., & Duguet, M. (1988) *Biochemistry 27*, 9102−9108.
Nikonowicz, E., Meadows, R., & Gorenstein, D. (1990) *Biochemistry 29*, 4193−4204.
Osapay, K., & Case, D. A. (1991) *J. Am. Chem. Soc. 113*, 9436−9444.
Osapay, K., & Case, D. A. (1994) *J. Biomol. NMR 4*, 215−230.
Paik, W., & Kim, S. (1980) *Protein Methylation*, John Wiley and Sons, New York.
Pearlman, D., Case, D., Caldwell, J., Seibel, G., Chandra Singh, U., Weiner, P., & Kollman, P. (1991) *AMBER 4.0*, University of California, San Francisco.
Piatini, U., Sorensen, O., & Ernst, R. (1982) *J. Am. Chem. Soc. 104*, 6800−6801.
Rance, M., Bodenhausen, G., Wagner, G., Ernst, R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479−485.
Rashin, A., Iofin, M., & Honig, B. (1986) *Biochemistry 25*, 3619−3625.
Reddy, T., & Suryanarayana, T. (1988) *Biochim. Biophys. Acta 949*, 87−96.
Richards, F. (1977) *Annu. Rev. Biophys. Bioeng. 6*, 151−176.
Richardson, J., & Richardson, D. (1989) in *Prediction of Protein Structure, & the Principles of Protein Conformation* (Fasman, G., Eds.) Plenum, New York.
Schellman, J., Lindorfer, M., Hawkes, R., & Grutter, M. (1981) *Biopolymers 20*, 1989−1999.
Shriver, J., & Edmondson, S. (1993) *Biochemistry 32*, 1610−1617.
Shriver, J. W., & Edmondson, S. P. (1994) *Methods Enzymol. 240*, 415−437.
States, D., Haberkorn, R., & Ruben, D. (1982) *J. Magn. Reson. 48*, 286−292.
Stetter, K. O., Fiala, G., Huber, G., Huber, R., & Segerer, A. (1990) *FEMS Microbiol. Rev. 75*, 117−124.
Vijay-Kumar, S., Bugg, C., & Cook, W. (1987) *J. Mol. Biol. 194*, 531.
Volkin, D., & Klibanov, A. (1987) *J. Biol. Chem. 262*, 2945−2950.
Weiner, P., & Kollman, P. (1981) *J. Comput. Chem. 2*, 287−303.
Weiner, S., Khllman, P., Case, D., Singh, U., Ghio, C., Alagona, G., Profeta, S., & Weiner, P. (1984) *J. Am. Chem. Soc. 106*, 765−784.
Weiner, S., Kollman, P., Nguyen, C., & Case, D. (1986) *J. Comput. Chem. 7*, 230−252.
White, S. (1992) *J. Mol. Biol. 227*, 991−995.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991) *J. Mol. Biol. 222*, 311−333.
Woese, C., Kandler, O., & Wheelis, M. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 4576−4579.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York.
Zuiderweg, E., Hallenga, K., & Olejniczak, E. (1986) *J. Magn. Reson. 70*, 336−343.

BI950833F