A Thermodynamic Scale for the β-Sheet Forming Tendencies of the Amino Acids

Catherine K. Smith,‡ Jane M. Withka,§ and Lynne Regan*,‡

Department of Molecular Biophysics and Biochemistry, Yale University, P.O. Box 208114, 266 Whitney Avenue, New Haven, Connecticut 06520-8114, and Pfizer Central Research, Eastern Point Road, Groton, Connecticut 06340

Received December 20, 1993; Revised Manuscript Received February 14, 1994®

ABSTRACT: The results of a study to measure the β-sheet forming propensities of the 20 naturally occurring amino acids are presented. The protein host for these studies is the 56 amino acid B1 domain of staphylococcal IgG binding protein G [Fahnestock, S. R., Alexander, P., Nagle, J., & Filpula, D. (1986) *J. Bacteriol. 167,* 870-880]. This protein was selected because it exhibits a reversible two-state thermal denaturation transition and its structure is known at high resolution. A suitable guest position in the protein was identified, and its neighboring environment was modified to minimize the potential for artifactual interactions. All 20 amino acids were individually substituted at the guest site, and their effect on the protein's thermal stability was determined. NMR was used to verify the structural integrity of several of the proteins with different amino acid substitutions at the guest site. The results of these studies provide a thermodynamic scale for the relative β-sheet forming propensities of the amino acids that shows a clear correlation with the β-sheet preferences derived from statistical surveys of proteins of known structure.

The solution of the "protein folding problem" requires a complete understanding of how a protein's primary amino acid sequence specifies its three-dimensional structure. Progress toward this goal necessitates a deconvolution of the individual factors that contribute to protein stability. The problem can be simplified, both conceptually and experimentally, if one considers the formation of secondary and tertiary structure separately. It is then possible to ask two separate questions: "What are the important factors governing the stability of α-helices and β-sheets?" and "How do the helices and sheets pack together to form tertiary structure?" Here we address the first of these questions, specifically by developing a thermodynamic scale to quantitate the β-sheet forming tendencies of the amino acids.

Statistical surveys of proteins of known structure describe the frequency of occurrence of each amino acid in an α-helix, β-sheet, or β-turn, and the results reveal a nonrandom distribution (Chou & Fasman, 1974, 1978). It follows that the amino acids have definite conformational preferences, and the extent to which these preferences favor the formation of a particular secondary structure is one of the important factors determining the protein stability. Building upon these observations, several groups have attempted to quantify the preferences for α-helix formation experimentally. Pioneering studies in this area were performed by Scheraga and colleagues, who devised the "host–guest" method for measuring the intrinsic preference of each amino acid to assume a helical conformation (Scheraga, 1978; Sueki et al., 1984).

Subsequent to these early experiments, it was demonstrated that short monomeric peptides of only 20 amino acids were able to form helices in solution, motivating several groups to make measurements of α-helical propensities using short peptides as the host molecule (Lyu et al., 1990; Padmanabhan et al., 1990). The rationale which underlies these later experiments derives from the original copolymer work of Scheraga (1978). Specifically, the stability of a standard protein or peptide is compared with that of mutants in which all the amino acids are substituted into the guest site individually. Changes in stability of the mutant relative to a standard host are measured through an analysis of their thermally or chemically induced denaturation curves.

The propensity of each amino acid for α-helix formation has been measured in the context of monomeric helical peptides, associating trimeric peptides and exposed surface helices in proteins (Padmanabhan et al., 1990; Lyu et al., 1990; O'Neil & DeGrado, 1990; Horovitz et al., 1992; Blaber et al., 1993). All of these studies have revealed measurable differences in the propensities of the amino acids to assume an α-helical conformation. In addition, these experimentally determined α-helical propensity scales correlate reasonably well with the propensity scales derived from statistical analyses of proteins of known structure. The range of free energy differences that are observed between the best and worst helix formers (excluding Gly and Pro) is about 0.7–0.9 kcal/mol in the peptide studies and increases to about 3 kcal/mol when Pro is included. A somewhat larger range of free energies but with a similar rank order has been derived from the protein studies.

In contrast to the experiments determining α-helical propensities, the study of β-sheets has not been as successful, principally because of the lack of an unaggregated model system which would allow detailed thermodynamic and structural characterization. Nevertheless, previous attempts to study β-sheet formation do exist. First, poly(L-lysine) under certain conditions yields a high molecular weight aggregate that has been assigned β-structure by CD.[1] Early work was performed with poly(L-lysine) with the aim of understanding the kinetics of β-sheet formation (Hartman et al., 1974). Second, Mutter and Altman (1985) performed a qualitative assessment of the β-structure forming properties of Ala, Ile, and Leu in a homooligopeptide conjugated with poly(ethylene glycol). Neither of these systems, however, proved appropriate for the development of a scale to describe β-sheet propensities.

* To whom correspondence should be addressed.
‡ Yale University.
§ Pfizer Central Research.
® Abstract published in *Advance ACS Abstracts,* March 15, 1994.

[1] Abbreviations: CD, circular dichroism; T_m, melting temperature; DSC, differential scanning calorimetry; IPTG, isopropyl thio-β-D-galactopyranoside; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser enhancement spectroscopy; TOCSY, total correlation spectroscopy; ppm, parts per million; 2D, two dimensional; TPPI, time proportional phase incrementation; UV, ultraviolet.

0006-2960/94/0433-5510$04.50/0

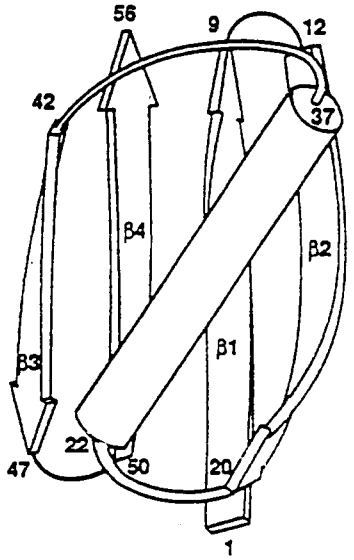

FIGURE 1: Ribbon diagram illustrating the overall protein fold of β1. Reproduced with permission (Gronenborn et al., 1991).

More recently, Kemp (1990) used a heteroaromatic template, diacylaminoepindolidione, linked to short peptides to investigate β-sheet formation. Several amino acids were tested in the conjugated peptide, and the identity of the substituted amino acid was shown to have a relatively strong influence on β-sheet stability (Gly < Ala < Phe, Val). This limited study correlated with the relative ranking of these amino acids in Chou and Fasman (1974, 1978) probability studies. Although sheet structure was detected by NMR, face to face packing of these ring structures precluded a detailed analysis of local stability effects. In addition, these molecules were only soluble in the presence of DMSO or other organic solvents.

Finally, the most complete attempt to delineate the β-sheet propensities is the work of Kim and Berg (1993). In their studies, an exposed residue on the β-strand of a zinc finger peptide was targeted as the guest site. The relative stability of each peptide was determined by competition with a second zinc-finger peptide in a Co(II)-binding assay. The amino acid rankings obtained in this study show a striking correlation (correlation coefficient 0.83, excluding Pro) with Chou–Fasman (Chou & Fasman, 1974, 1978) probability listings for β-sheet propensity. However, the substitutions in this study were made at a single, somewhat unusual guest site. Specifically, the residues at $i-2$ and $i+2$ relative to the guest site in the β-strand were both proline. Also, the range of values reported was small. There was a span of only 0.56 kcal/mol between the best (Ile) and worst (Gly) β-sheet forming residues and only 0.23 kcal/mol between the best and worst residues if Gly and Pro are excluded.

These interesting and important first observations emphasized the need to make similar measurements but in a more homogeneous sheet environment and in a protein rather than a peptide host. The β-sheet propensity studies presented here use the IgG F_c-binding domain B1 of staphylococcal protein G (hereafter referred to as β1) as the host. The structure of β1 has been determined by NMR and can be described as an "open faced sandwich" composed of four strands of β-sheet that are crossed diagonally by a single α-helix as shown in Figure 1 (Gronenborn et al., 1991). β1 is small (56 amino acids) and highly soluble, contains neither proline residues nor disulfide bonds, and displays a reversible two-state thermal denaturation transition (Gronenborn et al., 1991). Together, these properties made β1 an ideal choice for a model system in which to make the detailed thermodynamic and structural characterizations that are essential for the β-sheet propensity studies.

MATERIALS AND METHODS

Modeling. Solvent accessibility calculations were based on the structure of wild-type β1 determined by Gronenborn and colleagues (1991), the coordinates of which are deposited in the Brookhaven Protein Data Base (PDB accession number 2GB1). The modeling of different potential host environments was performed using the computer program What If (G. Vriend, EMBL). Accessible surface area calculations were performed using the computer program ACCESS with a 1.4-Å probe sphere radius (Lee & Richards, 1971).

Cloning and Protein Purification. A synthetic gene encoding β1 was synthesized and cloned into the T7 based expression plasmid, pET11aΔ (Lemmon et al., 1992; Studier et al., 1990), using standard cloning procedures (Sambrook et al., 1989). Codons were selected based on the highest *Esherichia coli* usage frequency (Sharp & Li, 1987), except where inclusion of a less common codon allowed the incorporation of a convenient restriction site. Expression of wild-type β1 in this system yielded a heterogeneous population of protein containing two species. Through N-terminal sequencing, these species were shown to be an unprocessed form of β1 retaining the N-terminal methionine and a processed form lacking the N-terminal methionine. This incomplete processing had been reported previously and has serious consequences: The two species differ in stability by 1.7 kcal/mol, and the heterogeneity of the N-terminus causes a duplication of many of the resonances in the ^1H NMR spectrum [A. Gronenborn, personal communication; Gronenborn et al., 1991]. Before initiating the β-sheet propensity studies, it was therefore necessary to make modifications to the sequence to obtain an homogeneous protein preparation.

Substrate specificity studies on the processing enzyme, N-terminal methionine aminopeptidase, indicate a clear recognition preference for Ala, Gly, Ser, and Thr at the second position following the N-terminal methionine (Boissel et al., 1988). Accordingly, substitution of Gln for Thr at position 2 of β1 should prevent cleavage of the N-terminal methionine. We synthesized the gene containing the T2Q mutation in β1 and expressed and purified the mutant protein. N-terminal amino acid sequencing of this material showed that the mutant β1-T2Q is completely unprocessed and retains methionine in the N-terminal position. All subsequent mutations and studies were made in the β1-T2Q background. Mutations at position 6 and 44 were introduced by cassette mutagenesis. Mutations at position 53 were made using a cassette randomized at position 53 in order to encode all the amino acids at this position. All mutations were verified by double-stranded DNA dideoxy sequencing (Sanger et al., 1977).

Plasmids encoding mutant proteins were transformed into the *E. coli* T7 expression strains HMS174 (DE3) or BL21 (DE3) (Studier et al., 1990). Transformed cells were grown in Luria broth to an OD_{600} of approximately 0.6 and were then induced for 2.5 h at 37 °C with 1 mM IPTG. β1 variants with Gly and Pro at the guest position were induced at 30 °C for 2.5 h because of the low protein yields encountered under higher temperatures. Cells were harvested by centrifugation, sonicated in 20 mM Tris-HCl, pH 8.0, and centrifuged again to remove cellular debris. Poly(ethylenimine) was added to the crude lysate to a final concentration of approximately 0.4% (v/v) while stirring to precipitate nucleic acids, which

were subsequently removed by centrifugation. The final cleared lysate was filtered through a 0.45-μm filter and applied to a Q-Sepharose high performance column (Pharmacia), equilibrated in 20 mM Tris-HCl, pH 8.0. β1 protein was eluted with a 0–35% gradient of 20 mM Tris-HCl, pH 8.0, 1 M NaCl over 400 mL. For most of the variants, the β1 protein that eluted from this column was sufficiently pure that no further purification was required. In the few examples where further purification was necessary, gel filtration on a Hi-Load Superdex 75 column (Pharmacia) in 20 mM Tris-HCl, pH 8.0, was performed. The purity of mutant proteins was monitored by SDS–polyacrylamide gel electrophoresis; all final protein preparations were at least 95% homogeneous by this assay. Typically, the final yield of purified protein was approximately 100 mg from a 4-L *E. coli* culture. Protein concentrations were determined by UV absorbance, using the published value for the extinction coefficient (A_{280} of 1.32 corresponds to a 1 mg/mL solution) (Alexander et al., 1992a), except for the β1 variant with Trp at the guest site, for which the extinction coefficient was corrected.

Thermal Denaturation. Thermal denaturation of the proteins was monitored by following the ellipticity of the proteins at 222 nm (θ_{obs}) as a function of temperature, on an Aviv Circular Dichroism Spectrophotometer model 62DS. A wavelength of 222 nm was chosen because at this point there is a large difference between the folded and unfolded protein (Figure 7), but there is little buffer absorbance, which would facilitate future chemical denaturation studies. Experiments were performed in 50 mM sodium acetate, pH 5.2, at a protein concentration of 0.5 mg/mL. These buffer conditions were chosen based on previous studies which reported that the protein was most stable at this pH (Alexander et al., 1992b). In addition, for β1 variants with Thr and Ala at the guest position, we verified that under these conditions the thermal denaturation transition was reversible and that the calculated van't Hoff enthalpy was equal to the calorimetric enthalpy (Hawkes et al., 1984). For β1-T2Q we determined that ΔH_{Cal} = 66.97 kcal/mol, $\Delta H_{van't\ Hoff}$ = 64.83 kcal/mol, and $\Delta H_{Cal}/\Delta H_{van't\ Hoff}$ = 1.03.

The thermal denaturation curves were analyzed according to a two-state transition model. Linear extrapolation of the pre- and posttransition baselines was performed using the curve-fitting program Kaleidagraph (Synergy Software). The fraction native (F_n) at each temperature was calculated using the equation $F_n = (\theta_{obs} - \theta_u)/(\theta_f - \theta_u)$, where θ_{obs} is the observed ellipticity at 222 nm, and θ_f and θ_u are the folded and unfolded ellipticities derived from the extrapolated baselines. The fraction unfolded (F_u) is given by $F_u = 1 - F_n$. These data allowed calculation of van't Hoff plots from which the T_m and ΔH were determined. The entropy, ΔS, was calculated at T_m, and these values of ΔH and ΔS were used to calculate ΔG for each protein at 333 K. The mutant containing Ala at position 53 was taken as the standard, and $\Delta\Delta G$ for each mutant was calculated relative to ΔG_{Ala}: $\Delta\Delta G_{\beta 1X} = \Delta G_{\beta 1X} - \Delta G_{\beta 1A}$. All thermal unfolding experiments were performed in triplicate and the results averaged. We estimate that the maximum error in the T_m to be ±0.4 °C and in ΔG to be less than 5%.

Calorimetry. Calorimetric measurements were performed on a Microcal MC-2 differential scanning calorimeter at a scan rate of 1 °C per minute in 50 mM sodium acetate, pH 5.2. Protein concentrations were determined by duplicate measurements of the protein absorbance at 280 nm. The calorimetric measurements were performed using protein samples of approximately 3 mg/mL. Data was analyzed using the program INDEP.

NMR Spectroscopy. Protein samples were concentrated to approximately 3.2 mM using a Centriprep 3 microconcentrator (Amicon). A PD-10 (Pharmacia) desalting column was then used to exchange the protein into a buffer appropriate for the NMR studies (90% 50 mM d_4-acetate, pH 5.2, 10% D_2O). Finally, proteins in this deuterated buffer were reconcentrated to 3.2–3.75 mM, using a Centricon 3 microconcentrator (Amicon).

Proton assignments were obtained by standard ^{1}H 2D homonuclear methods which included TOCSY (Braunschweiler & Ernst, 1983), NOESY (Jeener et al., 1979; Anil-Kumar et al., 1983), and DQFCOSY (Piantini et al., 1982; Shaka & Freedman, 1983; Rance et al., 1983) techniques. Experiments were carried out on a Bruker AMX600 at 20 °C. Additional experiments were run at temperatures between 10 and 30 °C as necessary to resolve degenerate resonances. All spectra were recorded in a phase-sensitive manner using TPPI methods (Marion & Wüthrich, 1983). In all spectra, 48–64 scans for each of the 512 t_1 values were collected with 4K points. Water suppression was achieved by presaturation during the 1.3–1.5 s relaxation delay. For the NOESY spectra, low-power presaturation was also used during the 100-ms mixing time for additional water suppression. The TOCSY spectra was obtained using the DIPSI II spin locking sequence (Shaka et al., 1988) and a mixing time of 75 ms.

Data were processed and analyzed on a Silicon Graphics Indigo using Felix Software (BIOSYM Technologies, San Diego). In the acquisition dimension, low-frequency deconvolution was applied to remove the residual water signal (Marion et al., 1989). Data were multiplied by a 60°-shifted squared sine bell or 90°-shifted sine bell apodization function prior to Fourier transformation. A polynomial baseline correction was applied to the rows of the transformed matrices. In the t_1 dimension, the data were zero filled to 1K points, and a 60°-shifted squared sine bell or 90°-shifted sine bell apodization function was applied before Fourier transformation.

RESULTS

Selection of the Guest Site and Design of the Host Environment. The position and local environment of the guest site are of key importance in ensuring that any difference in stability between the β1 variants reflects different β-sheet propensities of the amino acids rather than artifacts caused by other interactions. Figure 2 provides a view of the location and hydrogen bond interactions of all the β-strand residues. A cursory inspection of this figure suggests that several exposed β-strand positions might serve as a satisfactory guest site. However, position 53 was selected as the optimal site for the following reasons: First, it is solvent exposed, and its side chain extends from the face of the sheet that is opposite to the helix. Such a location avoids packing and helix–guest residue interactions. Position 53 is found in the middle of a central β-strand, which provides the most homogeneous sheet environment available. It is possible that the conformational requirements of the last hydrogen bonding position before a turn or of an edge strand may be different from those of the rest of the sheet. Furthermore, the amide and carbonyl hydrogens of position 53 are fully bonded. In comparison, on an edge strand, the carbonyl and amino groups of alternate residues are not hydrogen bonded to another strand and interact freely with solution. As a result, these side chains may possess slightly more conformational freedom. This extra mobility may compensate for less favorable β-sheet forming

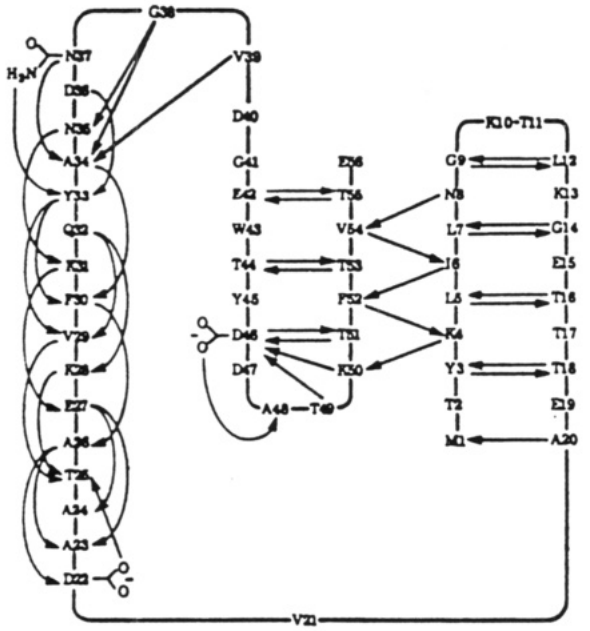

FIGURE 2: Diagrammatic representation of the hydrogen bonding pattern within β1. Hydrogen bonds are represented by arrows. Reproduced with permission (Gronenborn et al., 1991).

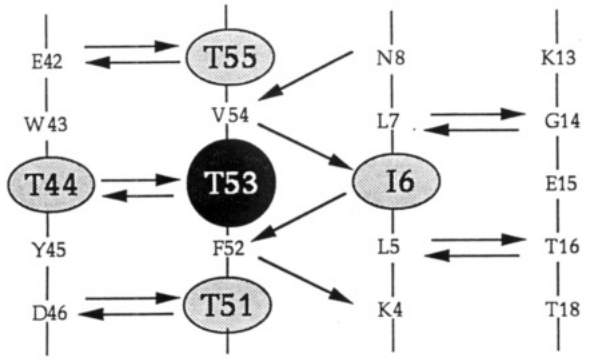

FIGURE 3: Illustration of the guest site and surrounding host environment in the four-stranded β-sheet region of β1. Hydrogen bonds are indicated by arrows. The guest site, position 53, is circled and highlighted in black. The nearest neighbor residues on the same face of the β-sheet, positions 6, 44, 51, and 55, are also circled, but highlighted in light gray.

residues and hence may mask the potentially small differences in actual conformational preferences.

With the guest position selected, it was then important to determine which surrounding mutations would best isolate the side chain of the guest residue in order to eliminate interfering interactions with neighboring residues. In wild-type β1, the four residues that are nearest neighbors to the guest site are residues T51 and T55 in the same strand and residues I6 and T44 on the two adjacent strands (Figure 3). In order to investigate the effect of mutating these four nearest neighbors on the solvent accessibility of the guest site, a computer modeling study was performed.

First a model was constructed in which residues 6, 44, 51, and 55 were changed to Gly. The solvent accessibilities of three different sized "test" residues at the guest site were calculated. Because this environment would maximally expose the guest residue in the context of the protein, these values were taken to represent an accessibility of the guest residue of 100%. In practice, however, a protein containing four Gly substitutions most likely would be too destabilized to be useful in these studies. Hence, the solvent accessibilities of the "test" guest residues were calculated with different combinations of residues at positions 6, 44, 51, and 55. These results were compared by expressing guest site accessibilities as a percentage of those calculated in the maximally exposed, all-glycine environment (Table 1).

There was little change in the accessibility of the guest residue in response to substitutions of Gly, Ala, or Thr at the same-strand neighbor positions, T51 and T55. As a result, it did not seem necessary to mutate these residues from their wild-type identity. In contrast, the identities of the cross-strand residues at positions 6 and 44 had a more substantial effect on the solvent accessibility of the guest residue. With Gly at positions 6 and 44, the accessibility of the guest residue was approximately 90% of the maximum, and with Ala at these positions it was about 70%. These results suggested that either the combination of I6G, T44G, T51, and T55 or I6A, T44A, T51, and T55 would provide the most appropriate neighbor environment. Because the single I6G mutation destabilized the protein by an unprecedented 25 °C in preliminary studies, the double Gly mutant I6G, T44G was predicted to be too destabilized for the proposed studies. The combination of mutations, I6A, T44A, T51, and T55, was then selected as the optimal host environment for it provided the best compromise between minimizing host–guest interactions, maximizing solvent accessibility of the guest site, and maintaining a reasonably stable protein.

Creation of β1 Proteins with Guest Site Substitutions. All 20 amino acids were substituted individually at position 53, in the context of the I6A, T44A, T51, T55 host environment. The details of the mutagenesis strategy and the protein purification protocol are described under Materials and Methods. The protein purification protocol used here differs from the published methods, which involve heat denaturation and/or elution from an affinity matrix at extremely low pH (Alexander et al., 1992a; Gronenborn et al., 1991). We sought to avoid the potential irreversible chemical modifications that may occur as a result of these harsh treatments and to ensure that the protein preparations used were as homogeneous as possible. For the majority of the β1 variants, it was possible to purify large quantities of protein for structural and thermodynamic characterization with the exception of the proteins containing Gly and Pro at the guest site. In these

Table 1: Summary of the Solvent Accessibility of Guest Residues in Different Host Environments[a]

	host				
	6 44 51 55 G G G G	6 44 51 55 G G T T	6 44 51 55 A A A A	6 44 51 55 A A T T	6 44 51 55 I T T T (wild type)
guest					
Ala	100%	96%	74%	68%	56%
Ile	100%	94%	81%	75%	57%
Phe	100%	88%	74%	71%	62%

[a] The left column lists the identity of the test guest residue in position 53. The top row shows the identity of the test host residues at positions 6, 44, 51, and 55.

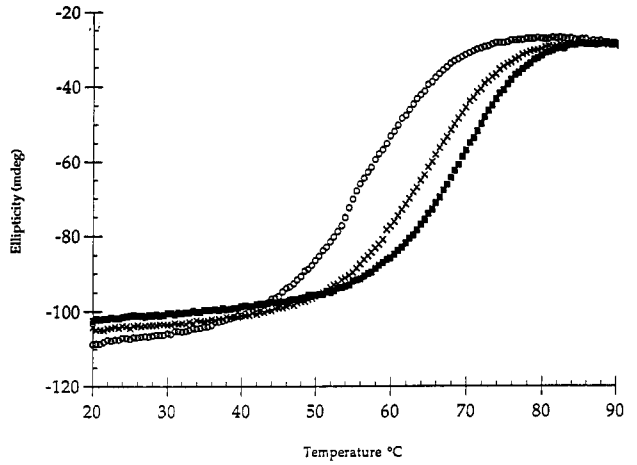

FIGURE 4: Typical raw data from thermal denaturation curves of β1 with three different amino acid substitutions at the guest position. The ellipticity at 222 nm is plotted against sample temperature. Open circles (○), Ala at the guest site; diagonal crosses (×), Met at the guest site; closed squares (■), Thr at the guest site.

cases, protein yields were reduced, and growth conditions were adjusted to obtain sufficient quantities of pure protein.

Stability Studies. The stability of the 20 proteins with different guest site substitutions was compared using CD to monitor their thermally induced denaturation transitions. Sample denaturation curves, showing the raw data for three proteins, are depicted in Figure 4. It is clear from the readily detectable differences in melting temperatures that protein stability is influenced quite substantially by the nature of the guest site substitution.

The results of the thermal denaturation studies were analyzed as described under Materials and Methods, yielding the melting temperature and free energy associated with the denaturation transition for each protein. Assuming all other interactions have been eliminated or at least minimized, the difference in free energy, $\Delta\Delta G$, between a particular variant and the standard, in this case Ala at the guest position, provides a quantitative measure of the relative intrinsic β-sheet forming preference of the amino acids. Hence, these $\Delta\Delta G$ values provide the basis for the quantitative scale of the β-sheet forming tendencies of the amino acids. We chose Ala in preference to Gly as the standard because it does not possess the unrepresentative rotational freedom of Gly and yet is the simplest amino acid with a side chain.

Table 2 shows the β-sheet propensity data, with the melting temperature and free energy difference relative to Ala for each guest site substitution. For comparison, the relative ranking of the amino acids from statistical analyses of proteins of known structure (Chou & Fasman, 1974, 1978) and from the Kim and Berg (1993) zinc finger study are also listed. Several features of Table 2 deserve comment. First, it is evident from an examination of the T_m and $\Delta\Delta G$ data obtained here that both give a clear ranking of the amino acids: All amino acids do not form β-structure equally well in this study. Moreover, the magnitude of differences observed between the best β-sheet former (Tyr) and the worst (Asp) is quite large, with a variation in the T_m of 18.3 °C and in the $\Delta\Delta G$ of 2.5 kcal/mol, excluding Gly and Pro. Second, the β-sheet propensities determined in our study show correlations with the statistically determined propensities. On a one-to-one basis, the correlation with the statistical probabilities is certainly not exact. Nevertheless, on a broader level, there are certain striking similarities between the scales. The best β-sheet forming residues, according to our study, are Val, Ile, Thr, Phe, Tyr, and Trp, and these amino acids are also clearly the residues that are most commonly found in β-structure in the statistical surveys (Chou & Fasman, 1974, 1978). Similarly, the worst β-sheet forming residues, according to our study, are Ala, Asp, Gly, and Pro, and again these are the residues that are found in β-structure most infrequently in statistical surveys. The existence of these correlations lends strong support to the contention that in the β1 studies we are indeed largely measuring intrinsic β-sheet forming propensities and not the artifactual effects of other interactions of the guest residues. Additional corroborating evidence for the general applicability of the $\Delta\Delta G$ values measured at the guest position can be found in a comparison of the stabilities of proteins β1 T2Q and β1 T2Q, I6A, T44A. The difference in stability of these two proteins can be accounted for by the sum of the $\Delta\Delta G$ values for Ile and Thr versus Ala measured at the guest position.

Table 2: Summary of the β-Sheet Propensity Data[a]

guest residue	T_m (°C)	$\Delta\Delta G$ (kcal/mol)	P_β	$\Delta\Delta G_{(Kim\&Berg)}$ (kcal/mol)
Tyr	69.22	−1.63	1.31	−0.50
Thr	68.67	−1.36	1.33	−0.48
Ile	67.78	−1.25	1.57	−0.56
Phe	67.68	−1.08	1.23	−0.55
Trp	65.73	−1.04	1.24	−0.48
Val	65.47	−0.94	1.64	−0.53
Ser	64.80	−0.87	0.94	−0.39
Met	64.26	−0.90	1.01	−0.46
Cys	63.99	−0.78	1.07	−0.47
Leu	62.47	−0.45	1.17	−0.48
Arg	62.41	−0.40	0.94	−0.44
Asn	61.88	−0.52	0.66	−0.38
His	60.96	−0.37	0.83	−0.46
Gln	60.90	−0.38	1.00	−0.40
Lys	60.65	−0.35	0.73	−0.41
Glu	58.81	−0.23	0.51	−0.41
Ala	57.05	0	0.79	−0.35
Asp	50.91	0.85	0.66	−0.41
Gly	45.95	1.21	0.87	0
Pro	<10	ND	0.62	0.23

[a] The amino acids are listed in the order of their β-sheet propensity. The T_m of the thermal denaturation transition and $\Delta\Delta G$ relative to that of the standard, β1 T2Q, I6A, T44A, T53A, are included. $\Delta\Delta G$ for Pro was not determined (ND) for reasons discussed in the text. Absolute values are as follows: for T2Q, ΔG = −9.96 kcal/mol (T_m = 81.89 °C); for T2Q, I6A, ΔG = −7.05 kcal/mol (T_m = 71.12 °C); for T2Q, I6A, T44A, T53, ΔG = −6.54 kcal/mol (T_m = 68.73 °C); and for T2Q, I6A, T44A, T53A, ΔG = −4.62 kcal/mol (T_m = 57.05 °C). In the table, ΔG is determined by calculating ΔH in the transition region, using the van't Hoff equation and then using this value to calculate ΔS at the T_m. This treatment assumes that, within the transition region, ΔH is independent of temperature. Accordingly, $\Delta\Delta G$ values are reported at a temperature that is within the transition region for all the mutants (60 °C). For the standard β1 containing Ala at the guest site, ΔG = −0.24 kcal/mol at 60 °C. Also shown are the P_β values for the probability of occurence of each amino acid in β-sheet in proteins of known structure (Chou & Fasman, 1974, 1978) and the $\Delta\Delta G$ values from the Kim and Berg (1993) zinc finger study, which are relative to Gly at the guest postion.

Structural Studies. For a complete interpretation of the β-sheet propensity results presented, it is important to show that the folded state of the proteins has not been significantly distorted by the different substitutions at the guest position. Specifically, it is essential to verify the structural integrity of the β-strand region for several of the proteins with different guest site substitutions, especially those that are severely destabilizing.

The structure of β1 was originally solved by NMR (Gronenborn et al., 1991), and the proton assignments for the two wild-type forms of β1 (Met-Thr- and Thr-) were available to us. It was therefore straightforward to reassign the spectra of the homogeneous T2Q wild-type and the T2Q,

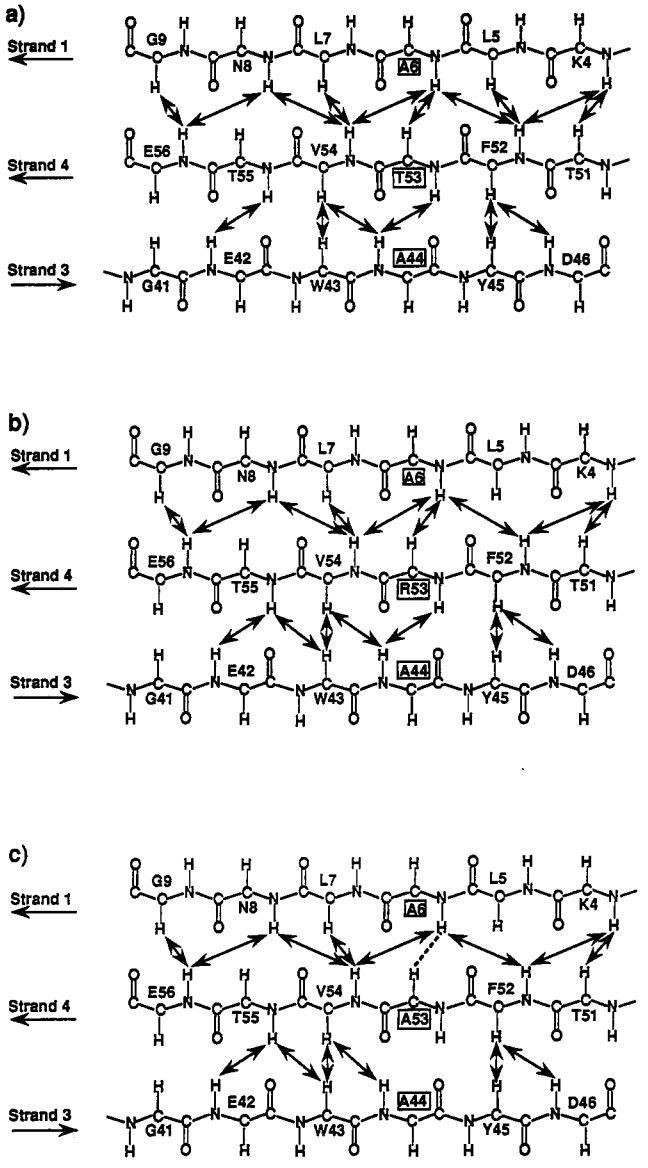

FIGURE 5: Schematic diagram illustrating the β-sheet regions for (a) the T2Q, I6A, T44A, T53 mutant, (b) the T2Q, I6A, T44A, T53R mutant, and (c) the T2Q, I6A, T44A, T53A mutant. Observed NOEs are indicated by solid double-headed arrows. The ambiguous NOE due to cross peak overlap is indicated by a dashed line in panel c.

I6A, T44A, 53X (X = R, T, and A) mutants following conventional homonuclear 2D NMR procedures (Wüthrich, 1986). In all samples, the β-strand regions were well defined with both the parallel and antiparallel β-strands showing characteristic strong sequential $NH_{(i+1)}/C\alpha H_{(i)}$ NOEs and also medium to weak long-range $NH_{(i+1)}/C\alpha H_{(j)}$ and $NH_{(i+1)}/NH_{(j-1)}$ NOEs. Additionally, strong $C\alpha H_{(i)}/C\alpha H_{(j)}$ NOEs were observed for antiparallel β-strands.

Figure 5 shows a schematic illustration of the β-sheet region and the observed NOEs defining antiparallel and parallel β-strands for the three variants of β1 that range in stability. Figure 5a illustrates the structural information for a highly stable variant containing Thr at the guest site (T_m = 68.7 °C). Figure 5b describes a moderately stable β1 mutant containing Arg at the guest site (T_m = 62.4 °C). Figure 5c details the standard β1 mutant that contains Ala at the guest site (T_m = 57.1 °C). In all cases, the antiparallel and parallel β-strands containing the guest and hosts sites are clearly maintained. However, at the experimental temperatures for the mutant containing Ala at the guest position, the $NH_{(A6)}/CaH_{(A53)}$ NOE cross peak involving the guest and host sites (Figure 5c, dashed line) cannot be unambiguously assigned because of overlap with the intraresidue $HN_{(A53)}/C\alpha H_{(A53)}$ NOE cross peak. Additionally, the amide chemical shifts of A44 and A53 are degenerate, and, consequently, the interstrand $HN_{(A44)}/NH_{(A53)}$ NOE cannot be observed.

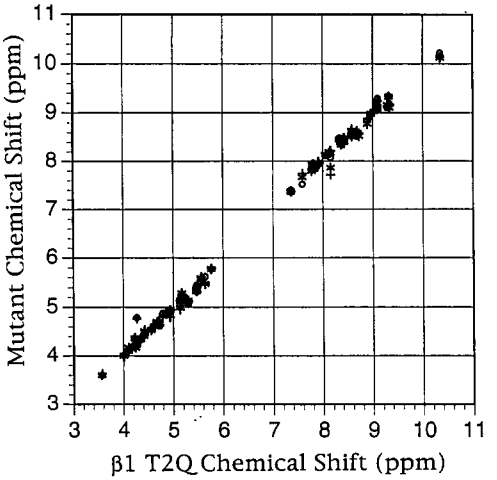

FIGURE 6: Comparison of the backbone amide and α-proton chemical shifts (ppm) for β-sheet residues (2–9, 12–20, 41–47, 50–56) for T2Q wild-type versus β1 T2Q, I6A, T44A, T53X. Open circles, X = Thr; crosses, X = Arg; and stars, X = Ala. Chemical shifts were determined at 20 °C with water referenced to 4.80 ppm. Chemical shifts for the alpha protons of L5 and A6 for the mutant containing Ala at the guest site (T2Q, I6A, T44A, T53A) were under the water signal at the experimental temperatures and were not included.

The similarity of proton chemical shifts for the wild-type and mutant forms provides additional evidence for the lack of structural perturbation as a result of mutated residues. In Figure 6, only very small variations in the backbone amide and alpha proton chemical shifts of β-sheet residues are observed for the T2Q wild-type relative to the T2Q, I6A, T44A, T53X derivatives. Additionally, chemical shifts for the α-helical and loop regions are also similar (data not shown), indicating that the global conformation of the protein is unaltered.

Typically, downfield-shifted backbone amide resonances are characteristic of β-sheet structure due to the network of interstrand hydrogen bonds. The amide protons of the guest site (residue 53) and the host sites (residues 6 and 44) can all potentially form interstrand hydrogen bonds (Figure 2). In this work, all amide chemical shifts for these resonances are shifted downfield (>9.0 ppm). Additionally, the amide chemical shifts of the T2Q wild-type and all β1 variants are similar regardless of the amino acid in the guest position. These results indicate that the β-sheet region is intact and that the amide resonances of residues 6, 44, and 53 are likely to be involved in interstrand hydrogen bonds.

Detailed structural characterization of the two most destabilized proteins was not possible by NMR because large quantities of these proteins could not be readily purified. CD was therefore used as an alternative means by which to compare their overall secondary structures. As shown in Figure 7, the CD spectra of β1 T2Q, I6A, T44A proteins with Thr and Gly at the guest position were very similar and contained the expected mixture of α- and β-secondary structure, suggesting that the overall structure of the Gly mutant was not grossly perturbed. In contrast, the CD spectrum of β1 T2Q, I6A, T44A with Pro at the guest position was quite different and indicated that this protein is significantly unfolded, even at low temperature. We were therefore unable to determine a

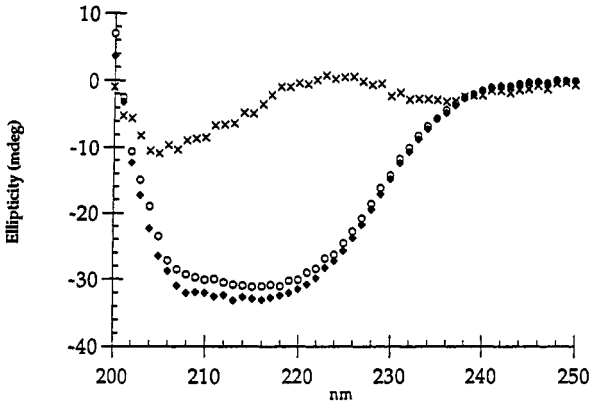

FIGURE 7: Comparison of the CD spectra of β1 with Thr, Gly, or Pro as the guest residue. Ellipticity is plotted against wavelength for protein samples of equal concentration (approximately 0.5 mg/mL). The measurements were made at 1 °C. Open circles (○), Gly at the guest site; closed diamonds (♦), Thr at the guest site; diagonal crosses (×), Pro at the guest site.

meaningful $\Delta\Delta G$ for the protein with Pro at the guest position.

DISCUSSION

We have described a thermodynamic study which demonstrates that there are experimentally measurable differences in how readily each amino acid is able to adopt a β-sheet conformation. Through NMR structural studies, we have shown that the structural integrity of the β-sheet region surrounding the guest site has not been compromised as a result of the mutations. The rank order of preferences obtained here clearly shows a broad correlation with the statistical values of β-sheet formation derived from surveys of proteins of known structure. It is important to note, however, that the Chou–Fasman probability values (Chou & Fasman, 1974, 1978) are averaged over all possible β-sheet environments, such as middle and edge positions as well as partially and fully hydrogen bonded positions. In contrast, the β1 propensity measurements are made in the most central, fully hydrogen bonded β-sheet environment possible. Hence, the lack of a one-to-one correlation between our rankings and the Chou–Fasman probabilities (Chou & Fasman, 1974, 1978) is not unexpected and may reflect both the influence of the averaging in the statistical survey data and also the very small differences between the sheet forming propensity of the amino acids in the middle range.

The range of free energy differences observed between the best and worst β-sheet forming residues is 2.5 kcal/mol, which is substantially greater than would be predicted from the Chou–Fasman probabilities (Chou & Fasman, 1974, 1978) and is also greater than the range of approximately 0.2 kcal/mol measured in the zinc finger study (Kim & Berg, 1993). However, it is possible that different positions in a β-sheet may have different conformational requirements. Hence, the predicted range of the Chou–Fasman probability rankings (Chou & Fasman, 1974, 1978) reflects the averaging of these different conformational stringencies and, as a result, would not necessarily correlate with the range determined in a study which targets a single sheet environment such as this one. In the Kim and Berg study (1993), a non-hydrogen-bonded, edge strand was targeted as the guest site. This guest site may be able to accommodate poor β-sheet forming residues more easily than a one in a central sheet position because it is not as fully constrained by hydrogen bonds. Furthermore, it is possible that, as a peptide, the zinc finger host is able to accommodate distortions caused by poor β-sheet formers more readily than is a stable protein. In this regard, we note that in α-helical propensity studies, the range of free energies observed for protein-based is greater than peptide-based hosts. The location of the guest site and the model system used in the Kim and Berg study (1993) may have contributed to the more narrow range of propensities than found in this study.

Thus, from statistical distributions and experimental studies of both α-helix and β-sheet propensities, it seems clear that there are intrinsic differences in how readily each amino acid can adopt helical or sheet structure. The structural basis for these differences is, however, far from clear. Two principal theories have been proposed to explain the experimentally observed α-helical propensities, the first of which is that side chain entropy is a dominant factor in determining propensities. When comparing side chain rotamer distributions in the α-helix versus extended, unfolded polypeptide, certain values of χ_1 are not populated by, for example, β-branched amino acids. It is therefore entropically costly for these amino acids to adopt a helical conformation, and this cost results in low helical propensities (Creamer & Rose, 1992). Secondly, an important factor contributing toward helical propensity is the hydrophobic effect. In an α-helix, the solvent accessibility of a given side chain is partially restricted by backbone atoms in the next turn of the helix. Amino acids such as Ala, Leu, and Met are able to shield a larger amount of hydrophobic surface from solvent when they adopt a helical conformation than when they are extended in a "random coil" (Blaber et al., 1993; Hermans et al., 1992; Horovitz et al., 1992). Consequently, these amino acids have high helical propensities.

In contrast, the difference between the environment an amino acid experiences in an extended β-sheet as opposed to an extended "random coil" is not as obvious. Although statistical surveys of natural proteins (Chou & Fasman, 1974, 1978) reveal that the amino acids most commonly found in β-sheets are the hydrophobic, β-branched residues, β-sheets are more often found fully buried than are α-helices. Hence, prior to the results of this study, one might have argued that the observed statistical distribution reflects only a hydrophobic requirement. Apparently, these statistical frequencies must also reflect the high propensity of these amino acids for a β-sheet conformation.

Future studies will focus upon investigating the underlying structural and energetic basis of the observed propensity differences through a more detailed characterization of the folded and unfolded forms of β1 containing different guest site substitutions.

ACKNOWLEDGMENT

We thank Julian Sturtevant and Yu-Feng Lin for performing the DSC measurements. We thank Kevin Gardner for his invaluable assistance in obtaining initial NMR spectra and his help in their interpretation. We thank Angela Gronenborn for providing the NMR assignments of wild-type β1 and for several useful and informative discussions. We thank Mark Lemmon for his generous gift of pET11aΔ. We thank Alice Marcy for technical advice concerning incomplete processing by methionine amino peptidase and Gary Schoenhals for advice concerning protein expression. We thank Joe Coleman, Kevin Gardner, Mike Klemba, Mary Munson, Fred Richards, Steve Smith, Gary Schoenhals, and Julian Sturtevant for helpful discussions and for comments on the manuscript. We thank Peter Kim for sharing with us the unpublished results from a similar study of β-sheet propensities. Finally, we thank Jim Matthew for his interest

and support. L.R. is a NSF National Young Investigator; C.K.S. is a NIH predoctoral fellow.

REFERENCES

Alexander, P., Orban, J., & Bryan, P. (1992a) *Biochemistry 31*, 7243–7248.
Alexander, P., Fahnestock, S., Lee, T., Orban, J., & Bryan, P. (1992b) *Biochemistry 31*, 3597–3603.
Anil-Kumar, R. R., Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1–6.
Blaber, M., Zhang, X.-J., & Matthews, B. W. (1993) *Science 260*, 1637–1640.
Boissel, J., Kasper, T. J., & Bunn, H. F. (1988) *J. Biol. Chem. 263*, 8443–8449.
Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–528.
Chou, P. Y., & Fasman, G. D. (1974) *Biochemistry 13*, 211–222.
Chou, P. Y., & Fasman, G. D. (1978) *Adv. in Enzymol. Relat. Areas Mol. Biol. 47*, 45–148.
Creamer, T. P., & Rose, G. D. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 5937–5941.
Fahnestock, S. R., Alexander, P., Nagle, J., & Filpula, D. (1986) *J. Bacteriol. 167*, 870–880.
Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T., & Clore, G. M. (1991) *Science 253*, 657–661.
Hartman, R., Schawaner, R. C., & Hermans, J. (1974) *J. Mol. Biol. 90*, 415–429.
Hawkes, R., Grutter, M. G., & Schellman, J. (1984) *J. Mol. Biol. 175*, 195–212.
Hermans, J., Anderson, A. G., & Yun, R. H. (1992) *Biochemistry 31*, 5646–5653.
Horovitz, A., Matthews, J. M., & Fersht, A. R. (1992) *J. Mol. Biol. 227*, 560–568.
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.
Kemp, D. S. (1990) *Trends Biotechnol. 8*, 249–255.
Kim, C. A., & Berg, J. M. (1993) *Nature 362*, 267–270.
Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 397–400.
Lemmon, M. A., Flanagan, J. M., Hunt, J. F., Adair, B. D., Bormann, B. J., Dempsey, C. E., & Engelman, D. M. (1992) *J. Biol. Chem. 267*, 7683–7689.
Lyu, P. C., Liff, M. I., Marky, L. A., & Kallenbach, N. R. (1990) *Science 250*, 669–671.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Marion, D., Ikura, M., & Bax, A. (1989) *J. Magn. Reson. 84*, 425–428.
Mutter, M., & Altman, K.-H. (1985) *Int. J. Pept. 26*, 373–380.
Padmanabhan, S., Marquesse, S., Ridgeway, T., Laue, T. M., & Baldwin, R. L. (1990) *Nature 344*, 268–270.
Piantinin, U., Sørensen, O. W., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800–6801.
Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, NY.
Sanger, F. S., Nicklen, A. R., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5470.
Scheraga, H. A. (1978) *J. Pure Appl. Chem. 50*, 315–324.
Shaka, A. J., & Freeman, R. (1983) *J. Magn. Reson. 77*, 274–293.
Shaka, A. J., Lee, C. J., & Pines, A. (1988) *J. Magn. Reson. 51*, 169–173.
Sharp, P. M., & Li, W.-H. (1987) *Nucleic Acids Res. 15*, 1281–1295.
Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60–89.
Sueki, M., Lee, S., Powers, S. P., Denton, J. B., Konishi, Y., & Scheraga, H. A., (1984) *Macromolecules 17*, 148–155.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley, New York.