Importance of Environment in Determining Secondary Structure in Proteins†

D. Vincent Waterhous and W. Curtis Johnson, Jr.*

Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331-7305

Received April 30, 1993; Revised Manuscript Received December 1, 1993®

ABSTRACT: We report here the effect of bulk solvent environment on the secondary structure of several peptides. In previous work, equivocal peptide sequences that are predicted to be α-helical from amino acid preference but are found to be β-strand in their proteins were shown to be α-helical in alcohol solvents and β-strand in nonmicellar sodium dodecyl sulfate (SDS) by circular dichroism (CD) spectroscopy [Zhong, L., & Johnson, W. C., Jr. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 4462–4465]. Here we show that equivocal sequences that are predicted to be β-strand but are found to be α-helical follow the same pattern; they are α-helical in alcohol solvents and β-strand in nonmicellar SDS. Furthermore, we investigated a control sequence with only a strong α-helical propensity and a control sequence with only a strong β-strand propensity. Both of these well-behaved sequences followed the same pattern as the equivocal sequences. The exceptionally stable $Y(EAAAK)_3A$ is an α-helix in all solvents, but analyses of the CD spectra indicate the loss of helix with an increase in β-strand and other structures on changing solvent from trifluoroethanol (TFE) to SDS, similar to the other peptides. We find that solvent is a very important factor in determining the secondary structure of an amino acid sequence *in vitro* and can override the propensity for a secondary structure due to sequence. This implies that the microsolvent seen by a secondary structure due to nonlocal interactions of amino acids from the tertiary structure of a protein, which we call environment, may be an important factor in determining the secondary structure of peptides and therefore should be considered to correctly predict the secondary structure of an amino acid sequence in proteins.

The sequence of a protein clearly determines its native structure, and sequence, in the form of local interactions, has been a popular factor for predicting secondary structure. Much of the research effort that has focused on *a priori* prediction of protein secondary structures from their primary sequences has been developed from the statistics of known protein structures (Chou & Fasman, 1978; Burgess et al., 1974; Lim, 1974; Garnier et al., 1978) or comparison of homologous sequences (Pongor & Szaley, 1985; Sweet, 1986; Nishikawa & Ooi, 1986; Levin et al., 1986; Zvelebil et al., 1987). These methods are rather successful, and this has stimulated the continued research effort into determining what factors control the folding of a sequence of amino acids into a globular protein. This effort has resulted in the development of some well-designed model proteins, which fold into predicted secondary (Lau et al., 1984; Eisenberg et al., 1986; Ho & DeGrado, 1987; Marqusee & Baldwin, 1987; Richardson & Richardson, 1987; Lyu et al., 1989), tertiary (Ho & DeGrado, 1987), and quarternary (Hill et al., 1990) structures *de novo*. Such successes make many researchers confident that we are solving the protein folding problem, and therefore that we will indeed be able to crack the second half of the genetic code.

Current *a priori* methods are about 70% successful at predicting secondary structure from primary sequence; this is much better than mere percentages imply, because many of the errors are at the ends of correctly predicted secondary structures. However, some sequences are predicted in one secondary structure from amino acid preferences but are found in another secondary structure in the protein. These are interesting sequences because they are demonstrated failures of the prediction methods. Furthermore, the prediction methods have not improved substantially since the pioneering work of Chou and Fasman (1974a,b), despite the increasing size of the structural data bank (Hayward & Collins, 1992). Since the information extracted by such statistical methods considers explicitly only local amino acid interactions (amino acids that are near neighbors in the primary sequence), the effect of nonlocal interactions that create a variable microsolvent in a protein, which we call environment, is averaged out in these algorithms. An important rationale for this research is that the specific environment seen by an amino acid may be important in determining secondary structure. Perhaps the microsolvent environment created by the tertiary structure of a protein (due to nonlocal amino acid interactions) must be considered explicitly when predicting secondary structure. Certainly it would be no trick for a polymer that can create an active site to also control the microsolvent environment around a peptide and thus influence its secondary structure. This would make the exact prediction very complicated, and would explain why current algorithms are not 100% successful. A discussion of this problem can be found in a recent paper by Rao et al. (1993).

The hierarchical model of protein folding has guided much of the initial direction and provided a useful framework for prediction of protein structure. This model states that an unfolded primary sequence of amino acids folds into a small number of regular and periodic secondary structures that subsequently fold into a specific tertiary conformation. This hierarchical model, however, seems too simple to account for all the data. A nonhierarchical approach has recently been proposed by Dill (1990). In his model the driving force for protein folding is the hydrophobicity of the amino acid sequence, and local secondary structures are a result of local hydrophobic and steric constraints, rather than vice versa. The local regular secondary structures thus formed provide a compact assembly of the amino acids, which are further stabilized by hydrogen bonding and other local forces.

† This work was supported by PHS Grant GM 21479 from the National Institutes of Health.

* To whom correspondence should be addressed.

® Abstract published in *Advance ACS Abstracts,* February 1, 1994.

0006-2960/94/0433-2121$04.50/0

What factors actually control the coalescing of the many possible sequences of amino acids into a limited number of regular secondary structures? Clearly, the local interactions due to sequence are the primary factor in determining the secondary structure of a peptide in a protein. However, it has been known for more than half a century that the conformation of repeating polypeptides and proteins is profoundly influenced by bulk solvent [see, for instance, Fasman (1967)]. The problem investigated here is whether the variation in microsolvent environment created by nonlocal amino acids in the tertiary structure of proteins could be another important factor in determining the secondary structure of a peptide and could account for the failure in some instances of algorithms that predict secondary structure from primary sequence alone. This hypothesis was tested *in vitro* in a previous work (Zhong & Johnson, 1992) by placing peptide sequences found in proteins in different organic solvents. Bulk solvent effects can be compared to the change in environment for many of the amino acids due to protein folding into a globular structure, although the environment in a folded protein might best be defined as a microsolvent. In the previous work, three equivocal peptide sequences that are predicted by the Chou and Fasman (1978) algorithm to be α-helical but are found to be β-strand in the crystallographic structure of their respective proteins were shown to be sensitive to solvent environment. The predicted α-helical structure was recovered in alcohol solvents; the β-strand structure of the sequence in the protein was produced in hydrophobic solvents, most reliably in nonmicellar SDS. This initial result substantiated our hypothesis that variation in the environment can account for the failures in predicting secondary structure from local primary structure.

In this work we extend our study and test the effect of bulk solvent on the opposite type of equivocal peptide sequence, sequences that are predicted to be β-strand with the Chou–Fasman algorithm (1978) but are found to be α-helical in the crystallographic structure of the protein. In order to understand to what extent we can generalize this effect, we also test three "well-behaved" peptide sequences, sequences that have a strong propensity for only a single secondary structure in proteins.

In general, we find that solvent can overcome the propensity for a secondary structure that is due to sequence. However, we find no single solvent that always reproduces the correct secondary structure of a peptide segment from a protein. This work is germane to determining whether, by measurement of the CD spectra of a peptide in two different solvents, one can decide if the sequence forms an α-helix, β-strand, or neither one in the protein from which it is derived, as predicted recently by others (Segawa et al., 1991; Luidens et al., 1993).

EXPERIMENTAL PROCEDURES

Peptide Sequences. Proteins in the Kabsch and Sander (KS) data base (Kabsch & Sander, 1983) were analyzed using the Chou–Fasman (CF) algorithm with a 64-protein data base (Chou, 1989), based on the program recently published by Prevelige and Fasman (1989). The average propensities for tetrads of amino acids in α-helices or β-strands, $\langle P_\alpha \rangle$ and $\langle P_\beta \rangle$, respectively, are given as output from the program. These values for the entire peptide sequence were then averaged for each structure ($\langle P_\alpha \rangle_{av}$ and $\langle P_\beta \rangle_{av}$) and were compared to determine the appropriate predicted secondary structure. The other standard CF rules were followed for determination of the predicted secondary structure (Prevelige & Fasman, 1989).

Comparison was made between the secondary structure predicted by the CF method and the structure assigned in the KS data base from X-ray diffraction data. Equivocal peptide sequences are those that have different secondary structures in the KS data base than those predicted by the CF algorithm. Peptide sequences were chosen on the basis of (1) having enough hydrophilic residues to be reasonably soluble, (2) being longer than 12 amino acids to help stabilize secondary structures, (3) having one aromatic amino acid to use for determining concentration, and (4) having no more than one aromatic amino acid to minimize effects on the circular dichroism, and subsequently on the analysis of this data. Equivocal sequences that are predicted to be β-strand but identified as α-helix in the protein are quite common. We chose three such sequences: 286–305 from carboxypeptidase A (CPA) (Lipscomb et al., 1968), 119–136 from horse hemoglobin (HHM) (Ladner et al., 1979), and 133–151 from thermolysin (TRM) (Colman et al., 1972).

Three additional sequences were studied: a synthetic sequence that has a large $\langle P_\alpha \rangle_{av}/\langle P_\beta \rangle_{av}$ ratio and is used in our laboratory to analyze α-helix potential, $(VAEAK)_3$; a sequence that is predicted and found (Brayer et al., 1979) to be β-strand, residues 108–122 from the protein α-lytic protease (αLP); and a sequence that is a highly stable α-helix developed by Marqusee and Baldwin (1987) and modified by them and others (Marqusee et al., 1989; Merutka et al., 1990), $Y(EAAAK)_3A$.

Peptide Synthesis and Purification. Peptide sequences were synthesized by solid-phase methods on an automated Applied Biosystems Model 431A peptide synthesizer using N^α-[9-fluorenyl(methoxycarbonyl)]- (FMOC-) protected amino acids. Terminal ends were blocked on most of the peptides studied by addition of an acetyl functional group to the N-terminus and an amine functional group to the C-terminus, as indicated in Table 1.

The crude peptides were purified by HPLC using a Hewlett-Packard LiChrospher 100 RP-18 reversed-phase analytical column (125 × 4 mm) or a Vydac C-18 reversed-phase semipreparative column (25 × 1 cm). Various elution gradients were used for purification of each peptide by modifying the percentage of acetonitrile (with 0.1% TFA) added to distilled water (with 0.1% TFA). Peptide elution was monitored at either 214 or 280 nm using a Hewlett-Packard 1040A diode-array detection system. The largest peak was collected and lyophilized. The purified peptide was verified as the correct peptide by amino acid analysis on a Beckman 126AA System Gold amino acid analyzer.

Stock Solutions. Purified samples of the peptides were lyophilized and then dissolved in deionized distilled water, ddH_2O (18-MΩ resistance), at a concentration of about 1 mg/mL. These samples were then dialyzed in Spectra/Por CE tubing with a molecular mass cutoff of 1000 Da and placed for a minimum of 24 h in ddH_2O, with three changes of ddH_2O made. The samples were removed and frozen in small plastic conical tubes as stock solutions.

[1] Abbreviations: αLP, α-lytic protease; CD, circular dichroism; CF, Chou–Fasman; CPA, carboxypeptidase A; ddH_2O, distilled deionized water; GdnHCl, guanidine hydrochloride; HHM, horse hemoglobin; HPLC, high-performance liquid chromatography; KS, Kabsch–Sander; MOPS, 3-(*N*-morpholino)propanesulfonic acid; OβG, octyl β-D-glucopyranoside; $\langle P_\alpha \rangle$, average of α-helix propensity for tetrads; $\langle P_\beta \rangle$, average of β-strand propensity for tetrads; $\langle P_\alpha \rangle_{av}$ and $\langle P_\beta \rangle_{av}$ average of the α-helix and β-strand propensity for tetrads across the whole sequence; SDS, sodium dodecyl sulfate; SVD, singular value decomposition; TFA, trifluoroacetic acid; TFE, 2,2,2-trifluoroethanol; UV-CD, ultraviolet circular dichroism; VS-SVD, variable selection-singular value decomposition; VUV-CD, vacuum ultraviolet circular dichroism.

Table 1: Sequences, Average Propensities for α-Helix and β-Strand,[a,b] and Molar Extinction Coefficients of the Peptides Studied[c]

sequence	residues	notation	P_α	P_β	ϵ190 (M^{-1} cm^{-1})
H_2N-I-I-P-T-A-Q-E-T-W-L-G-V-L-T-I-M-E-H-T-V-COOH	286–305	CPA	102	110	7600
acetyl-L-S-G-G-I-D-V-V-A-H-E-L-T-H-A-V-T-D-Y-amide	133–151	TRM	103	108	7800
HN-P-A-V-H-A-S-L-D-K-F-L-S-S-V-S-T-V-L-amide	119–136	HHM	102	105	8600
acetyl-G-Y-Q-C-G-T-I-T-A-K-N-V-T-A-N-amide	108–122	αLP	94	109	9300
acetyl-V-A-E-A-K-V-A-E-A-K-V-A-E-A-K-amide		$(VAEAK)_3$	128	87	8400
acetyl-Y-E-A-A-A-K-E-A-A-A-K-E-A-A-A-K-A-amide		$Y(EAAAK)_3A$	135	74	8150

[a] P_β and P_α values shown are the average $\langle P_\beta \rangle$ and $\langle P_\alpha \rangle$ values for the sequences as shown. [b] Underlined portions of the peptide sequences are where $\langle P_\beta \rangle$ is greater than $\langle P_\alpha \rangle$ for the sequences as found in their respective proteins. [c] Molar extinction coefficients of peptide sequences in 10 mM sodium phosphate, pH 7.2, on a per-amide basis.

Peptide Concentration. The concentration of the stock solution of a particular peptide was monitored by absorption and a knowledge of the molar absorptivity constant, ϵ, for the amide absorption at 190 nm in 10 mM sodium phosphate buffer. The ϵ_{190} was determined by calculating the concentration of a stock solution from measurement of the absorbance of that peptide in 6 M guanidine hydrochloride, on a Cary 15 spectrophotometer, and using the known extinction coefficient at 280 nm for the aromatic amino acids in that peptide (Elwell & Schellman, 1977). The concentration of the stock solution of the peptide $(VAEAK)_3$ was determined by amino acid analysis, as described above, since this peptide contains no aromatic chromophores to use for concentration determination by the guanidine hydrochloride method. The ϵ_{190} values on a per-amide basis are given in Table 1.

Solvents. All organic solvents used in these experiments were spectrophotometric or HPLC grade. Sodium phosphate buffer was made from Baker monobasic sodium phosphate, with pH adjusted to 7.2 using a solution of Aldrich semiconductor-grade sodium hydroxide, and was filtered through a Millipore 0.22-μm filter. A stock solution of sodium dodecyl sulfate (SDS) was made with Bio-Rad electrophoresis-grade reagent in ddH_2O, stored at 4 °C, and diluted to the proper concentration with ddH_2O, as needed. Stock solutions of 50 mM octyl β-glucoside (OβG) were made using octyl 1-*O*-(β-D-glucopyranoside) (98%) purchased from Aldrich and dissolved in ddH_2O. A MOPS solution containing 25 mM MOPS (pH 7.4) and 1 M sucrose was made using 50 mM 3-(*N*-morpholino)propanesulfonic acid, from Research Organics, Inc. (pH adjusted with 6 N HCl) and ultrapure, density gradient grade, sucrose (Schwarz/Mann Biotech) with final adjustment of the solution with ddH_2O to yield the correct concentration. A solution of 0.08% (w/v) digitonin and 0.016% (w/v) cholate (dig/cholate) was made by dissolving cholate (commercial cholate was repurified by chromatography and precipitation) and digitonin (Sigma) in 10 mM sodium phosphate buffer (pH 7.0). Guanidine hydrochloride (GdnHCl) solutions were made using ultrapure (>99%) material from Schwarz/Mann Biotech, dissolved in ddH_2O.

Circular Dichroism Measurements. Circular dichroism (CD) spectra were measured on a vacuum ultraviolet (VUV) CD spectropolarimeter (Johnson, 1971) to 178 nm or on a Jasco-J40 CD spectropolarimeter to 200 nm. The instruments were calibrated daily using a two-point calibration method (Chen & Yang, 1977) with (+)-10-camphorsulfonic acid, $\Delta\epsilon_{290.5} = 2.37$ M^{-1} cm^{-1} and $\Delta\epsilon_{192.5} = -4.95$ M^{-1} cm^{-1}. The cell path lengths were 50 or 100 μm for VUV-CD measurements and 200 μm to 2 cm for other CD measurements. Samples were run at 20 °C unless otherwise noted. Temperature was controlled by a thermoelectric cell holder. Data were collected every 0.5 or 1.0 nm on a computer and digitally stored. Spectra were smoothed using a cubic spline algorithm and are presented in $\Delta\epsilon$ units on a per-amide basis.

Aggregation. Concentration dependence of the secondary structure of these peptides was determined by changes of the CD spectra between 260 and 205 nm using the Jasco spectropolarimeter. The CD spectra of CPA and HHM in 90% TFE were found to be concentration-dependent, so aggregation helps stabilize the helical structure. The other four peptides had typical α-helix CD spectra in 90% TFE that were independent of peptide concentration. These four peptides were not further tested for concentration dependence of the CD spectra in 90% TFE at different temperatures; therefore, the possibility exists that a tight dimer may form that may stabilize the helical structure of these peptides. The β-strand-like CD of all the peptides in nonmicellar SDS depends on the ratio of SDS to peptide, and thus concentration cannot be used as a test for aggregation.

Analysis for Secondary Structure. The VUV-CD spectra were analyzed for secondary structure with an algorithm using singular value decomposition (SVD) combined with variable selection (VS) (Manavalan & Johnson, 1987) and a 33-protein basis set containing 11 β-proteins. The criteria for determining acceptable predictions of these peptide secondary structures are as described elsewhere (Toumadje, et al., 1992).

The VS-SVD basis set is predominantly protein data. It may not be the best basis set to estimate the secondary structure formed in peptides. We also estimate the percentage α-helix from $\Delta\epsilon$ at 222 nm and the percentage β-strand from overall CD magnitude, as previously described (Zhong & Johnson, 1992), which may be more accurate for structures assumed by short peptides.

RESULTS

Equivocal Peptides. The three equivocal peptides (CPA, TRM, and HHM) contain 17 of the 20 common amino acids. Each of their respective $\langle P_\beta \rangle_{av}$ values is greater than their $\langle P_\alpha \rangle_{av}$ value (Table 1), and they meet the other criteria to be predicted β-strands (Fasman, 1985), although they are found as α-helices in their respective protein environments. Figure 1 clearly shows that changing the bulk solvent can elicit either α-helix or β-strand secondary structures for these equivocal peptide sequences. The spectra of CPA, TRM, and HHM in 90% TFE are typical of α-helices. These spectra have two bands with negative ellipticity centered at 222 and 208 nm and a positive band centered near 190 nm. The spectra of these three peptides in nonmicellar SDS indicate β-structure. CPA and HHM show two bands: a negative band is centered at 216 nm and a positive band is centered near 194 nm, characteristic of a β-strand. The spectrum of TRM (Figure 1b) in nonmicellar SDS is not a typical β-strand CD spectrum. The spectrum displays two negative bands and a positive band of nearly twice the intensity of the negative band, like an α-helix. However, the first negative band is centered at 216 rather than 222 nm, similar to a β-strand. Each of these peptides has a typical random structure CD in buffer, as is

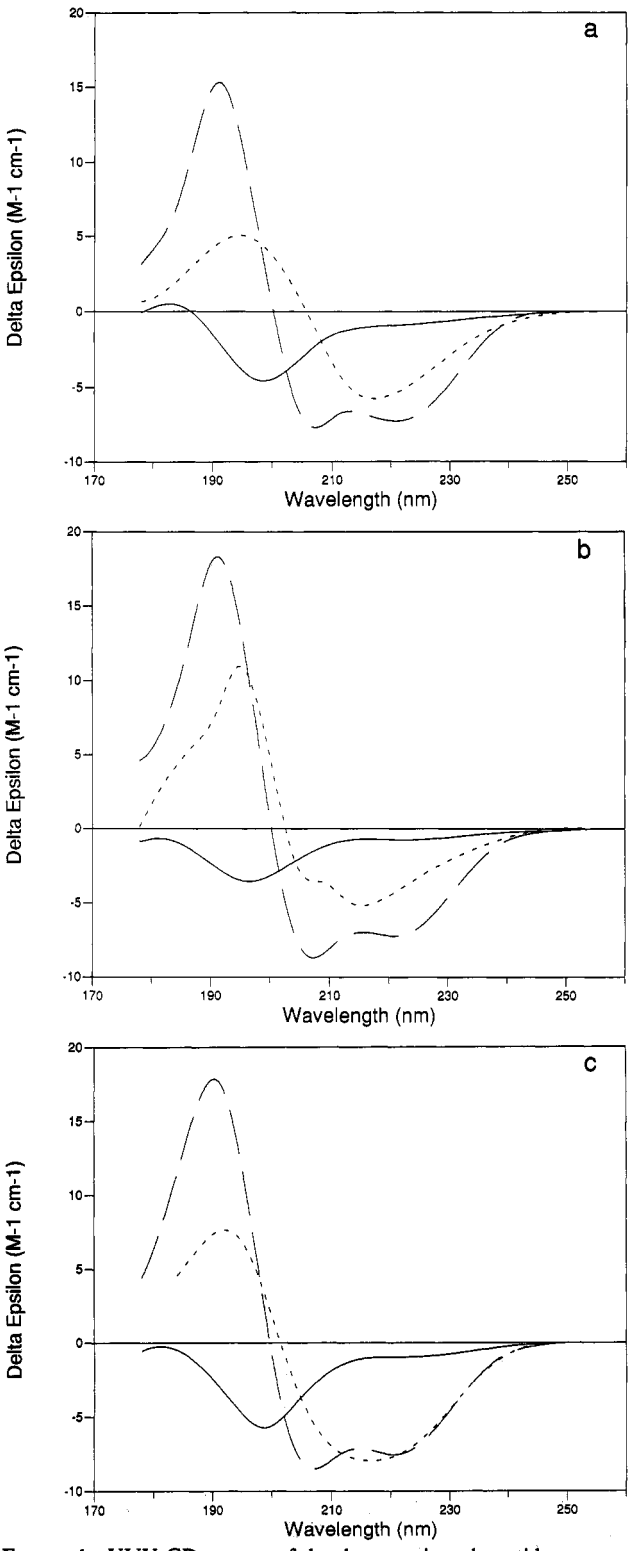

FIGURE 1: VUV-CD spectra of the three equivocal peptide sequences: (a) CPA in (—) 10 mM sodium phosphate, pH 7.2, (– – –) 90% (v/v) TFE, and (- - -) 6 mM SDS; (b) TRM in (—) 10 mM sodium phosphate, pH 7.2, (– – –) 90% (v/v) TFE, and (- - -) 6 mM SDS; (c) HHM in (—) 10 mM sodium phosphate, pH 7.2, 90% (v/v) TFE, and (- - -) 2 mM SDS at 52 °C. $\Delta\epsilon$ units on a per-amide basis expressed as M^{-1} cm^{-1}.

predicted for short peptides (Bierzynski et al., 1982; Shoemaker et al., 1985). These results parallel those reported for the converse equivocal peptide sequences, which are predicted to be α-helices but are found as β-strands in their respective proteins (Zhong & Johnson, 1992). That set of equivocal peptides are also found to be α-helical in TFE, β-strand in nonmicellar SDS, and randomly structured in aqueous buffer.

Well-Behaved Peptides. Two sequences that are well-behaved, in the sense that they have a strong propensity for only a single secondary structure, were chosen as controls for the *in vitro* studies. Table 1 lists the sequences chosen as the β-strand control, αLP, and the α-helix control, $(VAEAK)_3$. The data in Table 1 clearly indicate that αLP has a strong propensity to be a β-strand and no propensity for α-helix. This peptide is found as a β-strand in its native protein environment (Brayer et al., 1979). $(VAEAK)_3$ is a peptide sequence designed in our laboratory to study α-helix nucleation and propagation. Table 1 indicates that this sequence has a strong propensity to be α-helical in structure with no propensity to be a β-strand. Figure 2a shows the CD of αLP in 90% TFE, where this peptide has an α-helix spectrum, even though αLP clearly would not be predicted by the Chou–Fasman algorithm to form an α-helix structure. Figure 2b shows that the CD of $(VAEAK)_3$ in 2 mM SDS is a typical β-strand spectrum, yet this peptide sequence clearly would not be predicted by the Chou–Fasman algorithm to form a β-strand. Indeed, these controls behave exactly like the equivocal peptides: they are α-helix in TFE, β-strand in nonmicellar SDS, and random coil in buffer. The secondary structures formed for these peptides seem independent of the propensity in proteins, as predicted by Chou–Fasman, but they are dependent on the bulk solvent.

Finally, we chose $Y(EAAAK)_3A$ as an exceptionally stable structure which tends to be an α-helix, even in an aqueous buffer environment (Marqusee & Baldwin, 1987; Marqusee et al., 1989; Merutka et al., 1990). The solvent effects on $Y(EAAAK)_3A$ are summarized in Figure 2c; this peptide maintained an α-helix conformation in all solvents tested. The percent helix for this peptide in buffer was calculated to be 31% as determined from variable selection SVD (Table 2) or 39% as estimated from the $\Delta\epsilon_{222}$ (Zhong & Johnson, 1992). This range is in good agreement with the 38–44% helix estimated by the $\Delta\epsilon_{222}$ from previously reported results in phosphate buffer for this sequence (Merutka et al., 1990) and a similar sequence, $A(EAAAK)_3A$ (Marqusee & Baldwin, 1987; Marqusee et al., 1989).

The general trend of bulk solvent on the secondary structure of these peptides, and the three peptides previously studied (Zhong & Johnson, 1992), is that alcoholic solvents and micellar SDS promote α-helices, detergents and nonmicellar SDS promote β-strands, and random coil structures predominate in aqueous buffer. The secondary structure formed in acetonitrile is variable. The notable exception to this trend was $Y(EAAAK)_3A$, which remained helical in all solvents that we tested.

Quantitative Analysis of CD for Secondary Structure. Analysis of these CD spectra by VS-SVD (Table 2) is consistent with a qualitative visual analysis of the observed spectra. Each of the α-helical spectra analyze as helix with turns and little or no β-structure. The β-structure CD spectra analyze in a variable manner, as is found for β-proteins. These spectra analyze as having some helix component due to their negative ellipticity at 222 nm, yet none of these β-spectra have the two negative peaks at 222 and 208 nm, which characterize α-helices. Interestingly, the β-spectrum for TRM is reminiscent of an α/β protein, which has parallel β-strands interspersed by α-helices similar to the β–α–β motif predicted for this peptide (Table 1). The random structure spectra analyze as having little helix and a mix of β-strand, turns, and other structure. This is what should be expected for random structures, since random structures are actually a heteroge-

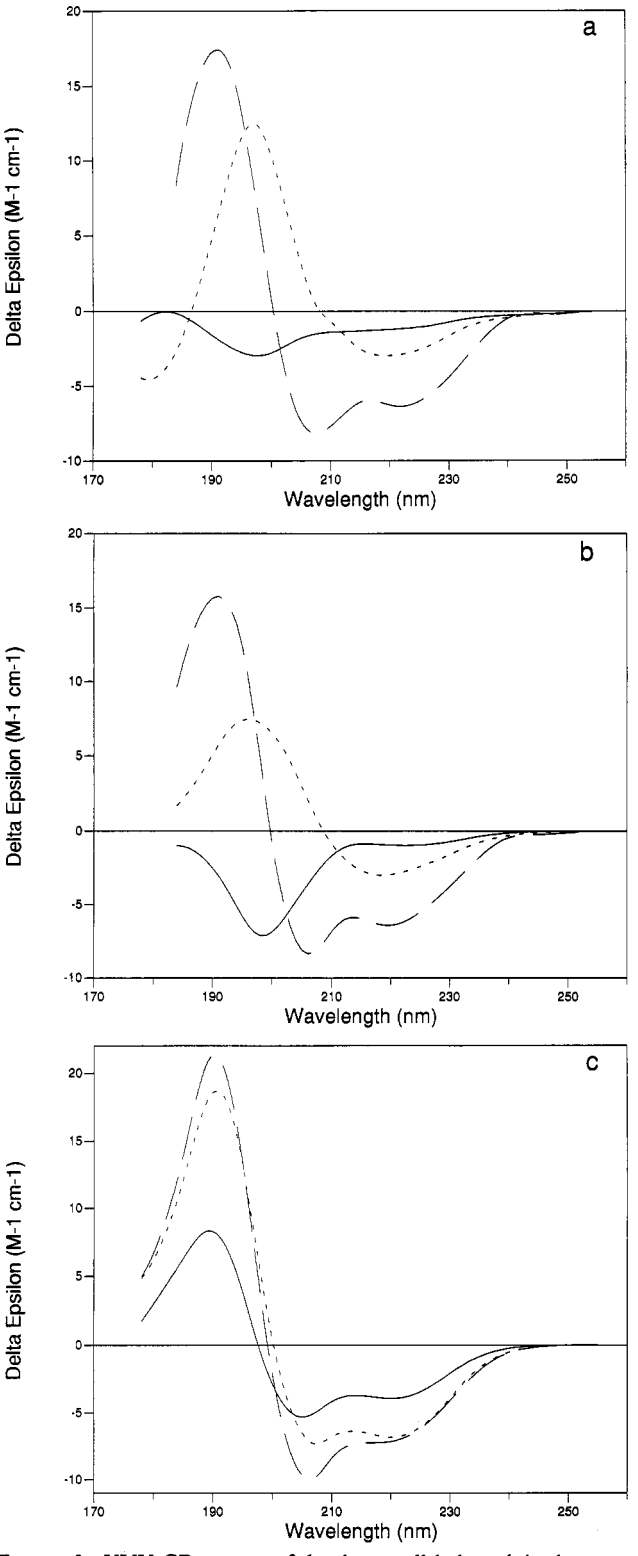

FIGURE 2: VUV-CD spectra of the three well-behaved, in the sense that they have a strong propensity for only a single secondary structure, peptide sequences. (a) αLP in (—) 10 mM sodium phosphate, pH 7.2, (– – –) 90% (v/v) TFE, and (- - -) 6 mM SDS; (b) $(VAEAK)_3$ in (—) 10 mM sodium phosphate, pH 7.2, (– – –) 90% (v/v) TFE, and (- - -) 2 mM SDS; (c) $Y(EAAAK)_3A$ in (—) 10 mM sodium phosphate, pH 7.2, (– – –) 90% (v/v) TFE, and (- - -) 2 mM SDS. $\Delta\epsilon$ units on a per-amide basis expressed as M^{-1} cm^{-1}.

neous population of low-energy conformers of the peptide sequence, and the ϕ–ψ angles typical of β-strands and turns are among this population of low-energy conformers.

Table 3 shows that the percent helix estimated from the $\Delta\epsilon_{222}$ is the same or somewhat larger than the estimate from VS-SVD (Table 2). Estimation of percent β-strand from the overall CD magnitude predicts much higher percentages of β-strand in these peptides (Table 3) relative to the VS-SVD method (Table 2). It is very likely that VS-SVD underestimates the percentage β-strand, since it strongly weighs in an α-helix component to fit the ellipticity at 222 nm, even if there is no distinct band at that wavelength (Table 2). It is also likely that the estimation of β-strand from the overall magnitude of the CD, as shown in Table 3, will be overestimated if there is substantial α-helical character in the peptide that increases the intensity of the 190-nm band.

DISCUSSION

Comparison of Predicted Structure for Peptides. It is interesting to look at these equivocal peptides with regard to the length, location, and dispersion of their secondary structure motifs. In Table 1 the segments of the peptide where the Chou–Fasman $\langle P_\beta \rangle$ values are greater than their $\langle P_\alpha \rangle$ values are underlined. The segments that are not underlined have α-helical propensity, since these regions have $\langle P_\alpha \rangle$ values greater than the $\langle P_\beta \rangle$ values and satisfy the other CF criteria. However, only HHM has a predicted α-helix segment long enough to meet the criteria of being able to nucleate this structure. These three peptides thus make up an interesting series. CPA has a central β-strand nucleating region in between two regions with α-helix propensity, α–β–α. TRM has two β-strand nucleating regions on either side of a pentamer with strong α-helix propensity, β–α–β. HHM is nearly evenly divided into an α-helical nucleation and propagation segment at the N-terminus and a β-strand nucleation and propagation segment at the C-terminus, α–β. The two well-behaved sequences, αLP and $(VAEAK)_3$, have propensity for only a single secondary structure, either β or α, respectively. Yet all five of these sequences form either α-helix, β-strand, or random coil dependent only on the solvent and independent of prediction for propensity, nucleation, or propagation in proteins. The same solvent dependence was obtained for three peptides that have virtually no β-strand nucleating or propagating potential even though they are found as β-strand in their respective proteins (Zhong & Johnson, 1992). Eight of nine peptide sequences that we have studied, containing various combinations of amino acids, all fold into α-helices in TFE and β-strands in nonmicellar SDS and are randomly structured in buffer.

Although $Y(EAAAK)_3A$ is α-helical in all solvents tested, it is also sensitive to the bulk solvent. The VS-SVD data for this peptide (Table 2) indicates that TFE promotes the most α-helix for this peptide: the CD spectrum analyzes as being totally helix and turn, 69% and 35%, respectively. In 2 mM SDS, where other peptides are β-strand, this peptide shows a loss of helix and turn, 64% and 21% respectively, compensated almost completely by an increase in β-strand to 14%. Finally, $Y(EAAAK)_3A$ further loses helix in buffer, to 31%, which is compensated by an increase in other structure (from 0% in TFE or SDS to 27%) while the relative amount of turns and β-strand stays almost the same (24% and 14%, respectively) as that found for this peptide in SDS. The VS-SVD method of analysis for CD spectra depends on a protein basis set, and may not be the best method for the analysis of peptides, but it quantitatively demonstrates what is qualitatively illustrated in the CD spectra, that the structures of these peptides are solvent-dependent. Monitoring $\Delta\epsilon_{222}$ in the CD spectra of $Y(EAAAK)_3A$ (Figure 2c) clearly shows the loss of α-helical structure upon changing the solvent from TFE to SDS to aqueous buffer, in agreement with the quantitative data

Table 2: Determination of the Percentage of Peptide in a Given Secondary Structure[a] by VS-SVD[b] Analysis of the CD Data

solvent	sequence	% H	% B	% T	% O	total (%)
90% TFE	CPA	59 ± 2	7 ± 2	19 ± 1	16 ± 1	101
	TRM	66 ± 1	5 ± 3	21 ± 2	9 ± 2	101
	HHM	63 ± 1	8 ± 1	13 ± 0	15 ± 1	99
	αLP	58 ± 1	11 ± 2	22 ± 1	12 ± 1	103
	$(VAEAK)_3$	69 ± 1	3 ± 2	27 ± 1	1 ± 1	100
	$Y(EAAAK)_3A$	69 ± 1	1 ± 1	35 ± 1	0 ± 1	105
2–6 mM SDS	CPA	26 ± 3	20 ± 2	5 ± 3	51 ± 5	101
	TRM	54 ± 1	15 ± 1	12 ± 1	20 ± 1	101
	HHM	35 ± 3	26 ± 3	9 ± 2	31 ± 1	101
	αLP	25 ± 2	49 ± 3	6 ± 2	22 ± 2	103
	$(VAEAK)_3$	28 ± 1	48 ± 2	0 ± 0	24 ± 1	100
	$Y(EAAAK)_3A$	64 ± 1	14 ± 2	21 ± 1	0 ± 1	99
10 mM sodium phosphate buffer, pH 7.2	CPA	10 ± 1	33 ± 1	23 ± 1	34 ± 1	100
	TRM	9 ± 1	26 ± 2	23 ± 1	42 ± 2	100
	HHM	5 ± 1	22 ± 2	29 ± 1	41 ± 1	97
	αLP	11 ± 1	24 ± 1	19 ± 1	44 ± 2	98
	$(VAEAK)_3$	1 ± 1	9 ± 2	43 ± 1	50 ± 2	102
	$Y(EAAAK)_3A$	31 ± 1	14 ± 4	24 ± 1	27 ± 2	96

[a] Abbreviations: H, α-helix; B, total β-strand, antiparallel and parallel; T, β-turn; O, other secondary structure. [b] VS-SVD: variable selection-singular value decomposition.

Table 3: Determination of the Percentage of Peptide in a Given Secondary Structure by Simple Methods[a]

sequence	$\Delta\epsilon_{222}$	% α-helix	$\Delta\epsilon_{pp}$[b]	% β-strand
CPA	−7.2	72	10.8	65
TRM	−7.2	72	16.2	98
HHM	−7.5	75	15.7	95
αLP	−6.4	64	15.5	94
$(VAEAK)_3$	−7.9	79	13.2	80
$Y(EAAAK)_3A$	−6.9	69	*c*	*c*

[a] $\Delta\epsilon_{222}$ data for determination of percent α-helix is taken from peptide structures in 90% TFE, and $\Delta\epsilon_{pp}$ data is taken from peptide structures in nonmicellar SDS. [b] $\Delta\epsilon_{pp}$, overall CD magnitude from minimum to maximum. [c] This peptide strand did not have a β-strand CD spectrum and was not estimated by this method.

obtained from VS-SVD (Table 2). However, this qualitative type of analysis does not provide any information with regard to what structures may form subsequent to the loss of the α-helical structure.

Related Research. There is other evidence for the phenomenon that a peptide sequence may adopt any of the three principal secondary conformations. Wu and Yang (1981) reported that certain natural peptides with both α- and β-potential would form either structure depending on whether SDS is above or below the micelle concentration. A study of pentapeptide sequences in the protein data base revealed that this length of peptides could not be used to predict a specific structure, since these sequences in different proteins were found in different secondary structures (Kabsch & Sander, 1984). Melittin, a 26-residue peptide found as an α-helix by X-ray crystallography, was found as α, β, or random coil conformation depending on the concentration of SDS (Kubota & Yang, 1986). Signal peptide sequences are reported to form α-helix or β-strand, depending on whether they are inserted in a micelle or bound to the surface (Cornell, et al., 1989). The flip-flop in structure between β-sheet and α-helix of gp120, the envelope protein of HIV-1, is implicated as a potential switch controlling the binding of this protein to the T-cell receptor protein CD4 (Reed & Kinzel, 1991). It has recently been reported that acid unfolding of cellular retinoic acid binding protein, a β-barrel protein, results in the development of substantial α-helix. This lead to speculation that nonnative α-helices may be important intermediates on the folding pathway of this protein (Liu et al., 1993).

Comparison of Bulk Solvent to Protein Environment. A protein can present any one of a number of microsolvent environments to a peptide. The environment will depend on the local amino acids, will depend on the nonlocal amino acids that are nearby because of tertiary structure, and will include water if the peptide is on the surface of the protein. Both the TFE and nonmicellar SDS solvents are mixed solvent systems and provide partly organic, partly aqueous environments. This seems to mimic some of the various environments that a peptide can encounter when it is part of a protein, because we find that one or the other of these two solvents will produce the unexpected secondary structure exhibited by an equivocal sequence in a protein.

TFE has a reputation for promoting α-helices (Nelson & Kallenbach, 1986, 1989; Merutka & Stellwagen, 1989). It has recently been described as helix-enhancing rather than helix-inducing, since a requirement for helix propensity was considered necessary to develop this secondary structure (Sönnichsen et al., 1992). This would seem to be a reasonable supposition, because there are reports of stable β-strands in TFE (Goodman et al., 1970; Balcerski et al., 1976; Kelly et al., 1977; Narayanan et al., 1986), thus demonstrating that peptides are not always α-helices in TFE. However, our results do not support the idea that a peptide strand must have helix propensity in proteins to be induced to a helical structure by TFE, since αLP forms 58% helix in TFE at ambient temperature, yet it has no helix propensity.

SDS is a surfactant that can provide a hydrophobic environment. Micellar SDS is well documented as a stabilizer of α-helices (Jirgensons, 1977, 1981; Wu & Yang, 1981, 1988; Wu et al., 1981; Gierasch, 1989). Our unpublished data from this and the previous work (Zhong & Johnson, 1992) support this finding. Interestingly, Yang and co-workers found that nonmicellar SDS tended to promote β-structures (Wu & Yang, 1981; Wu et al., 1981). We find that eight of the nine peptides we investigated also formed β-strands in nonmicellar SDS. This solvent is the best solvent for inducing β-strands. For example, 2 mM SDS is the only solvent that we investigated in which HHM developed a β-strand CD spectrum, and the sequence $(VAEAK)_3$, which has no β-strand propensity, has a CD spectrum in 2 mM SDS that analyzes as containing 48–80% β-strand (Tables 2 and 3). Furthermore, nonmicellar SDS not only promotes β-strand structure but also keeps this structure in solution. Most hydrophobic solvents tend to precipitate peptides, presumably because they promote β-structures that then aggregate.

The equivocal peptide sequences we have studied here and previously (Zhong & Johnson, 1992) represent the obvious exceptions to prediction methods. It is, perhaps, not surprising to be able to find a bulk solvent that mimics the protein environment for these peptides and produces the unexpected secondary structure found in the protein, but it is clear from our studies that no single solvent mimics the interior of a protein. Indeed, for some peptides we need TFE to obtain an environment that stabilizes the conformation found within the protein, while for other peptides we need nonmicellar SDS to obtain an environment that stabilizes the conformation found within the protein. The more general trend is that the bulk solvent tends to determine the conformation, almost regardless of the peptide sequence. Therefore, the idea that a single solvent is mimicking the interior of the protein, because one can achieve the same structure in that solvent as in the protein, may be fortuitous. It appears that a protein could control secondary structure through its microsolvent environment in addition to the control it achieves through local sequence.

Conclusions. These results are consistent with Dill's recent review of dominant forces in protein folding (Dill, 1990). Primary sequences collapse into a condensed structure due to nonlocal hydrophobic forces. This condensed structure can then undergo some final rearrangement of secondary structure to develop a compact structure that optimizes steric interactions, hydrophobic interactions, hydrogen bonds, and all other local forces.

Our work shows a major limitation of *a priori* protein folding prediction algorithms. These algorithms assume an average environment in the protein, but nonlocal amino acids can produce a wide range of microenvironments. Certainly the local microenvironment of a sequence may change dramatically as a result of initial global folding events. Our studies show how changing the solvent seen by a peptide can cause dramatic changes in conformation. Protein folding not only changes the local environment, removing many amino acids from the polar environment of the protein exterior; it must also introduce a microsolvent due to nonlocal amino acids that are now nearby due to tertiary structure. We, as of yet, have little insight into how the primary sequence predetermines the final microsolvent of the individual building blocks. However, we can conclude that environment effects must be part of an algorithm that correctly predicts folding of a primary sequence of amino acids into its final native structure.

ACKNOWLEDGMENT

We would like to acknowledge the support of the Central Services Laboratory of the Center for Gene Research and Biotechnology at Oregon State University for the syntheses and amino acid analyses of the peptides used for this paper.

REFERENCES

Balcerski, J. S., Pysh, E. S., Bonora, G. M., & Toniolo, C. (1976) *J. Am. Chem. Soc. 98*, 3740–3473.

Bierzynski, A., Kim, P. S., & Baldwin, R. L. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 2470–2474.

Brayer, G. D., Delbaere, L. T. J., & James, M. N. G. (1979) *J. Mol. Biol. 131*, 743–775.

Burgess, A. W., Ponnuswamy, P. K., & Scheraga, H. A. (1974) *Isr. J. Chem. 12*, 239–286.

Chen, G. C., & Yang, J. T. (1977) *Anal. Lett. 10*, 1195–1207.

Chou, P. Y. (1989) in *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G. D., Ed.) pp 549–586, Plenum Press, New York.

Chou, P. Y., & Fasman, G. D. (1974a) *Biochemistry 13*, 211–222.

Chou, P. Y., & Fasman, G. D. (1974b) *Biochemistry 13*, 222–245.

Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45–148.

Colman, P. M., Jansonius, J. N., & Matthews, B. W. (1972) *J. Mol. Biol. 70*, 701.

Cornell, D. G., Dluhy, R. A., Briggs, M. S., McKnight, C. J., & Gierasch, L. M. (1989) *Biochemistry 28*, 2789–2797.

Dill, K. A. (1990) *Biochemistry 29*, 7133–7155.

Eisenberg, D., Wilcox, W., Eshita, S. M., Pryciak, P. M., Ho, S. P., & DeGrado, W. F. (1986) *Proteins: Struct. Funct., Genet. 1*, 16–22.

Elwell, M. L., & Schellman, J. A. (1977) *Biochim. Biophys. Acta 494*, 367–383.

Fasman, G. D. (1967) in *Poly-α-Amino Acids*, Marcel-Dekker, New York.

Fasman, G. D. (1985) *J. Biosci. 8*, 15–23.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97–120.

Gierasch, L. M. (1989) *Biochemistry 28*, 923–930.

Goodman, M., Verdini, A. S., Choi, N. S., & Masuda, Y. (1970) in *Topics in Stereochemistry* (Eliel, E., Ed.) pp 69–166, N. L. Allinger.

Hayward, S., & Collins, J. F. (1992) *Proteins: Struct., Funct., Genet. 14*, 372–381.

Hill, C. P., Anderson, D. H., Wesson, L., DeGrado, W. F., & Eisenberg, D. (1990) *Science 249*, 543–546.

Ho, S. P., & DeGrado, W. F. (1987) *J. Am. Chem. Soc. 109*, 6751–6758.

Jirgensons, B. (1977) *Biochim. Biophys. Acta 473*, 352–358.

Jirgensons, B. (1981) *Makromol. Chem. Rapid Commun. 2*, 213–217.

Johnson, W. C., Jr. (1971) *Rev. Sci. Instrum. 42*, 1283–1286.

Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2577–2637.

Kabsch, W., & Sander, C. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 1075–1078.

Kelly, M. M., Pysh, E. S., Bonora, G. M., & Toniolo, C. (1977) *J. Am. Chem. Soc. 99*, 3264–3266.

Kubota, S., & Yang, J. T. (1986) *Biopolymers 25*, 1493–1504.

Ladner, R. C., Heidner, E. G., & Perutz, M. F. (1977) *J. Mol. Biol. 114*, 385.

Lau, S. Y. M., Taneja, A. K., & Hodges, R. S. (1984) *J. Biol. Chem. 259*, 13253–13261.

Levin, J. M., Robson, B., & Garnier, J. (1986) *FEBS Lett. 205*, 303–308.

Lim, V. I. (1974) *J. Mol. Biol. 88*, 857–894.

Lipscomb, W. N., Hartsuck, J. A., Reeke, G. N., Quiocho, F. A., Bethge, P. H., Ludwig, M. L., Steitz, T. A., Muirhead, H., & Coppola, J. C. (1968) *Brookhaven Symp. Biol. 21*, 24.

Liu, Z.-P., Rizo, J., & Gierasch, L. M. (1993) *J. Cell. Biochem. 17C*, 279.

Luidens, M. K., Aks, C. S., Zhu, Q-l., Smith, T. F., MacColl, R., & FIgge, J. (1993) *Pept. Res. 6*, 134–139.

Lyu, P. C., Marky, L. A., & Kallenbach, N. R. (1989) *J. Am. Chem. Soc. 111*, 2733–2734.

Manavalan, P., & Johnson, W. C., Jr. (1983) *Nature 305*, 831–832.

Manavalan, P., & Johnson, W. C., Jr. (1987) *Anal. Biochem. 167*, 76–85.

Marqusee, S., & Baldwin, R. L. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8898–8902.

Marqusee, S., Robbins, V. H., & Baldwin, R. L. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 5286–5290.

Merutka, G., & Stellwagen, E. (1989) *Biochemistry 28*, 352–357.

Merutka, G., Lipton, W., Shalongo, W., Park, S.-H., & Stellwagen, E. (1990) *Biochemistry 29*, 7511–7515.

Narayanan, U., Keiderling, T. A., Bonora, G. M., & Toniolo, C. (1986) *J. Am. Chem. Soc. 108*, 2431–2437.

Nelson, J. W. & Kallenbach, N. R. (1986) *Proteins: Struct., Funct., Genet. 1*, 211–217.

Nelson, J. W. & Kallenbach, N. R. (1989) *Biochemistry 28*, 5226–5261.
Nishikawa, K., & Ooi, T. (1986) *Biochim. Biophys. Acta 871*, 45–54.
Pongor, S. & Szaley, A. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 366–370.
Prevelige, P., & Fasman, G. D. (1989) in *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G. D., Ed.) pp 391–416, Plenum Press, New York.
Rao, S., Zhu, Q.-l., Vajda, S., & Smith, T. (1993) *FEBS Lett. 322*, 143–146.
Reed, J., & Kinzel, V. (1991) *Biochemistry 30*, 4521–4528.
Richardson, J. S., & Richardson, D. C. (1987) in Protein *Engineering* (Oxender, D. L., & Fox, C. F., Eds.) p 149, Alan, R. Liss, New York.
Segawa, S.-I., Fukuno, T., Fujiwara, K., & Noda, Y. (1991) *Biopolymers 31*, 497–509.
Shoemaker, K. R., Kim, P. S., Brems, D. N., Marqusee, S., York, E. J., Chaiken, I. M., Stewart, J. M., & Baldwin, R. L. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2349–2353.
Sönnichsen, F. D., VanEyk, J. E., Hodges, R. S., & Sykes, B. D. (1992) *Biochemistry 31*, 8790–8798.
Sweet, R. M. (1986) *Biopolymers 25*, 1565–1577.
Toumadje, A., Alcorn, S. W., & Johnson, W. C., Jr. (1992) *Anal. Biochem. 200*, 321–331.
Wu, C. C., & Yang, J. T. (1981) *Mol. Cell. Biochem. 40*, 109–122.
Wu, C. C., & Yang, J. T. (1988) *Biopolymers 27*, 423–430.
Wu, C. C., Ikeda, K., & Yang, J. T. (1981) *Biochemistry 20*, 566–570.
Zhong, L., & Johnson, W. C., Jr. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 4462–4465.
Zvelebil, M. J., Barton, G. J., Taylor, W. R., & Sternberg, M. J. E. (1987) *J. Mol. Biol. 195*, 957–961.