Journal of Pharmaceutical & Biomedical Analysis
Vol. 7, No. 10, pp. 1103–1119, 1989
Printed in Great Britain

0731–7085/89 $3.00 + 0.00

Review

Underlying assumptions in the estimation of secondary structure content in proteins by circular dichroism spectroscopy — a critical review

MARK C. MANNING

Department of Pharmaceutical Chemistry, The University of Kansas, Lawrence, KS 66045, USA

Abstract: Recombinant DNA technology has made possible the large-scale production of proteins for pharmaceutical applications. As a result, there has been a renaissance in methodology which can provide information on the structural stability and character of these materials. Circular dichroism (CD) spectroscopy, with its sensitivity to the secondary structure adopted by the polypeptide chain, is a powerful tool in this regard. Quantitative analysis of the CD spectra of proteins is now wide-spread, aided by the availability of such algorithms on commercial instrumentation. However, there are basic assumptions made when conducting these calculations, many of which have not been addressed or summarized. Some of these assumptions are independent of the selection of basis spectra and the algorithm employed. These assumptions are listed and the available data concerning their validity is presented and discussed.

Keywords: *Circular dichroism spectroscopy; secondary structure; structural analysis; proteins.*

Introduction

With the advent of biotechnology, the preparation, isolation, and purification of proteins for pharmaceutical applications is now feasible. Consequently, any method which can provide information regarding the structural integrity of proteins is of interest and importance. As a result, there has been a renaissance in spectroscopic methods as a non-destructive procedure for evaluating protein structure in solution. Circular dichroism (CD) spectroscopy is one such tool, which is sensitive to the overall conformation and folding of polypeptide chains. While CD and the related phenomenon of optical rotatory dispersion (ORD) have been known for decades [1], their application to proteins and polypeptides is relatively recent. Before beginning a discussion of CD spectral analysis, a description of the physical basis of CD is in order.

In a CD experiment, measurements are made by alternately passing left- and right-handed circularly polarized light through a sample, and determining the difference in absorption (see equation 1):

$$\Delta A = A_L - A_R. \tag{1}$$

Ultimately, this difference in absorption can be related to the difference in extinction coefficients, which is easily converted to molar ellipticity, $[\theta]$, a common unit for reporting the intensity of CD bands (equations 2 and 3):

$$\Delta A = \Delta \epsilon c l, \tag{2}$$

$$[\theta] = 3298\ \Delta \epsilon, \tag{3}$$

where c refers to the concentration and l the pathlength of the sample. In order to exhibit CD, a system must possess either local or global chirality. Achiral molecules will display no CD signal. As CD is an absorption process, the bands are characteristic of the electronic excited states of the molecule (as opposed to vibrational circular dichroism, VCD, or Raman optical activity, ROA, which involve vibronic excited states), with distinctive intensities and frequencies. Since it is a difference method, the bands also possess either a negative or positive sign. Many of the fundamental aspects of CD spectroscopy have been reviewed [1–7], particularly with regard to the CD of peptides and proteins [8–16].

One of the most widespread applications of CD is the determination of secondary structure composition of proteins in solution [16–44]. In this approach, an experimental far UV (250–175 nm) CD spectrum is deconvoluted into a linear combination of basis spectra, each representative of a secondary structural type. The relative contribution of each basis spectrum to the overall CD curve is proportional to the amount of that structural feature in the protein. Basis spectra are derived from CD spectra of proteins having a known secondary structure composition (as determined by X-ray crystallography). Employing one of many variable curve-fitting routines [16–44], basis spectra can be obtained.

A more detailed description of the procedure can be illustrated with a discussion of the singular value decomposition (SVD) approach of Johnson and co-workers [16–20]. First, the CD spectra from 16 proteins form a 16 × 42 matrix, **R** (42 data points comprised of molar ellipticities taken every 2 nm over the range 260–178 nm for each of 16 proteins). Secondary structure content forms a second matrix, **F**, which is 16 × 5, i.e. assuming there are five different secondary structures which can be resolved. This seems to be the optimal number for data taken to 178 nm [16, 17]. The range over which **R** is measured varies from method to method. However, extension of the data into the vacuum UV ($\lambda < 180$ nm) seems to provide greater information content and improved accuracy [16, 17, 37]. Thus, what is sought is the matrix, **X**, which can relate the two, so that

$$\mathbf{F} = \mathbf{X}\,\mathbf{R}. \tag{4}$$

The singular value decomposition framework proceeds via decomposition of the matrix, **R**, into the product of three different matrices (equation 5):

$$\mathbf{R} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^{\mathrm{T}}, \tag{5}$$

where **U** is a matrix of orthogonal basis vectors which represent the eigenvectors of $\mathbf{RR}^{\mathrm{T}}$, **V** is a matrix of the eigenvectors of $\mathbf{R}^{\mathrm{T}}\mathbf{R}$, $\mathbf{R}^{\mathrm{T}}$ is the transpose of **R**, and **S** is a diagonal matrix containing the square roots of the eigenvalues of eigenvectors common to both **U** and **V**. These are the singular values, which indicate the relative importance of each eigenvector. The vector components of **US** are referred to as the basis CD spectra, while the rows of $\mathbf{V}^{\mathrm{T}}$ contain the coefficients indicating the amount each vector contributes toward the original CD spectra (i.e. the **R** matrix).

According to SVD theory,

$$\mathbf{X} = \mathbf{F}\,\mathbf{V}\,\mathbf{S}^{+}\,\mathbf{U}^{\mathrm{T}}, \tag{6}$$

where $\mathbf{S}^{+}$ is the pseudo-inverse matrix of **S** (formed by replacing each diagonal element of **S** by its reciprocal). The product $\mathbf{VS}^{+}\mathbf{U}^{\mathrm{T}}$ is described as the generalized inverse. Once **X** is determined, it can be used to determine the secondary structure content of any protein, $\mathbf{f}_c$, from its CD spectrum, **c**, according to equation (7):

$$\mathbf{f}_c = \mathbf{Xc} = \mathbf{F}\,\mathbf{V}\,\mathbf{S}^{+}\mathbf{U}^{\mathrm{T}}\mathbf{c}. \tag{7}$$

However, further simplification is usually employed before implementation of matrix **X**. Strictly speaking, there may be as many basis CD spectra as there are proteins in the original database (16 in this case). The singular values of the **S** matrix suggest that only five of these are significant, that is, these five are sufficient to adequately reconstruct any of the original 16 spectra. Therefore, the **V**, **S**, and **U** matrices are truncated to include only the five essential basis spectra (these matrices are now designated $\mathbf{V}_5$, $\mathbf{S}_5^{+}$ and $\mathbf{U}_5$), so that equation (7) now becomes

$$\mathbf{f}_c = \mathbf{X}\,\mathbf{c} = \mathbf{F}\,\mathbf{V}_5\,\mathbf{S}_5^{+}\mathbf{U}_5{}^{\mathrm{T}}\mathbf{c}. \tag{8}$$

The transformation matrix, **X**, can now be used to obtain the amounts of any of these five secondary structure types from any experimental spectrum, **c**, via equation (8). Johnson *et al.* have ascribed these types to the α helix, parallel β sheet, antiparallel β sheet, β turn, and random coil, or unordered, conformations. The matrix manipulation is straightforward, and has been further simplified by Compton and Johnson [20], who have published tables of the general inverses, meaning that secondary structure contents can be obtained simply by taking the dot product of the experimental spectrum and the appropriate general inverse. While these tables can be used for data only down to 186 nm, Johnson has shown that the degree of confidence in the accuracy of the estimations is greatly improved upon extending the experimental data down to 178 nm [16, 17, 20]. Other workers have employed similar curve-fitting schemes, in which the basis spectra are either obtained from protein databases [21–35] or from model systems [36–44]. The major differences are in the choice of basis spectra and in the deconvolution algorithm [16]. Many of these programs are readily available, and some are even included with the software on commercial CD instruments.

However, secondary structure analysis by CD involves numerous assumptions, many of which are independent of the deconvolution method employed or the choice of basis spectra. Furthermore, these assumptions may not be obvious to the user. The validity of

these assumptions has rarely been addressed and a critical assessment of them is imperative. While the accuracy of various basis sets (i.e. reference spectra) has already been examined [10, 16, 45], this article will focus on the following four underlying assumptions which are endogenous to all CD deconvolution treatments:

(1) the secondary structure composition of the crystalline protein is retained in solution;
(2) the effect of tertiary structure is negligible (i.e. individual secondary structural elements do not interact and their contribution to the overall CD spectrum is additive);
(3) only the amide chromophores are responsible for the far UV CD spectrum (contributions from side chain chromophores are assumed to be zero for all proteins);
(4) the geometric variability of secondary structural units is assumed to be negligible, i.e. a single CD curve is sufficient to describe each type of secondary structure (e.g. one for all α helices, one for all β sheets, etc.).

Clearly, any of these may not be valid for a given protein. Many examples are known in which the CD spectrum of a protein does not deconvolute properly, leading to either erroneous estimates of secondary structure content or a poor agreement between the experimental and reconstructed spectra. Studies which address these assumptions will be presented and discussed. While this article investigates the potential shortcomings of CD analysis, it should be emphasized that these programs and algorithms often produce accurate and reliable results. In addition, it should not be inferred that the workers who developed these methods were ignorant of the potential problems associated with these assumptions. Spectral analysis often requires simplifications in order to obtain information in a tractable fashion. Rather, the goal of this article is to educate potential users regarding the quantitative analysis of CD spectral data, especially considering the increased demand for methods of structural characterization of recombinant protein products.

Discussion

Assumption 1. Retention of crystal structure in aqueous solution

X-Ray diffraction is taken to be the basis for determination of the secondary structure content of a protein. However, the solid state conformation may not be retained in solution. Calmodulin, in its fully coordinated state with four tightly bound calcium ions, adopts a dumbell-shaped structure in the solid state [46, 47]. Each lobe of the dumbell is connected by a long, continuous helical segment (residues 66–92) in the crystal. Small angle X-ray scattering measurements indicate that the radius of gyration of calmodulin in solution is inconsistent with the crystal structure [48, 49]. Time-resolved fluorescence measurements, in conjunction with CD studies, suggest that this central helix may not be a normal part of the solution structure [50]. In the presence of calcium, the molecule appears to become much more compact, presumably with the loss of this central helix. Only a single rotational correlation time was detected for the molecule, whereas an asymmetric structure, as found in the crystal, would exhibit two distinct values. Fourier transform IR spectroscopic measurements support the observations that the helix content in Ca_4–calmodulin in solution is inconsistent with the crystal structure [51]. Troponin C, a homologous protein which possesses a similar crystal structure, displays similar behaviour in solution [51]. Recently, mutants of troponin C, having portions of

the central helix deleted, were found to still be viable [52], suggesting that the central helix is not present or essential in solution. Rubredoxin is another protein whose estimated secondary structure content is quite different from that found in the crystalline state. The difference has been attributed to a change in overall conformation [37].

With the increased capabilities of nuclear magnetic resonance (NMR) spectroscopy to evaluate the three-dimensional (3-D) structure of proteins in solution [53], comparison to the conformation in the solid state is possible. Currently, the 3-D structure of more than 20 small proteins have been determined by NMR. So far, NMR studies on small proteins such as crambin and basic bovine pancreatic trypsin inhibitor (BPTI) have indicated only minor differences between solution and solid state structures [54–59]. However, larger discrepancies are seen in the structures of toxins [60–64]. The most detailed comparison of solution and crystal structures has been for α-bungarotoxin [61–64]. Major differences are observed in both the backbone and side chain conformations, especially in the region surrounding the invariant tryptophan residue. More importantly, discrepancies are seen in the amount of secondary structure, because the central β sheet appears to be more extensive in solution. Both this effect and rotation of the tryptophan to the other side of the β sheet will undoubtedly affect the CD spectrum.

Another important factor which may differ in solution and crystal environments is the degree of aggregation. Not only can associated monomers introduce additional interactions which can perturb the CD spectrum (see the section on assumption 2), but it can also introduce changes in the conformation of the peptide backbone. Insulin has been thoroughly studied in this regard. Existing in well-defined aggregation states from monomer to hexamer, the CD spectrum of insulin has been observed to vary significantly. Separating effects arising from monomer–monomer contacts from those arising from aggregation-induced distortions in the secondary structure is difficult, although molecular dynamics simulations [65] and theoretical investigations of the CD spectrum [66–69] indicate that both occur. Both effects appear to modulate the aromatic side chain rotational strengths as well (see the section on assumption 3). Certainly, crystal packing forces appear to alter the structure of the insulin monomer [65, 70–72]. Association of glucagon also causes changes in its CD spectrum [10, 73–76]. The aggregated form appears to possess a higher α helix content than the monomer. Further, acidic conditions induce a change to a β sheet-like structure. Distortions are observed in small proteins such as avian pancreatic polypeptide [77, 78] and complement C3a [79]. Recent NMR studies show that complement component C3a, possesses a solution structure which differs from that found in the crystal [79]. The length of the third α helix is shorter in solution, extending from residues 47 to 66, as opposed to residues 47 to 73 in the solid state. Helical structure is observed at the N-terminus of C3a in solution (residues 8–15), while residues 13–15 are distinctly nonhelical in the crystal, and the first 12 residues could not be found due to disorder and flexibility. These variations were attributed either to additional intermolecular contacts (mostly helix–helix packing) found within the crystal or the high salt conditions required for crystallization [79].

It is also possible that a protein can crystallize in different forms, each possessing unique secondary and tertiary structure. Lacking any additional data, deciding which form most accurately represents the solution conformation would be difficult (see the discussion of BPTI in the section on assumption 3). In the case of BPTI, comparison of the solution structure with single crystal X-ray diffraction data is complicated by the fact that BPTI has been crystallized in three different forms [80–82]. Each displays small

variations in conformation from the other, while maintaining the global folding pattern. This is not surprising considering that BPTI contains three disulphide bonds (out of 58 amino acid residues), and consequently, possesses a relatively rigid structure. Yet, if such heterogeneity can exist in the solid state for a highly constrained protein, other proteins would be even more susceptible to variation. Thus, it becomes a question of which crystal structure is representative of the solution conformation.

Crystallographic determination of secondary structure content is also not trivial [15–17, 83]. Accurate identification of which residues belong to a certain structure type varies with the method of definition and assignment. Levitt and Greer found significant differences in secondary structure composition of proteins analysed by their method and the values provided by the original description of the structures [83]. In the cases of low resolution structures, the problem becomes even more acute. For example, the original report of the crystal structure of ferricytochrome c (4.0 Å resolution), stated that the protein had "little or no α helix content" [84]. Subsequent refinement to 2.8 Å still placed the α helix content at <10% [85], despite numerous CD studies which estimated it to be 27–45% [86–91]. Finally, at 2.0 Å resolution, the number of residues listed as being in an α-helical conformation was 30 out of 104 total residues (~29%) [92]. Levitt and Greer calculate the amount to be closer to 40% [83].

Assumption 2. Contributions from secondary structures are additive (no contributions from tertiary structure)

Sensitivity of CD spectroscopy to the tertiary structure of proteins is still an open question. Close packing of secondary structure units may produce a system where significant through-space interactions may occur, leading to an alteration of the CD spectrum. Two common folding motifs are known where interaction between units should be maximal: the four-helix bundle, where four α helices are arrayed in antiparallel fashion [93–100], and the β-sheet sandwich, a face-to-face arrangement of two β sheets as described by Chothia and co-workers [100–105]. Both theoretical and experimental studies indicate that significant interaction occurs between members of these tertiary structures.

In the four-helix bundle, the α helices are typically aligned in an antiparallel fashion and are tilted ~20° one to another [93–100]. This arises from the packing of the hydrophobic residues on the interior of the bundle. While some dispute exists over the degree of stabilization afforded by an antiparallel arrangement of helix dipoles [93, 96, 99, 106, 107], it is agreed that the hydrophobic interactions account for a significant amount of the stabilization energy [108, 109]. Recently, a synthetic four-helix bundle was designed and synthesized by DeGrado and co-workers [110–112]. Other such structures have since been produced in other laboratories [113]. Initially, the bundle was formed by association of four independent helices, which displayed very little helical structure on their own, but showed significant amounts of helix upon assembly [111]. These helices have since been linked into a single polypeptide chain and formation of the bundle was still observed. From model building, the structure should be essentially entirely helical (>85%). Yet, the CD curve suggests a helix content of only 60–70% [111]. While unwinding of the helix temini is a possible explanation of the observed decrease in the CD signal, it is equally likely that there are interhelix interactions which modulate the CD intensity. Theoretical studies on two antiparallel helices are consistent with these observations. Calculations show that for two helices separated by 7–10 Å, the overall

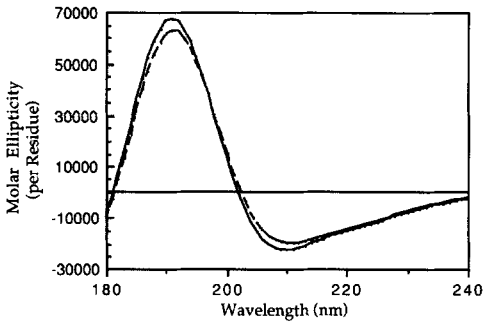

Figure 1
Calculated CD spectra for a single α helix (———) and for two antiparallel α helices separated by 8 Å (---------) (Cooper and Woody, unpublished data).

CD intensity is decreased by ~10% (see Fig. 1). The extent of change in the CD curve for four helices may be even more pronounced.

Systems composed of two α helices in close contact are also known. For example, the coiled-coil is a well established motif in biology [114, 115]. It involves two intertwined α helices, with a heptad repeat which allows the side chains to interdigitate themselves. Tropomyosin, a muscle protein, is one of the most widely studied coiled-coils. Decreases in the overall CD intensity, relative to a single α helix, are observed in these compounds [116–118], suggesting that helix–helix interactions are modulating the CD signal. Other coiled-coils behave in a similar fashion [119]. In order to account for this diminished intensity, the deconvolution algorithms factor in other secondary structures which have lower overall intensity (such as the β sheet), in order to produce an acceptable fit to the experimental data [119]. However, this may not produce a realistic description of the system. Another structure containing two α helices is the leucine zipper, found in certain DNA-binding proteins. Peptide models of the leucine zipper have been synthesized, and the helices were found to align in a parallel fashion [120, 121]. These dimers do not show a diminished CD curve [120], and the CD band positions and intensities are consistent with a system comprised entirely of α-helical peptide chains (Fig. 2). Amphiphilic α helices can associate as dimers [122–125] as well as tetramers [110–112]. Helical fragments from human growth hormone have been shown to associate, and yield an α-

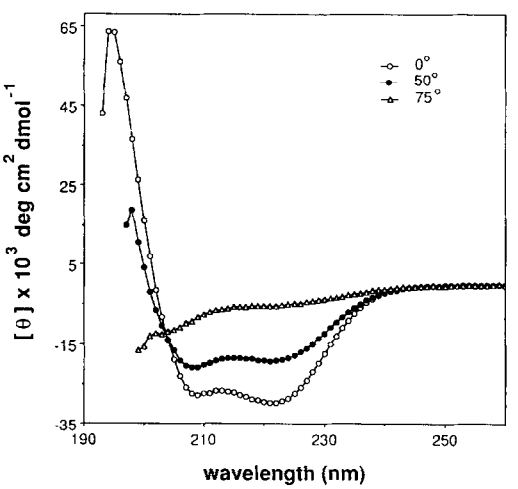

Figure 2
CD spectrum of a peptide model of the leucine zipper at 0°C (○), 50°C (●), and 75°C (△) [120]. (Reprinted with the permission of the author.)

helix-like CD spectrum [122–124]. However, the overall intensity, even in a 30% trifluoroethanol solution, is well below that anticipated for an entirely helical peptide [122]. Again, some unfolding may occur, but helix–helix interactions modulating the CD signal must be considered as a possible explanation. All of these analyses are complicated by the lack of a reliable standard for a fully α-helical peptide. Estimates of the molar ellipiticity at 222 nm, $[\theta]_{222}$, an indicator of helix content [26], range from −26,000 [122] to −38,000 [125] deg cm^2 dmol^{-1}.

To date, experimental measurements on isolated β-sheet sandwich structures has not been accomplished. Synthesis of a suitable, well-defined model for even a single β sheet is difficult in itself; construction of a multilayer structure would be even more problematic. Some models have been prepared, but their exact structure has yet to be determined [126–129]. There are some proteins which are comprised primarily of a β-sheet sandwich with little or no α helix [100–105], but they display widely variant CD spectra. Differences in strand length, the number of strands per sheet, and the extent of strand distortion are probably the major reasons for these discrepancies. All of these parameters have been calculated to have a marked effect of the CD spectra of β sheets [130, 131]. Theoretical studies on idealized β-sheet sandwiches and on the β-sheet sandwich portions of crystographically characterized proteins indicates that the intersheet interactions can modulate CD intensity by up to 10% [131]. However, the twisting of individual strands and the overall deformation of β sheets, both of which are observed in all of the proteins included in this study, leads to much larger effects on the CD spectra than the intersheet contributions [130, 131].

Assumption 3. Contributions from side chain chromophores in the far UV are negligible

The contributions from side chain chromophores, specifically aromatics and disulphides, can be sizeable, not only in the near UV (250–350 nm), where these chromophores display distinctive absorption and CD bands, but also in the far UV. Significant CD intensity in this region may compromise the accuracy of secondary structure analyses based upon this portion of the spectrum.

Of all the amino acid side chains, the indole group of tryptophan is expected to produce the largest CD signals, both in the near and far UV. Estimations of the molar ellipticity, [θ], range from 40,000 to 80,000 deg cm^2 dmol^{-1} in the far UV region [8, 37, 132]. In general, the CD of aromatic amino acids [tryptophan (Trp), tyrosine (Tyr), phenylalanine (Phe), and histidine (His)] has been widely reviewed [8, 10, 133, 134]. Of these, only tyrosine and tryptophan have been considered as contributing significantly to the far UV CD spectra of proteins. Woody has listed some globular proteins which seem to display sizeable contributions from tryptophan and tyrosine in the far UV [135]. An estimate of the rotational strength per aromatic group is provided. These observations were substantiated by calculations on Trp-containing dipeptides. An exhaustive search of conformational space for the Trp side chain suggested that there were accessible rotamers which could produce CD signals comparable to those observed in globular proteins [135]. Similar calculations had been conducted earlier for Tyr-, Trp- and Phe-containing peptides [136–146]. Unusual CD spectra (usually displaying strong positive features near 230 nm) have been observed for many proteins, and the anomalies have been postulated to arise from side chain contributions. These include concanavalin A [30, 147], DNase [148], BPTI [149, 150], erabutoxin b [151], avidin [152], gene 5 protein [153], and ACTH [154, 155]. Similarly, chymotrypsin and chymotrypsinogen, despite only minor differences in secondary structure by crystallography [156–159], display

markedly different far UV CD spectra at neutral pH [160]. The difference spectrum has a band at 230 nm indicative of aromatic contributions.

While aromatic side chains have been postulated to affect the far UV CD spectra of proteins and polypeptides, the ability to critically evaluate this phenomenon has been limited. Three major approaches have been described: (1) expanding the deconvolution algorithms to include aromatic side chain contributions, (2) site-directed mutagenesis, where the aromatic side chains have been replaced and a direct comparison of the wild type and mutant proteins can be accomplished, and (3) theoretical studies, where the CD spectra are calculated and individual side chain contributions can be determined.

Attempts to improve the secondary structure analysis of proteins by CD via inclusion of aromatic side chain contributions have been reported [8, 32, 37, 155]. While not including aromatic groups as a structural sub-type does not affect the generation of basis spectra, it does produce bases which include an average contribution from aromatic side chains. Therefore, unusual or large aromatic contributions may not be handled well. Whether this shortcoming really hinders accurate analysis of the secondary structure composition is doubtful. In fact, in many cases, they do not affect the analysis at all [8, 37, 155]. However, this may be due to the way in which the aromatic contributions are handled. Some workers have chosen to base the aromatic side chain basis spectra on simple dipeptides [37, 155], which undoubtedly ignores differences in side chain orientation and the interactions with the rest of the polypeptide chain. These effects have been shown to be significant in most proteins [135–137, 150, 154, 161]. A better approach is to generate a basis spectrum which represents the average aromatic side chain CD spectrum (ref. 32; W. C. Krueger, personal communication), and include it in the analysis as an additional secondary structural type. The added value of such modifications is still to be determined.

Proteins that do not deconvolute well mainly include those mentioned above, which display a prominent positive band in the 220–230 nm region. Presumably, these bands arise from the L_a band of tyrosine or tryptophan [8, 135, 136] or from disulphide-based transitions [134, 162–166]. No regular secondary structural feature, except the rare polyproline II helix, displays a positive band at these wavelengths. In fact, this structure has been invoked to account for the positive band in ACTH [155]. More likely, this feature is a result of aromatic side chain contributions. Surprisingly, one of the proteins which deconvolutes well without including side chains is concanavalin A [37], which has an extraordinary amount of aromatic residues and was previously found to be a difficult case [32]. Again, this is probably due to the difference in the handling of the aromatic side chain contributions.

Ability to manipulate the primary structure (i.e. sequence) of the protein via site-directed mutagenesis permits replacement of amino acids possessing aromatic side chains. Judicious substitutions which maintain the tertiary structure of the protein will allow the contribution of individual groups to the overall far UV CD to be determined. Proteins containing single Trp residues, such as growth hormone and interleukin-2 (IL-2), both pharmaceutically important materials, would be superb candidates for evaluating aromatic contributions in this region of the spectrum [12, 167]. Mutants of IL-2 have been synthesized, including one replacing the tryptophan with serine (Fig. 3). Dramatic differences are observed in the far UV CD spectra of wild type and the W121S mutant of IL-2 [167]. Whether this represents gross conformational changes or a significant tryptophan contribution in this region is still undetermined. Replacement of Trp with another large hydrophobic group would be less likely to disrupt the tertiary

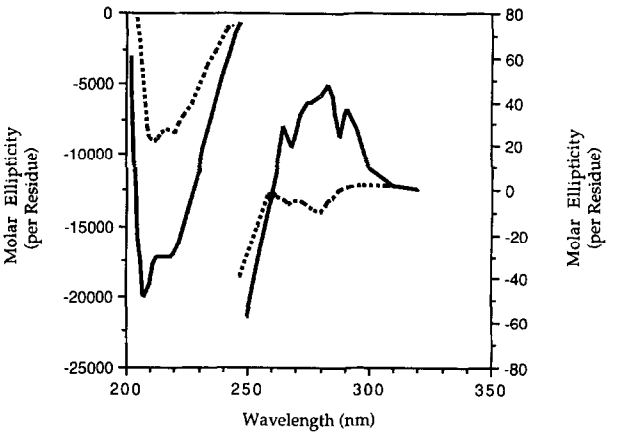

Figure 3
CD spectra of IL-2 (———) and the W121S mutant of IL-2 (■■) [167]. (Reprinted with the permission of the author.)

structure and present a better test for the rôle of Trp in perturbing the far UV CD spectrum of IL-2.

Detailed calculations on the impact of aromatic side chains on the far UV CD spectrum of proteins and polypeptides are numerous [66, 70, 130, 135–137, 146, 150]. Similar investigations have been conducted on small peptides and portions of proteins by Hooker and co-workers [138–145]. For example, detailed calculations have been conducted on the contributions of aromatic side chains to the far UV CD spectrum of BPTI [150]. As this protein contains no tryptophan or histidine, only phenylalanine and tyrosine needed to be considered. Theoretical studies demonstrated that any description of the far UV CD spectrum of BPTI based solely upon peptide backbone contributions gave an incorrect prediction of the experimental spectrum. Inclusion of tyrosine side chains did little to improve the fit. However, consideration of both tyrosine and phenylalanine produced a predicted curve very similar to the experimental data (Figs 4 and 5). Detailed analysis indicated that the phenylalanine side chains did not themselves provide significant rotational strength, but were essential in correctly modulating the interactions of other nearby groups [150]. The calculations also indicated that only those aromatic groups involved in the cluster of hydrophobic side chains (including Tyr^{21}, Tyr^{23}, Phe^{4}, Phe^{22}, and Phe^{45}) produced strong signals in either the far or near UV. Considering that aromatic side chains do tend to cluster in globular proteins [168], there may be many proteins who have a sizeable contribution to the overall CD arising from aromatic groups. Inaccuracies in the 240–260 nm region of the calculated CD spectrum are probably due to neglect of the disulphide chromophores [149].

Disulphides are predicted to possess electronic transitions in both the near and far UV [134, 162–166], which can produce strong CD signals due to the inherent chirality of the disulphide group. Despite the abundance of proteins containing this functional group, little is known about its higher energy electronic excited states. Experimentally, it is difficult to assess whether spectral changes are due to addition or removal of a particular chromophore. Site-directed mutagenesis allows such changes to be made, and the introduction of a disulphide bridge into T4 lysozyme has been accomplished [169, 170]. Differences are observed in the CD spectra of the mutant and wild-type proteins.

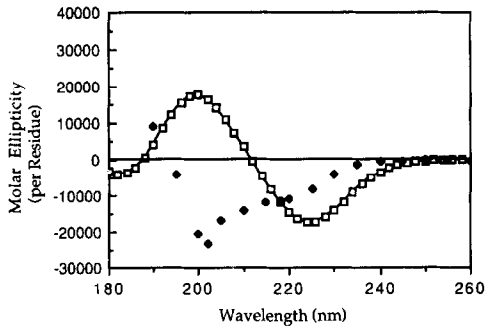

Figure 4
Comparison of the calculated (□) and experimental (◆) CD spectra for BPTI. Only contributions from the peptide backbone are included [150].

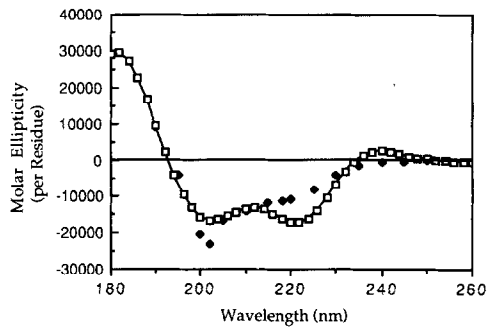

Figure 5
Comparison of the calculated (□) and experimental (◆) CD spectra for BPTI. Both backbone and side chain (tyrosine, phenylalanine) contributions are included [150].

Difference spectroscopy in the near UV does produce a reasonable spectrum, thought to arise from the engineered disulphide group [171]. Yet, whether the changes observed in the far UV are due to the disulphide transitions or to slight variation in secondary and tertiary structure is still undetermined. Difference spectra representing disulphide contributions in growth hormone have been determined by selective reduction or chemical modification of the protein [172–175]. As with T4 lysozyme, the disulphide signals were sizeable in the near UV, but small or not observed in the far UV.

Finally, examples of small peptides which adopt α-helical conformations in aqueous solution are now numerous [110–112, 120, 122–125, 176–184]. Frequently, CD spectroscopy is employed to estimate the amount of helical structure. Typically, the molar ellipticity at 222 nm is used as the indicator, following the correlation described by Chen *et al.* [24]. However, tyrosine, tryptophan, and disulphide groups can all produce strong CD signals in this region, making accurate determinations difficult or impossible. Close scrutiny of the CD spectra obtained for the N-terminal fragment from haemoglobin (residues 1–23) suggest that the single Trp is largely responsible for the discrepancy between NMR and CD estimates of α helix content [182]. Similarly, the C-terminal CNBr fragment from myoglobin contains two tyrosines. By CD, the amount of helical structure appears to be very small, whereas by NMR it is highly structured. Substitution of the N-terminal Tyr eliminates the interference caused by the aromatic side chain and restores the accuracy of the helical estimates by CD spectroscopy [183].

Prosthetic groups, such as haemes, may also affect the far UV CD spectra of proteins. Despite similar secondary structure composition (~80% α helix and no β structure), the CD spectra of haemoglobins and myoglobins are widely variant. While the shape of the

CD curve is maintained, the overall intensity can vary by 50%. Largest differences are observed for invertebrate and plant globins [185–187]. Considering the large dipole moment of an extended π system such as a haem, changes in haem orientation and position may produce an alteration of the far UV CD spectrum of haem proteins.

Assumption 4. Geometric variability does not affect the overall CD spectrum

Following the presentation of structural models for protein α and β structure by Pauling and Corey [188–190], the first high resolution X-ray crystal structures of protein demonstrated that both the α helix [191] and the β sheet [192] could be defined by a range of ϕ and ψ angles and not just a single set. Within this range, all of these structures would be nearly isoenergetic [193]. For example, the estimates of the average ϕ and ψ angles for the "typical" α helix span nearly 20° [194]. Even within a single helix, the distribution can be quite broad. Therefore, the question arises as to which set of backbone dihedrals will produce a representative α-helix-like CD curve. Moreover, whether structures at the extremes of each range will have differing CD spectra is still unknown. Again, the problem of constructing a well-defined model system with which to investigate these problems is apparent. Certain homopolymers are known to adopt α-helical conformations, but their exact structure (i.e. ϕ and ψ angles) is still undetermined and their solubility in water is quite limited. Extrapolating from data obtained in organic solvents to proteins in aqueous solution would further complicate the analysis. Not only is the heterogeneity in ϕ and ψ of importance, but difference in chain length should also produce variance in the CD spectra of α helices. Chang *et al.* attempted to account for this behaviour by scaling the α helix CD basis spectrum by the average length of a helix in globular proteins (10.5 residues) [26].

More specifically, certain features of the effect of varying chain length are of interest. First, knowledge of the minimum length required to produce an α-helix-like CD spectrum is essential. Conversely, the number of residues which are capable of simulating an infinite helix will be of interest. Experimental evidence on α helices indicates that approximately two turns of a helix are required to produce an α-helix-like CD spectrum [195–197]. This has been confirmed by theoretical studies [131, 198, 199], although some methods indicate that there is little length dependence in the α helix beyond three to four residues [200, 201]. Current deconvolution systems generate an average curve to represent each type of secondary structure, thereby ignoring the effects of structural variance.

While this problem occurs for all secondary structural types, it is even more acute for sheets and turns, where the number of geometric variables is even greater than with helices. As previously mentioned, the β sheet is susceptible to even more geometric variability, not only in chain length, but also the number of strands, the degree of twisting of each strand (as governed by the ϕ and ψ angles), and the overall deformation of the sheet. Theoretical studies have indicated that each of these parameters can have a dramatic effect on the far UV CD spectrum [130, 131]. Twisting of an individual strand produces the most pronounced effect, yielding a substantial increase in overall intensity. However, it must be noted that such twisting will distort the rest of the sheet, in order to maintain maximal hydrogen bonding and minimize contacts [100–105]. This makes it difficult to separate the effects of single strand twisting from that of the entire sheet. Finally, the characterization of the so-called "unordered" or "random" structure is the poorest of all. Considering the variety of conditions which can denature a protein and produce an unordered structure, the variability of this structural type must be extreme.

Conclusions

Circular dichroism spectroscopy is a powerful technique for evaluating the conformation of polypeptides and proteins. As an analytical tool, it can determine changes in secondary structure in a qualitative and even semi-quantitative fashion. Quantitative analysis of CD spectra for the amount of various secondary structural features is possible, but, as with any assay, the accuracy and proper interpretation of the results depends on an accurate understanding of the procedure. Deconvolution methods involve assumptions which are not necessarily apparent to the user. These are assumptions which are independent of the algorithm or basis set of spectra employed. Many of these have been presented here, along with a summary of the studies which pertain to the validity of these assumptions. While these assumptions may be justified in many cases, there are examples where the effects can be disastrous. Inspection of the shape of the protein CD curve can indicate whether accurate analysis can be accomplished. Spectra which do not resemble those found for "standard" secondary structures [9, 10] suggest that difficulties may be encountered, and caution should be exercised in the interpretation of these secondary structure estimates.

Acknowledgements — The author wishes to thank Dr Thomas Cooper, Dr Robert Woody, Dr Grace Ju, Dr V. Toome and Dr Peter Kim for providing results for inclusion in this publication. The comments of Dr Michael Mulkerrin, Dr Steven Shire, Dr William Krueger, Dr Peter Bayley, Dr W. Kevin Liddle, Dr Gerald Maggiora, Dr Jill Trewhella and, especially, Dr Robert Woody are greatly appreciated.

References

[1] S. F. Mason, *Molecular Optical Activity and the Chiral Discriminations*. Cambridge University Press, New York (1982).
[2] I. Tinoco Jr, *Adv. Chem. Phys.* **4**, 113–160 (1962).
[3] J. A. Schellman, *Acct. Chem. Res.* **1**, 144–151 (1968).
[4] I. Tinoco Jr and C. R. Cantor, *Meth. Biochem. Anal.* **18**, 81–203 (1970).
[5] W. C. Johnson Jr, *Rev. Scient. Instrum.* **42**, 1283–1286 (1971).
[6] P. M. Bayley, *Prog. Biophys. Molec. Biol.* **27**, 3–76 (1973).
[7] J. A. Schellman, *Chem. Rev.* **75**, 323–331 (1975).
[8] D. W. Sears and S. Beychok, in *Physical Principles and Techniques in Protein Chemistry*, Part C (S. J. Leach, Ed.), pp. 445–593. Academic Press, New York (1973).
[9] R. W. Woody, *J. Polym. Sci., Macromolec. Rev.* **12**, 181–321 (1981).
[10] R. W. Woody, in *The Peptides: Analysis, Synthesis, Structure*, Vol. 7 (V. J. Hruby, Ed.), pp. 15–114. Academic Press, Orlando (1985).
[11] Y. P. Myer, *Curr. Top. Bioenerg.* **14**, 149–188 (1985).
[12] T. G. Bewley, *Recent Prog. Horm. Res.* **35**, 155–213 (1979).
[13] B. Jirgensons, *Optical Activity of Proteins and Other Macromolecules*. Springer, Berlin (1973).
[14] V. Toome and M. Weigle, in *The Peptides: Analysis, Synthesis, Structure*, Vol. 4 (E. Gross and J. Meienhofer, Eds), pp. 85–184. Academic Press, New York (1981).
[15] J. T. Yang, C.-S. C. Wu and H. M. Martinez, *Meth. Enzym.* **130**, 209–269 (1986).
[16] W. C. Johnson Jr, *A. Rev. Biophys. Biophys. Chem.* **17**, 145–166 (1988).
[17] J. P. Hennessey Jr and W. C. Johnson Jr, *Biochemistry* **20**, 1085–1094 (1981).
[18] J. P. Hennessey Jr and W. C. Johnson Jr, *Analyt. Biochem.* **125**, 177–188 (1982).
[19] P. Manavalan and W. C. Johnson Jr, *Analyt. Biochem.* **167**, 76–85 (1987).
[20] L. Compton and W. C. Johnson Jr, *Analyt. Biochem.* **155**, 155–167 (1986).
[21] V. P. Saxena and D. B. Wetlaufer, *Proc. Natn. Acad. Sci. USA* **68**, 969–972 (1971).
[22] Y.-H. Chen and J. T. Yang, *Biochem. Biophys. Res. Commun.* **44**, 1285–1291 (1971).
[23] Y.-H. Chen, J. T. Yang and H. M. Martinez, *Biochemistry* **11**, 4120–4131 (1972).
[24] Y.-H. Chen, J. T. Yang and K. H. Chan, *Biochemistry* **13**, 3350–3359 (1974).
[25] C.-S. C. Wu and J. T. Yang, *Biochemistry* **15**, 3007–3014 (1976).
[26] C. T. Chang, C.-S. C. Wu and J. T. Yang, *Analyt. Biochem.* **91**, 13–31 (1978).
[27] S. W. Provencher and J. Glöckner, *Biochemistry* **20**, 33–37 (1981).
[28] S. W. Provencher, *Comput. Phys. Commun.* **27**, 213–227 (1982).

[29] S. W. Provencher, *Comput. Phys. Commun.* **27**, 229–242 (1982).
[30] I. A. Bolotina, V. O. Chekhov, V. Y. Lugauskas, A. V. Finkel'stein and O. Ptitsyn, *Molec. Biol. (Engl. Transl.)* **14**, 701–709 (1980).
[31] I. A. Bolotina, V. O. Chekhov, V. Y. Lugauskas and O. Ptitsyn, *Molec. Biol. (Engl. Transl.)* **14**, 709–715 (1980).
[32] I. A. Bolotina and V. Y. Lugauskas, *Molec. Biol. (Engl. Transl.)* **19**, 1154–1166 (1985).
[33] D. Grosse, S. Fermandjian, P. Fromageot, M. C. Khosla, R. R. Smeby and F. M. Bumpus, *Acta Biol. Med. Germ.* **29**, 777–791 (1972).
[34] D. Grosse, J. Malur, W. Meiske and K. R. H. Repke, *Biochim. Biophys. Acta* **359**, 33–46 (1974).
[35] J. B. Siegel, W. E. Steinmetz and G. L. Long, *Analyt. Biochem.* **104**, 160–167 (1980).
[36] H. Rosenkranz and W. Scholtan, *Hoppe-Seyler's Zeit. Physiol. Chem.* **352**, 896–904 (1971).
[37] S. Brahms and J. Brahms, *J. Molec. Biol.* **138**, 149–178 (1980).
[38] G. Holzwarth and P. Doty, *J. Am. Chem. Soc.* **87**, 218–228 (1965).
[39] N. Greenfield and G. D. Fasman, *Biochemistry* **8**, 4108–4116 (1969).
[40] J. Y. Cassim and J. T. Yang, *Biochem. Biophys. Res. Commun.* **26**, 58–64 (1967).
[41] C. C. Baker and I. Isenberg, *Biochemistry* **15**, 629–634 (1971).
[42] Y. I. Ramm, V. P. Gorenburg, A. B. Klionskii and Y. Y. Smirnova, *Biofizika* **25**, 559–560 (1980).
[43] Y. I. Ramm, V. P. Gorenburg, A. B. Klionskii and Y. Y. Smirnova, *Biofizika* **25**, 561–562 (1980).
[44] L. A. Holladay, *Biophys. Chem.* **7**, 41–49 (1977).
[45] A. Wollmer, W. Strassburger and U. Glattner, in *Modern Methods in Protein Chemistry — Review Articles* (H. Tschesche, Ed.), pp. 361–384. de Gruyter, Berlin (1983).
[46] T. S. Babu, J. S. Sack, T. J. Greenhough, C. E. Bugg, A. R. Means and W. J. Cook, *Nature* **315**, 37–40 (1985).
[47] R. H. Kretsinger, S. E. Rudnick and L. J. Weissman, *J. Inorg. Biochem.* **28**, 289–302 (1986).
[48] D. B. Heidorn, D. Torney and J. Trewhella, *Biophys. J.* **51**, 435a (1987).
[49] D. B. Heidorn and J. Trewhella, *Biochemistry* **27**, 909–915 (1988).
[50] P. M. Bayley, S. Martin and G. Jones, *FEBS Lett.* **238**, 61–66 (1988).
[51] J. Trewhella, W. K. Liddle, D. B. Heidorn and N. Strynadka, *Biochemistry* **28**, 1294–1301 (1989).
[52] Z. Dobrowolski, G.-Q. Xu and S. E. Hitchcock-DeGregori, *Biophys. J.* **55**, 274a (1989).
[53] K. Wüthrich, *NMR of Proteins and Nucleic Acids*. Wiley, New York (1986).
[54] T. Havel and K. Wüthrich, *Bull. Math. Biol.* **46**, 673–698 (1984).
[55] W. Braun and N. Go, *J. Molec. Biol.* **186**, 611–626 (1985).
[56] G. M. Clore, A. T. Brünger, M. Karplus and A. M. Gronenborn, *J. Molec. Biol.* **191**, 523–551 (1986).
[57] A. T. Brünger, G. M. Clore, A. M. Gronenborn and M. Karplus, *Proc. Natn. Acad. Sci. USA* **83**, 3801–3805 (1986).
[58] G. Wagner, W. Braun, T. F. Havel, T. Schaumann, N. Go and K. Wüthrich, *J. Molec. Biol.* **196**, 611–639 (1987).
[59] M. Vásquez and H. A. Scheraga, *J. Biomolec. Struct. Dyn.* **5**, 705–755 (1988).
[60] W. E. Steinmetz, P. E. Bougis, H. Rochat, O. D. Redwine, W. Braun and K. Wüthrich, *Eur. J. Biochem.* **172**, 101–116 (1988).
[61] R. A. Love and R. M. Stroud, *Protein Engng* **1**, 37 (1986).
[62] V. J. Basus, M. Billeter, R. A. Love, R. M. Stroud and I. D. Kuntz, *Biochemistry* **27**, 2763–2771 (1988).
[63] P. A. Kosen, J. Finer-Moore, M. P. McCarthy and V. J. Basus, *Biochemistry* **27**, 2775–2781 (1988).
[64] C. Schiffer, V. J. Basus, W. F. van Gunsteren and R. M. Stroud, *Biophys. J.* **55**, 149a (1989).
[65] P. Krüger, W. Strassburger, A. Wollmer and W. F. van Gunsteren, *Eur. Biophys. J.* **14**, 449–459 (1987).
[66] D. Mercola and A. Wollmer, in *Structural Studies on Molecules of Biological Interest* (G. Dodson, J. P. Glusker and D. Sayre, Eds), pp. 557–582. Clarendon Press, Oxford (1981).
[67] A. Wollmer, W. Strassburger, E. Hoenjet, U. Glattner, J. Fleischauer, D. A. Mercola, R. A. G. de Graaf, E. J. Dodson, G. G. Dodson, D. G. Smith, D. Brandenburg and W. Danho, in *Insulin — Chemistry, Structure, and Function of Insulin and Related Hormones* (D. Brandenburg and A. Wollmer, Eds), pp. 27–35. de Gruyter, Berlin (1980).
[68] E. H. Strickland and D. A. Mercola, *Biochemistry* **15**, 3875–3884 (1976).
[69] A. Wollmer, J. Fleischauer, W. Strassburger, H. Thiele, D. Brandenburg, G. Dodson and D. Mercola, *Biophys. J.* **20**, 233–243 (1977).
[70] C. Chothia, A. M. Lesk, G. G. Dodson and D. C. Hodgkin, *Nature* **302**, 500–505 (1983).
[71] J. F. Cutfield, S. M. Cutfield, E. J. Dodson, G. G. Dodson, C. D. Reynolds and D. Valley, in *Structural Studies on Molecules of Biological Interest* (G. Dodson, J. P. Glusker and D. Sayre, Eds), pp. 527–546. Clarendon Press, Oxford (1981).
[72] S. J. Wodak, P. Alard, P. Delhaise and C. Renneboog-Squilbin, *J. Molec. Biol.* **181**, 317–322 (1984).
[73] W. B. Gratzer, G. H. Beaven, H. W. E. Rattle and E. M. Bradbury, *Eur. J. Biochem.* **3**, 276–283 (1968).
[74] P. A. Srere and G. C. Brooks, *Archs Biochem. Biophys.* **129**, 708–710 (1969).
[75] B. Panijpan and W. B. Gratzer, *Eur. J. Biochem.* **45**, 547–553 (1974).

[76] S. Formisano, M. L. Johnson and H. Edelhoch, *Proc. Natn. Acad. Sci. USA* **74**, 3340–3344 (1977).
[77] P. Krüger, W. Strassburger, A. Wollmer and W. F. van Gunsteren, *Eur. Biophys. J.* **13**, 77–88 (1985).
[78] I. D. Glover, D. J. Barlow, J. E. Pitts, S. P. Wood, I. J. Tickle, T. L. Blundell, K. Takemoto, J. R. Kimmel, A. Wollmer, W. Strassburger and Y. S. Zhang, *Eur. J. Biochem.* **142**, 379–385 (1984).
[79] D. G. Nettesheim, R. P. Edalji, K. W. Mollison, J. Greer and E. R. P. Zuiderweg, *Proc. Natn. Acad. Sci. USA* **85**, 5036–5040 (1988).
[80] A. Wlodawer, J. Walter, R. Huber and L. Sjölin, *J. Molec. Biol.* **180**, 301–329 (1984).
[81] A. Wlodawer, J. Diesenhofer and R. Huber, *J. Molec. Biol.* **193**, 145–156 (1987).
[82] A. Wlodawer, J. Nachman, G. L. Gilliand, W. Gallagher and C. Woodward, *J. Molec. Biol.* **198**, 469–480 (1987).
[83] M. Levitt and J. Greer, *J. Molec. Biol.* **114**, 181–293 (1977).
[84] R. E. Dickerson, M. L. Kopka, J. Wienzierl, J. Varnum, D. Eisenberg and E. Margoliash, *J. Biol. Chem.* **242**, 3015–3018 (1967).
[85] R. E. Dickerson, T. Takano, D. Eisenberg, O. B. Kallai, L. Samson, A. Cooper and E. Margoliash, *J. Biol. Chem.* **246**, 1511–1535 (1971).
[86] D. W. Urry and P. Doty, *J. Am. Chem. Soc.* **87**, 2756–2758 (1965).
[87] D. W. Urry, *Proc. Natn. Acad. Sci. USA* **54**, 640–648 (1965).
[88] D. W. Urry, *J. Biol. Chem.* **242**, 4441–4448 (1967).
[89] T. Flatmark and A. B. Robinson, in *Structure and Function of Cytochromes* (K. Okunuki, K. Kamen and I. Sekuzu, Eds), pp. 318–328. University of Tokyo Press, Tokyo (1968).
[90] R. Zand and S. Vinogradov, *Archs Biochem. Biophys.* **125**, 94–97 (1968).
[91] K. Aki, T. Tagaki, T. Isemura and T. Yamano, *Biochim. Biophys. Acta* **122**, 193–201 (1966).
[92] T. Takano, B. L. Trus, N. Mandel, G. Mandel, O. B. Kallai, R. Swanson and R. E. Dickerson, *J. Biol. Chem.* **252**, 776–785 (1977).
[93] C. Chothia, M. Levitt and D. Richardson, *Proc. Natn. Acad. Sci. USA* **74**, 4130–4134 (1977).
[94] P. C. Weber and F. R. Salemme, *Nature* **287**, 82–84 (1980).
[95] K.-C. Chou, G. Nemethy and H. A. Scheraga, *J. Phys. Chem.* **87**, 2369–2381 (1983).
[96] K.-C. Chou, G. Nemethy and H. A. Scheraga, *J. Am. Chem. Soc.* **106**, 3161–3170 (1984).
[97] S. Furois-Corbin and A. Pullman, *J. Biomolec. Struct. Dyn.* **4**, 589–597 (1987).
[98] D. H. Ohlendorf, B. C. Finzel, P. C. Weber and F. R. Salemme, in *Protein Engineering* (D. L. Oxender and C. F. Fox, Eds), pp. 165–174. Liss, New York (1987).
[99] K.-C. Chou, G. M. Maggiora, G. Nemethy and H. A. Scheraga, *Proc. Natn. Acad. Sci. USA* **85**, 4295–4299 (1988).
[100] C. Chothia, *A. Rev. Biochem.* **53**, 537–572 (1984).
[101] C. Chothia and J. Janin, *Proc. Natn. Acad. Sci. USA* **78**, 4146–4150 (1981).
[102] C. Chothia, M. Levitt and D. Richardson, *J. Molec. Biol.* **145**, 215–250 (1981).
[103] C. Chothia and J. Janin, *Biochemistry* **21**, 3955–3965 (1982).
[104] C. Chothia and A. M. Lesk, *J. Molec. Biol.* **160**, 309–323 (1982).
[105] C. Chothia, *J. Molec. Biol.* **163**, 107–117 (1983).
[106] B. Honig, K. A. Sharp and M. K. Gilson, *J. Cell. Biochem.* **13A**, 40 (1989).
[107] R. F. Sheridan, R. M. Levy and F. R. Salemme, *Proc. Natn. Acad. Sci. USA* **79**, 4545–4549 (1982).
[108] T. J. Richmond and F. M. Richards, *J. Molec. Biol.* **119**, 537–555 (1978).
[109] J. W. Ponder and F. M. Richards, *J. Cell. Biochem.* **13A**, 59 (1989).
[110] D. Eisenberg, W. Wilcox, S. M. Eshita, P. M. Pryciak, S. P. Ho and W. F. DeGrado, *Proteins* **1**, 16–22 (1982).
[111] S. P. Ho and W. F. DeGrado, *J. Am. Chem. Soc.* **109**, 6751–6758 (1987).
[112] W. F. DeGrado, *Adv. Protein Chem.* **39**, 51–124 (1988).
[113] M. H. Hecht, D. C. Richardson and J. S. Richardson, *J. Cell. Biochem.* **13A**, 86 (1989).
[114] J. A. Talbot and R. S. Hodges, *Acct. Chem. Res.* **15**, 224–230 (1982).
[115] C. Coheen and D. A. Parry, *Trends Biochem. Sci.* **11**, 245–247 (1986).
[116] B. Bullard, D. A. Mercola and W. F. Mommaerts, *Biochim. Biophys. Acta* **434**, 90–99 (1976).
[117] B. Nagy, *J. Biol. Chem.* **252**, 4557–4563 (1977).
[118] S. Kumar, M. E. Holtzer and A. Holtzer, *Biophys. J.* **53**, 97a (1988).
[119] M. Farooq, W. T. Norton, W. T. Winter and N. M. Tooney, *Biophys. J.* **53**, 451a (1988).
[120] E. K. O'Shea, R. Rutkowski and P. S. Kim, *Science* **243**, 538–542 (1989).
[121] C. Abate, R. Gentz and T. Curran, *J. Cell. Biochem.* **13A**, 97 (1989).
[122] V. Roontga, R. Powers, C. Jones, M. J. Beakage, J. E. Shields and D. C. Gorenstein, *Biochemistry* **28**, 1048–1054 (1989).
[123] D. N. Brems, S. M. Plaisted, E. W. Kauffman and H. A. Havel, *Biochemistry* **25**, 6539–6543 (1986).
[124] D. N. Brems, S. M. Plaisted, H. A. Havel and C.-C. C. Tomich, *Proc. Natn. Acad. Sci. USA* **85**, 3367–3371 (1988).
[125] G. Merutka and E. Stellwagen, *Biochemistry* **28**, 352–357 (1989).
[126] R. Moser, R. M. Thomas and B. Gutte, *FEBS Lett.* **157**, 247–251 (1983).

[127] W. Kullman, *J. Med. Chem.* **27**, 106–115 (1984).
[128] J. S. Richardson and D. C. Richardson, in *Protein Engineering* (D. L. Oxender and C. F. Fox, Eds), pp. 149–163. Liss, New York (1987).
[129] C. B. Unson, B. W. Erickson, D. C. Richardson and J. S. Richardson, *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **43**, 1837 (1984).
[130] M. C. Manning and R. W. Woody, *Biopolymers* **26**, 1731–1752 (1987).
[131] M. C. Manning, M. Illangeskare and R. W. Woody, *Biophys. Chem.* **31**, 77–86 (1988).
[132] M. Shiraki, *Sci. Pap. Coll. Gen. Educ., Univ. Tokyo* **19**, 151–173 (1969).
[133] E. H. Strickland, *CRC Crit. Rev. Biochem.* **2**, 113–174 (1974).
[134] P. C. Kahn, *Meth. Enzym.* **61**, 339–378 (1979).
[135] R. W. Woody, *Proceedings of the Second International Conference CD* (M. Kajtar, Ed.), pp. 38–54. Budapest, Hungary (1985).
[136] R. W. Woody, *Biopolymers* **17**, 1451–1467 (1978).
[137] R. W. Woody, *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **39**, 1805 (1980).
[138] T. M. Hooker Jr and J. A. Schellman, *Biopolymers* **9**, 1319–1348 (1970).
[139] W. J. Goux, D. B. Cooke, R. E. Rodriguez and T. M. Hooker Jr, *Biopolymers* **13**, 2315–2329 (1974).
[140] J. W. Snow and T. M. Hooker Jr, *J. Am. Chem. Soc.* **96**, 7800–7806 (1974).
[141] W. J. Goux and T. M. Hooker Jr, *J. Am. Chem. Soc.* **97**, 1605–1606 (1975).
[142] P. E. Grebow and T. M. Hooker Jr, *Biopolymers* **14**, 871–881 (1975).
[143] J. W. Snow, T. M. Hooker Jr and J. A. Schellman, *Biopolymers* **15**, 977–997 (1976).
[144] W. J. Goux and T. M. Hooker Jr, *J. Am. Chem. Soc.* **102**, 7080–7087 (1980).
[145] R. W. Snyder and T. M. Hooker Jr, *Biopolymers* **21**, 547–563 (1982).
[146] A. K. Chen and R. W. Woody, *J. Am. Chem. Soc.* **93**, 29–37 (1971).
[147] M. S. Herrmann, C. E. Richardson, L. M. Setzler, W. D. Behnke and R. E. Thompson, *Biopolymers* **17**, 2107–2120 (1978).
[148] S. N. Timasheff and G. Bernardi, *Archs Biochem. Biophys.* **141**, 53–58 (1970).
[149] P. A. Kosen, T. E. Creigton and E. R. Blout, *Biochemistry* **20**, 5744–5754 (1981).
[150] M. C. Manning and R. W. Woody, *Biophys. J.* **53**, 96a (1988).
[151] A. Menez, F. Bouet, N. Tamiya and P. Fromageot, *Biochim. Biophys. Acta* **453**, 121–132 (1976).
[152] N. M. Green and M. D. Melamed, *Biochem. J.* **100**, 614–621 (1966).
[153] L. A. Day, *Biochemistry* **12**, 5329–5339 (1973).
[154] L. A. Holladay and D. Puett, *Biopolymers* **15**, 43–59 (1976).
[155] A. A. Makarov, V. M. Lobachev, Y. A. Pankov and N. G. Yesipova, *Biofizika* **26**, 951–952 (1981).
[156] B. W. Matthews, P. B. Sigler, R. Henderson and D. M. Blow, *Nature* **214**, 652–656 (1967).
[157] P. B. Sigler, D. M. Blow, B. W. Matthews and R. Henderson, *J. Molec. Biol.* **35**, 143–164 (1968).
[158] J. Kraut, H. T. Wright, M. Kellerman and S. T. Freer, *Proc. Natn. Acad. Sci. USA* **58**, 304–311 (1967).
[159] H. T. Wright, *J. Molec. Biol.* **79**, 1–11 (1973).
[160] J. McConn, G. D. Fasman and G. P. Hess, *J. Molec. Biol.* **39**, 551–562 (1969).
[161] P. L. Luisi, V. Rizzo, G. P. Lorenzi, B. Straub, U. Suter and R. Guarnaccia, *Biopolymers* **14**, 2347–2362 (1975).
[162] R. W. Woody, *Tetrahedron* **29**, 1273–1283 (1973).
[163] D. B. Boyd, *J. Am. Chem. Soc.* **94**, 8800–8804 (1972).
[164] J. Lindenberg and J. Michl, *J. Am. Chem. Soc.* **92**, 2619–2625 (1970).
[165] D. B. Boyd, *J. Phys. Chem.* **78**, 1554–1563 (1974).
[166] R. C. Hider, A. F. Drake and N. Tamiya, *Biopolymers* **27**, 113–122 (1988).
[167] L. Collins, W.-H. Tsien, C. Seals, J. Hakimi, D. Weber, P. Bailon, J. Hoskings, W. C. Greene, V. Toome and G. Ju, *Proc. Natn. Acad. Sci. USA* **85**, 7709–7713 (1988).
[168] S. K. Burley and G. A. Petsko, *Science* **229**, 23–28 (1985).
[169] L. J. Perry and R. Wetzel, Science **226**, 555–557 (1984).
[170] R. Wetzel, L. J. Perry, W. A. Baase and W. J. Becktel, *Proc. Natn. Acad. Sci. USA* **85**, 401–405 (1988).
[171] M. G. Mulkerrin, L. J. Perry and R. Wetzel, in *Protein Structure, Folding, and Design* (D. L. Oxender, Ed.), pp. 297–305. Liss, New York (1986).
[172] T. A. Bewley, J. Brovetto-Cruz and C. H. Li, *Biochemistry* **8**, 4701–4708 (1969).
[173] T. A. Bewley and C. H. Li, *Archs Biochem. Biophys.* **138**, 338–346 (1970).
[174] C. H. Li and T. A. Bewley, *Excepta Med. Int. Congr. Ser.* **381**, 85 (1970).
[175] T. Bewley, *Biochemistry* **16**, 209–215 (1977).
[176] S. Ihara, T. Ooi and S. Takahashi, *Biopolymers* **21**, 131–145 (1982).
[177] P. S. Kim and R. L. Baldwin, *Nature* **307**, 329–334 (1984).
[178] K. R. Shoemaker, P. S. Kim, D. N. Brems, S. Marqusee, E. J. York, I. M. Chaiken, J. M. Stewart and R. L. Baldwin, *Proc. Natn. Acad. Sci. USA* **82**, 2349–2353 (1985).
[179] M. A. Jiminez, J. L. Nieto, M. Rico and J. Santoro, *FEBS Lett.* **221**, 320–324 (1987).
[180] D. N. Brems, S. M. Plaisted, E. W. Kauffman, M. Lund and S. R. Lehrman, *Biochemistry* **26**, 7774–7778 (1987).

[181] L. R. McLean, K. A. Hagaman, T. J. Owen and J. L. Krstenansky, *Biophys. J.* **55**, 358a (1989).
[182] M. C. Peña, M. Rico, J. Jiminez, J. Herranz, J. Santoro and J. L. Nieto, *Biochim. Biophys. Acta* **957**, 380–389 (1989).
[183] P. E. Wright, *Biophys. J.* **55**, 224a (1989).
[184] P. R. Gooley and N. E. MacKenzire, *Biochemistry* **27**, 4032–4040 (1988).
[185] Y. Sugita, Y. Dohi and Y. Yoneyama, *Biochem. Biophys. Res. Commun.* **31**, 447–452 (1968).
[186] N. A. Nicola, E. Minasian, C. A. Appleby and S. J. Leach, *Biochemistry* **14**, 5141–5149 (1975).
[187] K. Shikama, T. Suzuki, Y. Sugawara, T. Katagiri, T. Tagaki and M. Hatano, *Biochim. Biophys. Acta* **701**, 138–141 (1982).
[188] L. Pauling and R. B. Corey, *Proc. Natn. Acad. Sci. USA* **37**, 235–240 (1951).
[189] L. Pauling and R. B. Corey, *Proc. Natn. Acad. Sci. USA* **37**, 241–250 (1951).
[190] L. Pauling and R. B. Corey, *Proc. Natn. Acad. Sci. USA* **37**, 251–256 (1951).
[191] J. C. Kendrew, R. E. Dickerson, B. E. Strandbert, R. G. Hart, D. R. Davies, D. C. Philips and V. C. Shore, *Nature* **185**, 422–427 (1960).
[192] C. C. F. Blake, D. F. Koenig, G. A. Mair, A. C. T. North, D. C. Philips and V. R. Sarma, *Nature* **206**, 757–761 (1965).
[193] G. N. Ramachandran and V. Sasisekharan, *Adv. Protein Chem.* **23**, 283–437 (1968).
[194] T. L. Blundell, D. Barlow, N. Borkakorti and J. M. Thornton, *Nature* **306**, 281–283 (1983).
[195] M. Goodman, A. S. Verdini, C. Toniolo, W. D. Philips and F. A. Bovey, *Proc. Natn. Acad. Sci. USA* **64**, 444–450 (1969).
[196] V. Madison and J. A. Schellman, *Biopolymers* **11**, 1041–1061 (1972).
[197] J. M. Becker and F. Naider, *Biopolymers* **13**, 1747–1750 (1974).
[198] R. W. Woody and I. Tinoco Jr, *J. Chem. Phys.* **45**, 4927–4945 (1967).
[199] M. C. Manning and R. W. Woody, *Biophys. J.* **51**, 280a (1987).
[200] J. Applequist, *Biopolymers* **20**, 387–397 (1981).
[201] J. N. Vournakis, J. F. Yan and H. A. Scheraga, *Biopolymers* **6**, 1531–1550 (1968).

[*Received for review 29 March 1989; revised manuscript received 18 April 1989*]