Determination of the Helix and β Form of Proteins in Aqueous Solution by Circular Dichroism†

Yee-Hsiung Chen,‡ Jen Tsi Yang,* and Kue Hung Chau§

ABSTRACT: The circular dichroism (CD) of a protein at any wavelength can be expressed as $X = f_H X_H^{\bar{n}} + f_\beta X_\beta + f_R X_R$ (1). The f's are the fractions of helix (H), β form, and unordered form (R). $\bar{n}$ refers to the average number of peptide units per helical segment in a protein molecule. The X parameters are determined from the CD spectra of five or eight proteins of known f's, using a least-squares method. They, in turn, can be used to estimate the f_H and f_β of a protein by fitting its CD spectrum (below 245 nm) with eq 1. Results of tests on eight proteins are good, provided that the $\bar{n}$ of the proteins is close to the overall $\bar{n}$ of the reference proteins (around 10). This restriction of a single $\bar{n}$ for all proteins can be removed by introducing a chain-length-dependent factor, k, for the helix, *i.e.*, $X_H^n = X_H^\infty(1 - k/n)$. The X_H^∞ is further expressed by three Gaussian bands. Three methods of analysis are proposed by (1) combining the CD spectrum of a protein with its optical rotation at 233 nm, (2) introducing a restriction of constant rotational strength for each CD band of an infinite helix, and (3) combining methods 1 and 2. They determine not only the fractions of helix and β form but also provide an estimate of the number and average size of helical segments in a protein molecule. Results of tests on eight proteins indicate that the proposed methods are superior to the use of eq 1 alone.

The helix, β-pleated sheet, and unordered form of proteins each have a characteristic CD[1] and ORD spectrum. We should be able to determine the fraction of each form in a protein molecule by summing up their contributions to the CD and ORD which should agree with the experimental spectra. The optical activities of the three conformations are assumed to be additive and can be expressed as seen in eq 1 (Saxena and Wetlaufer, 1971; Chen and Yang, 1971; Chen *et al.*, 1972). At any fixed wavelength, λ

$$X = f_H X_H + f_\beta X_\beta + f_R X_R \qquad (1)$$

with the sum of f's equal to unity and all f's ≥ 0. For the sake of brevity, the symbol (λ) is omitted after X, X_H, X_β, and X_R. Here X is the mean residue ellipticity, $[\theta]$, or mean residue rotation, $[m]$. The X_H, X_β, and X_R are the reference values for pure helix (H), β form, and unordered form (R). The f's are the fractions of the three forms in a protein molecule. Equation 1 is also applicable to the b_0 of the Moffitt equation (Moffitt and Yang, 1956); in this case, the term (λ) is of course absent. Equation 1 only applies when the contributions of non-peptide chromophores, if any, are insignificant in the wavelength range under study.

Attempts to fit the ORD and CD spectra of proteins with X_H, X_β, and X_R based on synthetic polypeptides (Greenfield *et al.*, 1967; Greenfield and Fasman, 1969; Magar, 1968; Rosenkranz and Scholtan, 1971) remain problematic. Questions have been raised as to the choice of poly(L-lysine) as a reference compound. For instance, the CD spectrum does vary in magnitude among different helical polypeptides. Even with poly(L-lysine), the β form prepared by adding sodium dodecyl sulfate to the coiled form (Sarkar and Doty, 1966) differs from that by mild heating of the helical form. Whether poly(L-serine) in 8 M LiCl is a better representation of the unordered form than coiled poly(L-lysine) is another matter to be settled. A new approach is to establish the reference values in eq 1 from proteins of known three-dimensional structure. Straus *et al.* (1969) resolved the CD spectra of myoglobin, hemoglobin, and lysozyme into several Gaussian bands for the three conformations. They found that the rotational strengths of each CD band for the protein helices differed from each other and were all less than those found for synthetic polypeptides. Saxena and Wetlaufer (1971) and Chen and Yang (1971) determined the X_H, X_β, and X_R in eq 1 from the measured CD, ORD, or b_0 of proteins of known structure. Assuming that distorted helices have a rotatory power intermediate between a helical and an unordered form, Saxena and Wetlaufer adopted as f_H the average of the lower (true α helices) and upper (true and distorted helices) limits. They found that the CD spectra of the helix and β form, based on three reference proteins, were close to, and the spectrum of the unordered form was considerably different from, the corresponding spectra of helical, β, and unordered poly(L-lysine).

In previous communications (Chen and Yang, 1971; Chen *et al.*, 1972), we used five reference proteins and computed the X_H, X_β, and X_R in eq 1 at different wavelengths by a least-squares method. Unlike Saxena and Wetlaufer (1971), we preferred maximum f_H values including distorted helices in our computations on the assumption that the rotatory power of the 3_{10}- and distorted helices were close to that of an α helix. This assumption seems to be supported by the recent theoretical calculations by Madison and Schellman (1972) of the optical activity of helical segments in four proteins. Our computed values of X_H, X_β, and X_R were significantly different from those of model polypeptides. In particular, the absolute values of X_H

† From the Department of Biochemistry and Biophysics and cardiovascular Research Institute, University of California, San Francisco, California 94143. *Received December 10, 1973.* This work was aided by NIGMS Grant GM-10880 and NHLI Grant HL-06285. Presented in part at the 57th Annual Meeting of American Societies for Experimental Biology, Atlantic City, N. J., April 1973 (Chen *et al.*, 1973) and the 9th International Congress of Biochemistry, Stockholm, July 1973 (Yang and Chen, 1973).

‡ Part of a dissertation submitted by Y.-H. Chen in partial fulfillment for the requirement of the degree of Doctor of Philosophy, University of California, San Francisco, Calif., 1973. Present address: Department of Biochemistry and Biophysics, University of California, Davis, Calif. 95616.

§ Present address: Radiopharmaceutics Division, Abbott Laboratories, North Chicago, Ill. 60064.

[1] Abbreviations used are: CD, circular dichroism; ORD, optical rotatory dispersion.

were smaller than those of helical poly(L-glutamic acid) and poly(L-lysine) (Cassim and Yang, 1970). This was attributed to the end effect on rotatory power of helices, as predicted by the theoretical calculations of Woody and Tinoco (1967). With X_H, X_β, and X_R thus determined, eq 1 was used conversely to estimate the fractions of helix, β, and unordered form from the CD spectrum of any protein.

In this work we first redetermine the reference values of X_H, X_β, and X_R in eq 1 as a result of recent refinement of the X-ray diffraction studies which has upgraded the percentages of helices and β form in some of the proteins we studied (see Chen *et al.*, 1972). Next, recognizing the importance of the end effect of helices, we introduce the following expression to account for the chain-length dependence of helices. At any fixed wavelength

$$X_H^n = X_H^\infty(1 - k/n) \quad (2)$$

where X_H^∞ is the reference value for a helix of infinite length, n the number of residues per helix, and k a wavelength-dependent constant ($1 > k/n > 0$). Implied in eq 1 is that the reference proteins used have a finite, overall average $\bar{n}$. For model polypeptides having n over 100, eq 2 would be unnecessary. We will also propose a refined treatment by removing this restriction on a single $\bar{n}$ for the five reference proteins. Such a refinement not only leads to a more reasonable determination of the fractions of helix and β form but also provides an estimation of the number and average size of helical segments in a protein molecule.

Method of Analysis

Theoretical calculations of the optical activity of a helix have resolved its CD spectrum into four optically active bands in the wavelength range of 180–250 nm (Woody, 1968). Experimentally, the CD spectrum of a helix in the 190–240-nm region can be resolved into three Gaussian bands at 222, 206, and 190 nm, which represent the n-π^*, π-$\pi_\parallel^*$ and π-$\pi_\perp^*$ transitions in accord with the linear dichroism study of helical polypeptide films (Holzwarth and Doty, 1965).

The rotational strength of each transition for the helix is chain length dependent (Woody and Tinoco, 1967; Vournakis *et al.*, 1968; Madison and Schellman, 1972). For a helix of n peptide units, the mean residue ellipticity of each of its Gaussian CD bands (j) can thus be expressed as (for details, see Chen, 1973)

$$[\theta_j]_H^n = [\theta_j]_H^\infty(1 - k_j/n) \quad (3)$$

and

$$[\theta_j]_H^\infty = [\theta_j^\circ]_H^\infty \exp[-(\lambda - \lambda_j)^2/\Delta_j^2] \quad (4)$$

where k_j, λ_j, and Δ_j are the chain-length-dependent factor, band position and half-bandwidth of the jth transition and $[\theta_j]_H^n$, $[\theta_j]_H^\infty$ and $[\theta_j^\circ]_H^\infty$ are the mean residue ellipticities of a helix of n and infinite residues and the maximum ellipticity of infinite length at wavelength λ_j. The contribution of an $n$mer helix is simply the sum of three Gaussian CD bands, $\Sigma[\theta_j]_H^n$ (eq 3 and 4). We do not use the CD of helical polypeptides for $[\theta]_H^\infty$ because ellipticities vary among the polypeptides studied and their averages do not necessarily represent $[\theta]_H$ for the proteins and because synthetic polypeptides resemble fibrous rather than globular proteins. The different environment effect on the exposed and buried amino acid residues could also change the positions and magnitudes of the three helical bands. For instance, helical polypeptides show a positive CD band at 191 nm, but myoglobin has a CD maximum at 193 nm.

The β form in protein molecules is usually far from ideal. Even in the ideal case, Woody (1969) showed that its ellipticity depends not only on the number of residues per segment but also on the number of strands and the polarity of the strands. No simple relationship such as eq 2 has been found for the β form. Accordingly, in our analysis $[\theta]_\beta$ represents a statistical average for the reference proteins used. The same is true for the ellipticity of the unordered form, $[\theta]_R$. Here again we do not use the $[\theta]_R$ of coiled polypeptides because the unordered form of proteins does not resemble the "random coils" of charged polypeptides, which are sensitive to the ionic strength of the solvent used. We do not attempt to resolve the CD bands of the β and unordered forms. On a molar basis the ellipticities of the helices dominate those of the β and unordered forms. Thus, our method of analysis for proteins containing a moderate amount of helices is not significantly affected by the use of statistical averages of the other two forms.

For each ith helical segment of n_i residues, we have at wavelength λ

$$[\theta]_{H,i}^{n_i} = \sum_{j=1}^{3} [\theta_j]_{H,i}^{n_i} \quad (5)$$

The mean residue ellipticity, $[\theta]$, of a protein can then be expressed as

$$[\theta] = \Sigma f_{H,i}[\theta]_{H,i}^{n_i} + f_\beta[\theta]_\beta + f_R[\theta]_R \quad (6)$$

where the summation represents the sum of weighted $[\theta]_H$'s for all helical segments. Since the number of helical segments, i, is simply related by

$$f_H = \sum_i n_i/N = i\bar{n}/N \quad (7)$$

where N is the total number of residues and $\bar{n}$ the average number of residues per segment of the protein molecule, it can be shown that by substituting eq 3, 5, and 7 into eq 6

$$[\theta] = (f_H - ik/N)[\theta]_H^\infty + f_\beta[\theta]_\beta + f_R[\theta]_R \quad (8)$$

or

$$[\theta] = f_H[\theta]_H^\infty(1 - k/\bar{n}) + f_\beta[\theta]_\beta + f_R[\theta]_R \quad (9)$$

or

$$[\theta] = f_H[\theta]_H^{\bar{n}} + f_\beta[\theta]_\beta + f_R[\theta]_R \quad (10)$$

where

$$[\theta]_H^\infty \equiv \sum_{j=1}^{3} [\theta_j^\circ]_H^\infty \exp[-(\lambda - \lambda_j)^2/\Delta_j^2] \quad (11)$$

and

$$k[\theta]_H^\infty \equiv \sum_{j=1}^{3} k_j[\theta_j^\circ]_H^\infty \exp[-(\lambda - \lambda_j)^2/\Delta_j^2] \quad (12)$$

Equation 10 is much the same as eq 1 but $[\theta]_H^{\bar{n}}$ is explicitly defined at a given $\bar{n}$. The five reference proteins we used had an overall $\bar{n}$ of about 10 to 11 (see Results and Discussion), so $[\theta]_H^{\bar{n}}$ will underestimate (or overestimate) f_H if the $\bar{n}$ of an unknown protein is smaller (or larger) than 10 to 11.

Equations 8, 9, and 10 are also applicable to ORD, so we can replace $[\theta]$ by $[m0$ or b_0 (of the Moffitt equation) and k by k' or k'' (another two constants). However, $[m]_H^\infty$, b_0^∞, k, and k'' can no longer be represented by eq 11 and 12. They must be determined experimentally (see Results and Discussion).

In the past, $[\theta]_H^\infty$ or $[m]_H^\infty$ based on synthetic polypeptides was used for estimating the helical content in a protein molecule (Greenfield and Fasman, 1969). This is equivalent to writing eq 9 as

$$X = f_H(\text{app})X_H^\infty + f_\beta X_\beta + f_R X_R \quad (13)$$

Here the X's can be either $[\theta]$'s or $[m]$'s at wavelength λ.

TABLE I: The Percentages of f_H and f_β of Eight Proteins Based on X-Ray and from CD Analysis.

Proteins	Form	X-Ray[a]	i	$\bar{n}$	CD[b]		
					$\bar{n} = 11$[c]	$\bar{n} = 9$[d]	$\bar{n} = \infty$[e]
Myoglobin	H	0.79	9	13.4	0.83	0.88	0.62
	β	0			0.05	0.03	0
Lysozyme	H	0.41	7	7.6	0.37	0.35	0.28
	β	0.16			0.11	0.27	0.09
Lactate dehydrogenase	H	0.45	12	12.5	0.39	0.40	0.29
	β	0.24			0.17	0.02	0.15
Papain	H	0.28	5	12.0	0.29	0.29	0.22
	β	0.14			0.15	0.09	0.14
Ribonuclease	H	0.19	3	8.0	0.25	0.22	0.19
	β	0.38			0.44	0.47	0.42
Insulin	H	0.45	3	7.7	0.39	0.39	0.29
	β	0.12			0.19	0.21	0.17
Cytochrome *c*	H	0.39	6	6.8	0.32	0.33	0.23
	β	0			0.19	0	0.19
Nuclease	H	0.24	3	12.0	0.36	0.35	0.26
	β	0.15			0.15	0.11	0.13

[a] References of the eight proteins (in descending order): (1) Kendrew *et al.* (1960); Kendrew (1962); (2) Blake *et al.* (1965); (3) Rossman *et al.* (1972); (4) Drenth *et al.* (1971); (5) Kartha *et al.* (1967); (6) Adams *et al.* (1969); Blundell *et al.* (1971); (7) Dickerson *et al.* (1971); (8) Arnone *et al.* (1969, 1971). [b] Based on CD spectra between 205 and 240 nm at 1-nm intervals. [c] Based on five reference proteins (see Table II). [d] Based on eight reference proteins. [e] Table II, except $[\theta]_H^{\bar{n}}$ is replaced by $[\theta]_H^{\infty}$ (see Table IV).

Clearly, the apparent f_H will be smaller than the true f_H because of the omission of the term $(1 - k/\bar{n})$ (*cf.* eq 9 and 10).

We solved the reference values of $X_H^{\bar{n}}$, X_β, and X_R in eq 10 at 1-nm intervals by using five or eight reference proteins to test whether the inclusion of the three additional proteins has any advantage (here again X can be $[\theta]$ or $[m]$). Once the X parameters are determined, eq 10 is used conversely to compute the f_H, f_β, and f_R of any protein from its CD or ORD spectrum (at 1-nm intervals). The computer program used was BMDX 85T, UCLA, with the conditions that Σf's = 1 and $1 \geq f_H$ (or f_β) ≥ 0. The f_R is replaced by $(1 - f_H - f_\beta)$ and could be negative, although in practice we have not observed such anomalous results. Our analysis can also be complicated by several factors such as the assignment of a residue as helical, β, or unordered and the rotatory contributions of non-peptide chromophores. Details have been discussed previously (Chen *et al.*, 1972).

The right side of eq 8 actually has 18 parameters (not including N and λ; *cf.* eq 11 and 12). In analyzing the CD of a reference protein, we can fix six parameters, that is, three f's and i from the X-ray diffraction results and $[\theta]_\beta$ and $[\theta]_R$ borrowed from the solution of eq 10. It is reasonable to assume that three CD bands for the helical segments in most, if not all, globular proteins may have the same band positions, λ_j's. Straus *et al.* (1969) found by curve resolving the CD spectra of several proteins that the three helical bands were all located around 223, 206, and 193 nm. For myoglobin the n-π^*, π-$\pi_\parallel^*$ and π-$\pi_\perp^*$ transitions were positioned at 223.4, 206.6, and 193.5 nm. If these values are applicable to other proteins, only nine parameters, $3[\theta_j^\circ]_H^\infty$'s, $3\Delta_j$'s, and $3k_j$'s, in eq 8 remain to be determined.

Equation 8 or 9 introduces two new parameters, i (or $\bar{n}$) and k (*cf.* eq 1). With known f's and i's or $\bar{n}$'s of the reference proteins, we can in principle determine $[\theta]_H^\infty$, $[\theta]_\beta$, $[\theta]_R$, and k in eq 8 or 9. They can in turn be used for computing the f's and i of any protein from its CD spectrum. In practice, however, this approach was unsatisfactory. The underlying assumption for using eq 8 or 9 to solve the $[\theta]$ reference values is that the $[\theta_j^\circ]_H^\infty$, λ_j, and δ_j's of the three helical CD bands are identical for all proteins (*cf.* eq 11 and 12). This is not so (see Results and Discussion). Another possibility is that the nonlinear fitting of eq 8 or 9 may simply not work, thus resulting in incorrect reference ellipticities.

Experimental Section

Eight proteins were studied in this work: sperm-whale myoglobin, egg-white lysozyme, dogfish lactate dehydrogenase, papain from papaya, bovine pancrease ribonuclease, bovine insulin, horse heart cytochrome *c*, and *Staphylococcus aureus* nuclease. The CD and ORD spectra of the eight proteins have been reported previously (Chen *et al.*, 1972). The CD data between 190 and 245 nm at 1-nm intervals will be provided upon request.

The BMDX 85T program was computed with an IBM system/360 at the computing center of the University of California at San Francisco.

Results and Discussion

Determination of $X_H^{\bar{n}}$, X_β, and X_R. Recent X-ray diffraction studies have revised and upgraded the fractions of helix and β form of two of the five reference proteins (papain and lactate dehydrogenase) that had been used in our previous work (Chen *et al.*, 1972). Table I (left half) lists the maximum values of f_H and f_β, including both true and distorted helices and β forms, for the eight proteins studied together with the number of helical segments, i, and average number of residues per segment, $\bar{n}$, of each protein. Ideally, the set of reference proteins should cover a wide range of f_H and f_β. In addition, the $\bar{n}$ of the helical segments of these proteins should preferably be within a narrow range because of the chain-length dependence of $X_H^{\bar{n}}$ ($[\theta]_H^{\bar{n}}$ or $[m]_H^{\bar{n}}$). A set of proteins that satisfy both conditions is, however, difficult to find. Thus, we solved eq 10 by (1)

TABLE II: The CD and ORD of Helix, β, and Unordered Forms Based on Five Proteins.[a]

[θ][a]				[m][a]			
λ (nm)	H	β	R	λ (nm)	H	β	R
190	70,100	−2,880	−20,300	197	64,100	−2,960	−5,130
193	77,000	12,900	−36,000	200	76,400	15,100	−21,600
196	68,200	14,000	−37,900	203	69,900	12,200	−20,900
199	37,200	6,050	−23,200	206	56,500	7,600	−18,500
201	16,800	8,810	−23,300	209	37,300	4,520	−13,000
204	−9,120	300	−11,300	212	27,800	3,710	−11,800
207	−22,300	−4,320	−5,770	215	22,300	−230	−9,520
210	−26,400	−8,190	−2,200	218	15,800	−4,380	−6,220
213	−24,800	−8,680	−850	221	9,620	−8,560	−3,630
216	−26,600	−9,210	1,230	224	2,080	−9,580	−2,280
219	−28,900	−6,890	1,720	227	−5,320	−8,660	−1,320
222	−30,000	−3,360	1,580	230	−11,600	−5,020	−1,360
225	−28,700	1,540	260	233	−13,800	−1,900	−1,740
228	−24,000	4,390	−480	236	−13,500	−1,090	−1,010
231	−17,300	4,580	−770	239	−11,700	−940	−670
234	−11,300	3,540	160	242	−9,340	−690	−950
237	−6,250	2,410	−90	245	−7,300	−330	−1,230
240	−2,950	3,370	−1,040	248	−5,650	−350	−1,340
243	−1,230	2,040	−930	251	−4,390	−470	−1,390

[a] Dimensions: (deg cm^2) dmol^{-1}.

choosing the same five reference proteins used previously and (2) including three additional proteins (last three in Table I; their three-dimensional structures are not completely clear and subject to future refinement). The eight proteins should give better statistical averages of the X-reference values. However, the $\bar{n}$ of 2 of the additional proteins was considerably smaller (6.8 and 7.7) than the overall average $\bar{n}$ (10.7) of the first set and $\bar{n}$ of the third protein was only slightly greater (12.0). This will affect the values of $X_H^{\bar{n}}$, especially when $\bar{n}$ is small, and also the value of X_β and X_R. Thus, the three additional proteins increase the range of $\bar{n}$ of the helical segments. The choice of these two sets of proteins is by necessity arbitrary, but the reference values so determined turn out to be comparable.

1. With Five Reference Proteins. Table II lists the computed CD and ORD of helix, β, and unordered forms from the experimental data of the first five proteins in Table I. Figures 1 and 2 show the corresponding spectra of the three conformations (points are used when a smooth curve cannot be drawn). The CD extrema of the helix are located at 222 (−), 209 (−), and 192 (+) nm and the ORD extrema at 233 (−) and 200 (+) nm. Their absolute values are about 5% smaller than the previously reported values (Chen *et al.*, 1972) because of the larger revised f_H's used here. The CD of the β form has a minimum at 216 nm and two small positive bands between 225 and 250 nm and the corresponding ORD has a minimum at 223 nm and a maximum around 200–202 nm with slight differences in magnitudes from those found previously. The computed CD and ORD of the unordered form differ considerably from those in the previous work (Chen *et al.*, 1972), especially in the region below 205 nm. Figure 1 still shows a CD minimum at 194 nm as before, but the new value is close to that of the poly(L-glutamic acid) "coil" at 197-nm minimum. Like the "coil," there is a small positive band at 218–219 nm. The one positive and two negative bands between 225 and 250 nm are too small to be significant. The ORD minimum of the unordered form is located at 202 nm and its magnitude is again close to that of the "coil." That the unordered form in this work seems to resemble the coiled form of synthetic polypeptides is intriguing. These new X_R values do not, however, affect our determinations of f_H, f_β, and f_R too much.

The new isorotational point of the β and unordered form remains at 233 nm and the corresponding isodichroic point is lo-

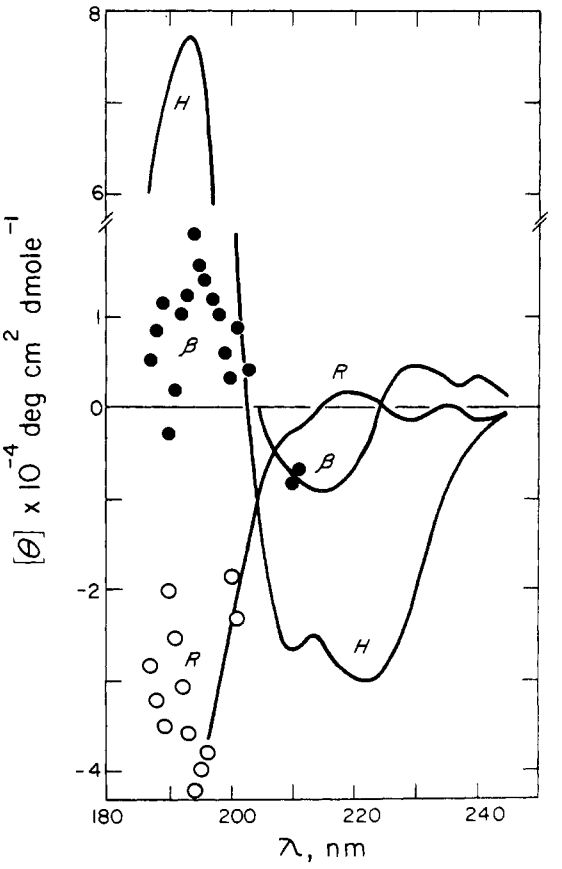

FIGURE 1: Circular dichroism of the helix (H), β, and unordered (R) form computed from the CD of five proteins.

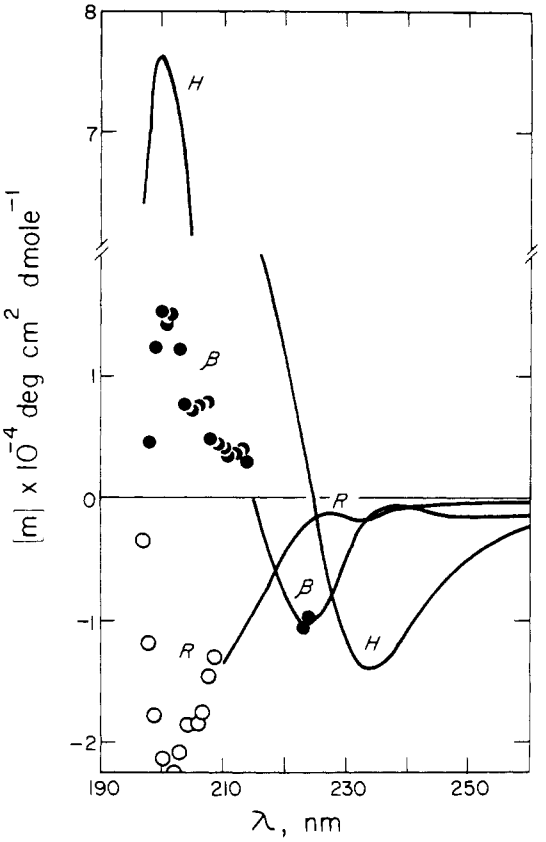

FIGURE 2: Optical rotatory dispersion of the helix (H), β, and unordered form (R) computed from the ORD of five proteins.

cated at 224 nm, which is still close to the helical minimum at 222 nm. Likewise, b_0 for the β and unordered forms was very small as compared with the helical b_0. Thus, $[m]_{233}$, $[\theta]_{222}$, and b_0 become essentially a linear function of f_H only. Equation 10 at the isodichroic or isorotational wavelength becomes

$$X = (X_H^{\bar{n}} - X_R)f_H + X_R \qquad (14)$$

with $X = [\theta]$ or $[m]$ (a similar equation can be written for b_0). Equation 14 provides a quick estimate of the helical content in a protein molecule (Myer, 1970; Chen *et al.*, 1972). However, the coefficients $(X_H^{\bar{n}} - X_R)$ and X_R will be affected by the choice of reference proteins and, furthermore, $X_H^{\bar{n}}$ is chain length dependent. Therefore, we will not attempt to revise the previously reported coefficients (Chen *et al.*, 1972).

2. With Eight Reference Proteins. Increasing the number of reference proteins from five to eight would reduce the dominance of the rotatory contributions of the helices in myoglobin, and the inclusion of insulin and cytochrome *c* lower the overall $\bar{n}$ of the reference proteins. According to eq 8 and 9, the absolute values of $[\theta]_H^{\bar{n}}$ are expected to be smaller than those based on five proteins. The CD profile of the helical form based on eight proteins is close to that shown in Figure 1, except that the absolute values at the extrema are reduced by several per cents; for example, $[\theta]_{222} = -27{,}900$, $[\theta]_{210} = -25{,}700$, and $[\theta]_{193} = 71{,}200$ (deg cm^2) dmol^{-1} (*cf.* Table II). On the other hand, the CD spectra of the β and unordered form do change significantly; for instance, the β minimum is shifted from 217 nm to about 210 nm with $[\theta]_{210} = -11{,}400$, which is more negative than that in Figure 1, and the *R* minimum from 196 to 193 nm with $[\theta]_{193} = -23{,}000$, which is considerably smaller in magnitude than that in Figure 1. Nevertheless, the optical activity of the helical form still dominates over that of β and unordered form in the wavelength range studied. The number of five or eight reference proteins used seems to be unimportant in the

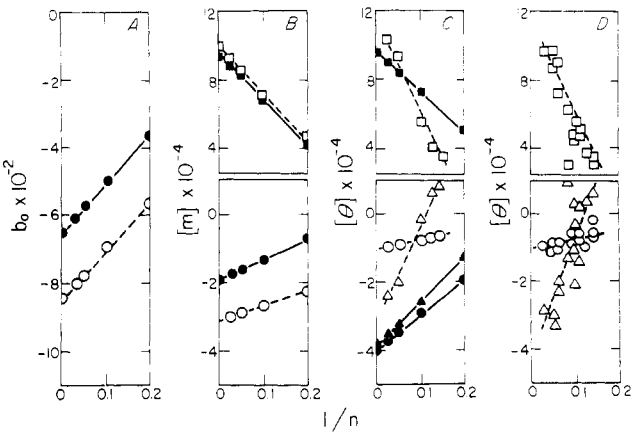

FIGURE 3: Variation of the optical activity of helices with chain length (n). Open symbols, theoretical calculations: A and B, Woody and Tinoco (1967); in B, circles at 233 nm and squares at 200 nm. C and D, Madison and Schellman (1972); circles at 222 nm, triangles at 207 nm, and squares at 194 nm; C, for ideal α helix; D, for helical segments in myoglobin, lysozyme, ribonuclease S, and α-chymotrypsin. Filled symbols, computed values of this work; in C, triangles and squares at 210 and 193 nm, respectively, instead.

computation of the three forms, unless the $\bar{n}$ of a protein to be studied happens to be closer to the overall $\bar{n}$ of the particular reference set (see below).

Determination of $f_H(\bar{n})$, f_β, and f_R. With the CD spectra of the three conformations based on five (Table II) or eight (not shown) reference proteins, we can determine the fractions of helix and β form of a protein molecule by fitting eq 10 with the experimental CD spectrum for the protein between 190 and 245 nm at 1-nm intervals. Previously (Chen *et al.*, 1972), we used only the CD above 205 nm because the data at shorter wavelengths were less precise. However, we have since found that the computed f's are little affected when the data between 190 and 205 nm are included in the computation. Our computed results of eight proteins are included in Table I (right half). With a few exceptions the agreement with the X-ray diffraction results is satisfactory. It seems that the results based on five reference proteins (Table II; $\bar{n} = 11$) are closer to the known structures of the proteins than those based on eight proteins $\bar{n} = 9$). The notable exception is cytochrome *c*, which shows significant amount of the computed β form based on $\bar{n} = 11$, whereas the X-ray results indicate no β form at all.

The length of helical segments in a protein molecule are not uniform and most segments are short. The $\bar{n}$ also varies from one protein to another, but the overall average of $\bar{n}$ of the few proteins tested was around 10 to 11: with five reference proteins $\bar{n} = 11.3$ and with eight proteins $\bar{n} = 10.6$. However, we will show later (Figures 4 and 5) that $[\theta]_H^{\bar{n}}$'s of the five and eight reference proteins are close to optical activities of helices having $\bar{n} = 11$ and 9, respectively (in previous work (Chen *et al.*, 1972), in which some of the shorter helical segments were not counted, $\bar{n} = 14$). The computed f_H values in this work are slightly larger than those reported in the previous work, because the $[\theta]_H^{\bar{n}}$ values with $\bar{n} = 11$ or 9 are smaller than those with $\bar{n} = 14$. On the other hand, the computed f_β's of this work are still close to previously published values (except cytochrome *c*).

According to eq 10, the estimate of helical content will be high or low according to how well the helical length of the protein agrees with the overall $\bar{n}$ of the chosen reference proteins. Thus, the computed f_H of myoglobin, which has an $\bar{n}$ larger than either 11 or 9, is overestimated and more so with $[\theta]_H^{\bar{n}}$ based on eight reference proteins. On the other hand, the computed f_H of insulin, with a smaller $\bar{n}$, is underestimated. There

TABLE III: Parameters for the Best Fit of Helical CD Bands of Eight Proteins.[a]

Proteins	$[\theta_j°]_H^\infty \times 10^4$ (deg cm^2) dmol^{-1}			Δ_j (nm)			k_j		
	n–π^*	π–$\pi_\parallel^*$	π–$\pi_\perp^*$	n–π^*	π–$\pi_\parallel^*$	π–$\pi_\perp^*$	n–π^*	π–$\pi_\parallel^*$	π–$\pi_\perp^*$
Myoglobin[b]	−3.68	−3.65	10.3	10.4	9.8	8.3	2.5	3.5	2.5
	−3.73	*−3.72*	*10.1*	*10.8*	*8.9*	*8.4*	*2.5*	*3.5*	*2.5*
Lysozyme	−3.90	−4.46	9.4	12.0	8.4	7.0	2.9	3.5	0.9
Lactate dehydrogenase	−4.00	−4.00	9.0	11.7	8.3	8.0	4.4	7.5	2.1
Papain	−4.00	−3.99	10.3	10.1	9.6	9.2	2.5	4.2	2.5
Ribonuclease	−4.77	−6.00	10.2	9.5	10.8	8.8	2.2	2.8	0.5
Insulin	−3.76	−4.70	12.0	11.3	7.0	10.3	2.4	2.7	2.5
Cytochrome *c*	−3.98	−4.28	12.8	9.8	7.5	10.0	1.8	3.1	2.2
Nuclease	−5.44	−4.15	13.0	11.1	7.7	8.4	1.6	2.2	0

[a] See text for details; $[\theta_j°]_H^\infty$, Δ_j, and k_j are defined in eq 11 and 12. [b] The italic values are based on a six-parameter solution with Δ_j's prefixed.

are some exceptions; for instance, ribonuclease gave an overestimated f_H and lactate dehydrogenase an underestimated f_H. The appearance of β form in cytochrome *c* based on five reference proteins is also unexpected.

The use of X_H^∞ based on synthetic polypeptides is essentially equivalent to solving f_H, f_β, and f_R by eq 13. Neglect of the term $(1 - k/\bar{n})$, which is less than one, would lower the computed f_H. Thus, the apparent f_H in the last column of Table I is smaller than f_H based on $X_H^{\bar{n}}$ with $\bar{n} = 11$ or 9. By multiplying the apparent f_H by $(1 - k/\bar{n})$ with $k = 3$ (to be shown later) and $\bar{n} = 10$ (eq 9) or by adding ik/N to the apparent f_H (eq 8), the corrected f_H of each protein in Table I will be close to the maximum f_H established from X-ray diffraction results. In the past, the CD and ORD of synthetic polypeptides as reference compounds were used for estimating the helical contents of proteins, which are often underestimated because of the neglect of this chain-length dependence factor.

Chain-Length Dependence of X_H^n. Previously we have shown that b_0, $[m]_{233}$, and $[\theta]_{222}$ of the helices based on the theoretical calculations of Woody and Tinoco (1967) vary linearly with $1/n$ (Chen *et al.*, 1972; also Figure 3A,B). This relationship is further substantiated by the recent calculations of the CD extrema at 222, 207, and 194 nm (Madison and Schellman, 1972) for several polypeptides and proteins (Figure 3C,D), except when the helical segments contain less than six peptide units. The agreement shown in Figure 3 between our computed values (filled symbols; to be described later) and theoretical calculations (open symbols) is demonstrated by the virtually identical slopes of the corresponding straight lines, although the numerical values do differ. If $\bar{n}$ for poly(L-glutamic acid), myoglobin, and lysozyme is assigned as infinite, 13.4 and 7.5, respectively, the rotational strengths of the n–π^* and two π–π^* transitions for the helices (Straus *et al.*, 1969) also vary linearly with $1/\bar{n}$. Of particular interest is the finding of Madison and Schellman (1972) that the calculated ellipticities of the helical segments of myoglobin, lysozyme, ribonuclease, and α-chymotrypsin are close to those of ideal helices of the same length (*cf.* Figure 3C,D). As far as optical activity is concerned, the distorted helices, whose dihedral angles may be deviated from the true ones, can be assumed to behave like the ideal helices of the same length. This finding reinforces our previous contention that maximum f_H's in proteins should be counted and used in the determination of $X_H^{\bar{n}}$. It seems not advisable to lower the f_H's by averaging the true and total (including distorted) helical segments so that the computed $X_H^{\bar{n}}$ can be raised to that of synthetic polypeptides.

Chain-Length Dependence Factor, k, of the Helix. In principle, the $[\theta]$ parameters and k at each wavelength in the simplified eq 8 or 9 can be determined from the experimental CD spectra of the five or eight reference proteins (Table I). The $[\theta]_H^\infty$'s so computed turned out to be unsatisfactory (for reasons given under Method of Analysis); they showed a single minimum instead of the characteristic double minimum for the helix and their absolute values did not agree with those of synthetic polypeptides except above 225 nm. Next, we tried to eliminate k as a variable by assigning some arbitrary positive value for k. When we confined k within a range of 2–4, the computed $[\theta]_H^\infty$'s resembled the typical CD of helical polypeptides. This defines the range of k.

With nine parameters (three f's, i, three λ_j's, and $[\theta]_\beta$ and $[\theta]_R$ from Table II) fixed, we can determine the remaining nine parameters in eq 8, 11, and 12 for each reference protein by a nonlinear least-squares fitting of its CD spectrum (see Methods of Analysis). These computed parameters are sensitive to any variation in the counting of f_H and i (especially when the f_H and i of a protein are small). Table III lists the computed maximum ellipticities, bandwidths, and k_j's of the three helical CD bands for the eight proteins. These parameters do vary from one protein to another, but they are within an acceptable range.

In a nonlinear fitting, the fewer parameters used, the more reliable the solution. Accordingly, we further fixed the three bandwidths found by Straus *et al.* (1969) for myoglobin and solved the remaining six parameters from the CD of myoglobin between 190 and 245 nm. The best-fit results (italic) are included in Table III for comparison. The $[\theta^\circ]_H^\infty$ and k_j's from nine- and six-parameter fittings are virtually identical. With these parameters (italic in Table III), we can calculate by eq 9, 11, and 12 the CD spectra for helices of different peptide units. The parameters of myoglobin were used because this protein has a high helical content, which is responsible for most of its optical activity, and no β form, and therefore the results are more reliable than those of other proteins. Through the Kronig–Kramers transform we can also convert the computed CD and ORD and determine the b_0 of the Moffitt equation from the ORD in the visible region (Cassim and Yang, 1970).

Table IV lists the mean residue ellipticities and rotations for a helix of infinite length, $[\theta]_H^\infty$ and $[m]_H^\infty$, together with the chain-length dependence factors, k and k'. Figures 4 and 5 show the CD and ORD for helices of various peptide units. The absolute values at any wavelength increase with the chain length and reach a plateau at about 40 residues. All CD spec-

TABLE IV: The Computed CD and ORD of a Helix of Infinite Length and Chain-Length-Dependent Factors, k and k'.[a]

CD			ORD		
λ (nm)	$[\theta]_H^\infty$	k	λ (nm)	$[m]_H^\infty$	k'
190	82,300	2.48	194	34,900	3.38
193	95,600	2.45	197	75,000	2.97
196	80,900	2.36	200	94,000	2.87
199	44,200	2.05	203	88,700	2.78
201	16,700	0.91	206	68,600	2.57
204	−16,400	4.61	209	46,900	2.15
207	−33,800	3.60	212	31,500	1.51
210	−38,500	3.35	215	23,200	0.97
213	−37,800	3.14	218	17,400	0.67
216	−37,100	2.89	221	9,960	
219	−38,400	2.69	224	150	
222	−39,500	2.57	227	−9,600	4.15
225	−38,600	2.53	230	−16,300	3.31
228	−31,100	2.51	233	−18,700	3.10
231	−22,200	2.50	236	−17,600	3.06
234	−13,600	2.50	239	−14,700	3.10
237	−7,050	2.50	242	−11,500	3.19
240	−3,120	2.50	245	−8,840	3.32
243	−1,170	2.50	248	−6,890	3.46

[a] Dimensions of $[\theta]_H^\infty$ and $[m]_H^\infty$: (deg cm²) dmol⁻¹.

tra have a double minimum at 222 and 210 nm and a maximum at 193 nm and the corresponding ORD's have a 233-nm minimum and a maximum at 200–201 nm. The magnitudes of these extrema for a helix of infinite length are close to those of synthetic polypeptides such as poly(L-glutamic acid) and poly(L-lysine) (Cassim and Yang, 1970). For ORD between

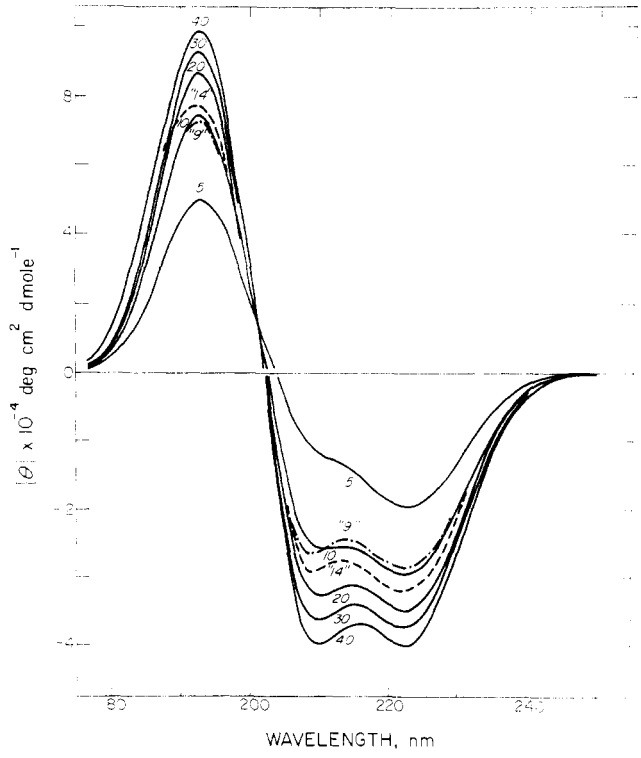

FIGURE 4: The computed circular dichroism for helices of different chain lengths. Numerals refer to the numbers of amino acid residues. Curve "9" is the statistical average based on five proteins (Table II); curve "14" is taken from a previous work (Chen *et al.*, 1972).

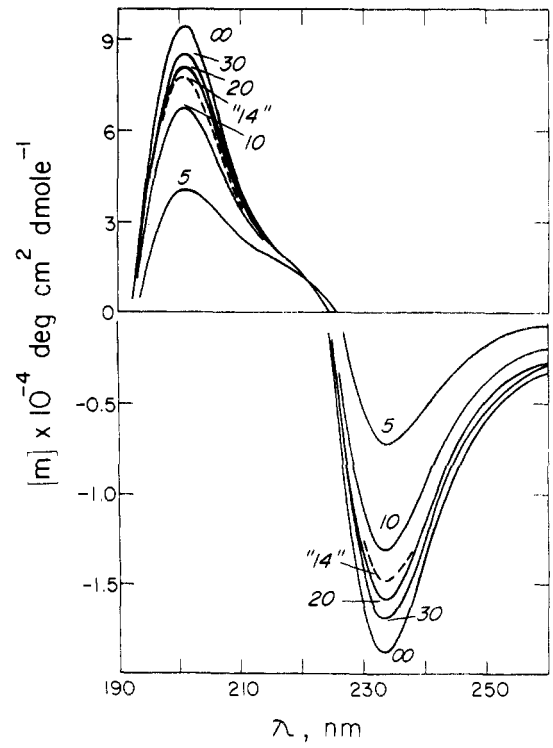

FIGURE 5: The computed optical rotatory dispersion for helices of different chain lengths. Numerals are the same as in Figure 4.

300 and 600 nm, the b_0 of an helix with 5, 10, 15, 20, 25, 30, 40, and ∞ residues is −340, −490, −540, −560, −570, −580, −600, and −630 (deg cm²) dmol⁻¹. If the wavelength range is limited to 400 and 600 nm, the corresponding b_0 varies from −360 to −650 (deg cm²) dmol⁻¹. Thus, the calculated b_0 of an infinite helix agrees well with the current accepted value of −630 based on synthetic polypeptides (Moffitt and Yang, 1956). We tested the linearship between X_H^n and $1/n$ for CD and ORD at 1-nm intervals and also for the b_0 values (see Figure 3). Equation 2 holds well in all three cases, except for CD near 200–203 nm and ORD near 221–224 nm (the crossovers in Figures 4 and 5). These results are all based on the CD of myoglobin.

Determination of f_H, f_β, and i *or* $\bar{n}$. The assumption of an overall average $\bar{n}$ that is applicable to any globular protein is of course an oversimplification, since no two proteins can have the same distribution of their helical segments except perhaps in a few cases such as myoglobin and hemoglobin. This restriction of a single $\bar{n}$ for all proteins, however, can be removed by fitting the CD of a protein with eq 8 or 9. Its solution would provide not only an estimate of f_H and f_β but also $\bar{n}$ and i in a protein molecule. Numerical fitting of eq 8 or 9 with many more parameters (*cf.* eq 11 and 12) than in eq 1 is of course more complicated; therefore, these computed results should be interpreted with caution.

With the maximum ellipticities and band positions of the three Gaussian bands and their chain-length dependence factors for the helices determined from the CD spectrum of myoglobin (Table IV), we can rewrite eq 9 (*cf.* eq 9, 11, and 12) as

$$[\theta] = f_H\{-3.73 \times 10^4 \times (1 - 2.50/\bar{n}) \exp[-(\lambda - 223.4)^2/\Delta_{\pi-\pi^*}{}^2] - 3.72 \times 10^4(1 - 3.50/\bar{n}) \exp[-(\lambda - 206.6)^2/\Delta_{\pi-\pi_\parallel^*}{}^2] + 10.1 \times 10^4(1 - 2.50/\bar{n}) \exp[-(\lambda - 193.5)^2/\Delta_{\pi-\pi_\perp^*}{}^2]\} + f_\beta[\theta]_\beta + f_R[\theta]_R \quad (15)$$

TABLE V: Determination of Helix and β Form of Eight Proteins by CD Methods.

Proteins	X-Ray Studies				Method 1				Method 2			
	f_H	f_β	i	$\bar{n}$	f_H	f_β	i	$\bar{n}$	f_H	f_β	i	$\bar{n}$
Myoglobin	0.79	0	9	13.4	0.80	0	9	14	0.77	0.02	8	14
Lysozyme	0.41	0.16	7	7.6	0.39	0.23	6	9	0.37	0.15	5	11
Lactate dehydrogenase	0.45	0.24	12	12.5	0.53	0.19	27[a]	7[b]	0.58	0.15	33[a]	6[b]
Papain	0.28	0.14	5	12.0	0.30	0.11	5	12	0.33	0.13	8	9
Ribonuclease	0.19	0.38	3	8.0	0.20	0.35	2	15[a]	0.22	0.34	2	15[a]
Insulin	0.45	0.12	3	7.7	0.42	0.39[a]	3	7	0.47	0.40[a]	2	10
Cytochrome *c*	0.39	0	6	6.8	0.34	0.24[a]	4	10	0.46	0.25[a]	8	6
Nuclease	0.24	0.15	3	12.0	0.38	0.15	6	10	0.35	0.17	5	11

[a] Grossly overestimated. [b] Grossly underestimated.

After replacing f_R by $1 - f_H - f_\beta$ and taking $[\theta]_\beta$ and $[\theta]_R$'s from Table II, only six parameters remain in eq 18. We set the boundary conditions of these parameters as $0 \leq f_H$ (or f_β) ≤ 1, $5 \leq \bar{n} \leq 15$ or higher, and $7 \leq \Delta_j$'s ≤ 14. Equation 15 can then be fitted with the ellipticities of a protein from 190 and 245 nm at 1-nm intervals. A computed value is considered unsatisfactory if the minimization of error process does not converge within the defined region of that parameter. However, because of low precision of the CD measurements below 200 nm, $\Delta_{\pi-\pi_\perp}*$ for the 190-nm band was assumed satisfactory even when its computed value reached the upper or lower limit.

We tested eq 15 with the eight proteins in Table I and found that the computed f_H and f_β's were close to the X-ray diffraction results except insulin and cytochrome *c* had too large f_β and f_H and lactate dehydrogenase and nuclease had too large f_H. But the computed $\bar{n}$'s were disappointing; in some cases they were much overestimated and in others underestimated. The Δ_j's were within the boundary conditions, although they varied from one protein to another. To further test the effect of variations in Δ_j's on other computed parameters, we deliberately fixed the three Δ_j's with values obtained from the CD of myoglobin (see Table III). The variables in eq 15 were thus reduced to 3 (f_H, f_β, and $\bar{n}$). The resultant f_H and f_β's thus determined were still close to those obtained with Δ_j's as variables and the $\bar{n}$'s deviated from the X-ray diffraction results. Thus, direct computations of eq 15 with the CD spectrum of a protein do provide reasonable estimates of its f_H and f_β, but not of its $\bar{n}$ or i. As a compromise we propose the following three approaches, although other possible methods of analysis are not ruled out.

METHOD 1: COMBINING CD SPECTRUM WITH $[m]_{233}$. We eliminate $\bar{n}$ from eq 15 by introducing another equation which itself is a function of f_H and $\bar{n}$. For instance, the virtually isorotational $[m]_{233}$ for the β and unordered form has an average value of -1820 (deg cm^2) dmol^{-1}. Likewise, the b_0 of the Moffitt equation is insignificant for the two conformations. Thus, we have

$$f_H \simeq ([m]_{233} - [m]_{\beta R})/([m]_H^\infty(1 - k'/\bar{n}) - [m]_{\beta R}) \quad (16)$$

and

$$f_H \simeq b_0^\infty(1 - k''/\bar{n})/(-630) \quad (17)$$

With the numerical values in Table II ($[m]_{\beta R}$ is an average of $[m]_\beta$ and $[m]_R$ at 233 nm), eq 16 can be expressed as

$$[m]_{233} \simeq [-18{,}700(1 - 3.10/\bar{n}) + 1820]f_H - 1820 \quad (18)$$

Likewise, eq 17 becomes

$$b_0 \simeq -630(1 - 2.28/\bar{n})f_H \quad (19)$$

where the chain-length dependence factor, k'', was found to be 2.28. A combination of eq 15 and 18 or eq 19 eliminates $\bar{n}$ and reduces the variables in eq 15 to five (f_H, f_β, and three Δ_j's), which can be fitted with the CD spectrum of a protein. We chose $[m]_{233}$ in our analysis because this quantity is simple to measure and of high precision. Once f_H is determined, $\bar{n}$ or i (see eq 7) can easily be calculated from eq 18.

Table V lists the best-fit results of f_H and f_β of eight proteins together with i and $\bar{n}$ of the helical segments under method 1. The computed f_H's are improved as compared with the CD results in Table I. The computed f_β's are reasonably good, except for two bad cases, insulin and cytochrome *c*. The computed $\bar{n}$ and thereby i values are more or less satisfactory, except that the $\bar{n}$ is too small for lactate dehydrogenase (its i is too large) and too large for ribonuclease. The computed Δ_j's (not listed) are all within the limits of 7–14 (*cf.* Table III).

METHOD 2: CD SPECTRUM WITH CONSTANT ROTATIONAL STRENGTHS. Instead of eliminating $\bar{n}$ as in method 1, we introduce into eq 15 a restriction on the rotational strength of the helical bands. For a Gaussian band, the rotational strength, R_j, is

$$R_j = 1.23 \times 10^{-42}[\theta_j^\circ]\Delta_j/\lambda_j \quad (20)$$

(Moscowitz, 1960). The small variations in the $[\theta_j^\circ]_H^\infty$ and Δ_j of each helical band among the eight proteins studied (Table III) suggest that the rotational strength of each band calculated from proteins for an infinite helix may be regarded as constant. From the CD data on myoglobin (Table III) we found that the R_j's for an infinite helix are -22.2×10^{-40}, -19.7×10^{-40}, and 53.9×10^{-40} erg cm^3 rad for n–π*, π–$\pi_\parallel$*, and π–$\pi_\perp$*, respectively. They agree well with those calculated from a synthetic helical polypeptide, poly(L-glutamic acid) (Cassim and Yang, 1970): -22×10^{-40}, -22×10^{-40}, and 57×10^{-40} erg cm^3 rad, respectively.

Replacing the $[\theta_j^\circ]_H^\infty$ by eq 20 (R_j in this case refers to an infinite helix), we can obtain an equation like eq 15, except that $(\lambda/\Delta_j$ rather than Δ_j becomes a new parameter

$$[\theta] = f_H\{-1800(\lambda/\Delta)_{n-\pi}*(1 - 2.50/\bar{n})\exp[-(\lambda/\Delta)^2_{n-\pi}*(\lambda/223.4 - 1)^2] - 1600(\lambda/\Delta)_{\pi-\pi_\parallel}*(1 - 3.50/\bar{n})\exp[-(\lambda/\Delta)^2_{\pi-\pi_\parallel}* \times (\lambda/206.6 - 1)^2] + 4370(\lambda/\Delta)_{\pi-\pi_\perp}*(1 - 2.50/\bar{n}) \times \exp[-(\lambda/\Delta)^2_{\pi-\pi_\perp}*(\lambda/193.5 - 1)^2]\} + f_\beta[\theta]_\beta + f_R[\theta]_R \quad (21)$$

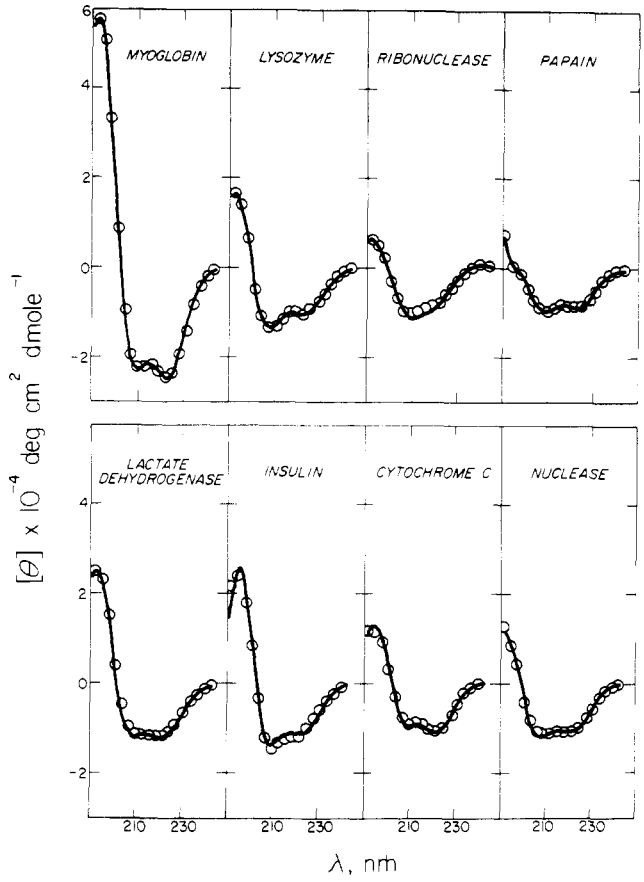

FIGURE 6: Comparison of the experimental and computed CD of eight proteins. Solid lines, experimental; points, computed. See text for details.

We set two boundaries conditions same as in method 1 and replace the third one concerning Δ_j's by $15 \leq (\lambda_j/\Delta_j)$'s ≤ 30. We solve the six unknowns (f_h, f_β, $\bar{n}$, and three (λ_j/Δ_j)'s of a protein from its CD spectrum. The results of eight proteins are included in Table V under method 2. The values of f_H and f_β based on the two methods agree fairly well, but the $\bar{n}$ values are uncertain in in several cases.

METHOD 3: COMBINATION OF METHODS 1 AND 2. By combining eq 18 and 21 and eliminating $\bar{n}$ we solve the five unknowns (f_H, f_β, and three (λ_j/Δ_j)'s) from the CD spectrum of protein. The results are essentially the same as those in Table V. Any one of the three methods appears to be superior to the method based on eq 10. The estimation of i and $\bar{n}$, which is not possible with other current methods, is still very rough. Nevertheless, our analysis clearly demonstrates the effect of chain length on the determination of helical content.

With the computed f_H, f_β, $\bar{n}$ (Table V, method 1), three Δ_j's (not listed), and $[\theta]_\beta$ and $[\theta]_R$ in Table II, we can reconstruct the CD spectra of the eight proteins studied according to eq 15 and compare them with their experimental results. The agreement is good in all cases, except perhaps for the spectra between 210 and 230 nm of cytochrome *c*, insulin and ribonuclease (Figure 6). With so many parameters in eq 15, it is always possible to find a best fit between the experimental and computed values, but this does not necessarily mean a correct solution for the computed curve. On the other hand, a poor fit does point up an imperfection in the method of analysis. Thus, the results of cytochrome *c*, insulin and ribonuclease in Figure 6 might indicate some complications that have escaped our attention and deserve further investigations.

Any method of analysis for CD and ORD of proteins must take into account both the theoretical basis and the existing experimental results of synthetic polypeptides and proteins. The three proposed methods are by necessity a compromise between the choice of as many parameters as desirable and the problem of computations with a few parameters that can still provide a reasonable solution. Thus, any such method is subject to future extensive tests and exceptions are almost certain to be found. At present these methods appear to be useful for proteins containing a moderate amount of helical content as reflected by the appearance of the characteristic double minimum and a single maximum in the CD spectrum (Figures 1 and 6). The eight proteins tested all have less than 60% unordered form. It remains to be investigated whether the methods are applicable to more unordered proteins. If a protein consists of only β and unordered form, obviously a simple equation such as eq 1 or 10 might suffice for the computation. Since the CD spectrum of the helix dominates that of the β and unordered form, any attempt to determine accurate X_β and X_R (Table II) would be difficult, if not impossible, unless the helical contribution to CD and ORD spectra can be accurately subtracted from the total optical activity. The computed f_β and f_R in Table V represent some statistical averages. The fairly good agreement with the X-ray results is indeed more than had been expected.

The present work has at least demonstrated the chain-length dependence of CD and ORD of the helices. Furthermore, $[\theta]_H^\infty$, $[m]_H^\infty$, and b_0^∞ are close to currently accepted values based on helical polypeptides of high molecular weight. The computed number of helical segments, i, and average number of residues, $\bar{n}$, are still problematic in spite of the good results shown in Table V. This is partly due to experimental difficulties. Equations 15 and 21 need data points below 200 nm, where the signal-to-noise ratio is usually low and the data are less precise than those above 200 nm. Thus, the reference values at low wavelengths are scattered (Figures 1 and 2), which in turn affect the determination of i and $\bar{n}$.

In our method of analysis we have neglected the rotatory contribution of non-peptide chromophores in the wavelength range of 190 and 240 nm. In most proteins they seem to constitute a small portion of the total optical activity, which is difficult to detect. This assumption may not be valid in some proteins such as carbonic anhydrases (work in progress) or conjugate proteins, which in turn could lead to erroneous computed results.

Although the results based on methods 1, 2, and 3 seem to be in good agreement, discrepancies are expected to appear when more proteins are studied. The choice of a correct answer would be difficult if all the boundary conditions are met and the mean-square error for the whole spectrum is comparable among the three methods. This is the uncertainty we still cannot solve at present.

For the sake of simplicity, the band positions of the three helical CD bands are predetermined from the data on myoglobin. Other proteins may have slightly different band positions. It is still a matter of conjecture whether they, especially the π-$\pi_\perp^*$ transition, should be adjusted accordingly. Since experimental measurements below 200 nm are often difficult, it might be necessary in some instances to prefix $\Delta_{\pi-\pi_\perp^*}$ at, say, 8.4 nm or $(\lambda/\Delta)_{\pi-\pi_\perp^*}$ at, say, 23 to solve eq 15 and 21 with data points between 205 and 240 nm. This again must await future investigations.

In summary, we have proposed a simple, although less rigorous, method of estimating the precentages of helical and β form (eq 1). For a more realistic approach, we take advantage of the chain-length dependence of the helical CD and ORD and propose three additional methods of analysis, which also

provide an estimate of the number of helical segments and average number of residues per segment. These methods are subject to future refinement and improvement. There are other different viewpoints on the CD ane ORD analyses such as the use of synthetic polypeptides as reference compounds. In some instances, they may also provide a reasonable estimate of the helical and β content. However, our belief is that the chain-length dependence of CD and ORD of the helical segments cannot be overlooked.

Added in Proof

At the Rehovot Symposium on Polyamino Acids, Polypeptides and Proteins, and Their Biological Implications, Israel, May 1974, R. W. Woody reported a theoretical treatment of the CD of β bends. The breadth of the allowed regions of the dihedral angles and the corresponding flexibility in β-turn conformation lead to many variants even under restricted criteria. The CD spectrum of most variants of the β turn resembles that of β-poly(L-lysine). In our method of analysis the residues of these β turns are counted as part of the unordered form. Whether a statistical term for these reverse turns should be separately listed in eq 1 or included in the term of the β form is a subject for future investigation. The counting of the β turn has some ambiguity too; each turn involves four residues but only three peptide bonds are used in theoretical calculations. Since the optical activity of helices dominates that of the β and unordered form our methods of analysis should remain valid for proteins containing a moderate amount of helical content.

References

Adams, M. J., Blundell, T. L., Dodson, E. J., Dodson, G. G., Vijayan, M., Baker, E. N., Harding, M. M., Hodgkin, D. C., Rimmer, B., and Sheat, S. (1969), *Nature (London) 224,* 491.

Arnone, A., Bier, C. J., Cotton, F. A., Day, V. W., Hazen, E. E., Jr., Richardson, D. C., Richardson, J. S., and Yonath, A. (1971), *J. Biol. Chem. 246,* 2302.

Arnone, A., Bier, c. J., Cotton, F. A., Hazen, E. E., Jr., Richardson, D. C., and Richardson, J. S. (1969), *Proc. Nat. Acad. Sci. U. S. 64,* 420.

Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Philips, D. C., and Sarma, V. R. (1965), *Nature (London) 206,* 757.

Blundell, T. L., Cutfield, J. F., Dodson, E. J., Dodson, G. G., Hodgkin, D. C., and Mercola, D. A. (1971), *Cold Spring Harbor Symp. Quant. Biol. 36,* 233.

Cassim, J. Y., and Yang, J. T. (1970), *Biopolymers 9,* 1475.

Chen, Y. H. (1973), Ph.D. Dissertation, University of California, San Francisco, Calif.

Chen, Y. H., and Yang, J. T. (1971), *Biochem. Biophys. Res. Commun. 44,* 1285.

Chen, Y. H., Yang, J. T., and Chau, K. H. (1973), *Fed. Proc., Fed. Amer. Soc. Exp. Biol. 32,* 495.

Chen, Y. H., Yang, J. T., and Martinez, H. (1972), *Biochemistry 11,* 4120.

Dickerson, R. E., Takano, T., Eisenberg, D., Kallai, O. B., Samson, L., Cooper, A., and Margoliash, E. (1971), *J. Biol. Chem. 246,* 1511.

Drenth, J., Jansonius, J. N., Koekoek, R., and Wolthers, B. G. (1971), *Enzymes 3,* 484.

Greenfield, N., Davidson, B., and Fasman, G. D. (1967), *Biochemistry 6,* 1630.

Greenfield, N., and Fasman, G. D. (1969), *Biochemistry 8,* 4108.

Holzwarth, G., and Doty, P. (1965), *J. Amer. Chem. Soc. 87,* 218.

Kartha, G., Bello, J., and Harker, D. (1967), *Nature (London) 213,* 862.

Kendrew, J. G. (1962), *Brookhaven Symp. Biol. 15,* 216.

Kendrew, J. G., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Philips, D. C., and Shore, V. C. (1960), *Nature (London) 185,* 422.

Madison, V., and Schellman, J. (1972), *Biopolymers 11,* 1041.

Magar, M. E. (1968), *Biochemistry 7,* 617.

Moffitt, W., and Yang, J. T. (1956), *Proc. Nat. Acad. Sci. U. S. 42,* 596.

Moscowitz, A. (1960), *in* Optical Rotatory Dispersion, Djerassi, C., Ed., New York, N. Y., McGraw-Hill, Chapter 12.

Myer, Y. P. (1970), *Res. Commun. Chem. Pathol. Pharmacol. 1,* 607.

Rosenkranz, H., and Scholtan, W. (1971), *Hoppe-Seyler's Z. Physiol. Chem. 352,* 896.

Rossman, M. G., Adams, M. J., Buehner, M., Ford, G. C., Hackert, M. L., Lentz, P. J., Jr., McPherson, A., Jr., Schevitz, R. W., and Smiley, I. E. (1972), *Cold Spring Harbor Symp. Quant. Biol. 36,* 179.

Sarkar, P. K., and Doty, P. (1966), *Proc. Nat. Acad. Sci. U. S. 55,* 981.

Saxena, I. P., and Wetlaufer, D. B. (1971), *Proc. Nat. Acad. Sci. U. S. 68,* 969.

Straus, J. H., Gordon, A. S., and Wallach, D. F. H. (1969), *Eur. J. Biochem. 11,* 201.

Vournakis, J. N., Yan, J. T., and Scheraga, H. A. (1968), *Biopolymers 6,* 1531.

Woody, R. W. (1968), *J. Chem. Phys. 49,* 4797.

Woody, R. W. (1969), *Biopolymers 8,* 669.

Woody, R. W., and Tinoco, I., Jr. (1967), *J. Chem. Phys. 46,* 4927.

Yang, J. T., and Chen, Y. H. (1973), *Proc. Int. Congr. Biochem., 9th,* 69.