Lipsett, M. (1964), *J. Biol. Chem. 239*, 1257.
Magee, W. S., Jr., Gibbs, J. H., and Zimm, B. H. (1963), *Biopolymers 1*, 133.
Naylor, R., and Gilham, P. T. (1966), *Biochemistry 5*, 2722.
Naylor, R., Ho, N. W. Y., and Gilham, P. T. (1965), *J. Amer. Chem. Soc. 87*, 4209.
Palacios, R., Palmiter, R. D., and Schimke, R. T. (1972), *J. Biol. Chem. 247*, 2316.
Palmiter, R. D., Palacios, R., and Schimke, R. T. (1972), *J. Biol. Chem. 247*, 3296.
Riley, M., Maling, B., and Chamberlin, M. J. (1966), *J. Mol. Biol. 20*, 359.
Steele, R. G. D., and Torrie, J. H. (1960), Principles and Procedures of Statistics, New York, N. Y., McGraw-Hill Book Company, p 353.
Stubbs, J. D., and Hall, B. D. (1968), *J. Mol. Biol. 37*, 289.
Szybalski, W. (1967), *in* Thermobiology, Rose, A. H., Ed., New York, N. Y., Academic Press, p 73.
Thomas, C. A., Jr. (1966), *Progr. Nucl. Acid Res. Mol. Biol. 5*, 315.
Uenoyama, K., and Ono, T. (1972), *J. Mol. Biol. 65*, 75.
Williams, A. R., and Askonas, B. A. (1967), *J. Mol. Biol. 23*, 201.

Determination of the Secondary Structures of Proteins by Circular Dichroism and Optical Rotatory Dispersion†

Yee-Hsiung Chen,‡ Jen Tsi Yang,* and Hugo M. Martinez

ABSTRACT: The rotatory contributions of a protein can be represented by $X = f_H f_H + f_\beta X_\beta + f_R X_R$, where X can be either the ellipticity or the rotation at any wavelength. The The f's are the fractions of the helix (H), β form and unordered form (R); the sum of f's is equal to unity and each f is greater than or equal to zero. With the f values of five proteins obtained by X-ray diffraction studies, the X's of these proteins at any wavelength are fitted by a least-squares method, which defines the X_H, X_β, and X_R. The circular dichroism and optical rotatory dispersion for the helix, β, and unordered form thus determined can be conversely used to estimate the secondary structure of any protein with X's at several wave lengths for the same equation. Results of tests on five additional proteins were satisfactory to good. The calculated spectra were also in satisfactory to good agreement with the experimental results. The β and unordered forms have the same ellipticity and rotation near 222 and 233 nm, respectively, and a very small b_0 of the Moffitt equation. Thus, as a first approximation, the helical content of a protein varies linearly with $[\theta]_{222}$, $[m]_{233}$, or b_0. The present approach is more realistic than previous methods using synthetic polypeptides as reference compounds.

Since the discovery of the optical activity of α helix (Doty and Yang, 1956; Yang and Doty, 1957) and the introduction of the Moffitt equation (Moffitt and Yang, 1956), optical rotatory dispersion (ORD) and, more recently, circular dichroism (CD) have been widely used for studying the conformation and conformational change of proteins and polypeptides in solution. At first, the use of ORD for estimating the helical content of a protein molecule was only suggestive, since X-ray diffraction studies had not yet revealed the three-dimensional structure of any protein (Crick and Kendrew, 1957). The current methods for estimating the helical content of a protein molecule such as the b_0 of the Moffitt equation, the reduced mean residue rotation at 233 nm, and the mean residue ellipticity at 222 nm are based on the use of synthetic polypeptides as model compounds and, furthermore, they completely neglect the optical activity of other structural elements such as the β form (see, for example, Yang, 1969). Recently, Greenfield *et al.* (1967) have attempted to fit the experimental ORD of proteins with a combination of the ORD of pure helix, β form, and random coils as deduced from synthetic poly(L-lysine). Magar (1968) has proposed a method of minimizing the variance between experimental and computed ORD curves for the determination of the three forms. Greenfield and Fasman (1969) have used the same principle for the minimization of the CD spectra. In the method of isodichroic points, Myer (1970) has estimated the fractions of the three forms in proteins from the CD measurements at three wavelengths corresponding to the three isodichroic points. In all cases poly(L-lysine) was used as a model compound for the three forms. All these treatments using synthetic polypeptides, however, lead to estimates that could differ considerably from the X-ray results. For instance, the three-dimensional structure of sperm-whale myoglobin showed 77% α helix and no β form (Kendrew *et al.*, 1960). The ORD of myoglobin, however, could best be fitted with a mixture of 54–55% α helix, 35–36% β form, and 10% coils (Greenfield *et al.*, 1967; Magar, 1968), and the isodichroic points method gave the same estimates as the ORD spectrum (Myer, 1970). The computer fit of the CD spectrum of myoglobin gave about 68% α helix and 5–8% β form, which were closer to the X-ray results (Greenfield and Fasman, 1969).

The use of synthetic polypeptides as model compounds for

† From the Cardiovascular Research Institute and Department of Biochemistry and Biophysics, University of California, San Francisco, California 94122. *Received May 30, 1972.* This work was aided by U. S. Public Health Service Grants GM-10880 and HL-06285. Presented in part at the 56th Annual Meeting of Federation of American Societies for Experimental Biology, Atlantic City, N. J., April 1972 (Yang and Chen, 1972).

‡ Part of a dissertation to be submitted by Y.-H. C. in partial fulfillment for the requirement of the degree of Doctor of Philosophy, University of California, San Francisco.

proteins remains unsatisfactory. X-Ray diffraction studies have shown that globular protein molecules contain only short helical and β segments and that the unordered portions of the molecules are rigid and compact. They are unlike synthetic polypeptides which can form long-chain helix and β form and extended, flexible coils. The rotatory properties of synthetic polypeptides, therefore, are not appropriate for describing those of proteins. Our contention is also partly based on the theoretical work of Woody and Tinoco (1967), who found that the rotational strengths of both the n–π^* and π–π^* transitions of α and 3_{10} helices are dependent on the chain length of the polypeptides, Vournakis *et al.* (1968) reached the same conclusion about the n–π^* transition of α helix. A further complication is that the magnitude of CD and ORD of even long-chain helical homopolypeptides can vary with the nature of the side groups (see, for example, Parrish and Blout, 1971). Recent work of Woody (1969) further shows that the optical activity of both parallel and anti-parallel β forms is dependent not only on the chain length but also on the number of strands.

We propose an alternative approach to the determination of helix, β, and unordered form of globular proteins by utilizing the fractions of the three forms of several proteins from X-ray diffraction studies and determining a new set of reference values for the helix, β, and unordered form, which in turn enable us to determine the secondary structures of a protein molecule. This approach has also been used by Saxena and Wetlaufer (1971). A preliminary account of our work (Chen and Yang, 1971) has shown significant differences in magnitude between the reference values based on proteins and those based on synthetic polypeptides.

Our interest in the ORD and CD of proteins is, of course, not confined to the determination of the helix, β, and unordered form *per se* in a protein molecule. This can be admirably solved by X-ray diffraction studies. Rather, we wish to investigate the conformational changes of the proteins in solution, in particular, the changes in secondary structure relative to biological functions under various experimental conditions. Toward this goal we must first have a tool for determining these structural elements by optical methods, *viz.*, ORD and CD.

Experimental Procedures

Materials. Crystallized sperm-whale myoglobin, papain from papaya (twice crystallized), bovine pancreas ribonuclease (five-times crystallized), horse heart cytochrome *c* (lyophilized powder), and bovine pancreas α-chymotrypsin (three-times crystallized) were purchased from Mann Research Laboratories. Egg-white lysozyme (three-times crystallized) and bovine insulin were obtained from Sigma Chemical Co., and *Staphylococcus aureus* nuclease (lyophilized powder) and bovine pancreas chymotrypsinogen A (five-times crystallized) from Worthington Biochemical Corp. Dogfish lactate dehydrogenase was a gift of Professor N. O. Kaplan and Dr. A. S. Levi. Myoglobin, lysozyme, ribonuclease, and papain were further purified by passing them in aqueous solution through a Sephadex G-75 column; the main fraction in each case was collected for optical studies. No detectable difference in optical activity has been found between the proteins treated in this manner and the untreated proteins obtained from commercial sources. All the proteins were dissolved in 0.01 M phosphate buffer (pH 6.8), except for insulin, which was dissolved in 0.1 M KCl at pH 2.1. The solutions were dialyzed against the buffer or salt solution overnight and clarified with a Millipore filter of 5 μ pore size before use. The protein concentrations were determined by micro-Kjeldahl nitrogen analyses, using the per cents of nitrogen based on available amino acid composition of the proteins. All chemicals were of reagent grade and water was double distilled.

TABLE I: Fractions of Helix, β, and Unordered Form of Five Proteins as Determined by X-Ray Diffraction Studies.[a]

Proteins	f_H	f_β	f_R	References
Myoglobin	0.77	0	0.23	Kendrew *et al.* (1960) and Kendrew (1962)
Lysozyme	0.29	0.16	0.55	Blake *et al.* (1965)
Lactate dehydrogenase	0.29	0.20	0.51	Adams *et al.* (1970)
Papain	0.21	0.05	0.74	Drenth *et al.* (1968)
Ribonuclease	0.19	0.38	0.43	Kartha *et al.* (1967)

[a] See text for details.

Methods. CD was measured with a Durrum-Jasco J-10 spectropolarimeter and ORD with a Cary 60 spectropolarimeter, both at 25° and under constant nitrogen flush. The CD instrument had been calibrated with an aqueous solution of *d*-10-camphorsulfonic acid (Eastman Kodak) and the ORD instrument with a freshly prepared aqueous solution of sucrose (National Bureau of Standards grade) (Cassim and Yang, 1969). The data were expressed in terms of mean residue ellipticity, $[\theta]$, for CD and mean residue rotation, $[m]$, for ORD, except that, following convention, reduced mean residue rotation, $[m']$, was used in the Moffitt equation (Moffitt and Yang, 1956). The mean residue weight of each protein was calculated from its amino acid composition. The experimental data were analyzed by the least-squares method with the use of a digital PDP-12 computer (Digital Equipment Corp.) of the Department of Biochemistry and Biophysics (for the early phase of this work) and the IBM 360/50 system of the Computer Center, both of the University of California, San Francisco.

Method of Analysis

The optical activity of a protein can be expressed as

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

with $f_H + f_\beta + f_R = 1$ and all f's ≥ 0. Here X can be the mean residue ellipticity, $[\theta]$, or the mean residue rotation, $[m]$, or the b_0 of the Moffitt equation. The X_H, X_β, and X_R, are the reference values that would be obtained if the protein molecule were made up of segments of pure helix (H), β form, and unordered form (R). The f's are the fractions of the helix, β, and unordered form in a protein molecule. We assume that all the helices are right-handed and exclude the collagen type of helix. Solving for the X parameters in eq 1 requires a minimum of three simultaneous equations with the use of the experimental values for three proteins whose f values can be deduced from X-ray diffraction studies. We choose, however, five proteins (see Table I) and fit the five experimental X values at any chosen wavelength (or b_0 values) in eq 1 by a linear least-squares method. The computations were repeated over a wide range of wavelengths. The CD and ORD

for the helix, β, and unordered forms thus determined can then be used conversely to determine the f_H, f_β, and f_R of any protein using X's at 20 or more wavelengths with the same equation. The computer program used was BMDX 85 T, UCLA (gauss–newton stepwise interation), with the conditions of (1) Σf's = 1, and (2) $1 > f_H, f_\beta > 0$. Since f_R was replaced by $1 - f_H - f_\beta$ in the computation, the computed sum of f_H and f_β could in rare cases be greater than unity, thus leading to a negative f_R. In practice, however, we have never observed such anomalous results. In the method of minimizing the variance between experimental and computed curves, Magar (1968) used a matrix inversion subroutine, whereas Greenfield and Fasman (1969) used a SIMQ subroutine (System/360, Scientific Subroutine Package (360A-CM-03X), IBM/Technical Newsletter). In these two programs only condition 1 was specified.

We are aware of several factors which could complicate our computations. First, the counting of amino acid residues in various secondary structures is not clear-cut. Usually we used the numbers given by the authors of the X-ray studies. We consider residues helical only when they form at least one turn of a helical segment. We make no distinction between Pauling–Corey's α helix and other helices such as the 3_{10} helix or distorted α helix, although the rotatory contributions of different helices may differ from one another to some extent. The uncertainty in determining the f's was also pointed out by Saxena and Wetlaufer (1971), who chose for the f_H in eq 1 a mean value between the lower limit representing true, regular α helix and the upper limit representing total helix including 3_{10} and distorted helices. We will show later that X_H is also chain length dependent and the X_H values given in this report represent the contributions of the helical segments that average three or four turns per segment based on the five proteins used in the computations. Second, the β form can be parallel or anti-parallel, or a mixture of both. For instance, one strand can be juxtapositioned to a second parallel strand and a third anti-parallel one. On the basis of Woody's theory (1969), we again make no distinction between the two types of the β form. Their rotatory contributions are also dependent on the number of strands, as well as on the number of residues per strand. A further complication is that the β structure is not fixed like the α helix. It can be so distorted that its backbone is not in one plane. Thus, its optical activity could differ significantly from that of a regular β sheet. We can only assume that there is a statistical average value for X_β in eq 1. Third, we also assume that there is a statistical average value for X_R in eq 1, although the actual X_R values could vary from one protein to another. Fourth, the rotatory contributions due to nonpeptide chromophores are assumed to be small as compared to those of peptide chromophores; therefore no attempt is made to include additional terms in eq 1 for these contributions. Obviously, for a protein having large Cotton effects of the aromatic groups, the disulfide bonds or prosthetic groups, this assumption may not be valid.

An explanation for the numerical values in Table I is in order. The myoglobin molecule has nine segments of α helix with 16, 16, 7, 7, 10, 10, 9, 19, and 24 residues (Kendrew *et al.*, 1960; Kendrew, 1962). In all, 118 of the 153 residues, or 77%, are helical. Dickerson and Geis (1969) pointed out that from the patterns of hydrogen bonding three additional residues, FG 1, GH 6, and HC 1, should be considered part of the helices to which they are connected. Their inclusion would bring the helical content up to 79%. Had this high value been used in our computations, the X_H values would have been slightly reduced. The lysozyme molecule, with 129 residues, has three fairly regular α helices at residues 5–15, 24–34, and 88–96 and another short helix of residues 80–85 (Blake *et al.*, 1965) which amount to 29% helix. In addition, residues 60–63, 97–101, and 109–115 tend to adopt an helical conformation of the same general dimensions and Ramachandran coordinates as the α or 3_{10} helices, which, however, were not counted in our computation. In the lysozyme molecule residues 41–48 and 49–54 form an anti-parallel pleated sheet and residues 55–60 form a much less regular third strand to the sheet; thus, the molecule contains about 16% β form. Lactate dehydrogenase, with 311 residues, has eight α-helical segments at residues 1–6, 32–40, 55–61, 95–106, 109–118, 153–156, 229–242, and 289–300 and a further 16 residues as a 3_{10} helix (Adams *et al.*, 1970); thus, the total helical content is 29%. Lactate dehydrogenase also has five well-defined strands of parallel β form at residues 22–26, 45–48, 70–72, 79–86, and 121–123 and three less well-defined strands of anti-parallel β form at residues 245–261, 262–276, and 277–283; thus, it contains about 20% β form. Papain, with 212 residues, has four stretches of α helices at residues 26–41, 50–56, 69–78, and 116–126 and a lone portion of the β form at residues 163–172 (Drenth *et al.*, 1968); thus, it has 21% helix and 5% β form. The ribonuclease molecule, with 124 residues, has three helical segments at residues 5–12, 28–35, and 51–58 and three strands of β sheet at residues 42–49, 71–92, and 94–110 (Kartha *et al.*, 1967); thus, the helical content is 19% and β form 38%.

After this work was completed, we learned of the latest interpretations of results of X-ray diffraction studies on lactate dehydrogenase and papain. As a result of the extention of the X-ray study of lactate dehydrogenase from 2.8 to 2.5 Å (Rossman *et al.*, 1972), the number of residues in α helical and β arrangement has been considerably increased. Even the primary structure has been expanded from 311 to 331 residues. The helical segments are located at residues 2–6, 33–44, 55–70, 84–89, 107–130, 141–154, 165–181, 227–245, 249–263, 308–329 and the β form at residues 23–28, 48–53, 77–81, 92–97, 134–139, 159–161, 188–192, 200–207, 267–279, 281–295, and 298–303. In all, 150 of the 331 residues, or 45%, are in helical structure and 79 residues, or 24%, in β structure. (Rossman *et al.* did not include segments 84–89 and 141–154 in their Table II, which would have lowered the percentage in helical arrangement to 39%. However, they did show these two helical segments in their Figure 1.) Drenth *et al.* (1971) reported five stretches of α helices instead of four in the papain molecule at residues 24–41, 49–57, 67–78, 117–128, and 137–143. In addition, they noted that the papain molecule contains a distorted pleated-sheet structure of about 30 residues. Thus, the fractions of helix and β form are increased from 21 to 28% helix and from 5 to 14% β form. Clearly, these new results would require some modification in our computed reference spectra. However, these computed values are heavily weighted by the contribution of myoglobin, which has a higher helical content than the other four reference proteins used in this work (see Results). Furthermore, the X-ray results might still undergo minor revisions. Therefore, recomputation of our reference spectra seems unnecessary at this time, particularly since our main conclusions will remain valid despite some quantitative adjustments.

Results

CD and ORD of Reference Proteins. Tables II and III list the CD and ORD of the five reference proteins (Table I),

TABLE II: Circular Dichroism of Five Proteins ([θ] in (deg cm²)/dmole).[a]

Wavelength (mμ)	Myoglobin	Lysozyme	Lactate Dehydrogenase	Papain	Ribonuclease[b]
190	52,900	17,100	22,700	8,660	5,840
191	54,400	17,500	23,900	7,240	6,210
193	54,500	17,000	25,100	3,150	5,840
195	51,100	14,700	23,000	−160	4,890
197	40,100	10,500	18,400	−630	2,630
199	25,100	4,060	11,200	−980	0
201	8,660	−4,690	3,860	−5,840	−3,580
203	−3,800	−10,300	−1,100	−6,640	−6,560
205	−14,000	−12,400	−6,730	−7,760	−8,600
208	−21,800	−13,500	−10,900	−9,620	−10,700
210	−22,800	−13,000	−11,500	−9,530	−11,300
213	−21,200	−11,100	−11,000	−8,780	−10,400
216	−22,300	−10,400	−11,400	−8,450	−10,100
218	−23,500	−10,400	−11,700	−8,510	−9,650
220	−24,400	−10,400	−12,200	−8,780	−9,060
222	−24,900	−9,720	−12,300	−9,020	−8,090
224	−24,300	−9,230	−11,800	−8,910	−6,860
227	−21,400	−8,530	−9,930	−7,940	−4,620
229	−17,900	−7,800	−8,260	−6,150	−3,310
231	−14,100	−6,710	−6,580	−4,660	−2,240
234	−8,800	−4,150	−4,460	−2,590	−610
237	−4,780	−2,200	−2,700	−1,550	−60
240	−2,290	−1,220	−1,370	−990	630
243	−980	−720	−690	−580	390
246	−360	−440	−290	−310	60
250	−70	−300	−60	−140	−290

[a] The protein concentrations were determined by nitrogen analyses using the following percentages of nitrogen: myoglobin, 17.4; lysozyme, 18.8; lactate dehydrogenase, 15.9; papain, 16.1; ribonuclease, 17.5. [b] Saxena and Wetlaufer (1971) used the data of Greenfield and Fasman (1969) and Timasheff and Stevens (1969). An inadvertent error has now been found (S. Timasheff, private communication); what Saxena and Wetlaufer quoted should be reduced mean residue ellipticity, [θ'], not mean residue ellipticity, [θ].

TABLE III: Optical Rotatory Dispersion of Five Proteins ([m] in (deg cm²)/dmole).

Wavelength (mμ)	Myoglobin	Lysozyme	Lactate dehydrogenase	Papain	Ribonuclease
194	24,000	12,900	12,800		7,610
196	42,000	17,900	22,200		9,690
198	55,400	21,500	27,200	15,000	12,600
200	57,900	22,500	28,500	11,800	13,800
202	55,000	20,400	27,600	10,600	12,300
205	46,500	13,400	23,000	9,990	8,400
208	32,300	7,660	15,800	6,450	5,070
211	21,500	4,920	11,400	3,390	2,100
214	16,300	2,850	8,870	2,220	980
217	13,400	1,730	6,300	1,620	−1,040
219	9,940	780	4,400	920	−2,290
221	6,560	180	2,400	0	−3,550
223	2,530	−1,140	50	−1,150	−4,830
225	−1,470	−1,850	−1,760	−2,690	−5,550
227	−5,390	−2,620	−3,830	−4,050	−5,930
229	−9,140	−3,810	−5,370	−5,070	−6,000
231	−11,400	−4,950	−5,940	−5,570	−5,710
233	−12,100	−5,840	−6,060	−5,440	−5,150
235	−11,800	−5,860	−5,990	−4,910	−4,560
238	−10,400	−5,110	−5,510	−4,180	−3,580
241	−8,290	−4,230	−4,840	−3,620	−2,810
244	−6,620	−3,490	−4,040	−3,130	−2,330
247	−5,240	−2,850	−3,200	−2,690	−2,000
250	−4,130	−2,380	−2,650	−2,320	−1,730
253	−3,390	−2,050	−2,230	−2,000	−1,540
256	−2,850	−1,790	−1,900	−1,780	−1,430
260	−2,300	−1,570	−1,600	−1,540	−1,200

which we used for determining the X_H, X_β, and X_R in eq 1. The values are the average of the results of at least duplicate experiments for each protein. The errors for $[\theta]_{222}$ and $[m]_{233}$ were within 2%, but the data below 205 nm were less precise because of low signal to noise ratios at shorter wavelengths. Since the CD and ORD of proteins as reported from various laboratories were sometimes at considerable variance, we believe it desirable to present the data listed in Tables II and III for quantitative comparison and possible future revision.

Ths ORD data between 340 and 550 nm were fitted with the Moffitt equation (Moffitt and Yang, 1956), using a least-squares method. The b_0 values were: lysozyme, −163; lactate dehydrogenase, −181; papain, −123; and ribonuclease, −91 (deg cm²)/dmole. The b_0 of myoglobin was not determined from its ORD in the visible region because of the presence of the Cotton effects due to the Soret bands. Attempts have been made to fit the ORD data between 240 and 310 nm with the Moffitt equation using a λ_0 different from the customary 212 nm (Urnes, 1963), but the b_0 so determined cannot be compared to the b_0 obtained in the conventional way. Thus, we adopted a different approach to this problem. Previously, Cassim and Yang (1970) have shown that the ORD in the visible region, and therefore the b_0, of helical poly(L-glutamic acid) can be calculated from the CD in the ultraviolet region by means of the Kronig–Kramers transform. Using the same procedure, we obtained a b_0 of −447 (deg cm²)/dmole for myoglobin (Chen and Yang, 1971). In this calculation we have assumed that the CD contributions due to non-peptide chromophores, especially the heme in this case, is negligible in the ultraviolet region (see, for example, Strauss *et al.*, 1969).

Determination of X_H, X_β, and X_R. We first attempted to solve eq 1 with the data on three proteins (see also Saxena and Wetlaufer (1971) who solved three simultaneous equations using the CD of myoglobin, lysozyme, and ribonuclease). Table IV lists the solutions obtained with several sets of three proteins. Clearly, the X_H, X_β, and X_R values so determined vary from one set of proteins to another. This uncertainty makes any quantitative analyses difficult. Such variations, however, are minimized when myoglobin is included in any set of three proteins. The X parameters thus determined seemed to be heavily weighted by the inclusion of myoglobin, which has the highest helical content of the five proteins studied and therefore the strongest helical CD and ORD in the ultraviolet region. The desir-

TABLE IV: Computations of X_H, X_β, and X_R Based on the Sets of Three Proteins.

Proteins[a]	$[m]_{233}$			$[\theta]_{222}$			b_0		
	H	β	R	H	β	R	H	β	R
M, L, P	−14,700	−1,410	−3,270	−31,100	−9,390	−4,000	−579	48	−5
M, L, R	−15,300	−4,500	−1,230	−32,800	−6,650	1,550	−578	60	−9
M, D, R	−15,200	−4,110	−1,620	−29,800	3,100	−8,370	−548	158	−109
L, P, R	−9,990	−3,730	−4,260	−18,300	−4,570	−6,680	−588	58	−3
D, P, R	−12,100	−3,320	−3,690	−41,900	−10	−310	−786	97	50

[a] Abbreviations used are: M, myoglobin, L, lysozyme; D, lactate dehydrogenase; P, papain; and R, ribonuclease.

TABLE V: CD and ORD of Helix, β, and Unordered Form Based on Five Proteins.[a]

λ (nm)	$[\theta]$			λ (nm)	$[m]$		
	H	β	R		H	β	R
190	71,500	−6,940	−5,050	194	27,900	−6,170	11,000
193	76,600	3,480	−12,700	197	65,900	−5,200	5,130
196	67,400	1,260	−14,700	200	77,500	9,400	−3,510
199	36,400	−1,260	−9,480	203	70,800	5,930	−3,730
201	15,200	−10	−11,500	206	56,900	−50	−3,360
204	−10,900	−6,100	−7,090	209	37,500	−950	−2,670
207	−24,300	−8,220	−5,590	212	26,500	−1,930	−2,660
210	−28,300	−10,400	−4,210	215	22,000	−4,680	−2,430
213	−26,400	−9,680	−3,500	218	15,300	−7,770	−1,260
216	−28,100	−9,320	−2,480	221	9,240	−10,200	−1,020
219	−30,500	−6,850	−2,350	224	1,410	−10,600	−1,170
222	−31,500	−2,670	−2,780	227	−6,000	−8,290	−1,940
225	−30,000	2,000	−3,380	230	−12,500	−4,940	−2,470
228	−25,000	4,130	−3,120	233	−14,800	−2,630	−2,690
231	−17,900	3,920	−2,470	236	−14,100	−1,490	−2,340
234	−11,300	3,740	−1,510	239	−12,100	−830	−2,170
237	−6,100	2,830	−1,160	242	−9,150	−390	−2,210
240	−2,780	3,450	−1,260	245	−7,460	−200	−2,040
243	−1,110	2,000	−900	248	−5,800	−310	−1,860

[a] The computed values are listed at 3-nm intervals here, but for the determination of f_H, f_β, and f_R one data point per nanometer was used over a specified range of wavelengths. See the footnotes of Table VII.

ability of using additional calibrating proteins and estimates of uncertainty has also been discussed by Saxena and Wetlaufer (1971), although they evaluated the X parameters from only three proteins and did not use statistical tests for best fit. To minimize the variations of the computed X parameters, we fitted eq 1 with the data on the five proteins (see Method of Analysis). The results, of course are subject to future refinement when the results of X-ray diffraction studies of additional proteins become available. A few proteins that have been studied by X-ray crystallography were not chosen in our computations, mainly because the fractions of helix or β form, or both, were still not completely determined at the time this work was begun. (Carboxypeptidase as excluded because it appeared to be unstable in solution when exposed to ultraviolet light; the solution became slightly turbid during the CD and ORD measurements. No such complication arose for measurements in the visible region.)

Table V lists the CD and ORD of helix, β, and unordered form based on the five reference proteins. Figures 1–4 show the CD and ORD spectra of the three conformations. Points were plotted for portions of the curves when a smooth curve could not be drawn. The standard deviations were relatively small for the helix, but large for the β and unordered form, especially on the short-wavelength side (Table VI). The X_H, X_β, and X_R values are based on the f's shown in Table I and would vary if different f's were used. As expected, these parameters were dominated by the f_H chosen for myoglobin, which has the highest helical content among the five proteins. If the f_H of any of the other four proteins was slightly altered, the computed X_H values remained essentially unchanged. The X_β and X_R did change when the reference f's were slightly changed, but they did not seriously alter the computed f_H, f_β, and f_R, to be described later.

The computed CD and ORD of the helical form (Figures 1 and 2) closely resemble those of synthetic polypeptides, with a double CD minimum at 222 and 208–209 nm, an ORD mini-

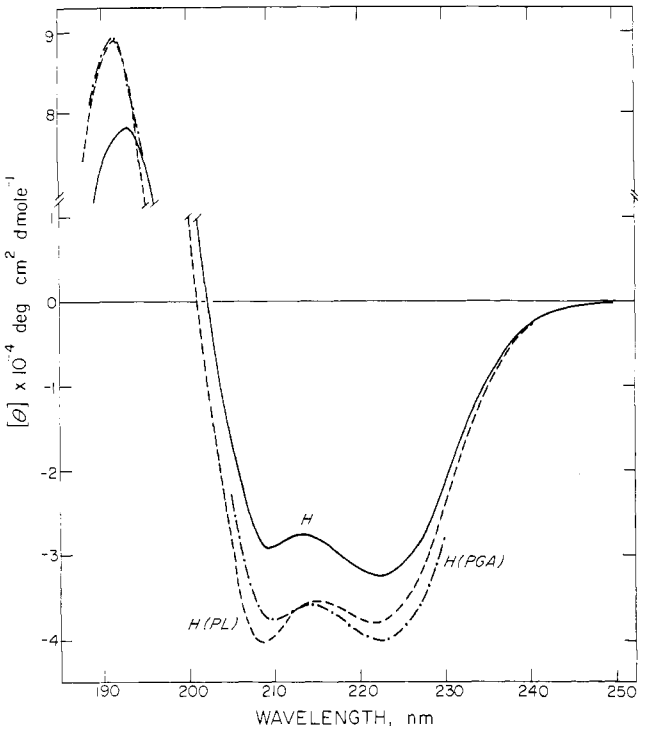

FIGURE 1: Circular dichroism of the computed helix (H) based on five proteins. Incuded for comparison: H(PGA), helical poly(L-glutamic acid); H(PL), helical poly(L-lysine).

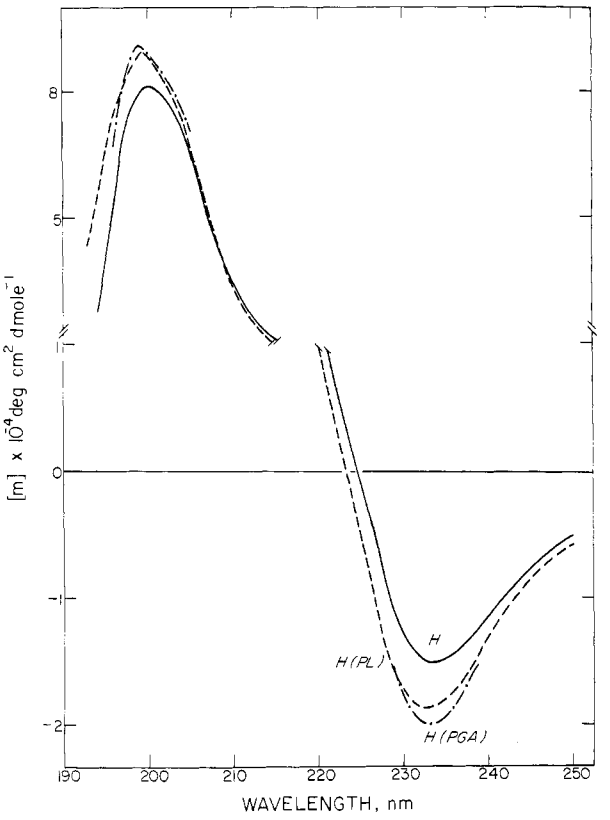

FIGURE 2: Optical rotatory dispersion of the computed helix based on five proteins. Symbols as in Figure 1.

mum at 233 nm, a CD maximum at 193 nm, and an ORD maximum at 200 nm. The last two figures are 2 nm higher than those observed for synthetic polypeptides, but are essentially identical with those found for proteins; for instance, myoglobin shows a CD maximum at 192 nm and an ORD maximum at 200 nm. This red shift probably results from the π–π^* transition of the peptide chromophore in an apolar protein interior or from the presence of short helical segments as well as helices other than the α helix (Straus *et al.*, 1969). The computed curves are about 20% smaller in magnitude than the curves, for synthetic polypeptides, which can be attributed to the end effects of the short helical segments found in protein molecules (see Discussion).

The CD and ORD of the unordered form of proteins and of synthetic polypeptides differ significantly (Figures 3 and 4). The computed curve has a CD minimum at 194 nm whose magnitude is only about one-third of that of synthetic polypeptides at 197-nm minimum. The corresponding ORD minimum at 202 nm is also much smaller than that of synthetic polypeptides at 205 nm. This finding is not unexpected since the so-called "coiled" form of poly(L-glutamic acid) and poly(L-lysine) are highly extended because of electrostatic repulsion and therefore differ from the compact, rigid unordered form in a protein molecule. The CD of synthetic polypeptides also shows a small positive band at 218–219 nm and a much smaller negative band near 240 nm, both of which are absent from the computed curve. Instead, we observed a small negative band located near 225 nm for the computed unordered form.

The CD of the computed β form has a minimum at 214 nm and a maximum at 194 nm, which are close to the 217- to 218- and 196-nm extrema for the β form of poly(L-lysine) (Sarkar and Doty, 1966; Townend *et al.*, 1966) and silk fibroin (Iizuka and Yang, 1966). The corresponding ORD has a minimum at 223 nm and a maximum at 200 nm, as compared to 230–231 and 205 nm for β-poly(L-lysine) (Sarker and Doty, 1966; Davidson *et al.*, 1966) and β-silk fibroin (Iizuka and Yang, 1966). But the magnitude of the computed extrema is only one-third to one-fourth of that for β-poly(L-lysine). Furthermore, the computed CD curve appears to have two small positive maxima between 225 and 250 nm, which are absent for β-poly(L-lysine). Our computed values are the statistical averages based on the five reference proteins and therefore are subject to future revision.

Determination of f_H, f_β, and f_R. With the CD and ORD spectra of the helix, β, and unordered form (Table V), we can conversely determine the f_H, f_β, and f_R of a protein molecule by fitting eq 1 with the experimental CD or ORD data for

TABLE VI: Standard Deviations of the Computed CD and ORD Extrema of Helix, β, and Unordered Form.

Conformation	λ (nm)	[θ]	SD	λ (nm)	[m]	SD
Helix	193	76,600	±11,200	200	77,500	±7,510
	209	−28,400	±1,530	233	14,800	±500
	222	−31,500	±1,950			
β form	194	8,240	±25,800	200	9,400	±16,100
	214	−10,300	±1,280	223	−11,300	±4,550
	230	4,240	±3,390			
Unordered	200	−7,870	±6,970	202	−4,080	±8,420
	223	−2,860	+2,170	221	−1,020	±2,850
				232	−2,690	±1,000

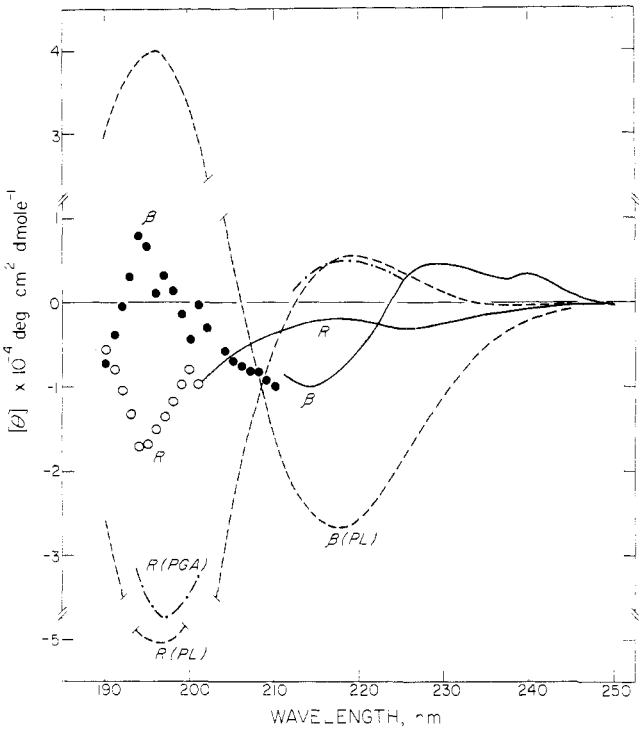

FIGURE 3: Circular dichroism of the computed β and unordered (R) form based on five proteins. Included for comparison: β(PL), β-poly(L-lysine); R(PL), unordered poly(L-lysine); R (PGA), unordered poly(L-glutamic acid). See text for details.

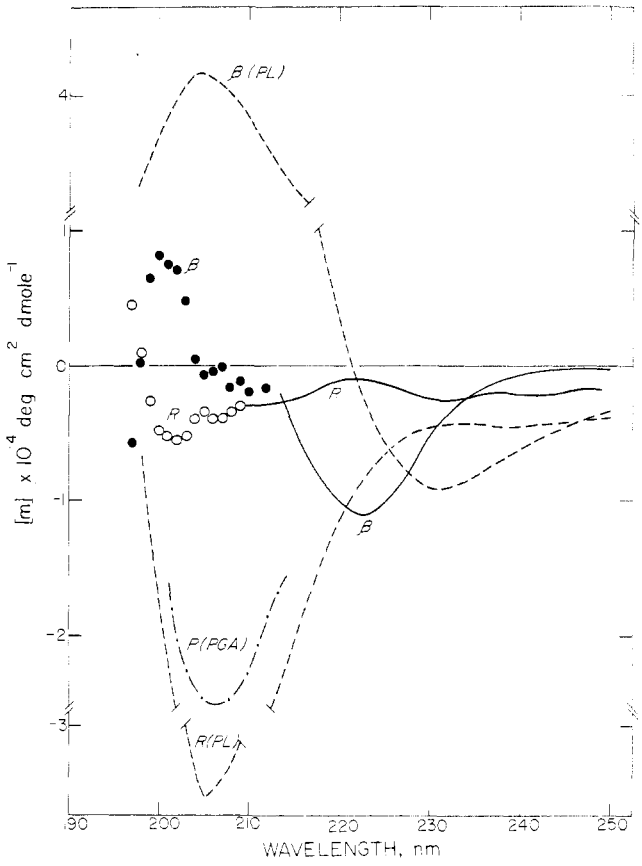

FIGURE 4: Optical rotatory dispersion of the computed β and unordered form based on five proteins. Symbols as in Figure 3.

TABLE VII: Determination of f_H and f_β of Several Proteins from CD and ORD Spectra.

Proteins	Form	X-Ray[a]	CD[b]	ORD[c] 1	ORD[c] 2
Myoglobin	H	0.77	0.77	0.77	0.77
	β	0	0.02	0.03	0.05
Lysozyme	H	0.29	0.29	0.22	0.25
	β	0.16	0.16	0.11	0
Lactate dehydrogenase	H	0.29	0.31	0.41	0.30
	β	0.20	0.06	0.06	0.08
Papain	H	0.21	0.21	0.20	0.22
	β	0.05	0.10	0.14	0.19
Ribonuclease	H	0.19	0.18	0.16	0.20
	β	0.38	0.44	0.46	0.51
Insulin	H	0.22–0.45	0.31	0.32	0.25
	β	0.12	0.18	0.06	0.07
Nuclease	H	0.24	0.27	0.27	0.27
	β	0.15	0.10	0.12	0.08
Cytochrome *c*	H	0.11–0.39	0.27	0.36	0.26
	β	0	0.06	0.04	0.08
α-Chymotrypsin	H	0.09	0.08	0.04	0.11
	β		0.10	0.12	0
Chymotrypsinogen A	H	0.06	0.09	0	0.08
	β		0.36	0.41	0.26

[a] See text for details. [b] Based on reference values between 205 and 240 nm in Table V at 1-nm intervals: The standard deviations for all f's were less than 0.04. [c] Based on reference values in Table V at 1-nm intervals: (1) between 210 and 240 nm, (2) between 225 and 240 nm. The standard deviations for all f's were less than 0.04.

the protein. We used only the CD between 205 and 240 nm and the ORD data for the protein. We used only the CD between 205 and 240 nm and the ORD between 210 and 240 nm for our computations (one data point per nanometer), mainly because the data at shorter wavelengths were less precise and would therefore introduce more errors.

To test the reliability of the method of analysis, we first determined the fractions of helix, β, and unordered form of the five reference proteins. The computed f_H's reproduced extremely well with what had been used originally (Table VII). Since the X_β and X_R values are less certain than the X_H's because of the dominance of the latter, the agreement between the computed β and unordered forms and those deduced from X-ray studies should be regarded as fairly good. One exception was lactate dehydrogenase whose computed f_β of 0.06 was much lower than the f_β of 0.20 from the X-ray study. The computed f_β of 0.02 for myoglobin was probably within the errors of such analysis. In general, the CD and ORD data give about the same answers. In a few cases where a discrepancy exists, as in the case of lysozyme and lactate dehydrogenase, the CD results seem to provide answers closer to the actual values than the ORD results. This is probably due to the fact that both positive and negative optical rotations of that helical form including those near the crossover (zero rotation) were used in the computations, whereas only negative CD data of the helical form between 205 and 240 nm were used.

With the computed f's (Table VII), we can reconstruct the CD and ORD spectra of the five reference proteins, using the reference values listed in Table V. Figures 5 and 6 clearly show that the agreement between the experimental and cal-

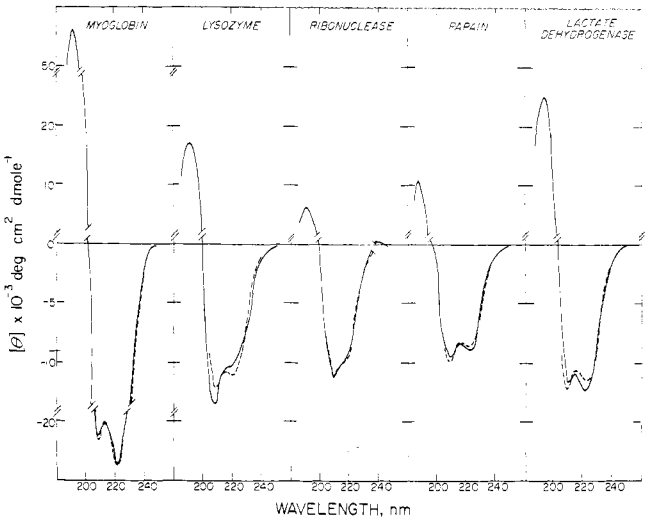

FIGURE 5: Comparison of the experimental and computed CD of reference proteins. Solid line, experimental; broken line, computed (based on data in Figures 1 and 3).

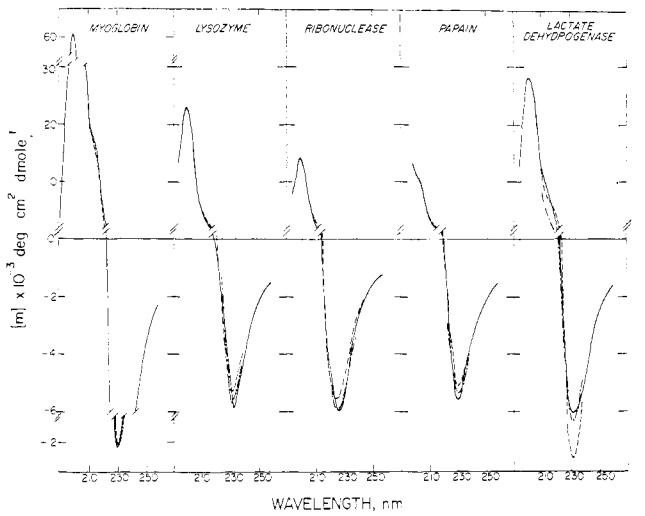

FIGURE 6: Comparison of the experimental and computed ORD of reference proteins. (——) Experimental; (– – –) computed (based on data in Figures 2 and 4 between 210 and 240 nm); (–·–) computed (based on data in Figures 2 and 4 between 225 and 240 nm). See text for details.

culated curves is good in all cases, except for the ORD of lactate dehydrogenase near the 233-nm minimum. In an attempt to improve the curve fitting, we redetermined the f's from the ORD over a narrower range of wavelength, 225–240 nm instead of 210–240 nm (Table V), thus excluding the dextrorotations and crossover (zero rotation). We then recalculated the ORD with these new f's which did improve the curve fitting over the narrower range of wavelength. Also the new f_H values for lysozyme and lactate dehydrogenase agreed well with those based on CD. But the f_β for lysozyme dropped to zero, which was obviously unsatisfactory.

Next, we tested our method of analysis on five additional proteins: insulin, nuclease, cytochrome *c*, α-chymotrypsin, and chymotrypsinogen A. The fractions of helix and β form of these proteins based on X-ray studies are counted as follows. The insulin molecule has two short stretches of near helical conformation between residues A2-6 and A13-19 and another three turns of α helix between B9–19 (Adams *et al.*, 1969; Blundell *et al.*, 1972). In all, 23 of the 51 residues, or 45%, are helical, but the true α helix in the β chain is only 22%. In the insulin dimer, residues B23–28 of the two subunits are arranged anti-parallel and appear to form a hydrogen-bonded pleated sheet over at least part of their lengths; thus, 12% is counted as the β form. Nuclease, with 149 residues, has three helical segments at residues 54–67, 99–106, and 122–134, that is, 24% helix, and three strands of anti-parallel β-pleased sheet at residues 12–19, 21–27, and 30–36, that is, 15% β form (Arnone *et al.*, 1969, 1971), Cytochrome *c*, with 104 residues, has one true segment of α helix at residues 92–102, that is, 11% (Dickerson *et al.*, 1971). But there are also five short segments which are gently helical and which probably have Ramachandran angles close to those of an α helix at residues 9–13, 14–18, 49–54, 62–70, and 71–75. If they are included in the counting, the percentage of helix would be brought up to 39%. Cytochrome *c* has no β sheet. α-Chymotryspin has a straight rod of α helix at residues 235–245 and an α-helix-like segment at residues 164–173, which is distorted by the existence of a 3_{10}-like hydrogen bond in the middle of it (Matthews *et al.*, 1967; Sigler *et al.*, 1968; Blow *et al.*, Birkloft *et al.*, 1969; Blow, 1971). In all, 21 of the 245 residues are helical, that is, 9%. The structure of the anti-parallel pleated-sheet type is rather common in the α-chymotrypsin molecule, although there are no extensive regions of regular pleated sheet. Thus, the counting of the β form is made difficult in this case. Chymotrypsinogen and all forms of the chymotrypsins have similar if not identical conformations (Kraut *et al.*, 1967; Freer *et al.*, 1970; Kraut, 1971). In the zymogen molecule the helical segments between residues 164–173 is further distorted so that it probably cannot form more than one turn of helix at residues 164–168. This is the only part of the zymogen molecule where an alteration of helical content may occur upon activation (Kraut, 1971). Thus, we counted 15 helical residues or 6% in chymotrypsinogen A. The amount of the β form is still not defined at this time, but there appear to be large sections of distorted anti-parallel pleated sheet. As Kraut (1971) pointed out, the model of chymotrypsinogen is still in a relatively unrefined state compared to the α-chymotrypsin model.

Our computed results on the five additional proteins, also listed in Table VII, are satisfactory to good as compared to the X-ray results. We found 27% helical content for nuclease, which is close to the X-ray finding of 24% (Arnone *et al.*, 1971). The reasonable results on the helical content of α-chymotrypsin and chymotrypsinogen A seem to suggest that our method of analysis is also applicable to proteins whose helical content is less than 10%. This enzyme and its zymogen seemed to show a marked difference in the estimated f_β. Kraut (1971) pointed out that at least 21 amino acid residues are shifted by 3.6 Å when chymotrypsinogen is activated to α-chymotrypsin. This movement must be accompanied by some conformational changes, but whether such changes could account for our different f's remains uncertain. We must wait for the refined results of X-ray studies on chymotrypsinogen. We also do not know whether this difference represents the failure of our analysis for the β form. Our computed f_H for insulin is within the range of X-ray results. It is expected to be larger than the lower limit of 22% true helix, as the two short segments of near helical conformation in the A chain must have also contributed to the optical activity of helices. Alternatively, the distorted helices could have different rotatory contributions from the true ones, a factor that is not considered in this work. In addition, the chain lengths of

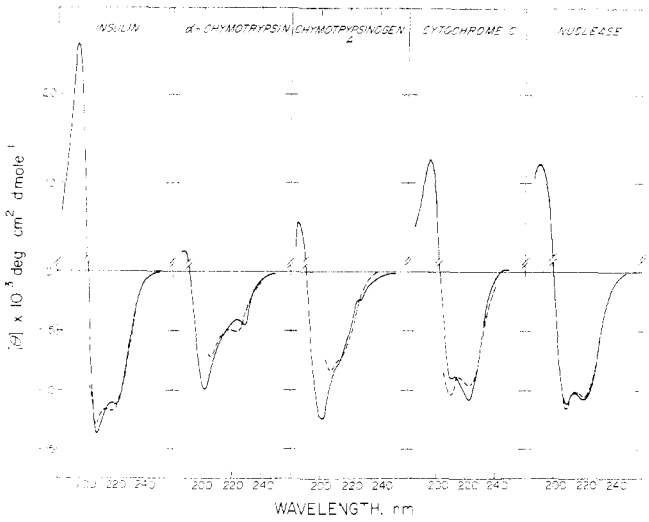

FIGURE 7: Comparison of the experimental and computed CD of five additional proteins. Solid line, experimental; broken line, computed (based on data in Figures 1 and 3).

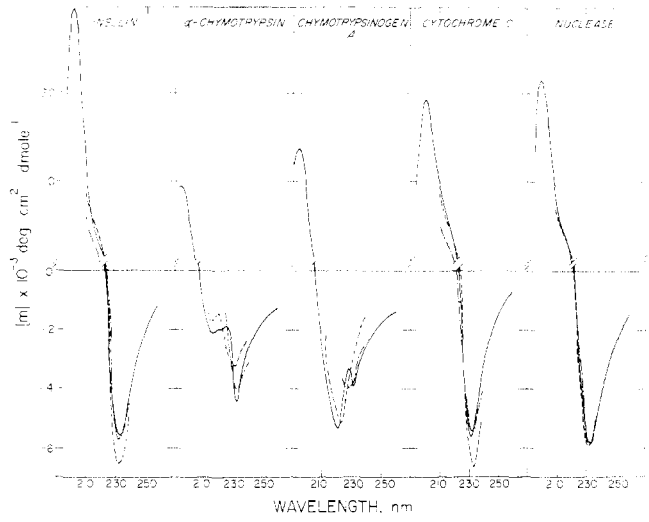

FIGURE 8: Comparison of the experimental and computed ORD of five additional proteins. Symbols as in Figure 6.

two of the three helical segments in the insulin molecule are shorter than the average length of the helical segments in the five reference proteins, which would result in an underestimate of the helical content by our method of analysis (see Discussion). Cytochrome *c* had cast doubts on the usefulness of the ORD analysis. A few years ago ORD studies of cytochrome *c* indicated a 17–27% helical content (Urry and Doty, 1965; Flatmark and Robinson, 1968; Zand and Vinogradov, 1968), whereas X-ray studies at 4-Å resolution indicated little or no α helix (see, for example, Beychok, 1968). Our present computed f_H of cytochrome *c* is, however, within the limits of the X-ray results. Here again our underestimated f_H as compared to the upper limit could be attributed to the five imperfect, short helical segments in addition to one true helix and also to the chain-length dependence of optical activity of these short helical segments as in the case of insulin (see Discussion).

Figures 7 and 8 compare the experimental CD and ORD spectra of the five additional proteins to the calculated spectra based on the computed *f*'s (Table VII). The agreement is satisfactory to good in all cases; in general, the CD data were in better agreement than the ORD data. The notable exceptions are α-chymotrypsin and chymotrypsinogen A, for which both the calculated CD and ORD began to deviate from the experimental values below 210 nm. For insulin and cytochrome *c* the calculated rotations near the 233-nm minimum were much larger in magnitude than the experimental rotations. As in the case of lactate dehydrogenase, we then redetermined the *f*'s using the reference values between 225 and 240 nm (Table V) and recalculated the ORD spectra. This procedure did improve the curve fitting in the region of 233-nm minimum, but it caused deviations below 225 nm. Note that the magnitude of the helical ORD between 210 and 220 nm is comparatively larger than that between 225 and 240 nm (Figure 2). The computation of the *f*'s would therefore be weighted by the reference values at shorter wavelengths. This, in turn, could result in a good curve fitting between 210 and 220 nm, but not necessarily at longer wavelengths. No such difficulties, however, were encountered in the curve fitting of the CD data.

In previous work (Chen and Yang, 1971), we have shown that a quick estimate of the helical content of globular proteins can be obtained from the following relationships, whose coefficients are subject to minor revisions in the future

$$b_0 = -580 f_H \quad (2a)$$

$$[m]_{233} = -12,700 f_H - 2520 \quad (2b)$$

$$[\theta]_{222} = -30,300 f_H - 2340 \quad (2c)$$

This is possible because the mean residue ellipticity and rotation of the β and unordered from virtually coincide near 222 and 233 nm, respectively (Figures 3 and 4), and (that) the b_0 of the β and unordered form is insignificant as compared to that of the helix. Thus, eq 1 in these cases is reduced to a linear function of f_H only. Table VIII shows the estimated helical contents of several proteins based on equations 2a–c which compare favorably to those listed in Table VII. Equation 2c is essentially based on the same principle as Myer's isodichroic points method (1970), although he found the point of crossing for the β form and "random coil" of poly(L-lysine) at 208 nm instead. For more precise determination of the helix, β, and unordered form of a protein, the method of analysis reported in this work is preferred over 2a–c.

Discussion

Inherent in our method of analysis is the assumption that the structure of a protein is identical in solid state and in aqueous solution. Whether this is true was recognized as far back as 1949 by Kendrew (see also the reviews by Urnes, 1963, and Rupley, 1969). For our purpose we merely wish to point out that wet crystals of proteins containing considerable amounts of salt and solvent were used for X-ray diffraction studies. It therefore seems reasonable to assume that the structure of the protein molecules remains unchanged when they are crystallized out from solution. True, the conformations of many synthetic polymers are sensitive to the solvent composition and salt concentration. But the protein molecules, unlike flexible polymers, are compact and rigid. It seems improbable, although not impossible, that the structure of protein molecules would undergo drastic changes when their crystals dissolve in aqueous solution. This, of

TABLE VIII: Estimation of Helical Contents of Several Proteins from b_0, $[m]_{233}$, and $[\theta]_{222}$.[a]

Proteins	X-Ray	b_0		$[m]_{233}$		$[\theta]_{222}$	
		This Work	Previous[b]	This Work	Previous[c]	This Work	Previous[d]
Myoglobin	0.77	0.77	0.71	0.78	0.57	0.77	0.66
Lysozyme	0.29	0.28	0.26	0.26	0.19	0.24	0.33
Lactate dehydrogenase	0.29	0.31	0.29	0.28	0.20	0.33	0.24
Papain	0.21	0.21	0.20	0.23	0.17	0.22	0.19
Ribonuclease	0.19	0.16	0.14	0.21	0.15	0.19	0.23
Insulin	0.22–0.45	0.36	0.33	0.24	0.17	0.30	0.33
Nuclease	0.24	0.28	0.26	0.26	0.19	0.28	0.25
Cytochrome *c*	0.11–0.39			0.24	0.17	0.28	0.16
α-Chymotrypsin	0.09			0.16	0.11	0.06	0.13
Chymotrypsinogen A	0.06			0.11	0.07	0.05	0.19

[a] The calculations in this work were based on eq 2a–c. [b] Based on $f_H = -b_0/630$ (Moffitt and Yang, 1956). [c] Based on $f_H = -([m']_{233} + 2{,}000/13{,}000$ (Yang, 1967). Here $[m']_{233} = 0.764[m]_{233}$. [d] Based on measurements at 208 nm: $f_H = -([\theta]_{208} + 4{,}000)/29{,}000$ (Greenfield and Fasman, 1969).

course, does not rule out local fluctuation of conformation at certain sites of the protein molecules. But our method of analysis remains valid as long as the secondary structures of the protein molecules are not affected by transition from crystal to solution and *vice versa*.

Virtually all the space within a protein molecule is taken up by amino acid residues. Thus, the optical activity of the peptide chromophores will be affected not only by the ordered structures such as helices and β forms but also by the neighboring residues that interact with one another. Equation 1 ignores these interactions, whose rotatory contributions are assumed to be insignificant compared to those of the structural elements. In our treatment we have also ignored the optical activity of non-peptide chromophores, which we assume to be negligible. Obviously, in cases where the Cotton effects of the aromatic groups or disulfide bonds could significantly alter the CD and ORD spectra, it might be difficult to apply our method of analysis.

In principle, eq 1 can be further expanded to include more than three terms on the right side of the equation. For example, the X_H for α and 3_{10} helices can be separated into two terms. The same may be true for the β form, although it is too complex to contemplate at present. If necessary, we can also include terms for the optical activity of non-peptide chromophores but their contributions may vary from one protein to another, thus making any quantitative analysis difficult. Such refinement might be considered when the results of X-ray diffraction studies on many other proteins become available. The results of any numerical analysis depends on the functions used as in eq 1. Our method of analysis is of necessity oversimplified. Unless rotatory contributions other than those described in eq 1 are comparatively small, the X_H, X_β, and X_R so determined might be subject to question. But the effects of non-peptide chromophores and of interactions between side-chain groups and between side-chain groups and peptide units will very likely be random in sign and are expected to compensate each other considerably. Thus, we believe that at least our X_H values do reflect the optical activity of the helical segments in proteins.

It remains to explain the difference in magnitude between the CD and ORD of the helix computed from the reference proteins and those based on synthetic polypeptides. For instance, the b_0, $[m]_{233}$, and $[\theta]_{222}$ of the proteins were lower by 10–20% than the corresponding values for helical poly(L-glutamic acid) and poly(L-lysine) (Cassim and Yang, 1970). Aside from the imperfection of certain helical segments in the reference proteins, which could result in different optical activity, the above difference could largely be due to the end effects of the polypeptide chain on its rotatory properties. Unlike synthetic polypeptides, the protein molecules contain only short helical segments. Figure 9 shows the distribution of the number of residues per segment of the five reference proteins employed in this work. Most of the segments appear to have 6–12 residues. The average number of residues per segment is 13.5 for myoglobin (assuming E1 and E2 are two separate segments), 9.1 for lysozyme, 10 for lactate dehydrogenase, 11 for papain, and 8 for ribonuclease; the overall average for 5 proteins is about 11 residues/segment. In contrast, synthetic polypeptides may have as many as several hundred residues in a single chain, the end effects of which are negligible for all practical purposes. Recent studies by Yaron *et al.* (1971) on the oligopeptides of L-lysine clearly showed a chain-length dependence of optical rotations of these oligomers. The investigators concluded that the onset of helicity began at dodecamer for poly(L-lysine) in aqueous solution. These oligopeptides form only partial helices, and therefore

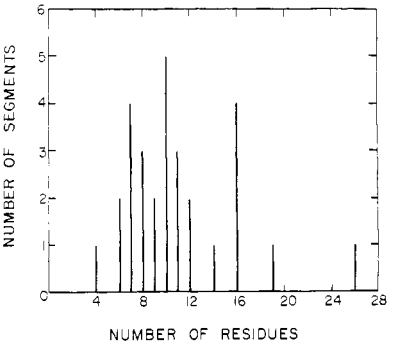

FIGURE 9: Distribution of the number of residues per helical segment for myoglobin, lysozyme, lactate dehydrogenase, papain, and ribonuclease.

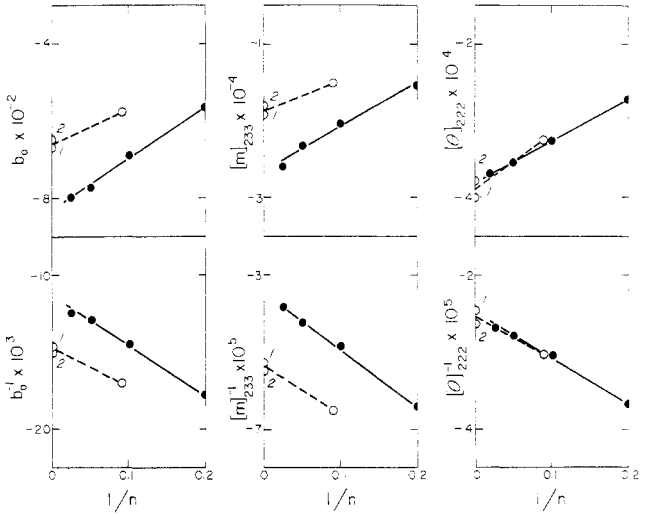

FIGURE 10: Variation of b_0, $[m]_{233}$, and $[\theta]_{222}$ of the helix with chain length. Closed circles, theoretical values of Woody and Tinoco (1967); open circles, computed value at $1/n = 0.9$. Open half-circles 1 and 2 refer to experimental value of poly(L-glutamic acid) and poly(L-lysine).

cannot be compared directly to the short, rigid helical segments in a protein molecule. The theoretical work of Woody and Tinoco (1967), however, does show that the optical activity of a helical polypeptide (poly(L-alanine)) is chain-length dependent.

We found that Woody and Tinoco's calculated values of b_0, $[m]_{233}$, and $[\theta]_{222}$ for 5-, 10-, 20-, and 40-mers can be fitted with

$$X_H = X_H^{\infty}(1 - k_1/n) \quad (3a)$$

$$1/X_H = (1/X_H^{\infty})(1 + k_2/n) \quad (3b)$$

Here X_H^{∞} is the rotatory parameter for a helix of infinite length, that is, the number of residues, n, goes to infinity, and k_1 and k_2 are two constants to be determined. We calculated $[\theta]_{222}$ from the adjusted rotational strength for the n-π^* transition of an α helix in Table II of the report by Woody and Tinoco (1967), using an assigned bandwidth of 2250 cm^{-1} for a gaussian band. We read off the b_0 and $[m]_{233}$ from Figures 6 and 7 of the same paper. (Woody and Tinoco's b_0 was calculated by presetting $\lambda_0 = 215.2$ nm in the Moffitt equation instead of the customary 212 nm. These investigators reported $[m']_{233}$, which we converted back to $[m]_{233}$.) Thus, the theoretical calculations of Woody and Tinoco (1967) give a linear plot of X_H *vs.* $1/n$ or a linear double-reciprocal plot of $1/X_H$ *vs.* $1/n$ (Figure 10). Experimentally, Yaron *et al.* (1971) also found such a linear relationship between $[m]_{233}$ and $1/n$ for the oligopeptides of L-lysine.

Included in Figure 10 are the helical b_0, $[m]_{233}$, and $[\theta]_{222}$ of this work, assuming an average n if 11, and also those of helical poly(L-glutamic acid) and poly(L-lysine) for which $1/n$ approaches zero. Although the theoretical values differ significantly from the experimental values, their slopes, $-k_1X_H^{\infty}$ or k_2/X_H^{∞}, are close to each other in all three cases. Admittedly, the experimental data were limited to only two n values, and the use of an average n of 11 for proteins is arbitrary and will almost certainly be subject to revision in future investigations. We believe that eq 3 is a good representation of the X_H^{∞} parameters. (Currently, the helical b_0 has been taken as -630 (deg cm^2)/dmole. This value was orginally based on the ORD of poly(γ-benzyl L-glutamate) in organic solvents and poly(L-glutamic acid) in acidic dioxane–water mixture. Since then the magnitude of b_0 (helix) for poly(L-glutamic acid) in aqueous solution has been raised to about -670 to -690 (see, for example, Cassim and Yang, 1970). Similarly, the b_0 (helix) for poly(L-lysine) in aqueous solution was found to be about -670. Thus, the helical b_0^{∞} for proteins in aqueous solution would probably be closer to -700 than to -630.)

To take into consideration the variation in n of the helical segments in a protein molecule, the term $f_H X_H$ in eq 1 can be rewritten as

$$\sum f_{H,i} X_H^{\infty}(1 - k/n_i) = f_H^{app} X_H^{\infty}(1 - k/n) \quad (4)$$

Here f_H^{app} is the apparent fraction of helix on the basis of an average n residues per segment, 11 in this work; i is the number of helical segments in a protein molecule, and n_i the number of residues in the ith segment. If n_i's are smaller (or larger) than n, f_H^{app} in eq 4 will be smaller (or larger) than f_H, that is, the helical content will be underestimated (or overestimated). Take, for example, insulin, which has three helical segments of 5, 7, and 11 residues. Two of these chain lengths are shorter than the average n of 11, thus leading to an underestimate of f_H according to eq 4. Likewise, cytochrome c has five short helical segments of 5, 5, 6, 9, and 5 residues in addition to one true helix of 11 residues. This again would make f_H^{app} smaller than the true f_H. We found that k in eq 4 was of the order of 2–4 over the entire wavelength range of the CD spectra of helices. Since the assignment of 11 to n is arbitrary and since there are not enough experimental points in Figure 10 to warrant a precise determination of k, no quantitative evaluation of such a correction for f_H^{app} will be attempted in this work.

After the submission of our paper, we learned of the work of Madison and Schellman (1972). Their theoretical calculations of the optical activity of ordered polypeptides overlap the studies by Woody (1968, 1969) and Pysh (1970). Calculations on myoglobin, lysozyme, ribonuclease S, and α-chymotrypsin clearly show the chain-length dependence of optical activity of the helical segments in proteins. Some of the distorted helices were found to have greater optical activities than the ideal helices; this reinforces our contention that the counting of a maximum helical content including α,3_{10}, and distorted helices is preferred in the computation of the reference values X_H (eq 1). The CD of the β regions of proteins differs markedly from that for ideal β forms because of a breakdown in symmetry, thus making their quantitative estimates difficult.

The method of analysis proposed in this work does provide a reasonable means of determining the helical, β, and unordered form of protein molecules in aqueous solution. The computed CD and ORD spectra of the three forms differ significantly from those of synthetic polypeptides. We believe that our use of proteins as reference standards is more realistic than the use of synthetic polypeptides. The difference in magnitude between the computed values of the helical form and experimental results of helical polypeptides can largely be accounted for by chain-length dependence. However, the reference spectra of the three forms are still subject to future revision and refinement as the interpretations of the X-ray results are modified. Our method of analysis represents only a new attempt to determine the secondary structures of proteins by optical activity.

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.

Adams, M. J., Ford, G. C., Koekoek, R., Lentz, P. J., Jr., McPherson, A., Jr., Rossmann, M. G., Smiley, I. E., Schevitz, R. W., and Wonacott, A. J. (1970), *Nature (London) 227*, 1098.

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.

Beychok, S. (1968), *Annu. Rev. Biochem. 37*, 437.

Birkloft, J. J., Matthews, B. W., and Blow, D. M. (1969), *Biochem. Biophys. Res. Commun. 36*, 131.

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.

Blow, D. M. (1971), *Enzymes 3*, 185.

Blow, D. M., Birkloft, J. J., and Hartley, B. S. (1969), *Nature (London) 221*, 337.

Blundell, T. L., Cutfield, J. F., Dodson, E. J., Dodson, G. G., Hodgkin, D. C., and Mercola, D. A. (1972), *Cold Spring Harbor Symp. Quant. Biol. 36*, 233.

Cassim, J. Y., and Yang, J. T. (1969), *Biochemistry 8*, 1947.

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

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

Cotton, F. A., and Hazen, E. E., Jr. (1971), *Enzymes 4*, 153.

Crick, F. H. C., and Kendrew, J. C. (1957), *Advan. Protein Chem. 12*, 133.

Davidson, B., Tooney, N., and Fasman, G. D. (1966), *Biochem. Biophys. Res. Commun. 23*, 156.

Dickerson, R. E., and Geis, I. (1969), The Structure and Function of Proteins, New York, N. Y., Harper & Row.

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

Doty, P., and Yang, J. T. (1956), *J. Amer. Chem. Soc. 78*, 478.

Drenth, J., Jansonius, J. N., Koekoek, R., Swen, H. M., and Wolthers, B. G. (1968), *Nature (London) 218*, 929.

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

Flatmark, T., and Robinson, A. B. (1968), *in* Structure and Function of Cytochromes, Okunuki, K., Kamen, M. D., and Sekuzu, I., Ed., Tokyo, University of Tokyo Press, p 318.

Freer, S. T., Kraut, J., Robertus, J. D., Wright, H. T., and Xuong, N. H. (1970), *Biochemistry 9*, 1997.

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

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

Iizuka, E., and Yang, J. T. (1966), *Proc. Nat. Acad. Sci. U. S. 55*, 1175.

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

Kendrew, J. G. (1949), *in* Haemoglobin, Roughton, F. J. W., and Kendrew, J. C., Ed., London, Butterworths, p 148.

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.

Kraut, J. (1971), *Enzymes 3*, 165.

Kraut, J., Wright, H. T., Kellerman, M., and Freer, S. T. (1967), *Proc. Nat. Acad. Sci. U. S. 58*, 304.

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

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

Matthews, B. W., Sigler, P. B., Henderson, R., and Blow, D. M. (1967), *Nature (London) 214*, 652.

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

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

Parrish, J. R., Jr., and Blout, E. R. (1971), *Biopolymers 10*, 1491.

Pysh, E. S. (1970), *J. Chem. Phys. 52*, 4723.

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.

Rupley, J. A. (1969), *in* Structure and Stability of Biological Macromolecules, Timasheff, S. N., and Fasman, G. D., Ed., New York, N. Y., Marcel Dekker, Chapter 4.

Sarkar, P., 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.

Sigler, P. B., Blow, D. M., Matthews, B. W., and Henderson, R. (1968), *J. Mol. Biol. 35*, 143.

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

Timasheff, S. N., and Stevens, L. (1969), *Trans. Bose Res. Inst. Calcutta 31*, 50.

Townend, R., Kumosinski, T. F., Timasheff, S. N., Fasman, G. D., and Davidson, B. (1966), *Biochem. Biophys. Res. Commun. 23*, 163.

Urnes, P. (1963), Ph.D. Dissertation, Harvard University, Cambridge, Mass.

Urry, D. W., and Doty, P. (1965), *J. Amer. Chem. Soc. 87*, 2756.

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. (1967), *in* Poly-α-amino Acids, Fasman, G. D., Ed., New York, N. Y., Marcel Dekker, Chapter 6.

Yang, J. T. (1969), *in* A Laboratory Manual of Analytical Methods of Protein Chemistry, Alexander, P., and Lundgren, H. P., Ed., New York, N. Y., Pergamon Press, Chapter 2.

Yang, J. T., and Chen, Y.-H. (1972), *Fed. Proc., Fed. Amer. Soc. Exp. Biol. 31*, 469.

Yang, J. T., and Doty, P. (1957), *J. Amer. Chem. Soc. 79*, 761.

Yaron, A., Katchalski, E., Berger, A., Fasman, G. D., and Sober, H. A. (1971), *Biopolymers 10*, 1107.

Zand, R., and Vinogradov, S. (1968), *Arch. Biochem. Biophys. 125*, 94.