248, 3313.
Lowry, O. H., Rosebrough, N. J., Farr, A. L., and Randall, R. J. (1951), *J. Biol. Chem. 193*, 265.
Mathews, B. W., Colman, P. M., Jansonius, J. N., Titani, K., Walsh, K. A., and Neurath, H. (1972), *Nature (London), New Biol. 238*, 41.
McCubbin, W. D., and Kay, C. M. (1973), *Biochemistry 12*, 4228.
Moore, S. (1963), *J. Biol. Chem. 238*, 235.
Moore, S., and Stein, W. H. (1963), *Methods Enzymol. 6*, 819.
Morrison, W. R. (1964), *Anal. Biochem. 7*, 218.
Nagy, B., Greaser, M. L., and Gergely, J. (1973), *Fed. Proc., Fed. Amer. Soc. Exp. Biol. 32*, 570 Abstr.
Nanninga, L. B., and Kempen, R. (1971), *Biochemistry 10*, 2449.
Nockolds, C. E., Kretsinger, R. H., Coffee, C. J., and Bradshaw, R. A. (1972), *Proc. Nat. Acad. Sci. U. S. 69*, 581.
Nyman, P-O., and Lindskog, S. (1964), *Biochim. Biophys. Acta 85*, 141.
Pauling, L. (1960), The Nature of the Chemical Bond and the Structure of Molecules and Crystals, Ithaca, N. Y., Cornell University Press, p 514.
Scatchard, G. (1949), *Ann. N. Y. Acad. Sci. 51*, 660.
Schubert, J. S., Russell, E. R., and Myers, L. S., Jr. (1950), *J. Biol. Chem. 185*, 387.
Sillén, L. G., and Martell, A. E. (1971), *Chem. Soc. Spec. Publ. No. 25*, 14.
Sipos, T., and Merkel, J. R. (1970), *Biochemistry 9*, 2766.
Sottacassa, G., Sandri, G., Panfili, E., de Bernard, B., Gazzotti, P., Vasington, F. D., and Carofoli, E. (1972), *Biochem. Biophys. Res. Commun. 47*, 808.
Stauffer, C. E., and Treptow, R. S. (1973), *Biochim. Biophys. Acta 295*, 457.
Steinhardt, J. and Reynolds, J. A. (1969), Multiple Equilibria in Proteins, New York, N. Y., Academic, p 26.
Svennerholm, L. (1957), *Biochim. Biophys. Acta 24*, 604.
Taniuchi, H., Anfinsen, C. B., and Sodja, A. (1967), *Proc. Nat. Acad. Sci. U. S. 58*, 1235.
Teo, T. S., and Wang, J. H. (1973), *J. Biol. Chem. 248*, 5950.
Warren, L. (1959), *J. Biol. Chem. 234*, 1971.
Wasserman, R. H., and Corradino, R. A. (1971), *Annu. Rev. Biochem. 40*, 501.
Wasserman, R. H., Corradino, R. A., and Taylor, A. N. (1968), *J. Biol. Chem. 243*, 3978.
Wasserman, R. H., and Taylor, A. N. (1966), *Science 152*, 791.
Wasserman, R. H., and Taylor, A. N. (1972), *Annu. Rev. Biochem. 41*, 179.
Wasserman, R. H., Taylor, A. N., and Fullmer, C. S. (1974), *Biochem. Soc. Spec. Publ. No. 3* (in press).
Wilcox, P. E. (1967), *Methods Enzymol. 11*, 63.
Yamada, S., and Tonomura, Y. (1972), *J. Biochem. (Tokyo) 72*, 417.
Zak, B., Moss, N., Boyle, A. J., and Zlatkis, A. (1954), *Anal. Chem. 26*, 776.

Practical Aspects of Calculating Protein Secondary Structure from Circular Dichroism Spectra†

Thomas D. Barela‡ and Dennis W. Darnall*,§

ABSTRACT: Protein secondary structure fractions were calculated from analysis of circular dichroism (CD) spectra assuming that the protein CD spectrum between 190 and 240 nm could be described by a linear combination of model secondary structure spectra (the α helix, β structure, and random coil of poly-L-lysine). This method is identical with that reported by Greenfield and Fasman (*Biochemistry 8*, 4108 (1969)). Wavelength dependence of calculated secondary structure fractions was determined by subdividing the 191–240-nm wavelength region into smaller subregions and calculating structure fractions based on the smaller subregions. Standard deviations for calculated structure fractions were generally less than 4%. Between 190 and 240 nm a particular secondary structure fraction could be varied as much as 39% by altering the wavelengths chosen for analysis. As judged by higher standard deviations about the calculated structure fractions, as well as calculation of negative structure fraction values, the 220–240-nm segment was the region of greatest error. We concluded that the greatest uncertainty associated with the use of this method was the choice of the experimental wavelengths used in the analysis, and further, that no choice of wavelengths could be made which would be useful for all proteins.

Optical rotatory properties of the protein polypeptide backbone have been extremely valuable for examining protein secondary structures in solution. Greenfield *et al.* (1967) and Magar (1968) quantitated secondary structures of proteins from optical rotatory dispersion (ORD) spectra by a least-squares regression analysis of a protein ORD spectrum assuming a linear relationship between the protein spectrum and spectra of the α helix, β structure, and random coil of poly-L-lysine. Technical advantages of circular dichroism (CD) over optical rotatory dispersion led to an analogous

† From the Department of Chemistry, New Mexico State University, Las Cruces, New Mexico 88003. *Received November 12, 1973.* This work was supported in part by Grant AM-13665 from the National Institute of Arthritis and Metabolic Diseases, U. S. Public Health Service. Presented in part at the 166th Annual Meeting of the American Chemical Society, Chicago, Ill., Aug 1973. This work constitutes a portion of the Ph.D. Thesis of T. D. B., New Mexico State University.

‡ Present address: The Biomedical Division, The Samuel Roberts Noble Foundation, Inc., Ardmore, Okla. 73401.
§ Recipient of U. S. Public Health Service Research Career Development Award, GM-32014.

treatment of CD spectra (Greenfield and Fasman, 1969), with some improvement in the correspondence between the calculated structure estimates and those determined by X-ray crystallography.

For the calculation of ordered protein secondary structures by CD, ordered synthetic polypeptides have been used as models for similar structures in proteins. The validity of this practice has been questioned by many investigators. For instance, the CD spectrum of poly-L-lysine at pH 11 was chosen as the model for α-helix protein structures (Greenfield *et al.*, 1967; Magar, 1968; Greenfield and Fasman, 1969); however, other polypeptides also were available in helical conformations (Breslow *et al.*, 1965). In addition, X-ray crystallographic analysis indicated that α-helixal segments in proteins were short compared to extended helices in polypeptides. This disparity was magnified by reports that helix optical rotatory strength depended upon helix chain length (Woody and Tinoco, 1967; Vournakis *et al.*, 1968; Tinoco *et al.*, 1963).

Objections concerning use of β-structured poly-L-lysine produced by heating in water at pH 11 as the model for protein β structure follow the same reasoning as those to the choice of the α-helix model. Sarkar and Doty (1966) and Li and Spector (1969) found that the spectrum of β-structured poly-L-lysine produced in sodium dodecyl sulfate was different from that produced in water at pH 11 (Townend *et al.*, 1966). In addition, β-structure optical activity showed side chain and solvent dependency (Tooney and Fasman, 1968; Quadrifoglio and Urry, 1968; Fasman and Potter, 1967).

The random coil model spectrum has been the object of the most persistent objections. X-Ray crystallography has revealed that unordered segments of proteins are constrained, in contradistinction to the flexible extended synthetic polypeptides which model this structure. Furthermore, there was little correspondence between dichroism spectra of random structured synthetic polypeptides and fully denatured proteins (Dearborn and Wetlaufer, 1970).

In addition to these considerations, calculation of protein secondary structures based only on polypeptide backbone circular dichroism have not considered known optical activity of disulfide bonds and aromatic amino acids (Urry, 1970; Urry *et al.*, 1968; Simons and Glazer, 1967) that certainly must contribute to protein spectra in the 190–240-nm region.

The purpose of our studies was to determine the extent to which these objections compromised the usefulness of circular dichroism for estimating ordered structure in proteins. To examine this, we tested a simple hypothesis: if model spectra can be used to reconstruct protein spectra in the range 190–240 nm, then any combination of wavelengths within this region should calculate the same structure fractions. When structure fractions were calculated using combinations of different wavelengths, the structure fractions were markedly changed.

Methods

Circular dichroism spectra were taken from previous reports. The myoglobin and poly-L-lysine spectra were reported by Greenfield and Fasman (1969). Spectra for lysozyme, ribonuclease, lactate dehydrogenase, and papain were reported by Chen *et al.* (1972). Spectra were supplied to us as tabulated mean residue ellipticities at regular intervals except for poly-L-lysine values which were tabulated at irregular intervals. Mean residue ellipticities at intermediate wavelengths were determined by graphing tabulated values in accordance with the published complete spectra.

Mathematical procedures were essentially identical with those described by Magar (1968). The protein spectrum was expressed as a linear combination of secondary structure model spectra (eq 1), subject to the constraint (eq 2)

$$a[\theta]_\lambda^\alpha + b[\theta]_\lambda^\beta + c[\theta]_\lambda^{RC} = [\theta]_\lambda^{Prot} \quad (1)$$

$$a + b + c = 1.000 \quad (2)$$

where a, b, and c are the secondary structure fractions for the α helix, β structure, and random coil, respectively, and $[\theta]_\lambda^\alpha$, $[\theta]_\lambda^\beta$, and $[\theta]_\lambda^{RC}$ are the mean residue ellipticities at any wavelength, λ, for the α helix, β structure, and random coil model spectra, respectively. $[\theta]^{Prot}$ is the protein mean residue ellipticity at the same wavelength, λ. An equation of the same form as eq 1 was formed at 1-nm intervals throughout the 191–240-nm region.

The experimental approach involved dividing up the entire 191–240-nm wavelength region into smaller regions; the equations used for each region were termed the "regression model" for that region. Ellipticities for protein and model secondary structure spectra within the smaller regions were used to calculate the protein secondary structure fractions as a function of wavelength. If one can describe the protein spectrum in terms of poly-L-lysine secondary structure models, then each of the smaller regions should calculate the same secondary structure fractions. Calculations of large standard deviations for secondary structure fractions for a particular wavelength region indicated a lack of correspondence between the model structures and the protein.

Secondary structure fractions were calculated by least-squares regression with the aid of an IBM 360-65 computer. Numerical methods for least-squares regression and calculation of standard deviations are outlined in the National Bureau of Standards Handbook 91 (Natrella, 1963).

Residual analysis was performed by the criteria set forth by Draper and Smith (1966). A residual (eq 3) was defined as the difference between the experimentally determined protein ellipticity at any wavelength λ, $[\theta]_\lambda^{Prot}$, and the ellipticity

$$\text{residual} = [\theta]_\lambda^{Calcd} - [\theta]^{Prot} \quad (3)$$

calculated by eq 1 subsequent to determining secondary structure fractions, $[\theta]_\lambda^{calcd}$. Standardized residuals were calculated and plotted by the FIT command of the OMNITAB II statistical package (U. S. National Bureau of Standards, OMNITAB II version 5.00 May, 1970).

Results

Tables I–V give calculated secondary structure fractions for the five proteins examined. Since our results were derived from identical testing of five protein spectra, when possible, observations pertinent to one protein will be extended to other proteins without detailed descriptions. To facilitate easy cross-referencing of tables, each line contains a separate regression model and a similarly numbered line in another table contains the results for the same regression model using a different protein. For example, lines 2 in Tables I–V all contain the results for the regression model formed over the entire 191–240-nm region, while line 7 contains the results using the regression formed over the smaller 205–225-nm region. Unless specifically noted, regression equations were formed at 1-nm intervals, *e.g.*, the 191–240-nm model in line 2 contained 50 equations, while the 205–225-nm model in line 7 contained 21 equations. Each table also contains a line indicating the reported secondary structure fractions as determined: by X-ray

TABLE I: Calculated Structure Fractions for Lysozyme.

Region (nm)	Fraction α	Fraction β	Fraction RC	σ_α	σ_β	σ_{RC}
1. X-ray[a]	0.28–0.42	0.10	0.62–0.48			
2. 191–240[b]	0.281	0.335	0.383	0.018	0.026	0.011
3. 191–230	0.282	0.334	0.384	0.021	0.029	0.013
4. 191–220	0.309	0.307	0.384	0.022	0.030	0.013
5. 200–240	0.262	0.274	0.465	0.013	0.018	0.012
6. 205–240	0.289	0.181	0.531	0.009	0.018	0.011
7. 205–225	0.288	0.182	0.530	0.012	0.023	0.015
8. 211–240	0.270	0.211	0.519	0.019	0.038	0.021
9. 221–240	0.370	−0.064	0.693	0.026	0.064	0.039
10. 205–240[c]	0.284	0.201	0.515	0.016	0.031	0.020
11. 208–240[c]	0.324	0.112	0.564	0.021	0.045	0.026
12. 191–200	0.238	0.426	0.335	0.019	0.026	0.009
13. 201–210	0.326	0.232	0.442	0.019	0.016	0.017
14. 211–220	0.340	0.098	0.561	0.034	0.061	0.028
15. 221–230	0.344	−0.001	0.659	0.052	0.128	0.076
16. 231–240	0.357	0.018	0.625	0.029	0.092	0.064
17. Greenfield[d]	0.285	0.111	0.604			
18. Rosenkrantz[e]	0.26	0.06	0.68			

[a] Phillips (1967). [b] Models formed from equations at 1-nm intervals. [c] Regression model formed from wavelengths at the irregular intervals reported by Greenfield and Fasman (1969). [d] Greenfield and Fasman (1969). [e] Rosenkrantz and Scholtan (1971).

TABLE II: Calculated Structure Fractions for Myoglobin.

Region (nm)	Fraction α	Fraction β	Fraction RC	σ_α	σ_β	σ_{RC}
1. X-ray[a]	0.65–0.77	0.00	0.23–0.32			
2. 191–240[b]	0.729	0.012	0.259	0.015	0.021	0.009
3. 191–230	0.725	0.016	0.259	0.017	0.023	0.010
4. 191–220	0.724	0.018	0.258	0.020	0.027	0.012
5. 200–240	0.698	0.012	0.290	0.013	0.018	0.012
6. 205–240	0.669	0.093	0.238	0.016	0.033	0.021
7. 205–225	0.643	0.113	0.245	0.009	0.019	0.012
8. 211–240	0.820	−0.218	0.398	0.036	0.075	0.040
9. 221–240	1.033	−0.732	0.699	0.049	0.012	0.074
10. 205–240[c]	0.643	0.134	0.223	0.023	0.045	0.029
11. 208–240[c]	0.685	0.047	0.270	0.036	0.076	0.044
12. 191–200	0.807	−0.083	0.276	0.040	0.055	0.018
13. 201–210	0.634	0.028	0.338	0.026	0.023	0.024
14. 211–220	0.548	0.266	0.186	0.022	0.040	0.018
15. 221–230	1.043	−0.760	0.721	0.080	0.195	0.116
16. 231–240	0.863	−0.115	0.252	0.080	0.255	0.178
17. Greenfield[d]	0.685	0.047	0.270			
18. Rosenkrantz[e]	0.64	0.00	0.36			

[a] Kendrew *et al.* (1960). [b] Models formed from equations at 1-nm intervals. [c] Regression model formed from wavelengths at the irregular intervals reported by Greenfield and Fasman (1969). [d] Greenfield and Fasman (1969). [e] Rosenkrantz and Scholtan (1971).

(line 1); by Greenfield and Fasman (1969) who used irregular wavelength intervals to form the regression equations (line 17); and by Rosenkrantz and Scholtan (1971) who replaced the poly-L-lysine random coil model spectrum with a random coiled poly-L-serine spectrum (line 18).

Greenfield and Fasman (1969) and Rosenkrantz and Scholtan (1971) used only the 208–250-nm wavelength region in their analysis because of expected nonpeptide optical activity outside of this region. Myer (1970), by utilizing isodichroic points, did not consider wavelengths below 200 nm. Chen *et al.* (1972) using a related technique (discussed later) did not consider chromophores other than the peptide-amide bond. Our initial efforts, therefore, were intended to verify that the 205–240-nm region would be best. By subdividing the 205–240-nm region into smaller regions (lines 6–11, 14–16 in Tables I–V) we found that calculated structure fractions were different for each region and the range was much more than could be attributed to statistical error alone. For example, compare lines 6, 7, and 9 in Tables I–V, 205–240-, 205–225-, and 221–240-nm regression models, respectively. These three regression models were formed from increasing proportions of wavelengths above 220 nm. Calculated structure

TABLE III: Calculated Structure Fractions for Ribonuclease.

Region (nm)	Fraction α	Fraction β	Fraction RC	σ_α	σ_β	σ_{RC}
1. X-ray[a]	0.06–0.18	0.36	0.46–0.58			
2. 191–240[b]	0.139	0.426	0.435	0.014	0.020	0.009
3. 191–230	0.147	0.417	0.437	0.014	0.019	0.008
4. 191–220	0.169	0.393	0.438	0.012	0.017	0.007
5. 200–240	0.142	0.396	0.426	0.016	0.022	0.014
6. 205–240	0.159	0.338	0.503	0.021	0.043	0.028
7. 205–225	0.191	0.314	0.496	0.015	0.031	0.020
8. 211–240	−0.049	0.754	0.294	0.041	0.084	0.045
9. 221–240	−0.262	1.263	−0.001	0.050	0.124	0.075
10. 205–240[c]	0.164	0.360	0.476	0.030	0.060	0.038
11. 208–240[c]	0.178	0.329	0.493	0.052	0.111	0.064
12. 191–200	0.082	0.515	0.403	0.010	0.014	0.005
13. 201–210	0.225	0.351	0.424	0.013	0.012	0.012
14. 211–220	0.264	0.200	0.535	0.049	0.089	0.041
15. 221–230	−0.213	1.154	0.061	0.034	0.082	0.048
16. 231–240	−0.091	0.568	0.523	0.101	0.323	0.225
17. Greenfield[d]	0.093	0.326	0.581			
18. Rosenkrantz[e]	0.10	0.31	0.59			

[a] Kartha *et al.* (1967). [b] Models formed from equations at 1-nm intervals. [c] Regression model formed from wavelength at the irregular intervals reported by Greenfield and Fasman. [d] Greenfield and Fasman (1969). [e] Rosenkrantz and Scholtan (1971).

fractions for any one protein show considerable variance between these three regression models. For lysozyme, myoglobin, and ribonuclease, negative structure fractions were calculated in the 221–240-nm region. Negative structure fractions were physically meaningless, and interpreted to mean that the model spectra did not describe adequately the protein spectrum in that region. In most cases where negative structure fractions were not calculated, variations between calculated structure fractions exceeded associated error terms. For lactate dehydrogenase (Table V), the 205–240-nm regression model calculated a β-structure value of 38.6%, while the 221–240-nm regression model calculated a 23.0% value. Standard deviations were 1.9 and 5.8%, respectively. Error terms can only account for 7.7% of the 15.6% variation in structure fractions.

The above studies indicated that regression models formed from increasing proportions of wavelengths above 220 nm did not describe the protein spectrum very well. We then divided the entire 191–240-nm region into five 10-nm regression models (lines 12–16). Negative structure fractions were calculated by either the 221–230- or 231–240-nm regression models (lines 15 and 16) for all five proteins. Only in the case of myoglobin did any of the other three 10-nm regression models result in calculation of negative structure fractions.

To examine the 220–240-nm wavelength region further, we formed the 191–240-, 191–230-, and 191–220-nm regression models. We expected the 191–220-nm model to calculate much different structure fractions since this regression model did not contain any wavelengths above 220 nm. However, structure fractions calculated by these three models were nearly identical. Although at first surprising, this observation can be explained by the matrix multiplications necessary in the least-squares regression (Natrella, 1963). The ellipticities of all spectra, both structure models and protein, were much larger below 200 nm than above (Greenfield and Fasman, 1969; Dearborn and Wetlaufer, 1970; Chen *et al.*, 1972). These larger numbers (below 200 nm) when multiplied and squared in the regression calculation exerted a dominating influence on the lower values above 220 nm.

Since the method of Greenfield and Fasman is widely used, it was of interest to examine the numerical variations associated with its use. Greenfield and Fasman (1969) formed their regression equations from 13 irregularly spaced wavelengths: 208, 210, 211, 214, 215, 217, 220, 222, 225, 230, 234, 238, and 240 nm. These wavelengths generally describe the peaks and troughs in the poly-L-lysine structure model spectra. Our analysis using these irregular wavelengths is given in line 11. A similar model employing a single additional equation at

TABLE IV: Calculated Structure Fractions for Papain.

Region (nm)	Fraction α	Fraction β	Fraction RC	σ_α	σ_β	σ_{RC}
1. X-ray[a]	0.20	0.05	0.74			
2. 191–240[b]	0.150	0.381	0.469	0.010	0.014	0.006
3. 191–230	0.151	0.380	0.469	0.012	0.016	0.007
4. 191–220	0.146	0.385	0.469	0.014	0.020	0.008
5. 200–240	0.160	0.367	0.474	0.008	0.012	0.007
6. 205–240	0.174	0.320	0.506	0.008	0.017	0.011
7. 205–225	0.172	0.320	0.508	0.010	0.020	0.013
8. 211–240	0.223	0.215	0.562	0.015	0.031	0.016
9. 221–240	0.159	0.386	0.455	0.035	0.087	0.052
10. 205–240[c]	0.154	0.357	0.489	0.013	0.026	0.017
11. 208–240[c]	0.201	0.255	0.554	0.013	0.028	0.016
12. 191–200	0.088	0.470	0.442	0.042	0.058	0.019
13. 201–210	0.178	0.350	0.473	0.017	0.015	0.015
14. 211–220	0.207	0.237	0.556	0.015	0.027	0.013
15. 221–230	0.211	0.266	0.524	0.043	0.105	0.062
16. 231–240	0.258	−0.047	0.789	0.034	0.110	0.077

[a] Adams *et al.* (1970). [b] Models formed from equations at 1-nm intervals. [c] Regression model formed from wavelengths at the irregular intervals reported by Greenfield and Fasman (1969).

TABLE V: Calculated Structure Fractions for Lactate Dehydrogenase.

Region (nm)	Fraction α	Fraction β	Fraction RC	σ_α	σ_β	σ_{RC}
1. X-ray[a]	0.45	0.20–0.25	0.30–0.35			
2. 191–240[b]	0.217	0.538	0.245	0.017	0.023	0.010
3. 191–230	0.217	0.538	0.245	0.019	0.027	0.012
4. 191–220	0.232	0.523	0.245	0.023	0.031	0.013
5. 200–240	0.185	0.496	0.319	0.012	0.016	0.010
6. 205–240	0.221	0.386	0.393	0.009	0.019	0.012
7. 205–225	0.213	0.392	0.395	0.011	0.022	0.014
8. 211–240	0.314	0.191	0.495	0.012	0.024	0.013
9. 221–240	0.303	0.230	0.468	0.023	0.058	0.035
10. 205–240[c]	0.201	0.422	0.377	0.017	0.033	0.021
11. 208–240[c]	0.258	0.299	0.444	0.018	0.037	0.021
12. 191–200	0.234	0.545	0.221	0.012	0.017	0.006
13. 201–210	0.230	0.476	0.294	0.017	0.015	0.016
14. 211–220	0.253	0.294	0.453	0.021	0.039	0.018
15. 221–230	0.273	0.301	0.428	0.009	0.023	0.014
16. 231–240	0.377	−0.009	0.632	0.071	0.228	0.159

[a] Rossman *et al.* (1972). [b] Models formed from equations at 1-nm intervals. [c] Regression model formed from wavelengths at the irregular intervals reported by Greenfield and Fasman (1969).

205 nm is given in line 10. Among the five proteins examined, four showed β-structure changes of 9–13% when this single additional equation was included, ribonuclease excepted. In three of these four proteins, standard deviations could not account for the β-structure variation, myoglobin excluded. We also have compared the model in line 10 (14 equations at irregular wavelength intervals) with a model over the same 205–240-nm region using 1-nm intervals (line 6). In all cases, the two models can be considered to calculate the same structure fractions; any variation can be accounted for by standard deviations. However, standard deviations are lower for the structure fractions calculated in line 6. Thus, the wavelengths chosen by Greenfield and Fasman (1969) describe the 205–240-nm region as well as 36 equations in this same region.

The previous results suggested that numerical values of the calculated structure fractions were wavelength dependent. Therefore, we attempted to elucidate the nature of this wavelength dependence through residual analysis (see Methods). We sought apparent trends in residual patterns and attempted to correlate them to known secondary structure elements determined by X-ray analysis. No correlation was apparent as each protein residual pattern was unique. A composite residual pattern for the residuals calculated by the 191–240-nm regression models (line 2) is given in Figure 1. Analyses for four additional proteins are also included in this figure: myoglobin from Saxena and Wetlaufer (1970), myoglobin from Chen *et al.* (1972), lysozyme from Dearborn and Wetlaufer (1970), and glyceraldehyde-3-phosphate dehydrogenase from Barela and Darnall (1971). The cyclic nature of Figure 1 cannot be explained easily and must be assumed to indicate that any necessary corrections to the general regression equation (eq 1) are complex (Draper and Smith, 1966). Increased residual scatter at lower wavelengths suggest that a weighted least squares is a necessary addition to the regression analysis (Draper and Smith, 1966). The latter observation is not unusual since spectropolarimeter signal to noise ratio decreases at lower wavelengths.

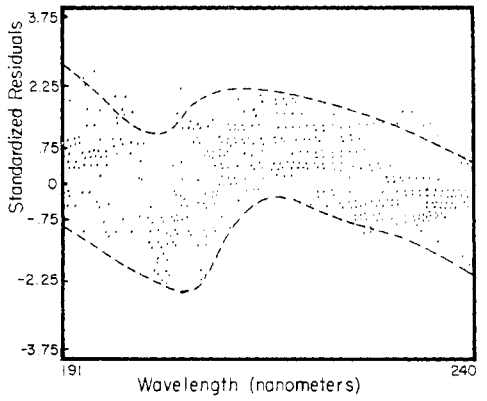

FIGURE 1: Standardized residual plot. Standardized residuals were calculated (see Methods) for nine protein spectra from the 191–240-nm regression model (except 205–240-nm regression model for glyceraldehyde-3-phosphate dehydrogenase).

Discussion

The data presented in this report were intended to clarify interpretation of protein secondary structure contents calculated using circular dichroism spectra of poly-L-lysine conformations as model spectra for similar conformations in proteins. Our studies indicated that protein secondary structure fractions calculated from circular dichroism spectra varied considerably depending upon the particular wavelength region chosen. The range for lysozyme β structure was the largest at 40.6% (Table VI), while the smallest was lactate dehydrogenase α helix at 12.9%. In most cases the standard deviations of the structure fractions were less than 4%; hence, the range of structure fractions is a better estimate of the associated errors than calculated standard deviations. When compared to X-ray determined structures, no single wavelength region gave the best correspondence, *e.g.*, for myoglobin, structure fractions calculated by the 191–240-nm regression model (Table II, line 2) correspond well with X-ray values, but for lysozyme, papain, and lactate dehydrogenase, similar correspondence is poor. Since different wavelength regions resulted in highly variable structure fractions, the 190–240-nm protein circular dichroism spectra cannot be

explained in terms of peptide bond optical activities of the α helix, β structure, and random coil of poly-L-lysine.

Several explanations for the above observations are: (1) poly-L-lysine secondary structure models may not be representative of similar structures in proteins; (2) additional secondary structures, such as 3_{10} helix, should be included in the general regression equation (eq 1); and (3) protein structures (disulfides, prosthetic groups, aromatic amino acids) other than the peptide backbone exhibit optical activity in the 190–240-nm region. Furthermore, our standardized residual plots suggest that the necessary changes may be complex. The first two possibilities have been examined extensively in previous reports (Greenfield and Fasman, 1969; Rosenkrantz and Scholtan, 1971; Magar, 1971; Chen *et al.*, 1972).

Disulfide bonds of proteins are inherently dissymmetric and may give rise to optical activity in the 190–240-nm region (Urry, 1970; Urry *et al.*, 1968; Carmack and Neubert, 1967). The ellipticities for disulfide optical activity are usually low (200–600° cm²/dmol) in relation to the peptide ellipticities below 230 nm, but between 230 and 240 nm, peptide derived optical activity decreases to near zero and, hence, disulfide optical activity may be significant. Since disulfide optical activity has not been included in the regression model, the regression model has difficulty (negative structure fractions, high standard deviations) accounting for the added protein ellipticity. Since lysozyme, ribonuclease, and papain all contain disulfide bonds (Canfield and Liv, 1965; Smyth *et al.*, 1963; Drenth *et al.*, 1968), the lack of a disulfide term in the general regression equation may account for the negative structure fractions calculated using the 221–230- and 231–240-nm regression models. The screw sense of disulfide bonds determines the sign of the resulting Cotton effect and the dihedral angle determines the magnitude of the optical activity (Urry, 1970). Hence, additional terms to describe disulfide contributions must consider both dihedral angle and chirality. At any particular wavelength, the possible correction term(s) for disulfide contributions would be infinite since the magnitude would necessarily be dependent upon the dihedral angle.

Disulfide optical activity alone cannot account for all the difficulties encountered in the 220–240-nm region since negative structure fractions were calculated using the 221–230 and/or 231–240-nm regression models for myoglobin and lactate dehydrogenase, proteins which do not contain disulfide bonds (Kendrew *et al.*, 1960; Adams *et al.*, 1970). Aromatic amino acid circular dichroism between 220 and 250 nm has been demonstrated for poly-L-tyrosine (Beychok and Fasman, 1964; Simons and Glazer, 1967), oxytocin (Simons and Glazer, 1967), immunoglobulin-G (Dorrington and Smith, 1972), and several proteins (Beychok, 1966, 1968). In addition, the amino acids tryptophan, phenylalanine, and tyrosine all exhibit optical activity in the 200–300-nm region of the spectrum (Beychok, 1967). Since all proteins mentioned in this study contain aromatic amino acids, it is probable that side chain optical activity may contribute considerably to the negative structure fractions calculated particularly when greater proportions of wavelengths above 220 nm were included in regression models. Aromatic amino acid optical activity is dependent upon solvent, local environment, and pH (Fasman and Potter, 1967). Hence, additional terms representative of aromatic amino acid optical activity will also be numerous.

In the case of proteins, such as myoglobin, which have prosthetic groups which absorb in the ultraviolet region of the spectrum, there is an added complication in that these transitions may be optically active and introduce additional errors in the calculation of secondary structures. Urry (1967) showed very clearly that there are optically active heme transitions in the 200–250-nm region in the heme undecapeptide from cytochrome *c*.

TABLE VI: Ranges of Calculated Secondary Structure Fractions.[a]

Protein	α Helix	β Structure	Random Coil
Lysozyme	0.132	0.406	0.290
Myoglobin	0.181	0.255	0.152
Ribonuclease	0.182	0.315	0.132
Lactate dehydrogenase	0.129	0.354	0.247
Papain	0.135	0.255	0.120

[a] Range is defined as the difference between the largest value for a structure fraction calculated from any regression model and the smallest value from any regression model. Regression models in which any negative structure fractions were calculated are not included. Summary from Tables I–V.

As indicated above, the variations between structure fractions calculated by different wavelength regions between proteins did not depict any consistent patterns, *i.e.*, the same wavelength region would not give the "best fit" for all proteins. An analysis of residuals was intended to clarify the nature of the corrections to the general regression equation (eq 1). Figure 1 indicates that the corrections, other than the need of a weighted least squares, are probably complex. We have examined over 150 residual plots with nine protein spectra and found that almost all spectra were unique in some wavelength region. These studies further support the conclusions based on variations of the secondary structure fractions and associated error terms.

Several recent studies (Dearborn and Wetlaufer, 1970; Chen *et al.*, 1972) have recognized that poly-L-lysine model spectra may not be representative of similar protein conformation spectra and, thus, have calculated new conformation model spectra based on X-ray determined structure fractions. These new conformation spectra were considerably different from corresponding poly-L-lysine spectra and also dependent on the combinations of proteins used (Chen *et al.*, 1972). Side chain optical activity was not represented in these calculations. Based on our reported difficulties associated with the 220–240-nm region and arguments presented above, side chain optical activity may be the cause for variations seen in these studies also. The fact that peptide structures are not as optically active in the 231–240-nm region as at lower wavelength also complicates the situation.

Conclusion

Poly-L-lysine conformational circular dichroism spectra alone (or possibly not at all) cannot be used to calculate protein secondary structure contents since any change in the wavelengths comprising the regression model may result in up to 40% change in an estimated secondary structure. *Under circumstances where X-ray structures are not known, this large error term would render the calculated structure fractions almost meaningless.* Although the preceding statement may seem pessimistic in view of the fact that CD measurements have yielded satisfactory estimates of secondary structure for certain proteins, the fact remains that different

amounts of secondary structures can be calculated for these same proteins depending upon the wavelength regions used in the analysis. We therefore suggest that a great deal of caution needs to be exercised before an interpretation of the CD spectrum of an unknown protein is made. At the very least, the rather common practice of calculating the amount of α helix in a protein from ellipticity measurements at 222 nm should be discontinued.

Acknowledgments

We thank Kayreen Barela, mathematician, White Sands Missile Range, N. M., for assistance in computer programming and design of regression models, Dr. Gerald Fasman for sending the ellipticity values for myoglobin, and Dr. J. T. Yang for allowing us to use his ellipticity values prior to publication.

References

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

Barela, T. D., and Darnall, D. W. (1971), *Biochim. Biophys. Acta 236*, 593.

Beychok, S. (1966), *Science 154*, 1288.

Beychok, S. (1967), *in* Poly α-Amino Acids, Fasman, G. D., Ed., New York, N. Y., Academic press.

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

Beychok, S., and Fasman, G. D. (1964), *Biochemistry 3*, 1675.

Breslow, E., Beychok, S., Hardman, G. D., and Gurd, F. N. (1965), *J. Biol. Chem. 240*, 304.

Canfield, R., and Liv, A. K. (1965), *J. Biol. Chem. 240*, 1977.

Carmack, M., and Neubert, L. A. (1967), *J. Amer. Chem. Soc. 89*, 7134.

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

Dearborn, D. G., and Wetlaufer, D. B. (1970), *Biochem. Biophys. Res. Commun. 39*, 314.

Dorrington, K. J., and Smith, B. R. (1972), *Biochim. Biophys. Acta 263*, 70.

Draper, N. R., and Smith, H. (1966), Applied Regression Analysis, New York, N. Y., Wiley.

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

Fasman, G. D., and Potter, J. (1967), *Biochem. Biophys. Res. Commun. 27*, 209.

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

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

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

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

Li, L. K., and Spector, A. (1969), *J. Amer. Chem. Soc. 91*, 220.

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

Magar, M. E. (1971), *J. Theor. Biol. 33*, 105.

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

Natrella, N. G. (1963), *Nat. Bur. Stand. (U. S.), Handbk. No. 91*, 6-1.

Phillips, D. C. (1967), *Proc. Nat. Acad. Sci. U. S. 57*, 484.

Quadrifoglio, F., and Urry, D. W. (1968), *J. Amer. Chem. Soc. 90*, 2760.

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

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

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

Simons, N. S., and Glazer, A. N. (1967), *J. Amer. Chem. Soc. 89*, 5040.

Smyth, D. G., Stein, W. H., and Moore, S. (1963), *J. Biol. Chem. 238*, 227.

Tinoco, I. J., Woody, R. N., and Bradley, D. F. (1963), *J. Chem. Phys. 38*, 1317.

Tooney, N. M., and Fasman, G. D. (1968), *J. Mol. Biol. 36*, 355.

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

Urry, D. W. (1967), *J. Biol. Chem. 242*, 4441.

Urry, D. W. (1970), Spectroscopic Approaches to Biomolecular Conformation, Chicago, Ill., American Medical Association.

Urry, D. W., Quadrifoglio, F., Walter, F., and Schwartz, I. L. (1968), *Proc. Nat. Acad. Sci. U. S. 60*, 967.

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

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