γ_1C-4 have no γ_2 analogues. Together these fragments account for ~180 residues and probably originate from the Fc region of the γ_1 chain. Fragment γ_1C-3 was the only γ_1 fragment bearing sugar residues and these are probably located near residue N-293 in the Fc region where an oligosaccharide is attached to the γ_2 chain (Tracey and Cebra, 1974).

Amino acid composition data (Leslie and Cohen, 1970) of H chains suggest that the γ_1 chain has one more methionyl residue than the γ_2 chain. This finding is supported by our isolation of nine CNBr fragments from γ_1 chain, whereas eight fragments account for the γ_2 chain. Thus it is likely that our nine fragments do account for the entirety of the γ_1 chain, as the γ_2 fragments have been rigorously aligned and shown to account for the whole heavy chain (Benjamin et al., 1972).

The probable alignment of the γ_1 chain CNBr fragments and the arrangement of disulfide bonds within the IgG1 molecule have almost been completed (S. Graziano, S. Liu, D. Tracey, unpublished data). With the framework of the IgG1 molecule thus set out, we have begun direct comparisons of the V regions of IgG1 and IgG2 antibodies of the same ligand binding specificity using selective modification by affinity labels and primary structural analysis (T. Trischmann, M. Ricardo, S. Liu, and D. Tracey, unpublished observations).

References

Ambler, R. P. (1965), *Biochem. J. 96,* 320.

Asherson, G. L., and Stone, S. H. (1965), *Immunology 9,* 205.

Benjamin, D. C., Hussain, Q. Z., and Cebra, J. J. (1972), *Biochemistry 11,* 3641.

Benacerraf, B., Ovary, Z., Bloch, K. J., and Franklin, E. C. (1963), *J. Exp. Med. 117,* 937.

Birshtein, B. K., and Cebra, J. J. (1971), *Biochemistry 10,* 4930.

Birshtein, B. K., Turner, K. J., and Cebra, J. J. (1971), *Biochemistry 10,* 1.

Cebra, J. J., Koo, P. H., and Ray, A. (1974), *Science 186,* 263.

Cebra, J. J., Ray, A., Benjamin, D., and Birshtein, B. (1971), *Prog. Immunol. 1,* 269.

Cunningham, B. A., Pflumm, M. N., Rutishauser, U., and Edelman, G. M. (1969), *Proc. Natl. Acad. Sci. U.S.A. 64,* 997.

Dreyer, W. J., and Bennett, J. C. (1965), *Proc. Natl. Acad. Sci. U.S.A. 54,* 864.

Eisen, H. N. (1964), *Methods Med. Res. 10,* 95.

Kinard, F. E. (1957), *Rev. Sci. Instrum. 28,* 293.

Lamm, M. E. (1969), *Immunochemistry 6,* 235.

Lamm, M. E., Lisowska-Bernstein, B., and Nussenzweig, V. (1967), *Biochemistry 6,* 2819.

Leslie, R. G. Q., and Cohen, S. (1970), *Biochem. J. 120,* 787.

Leslie, R. G. Q., Melamed, M. D., and Cohen, S. (1971), *Biochem. J. 121,* 829.

Liu, S. H., Koo, P. H., and Cebra, J. J. (1974), *J. Immunol. 113,* 677.

Melamed, M. D. (1973), *Immunochemistry 9,* 1233.

Oettgen, H. F., Binaghi, R. A., and Benacerraf, B. (1965), *Proc. Soc. Exp. Biol. Med. 118,* 336.

Oliveira, B., Osler, A. G., Siraganian, R. P., and Sandberg, A. L. (1970), *J. Immunol. 104,* 320.

Press, E. M., and Hogg, N. M. (1969), *Nature (London) 223,* 807.

Ray, A., and Cebra, J. J. (1972), *Biochemistry 11,* 3647.

Tracey, D. E., and Cebra, J. J. (1974) *Biochemistry 13,* 4796.

Trischmann, T., and Cebra, J. J. (1974), *Biochemistry 13,* 4804.

Turner, K. J., and Cebra, J. J. (1971), *Biochemistry 10,* 9.

White, R. G., Jenkins, G. C., and Wilkinson, P. C. (1963), *Int. Arch. Allergy Appl. Immunol. 22,* 156.

On the Analysis of Circular Dichroic Spectra of Proteins†

Carl. C. Baker‡ and Irvin Isenberg*

ABSTRACT: A new method is presented for analyzing circular dichroism spectra. The method employs integrals over the data and calculates the α-helical, β-sheet, and random coil content of the proteins from such integrals. It is shown that the analyzed α-helical content is usually reliable to within 5%, β-sheet values are somewhat less reliable, and random coil values are least reliable. Curve fitting techniques are shown to be misleading. The method has a number of advantages over existing procedures.

Circular dichroism (CD) is widely used to obtain information on the secondary structure of proteins and nucleic acids in solution. The technique is sensitive, and only dilute solutions are needed. The raw CD data themselves, however, are not of direct concern to an investigator. What is of interest is the content of various types of secondary structure.

It has become customary to divide CD spectra into three types—α helical, β sheet, and random coil. Random coil is a miscellaneous category which bears no necessary relationship to a hydrodynamic random coil, nor does it necessarily

† From the Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331. *Received July 25, 1975.* Supported by U.S. Public Health Service Grant CA-10872.

‡ Present address: The Rockefeller University, New York, New York.

mean that there is no short-range order in the protein when it is in such a conformation.

The determination of the fraction of protein in the various forms is by no means trivial, and several methods have been used to make such a resolution. Greenfield and Fasman (1969) used the CD spectra of poly(L-lysine) in the α-helix, β-sheet, and random coil conformations as reference spectra and estimated the amount of α helix in proteins from the ellipticity at 208 nm. The experimental curve was then compared with synthetic curves containing that amount of α helix and different amounts of β sheet and random coil. Greenfield and Fasman also used a method which minimized the variance between the experimental CD and linear combinations of poly(L-lysine) basis spectra. The condition was imposed that the sum of the fractions of each conformation must be unity and the criterion for a good analysis was a satisfactory comparison with the known X-ray structure.

Johnson et al. (1972) and Hanlon et al. (1974) fit the CD of chromatin with A, B, and C type DNA. One variable was eliminated by setting the sum equal to unity, and simultaneous equations were solved at pairs of wavelengths and averaged. Linear least-squares regression was also used over the whole range and the fractions of all components were forced to sum to unity. In later work (Hanlon et al., 1975) this last condition was dropped. This was a marked improvement in the method since the unity sum requirement could then be used as a criterion for the validity of an analysis, rather than an assumption.

Rather than using model polypeptide reference spectra, Chen et al. (1972, 1974) combined x ray and CD data for five proteins and, using a least-squares method, computed α-helix, β-sheet, and random coil reference spectra for that set of proteins. They also resolved the α-helix CD spectrum into two Gaussian bands and included chain length dependence factors for the helix. These CD reference spectra, along with optical rotatory dispersion (ORD) data and the restriction of constant rotational strength across a helical band were combined to analyze spectra for the various components.

All of these methods have one problem in common: the choice of valid reference spectra. Such a choice is by no means obvious. The rotational strengths of the α-helix transitions are chain length dependent (Woody and Tinoco, 1967); the CD spectra of various homopolymer α helices differ (Chen et al., 1972) and the CD spectra of β sheet are dependent on the chain length and number of strands (Woody, 1969). Furthermore, various random coils seem to be quite different (Chen et al., 1974) and, in addition, most globular proteins may have considerable amounts of β turns (Crawford et al., 1973).

In this paper we present a method for analyzing CD spectra which appears to have a number of advantages over existing techniques. Although it still suffers from the necessity of choosing reference spectra, it provides criteria for discriminating between them.

It should be emphasized that this method is not a curve-fitting procedure. Instead of curve fitting, it substitutes certain integrals over the data for the data themselves. In this it resembles the method of moments (Isenberg and Dyson, 1969; Isenberg et al., 1973; Isenberg, 1973), but the integrals are different. From the integrals one calculates the parameters of interest. We will discuss why we feel this method is superior to curve fitting and shall give examples showing that curve fitting can be misleading.

Theory

The following formalism can be used to resolve any curve which is a linear combination of n linearly independent reference curves. Let Ψ be the observed spectrum, $\psi_1, \psi_2, \ldots, \psi_n$ be the n reference spectra of which it is composed, and $A_1, A_2, \ldots A_n$ be the absolute amounts of each reference spectrum. Then

$$\Psi = \sum_{i=1}^{n} A_i \psi_i \tag{1}$$

We now define n functions, $\phi_1, \phi_2, \ldots \phi_n$, as linear combinations of the ψ_i's:

$$\begin{aligned} \phi_1 &= \sum_{i=1}^{n} \xi_{1i}\psi_i \\ \phi_2 &= \sum_{i=1}^{n} \xi_{2i}\psi_i \\ \phi_n &= \sum_{i=1}^{n} \xi_{ni}\psi_i \end{aligned} \tag{2}$$

and determine the ξ_{ij} by the conditions

$$\int_{\lambda_1}^{\lambda_2} \phi_i \psi_j d\lambda = \delta_{ij} \tag{3}$$

where

$$\begin{aligned} \delta_{ij} &= 1 \qquad \text{if } i = j \\ \delta_{ij} &= 0 \qquad \text{if } i \neq j \end{aligned}$$

For any ϕ_i this condition yields n linear equations in the n unknowns $\xi_{i1}, \xi_{i2}, \ldots \xi_{in}$, and ϕ_i is thereby determined.

Given ϕ_i and any experimental spectrum Ψ, A_i may be easily determined by

$$A_i = \int_{\lambda_1}^{\lambda_2} \phi_i \Psi d\lambda \tag{4}$$

For reasons that will become clear in the next section, the functions ϕ_i may be called reciprocal functions.

The most obvious advantage of the method is that it uses only integrals of experimental data so that random noise tends to be averaged away. However, there are additional advantages since there arise, naturally, a number of tests for a proper choice of reference spectra. The sum of the amplitudes, A_i, must equal unity. The values of A_i should be independent of the choice of integration range. No A_i should be negative. If n reference spectra are present, an analysis for $n + 1$ reference spectra should give a zero amplitude for the extra spectrum.

As will be seen, these conditions are stringent. They are sufficient, in themselves, to eliminate many choices of sets of reference spectra.

In addition to these advantages, there are others that are less obvious. It turns out, for example, that the calculated α-helical content is relatively independent of the reference spectra chosen and, in such cases, one may accept these values with relatively high confidence. Furthermore, as will be seen, the method permits one to take an overall view of the present state of CD and the weaknesses as well as the strengths of the technique become apparent.

An understanding of the method is greatly aided by a vector interpretation of the formalism.

Vector Interpretation of the Analysis. Consider a space of n dimensions which is spanned by n vectors, which we label $\vec{\psi}_1, \vec{\psi}_2, \ldots \vec{\psi}_n$. We define the length of a vector by

Table I: Synthetic Data using Poly(L-lysine) Basis Spectra[a] ($A_\alpha = 0.385, A_\beta = 0.255, A_R = 0.36$).

Ref Spectra	Range (nm)	% α	% β	% R	Total %
α,[b] β,[b] R[b]	204–244	38.5	25.5	36.0	100.0
	208–244	38.5	25.5	36.0	100.0
α,[b] β[b]	204–244	49.7	−8.1		41.5
	206–246	46.9	−1.4		45.6
	208–244	43.6	6.3		49.9
α,[b] R[b]	204–244	47.0		12.0	59.0
	206–246	47.0		10.6	57.6
	208–244	46.8		9.3	56.1
α,[b] R[c]	202–246	46.1		22.2	68.4
	204–244	46.2		19.3	65.5
	206–246	46.2		15.1	61.3
α,[b] β,[b] R[c]	204–244	38.5	19.3	49.1	107
	206–246	38.1	20.1	44.5	103
	208–248	37.9	20.7	37.9	98.5
α,[c] β,[c] R[c]	204–244	51.9	20.9	46.1	119
	206–246	51.9	21.1	41.1	114
	208–248	51.8	20.0	28.4	100

[a] Greenfield and Fasman (1969). [b] Poly(L-lysine). [c] From five globular proteins (Chen et al., 1974).

$$|\vec{\psi}_i| = \left(\int_{\lambda_1}^{\lambda_2} \psi_i^2 d\lambda\right)^{1/2} \quad (5)$$

and the angle between two vectors by

$$\cos\theta = \left(\int_{\lambda_1}^{\lambda_2} \psi_i \psi_j d\lambda\right) \Big/ |\vec{\psi}_i||\vec{\psi}_j| \quad (6)$$

The set of vectors will usually be nonorthogonal. We now define a set of reciprocal vectors, $\vec{\phi}_1, \vec{\phi}_2, \ldots \vec{\phi}_n$, by

$$\vec{\phi}_i \cdot \vec{\psi}_j = \delta_{ij} \quad (7)$$

It may be seen that $\vec{\phi}_1$ is normal to the set $\vec{\psi}_2, \vec{\psi}_3, \ldots \vec{\psi}_n$ and $\vec{\phi}_j$ is normal to $\vec{\psi}_1, \vec{\psi}_2, \ldots \vec{\psi}_{j-1}, \vec{\psi}_{j+1}, \ldots \vec{\psi}_n$.

Any CD spectrum, Ψ corresponds to a vector $\vec{\Psi}$. If

$$\vec{\Psi} = \sum_{i=1}^{n} A_i \vec{\psi}_i \quad (8)$$

then

$$\vec{\phi}_j \cdot \vec{\Psi} = A_j \quad (9)$$

Thus the amplitudes are obtained by a simple projection of the experimental data onto the appropriate reference axis.

The vector formalism makes the following considerations clear. In general, the value of any particular amplitude will depend on the choice of all of the reference spectra. However, if two vectors $\vec{\psi}_1$ and $\vec{\psi}_2$ are almost orthogonal then the determination of A_1 will be relatively insensitive to a choice of $\vec{\psi}_2$. We believe that this is one of the reasons why, as will be seen, the computed α-helical content is insensitive to the choice of a random coil reference spectrum since these two vectors are almost orthogonal (89° for the random and α-helical poly(L-lysine) spectra, for example). The other reason is that the α-helical spectrum is much larger than the random coil spectrum in the 205–240-nm range. It is for this reason that the random coil content is not independent of the choice of the α-helical reference curve.

All calculations were done on an HP9821A calculator and integration was by Simpson's rule. The reference spectra used were those of α, β, and random coil forms of poly(L-lysine) and the corresponding spectra calculated by Chen et al. (1974) from the x-ray data and CD of five proteins. In addition, we have used the spectrum of the elastin β turn (Urry et al., 1974). As will be seen, the CD spectrum of a histone in 0.001 M HCl was found to be better as a random coil spectrum for that histone than either of the above random coil spectra. The CD spectra of myoglobin, lysozyme, lactate dehydrogenase, papain, and ribonuclease came from Chen et al. (1972). The spectra of histones and histone complex were taken from papers from this laboratory (Li et al., 1972; Wickett et al., 1972; D'Anna and Isenberg, 1972, 1973, 1974a–d).

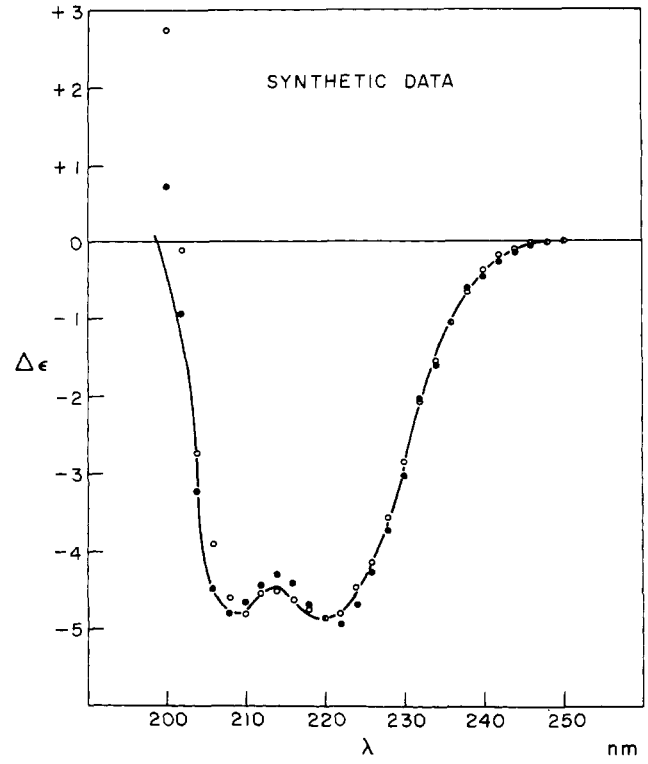

FIGURE 1: Synthetic data composed of 38.5% α helix, 25.5% β sheet, and 36% random coil using the poly(L-lysine) spectra (—); calculated spectrum from the analysis over the range 208–248 nm using the reference spectra of Chen et al. (1973) (○) and calculated from the analysis over the range 204–244 nm, using only the α helix and random coil of poly(L-lysine) (●). Note that poor parameters can still give curves that look like a good fit.

Results

Synthetic Data. Table I shows various analyses of a synthetic linear combination of the poly(L-lysine) spectra with $A_\alpha = 0.385$, $A_\beta = 0.255$, and $A_R = 0.360$. If a two-component analysis is made and either the β-sheet or the random coil vector is left out of the analysis, then the amounts of random coil or β-sheet forms are very much in error. However, the amount of α-helix changes less than 10%, even though a very serious error has been made. We note that the sum test is bad if one of the components is missing.

If the random coil spectrum of Chen et al. (1974) is used, the helix content is very close to the actual content, the β-sheet value is within 5%, and the sum comes close to 100%. Use of all three of the spectra given by Chen et al. yields an analysis in which the helical content is high by about 14% which is to be expected since this helix spectrum is less in magnitude than a long homopolymer helical spectrum. Note that when wrong reference spectra are used, a large variation in the analysis results from a small change in the integration range.

Figure 1 gives dramatic, and at first sight surprising, evi-

Table II: Myoglobin[a] (79% Helix, 0% β from X-ray Data).

Range (nm)	% Helix[b]	% β Sheet[b]	% Random	Total
200–248	83.5	1.1		84.6
200–248	82.6	7.5	17.1[b]	107
208–244	82.8	6.4	21[b]	110
200–248	83.4	7.3	9.5[c]	100
208–244	83.9	6.4	14.0[c]	104

[a] Chen et al., 1972. [b] From five globular proteins (Chen et al., 1974). [c] Poly(L-lysine) (Greenfield and Fasman, 1969).

Table III: Lysozyme[a] (41% Helix, 16% β Sheet from X Ray).

Range (nm)	% Helix[b]	% β Sheet[b]	% Random	Total
200–248	36.6	17.3	68.3[b]	122
204–244	36	18	73.5[b]	128
206–246	36.2	17.7	77.9[b]	132
200–248	39.7	18.3	41[c]	99
202–246	39.7	18.3	42.1[c]	100
204–244	39.6	18.3	45[c]	103

[a] Chen et al., 1972. [b] From five globular proteins (Chen et al., 1974). [c] Poly(L-lysine) (Greenfield and Fasman, 1969).

Table IV: Lactate Dehydrogenase[a] (45% Helix, 24% β Sheet from X Ray).

Range (nm)	% Helix[b]	% β Sheet[b]	% Random	Total
200–248	40	7.7	8.2[b]	56
208–244	40.4	7.7	11.9[b]	60
200–248	40.5	7.6	4.5[c]	53
204–244	40.6	7.8	2.7[c]	51
208–244	41.0	7.7	7.4[c]	56

[a] Chen et al., 1972. [b] From five globular proteins (Chen et al., 1974). [c] Poly(L-lysine) (Greenfield and Fasman, 1969).

Table V: Papain[a] (28% Helix, 18% β Sheet from X Ray).

Range (nm)	% Helix[b]	% β Sheet[b]	% Random	Total
200–248	29.7	7.9	51[b]	89
204–244	30	9.8	32.6[b]	72.6
208–244	30.2	10.2	37.1[b]	77.6
200–248	32	8.3	30[c]	70
204–248	31.8	9.7	19.3[c]	61
208–244	32.0	10.0	21.4	63.5

[a] Chen et al., 1972. [b] From five globular proteins (Chen et al., 1974). [c] Poly(L-lysine) (Greenfield and Fasman, 1969).

dence of how badly a curve fitting criterion may mislead the investigator. Two analyses, one moderately bad and one very poor, both yield curves with are very close to the known spectrum. As will be discussed later, curve fitting criteria, whether by a least-mean-squares method, or by visual inspection, can be grossly in error.

Globular Proteins. Tables II–VI give sample analyses for myoglobin, lysozyme, lactate dehydrogenase, papain, and ribonuclease. The helical content of myoglobin is in good agreement with x-ray data. On the other hand, the β-sheet content is overestimated. The poly(L-lysine) random coil gave a better analysis by the sum rule than did the Chen et al. random coil. Both analyses give very good visual fits (Figure 2).

Table VI: Ribonuclease[a] (19% Helix, 38% β Sheet from X Ray).

Range (nm)	% Helix[b]	% β Sheet[b]	% Random	Total
200–248	23	48	58[b]	129
204–244	23	49	48[b]	120
208–244	23.5	49.5	48.9[b]	122
200–248	26	48	34.5[b]	109
204–244	26	49	29[c]	104
204–244	26.1	49.4	31.2[c]	107

[a] Chen et al., 1972. [b] From five globular proteins (Chen et al., 1974). [c] Poly(L-lysine) (Greenfield and Fasman, 1969).

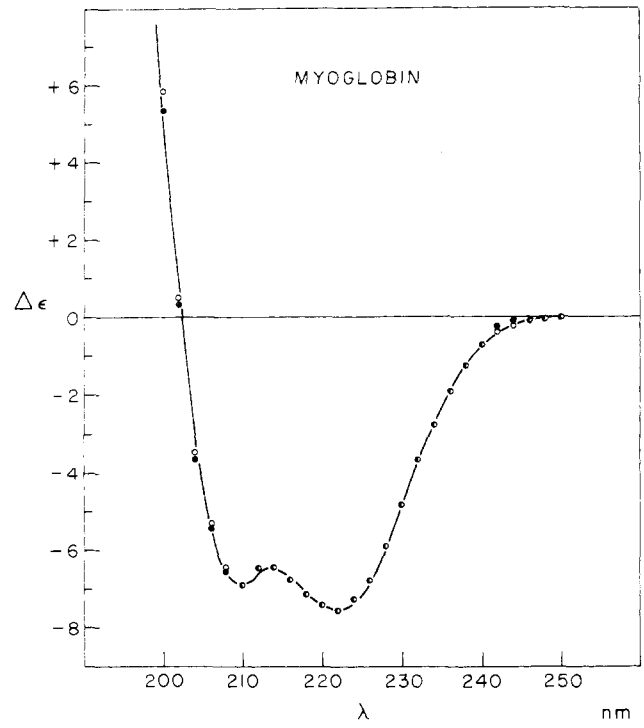

FIGURE 2: Myoglobin (—); calculated from analysis using Chen et al. basis spectra (●); using helix and β sheet of Chen et al. and random coil of poly(L-lysine) (○).

The calculated helical and β-sheet contents of lysozyme agree well with the x-ray values. The helix and β-sheet values are relatively invariant to the integration range and choice of a random basis spectra, but the random coil of poly(L-lysine) gives the best analysis. We note that the visual fit is not nearly as good as that for myoglobin (Figure 3).

Analyses of lactate dehydrogenase spectra give a helical content which is only 5% low, but the β-sheet content is grossly underestimated, the values sum to only 50–60% and the analysis is unsatisfactory. Obviously one or both of the β-sheet and random coil Chen et al. reference spectra are invalid for lactate dehydrogenase. Inclusion of the elastin β turn as a fourth basis spectrum helped the analysis somewhat (not shown), which is consistent with the high β-turn content of the dehydrogenases (R. W. Woody, private communication), but it is not valid to use both β-turn spectrum and the Chen et al. random since all β turns were included as random in the analysis for the Chen et al. basis spectra. Also the high variability of β-turn CD (R. W. Woody, private communication) makes it impossible to pick a general β-turn spectrum. We have therefore left it out of the analyses. It should be noted that lactate dehydrogenase gives a

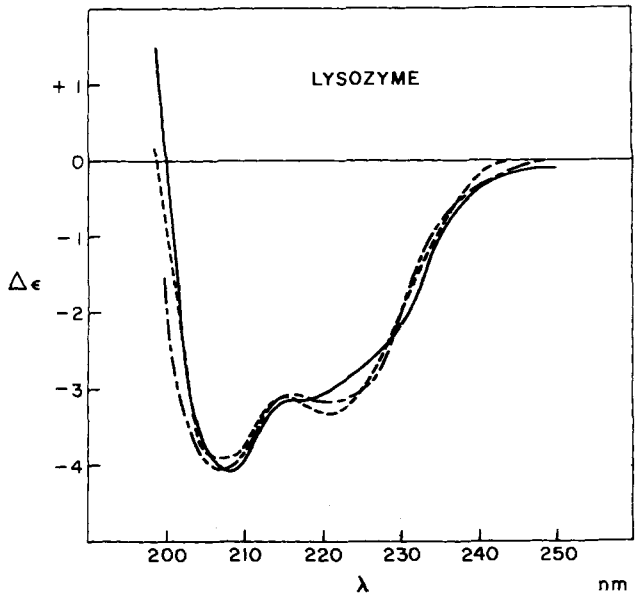

FIGURE 3: Lysozyme (—); calculated from analysis over range 206–246 nm, using Chen et al. reference spectra (— - —); from analysis over range 202–246 nm using helix and β-sheet spectra of Chen et al. and random coil of poly(L-lysine) (- - -).

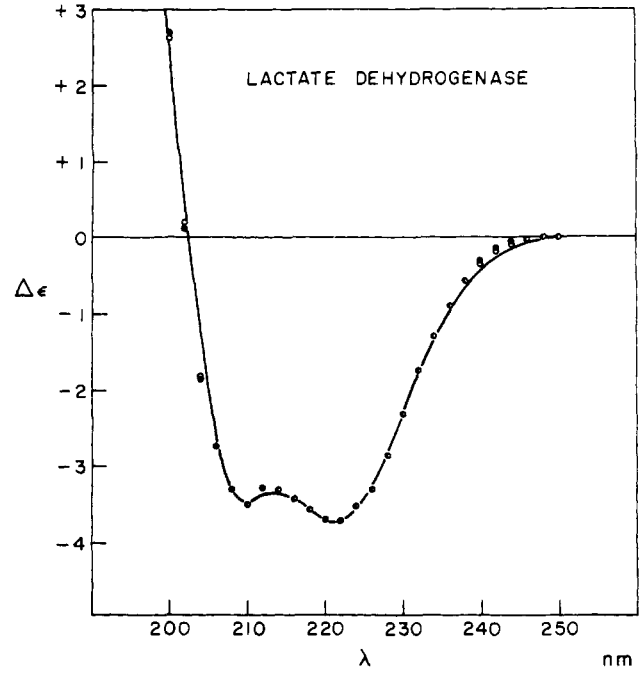

FIGURE 4: Lactate dehydrogenase (—); calculated from analysis over range 208–244 nm using Chen et al. reference spectra (O); using Chen et al. helix and β-sheet reference spectra and random coil spectrum of poly(L-lysine) (●).

very good visual fit even though the analysis is poor (Figure 4).

The papain CD analysis gave a close agreement of the helical content with x-ray values but underestimated the β-sheet content. The Chen et al. random coil gave a better analysis than the poly(L-lysine) random coil because papain dominated the analysis for the Chen et al. random spectrum.

The analysis of ribonuclease spectra overestimates the helical and β-sheet content of ribonuclease. The Chen random gives a better helical content, but the poly(L-lysine) random coil gives a better content sum.

Histone Conformations. We have reanalyzed the CD spectra of histones and histone complexes taken in a set of studies of this laboratory (Li et al., 1972; Wickett et al., 1972; D'Anna and Isenberg, 1972, 1973, 1974a–d). As with the proteins previously discussed in this paper, the analyzed values of the α-helical content were not sensitive to the choice of reference spectra.

As may be seen in Table VII, the percent α-helical contents were close to the values previously found by a difference spectra method (Li et al., 1972; D'Anna and Isenberg, 1974c).

This close agreement is fortuitous, however. Histones, in the folded but unaggregated form, have no β sheet. Furthermore, the random coil spectrum in the 217–235-nm region is small compared to the α-helical spectrum. The calculation of α-helical content is therefore insensitive to the mode of analysis. This, of course, will not be true, in general. Furthermore, even here, despite the agreement, the present method is superior to the previous analyses. Whereas previously, it was necessary to assume that the folded histone molecules had negligible β-sheet content, here we demonstrate it.

Of some interest is the influence of the choice of random coil reference spectra on the sum test. Use of the Chen et al. (1974) random coil spectrum produced unsatisfactory sum tests, with sums always above 1.4 and sometimes as high as 1.7. For the same reason, the poly(L-lysine) random coil spectrum was unsatisfactory, the sums ranging from 0.08 to 1.7. However, the use of poly(L-lysine) α-helix and β-sheet spectra plus the random spectrum of the particular histone in 0.001 M HCl always gave very good sums, and the computed parameters varied very little with changes in the chosen integration span.

Table VII: Conformational Changes of the Histones.[a]

Histone	Phosphate Concn used in Extrapolation (mM)	Fast Step % α Helix (This Method)	% α Helix (Difference Spectra Method)[b]	Sum Test
f2a2	3.3	11	13	1.1
f2b	26	14	14	0.97
f3	3.2	12.5	15	0.97
f2a1	10	16	15	0.94

[a] Poly(L-lysine) α helix and β sheet and the histone random coil in 0.001 M HCl were used for both methods of analysis. [b] D'Anna and Isenberg (1974c).

We obtained the values of Table VII by extrapolating our data to infinite salt concentrations where, presumably, all of the molecules have undergone a fast-step folding.

If the folding fits a two-state model, as it does for f2a1, (Li et al. 1972; Wickett et al., 1972), f2b (D'Anna and Isenberg, 1972), and f3 (D'Anna and Isenberg, 1974c) then the conformation of the infinite salt state can be calculated from the equilibrium constant and one CD spectrum. Let f be the fraction of molecules in state 2, ψ^∞ be the CD of state 2, ψ be the CD for a particular salt concentration, and state 1 be the random form. Then

$$\psi = A_\alpha\psi_\alpha + A_\beta\psi_\beta + A_R\psi_R$$

$$\psi^\infty = A_\alpha{}^\infty\psi_\alpha + A_\beta{}^\infty\psi_\beta + A_R{}^\infty\psi_R$$

$$\frac{1}{f} = 1 + \frac{1}{K[\text{salt}]}$$

Table VIII: Histone Complexes.[a]

Complex	% α Helix	% β Sheet	Sum Test
f3–f2a1	18.1 (17.1)[b]	0 (0)[b]	0.90
f2a2–f2b	22.7 (18.6)[c]	0 (0)[b]	0.84
f2b–f2a1	21 (17.2)[d]	0 (0)[b]	0.93

[a] All analyses used α helix and β sheet of poly(L-lysine) and the random form of the histone mixture in HCl. [b] D'Anna and Isenberg (1974d). [c] D'Anna and Isenberg (1974b). [d] D'Anna and Isenberg (1973).

and

$$\psi = f\psi^{\infty} + (1 - f)\psi_R$$

Therefore

$$A_{\alpha}^{\infty} = A_{\alpha}/f$$

and

$$A_{\beta}^{\infty} = A_{\beta}/f$$

Histone f2a2 does not satisfy a two-state model (D'Anna and Isenberg, 1974a); for this histone values of A_{α}^{∞} and A_{β}^{∞} were estimated by assuming an effective equilibrium constant of 526 M^{-1} equal to the inverse concentration for which $f = 0.5$ (D'Anna and Isenberg, 1974c).

The CD spectra of the strong histone complexes were taken under conditions where almost all of the histone molecules were complexed; no extrapolations are needed for these.

Table VIII compares the α-helix and β-sheet content for the histone complexes calculated by this method and by the difference spectra method. Most of the values agree with one another within experimental error.

Discussion

Curve fitting is widely used to analyze data obtained by many techniques. A fit is usually made either by intuitive visual criteria or by some form of least-mean-squares fitting. However, such curve-fitting methods involve a set of important assumptions, which are usually unstated. The most important assumption is that random errors determine the accuracy of the data and are much more important than consistent errors. If this is not true, if consistent errors dominate, then least-mean-squares fitting will not determine the significant parameters in the data. The second important assumption is that if the analyzed parameters yield curves that "look good" then the parameters are believable.

Neither assumption is valid for analyzing CD data. The most important source of error is that we do not know which reference spectra to choose. This dominates the problem of analysis and this leads to consistent error, not random error. Furthermore, as shown in many of the figures, poor parameters can still yield curves that "look good". Such results remind one of fluorescence time decay analyses where again it is the consistent errors, and not the random errors, which dominate and limit analyses (Isenberg et al., 1973; Isenberg, 1973; Small and Isenberg, 1976).

The present method does not curve fit data. Instead it provides a set of criteria for making some sort of judgment on the trustworthiness of an analysis. It may be seen that the analyzed α-helical content is fairly reliable, β sheet somewhat less reliable, and the random coil content least reliable.

The work presented here shows that there is something important still missing in our understanding of CD data. It may be that our knowledge of β turns is still too scanty, or perhaps important interactions *between* segments of defined secondary structure should be considered. Consider lactate dehydrogenase, for example. The analyses are very poor (Table IV) but we do not know why. Nevertheless it is perhaps encouraging that the overwhelming majority of analyses for α-helical content are excellent. As a good rule of thumb, we may apparently believe such values to within something like 5%.

References

Chen, Y., Yang, J. T., and Chan, K. H. (1974), *Biochemistry 13,* 3350.

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

Crawford, J. L., Lipscomb, W. N., and Schellman, C. G. (1973), *Proc. Natl. Acad. Sci. U.S.A. 70,* 5381.

D'Anna, Jr., J. A., and Isenberg, I. (1972), *Biochemistry 11,* 4017.

D'Anna, Jr., J. A., and Isenberg, I. (1973), *Biochemistry 12,* 1035.

D'Anna, Jr., J. A., and Isenberg, I. (1974a), *Biochemistry 13,* 2093.

D'Anna, Jr., J. A., and Isenberg, I. (1974b), *Biochemistry 13,* 2098.

D'Anna, Jr., J. A., and Isenberg, I. (1974c), *Biochemistry 13,* 4987.

D'Anna, Jr., J. A., and Isenberg, I. (1974d), *Biochemistry 13,* 4992.

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

Hanlon, S., Brudno, S., Wu, T. T., and Wolf, B. (1975), *Biochemistry 14,* 1648.

Hanlon, S., Johnson, R. S., and Chan, A. (1974), *Biochemistry 13,* 3963.

Isenberg, I. (1973), *J. Chem. Phys. 59,* 5696.

Isenberg, I., and Dyson, R. D. (1969), *Biophys. J. 9,* 1337.

Isenberg, I., Dyson, R. D., and Hanson, R. (1973), *Biophys. J. 13,* 1090.

Johnson, R. S., Chan, A., and Hanlon, S. (1972), *Biochemistry 11,* 4347.

Li, H. J., Wickett, R., Craig, A. M., and Isenberg, I. (1972), *Biochemistry 11,* 2952.

Small, E., and Isenberg, I. (1976), *Biopolymers* (in press).

Urry, D. W., Long, M. M., Ohnishi, T., and Jacobs, M. (1974), *Biochem. Biophys. Res. Commun. 61,* 1427.

Wickett, R., Li, H. J., and Isenberg, I. (1972), *Biochemistry 11,* 2952.

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

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