Biophysical Chemistry 7 (1977) 41–49
© North-Holland Publishing Company

CIRCULAR DICHROISM OF INSULIN A CHAIN IN WATER: TRIFLUOROETHANOL MIXTURES. USE OF LINEAR AND NONLINEAR LEAST SQUARES ANALYSIS

Leslie A. HOLLADAY
Department of Biochemistry, Vanderbilt University, Nashville, Tennessee 37232, USA

Received 8 March 1977

Circular dichroic spectra obtained for S-carboxymethylated insulin A chain in water : trifluoroethanol mixtures show that a marked conformational transition occurs as the concentration of trifluoroethanol is increased to (v/v) 83% to give a polypeptide containing about 43% of the residues in the α-helical conformation. Several proposed methods of analysis, including both linear and two nonlinear least squares methods, were unable to quantitate the amount of β-structure present in the polypeptide in 83% trifluoroethanol. Examination of the methods of analysis lead to the conclusion that the current models for far UV CD analysis are not adequare for the data obtained in this study and shows that nonlinear least squares procedures may lead to erroneous conclusions.

1. Introduction

The analysis of the far UV CD spectra of globular proteins has been the subject of many investigations which have sought to improve the reliability of the estimates obtained for the secondary structure content. Of much current interest are the structure-conformation-function relationships of small polypeptide hormones. Studies which seek to determine the degree and stability of secondary structure content of small polypeptides may yield useful information on the relationship of conformation and biological activity. In this context this study seeks to explore the usefulness of proposed methods of analysis for the estimation of the secondary structure content of small polypeptides.

A recent report suggests that the degradation of insulin by glutathione-insulin transhydrogenase of the fat cell is linked to the increase in phosphodiesterase activity observed on addition of insulin [1]. This work suggests that the separate chains of insulin may mediate the increase in phosphodiesterase activity. In view of this, it becomes of interest to examine the conformation of the insulin chains in both aqueous solution and in solvents which promote conformational transitions. Therefore, this study will explore the capacity of the A chain of insulin to assume a well defined structure in solvents which mimic a membrane environment by virtue of their hydrophobic characteristics. Such studies should lead to a better understanding of the relationship between conformation and biological activity for the various functions of the insulin molecule.

2. Materials and methods

S-carboxymethylated (S-CM) porcine insulin A chain was purchased from Schwarz/Mann (lot 22-1660). Contamination with B chain was completely absent as judged by the lack of even a trace of phenylalanine and basic amino acids after hydrolysis in 4 N methanesulfonic acid for 20 h at 110°C. Guanidinium chloride (Gdm·Cl) was Heico ultra-pure grade. 2,2,2-trifluoroethanol (TFE) was spectroscopic grade from Matheson, Coleman, and Bell. Other salts were of reagent quality. Stock solutions of A chain were prepared in 0.5 M KCl, 50 mM Tris·HCl with the pH adjusted to 7.0. Dilutions were made such that the final solutions were 0.1 M KCl, 10 mM Tris·HCl, with TFE or Gdm·Cl present. Concentrations of stock solutions were determined spectrophotometrically, with corrections for light-scattering [2], using an extinction coefficient for the tyrosyl chromophore in water of 1400 ℓ mole^{-1} cm^{-1} [2]. CD spectra were obtained with a Cary 60 spectropolarimeter equipped with a model 6002 CD attach-

ment. Pathlengths of 0.05 to 2.0 cm were used; the sample temperature was maintained at 25.0 ± 0.2°C by a circulating water bath and thermostated cell compartment.

3. Mathematical methods

The methods used in the analysis of the far UV CD data are described as follows. The basic assumptions common to these methods pertain to the equation

$$[\theta]_\lambda^{\exp} = f_\alpha [\theta]_\lambda^\alpha + f_\beta [\theta]_\lambda^\beta + f_R [\theta]_\lambda^R, \quad (1)$$

where the f's denote the fraction of a particular conformation, e.g., α-helix (α), β-structure (β), or the aperiodic conformer (R), and $[\theta]_\lambda^{\exp}$ refers to the experimental ellipticity at a wavelength λ.

One common method used to estimate f_α, f_β, and f_R from CD spectra is the linear least squares method described by Chen et al. [3–5] In this approach the value of f_R is determined as $1-f_\alpha-f_\beta$. A set of reference spectra ($[\theta]_\lambda^\alpha$, $[\theta]_\lambda^\beta$, $[\theta]_\lambda^R$) is chosen and a linear least squares fit of eq. (1) is used. In our implementation, the sum of f_α, f_β, f_R was constrained to be unity, but f_α and f_β were not required to be $\geqslant 0$.

As a check on this method of analysis, a completely unconstrained least squares fit may also be computed. In this case, a completely suitable reference spectra set would return a reasonable fit to the data along with a sum for f_α, f_β, and f_R close to unity.

The above two methods may be used to estimate the secondary structure content of proteins under a defined set of solution conditions. Frequently it is possible to obtain a series of far UV CD spectra for a polypeptide by varying the temperature and solvent, such that a family of curves are obtained. For such a family, eq. (1) is still considered to be applicable for each curve in the set. Markussen and Vǿlund [6] pointed out that in such a situation it becomes possible to fit the entire family of spectra to eq. (1), using the criterion of least squares. With this approach it becomes possible to solve also for the estimated random coil spectrum.

This approach arises from the hypothesis that while the reference spectra for the completely α-helical and completely β-sheet forms may be regarded as accurate, the reference spectra for the aperiodic form is uncertain. This uncertainty arises from the concept that differing proteins may possess differing aperiodic CD spectra due to variability in disulfide content, aromatic contributions to the far UV CD spectrum, and amino acid sequence.

Thus, for the nonlinear approach, $[\theta]_\lambda^R$, in eq. (1) is regarded as part of the parameter set to be estimated. Their method used the NILES procedure described by Wold [7]. In this algorithm, an initial guess is used for $[\theta]_\lambda^R$, and a series of successive linear least squares steps leads to the final estimates for f_α, f_β, f_R, and $[\theta]_\lambda^R$. Convergence proofs for this algorithm have not been presented, and only empirical evidence exists that convergence occurs. The sum $f_\alpha + f_\beta + f_R$ is constrained to be one.

The fourth method is the nonlinear least squares algorithm of Marquardt [8,9] as implemented in the XZSSQ routine of the International Mathematical and Statistical Library [10]. This implementation uses a forward-finite difference method for estimation of the required partial derivatives. Convergence was considered obtained when the residual sum of squares for two successive iterations had a relative difference less than 10^{-6} to 10^{-4}. Restrictive convergence criteria (10^{-6}) were employed when the initial guess for the parameters was such that the value for the Marquardt parameter became very large, leading to extremely short steps in the first few iterations.

The equation used for the Marquardt method was modified to constrain $f_\alpha, f_\beta \geqslant 0$. The equation fitted was

$$[\theta]_\lambda^{\exp} = (f'_\alpha)^2 [\theta]_\lambda^\alpha + (f'_\beta)^2 [\theta]_\lambda^\beta + [1-(f'_\alpha)^2-(f'_\beta)^2][\theta]_\lambda^R . \quad (2)$$

Upon convergence, f_α for a particular spectrum is given by $(f'_\alpha)^2$; f_β by $(f'_\beta)^2$. This process ensures positive values for f_α and f_β.

Approximate standard errors for the parameter estimates were computed using the assumption that in the neighborhood of the solution that the function on the right side of eq. (2) may be approximated by a linear one-term Taylor series and that the variance in $[\theta]_\lambda^{\exp}$ is uncorrelated and uniform. Under these assumptions the covariance matrix is given by [11]

$$V_\theta = 2\sigma^2 N^{-1} = 2\sigma^2 \left[\sum_{\mu=1}^{n} (\partial f_\mu/\partial\theta)(\partial f_\mu/\partial\theta)^T \right]^{-1} . \quad (3)$$

In eq. (3), σ^2 is estimated from the residual sum squares, and f_μ is the function defined by the right side of eq. (2), where $[\theta]$ is the set of parameters being estimated. For the constrained equation (2) an additional problem arises in that we obtain f'_α, not f_α. In order to estimate the standard deviation for f_α, knowing an approximate standard deviation for f'_α, we assume that the variance of f_α is given by [12]

$$V(f_\alpha) = [g'(f'_\alpha)]^2 V(f'_\alpha)\,, \tag{4a}$$

$$g(f_\alpha) = (f'_\alpha)^2 \tag{4b}$$

Thus, knowing the standard deviation of f'_α and $f_{\beta'}$, it is possible to compute the standard deviation of f_α and f_β for each curve. This last approximation was checked by analyzing several data sets using an unconstrained algorithm. The approximate standard errors obtained were in good agreement.

4. Results

The CD spectra of S-CM insulin A chain are shown in fig. 1. As the concentration of TFE is increased from 0% to 83% (v/v), the polypeptide appears to undergo a considerable conformational change, with the appearance of a spectrum in 83% TFE which resembles that of a protein with about 50% α-helicity [3]. At intermediate TFE concentrations, a distinct shoulder appears at about 213 nm; this is indicative of β-structure. Near UV CD spectra were obtained in aqueous buffer, 6 M Gdm·Cl, and 80% (v/v) TFE (all solutions contained 0.1 M KCl, 10 mM Tris·HCl, pH 7). A drastic alteration in the near UV CD spectrum is seen as the solvent is changed from water to either 80% TFE or 6 M Gdm·Cl.

Given the family of far UV CD spectra in fig. 1, it is of interest to quantitate the changes in secondary

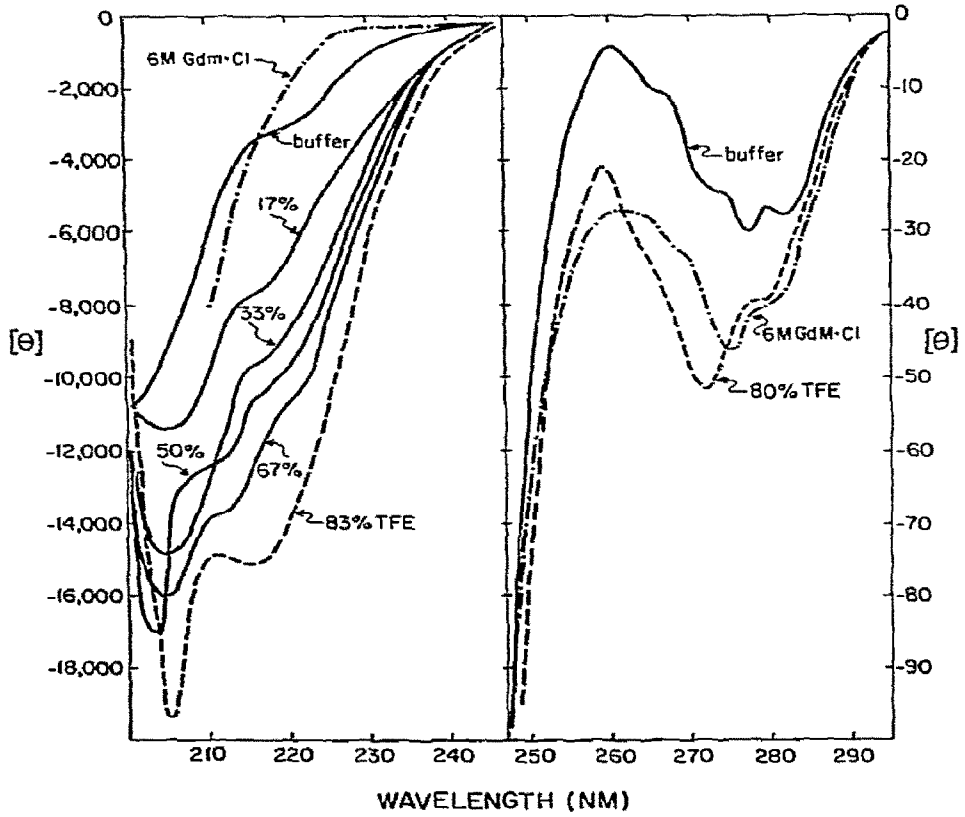

Fig. 1. CD spectra of S-CM insulin A chain. Left panel: Far UV CD spectra obtained in water, 6M Gdm·Cl, and water: TFE mixtures (%, v/v). Right panel: Near UV CD spectra obtained in water (——); in 6 M Gdm·Cl (—·—), and in 80% (v/v) TFE (– – –). All solvents contained 0.1 M KCl, 10 mM Tris·HCl, pH 7.

structure content accompanying solvent alterations. The CD data, which extend (except for the Gdm·Cl curve) with reasonable precision to 200 nm, cover a fairly wide range of ellipticity values. In attempts to analyze these data, two wavelength ranges were selected: 243–201 nm, and 243–210 nm. The former range excludes the Gdm·Cl data; the latter includes all seven curves, but excludes the 210–200 nm region where marked changes occur with increasing TFE concentrations. It has been suggested that the 210–200 nm region is unsuitable for secondary content analysis because of the large chain-length dependence of the rotational strength of the 208 nm α-helix CD band [13]. For all analyses below, the reference spectra of Chen et al. [5] were used.

Results obtained by linear least squares where f_α and f_β are not constrained to be 0 or greater) are given in table 1. For the 243–201 nm interval, the extremely poor fit, especially to the data obtained using high TFE concentrations, is apparent. In the 243–210 nm interval, the lack of fit is about twice the experimental error in $[\theta]_\lambda^{exp}$. These results suggest that the CYC basis spectra set [5] are not adequate for this polypeptide and that a good deal of the difficulty arises from the 201–210 nm region.

Table 2 gives the results obtained using unconstrained least squares. The much better fit to the data results from the deletion of the constraint that $f_\alpha + f_\beta + f_R = 1$. This method of analysis suggests that the reference spectra set [5] used are not adequate for this polypeptide, as the sum $f_\alpha + f_\beta + f_R$ is greater than 1.5 for four of the six spectra in the 243–201 nm interval. Dropping the 210–201 nm region marginally improves the agreement between $f_\alpha + f_\beta + f_R$ and 1.

Both tables 1 and 2 give results which suggest that the far UV CD spectra of S-CM insulin A chain cannot be adequately described by the CYC basis spectra set. This point leads to the hypothesis that the basis spectra for the α-helical and β-sheet conformations may be correct, but that the random coil spectrum may not be suitable for this analysis. Thus, further analysis will pool all of the data in either the 243–201 nm range or the 243–210 nm range and will solve for the estimated random coil spectrum of S-CM insulin A chain as well as for the secondary structure content of each curve subject to the constraint that $f_\alpha + f_\beta + f_R$ sum to one and that $f_\alpha, f_\beta \geq 0$.

Results obtained using the NILES algorithm proposed by Markussen and Vǿlund are given in table 3 for both wavelength ranges. An approximate two fold reduction in the sum of squared residuals is obtained for both wavelength intervals. Convergence was considered obtained when the residual sum of squares for two successive iterations had a relative difference of $\leq 10^{-5}$. It is apparent from examination of $[\theta]_{res}$ for the various curves that the fit obtained to the data for S-CM insulin A chain in 0.1 M KCl, 10 mM Tris, pH 7, is particularly poor, and that the fit becomes

Table 1
Analysis by constrained [a] least squares

(A) interval 243–201 nm							
% (v/v) TFE:	0	17	33	50	67	83	
f_α	0.10	0.23	0.32	0.34	0.38	0.47	
f_β	0.24	0.12	0.00	−0.09	−0.08	−0.14	
f_R	0.66	0.65	0.68	0.75	0.70	0.67	
$[\theta]_{res}$ [b]	598	1199	2327	2625	2917	3569	
(B) interval 243–210 nm							
% (v/v) TFE:	0	17	33	50	67	83	0 (6 M Gdm·Cl)
f_α	0.10	0.21	0.28	0.30	0.34	0.41	0.07
f_β	0.21	0.28	0.32	0.33	0.37	0.42	0.32
f_R	0.69	0.51	0.40	0.37	0.29	0.17	0.61
$[\theta]_{res}$ [b]	287	617	795	654	852	681	1019

[a] $\Sigma f_\alpha + f_\beta + f_R$ is constrained to be unity, but f_α and f_β are not constrained to be ≥ 0.
[b] $[\theta]_{res}$ is calculated as $[(\Sigma_{i=1}^{n} r_i^2)/(n-p)]^{1/2}$, where n is the number of data points (3 nm intervals), and p is 2.

Table 2
Analysis by unconstrained least squares

(A) interval 243–201 nm							
% (v/v) TFE:	0	17	33	50	67	83	
f_α	0.10	0.21	0.28	0.30	0.34	0.42	
f_β	0.23	0.31	0.39	0.35	0.43	0.47	
f_R	0.65	0.77	0.93	1.02	1.01	1.04	
Σf_i	0.98	1.30	1.60	1.69	1.79	1.93	
$[\theta]_{res}$ [a]	619	560	792	873	669	1290	
(B) interval 243–210 nm							
% (v/v) TFE:	0	17	33	50	67	83	0 (6 M Gdm·Cl)
f_α	0.10	0.21	0.28	0.30	0.34	0.41	0.07
f_β	0.21	0.30	0.37	0.36	0.42	0.44	0.38
f_R	0.63	0.84	0.98	0.81	0.86	0.46	1.32
Σf_i	0.93	1.35	1.63	1.47	1.62	1.31	1.77
$[\theta]_{res}$ [a]	291	492	336	384	488	608	536

[a] $[\theta]_{res}$ is calculated as $[(\Sigma_{i=1}^{n} r_i^2)/(n-p)]^{1/2}$, where n is the number of data points, and p is 3.

Table 3
Analysis by constrained NILES algorithm

(A) interval 243–201 nm							
% (v/v) TFE:	0	17	33	50	67	83	
f_α	0.09	0.21	0.30	0.32	0.37	0.44	
f_β	0.49	0.33	0.19	0.12	0.09	0.00	
f_R	0.42	0.46	0.52	0.56	0.54	0.55	
$[\theta]_{res}$ [a]	2255	997	458	593	908	1905	
overall $[\theta]_{res}$ [b] = 1515.3							
(B) interval 243–210 nm							
% (v/v) TFE:	0	17	33	50	67	83	0 (6 M Gdm·Cl)
f_α	0.12	0.23	0.29	0.32	0.35	0.42	0.09
f_β	0.05	0.13	0.20	0.21	0.27	0.35	0.17
f_R	0.83	0.64	0.51	0.47	0.38	0.23	0.74
$[\theta]_{res}$ [a]	956	273	309	227	495	524	409
overall $[\theta]_{res}$ [b] = 560.7							

[a] $[\theta]_{residual}$ is calculated as $[(\Sigma_{i=1}^{n} r_i^2)/(n-2)]^{1/2}$, where n is the number of data points for that curve.
[b] $[\theta]_{residual}$ is calculated as $[(\Sigma_{i=1}^{m} r_i^2)/(m-p)]^{1/2}$, where m is the total number of data points, and p is the total number of parameters determined.

very unsatisfactory as the lower end of the wavelength range is increased from 210 nm to 201 nm. While the estimates for f_α agree closely for parts A and B of table 3, the values for f_β differ greatly. The operation of the NILES algorithm does not conveniently supply estimates for the standard errors of the parameters.

The Marquardt method was chosen because of its proven ability to converge with reasonable rapidity from poor guesses. Moreover, at the termination of the procedure the matrix N^{-1} is available for examination. The great difficulty with any nonlinear least squares algorithm is that convergence to the global minimum may not be obtained if either the sum of squares surface is poorly conditioned or if local mini-

ma lie in the region of the initial guess. We chose to investigate the sum of squares surface by using two sets of initial guesses for the parameter estimates.

In parts A and C of table 4 the initial guess for the aperiodic spectrum was the spectrum reported for the aperiodic conformer by Chen et al. [5], together with the values for f_α and f_β from table 1 with negative values set to zero. It is clear that two initial starting points have led to convergence at two different apparent minima, with considerably different values for f_β. It should be noted that the differences in values for f_β in parts A and B of table 4 are very nearly all

Table 4
Analysis by constrained Marquardt algorithm

% (v/v) TFE	f_α (initial)	f_α (final)	est.S.D.	f_β (initial)	f_β (final)	est.S.D.	$[\theta]_{res}$ [a]
(A) interval 243–201 nm							
0	0.10 [b]	0.10	0.05	0.24	0.31	0.19	1897
17	0.23	0.23	0.08	0.12	0.17	0.14	815
33	0.32	0.31	0.10	0.00	0.04	0.07	547
50	0.34	0.32	0.10	0.00	0.00	0.01	718
67	0.38	0.36	0.10	0.00	0.01	0.02	1045
83	0.47	0.44	0.11	0.00	0.00	0.00	1988
overall $[\theta]_{res}$ = 1445.3							
(B) interval 243–201 nm							
0	0.09 [c]	0.08	0.04	0.49	0.50	0.16	2033
17	0.21	0.21	0.07	0.33	0.35	0.14	944
33	0.30	0.30	0.08	0.19	0.22	0.11	450
50	0.32	0.32	0.08	0.12	0.15	0.09	541
67	0.37	0.37	0.08	0.09	0.13	0.08	784
83	0.44	0.45	0.09	0.00	0.06	0.06	1473
overall $[\theta]_{res}$ = 1309.1							
(C) interval 243–210 nm							
6 M Gdm·Cl	0.07 [b]	0.04	0.02	0.21	0.23	0.17	250
buffer	0.10	0.06	0.02	0.21	0.23	0.15	866
17% (v/v) TFE	0.21	0.19	0.04	0.28	0.30	0.16	262
33% (v/v) TFE	0.28	0.26	0.04	0.32	0.33	0.17	411
50% (v/v) TFE	0.30	0.29	0.04	0.33	0.33	0.17	323
67% (v/v) TFE	0.34	0.33	0.05	0.37	0.37	0.18	613
83% (v/v) TFE	0.41	0.41	0.05	0.42	0.42	0.19	565
overall $[\theta]_{res}$ = 560.7							
(D) interval 243–210 nm							
6 M Gdm·Cl	0.09 [c]	0.09	0.02	0.17	0.13	0.12	249
buffer	0.12	0.12	0.03	0.05	0.05	0.07	867
17% (v/v) TFE	0.23	0.23	0.04	0.13	0.13	0.12	262
33% (v/v) TFE	0.29	0.29	0.04	0.20	0.19	0.15	411
50% (v/v) TFE	0.32	0.32	0.04	0.21	0.21	0.16	323
67% (v/v) TFE	0.35	0.35	0.04	0.27	0.27	0.18	613
83% (v/v) TFE	0.42	0.42	0.05	0.35	0.36	0.20	565
overall $[\theta]_{res}$ = 560.7							

[a] For examination of the lack of fit to each separate curve $[\theta]_{res}$ is taken as $[(\Sigma_{i=1}^{n} r_i^2)/(n-2)]^{1/2}$, where n is the number of data points for that curve.

[b] The initial guess was taken from the parameter estimates obtained from the linear least squares method, along with the CYC estimate for the aperiodic spectrum.

[c] The initial guess was taken from the final values of the parameter estimates obtained by the NILES algorithm of Markussen and Vølund.

within one estimated standard error. The most striking aspect of table 4 is the fact that the Marquardt algorithm succeeded in considerably reducing the sum of squared residuals (part B) below that corresponding to the solution obtained by the NILES algorithm (table 3, part A). Thus, the NILES algorithm appears to have reached a point on the sum of squares surface at which further iterations failed to substantially reduce the sum of squares, while the Marquardt algorithm, beginning at that point, fairly rapidly reached a lower apparent minimum.

In parts C and D of table 4 are shown results for the 243–210 nm wavelength range when two initial starting points are chosen. In this case, two differing solutions were obtained, having an identical value for the residual sum of squares, but very different estimates for the content of β-sheet under the various solution conditions. Again, the differences in the estimates for f_β are of the same size as the estimated standard errors for f_β. The estimates obtained for α-helicity for all the spectra, for all solutions in table 4, appear in close agreement, with differences being of about the same size as the estimated standard error in f_α. The estimates obtained for the aperiodic CD spectrum of S-CM insulin A chain are shown in fig. 2. For the 243–210 nm interval, the estimated standard error for $[\theta]_\lambda^R$ was about 1800 deg cm^2/dmole, and for the 243–201 nm interval about 4900 deg cm^2/dmole. The estimates obtained for the aperiodic spectrum by the Marquardt algorithm are shown in fig. 2.

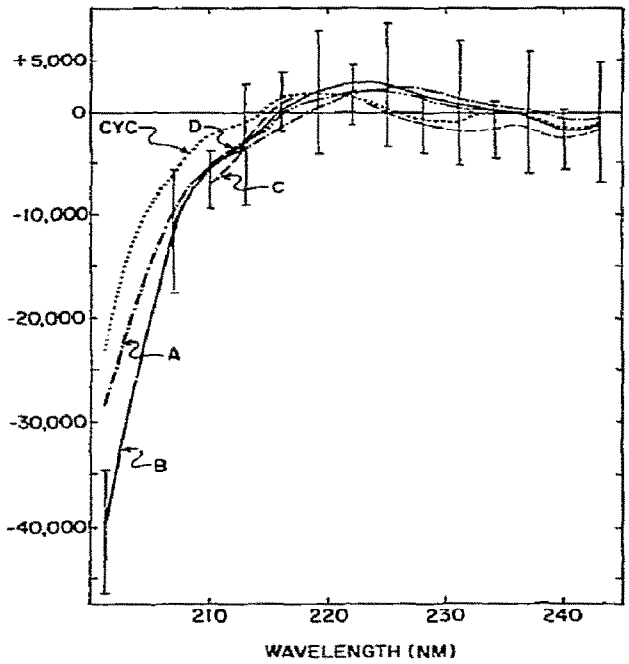

Fig. 2. The estimated CD spectrum of the aperiodic conformer as calculated using the Marquardt algorithm. Curves A, B, C and D refer to the estimates obtained in parts A, B, C and D of table 4. Standard error bars given for curves A and D are comparable to those obtained for curves B and C, respectively (not shown for clarity). The dotted line is the CD spectrum of the aperiodic conformer given by Chen et al. [5].

5. Discussion

Several assumptions are presently required to estimate secondary structure content from a CD spectrum. Firstly, it is assumed that the optical activity associated with the far UV transitions of the aromatic chromophores are negligible compared with the peptide contribution. Secondly, the assumption that a unique reference spectra set, computed from the known secondary structure contents (obtained via X-ray diffraction) and CD spectra of standard proteins, will be suitable for the analysis of a particular polypeptide is made. Thirdly, it is suggested that if a family of curves is available, the aperiodic conformer will have the same CD spectrum under the various solution conditions used, and that pooled data can be used to obtain more accurate estimates for the secondary structure content if the CD spectrum of the aperiodic conformer becomes part of the parameter set to be estimated. The CD results obtained with the S-CM insulin A chain indicate that most, if not all, of these assumptions are invalid for this polypeptide.

The results given in tables 1 and 2 demonstrate the unsuitability of the basis spectra set of Chen et al. [5] for this polypeptide. The data also show that the problem with the basis spectra set is particularly severe below 210 nm. At this point it seems plausible to suppose that a different reference spectrum for the aperiodic conformer might remedy this extreme lack of fit and uncertainty in f_β values. It appears that the CYC aperiodic reference spectrum is unsuitable, and an examination of the CD spectra of the S-CM insulin A chain in buffer and in 6 M Gdm·Cl reveals that in buffer the polypeptide may possess a small amount of or-

dered structure and, thus, this spectrum could not be considered as the aperiodic reference spectrum. The data for S-CM insulin A chain in 6 M Gdm·Cl are reasonably accurate only to about 215 nm; use of this data as a reference spectrum for the aperiodic conformer would exclude much of the data from the analytical process.

Therefore, we are led to consider the next possibili ty. This is to permit the aperiodic spectrum to be part of the parameter set to be estimated, assuming that it remains constant as solution conditions are changed. Because of the fact that nonlinear regression procedures have many hazards including the problem of the initial guess for the parameters, possible convergence problems, and possible existence of local as well as global minima, two different algorithms were explored. The results obtained suggest for the set of far UV CD data analyzed here that the following generalizations are valid. The most straightforward process for obtaining an initial guess for the parameter estimates, linear least squares results, along with the aperiodic spectrum of Chen et al. [5], seems to be adequate for the Marquardt algorithm. The number of iterations required for each algorithm to reach convergence were about the same, however, the computer time per iteration is larger for the Marquardt algorithm. The residual sum of squares surface for both the 243–201 nm and the 243–210 nm wavelength interval appears to be very poorly conditioned. For the 243–210 nm interval, multiple minima, with the same value for the residual sum of squares, exist. For both wavelength intervals, the likely error in f_α seems to be about 0.05–0.10, while the uncertainty in f_β is considerably larger, 0.10 to 0.20. In short, it appears that once the aperiodic spectrum is added to the parameter set to be estimated, the amount of β-sheet becomes very difficult to estimate with any reasonable degree of precision.

Examination of the lack of fit obtained in parts A and B of table 4 suggest that even allowing the aperiodic spectrum to vary in order to increase the degree of fit to the experimental data still does not produce an acceptable fit to most of the spectra. There are two possible explanations for this lack of fit. First, the assumption that the aperiodic spectrum remains constant as the solvent composition is changed may not be valid. Examination of the near UV CD spectra in fig. 1 shows a substantial alteration in the tyrosyl near UV CD bands as the solvent is changed. It is possible that similar changes occur in the tyrosyl contribution to the far UV CD spectrum, negating the assumption of a constant aperiodic spectrum. A second possibility is that the reference spectra reported by Chen et al. [5] for the α-helical and β-sheet conformers may not suitably describe the peptide chromophore of S-CM insulin A chain in the 210–201 nm region. No firm evidence exists to rule out either of these two possibilities.

In summary, while this study has elucidated some problems involved in several proposed methods for far UV CD analysis, the various methods which have been examined have all failed to provide reasonably precise estimates for the content of β-sheet structure in S-CM insulin A chain in various concentrations of TFE. This study illustrates the hazards involved in the nonlinear least squares method. It is clear that estimates obtained for f_α and f_β from a single initial starting guess should not be accepted without careful consideration of the problems of the large standard errors in f_β, the very ill-conditioned sum of squares surface, and the possible existence of multiple, equally good, minima on the surface.

S-CM insulin A chain appears to possess a very small amount of structure in aqueous buffers, to be essentially unordered in 6 M Gdm·Cl, and to undergo a conformational transition as the concentration of TFE is increased to 83% to give a structure which is about 45% α-helical, with an essentially indeterminate amount of β-sheet present.

Acknowledgement

This investigation was supported by the Vanderbilt University Diabetes-Endocrinology Center (AM-17026) and AM-15838. The excellent services of the Vanderbilt University Computer Center are gratefully acknowledged.

References

[1] B.H. Phelps and P.T. Varadani, Federation Proceedings 35 (1976) 1628.
[2] G.H. Beaven and E.R. Holiday, Advan. Protein Chem. 7 (1952) 319.
[3] Y. Chen and J.T. Yang, Biochem. Biophys. Res. Commun. 44 (1971) 1285.

[4] Y. Chen, J.T. Yang and H.M. Martinez, Biochemistry 11 (1972) 4120.
[5] Y. Chen, J.T. Yang and K.H. Chau, Biochemistry 13 (1974) 3350.
[6] J. Markussen and A. Vølund, Intern. J. Peptide Protein Res. 7 (1975) 47.
[7] H. Wold, in: Festschrift for J. Neyman, ed. F.N. David (Wiley, New York, 1966) pp. 411–444.
[8] D.W. Marquardt, J. Soc. Ind. Appl. Math. 11 (1963) 431.
[9] K.M. Brown and J.E. Dennis, Numerische Mathematik 18 (1972) 289.
[10] T.J. Aird and H.L. Darilek, in: International Mathematics and Statistics Library, ZXSSQ-1-ZXSSQ-6, IMSL, Houston (1975).
[11] Y. Bard, Nonlinear parameter estimation (Academic Press, New York, 1974) pp. 176–179.
[12] G.N. Wilkinson, Biochem. J. 80 (1961) 324.
[13] R. Grosse, J. Malur, W. Meiske and K.R.H. Repke, Biochim. Biophys. Acta 359 (1974) 33.