

CCA-1852

YU ISSN 0011-1643

UDC 541.653

Original Scientific Paper

Convex Constraint Decomposition of Circular Dichroism Curves of Proteins[†]

András Perczel and Miklós Hollósi*

The Department of Organic Chemistry, L. Eötvös University, P.O.B. 325, H-1445 Budapest, Hungary

Gábor Tusnády

Mathematical Institute of the Hungarian Academy of Sciences, Budapest, Hungary

Gerald D. Fasman

Graduate Department of Biochemistry, Brandeis University, Waltham, MA 02254

Received November 15, 1988

A new algorithm, called convex analysis, has been developed to deduce the chiral contribution of the common secondary structures directly from experimental circular dichroism (CD) curves of a large number of proteins. The analysis is based on CD data reported by Yang *et al.*¹ Test runs were performed on sets of artificial protein spectra created by the Monte Carlo technique using poly-L-lysine based component spectra. Application of the decomposition algorithm for the created sets of spectra resulted in component spectra $[B(\lambda, i)]$ and weights $[C(i, k)]$ with excellent Pearson correlation coefficients (r).² The algorithm, independent of X-ray data, revealed that the CD spectrum of a given protein is composed of at least four independent sources of chirality. Three of the computed component curves show remarkable resemblance to the CD spectra of known protein secondary structures. This approach yields a significant improvement compared to the eigenvector analysis of Hennessey and Johnson.³ The new method is a useful tool not only in analyzing CD spectra but also in treating other decomposition problems where an additivity constraint is valid.

INTRODUCTION

Circular dichroism (CD) spectroscopy has been widely used for the analysis of the secondary structure of proteins due to its sensitivity in distinguishing the presence and proportion of α -helical, β -pleated sheet and unordered conformations.^{1,4} Alternative approaches have been developed to define the spectral contribution of the basic components. Computation was performed by using reference spectra derived from model polypeptides⁵⁻⁷ or proteins.^{2,8-14}

[†] This work was supported by the Institute of Science Management and Informatics, Ministry of Education, Budapest, Hungary, and the National Science Foundation, Washington, D. C. USA, DMB 87 13193.

This is publication # 5 from the Graduate Department of Biochemistry, Brandeis University, Waltham, MA, 02254.

Provencher and Glöckner¹⁵ analyzed the experimental CD curves as linear combinations of the spectra of proteins whose structure was determined by X-ray diffraction. Hennessey and Johnson³ used basically the same approach, but applied an eigenvector method of multicomponent matrix analysis.¹⁶ The variable selection method¹⁷ added the flexibility of the Provencher and Glöckner approach to the original analysis by Hennessey and Johnson. The latter methods avoid difficulties arising from the selection of suitable reference spectra, but have the disadvantages which are characteristic of all methods deriving basis spectra from proteins of known secondary structure.¹

Making use of a new algorithm, an approach was developed, called convex constraint analysis, to deduce the spectral contribution of the common secondary structures directly from experimental CD curves of a large number of proteins.*

EXPERIMENTAL

Data Basis

The analysis is based on CD data published by Yang *et al.*¹ The first set of data summarized in Table VIII of this excellent review gives the mean residue ellipticities of eighteen proteins in the wavelength range of 240–190 nm (Chang *et al.*²). Table IX contains $\Delta\epsilon$ data of fifteen proteins and one helical polypeptide (poly-L-glutamic acid) which were measured over the range of 260–178 nm and previously used as the basis of the analysis of Hennessey and Johnson.³ The CD parameters of poly-L-glutamic acid were excluded from the analysis reported here.

Method of Analysis

The circular dichroism (CD) of proteins is measured as a function of the wavelength. By keeping all the other factors constant and neglecting the chiral contribution of non-peptide chromophores, this function will depend only on the chiroptical properties of the various secondary structures. The experimental CD spectra are represented by functions $A(\lambda, k)$, where λ stands for the wavelength and k is the serial number of the protein. By assuming additivity of the chiral contribution of conformations like α -helix, β -pleated sheet, turns, *etc.*

$$A(\lambda, k) = \sum_{i=1}^P B(\lambda, i) C(i, k) + \text{noise}, \quad (1)$$

where P is the number of the allowed conformational components, $B(\lambda, i)$ is the CD of the i th component at wavelength λ , and $C(i, k)$ is the weight (proportion) of the i th component in the k th protein.

The Convex Constraint Analysis of CD Curves

Let us denote the investigated proteins by S_1, \dots, S_N where N is the number of proteins. We shall suppose that

$$\text{a) } \sum_{i=1}^P C(i, k) = 1, \quad k = 1, 2, \dots, N$$

$$\text{b) } C(i, k) \geq 0, \quad i = 1, 2, \dots, P$$

c) the noise is proportional with

$$E(\lambda) = 1/2N \sum_{k=1}^N A^2(\lambda, k) + 1/M \sum_{j=1}^M \sum_{k=1}^N A^2(\lambda_j, k), \quad (2)$$

where $\lambda_1 < \lambda_2 < \dots < \lambda_M$ are discrete wavelengths.

The first step in the calculation is the factor analysis on the weighted covariance function.

$$Q(k, n) = \sum_{j=1}^M 1/E(\lambda_j) [A(\lambda_j, k) - \bar{A}(\lambda_j)] [A(\lambda_j)] [\bar{A}(\lambda_j, k) - \bar{A}(\lambda_j)] \quad (3)$$

where

$$\bar{A}(\lambda) = (1/N) \sum_{k=1}^N A(\lambda, k). \quad (4)$$

Let us denote the eigenvalues of $Q(k, n)$ by

$$\varepsilon(1) \geq \varepsilon(2) \geq \dots \geq \varepsilon(N-1) \geq \varepsilon(N) = 0.$$

The variance of the noise is then estimated by

$$\sigma^2(P) = (1/DF) \sum_{k=p+1}^{N-1} \varepsilon(k), \quad (5)$$

where

$$DF = M(N-1) - P(N+M-1-P). \quad (6)$$

If the number of components or factors (P) is chosen apriori and P is small enough, it is more economic to use the so-called QR algorithm for calculating the main factors. It is based on the following well-known fact. For given matrices Q, R the difference

$$\|Q - RU\|^2 \quad (7)$$

is minimized by

$$U_0 = (R^T R)^{-1} R^T Q,$$

i. e.

$$\|Q - RU\|^2 \geq \|Q - RN_0\|^2 \quad (8)$$

which holds true for any U , where the norm is the usual l_2 -norm:¹⁸

$$\|A\|^2 = \sum_{j=1}^M \sum_{k=1}^N A^2(\lambda_j, k) \quad (9)$$

Constraints a) and b) are necessary but not sufficient. As a matter of fact they are not real constraints because one can substitute the components $B(\lambda, i), C(i, k)$ with

$$\begin{aligned} B(\lambda, i) &= \sum_{n=1}^P B(\lambda, n) S^{-1}(n, i) \\ C(i, k) &= \sum_{n=1}^P S(i, n) C(n, k) \end{aligned} \quad (10)$$

where S is an arbitrary invertible matrix.

For the sake of uniqueness it is proposed to use a transformation matrix here with the largest possible determinant which is equivalent to embedding the points

$$\{C(n, k), n = 1, \dots, P\}, k = 1, \dots, N$$

of the P -dimensional Euclidean space in a simplex of the smallest volume. The embedding procedure can be performed by a volume-minimizing algorithm detailed in the Appendix.

Returning to Eq. (1), one can approximate $B(\lambda, j)$ for example, with a Gaussian function

$$g_i(\lambda) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{(\lambda - m_i)^2}{2\sigma_i^2}\right)$$

and create a complete set of $\{m(1), \sigma(1), \dots, m(l), \sigma(l)\}$ where l is a well chosen value. This will generate

$$\sum_{\lambda=\lambda_1}^{\lambda_n} [B(\lambda, j) - \sum_{i=1}^l g_i(\lambda) c_i]^2 \rightarrow \min$$

as a function to be minimized with a common curve fitting procedure.

Test Runs

Test runs were performed to evaluate the new algorithm. $C(i, k)$'s were chosen by the Monte Carlo technique, but the P summits of the simplex were pin-pointed with the vector $C(1, 0, \dots, 0), \dots, C(0, 0, \dots, 1)$, and a varied number of $B(\lambda, i)$'s with different P 's were used (Table I). In series of test runs for $P=3$ the CD spectra of α -helix, β - and unordered forms of poly-L-lysine⁵ were used as $B(\lambda, i)$. For $P=4$ the CD spectrum of a type I β -turn model¹⁹ was also added to the data base. Application of the decomposition algorithm for the created sets of $A(\lambda, k)$'s resulted in C 's and B 's (Eq. 1) with correlation coefficients (r) higher than 0.9.

Correlation coefficients ($-1 \leq r \leq 1$) near 1 indicate the success of the decomposition.² When the summits were not fixed so strictly,

$$C(v, \frac{1-v}{P-1}, \dots, \frac{1-v}{P-1}), \dots, C(\frac{1-v}{P-1}, \dots, v)$$

where v , the pin-pointing coefficient, is 0.75, the correlation coefficients still remained reasonable (Table I).

TABLE I

Pearson correlation coefficients (r)^a for N component sets of test spectra created by the Monte Carlo technique

Pin-pointing coefficient (v)	Number of basis spectra used (P)	N			
		10	12	14	16
0.97	3	0.967	0.984	0.957	0.969
	4	0.938	0.958	0.936	0.900
0.75	3	0.972	0.976	0.989	0.990
	4	0.970	0.979	0.945	0.961

^a As defined by Chang *et al.*²

RESULTS AND DISCUSSION

Prompted by the success of these results, the decomposition algorithm for the CD parameters of proteins listed in the data bases used was performed.¹ The numerical CD values measured in two laboratories supplied a reliable and relatively broad data set for the analysis. On the other hand, the use of these data enabled a direct comparison of results herein with those of Provencher and Glöckner¹⁵ and Hennessey and Johnson³. The method of Hennessey and Johnson,³ not considering constraints a and b , yielded the five most significant basis spectra. The use of a sixth orthogonal basis spectrum hardly improved the analysis and the involvement of more than five of the 16 basis spectra had practically no influence on the spectral features of the

five most important ones. This is not the case in the convex analysis introduced in this paper, the $B(\lambda, i)$ functions were found to depend, as a whole, on P . That is why the choice of the number of the allowed components is of intrinsic importance. In Figure 1 the dependence is shown of $\sigma(P)$ (standard deviation in thousandths) (Eq. 5) from P . The computed $B(\lambda, i)$ functions are given in Figures 2a—3d.

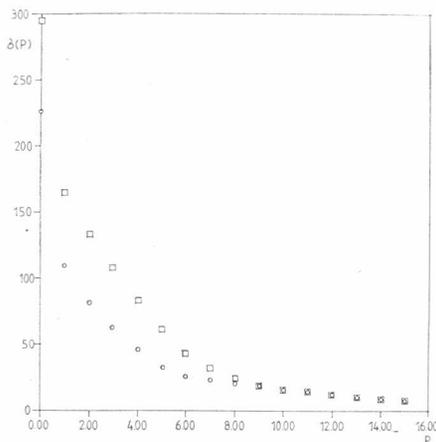


Figure 1. The dependence of standard deviation $\sigma(P)$ in thousandths from the number of the pure components, P .

○, database 240—190 nm (Chang *et al.*,² cf. Table VIII in the review of Yang *et al.*¹).
 □, database 260—178 nm by Hennessey and Johnson³ (cf. Table IX in Yang *et al.*¹).

For $P = 3$ the method generates common component curves which do not resemble the CD spectra of any of the basic secondary structures of proteins (Figures 2a and 3a). This gives strong support to the assumption that the given sets of experimental CD curves are composed of the chiral contribution of more than three conformations. However, even in this unrealistic case, the number of transitions in the component curves agrees with that of experimental CD spectra measured in the 190—260 nm spectral region. This seems to be one of the most important results of our analysis.

For $P = 4—6$, not depending on P and the database used, two of the generated curves, types *A* and *C*, feature the shape of the CD of α -helix and unordered conformation, respectively, with λ_{\max} and $[\theta]$ or $\Delta\epsilon$ values comparable to literature data. Spectral parameters of the $B(\lambda, i)$ component spectra are summarized in Table II. The exciton couplet of the α -helix-like curve *A* appears in the 190—210 nm spectral region with an $n \rightarrow \pi^*$ transition above 220 nm. All spectra labelled by *C* resemble each other. They show a negative maximum below 200 nm as observed in the spectra of unordered polypeptides and proteins.

The curves of type *B* can be correlated with the CD spectrum of the β -form which is well known to be an assemblage of more than one somewhat different secondary structures (parallel, antiparallel, twisted, *etc.*). Therefore it is not surprising that the shape of type *B* functions changes with the data

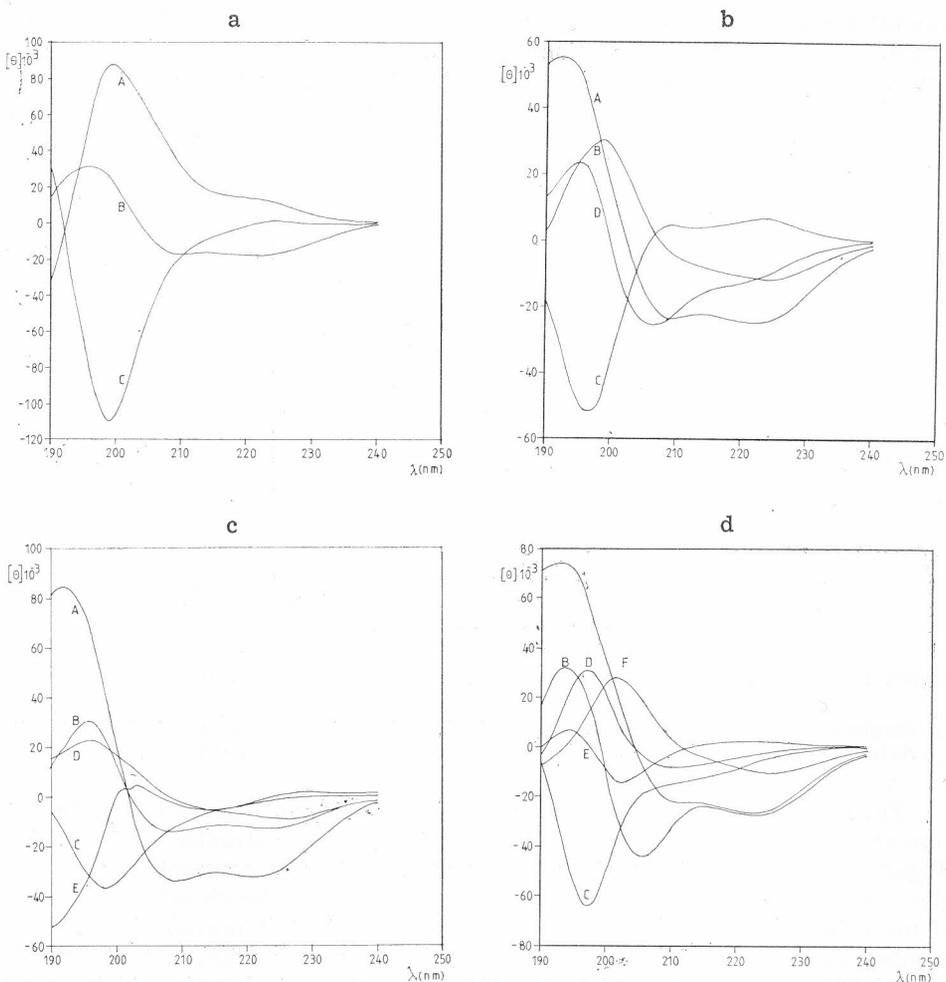


Figure 2. Convex analysis of CD data ($[\theta] \times 10^{-3}$) on 18 proteins, wavelength range: 240–190 nm (Chang *et al.*², cf. Table VIII in the review of Yang *et al.*¹). a, $P = 3$; b, $P = 4$; c, $P = 5$; d, $P = 6$. A (α -type), B (β -type), C (unordered), D [type I(III) turn], E («unlabelled» additional chiral contribution), F («unlabelled» additional chiral contribution).

set used (Figures 2b–3d). It seems to be more conservative for the $\pi \rightarrow \pi^*$ than for the $n \rightarrow \pi^*$ transition.

By increasing the number of the allowed components ($P \geq 4$), the algorithm reveals a component curve (*D*), which may correspond to the CD contribution of type I(III) β -turns and/or a 3_{10} helical structure. Based on the statistical data on the X-ray structure of proteins²⁰ type I and type III β -turns represent the mostly populated group of turn conformations. They account for 60–70% of these structures. According to comparative CD data on linear and bridged model peptides, type I(III) turns show α -helix-like CD spectra with bands of lower intensity.¹⁹ Characteristic CD parameters (λ , $[\theta]$), of typical type I(III) β -turns are in qualitative agreement with those of com-

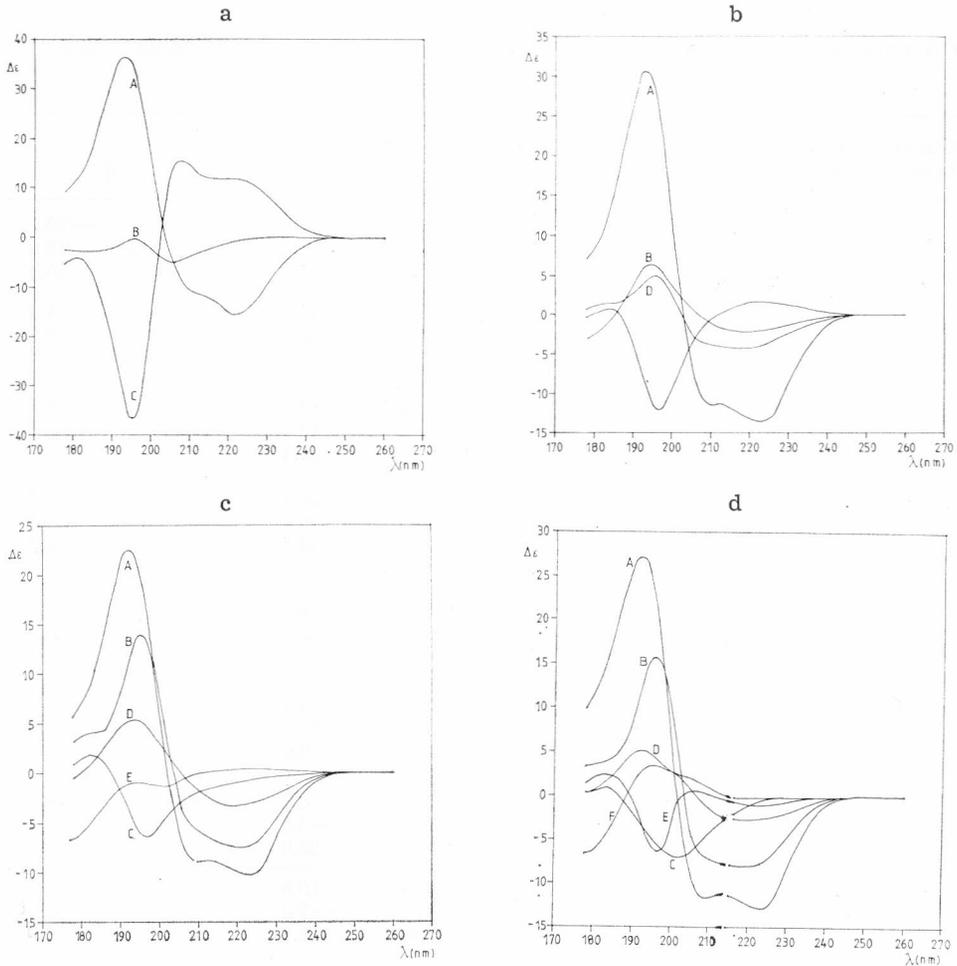


Figure 3. Convex analysis of CD data ($\Delta\epsilon$) on 15 proteins. Measurements were performed over the wavelength range of 260–178 nm by Hennessey and Johnson.³ (cf. Table IX in Yang *et al.*¹). A...F as in Figure 2.

a, $P = 3$; b, $P = 4$; c, $P = 5$; d, $P = 6$.

ponent curve *D* (Table II). 3_{10} helices are repeating type III β -turns and also show α -helix-like spectra with decreased band intensities²¹ which are close to those of type I(III) turns.

As for type II β -turns, the majority were reported to show class B CD spectra,⁴ with a negative band above 220 nm and a positive one below 200 nm. The positive band of the β -pleated sheet conformation also appears between 195–200 nm with a weaker negative band near 217 nm.¹ Taking into consideration that the latter band may be red shifted to the 220–225 nm region, the algorithm is not expected to reveal the chiral contribution of this type of β -turns. (For $P = 4$ and 5, it is probably merged into component curves *B* and *D*.)

TABLE II

CD parameters (λ , nm; $[\Theta] \times 10^{-3}$, mean residue ellipticity) of calculated component curves (B (λ , i))

Curve type	Data base ^a	Number of components (P)					
		4		5		6	
A	(1)	223	-24.0	223	-31.5	223	-28.1
		208.5	-23.1	208	-33.0	210sh	-22.0
		193	56.0	193	84.0	195	75.0
	(2)	223	-44.2	223	-34.5	223	-44.0
		210	-37.7	209	-29.7	209	-39.0
		193.5	100.8	194	73.0	194	88.5
B	(1)	225	-11.2	224	-12.0	223	-28.5
		~ 210sh	-5.5	208	-13.5	207	-45.2
		198	30.5	197	31.4	194	34.1
	(2)	218.5	-7.5	222	-24.4	219.5	-26.7
				~ 210sh	-20.0	210	-24.5
		195	20.9	197	44.5	196	53.4
C	(1)	224	8.5				
		208	5.1	~ 212sh	-8.5	~ 211sh	-15.0
		197	-52.0	198	-36.0	197	-64.0
	(2)	224	5.7			225	1.0
				~ 205sh	-9.5		
		197	-38.8	197	-21.5	202	-23.0
			184	6.0			
D	(1)	~ 221	-12.0	225.5	-8.2	~ 219sh	-4.5
		206	-24.9	~ 212sh	-3.5	209.5	-7.7
		195	23.0	197	23.0	197	31.1
	(2)	223.5	-13.8	220	-10.8		
		213.5sh	-13.0	~ 211sh	-7.5	216	-7.5br
		196	16.2	194	18.3	192	16.5
E	(1)			~ 231	2.1	~ 222	2.6br
				213.5	-5.3	203	-15.5
				202.5	5.4	194	8.0
	(2)			~ 226	2.7br	223	-4.0
				201.5	-5.1	205.5	2.8
						197	-21.0
				185	8.3		
F	(1)					226	-10.5
						~ 211sh	-0.5
						202	28
	(2)					~ 226	3.0
					217	-1.0	
					196	12.0	

^a For data bases (1) and (2) see Tables VIII and IX, respectively, of Yang *et al.*¹ $\Delta\epsilon$ values resulting from data set (2) are converted into $[\Theta]$ for the sake of comparison.

The curves *E* and *F* have been called »unlabelled« component spectra. They may reflect the presence of other conformations, or more generally, additional sources of optical activity. It is expected, however, that these component curves will be correlated with further chiroptically different structural elements on the basis of the X-ray structure and/or amino acid composition of the proteins whose experimental CD curves served as the database of the decomposition procedure.

Our analysis is rather sensitive to the data set used. There are only 9 proteins whose CD spectra were measured in both laboratories (cf. Tables VIII and IX in the review of Yang *et al.*¹). Their CD parameters are generally in good agreement except for small differences in the intensity of the bands. In some cases there are, however, significant differences regarding even the number of transitions observed in the 260—200 nm region. According to the data¹ in Table VIII, ribonuclease S, cytochrome *c* and subtilisin BPN' feature two negative extrema above 190 nm which is in contradiction with the single negative transition given in Table IX. These differences together with the consequences of database truncation to 190 nm (Table VIII) may give an explanation for the somewhat differing results of our analysis using data from Table VIII or IX, respectively. (The general effect of database truncation is discussed in detail by Hennessey and Johnson³).

The variable selection method¹⁷ improves the analysis by removing proteins from the basis set whose CD spectra contain chiral contributions not found in the CD curve of the protein being decomposed. The convex analysis circumvents this problem by using a noise term (Eq. 1) to eliminate individual chiral components. The main power of the method reported here is that it does not make use of X-ray data in the manner of Provencher and Glöckner.¹⁵ The proportion of differing and chiroptically independent secondary structural units may be similar in the crystalline state and in solution, but it is not necessarily the same. On the other hand, we agree with the above authors that constraints a and b [$\sum_{i=1}^P C(i, k) = 1$ and $C(i, k) \geq 0$] are not artificial. Originating in the homochirality of the amino acid units, the CD curve of none of the known protein secondary structures relates to any of the others as mirror images. This is true even of the CD of β -turns. The spectrum of type II' β -turns is, for example, not the mirror image of type II spectrum, but resembles that of the predominant type I and III ones.⁴

Our analysis provides evidence that the CD spectra of proteins contain independent information of at least four different secondary structures. Without making use of any basis spectra originating from model systems of known secondary structure, the analysis herein gives independent proof of the existence of the different secondary structural elements with individual CD spectra. These spectra proved to be identical or, at least, similar to those found in model polypeptides. The new mathematical approach is a useful tool, not only in analyzing CD spectra, but also in other cases where the additivity constraint (a) is valid.

APPENDIX

Let us suppose that the constraints a, b are already met. The S matrix used for the transformation may be constructed by the iteration

$$S_{n+1} = T_n S_n$$

where

$$T_n \text{ has the form } T_n(i, j) = \begin{cases} \gamma(i) & \text{if } i = j \in H_n \\ 0 & \text{if } j \neq i \in H_n \\ 1 & \text{if } i = j \in G_n \\ 1 - \gamma(j) & \text{if } j \neq i \in G_n \end{cases} \quad (11)$$

where the set G_n has a single index ϱ_n , $H_n = \{1, 2, \dots, P\}/G_n$, and ϱ_n is the index of one specified row playing a discriminated role in the algorithm. [One possibility for the choice of ϱ_n is $\varrho_n = n \pmod{P}$]. One can easily check that

$$\det(T_n) = \pi \prod_{i \in H_n} \gamma(i) \quad (12)$$

thus T_n may be chosen as a solution for the next problem.

For the sake of simplicity let us speak about the first step. Given the integer $\varrho = \varrho_1$, $1 \leq \varrho \leq P$, find the variables $\{\gamma(i), i \in H\}$, such that

$$\sum_{i \in H} \gamma(i) C(i, k) \leq 1 \quad k = 1, \dots, N \quad (13)$$

and

$$w = \pi \prod_{i \in H} \gamma(i) \rightarrow \max \quad \text{now } H = \{1, 2, \dots, P\} \setminus \{\varrho\}$$

For arbitrary $(P - 1)$ dimensional vectors

$$\{d(i), i \in H\}$$

such that

$$d(i) = \sum_{k=1}^N \delta(k) C(i, k), \quad i \in H \quad (14)$$

with some constants $\delta(k) \geq 0$, $\sum_{k=1}^N \delta(k) = 1$ we have

$$\sum_{i \in H} \gamma(i) d(i) \leq 1.$$

If we had only the last constraints, the product w would be maximized by

$$\gamma(i) = \frac{1}{(P - 1) d(i)} \quad (15)$$

and one can show that the original problem is equivalent to

$$\pi \prod_{i \in H} d(i) \rightarrow \max$$

which can be solved e.g. by a gradient iteration. Let us label the convex hypersurface by D ;

$$D = \{d: \exists \delta \in \Delta_N, d(i) = \sum_{k=1}^N \delta(k) \cdot C(i, k)\}$$

where

$$\Delta_N = \{\delta(k) \geq 0, \sum_{i=1}^N \delta(k) = 1\}$$

The transformation

$$\log \pi \prod_{i \in H} d(i) = \sum_{i \in H} \log d(i) \quad (16)$$

leads to the following equation

$$f[\delta(1), \delta(2), \dots, \delta(N)] = \sum_{i \in H} \log d(i) \quad (17)$$

which is to derive:

$$\frac{\partial f}{\partial \delta(k)} = \sum_{i \in H} \frac{C(i, 1)}{d(i)}, \dots, \sum_{i \in H} \frac{C(i, k)}{d(i)}, \dots, \sum_{i \in H} \frac{C(i, N)}{d(i)} \quad (18)$$

where

$$1 \leq k \leq N$$

From there the successive $\delta_{n+1}(k)$ can be calculated as

$$\delta_{n+1} = \delta_n + \alpha \frac{\partial f}{\partial \delta_n(\alpha)}$$

where

$$0 \leq \alpha \leq 1.$$

REFERENCES

1. J. T. Yang, C.-S. C. Wu, and H. M. Martinez, in *Methods in Enzymology*, Vol. 130, Academic Press (1986), pp. 208—269.
2. C. T. Chang, C.-S. C. Wu, and J. T. Yang, *Anal. Biochem.* 91 (1978) 12.
3. J. P. Hennessey and W. C. Johnson, Jr., *Biochemistry* 20 (1981) 1085.
4. R. W. Woody, in *The Peptides*, Vol. 7, (Hruby, V. J., ed.) Academic Press (1985), pp. 15—113.
5. N. Greenfield and G. D. Fasman, *Biochemistry* 8 (1969) 4108.
6. H. Rosenkranz and W. Scholtan, *Z. Physiol. Chem.* 352 (1971) 896.
7. S. Brahm and J. Brahm, *J. Mol. Biol.* 138 (1980) 149.
8. V. P. Saxena and D. B. Wetlauffer, *Proc. Nat. Acad. Sci. USA* 68 (1971) 969.
9. Y. H. Chen and J. T. Yang, *Biochem. Biophys. Res. Comm.* 44 (1971) 1285.
10. R. Grosse, J. Malur, W. Meiske, and K. R. H. Repke, *Biochim. Biophys. Acta* 359 (1974) 33.
11. Y. H. Chen, J. T. Yang, and K. H. Chan, *Biochemistry* 13 (1974) 3350.
12. J. Markussen and A. Vølund, *Int. J. Peptide Protein Res.* 7 (1975) 47.
13. J. B. Siegel, W. E. Steinmetz, and G. L. Long, *Anal. Biochem.* 104 (1980) 160.
14. I. A. Bolotina, V. O. Chekhov, V. Lugauskas, and O. B. Ptitsyn, *Mol. Biol. (USSR)* 14 (1980) 709 (Engl. transl.).
15. S. W. Provencher and J. Glöckner, *Biochemistry* 20 (1981) 33.
16. D. Lloyd, *Ph. D. Thesis*, University of California (1969), Berkeley.
17. P. Manavalan and W. C. Johnson, Jr., *Anal. Biochem.* 167 (1987) 76.
18. M. Balla and G. Tusnády, *Periodica Mathematica Hungarica* 16 (1985) 201.
19. M. Hollósi, K. E. Kövér, S. Holly, and G. D. Fasman, *Biopolymers* 26 (1987) 1527.
20. P. Y. Chon and G. D. Fasman, *J. Mol. Biol.* 115 (1977) 135.
21. T. S. Sudha, E. K. S. Vijayakumar, and P. Balaram, *Int. J. Peptide Protein Res.* 22 (1983) 464.

SAŽETAK

Rastavljanje konveksne napetosti u krivuljama cirkularnog dikroizma proteina

András Perczel, Miklós Hollósi, Gábor Tusnády i Gerald D. Fasman

Razvijen je novi algoritam, nazvan konveksna analiza, za određivanje kiralnog doprinosa uobičajenih sekundarnih struktura, neposredno iz eksperimentalnih CD krivulja većeg broja proteina. Analiza se zasniva na CD podacima Yanga *et al.* (1986). Prvi pokušaji provedeni su na skupini umjetnih proteinskih spektara kreiranih Monte Carlo tehnikom, koristeći sastavne spektre zasnovane na poli-L-lizinu. Primjena algoritma rastavljanja na ostvareni niz spektara rezultirao je

djelomičnim spektrima $[B(\lambda, i)]$ i težinama (udjelima) $[C(i, k)]$ uz izvanredan Pearsonov korelacijski koeficijent (r) [C. T. Chang, C. S. C. Wu, J. T. Yang, (1978) sonov korelacijski koeficijent (r) [C. T. Chang, C. S. C. Wu, J. T. Yang, *Anal. Biochem.* **91** (1978) 12].

Neovisno o podacima rentgenske analize ovaj je algoritam pokazao da je CD spektar određenog proteina sastavljen od barem četiri različita neovisna izvora kiralnosti. Tri od izračunane komponente krivulje pokazale su znatnu sličnost s CD spektrima poznatih proteinskih sekundarnih struktura. Ovaj pristup predstavlja znatno poboljšanje u odnosu na analizu svojstvenih vektora J. P. Hennessey and W. C. Johnson, *Biochemistry* **20** (1981) 1085]. Nova metoda nije samo upotrebljiva za analizu CD spektara nego i za razmatranje drugih problema rastavljanja, gdje vrijedi aditivnost.