

# Evidence for the Validity of Three-Component Fitting of Protein Circular Dichroism Spectra

W. H. Bannister and J. V. Bannister

Department of Physiology and Biochemistry, Royal University of Malta, Msida

(Z. Naturforsch. 29 c, 9–11 [1974]; received September 5, 1973)

Protein Circular Dichroism Spectra, Three-Component Basis, Matrix Rank Analysis, Curve Fitting

The validity of three-component fitting of protein circular dichroism (CD) spectra in the peptide absorption region with basis spectra for helical,  $\beta$ -sheet and unordered structure is examined by matrix rank analysis of protein CD spectra. It is shown that a three-component basis can be justified for several proteins within the limits of experimental error of the spectra. It is suggested that the fitting of protein CD spectra with basis spectra derived from reference proteins should be preceded by matrix rank analysis to show that the CD spectra of the reference and unknown proteins have the same basis.

## Introduction

The main forms of secondary structure in protein molecules are known to be helical,  $\beta$ -sheet and unordered structures. Since the discovery of the optical activity of these structures, proposals have been made to determine their proportions in globular proteins from optical rotatory dispersion (ORD) and circular dichroism (CD) spectra. An interesting approach in recent years has been to consider protein ORD and CD spectra in the region of peptide absorption as a linear combination of three basis spectra due to helical,  $\beta$ -sheet and unordered structure<sup>1,2</sup>. It is possible to compute statistically average basis spectra from the ORD or CD spectra of a set of reference proteins whose structural composition has been determined by X-ray diffraction analysis<sup>3,4</sup>. These spectra are preferable to the spectra of synthetic poly- $\alpha$ -amino acids in  $\alpha$ -helical,  $\beta$ -sheet and random coil conformations<sup>1,2,5</sup> for fitting protein ORD and CD spectra<sup>3,4</sup>. The assumption of only helical,  $\beta$ -sheet and unordered structure contributions in protein ORD and CD in the peptide absorption region is obviously a first approximation since it excludes the Cotton effects of any other secondary structure and non-peptide chromophores. These Cotton effects must be negligible for the assumption to hold. We here examine the validity of current three-component fitting of protein CD spectra by considering the intrinsic numerical evidence in some spectra.

Requests for reprints should be sent to W. H. Bannister, Department of Physiology, The Royal University of Malta, Msida, Malta.

Regarding protein CD spectra as vectors whose components consist of the mean residue ellipticity at a series of wavelengths, we can represent the spectra of a set of proteins as a matrix whose columns consist of the individual spectra. If the spectra have a helical,  $\beta$ -sheet and unordered structural basis, we have the matrix equation

$$AC = D,$$

where  $A$  is an  $n \times 3$  matrix whose columns are the basis spectra for the helical,  $\beta$ -sheet and unordered structures,  $C$  is a  $3 \times m$  matrix whose columns consist of the proportions of helical,  $\beta$ -sheet and unordered structure in each protein, and  $D$  is an  $n \times m$  matrix whose columns consist of the spectra of the  $m$  proteins. By a well-known theorem of matrix algebra<sup>6,7</sup>, the rank of  $D$  is equal to the rank of  $A$  or  $C$  whichever is smaller. Since the rank of both  $A$  and  $C$  can be no larger than three by the assumption of a three-component basis, the rank of  $D$  can be no larger than three. The rank of  $C$ , and therefore  $D$ , can of course be less than three if one (or more) of the secondary structures is absent in all proteins of the set forming the basis for this treatment, or its spectrum can be expressed as a linear combination of the spectra of the other secondary structures in the whole set of proteins.

## Methods

Matrix rank analysis was carried out by the method of Wallace and Katz<sup>8</sup>. In this method a companion error matrix is set up. The matrix of experimental data is reduced by Gaussian elimina-



Dieses Werk wurde im Jahr 2013 vom Verlag Zeitschrift für Naturforschung in Zusammenarbeit mit der Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. digitalisiert und unter folgender Lizenz veröffentlicht: Creative Commons Namensnennung-Keine Bearbeitung 3.0 Deutschland Lizenz.

Zum 01.01.2015 ist eine Anpassung der Lizenzbedingungen (Entfall der Creative Commons Lizenzbedingung „Keine Bearbeitung“) beabsichtigt, um eine Nachnutzung auch im Rahmen zukünftiger wissenschaftlicher Nutzungsformen zu ermöglichen.

This work has been digitalized and published in 2013 by Verlag Zeitschrift für Naturforschung in cooperation with the Max Planck Society for the Advancement of Science under a Creative Commons Attribution-NoDerivs 3.0 Germany License.

On 01.01.2015 it is planned to change the License Conditions (the removal of the Creative Commons License condition "no derivative works"). This is to allow reuse in the area of future scientific usage.

tion with complete pivoting to an equivalent matrix with zero elements below the principal diagonal and the error matrix is transformed by taking into account the propagation of error in each reduction step. The rank is given by the number of nonzero elements on the principal diagonal of the reduced data matrix, the criterion for a nonzero element being that it exceeds in absolute value the corresponding element on the principal diagonal of the transformed error matrix. In the present work a nonzero element was assumed when it exceeded its estimated error three times in absolute value. A 1% coefficient of variation in the elements of the matrices of experimental CD spectra was first assumed to set up error matrices and carry out basic calculations. At any other percentage level of error the elements of the transformed error matrix are in proportion to the elements of the matrix found for 1% error according to the ratio of the error level and 1%.

The lower and upper limits of error consistent with the determined rank were found for matrices of sets of CD spectra of the following proteins: Myoglobin, lysozyme, ribonuclease, lactate dehydrogenase and papain<sup>4</sup>; whole histone and histone fractions from chicken erythrocytes<sup>9</sup>;  $\beta$ -lactamases I and IIp from *Bacillus cereus*<sup>10</sup>; bovine erythrocuprein<sup>11</sup> and human erythrocuprein<sup>12</sup>. The spectra were the values of mean residue ellipticity at 2.5 nm intervals between 205 and 230 nm.

Matrices were set up and labelled as follows: D1 — myoglobin, lysozyme, ribonuclease, lactate dehydrogenase and papain spectra; D2 — eight whole histone spectra at different ionic strengths and pH values; D3 — six spectra of histone fraction 2a1; D4 — six spectra of histone fraction 2a2; D5 — seven spectra of histone fraction 2b; D6 — spectra of myoglobin, lysozyme, ribonuclease, whole histone (ionic strength 0.1, pH 7.5) and histone fractions 2a1, 2a2 and 2b (ionic strength 0.05, pH 7.5); D7 — matrix D6 plus  $\beta$ -lactamase I and IIp spectra; D8 — matrix D7 plus bovine and human erythrocuprein spectra; D9 — matrix D1 plus  $\beta$ -lactamase I and IIp and bovine and human erythrocuprein spectra.

The basis spectra for helical,  $\beta$ -sheet and unordered structure calculated by Chen, Yang and Martinez<sup>4</sup> from the spectra of myoglobin, lysozyme, ribonuclease, lactate dehydrogenase and papain were used to obtain a linear least-squares approximation of the spectra of  $\beta$ -lactamases I and IIp and bovine and human erythrocuprein. The fitting was done for data points at 1 nm intervals between 205 and 230 nm.

The spectra for all computations were taken from published graphical or numerical data<sup>4, 9-12</sup>. Profiles of spectra were read by means of a mechanical co-ordinate digitizer with a resolution of 0.1 mm. Five-point Lagrangian interpolation was used on numerical data as necessary to obtain data points at intervals of 2.5 or 1 nm. All computations were performed by means of a Hewlett-Packard 9100B Calculator equipped with a 9101A Extended Memory.

## Results and Discussion

The ranks of the matrices examined are given in Table I. For each matrix the rank ( $r$ ) is  $r + 1$  below the lower limit of the error given and  $r - 1$  above the upper limit.

Table I. Results of matrix rank analysis of protein CD spectra.

Matrix *	Rank	Error limits [%]
D1	three	1.6— 8.0
D2	two	2.0—11.8
D3	two	3.4—12.1
D4	two	1.1— 6.6
D5	two	1.9—10.9
D6	three	2.8— 8.0
D7	three	3.2— 8.0
D8	three	3.2— 8.0
D9	three	4 — 8.0

\* See Methods section.

Dalglish<sup>9</sup> has previously reported the matrices of histone CD spectra to be of rank three. He found the rank to increase to four with inclusion of myoglobin, lysozyme and ribonuclease spectra, remain four on further addition of  $\beta$ -lactamase spectra, and increase to five on still further addition of erythrocuprein spectra. These findings were considered to cast doubt on three-component fitting of protein CD spectra. However, in the light of the present computations it is difficult to uphold these conclusions without assuming an experimental error below 1.1 to 3.4% (see Table I), as the case may be, in the CD spectra. Unfortunately the error considerations in the work of Dalglish<sup>9</sup> were not given.

The matrix rank determinations made in the present work support the three-component fitting of protein CD spectra by calculated basis spectra. The calculation of basis spectra for helical,  $\beta$ -sheet and unordered structure from the CD spectra of myoglobin, lysozyme, ribonuclease, lactate dehydrogenase and papain as reference proteins according

to Cheng, Yang and Martinez<sup>4</sup> is justified by the finding of rank three for a matrix of the protein spectra. The spectra of Chen, Yang and Martinez<sup>4</sup> were reproducible to within 5% (Yang, personal communication). This is in the centre of the error range found for a matrix of the spectra to be of rank three (Table I).

It would appear that  $\beta$ -lactamases I and IIp from *Bacillus cereus* and bovine and human erythrocyte have the same three-component basis for their CD spectra in the 205–230 nm wavelength region as the reference spectra of Chen, Yang and Martinez<sup>4</sup>, since a matrix of all the spectra in question was determined to be of rank three within

Table II. Structural content of *Bacillus cereus*  $\beta$ -lactamases and bovine and human erythrocyte as estimated by least-squares fitting of the CD spectrum between 205 and 230 nm with basis spectra derived from the CD spectra of myoglobin, lysozyme, ribonuclease, lactate dehydrogenase and papain.

Protein	Helix [%]	$\beta$ -Sheet [%]	Un-ordered structure [%]	RMS*
$\beta$ -Lactamase I	59.5	34.4	6.1	300
$\beta$ -Lactamase IIp	47.8	43.9	8.3	190
Bovine erythrocyte	2.2	32.5	65.3	240
Human erythrocyte	5 **	35.8	59.2	140

\* Root mean square of residuals for fitted spectrum (mean residue ellipticity).

\*\* Negative with respect to least squares.

error limits compatible with experimental error (Table I). This was supported by the results of least-squares fitting of the  $\beta$ -lactamase and erythrocyte spectra with the basis spectra of Chen, Yang and Martinez<sup>4</sup>. The results obtained are given in Table II. The goodness of the fits achieved is indicated by the low values of the root mean squares of the residuals for the fitted spectra. The results show that bovine and human erythrocyte contain little or no helical structure and a mixture of  $\beta$ -sheet and unordered structure with preponderance of the latter. The  $\beta$ -lactamases appear to have little unordered structure and to contain helical and  $\beta$ -sheet structure with some preponderance of the former.

The significance of rank two determined for the chicken erythrocyte histone spectra (Table I) is not clear from the present work. Curve fitting of the spectra with basis spectra for helical,  $\beta$ -sheet and unordered structure was not carried out because the mean residue ellipticities for the histone spectra<sup>9</sup> were only available at 5 nm intervals.

The present work suggests a numerical precaution before curve fitting of the CD spectra of proteins of unknown structure with basis spectra for helical,  $\beta$ -sheet and unordered structure derived from a set of reference proteins. This is to check that a matrix of the CD spectra of the reference and unknown proteins is of rank three which shows that the unknown proteins have the same three-component basis for their CD spectra as the reference proteins.

<sup>1</sup> N. Greenfield, B. Davidson, and G. D. Fasman, *Biochemistry* **6**, 1630–1637 [1967].

<sup>2</sup> N. Greenfield and G. D. Fasman, *Biochemistry* **8**, 4108–4116 [1968].

<sup>3</sup> V. P. Saxena and D. B. Wetlaufer, *Proc. nat. Acad. Sci. USA* **68**, 969–972 [1971].

<sup>4</sup> Y.-H. Chen, J. T. Yang, and H. M. Martinez, *Biochemistry* **11**, 4120–4131 [1972].

<sup>5</sup> H. Rosenkranz and W. Scholtan, *Hoppe-Seyler's Z. Physiol. Chem.* **352**, 896–904 [1971].

<sup>6</sup> A. C. Aitken, *Determinants and Matrices*, 9th ed., p. 96, Oliver and Boyd, Edinburgh 1956.

<sup>7</sup> R. M. Wallace, *J. phys. Chem.* **64**, 899–901 [1960].

<sup>8</sup> R. M. Wallace and S. M. Katz, *J. phys. Chem.* **68**, 3890–3892 [1964].

<sup>9</sup> D. G. Dalgleish, *FEBS Lett.* **24**, 134–136 [1972].

<sup>10</sup> D. G. Dalgleish and A. R. Peacocke, *Biochem. J.* **125**, 155–158 [1971].

<sup>11</sup> E. Wood D. Dalgleish, and W. Bannister, *Eur. J. Biochem.* **18**, 187–193 [1971].

<sup>12</sup> W. H. Bannister, D. G. Dalgleish, J. V. Bannister, and E. J. Wood, *Int. J. Biochem.* **3**, 560–568 [1972].