MODIFIED FACTOR ANALYSIS OF THE CIRCULAR DICHROISM SPECTRA, APPLIED TO A SERIES OF CYCLODIPEPTIDES CONTAINING L-PROLINE

Petr PANČOŠKA*, Ivo FRIČ and Karel BLÁHA

Institute of Organic Chemistry and Biochemistry, Czechoslovak Academy of Sciences, 166 10 Prague 6

Received November 9th, 1977

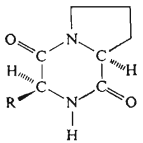
A mathematical model has been developed which uses the theory of Hilbert spaces and the modified method of the principal components of the factor analysis to the determination of the minimum number of subspectra sufficient to describe a set of experimental CD curves of a series of structuraly related compounds. The use of this method has been demonstrated on the CD spectra of nine cyclodipeptides containing (besides L-proline) glycine, L or D alanine, L or D valine, L or D leucine, and L or D tert-leucine, measured in acetonitrile and 2,2,2-trifluoroethanol. Relation is discussed between the calculated subspectra and the structural characteristics of the measured systems.

Due to the extreme sensitivity of the circular dichroism spectra (CD spectra) to the changes in the spatial arrangement of molecules, measuring the chiroptical properties becomes an efficient tool for structure elucidation of optically active compounds[1]. One of the ways of detailed analysis of relations between the structure and the CD spectra of a given type of substances is the study of chiroptical properties under various conditions of the sets of model compounds which differ only by definite structural features[2]. The number of the CD curves which have to be processed is considerable and they are in general of complex character. Mathematical methods for the analysis and data processing can be used to facilitate studies at this stage. The factor analysis has been successfully applied in chemistry to process sets of discrete data (*e.g.*, intensities of the m/e peaks in mass spectra[3-7]). The generality of the mathematical framework of this method[8-11] makes it possible to modify it for different special requirements.

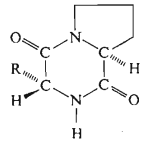
The aim of this paper is to formulate a mathematical procedure in the analysis of the CD spectra of rather extensive sets of structuraly related compounds. This treatment is based on the theory of Hilbert spaces and it extends the application of the factor analysis to the sets of continuous functions[12] which represent the CD spectra. The CD spectra have been evaluated by this method of a series of diastereoiso-

* Present address: Faculty of Mathematics and Physics, Department of Chemical Physics, Charles University, 121 16 Prague.

meric cyclodipeptides of the cyclo(L-Pro-X) type where X are the amino-acid residues of glycine, alanine, leucine, valine, and tert-leucine (*I*–*IX*). The results of this treatment have been discussed with respect to the method used as well as from the point of view of stereostructural considerations.



cis-series



trans-series

I, R = H [cyclo (L-Pro-Gly)]
II, R = CH_3 [cyclo (L-Pro-L-Ala)]
IV, R = $CH(CH_3)_2$ [cyclo (L-Pro-L-Val)]
VI, R = $CH_2CH(CH_3)_2$ [cyclo(L-Pro-L-Leu)]
VIII, R = $C(CH_3)_3$ [cyclo(L-Pro-L-Tle)]

III, R = CH_3 [cyclo(L-Pro-D-Ala)]
V, R = $CH(CH_3)_2$ [cyclo(L-Pro-D-Val)]
VII, R = $CH_2CH(CH_3)_2$ [cyclo(L-Pro-D-Leu)]
IX, R = $C(CH_3)_3$ [cyclo(L-Pro-D-Tle)]

THEORETICAL

Let us formulate first from the mathematical point of view the concept of the formation of experimental CD curves of a sufficiently large set of analogous compounds. In the first step we consider the examined compounds as pure chemical individua, forming a set of isolated objects. In the structure of these compounds, all the possible interactions are included governing effects which can be measured. The second step involves the choice of conditions for the measurements; many of the above mentioned interactions are manifest here. We shall confine only to those which can be studied by the CD spectroscopy. This restriction can be expressed by the conception that a set of structural, physical and chemical factors being the inherent property of the measured system (solution) leads to a formation of one or more chiral entities with electronic transitions yielding dichroic bands. These chiral entities are molecular segments, molecules, molecular aggregates, solvated molecules *etc.* With the series of analogous compounds we may expect only a limited number of chiral entities in the measured systems (*e.g.* in cyclodipeptides several conformers and solvates all of them containing the same chromophore – amide group). We can say that we have mapped the set of individual compounds onto a set of different chiral entities.

The third and final step represents the actual measurement of the CD curves. Each chiral entity which is present in the measured sample of a given compound is characterized by a certain CD curve; we shall call this curve the CD subspectrum.

An experimentally measured CD spectrum is therefore a superposition of the individual CD subspectra. The observed changes in the experimental CD curves can be interpreted as arising from changes in the participation of the chiral entities due to the molecular structure modifications of the measured substances or due to the changes of intra and intermolecular interactions. The original set of compounds is mapped onto a set of CD curves composed from the subspectra corresponding to the common chiral entities.

The possibility of the determination of subspectra follows from the fact that a set of n CD spectra $\Theta_i(\lambda)$ has the property of a real finite-dimensional Hilbert space $\boldsymbol{H}_p$ ($p \leqq n$) of functions $\Theta_i(\lambda)$ (ref.[13]). The operation of addition corresponds here to the principle of spectra superposition; zero element is the CD curve of an optically inactive sample; the inverse element to each CD curves exists: the CD curve of an enantiomer. A scalar product of two spectra $\Theta_i(\lambda)$ and $\Theta_k(\lambda)$ can furthermore be defined in the space $\boldsymbol{H}_p$:

$$\langle \Theta_i(\lambda) \mid \Theta_k(\lambda) \rangle = \int \Theta_i(\lambda) \, . \, \Theta_k(\lambda) \, \mathrm{d}\lambda \, . \tag{1}$$

Without the loss of generality, we can use the normalized spectra $\mathbf{Y}_i(\lambda)$ defined as

$$\mathbf{Y}_i(\lambda) = \frac{\Theta_i(\lambda)}{\| \Theta_i(\lambda) \|} , \tag{2}$$

where

$$\| \Theta_i(\lambda) \| = \left(\int [\Theta_i(\lambda)]^2 \, \mathrm{d}\lambda \right)^{1/2} . \tag{3}$$

The scalar product of the normalized spectra has then the meaning of their correlation coefficient[14],

$$\langle \mathbf{Y}_i(\lambda) \mid \mathbf{Y}_k(\lambda) \rangle = r_{ik} \, . \tag{4}$$

It holds $r_{ki} = r_{jk}$, $r_{ii} = 1$ and $r_{ik} \in \langle 0, 1 \rangle$, *i.e.* r_{ik} satisfy all the requirements for the properties of a scalar product of vectors in the Hilbert space $\boldsymbol{H}_p$. The finite dimension of this space guarantees its completeness[13]. The possibility of finding a decomposition of a set of experimental CD curves into subspectra follows from the theory of the orthogonal decomposition of the Hilbert space. Each vector from this space – *i.e.* each CD curve in our case – can be uniquelly expressed as a sum of its projections into mutually orthogonal (and disjunctive) subspaces $\boldsymbol{F}_j$ the direct sum of which is again the original space[13,15], *i.e.*

$$\mathbf{Y}_i(\lambda) = \sum_{j=1}^{p} \mathbf{y}_{ji}(\lambda) \, , \tag{5}$$

where

$$\mathbf{y}_{ji}(\lambda) \in \mathbf{F}_j\,, \quad \mathbf{H}_p = \sum_{j=1}^{p} + \mathbf{F}_j\,. \qquad (6a), (6b)$$

The inverse statement is that each vector (spectrum) $\mathbf{Y}_i(\lambda) \in \mathbf{H}_p$ can be uniquelly decomposed according to Eq. (5), satisfying conditions (6a) where (6b) represents the only "complete" decomposition of the space $\mathbf{H}_p$ of the CD spectra into subspaces generated by the subspectra $\mathbf{y}_{ji}(\lambda)$; according to the above described model, the $\mathbf{y}_{ji}(\lambda)$'s correspond to different chiral entities. In agreement with the physical meaning of the subspectra, they can only be realized either as the experimental CD curves directly or as their linear combinations. We shall introduce the subspectra in a mathematical form which would allow for a quantitative evaluation of the participation of the chiral entities in individual samples; we shall seek for $\mathbf{y}_{ji}(\lambda)$ in the form of a multiple of the net subspectra $\vartheta_j(\lambda)$:

$$\mathbf{y}_{ji}(\lambda) = \alpha_{ij}\,\vartheta_j(\lambda)\,. \qquad (7)$$

Our goal is to find a projection operator $\hat{O}$ (refs.[15-17]) which projects the experimental CD curves into the orthogonal subspaces formed by the net subspectra ϑ_j. It holds then that

$$\langle \vartheta_j \,|\, \vartheta_k \rangle = \delta_{jk}\lambda_j\,, \qquad (8)$$

where the meaning of λ_j will be given below. The symmetric matrix the elements of which are the scalar products of the pairs of the net subspectra ϑ_j (correlation matrix, matrix of the overlap integrals) will be a diagonal matrix in this case. It follows that the net subspectra ϑ_j $(j = 1, 2, \ldots, p)$ have the meaning of the eigenvectors of the Hermitian operator $\hat{R}$ represented in a certain basis of $\mathbf{H}_p$ by a symmetric matrix $\mathbf{R}$ of the correlation coefficients[13], see Eq. (4). The following relation holds[17] for the projection operator which projects an arbitrary vector from the space $\mathbf{H}_p$ into the basis formed by the eigenvectors of the operator $\hat{R}$ in the space $\mathbf{H}_p$:

$$\hat{R}\hat{O} = \hat{O}\mathbf{\Lambda}\,, \qquad (9)$$

where $\mathbf{\Lambda}$ is a diagonal matrix of the eigenvalues λ_j of the operator $\hat{R}$. By solving the eigenvalue problem for the correlation matrix (overlap matrix) we therefore obtain the matrix representation $\mathbf{O}$ of the projection operator $\hat{O}$. The matrix of the net subspectra $|\mathfrak{I}\rangle$ will be determined by the effect of the operator $\hat{O}$ on the matrix $\mathbf{Y}$ the columns of which are the experimental CD curves

$$|\mathfrak{I}\rangle = |\mathbf{Y}\rangle\,\mathbf{O}\,. \qquad (10)$$

The matrix $\boldsymbol{Y}$ can be expressed from Eqs (*5*) and (*7*) in the form

$$|\boldsymbol{Y}\rangle = |\vartheta\rangle\,\boldsymbol{\alpha}\,, \tag{11}$$

where $\boldsymbol{\alpha}$ is the matrix of the coefficients α_{ij} in Eq. (*7*). It is obvious that by substituting for $|\vartheta\rangle$ from Eq. (*10*) into Eq. (*11*), and making use of the symmetry of the correlation matrix $\mathbf{R}$, we obtain[11] for the matrix of the coefficients α_{ij}

$$\boldsymbol{\alpha} = \mathbf{O}^{-1} = \mathbf{O}^{\mathrm{T}}\,. \tag{12}$$

The following relation holds for the matrix of the scalar products of the subspectra according to Eqs (*10*) and (*9*):

$$\langle\vartheta\,|\,\vartheta\rangle = \mathbf{O}^{\mathrm{T}}\langle\boldsymbol{Y}\,|\,\boldsymbol{Y}\rangle\,\mathbf{O} = \mathbf{O}^{\mathrm{T}}\mathbf{R}\mathbf{O} = \boldsymbol{\Lambda}\,; \tag{13}$$

net subspectra ϑ_j are therefore the eigenvectors of the operator $\hat{R}$.

Eqs (*9*)–(*11*) represent a computation scheme for the decomposition of a set of CD curves into the subspectra corresponding to the initial assumptions on the chiral entities. The introduction of the formalism of the Hilbert space secures the unambiguity of this solution.

This treatment holds for a wide class of vectors suppose they satisfy the axioms defining the Hilbert space. The distance of a vector $|\boldsymbol{y}\rangle$ from its projection $|\boldsymbol{g}\rangle = \hat{O}\,|\,\boldsymbol{y}\rangle$ defined by the norm $\|\,|\boldsymbol{y}\rangle - |\boldsymbol{g}\rangle\,\|$ assumes the minimum values, *i.e.* the best least squares approximation to a vector $|\boldsymbol{y}\rangle$ in a particular subspace is given by[13,15] $\hat{O}\,|\,\boldsymbol{y}\rangle$. We can say also that the basis vectors under consideration are oriented so that projections of an arbitrary vector from $\boldsymbol{H}_p$ into their direction assume maximum values.

The method of principal components of the factor analysis[8,10] is used to evaluate sets of experimental data arranged in two-dimensional matrices. Columns and rows of these matrices represent different objects and conditions for which these data have been obtained. It is assumed[11,18] that each matrix element is given as a linear combination of a limited number of factors which are products of two cofactors. Each experimental point is considered as a random deviation of the observed (physical) effect from its zero or mean value. Information which is involved in the changes of the measured values can be described by means of the dispersion of these random data, *i.e.* by the normalized mean value of the sum of the squares of their deviations from the zero or mean value. From the geometrical point of view, this dispersion is identical with the square of the norm (magnitude) of a vector given by a series of m measured quantities. The factors (principal components) are determined in such a way that each calculated factor describes the maximum of the above indicated

dispersion with the conservation of correlations between individual data; in other words, projections of a series of m data into the directions of those factors are required to have a maximum value. Such a requirement leads to a solution of the eigenvalue problem of the correlation matrix, *i.e.*, to a solution of Eq. (*9*) in a matrix representation and to the determination of the principal components according to Eq. (*10*). Considerations having a different theoretical background – algebraical and statistical – therefore lead to the same solution of the problem which thus involves both aspects.

The basic difference in the physical meaning of results must be mentioned between the factor analysis in the case of discrete data and that of continuous functions (*e.g.*, the CD spectra). In the first case, each element of a series of m values (*i.e.*, of a column matrix) can be a different linear combination of the common components. It follows from this fact that the calculated factors have no direct physical meaning and transformation methods such as *e.g.* the target transformation[11] must be used to identify them. In algebraic terms, the projection operator $\hat{O}$ must be changed to transform experimental data into a set of physically meaningful data. On the basis of additional assumptions, these data are considered to be the desired factors. In the second case the matrix elements are the values of continuous functions (for computational purposes evaluated at discrete values of the argument) and within a given column all these elements represent the same linear combination of the principal components (vectors). These components have then directly the meaning of *e.g.* the subspectra.

The main contribution of the algebraic approach is the determination of the mathematical significance of the eigenvector matrix of the correlation matrix $\mathbf{R}$ as a representation of the projector $\hat{O}$ which converts the data into a set of mutually orthogonal vectors with the desired properties. The statistical approach makes it possible to see the significance of the calculated quantities in more detail; it also allows to solve problems related to the experimental errors of the measured data. The last mentioned aspect is related to the determination of the number of factors (subspectra). From the mathematical point of view[13], the number of subspectra equals according to Eq. (*6*) to the dimension of the space $\boldsymbol{H}_p$, *i.e.*, to the number of nonzero orthogonal, one dimensional subspaces $\boldsymbol{F}_j$. It can be found with the use of Eqs (*6*) and (*8*) that this number equals at the same time to the number of nonzero eigenvalues of the matrix $\mathbf{R}$. Each measurement is subject to an experimental error which in a random way individualizes each curve. There are therefore subspaces in $\boldsymbol{H}_p$ corresponding to these random changes of the individual spectra from a given set. A graphical comparison of the calculated and experimental curve has been used to separate the random vectors from the subspectra in processing the CD spectra. Different criteria for the determination of the number of factors have been given by Malinowski[19,20]. The subspectrum with the largest eigenvalue represents – from the point of view of a statistical approach[9] – the most significant deviation of the measured effect from the

zero value which corresponds to the optically inactive sample, *i.e.*, to such deviation which occurs practically in all experimental spectra from a given set. This is the CD curve of an arbitrary substance in a set deprived from those contributions which are changing in respect to the mean value within a set. These variable contributions are described by the remaining subspectra.

EXPERIMENTAL

The CD spectra have been measured[21] of the analytically pure cyclodipeptides *I*—*IX* (ref.[21]) in 2,2,2-trifluoroethanol and acetonitrile (Figs 1—3). Spectra were recorded with the Roussel–Jouan Dichrographe model CD 185/II in quartz cells with the optical path from 1 cm to 0·01 cm at room temperature (22—25°C). Concentrations of samples varied from 0·5 to 0·8 mg/2 ml (approximately $4 . 10^{-3}$M). Spectra have been converted into a digital form with a computer program[22] using the Hewlett–Packard 9862 A plotter on-line with the Hewlett–Packard 9830 A calculator. Spectra have been recorded at 1 nm interval in the 250—186 nm regions. Recorded values have been transformed into the units of molar ellipticity (deg cm^2 $dmol^{-1}$) and they have not been corrected for the index of refraction of the solvent.

The CD spectra have been processed with a modified program for the factor analysis of chemical data using a double precision arithmetics; the basic subroutine of the program[23] has been provided to us by Prof. D. G. Howery (City University of New York, U.S.A.). The integrals (*4*) have been calculated numerically using a Simpson method. The matrices of the subspectra (in the form of tables of intensities at individual wavelengths) and the matrix of the coordinates α have been converted into the memory of the Hewlett–Packard 9830 A calculator. calculated and experimental curves have been compared graphically on a Hewlett-Packard 9862 A plotter.

TABLE I
Criteria for the Determination of the Minimum Number of Subspectra

Subspectrum *j*	Eigenvalue	Partial dispersion in %, see[a]	Total dispersion see[b]	Mean deviation see[c]
1	16·38500	91·03	91·03	4 541
2	1·08190	6·01	97·04	2 522
3	0·40867	2·27	99·31	1 416
4	0·07102	0·395	99·70	923
5	0·03835	0·213	99·92	444
6	0·00949	0·053	99·97	284

[a] The percents of dispersion described by the *j*-th subspectrum; [b] the percents of dispersion described by the 1st till the *j*-th subspectrum; [c] values in molar ellipticities; mean deviation is given as the arithmetical mean value of the mean deviations of all experimental spectra and spectra calculated by adding the 1st till the *j*-th spectrum at each wavelength.

RESULTS

The experimental CD curves are given in Figs 1 – 3. The Cotton effects can be divided in all cases into two types which correspond to $\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$ transitions of the amide group. In the 185 – 210 nm region ($\pi \rightarrow \pi^*$ transition) a pair of negative and positive band can be found. The negative band is characterized by a variable position (from 185 to 193 nm) and intensity. The positive band is located at longer wavelengths (205 – 210 nm) and it is less intense in general; the variability of the position and intensity of this band is smaller in comparison with the negative band. The Cotton effects in the 210 – 260 nm region correspond to the $n \rightarrow \pi^*$ transition of the amide group; they exhibit a considerable variability of the shape characteristics.

The percentual description of the partial and total information involved in the individual subspectra and their sums expressed according to the theory of factor analysis[8-11] by the six greatest eigenvalues belonging to the most significant subspectra is given in Table I. It was found that four subspectra are sufficient to describe

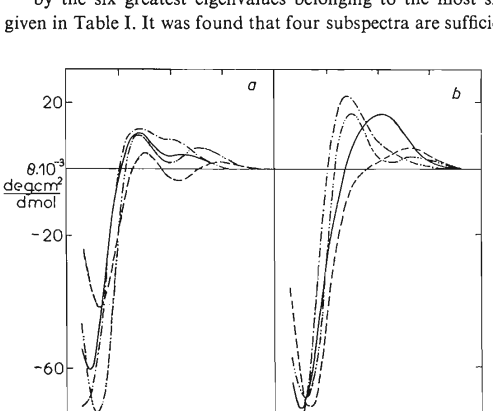


FIG. 1

CD Spectra of (*a*) Cyclo(L-Pro-L-Ala) (*II*) in 2,2,2-Trifluoroethanol(———) and in Acetonitrile (– – – –) and Cyclo(L-Pro-D-Ala) (*III*) in 2,2,2-Trifluoroethanol (—·—·) and in Acetonitrile (—··—··), (*b*) Cyclo(L-Pro-L-Val) (*IV*) in 2,2,2-Trifluoroethanol (———) and in Acetonitrile (— — —) and Cyclo(L-Pro-D-Val) (*V*) in 2,2,2-Trifluoroethanol (—·—·) and in Acetonitrile (—··—··)

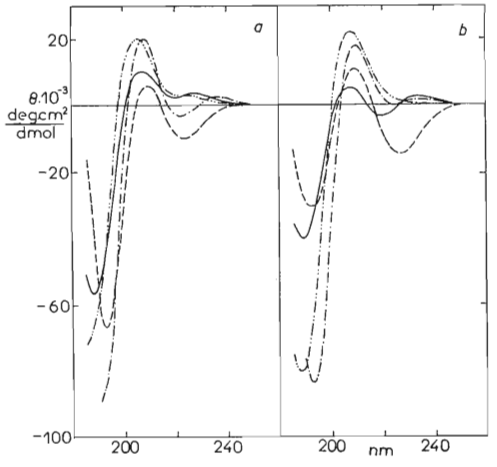


FIG. 2

CD Spectra of (*a*) Cyclo(L-Pro-L-Leu) (*VI*) in 2,2,2-Trifluoroethanol (———) and in Acetonitrile (— — —) and Cyclo(L-Pro-D-Leu) (*VII*) in 2,2,2-Trifluoroethanol (— ·· — ··) and in Acetonitrile (— · — ·) (*b*) Cyclo(L-Pro-L-Tle) (*VIII*) in 2,2,2-Trifluoroethanol (———) and in Acetonitrile (— — —) and Cyclo(L-Pro-D-Tle) (*IX*) in 2,2,2-Trifluoroethanol (— ·· — ··) and in Acetonitrile (— · — ·)

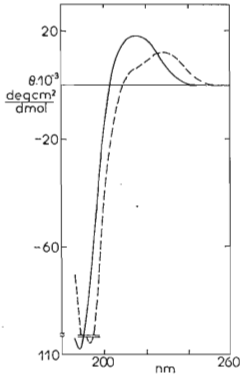


FIG. 3

CD Spectra of Cyclo(L-Pro-Gly) (*I*) in 2,2,2-Trifluoroethanol (full curve) and in Acetonitrile (dashed curve)

eighteen experimental CD curves with the experimental accuracy[12,24]* (Fig. 4). Starting with the fifth subspectrum, the subspectra can be considered not to be significant.

For the sake of comparison of the relative participations of the individual chiral entities in different samples, we define the weight of the j-th subspectrum in the i-th CD curve by the expression

$$W_{ij} = \frac{\alpha_{ij}}{\sum_{j=1}^{4} |\alpha_{ij}|} . 100, \qquad (14)$$

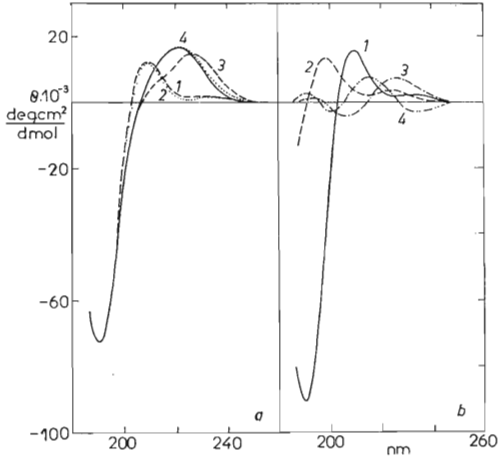


FIG. 4

(*a*) Comparison of the Experimental CD Curve of *IV* in 2,2,2-Trifluoroethanol (full curve) with the Sum Curves of the First Subspectrum (dashed curve *1*), the First Two Subspectra (dotted curve *2*), the First Three Subspectra (dashed curve *3*) and All Four Subspectra (dotted curve *4*) (*b*) Decomposition of the Experimental CD Curve of the Substance *I* in 2,2,2-Trifluoroethanol into Subspectra *1*—*4*

* A graphical comparison of the calculated and experimental values is an independent method for the determination of the number of subspectra. This method makes it also possible to estimate the probable experimental error of measurement. In the present case we obtain a dependence of the error of measurement on the wavelength which is in complete agreement with the dependence obtained by a statistical processing of one hundred repeated measurements[24].

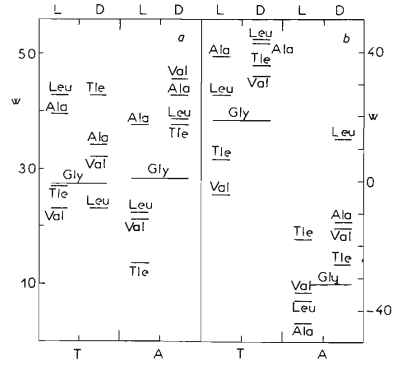


FIG. 5

Comparison of the Weights of the First (*a*) and Second (*b*) Subspectrum for Individual Substances from the Studied Set (symbols L and D denote the absolute configuration of the non-proline amino-acid residue indicated by the symbol) and for Both Solvents

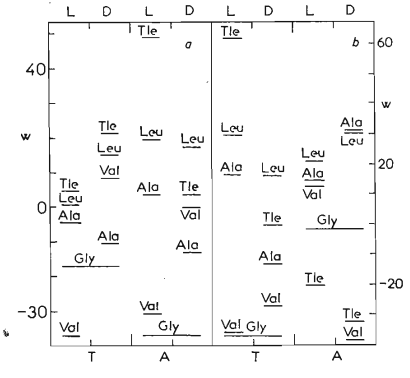


FIG. 6

Comparison of the Weights of the Third (*a*) and Fourth (*b*) Subspectrum for Individual Substances from the Studied Set (symbols L and D denote the absolute configuration of the non-proline amino-acid residue indicated by the symbol) and for Both Solvents

where W_{ij} is the percentual contribution of the subspectrum to the total optical activity of a given sample. The total value is expressed within the whole set by means of a sum of the absolute values of the coefficients of the subspectra which are necessary for a description of this CD curve. In the sense of the model introduced above, this quantitative characteristics is connected with the structural features of the measured system. This is obvious from Figs 5 and 6 where the relative distribution of the weights of individual subspectra is given for the processed set of spectra.

The calculated first subspectrum determines the common features of all CD curves of compounds *I*–*IX* in both solvents: in the region of the $\pi \rightarrow \pi^*$ transition there is an intense negative Cotton effect at 190 nm and a positive Cotton effect at 209 nm. The $n \rightarrow \pi^*$ region is modelled exclusively by further subspectra. The weights of the first subspectrum are positive in all cases; in *cis*-disubstituted cyclodipeptides, their values for a given X are lower in acetonitrile than in 2,2,2-trifluoroethanol. The inverse is true for the *trans*-disubstituted cyclodipeptides cyclo(L-Pro-D-X) except compound *IX* (Fig. 5*a*). There is no apparent relation to the character of the side chain.

The second subspectrum contributes to the description of the experimental curve in the short wavelenght $\pi \rightarrow \pi^*$ region and in the $n \rightarrow \pi^*$ region. The second subspectrum differs from all others by a pregnant dependence of the sign of its weight on the solvent: there are positive weights in 2,2,2-trifluoroethanol (except compound *IV*), negative weights in acetonitrile (except compound *VII*). For a given solvent, the weight of the second subspectrum depends on the structural properties of the molecule alone. In the series of *cis*-isomers, the effect of the side chain is so pregnant that the sequence of compounds is reversed in trifluoroethanol and acetonitrile. In the *trans*-series, the sequence is not changed by the choice of the solvent and at least in 2,2,2-trifluoroethanol, the weights of the second subspectrum are very similar in all compounds (Fig. 5*b*).

The third subspectrum describes contributions to the long wavelength $\pi \rightarrow \pi^*$ band and to the $n \rightarrow \pi^*$ region. The fourth subspectrum describes solely contributions in the $n \rightarrow \pi^*$ region. Certain complementarity of these subspectra is obvious from the comparison of the weights (Figs 6*a*, 6*b*) in *cis*-disubstituted cyclodipeptides: the relative participation of the third subspectrum changes in the largest interval in acetonitrile while the largest changes of the fourth spectrum occur in trifluoroethanol. The relative changes in the participation of the subspectra of the *trans*-isomers are less pregnant going from trifluoroethanol to acetonitrile. However, a detailed discussion of the quantitative results is not justified. Even the relatively large weights of the third and fourth subspectra need not mean their significant contribution to the CD curves.

These results are stable with respect to deletion of a certain number of input experimental CD curves. Identical results have been obtained by processing a set from which the CD curves have been removed of the substances *II*, *III* and *VII* measured in both solvents.

For experimental verification, systems with a significant majority of one chiral entity would be needed. In these cases only the direct comparison could be performed of the experimental CD curves with the calculated subspectra. In principle, the relative population of chiral entities differing in energy can be influenced by the temperature changes. However, several simultaneous processes (conformation changes, solvation *etc.*)[25] can result from temperature changes and chiral entities which are present in the solution at lower temperatures need not be identical with the chiral entities at higher temperatures, although their number can be lower.

In the set of substances *I*–*IX* it was possible to use a special position of the CD spectrum of cyclo(L-Pro-L-Tle) (*VIII*) in 2,2,2-trifluoroethanol. It follows from the calculation that this spectrum is described solely by the contributions from the first and fourth subspectrum. On the other hand it can be assumed that due to the large steric effect of the tert-butyl group in compound *VIII*, the formation of a large number of chiral entities will be considerably reduced in this case. It seems therefore reasonable to assign the chiral entity giving the fourth subspectrum to the chiral entity predominant in the solution of compound *VIII* in trifluoroethanol. With the decreasing temperature, the CD curve of *VIII* changes considerably; the longest wavelength positive band disappears and the intensity of the negative band at 226 nm increases significantly. The same effect can be observed also for cyclo(L-Pro-L-Leu) (*VI*) in trifluoroethanol. The CD curve of this compound contains 27% contribution from the fourth subspectrum with the same sign as in compound *VIII*. On the contrary, if the calculated fourth subspectrum was found to have the opposite sign with respect to the sign of cyclodipeptides *VIII* and *VI* (compounds *I*, *IV*, *V*, and *IX*), the intensity of the positive band increases in the investigated wavelength region with the decreasing temperature[21]. It was therefore possible to distinguish compounds *VIII* and *VI* from other substances of the set.

In a similar way we could assign the chiral entity giving the third subspectrum to the chiral entity which predominates in the acetonitrile solution of *VIII* at room temperature because the experimental curve is contributed under these conditions almost exclusively by the first and third subspectrum.

DISCUSSION

Let us now characterize those structural features of cyclodipeptides *I*–*IX* which lead to a high degree of similarity between their CD curves expressed by the first subspectra. The investigated 2,5-piperazindiones *I*–*IX* are formed by the two homoconjugated *cis*-amide groups. Molecular fragment involving the pyrrolidine ring and the amide group including its nitrogen atom is the decisive factor introducing the main conformational feature into the molecule, *i.e.* the restriction of the flexibility of the cyclodipeptide ring and its fixation in one of the possible chiral boat forms[26].

In this conformation, not only the proline side chain is in an unfavourable interaction with the oxygen atom of the amide group but also the *cis*-oriented substituent at the second α-carbon atom. The *trans*-substituent is in a more favourable pseudoaxial position. The second boat conformation is not probable as follows from the results of X-ray analyses[26,27], chemical reactions[28,29], and model considerations. The X-ray studies as well as the calculations of the energy of conformers for different cyclodipeptides containing L-proline residue[26,30] confirm that the steric arrangement of the pyrrolidine ring as well as that of the 2,5-piperazindione ring is roughly the same in all these compounds. In agreement with the concept of the basic chiral entity, *i.e.* of the cyclodipeptide ring in the boat conformation of a given molecular chirality, the calculated first subspectrum is identical with the experimental CD curve of cyclo(L-Pro-D-Val) (*V*) in acetonitrile. In this solution a simple situation can be assumed: *i*) side chain in the energetically favourable pseudoaxial position which does not require a special arrangement of the rest of the molecule; *ii*) interaction with a solvent having limited sterical requirements.

Solvent contributions to the experimental CD curves are further described by the subspectra 2–4. The infrared-spectra investigation showed that both solvents can form solvates with cyclodipeptides[31]. 2,2,2-Trifluoroethanol as a proton donor can be hydrogen-bonded to the oxygen atoms of both amide groups; its role as a proton acceptor is not significant. Acetonitrile, on the contrary, acts as a proton acceptor in interaction with the N—H group.

There is no experimental curve which would correspond to the second subspectrum. This reflects probably the presence (in 2,2,2-trifluoroethanol) or absence (in acetonitrile) of solvent molecules bound to the proline carbonyl oxygen. Steric interactions of the bound solvent can cause local changes of the conformation in the area of the proline residue in all compounds from the set; electronic factors can further modify the geometry of the amide group. The different nature of this action is described by different signs of the second subspectrum in 2,2,2-trifluoroethanol and in acetonitrile. Interactions of the solvent with groups in the neighbourhood of the non-proline amino-acid residue exhibit in *cis*-isomers *II*, *IV*, *VI*, and *VII* a steric interaction of the bonded solvent molecules with energetically less favourably located non-proline side chain. The consequence of this interaction may be the twisted boat conformation with non-planar amide groups[32]. The sense of the twist should be enantiomeric for the solvent attached to the carbonyl group in comparison with the solvent attached to the NH group. We assume that these effects determine the order of participation of the second subspectrum in dependence on the side chain bulkiness within the series of *cis*-isomers. The special nature of the solvent interaction manifests in the third subspectrum for acetonitrile, and in the fourth subspectrum in 2,2,2-trifluoroethanol. In the case of *trans*-isomers *III*, *V*, *VII*, and *IX*, these steric interactions with the bonded solvent cannot be significant. This is in agreement with the small participation of the third and fourth subspectrum in the

CD curves of these compounds and with the fact that the weights of the second subspectrum are close.

From the methodical point of view, the above described decomposition of the set of experimental CD spectra into subspectra represents an alternative to a more customary decomposition of the experimental CD spectra into individual dichroic bands[33,34]. Treatment which is supported by the theory of Hilbert spaces allows to use as a main source of information the total change of all the CD curves of a set and their mutual correlations. Only experimental data are involved in the calculation and it is not necessary to introduce assumptions on the number, shape and parameters of the dichroic bands. In that respect the mathematical method described in this paper approaches the usual empirical evaluation of spectral data of a series of analogous compounds. Confrontation of the results of the mathematical analysis with the empirical analysis of the studied series of compounds has shown an agreement of basic conclusions on the conformation of these substances. However, the true advantage of the presented approach to the evaluation of CD data lies in the possibility to quantify the effects of some structural factors.

We are grateful to Prof. D. G. Hovery, the City University of New York, U.S.A., for providing us the program for factor analysis. We consider as highly valuable comments to this paper from Drs J. Štěpánek and V. Mikoláš, Mathematical and Physical Faculty, Charles University, Prague.

REFERENCES

1. Crabbée P.: *Optical Rotatory Dispersion and Circular Dichroism in Organic Chemistry*. Holden-Day, San Francisco 1965.
2. Ivanov V. T. Lavrinovich I. A., Kogan G. A., Meshcheryakova E. A., Portnova S. L., Isabaev M. D., Kostetskii P. V., Ovchinnikov Yu. A.: Khim. Prir. Soedin. *11*, 202 (1975).
3. Ritter G. L., Lowry S. R., Isenhour T. L., Wilkins C. L.: Anal. Chem. *48*, 591 (1976).
4. Malinowski E. R., McCue M.: Anal. Chem. *49*, 284 (1977).
5. Rozett R. W., Petersen E. M.: Anal. Chem. *47*, 2377 (1975).
6. Daves J. E., Shepard A., Stanford N., Rogers L. B.: Anal. Chem. *46*, 821 (1974).
7. Justice J. B., Isenhour T. L.: Anal. Chem. *47*, 2286 (1975).
8. Rummel R. J.: *Applied Factor Analysis*. Northwestern University Press, Evanston 1970.
9. Harman H. H.: *Modern Factor Analysis*. University Chicago Press, Chicago 1967.
10. Überla K.: *Faktoren-Analyse*. Springer, Berlin 1971.
11. Weiner P. H., Malinowski E. R., Levinstone A. R.: J. Phys. Chem. *74*, 4537 (1970).
12. Pančoška P.: *Thesis*. Charles University, Prague 1977.
13. Fano G.: *Mathematical Methods of Quantum Mechanics*, Chap. 5. Mc Graw-Hill, New York 1971.
14. Magar M. E.: *Data Analysis in Biochemistry and Biophysics*. Academic Press, New York 1972.
15. Roby K. R. in the book: *Wave Mechanics, the First Fifty Years* (W. C. Price, S. S. Chissick, T. Ravensdale, Eds), Chap. 4. London 1973.
16. Halmos P. R.: *Finite Dimensional Vector Spaces*. Van Nostrand, New York 1958.
17. Lövdin P. O.: Preprint No 125, Quantum Chemistry Group, Uppsala, Sweden 1964.

18. Hovery D. G.: Int. Lab. *1976*, March/April, 11.
19. Malinowski E. R.: Anal. Chem. *49*, 606 (1977).
20. Malinowski E. R.: Anal. Chem. *49*, 612 (1977).
21. Frič I., Pančoška P., Bláha K.: Unpublished results.
22. Vítek A.: Program No 12. Program Library, Institute of Organic Chemistry and Biochemistry, Czechoslovak Academy of Sciences, Prague.
23. Malinowski E. R.: *Thesis*. Stevens Institute of Technology, Hoboken, U.S.A. 1969.
24. Pančoška P., Frič I., Bláha K.: This Journal *44*, 1312 (1979).
25. Snatzke G., in the book: *Optical Rotatory Dispersion and Circular Dichroism in Organic Chemistry* (G. Snatzke, Ed.), p. 335. Heyden, London 1967.
26. Ramani R., Sasisekharan V., Venkatesan K.: Int. J. Peptide Protein Res. *9*, 277 (1977).
27. Madison V., Young P. E., Blout E. R.: J. Amer. Chem. Soc. *98*, 5358 (1976).
28. Vičar J., Bláha K.: This Journal *38*, 3304 (1973).
29. Liberek B., Bednarek M., Kitowska A., Kozlowska M., Macikowska A. in the book: *Peptides* 1976, *Proceedings of the 14th European Peptide Symposium*, Wépion 1976 (A. Loffet, Ed.), p. 641. Editions de l'Université de Bruxelles, Brussels 1976.
30. Young P. E., Madison V., Blout E. R.: J. Amer. Chem. Soc. *98*, 5365 (1976).
31. Smolíková J., Bláha K.: Unpublished results.
32. Frič I., Maloň P., Tichý M., Bláha K.: This Journal *42*, 678 (1977).
33. DeTar D. F.: *Computer Programs for Chemistry*, Vol. IV. Academic Press, New York 1972.
34. Madison V., Schellman J.: Biopolymers *9*, 511 (1970).

Translated by D. Papoušek.