STATISTICAL ACCURACY TEST IN MEASUREMENTS OF CIRCULAR DICHROISM

Petr PANČOŠKA*, Ivo FRIČ and Karel BLÁHA

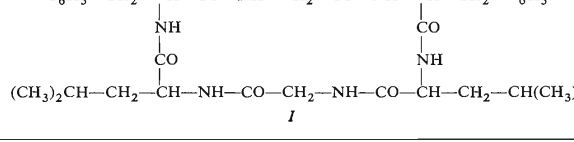
Institute of Organic Chemistry and Biochemistry, Czechoslovak Academy of Sciences, 166 10 Prague 6

Received May 3rd 1978

A collection of one hundred circular dichroism measurements (in acetonitrile) of the model cyclohexapeptide cyclo(glycyl-L-phenylalanyl-L-leucyl-glycyl-L-leucyl-L-phenylalanyl) has been evaluated using standard statistical methods. The reproducibility of the experimental intensity values is discussed with respect to the sample concentration and the particular regions of the CD curve. The maximum probable error in the intensity reading varies with wavelength from ± 20 to ± 10000 deg cm^2 dmol^{-1} in the spectral range from 274 to 196 nm, *i.e.* by 5-15% of the observed intensity.

In interpreting circular dichroism (CD) spectra, the intensity of the observed phenomenon plays an important role[1]. In order to make use of this spectral information, it is necessary to know the error in the determination of intensity both for various points or segments of the curve, and for various conditions of recording. A procedure has been recently proposed for testing the accuracy of CD spectral measurement utilising a combination of standard statistical methods. On confrontation with experimental data, this method makes it possible to express the results of calculations in the form of practically useful conclusions (optimal conditions for the measurement, significance test of quantitative data, *etc.*).

For this statistical study, cyclo(glycyl-L-phenylalanyl-L-leucyl-glycyl-L-leucyl-L-phenylalanyl) (*I*) was selected as a model substance[2,3]. The CD curve of this cyclohexapeptide exhibits dichroic bands representing $n-\pi^*$ and $\pi-\pi^*$ electronic transitions of several amide groups, as well as those of aromatic chromophores.



I

* Present address: Institute of Mathematics and Physics, Department of Chemical Physics, Charles University, Prague.

Similar to other cyclopeptides, compound *I* is a suitable model for investigations of the spatial arrangement of peptide chains[4,5]. Furthermore, CD curves of cyclohexapeptide *I* are representative of common types of CD spectra of peptides and proteins[6], and may be considered as a typical example.

EXPERIMENTAL

CD spectra were measured on a Roussel–Jouan Model CD-185/II apparatus at room temperature in non-thermostated quartz cells with an optical pathlength of 1 cm, and 0·01 cm. The maximum sensitivity ($1 . 10^{-5}$ of dichroic absorption units per millimeter of reading on the abscissa scale), a scanning rate of 0·125 nm s^{-1} and a time constant of 4 s were used. Acetonitrile solutions of compounds *I* were prepared in ten different concentrations (Table 1), and ten CD spectral records of each solution were taken. In each measurement, the random-noise level was lowered by repeating the record. For all measurements at wavelengths shorter than 200 nm, the optical system of the apparatus was flushed with nitrogen. The spectra were digitized by means of a programme and a Hewlett–Packard 9862A XY-plotter as digitizer, in combination with a Hewlett–Packard 9830A calculator. The programme makes possible vertical differentiation of 0·25 nm and ensures minimum error in determination of the wavelength. The reading interval was 2 nm in the 274—196 nm range. With the use of an additional programme[2], the data were transformed to molar ellipticities deg cm^2 $dmol^{-1}$. This programme also contains corrections for irregularities in the curve where one spectral section overlap with a contiguous one measured in a cell of a different thickness.

RESULTS AND DISCUSSION

From a sample space of four thousands molar ellipticity values (one hundred spectra, forty molar ellipticity values each), the following data were calculated*: total arithmetic means R_j for each wavelength considered (from one hundred measurements of the CD curve intensity at a given point), partial arithmetic means R_{ij} (from repeated intensity measurements with the respective sample concentrations), corresponding total and partial variances, standard deviations, and total and partial confidence intervals with the use of critical values of the normal or Student's distribution[7,8].

In the first part, accuracy and reproducibility of CD spectral record were examined at various concentrations, and the existence of a most advantageous concentration for the present purpose was verified. It was assumed that no solute aggregation occured in the sample solutions. As the criterion of accuracy, the variance of results was used in ten repeated curve evaluations for each concentration. By means of the F-test[7,8], the 45 possible pairs of variances of these results were examined at each measurement point, and the acceptance or rejection of the null hypothesis on the identity of variances was determined, *i.e.*, it was investigated whether the difference

* In application of the estimate theory and the theory of hypothesis testing, a 99% confidence level was always used.

in accuracy of determinations of the intensity of the CD spectrum in a given pair of concentrations is statistically significant or insignificant. The total number of spectral points with a statistically significant difference of variances for the given combination of concentrations is shown in Table I (F-test). Information on the reproducibility of CD spectral records taken under the same conditions with solutions of different concentrations was supplied by the t-test[7,8] on the identity of selected means, represented by partial arithmetic means R_{ij}. The calculation was programmed analogously to the F-test, the expressions for the determination of critical regions being modified according to the actual situation into the form for an identical, as well as statistically significantly different, variance. The frequencies of statistically significantly different arithmetic means R_{ij} for the respective combination of concentrations are shown in Table I (t-test).

It may be inferred from the results of the F-test that the accuracy of the measurement at the lowest concentration cannot be compared with that at the other concentrations. At the lowest concentration, significant differences in variances amounted to 10–60% of the total number of points. The other concentrations with 0–15% of significant differences were equivalent and acceptable for the present purpose. The distribution of significant differences in accuracy with respect to wavelengths was statistically uniform over the whole spectral region; in other words, using variances as a criterion of accuracy, the effects of the shape of the CD curve, of the recording technique or of the noise cannot be evaluated in detail[2].

TABLE I

Rejection Frequencies of the Null Hypothesis on Identity of Variances (F-test, above the diagonal) and on Identity of Means (t-test, under the diagonal) in Sets of 10 Measurements at Various Concentrations

Concentration mg/2 ml	0·450	0·720	0·860	1·060	1·340	1·500	1·727	1·860	1·927	2·300
0·450	—	4	11	7	9	14	24	14	16	17
0·720	0	—	1	1	1	2	6	4	3	4
0·860	14	6	—	1	3	2	6	4	5	2
1·060	7	3	6	—	1	2	3	2	5	2
1·340	4	5	10	1	—	1	3	0	1	0
1·500	7	3	12	2	0	—	5	4	4	2
1·727	7	10	15	1	0	2	—	3	2	2
1·860	4	6	16	6	1	4	0	—	0	0
1·927	11	15	15	7	6	2	6	12	—	0
2·300	8	14	14	7	2	3	2	6	2	—

More detailed information was supplied by processing the results of the t-test. The differences between the selected concentration of the sample and other concentrations were correlated with the frequency of statistically significant differences in R_{ij} values. The occurrence of minima in these dependences and their positions confirm the existence of an optimum concentration (in the present case, 1·0 to 1·5 mg of the cyclohexapeptide *I* per 2 ml of the solution, *i.e.*, an approximately 0·005M solution referred to one amino-acid residue). As shown by analysis of the frequency distribution of statistically significant differences in the respective R_{ij} values with respect to the wavelengths (Fig. 1), the reproducibility of quantitative results was very low in some segments of the CD curve. This may be ascribed to the following features of the CD curve: *i*) Low absolute value of the reading (at about 248 nm on Fig. 1), *ii*) relatively high noise, and *iii*) merging of spectral sections measured in cells of different optical path length (at about 220 nm in Fig. 1). These results confirm in general (quantitatively in the case of substance *I*) concepts on alterations in reproducibility of results of CD spectral measurements.

It was further goal to investigate the limits of error of readings of intensity of CD curves in practice, when only small amounts of substance are available or time does not permit large numbers of measurements for statistical evaluation. According

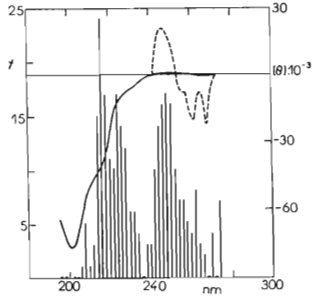


FIG. 1

Frequency Distribution of Statistically Significant Differences of R_{ij} Values in Sets of Ten Measurements at Various Concentrations

The dashed line denotes a fifty-fold expansion of the corresponding part of the curve.

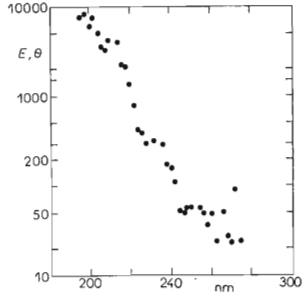


FIG. 2

Dependence of the Maximum Probable Error on the Wavelength in Determinations of the CD Curve Intensity

Numerical error values were obtained from confidence intervals around R_j for a set of measurements at an optimum concentration.

to the suggested methods firstly the CD spectrum, in terms of arithmetic means R_j, was taken as equivalent to the „real" CD curve. This idealised curve therefore had no experimental error in the given intensity range. (The R_j values are the best attainable estimates of ideal values of molar ellipticities, error in R_{ij} is 1/10th of that of separate intensity measurements). Furthermore, the confidence intervals about R_j were calculated from critical values of Student's distribution in sub-groups measured at the same concentration – *i.e.* with the same distribution of experimental error.

In the second part, we examined the probability (using relative frequency as an estimate[7]) with which the reading of intensity will fall into the corresponding confidence interval. It was found that about 60–70% of all observed values lie within these intervals, and that about 90–100% of all intensity values lie within twice the interval (irrespective of the wavelength in both cases). It can therefore be argued that if results from 100 CD intensity readings fall within a given limit, it is highly probable (according to calculations[2] this probability approaches 1·0) that this interval will also contain the result of an arbitrary CD intensity measurement of the same (or of a similar) compound. Consequently, the magnitude of these intervals can be considered as a rough estimate of the maximum probable error of the particular measurement. In accordance with Part 1, the magnitude of this error depends on concentration. Fig. 2 shows the dependence of the probable error (calculated from double the confidence interval values) on the wavelength for the optimum concentration. The shape of dependencies in the case of other concentrations is similar. The magnitude of relative errors, calculated from probable errors, varies from 5 to 12% (low intensities near the cross-point at 240 nm being neglected).

On the basis of this test, the effect of experimental conditions (particularly that of concentration) on the accuracy of numerical values of ellipticities may be evaluated. Part 2 of the test may be regarded as a basis for discussion of quantitative aspects of CD phenomena in the field of peptides and proteins.

We can mention here the following example: The value of the mean experimental error, calculated in this paper as the arithmetic mean of the probable experimental error ($1\,000$ deg cm^2 dmol^{-1}), has been used for the estimation of the number of main components in the factor analysis of CD spectra of the proline containing cyclodipeptides[9]. The estimation of the probable range of error should be valid for each single measurement of CD curves within the studied structural type of compound.

REFERENCES

1. Caldwell D. J., Eyring H.: *The Theory of Optical Activity*. Wiley-Interscience, New York 1971.
2. Pančoška P.: *Thesis*. Charles University, Prague 1977.
3. Yang Chen-su, Bláha K., Rudinger J.: This Journal *29*, 2633 (1964).
4. Bláha K., Frič I., Rudinger J.: This Journal *34*, 3497 (1969).
5. Bláha K., Buděšínský M.: Int. J. Peptide Protein Res. *5*, 399 (1973).
6. Jirgensons B.: *Molecular Biology, Biochemistry and Biophysics*, Vol. 5. *Optical Activity of Proteins and Other Macromolecules*, 2nd Ed. Springer, Berlin 1973.
7. Korn G. A., Korn T. M.: *Mathematical Handbook for Scientists and Engineers*. McGraw-Hill, New York 1968.
8. Magar M. E.: *Data Analysis in Biochemistry and Biophysics*. Academic Press, New York 1972.
9. Pančoška P., Frič I., Bláha K.: This Journal *44*, 1296 (1979).