J|A|C|S
COMMUNICATIONS

Published on Web 01/22/2003

Determination of the Enantiomeric Composition of Guest Molecules by Chemometric Analysis of the UV−Visible Spectra of Cyclodextrin Guest−Host Complexes

Kenneth W. Busch, Isabel Maya Swamidoss, Sayo O. Fakayode, and Marianna A. Busch*

Department of Chemistry and Biochemistry, Baylor University, P.O. Box 97348, Waco, Texas 76798

Received February 15, 2002; E-mail: Marianna_Busch@baylor.edu

The need for improved strategies for the assessment of enantiomeric purity arises from increased pressure on the pharmaceutical industry by government agencies for documentation on the pharmacological effects of individual enantiomers and the simultaneous demand in drug development for determination of enantiomeric excess in large combinatorial libraries.[1] For high throughput screening strategies, slow chromatographic methods are not attractive, and rapid spectroscopic techniques appear most promising.[2] In this study, we report the development of a new strategy for the quantitative determination of enantiomeric purity that combines UV spectroscopy, cyclodextrin guest−host complexation, and chemometric modeling.

Previous work on the development of rapid enantioselective sensors has revealed that chromogenic enantioselective chiral hosts are capable of discriminating between enantiomers of chiral guests through a change in the visible absorption spectrum of the enantioselective complex (i.e., a color change).[3] Under this strategy, the complexation of one enantiomer of a chiral substrate with a chiral host results in a visible spectral shift and/or the formation of an entirely new visible band, while little or no color change is observed when the other enantiomer complexes with the chiral host.

Because chromogenic enantioselective response of the type just described need not be restricted to the visible region of the spectrum, we thought it worthwhile to determine whether similar spectral effects could be observed in other spectral regions. To test this hypothesis, we chose to study the changes in the UV spectra of guest−host complexes formed with various cyclodextrins.

Cyclodextrins (CDs) are homochiral barrel-shaped sugar molecules that can form transient diastereomeric guest−host complexes with chiral guest molecules. Because the complexes that form are diastereomeric, they have different physical properties,[4] which should, in turn, result in small changes in their spectra.

In general, we have observed that when a guest molecule is added to an aqueous solution of cyclodextrin to form a guest/host complex, a weak, more or less prominent, tail or shoulder is observed to form on the long-wavelength side of the absorption band of the complex. It is in this tail or shoulder region where the effect of enantiomeric composition of the guest molecule has the most pronounced effect on the spectrum.

Figure 1 shows spectral data (250−500 nm) in this shoulder region for a series of nine aqueous solutions (pH 12, selected for solubility considerations) containing a fixed concentration of β-CD (30 mM) and a fixed concentration (15 mM) of 2-phenylglycine (ϕ-Gly). While the total concentration of ϕ-Gly was fixed, the enantiomeric composition of the ϕ-Gly used to prepare the solutions was varied from mol fraction [(*R*)-ϕ-Gly] 0.5 to 0.9, using different amounts of commercially available enantiomers (99%, Aldrich, Milwaukee, WI).

Figure 1 clearly shows that the spectra obtained for the different samples depend on the enantiomeric composition of the ϕ-Gly guest molecule. In general, however, small spectral variations of the type shown in Figure 1 are often dismissed as having little, if any, utility for predicting the composition of a sample because the variations are small, the bands all overlap, and the spectra do not appear to show a consistent trend (such as an offset) with composition. Thus, while Figure 1 clearly demonstrates that spectral changes occur as a result of differences in the enantiomeric composition of the ϕ-Gly guest molecule, it is difficult to discern by simple visual inspection how these spectral and compositional changes are correlated. In such situations, chemometric methods, such as multivariate regression, offer a variety of powerful techniques for revealing hidden relationships in data that are not immediately apparent.[5]

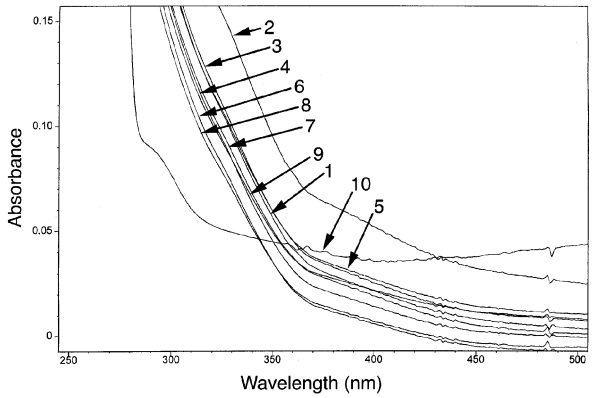

Figure 1. Absorption spectra of solutions (pH 12) containing 30 mM β-CD and 15 mM ϕ-Gly of various enantiomeric composition (**1**−**9**). Mol fraction of *R*-ϕ-Gly: **1**, 0.460; **2**, 0.500; **3**, 0.566; **4**, 0.600; **5**, 0.634; **6**, 0.700; **7**, 0.800; **8**, 0.854; **9**, 0.900. Spectrum of 30 mM β-CD (**10**).

In the present study, spectral information was treated by a partial least-squares (PLS-1) approach. The PLS-1 algorithm is especially powerful as a means of regression because both the *X*- and the *Y*-data are actively involved in the construction of the new basis set made up of PLS components. In this way, the PLS regression algorithm focuses on those aspects of the data that are most important in predicting *Y*.

Multivariate modeling is a two-step procedure. In the first or calibration phase, a mathematical model in the form of a regression vector is determined with a training set of samples whose *Y*-variable is known. Equation 1 shows the form of the regression vector

$$X_R = k_0 + k_1A_1 + k_2A_2 + ... + k_nA_n \qquad (1)$$

where X_R is the unknown mol fraction of *R*-ϕ-Gly in the sample, k_i are the coefficients of the regression vector, and A_i are the absorbances at the different *i* wavelengths ($i = 1, ..., n$) for a given unknown sample.

In our study, the training set consisted of the measured spectral data shown in Figure 1 along with the known enantiomeric

10.1021/ja025947a CCC: $25.00

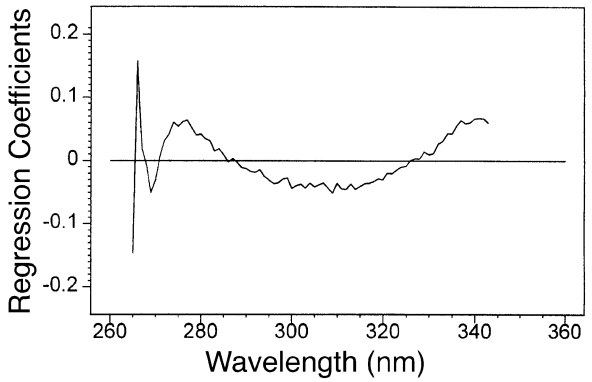

Figure 2. Regression coefficients for the PLS-1 model for *R*-ϕ-Gly as a function of wavelength from 260 to 345 nm at pH 12.

Table 1. Prediction Results Obtained with the Regression Model for *R*-ϕ-Gly

known mol fraction of *R*-ϕ-Gly	predicted mol fraction of *R*-ϕ-Gly	relative error, %
0.528	0.515	−2.46
0.620	0.581	−6.29
0.673	0.673	0
0.727	0.753	3.58
0.827	0.873	5.56
0.873	0.873	0

composition of the laboratory-prepared samples. Figure 2 shows the regression coefficients as a function of wavelength for the model for *R*-ϕ-Gly. A similar plot of the regression coefficients obtained for *S*-ϕ-Gly is C_2-symmetric about the abscissa (see the Supporting Information).

For the example cited in this study, a plot of enantiomeric composition of the ϕ-Gly samples predicted by the model versus their known enantiomeric composition gave a straight line with a correlation coefficient of 0.955, a slope of 1.05, and an offset of -5.61×10^{-4}. A perfect model would have a correlation coefficient of 1, a slope of 1, and an offset of zero.

While the model obtained for the ϕ-Gly samples is impressive, the real test of any regression model lies in its ability to predict future samples. In the second or validation phase, the mathematical model developed for the training set of samples is used to predict the enantiomeric composition of another independently obtained set of samples whose enantiomeric composition is also known. Here, the spectral data for the validation set of samples are obtained, and eq 1 is used to predict the enantiomeric composition of the samples in the set from the measured spectral data. In this phase, the values of the *Y*-data predicted by the model are compared with the known values for the validation set. Table 1 shows the results obtained when the regression model developed in phase I was used to predict the enantiomeric composition of six new independently prepared samples. As shown in Table 1, the prediction results are in good agreement with the known values for the samples (average of the absolute values of the relative errors is 3% over the mol fraction range of 0.5−0.9).

To date, in addition to ϕ-Gly, we have successfully used this strategy for the determination of the enantiomeric composition of the following compounds, all of which gave similar results: tartaric acid, glycidyl butyrate, aspartic acid, phenylalanine, and arabinose. Naturally, a specific regression model must be developed for each guest molecule using the procedure described in this communication.

The diversity of compounds that have been successfully used suggests that the method is quite general. Depending on the guest molecule, different hosts (α-, β-, or γ-CD) may give somewhat better results in terms of correlation coefficients and prediction ability with future samples. Because the method depends solely on the changes produced in the spectrum of the complexed guest molecule as a result of differences in the binding conditions for the different enantiomeric forms of the guest molecule with the chiral host, it does not assume or depend on any particular stoichiometry of the guest/host complex. Whatever guest/host complexes may be present in the solution, they are not expected to vary because the concentrations of the guest and the host are fixed for the experiment.

We have found the regression models developed as described in this communication to be extremely robust. In the case of tartaric acid, the regression model has continued to correctly predict the enantiomeric composition of unknown samples for up to 6 months without need for recalibration. Moreover, because the chiral analysis method described in this communication does not depend on the specific rotation of the target molecule, it should be especially valuable for compounds where polarimetric determinations are problematic due to small specific rotations.

Acknowledgment. Support from the Robert A. Welch Foundation of Houston, TX, is gratefully acknowledged (grant no. AA-1446).

Supporting Information Available: Additional experimental information on ϕ-Gly and aspartic acid (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Maier, N. M.; Franco, P.; Lindner, W. *J. Chromatogr., A* **2001**, *906*, 3. (b) Reetz, M. T. *Angew. Chem.* **2001**, *40*, 284.
(2) Finn, M. G. *Chirality* **2002**, *14*, 534.
(3) (a) van Delden, R. A.; Feringa, B. L. *Chem. Commun.* **2002**, 174. (b) Kubo, Y.; Maeda, S.; Tokita, S.; Kubo, M. *Nature* **1996**, *382*, 522. (c) Kaneda, T.; Hirose, K.; Misumi, S. *J. Am. Chem. Soc.* **1989**, *111*, 742. (d) Kubo, Y. *Synlett* **1999**, *2*, 161.
(4) Eliel, E. L.; Wilen, S. H. *Stereochemistry of Organic Compounds*; John Wiley: New York, 1994.
(5) Martens, H.; Næs, T. *Multivariate Calibration*; John Wiley: New York, 1989.

JA025947A