Journal of Chromatography, 466 (1989) 364–370
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROM. 21 171

Note

Chemometric approach to explain the liquid chromatographic retention of some chiral indoles on swollen microcrystalline triacetylcellulose

PER ERLANDSSON
Technical Analytical Chemistry, Chemical Centre, University of Lund, P.O. Box 124, S-221 00 Lund (Sweden)
ROLAND ISAKSSON*
Division of Organic Chemistry 3, Chemical Centre, University of Lund, P.O. Box 124, S-221 00 Lund (Sweden)
INGEMAR NILSSON
Hässle Research Laboratories, Department of Organic Chemistry, S-431 83 Mölndal (Sweden)
and
SVANTE WOLD
Research Group for Chemometrics, Institute of Chemistry, University of Umeå, S-901 87 Umeå (Sweden)
(First received September 7th, 1988; revised manuscript received December 8th, 1988)

During the last decade triacetylcellulose[1] (TAC) has been used successfully as a chiral stationary phase for liquid chromatography to resolve a great number of chiral compounds of different classes. However, little is known about the mechanisms that govern the interaction between the stationary phase and the optically active solute. It is not possible to predict either resolution or retention orders with the present knowledge of these mechanisms.

It has been suggested by Staudinger and Döhle[2] and Hesse and Hagel[1] that the solutes enter cavities between the laminae in the crystalline cellulose to form inclusion complexes. Several experimental findings support this idea, *e.g.*, the difference in retention between benzene and 1,3,5-tri-*tert.*-butylbenzene[3]. The unsubstituted benzene is supposed to fit between the laminae and is retained whereas 1,3,5-tri-*tert.*-butylbenzene is too bulky to enter between these laminae and is therefore not retained. Also, Blaschke *et al.*[4] reported a dependence of the resolution on the size of the solute. Drugs which are not retained are not resolved. Blaschke *et al.* concluded that the size of these solutes is too large to allow their entry into the cavities of the cellulose.

Hesse and Hagel[1] and Francotte *et al.*[5] have found that the tertiary structure of the cellulose is important for selectivity. It was found that, compared with the corresponding crystalline TAC, amorphous TAC more or less completely loses its resolving power, presumably because of a loss of the chiral cavities in the amorphous material. The idea of chiral cavities is not conclusive, however. Okamoto *et al.*[6] found that the retention order of the enantiomers of Troegers base on microcrystalline TAC were reversed when columns prepared by coating dissolved TAC on silica were used.

0021-9673/89/$03.50

The native structure of cellulose, cellulose I, is lost in the preparation of these columns and is presumably replaced with another kind of cellulose, the morphology of which is not yet known. We have also shown[7] that it is possible to separate the enantiomers of even larger molecules on TAC than those studied by Blaschke *et al.* A possible explanation of this could be that partial inclusion in such cavities is sufficient to afford stereoselectivity. Another possible explanation is that the entire surface of the cellulose is utilized in the chiral recognition process.

Recently, Schulze and König[8] reported enantiomer separations on silica gel to which monosaccharides were covalently bonded. Glucose, a monosaccharide, is the structural unit of cellulose. These findings, together with the great variety of structures of compounds resolved into enantiomers on TAC, indicate that several different chiral recognition mechanisms have to be considered.

Multivariate statistical (chemometric) methods for the evaluation of data in chemistry[9] have been extensively used in studies of the quantitative structure–activity relations (QSAR) of drugs and other biologically active compounds. The same approach has also been used to study the relationship between molecular structure and chromatographic behaviour and has been termed quantitative structure–retention relationships (QSRR)[10]. Among the multivariate methods, multiple linear regression (MLR), an extension of ordinary least squares, has been most commonly used. Partial least squares in latent variables (PLS), developed by Wold and co-workers[11,12], is based on the projection of the original multivariate data matrices down on smaller matrices (T) with orthogonal columns.

$$X = TP' + E$$

$$Y = TQ' + F$$

Here the $n \times p$ matrix X is projected down on the $n \times A$ matrix T (score matrix) by the $p \times A$ projection matrix P leaving the residuals E. Analogously, Y ($n \times q$) is projected on T by the $q \times A$ projection matrix Q leaving the residuals F. P and Q are often called loading matrices. The calculations also gives an auxiliary matrix W (PLS weights), which is similar to P. For a new case (here a compound), the values of t_a ($a = 1,2, \ldots, A$) —one additional row in the matrix T — are computed from its x-vector and the matrices P and W. This t-vector gives predicted values of y for this case (compound) as tQ'. The projections are chosen to give a maximum correlation between the descriptor (X) and the response matrices (Y), under the condition that X and Y are well approximated by TP' and TQ', respectively. Determination of the significant number of model dimensions (A) is made by cross-validation[11,12]

PLS has several advantages over MLR. MLR cannot handle co-linearities in X whereas PLS concentrates the systematic variation in X into fewer factors (columns in T) than variables in the original matrix X. The PLS projection can be calculated regardless of the number of variables in X, *i.e.*, the number of descriptors may be large than the number of objects (compounds). Projections involving many variables are stable provided that the number of model dimensions (latent variables) extracted is less than about one third of the number of compounds[13]. PLS has been extensively described in the literature and today has a firm statistical basis[14].

To be able to carry out a PLS study of this kind, it is necessary to have a

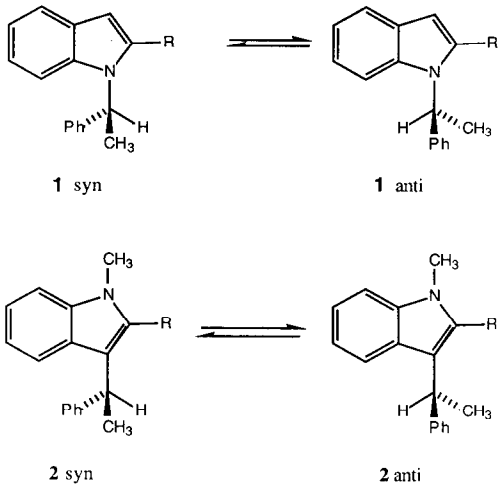

Fig. 1. Chiral indole derivatives **1** and **2**. For explanation of R, see Table I. Ph = Phenyl.

material that shows a systematic variation in both structure and chromatographic behaviour. The chiral indole derivatives[15] (Fig. 1) used in this study are good candidates, as their absolute configurations and conformations are known[16,17,18]. Further, the retention, separation and conformations are sensitive to the substitution pattern on the indole framework. In this paper, only the results for the phenethylamino derivatives are reported.

The purpose of this investigation is (a) to use PLS as a tool to analyse retention data for chiral resolutions so as to be able to predict retention and retention orders between enantiomers and (b) to find relationships between the structures of the solutes and their retention that can be used to formulate a model for chiral recognition on TAC.

EXPERIMENTAL

The syntheses of indoles **1a–e** and **2a–e** have been reported previously[15].

The separations were performed by liquid chromatography on swollen microcrystalline TAC as described[15]. The capacity factors (k') for **1b**, **1d** and **1e** were calculated by fitting the experimental chromatograms to skewed overlapping Gaussian curves.

CALCULATIONS

The three-dimensional representations of the indoles were obtained from molecular mechanics calculations (MMP2 force field)[19]. Coordinate transformations and calculations of the non-tabulated molecular descriptors were carried out with the use of the MIMIC (methods for interactive modelling in chemistry) system[20] available at the Chemical Centre in Lund. The absolute configurations were obtained by comparing CD data with those from compounds of known absolute configuration prepared from chiral precursors[16,17].

Molecular descriptors (matrix X)

Two kinds of molecular descriptors were used: (i) tabulated values[21] describing the length, breadth, Hammett σ_m and σ_p, lipophilicity and molar refractivity of R, and also the number of hydrogen bonding sites, and (ii) calculated values from the modelled three-dimensional coordinates such as coordinates for the six-membered ring in the indole part, coordinates of the indole nitrogen, coordinates for the phenyl, methyl and hydrogen connected to the asymmetric carbon, Van der Waals area and volume of the whole molecules and of substituents R, total dipole moment and its three-dimensional direction. These variables comprise the *X*-block in the PLS model, each row corresponding to one compound and each column corresponding to one descriptor.

The chromatographic behaviour of the solutes was described by the logarithm of the capacity factor, log k', and was used as response data in the *Y*-block of the PLS model.

Calculations of the PLS model were carried out with a Victor V-286 microcom-

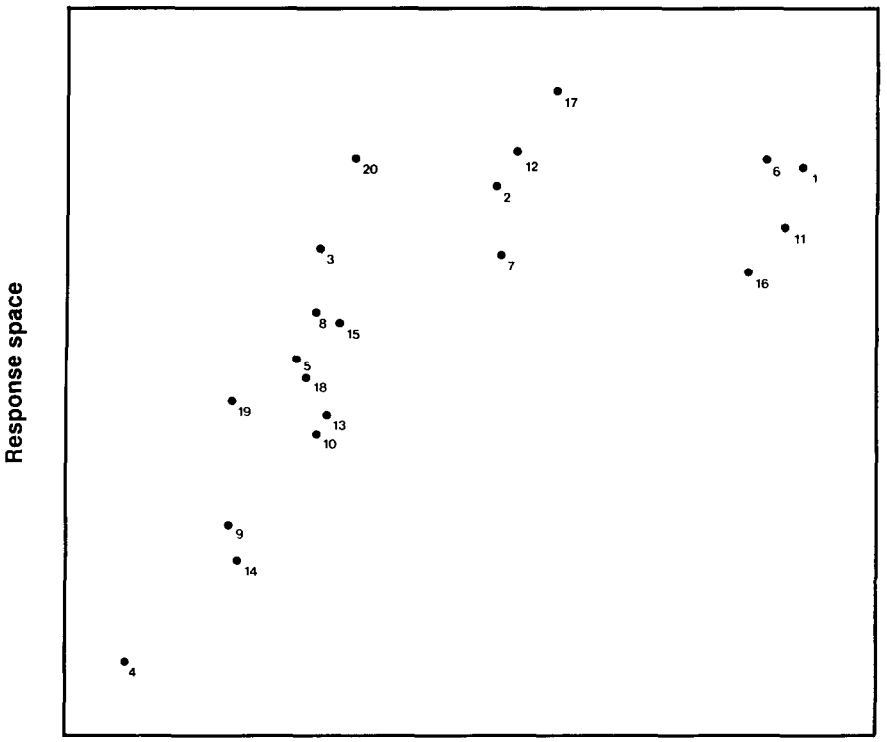

Fig. 2. Score plot of the first dimension of the PLS model. The *x*-axis is the most significant direction in descriptor space and the *y*-axis the most significant direction in response space (chromatographic retention data). 1 = **1a**-*S*; 2 = **1b**-*S*; 3 = **1c**-*S*; 4 = **1d**-*S*; 5 = **1e**-*S*; 6 = **2a**-*S*; 7 = **2b**-*S*; 8 = **2c**-*S*; 9 = **2d**-*S*; 10 = **2e**-*S*; 11 = **1a**-*R*; 12 = **1b**-*R*; 13 = **1c**-*R*; 14 = **1d**-*R*; 15 = **1e**-*R*; 16 = **2a**-*R*; 17 = **2b**-*R*; 18 = **2c**-*R*; 19 = **2d**-*R*; 20 = **2e**-*R*. Compounds 1, 6, 11 and 15 have an *anti* conformation whereas the others have a *syn* conformation.

puter and the SIMCA program package, which is available from Sepanova (Enskede, Sweden) and also from Principal Data Components (Columbia, MO, U.S.A). Details of the computational procedures and molecular descriptors will be reported in a subsequent paper.

RESULTS AND DISCUSSION

Two significant dimensions were obtained from the PLS calculations describing 80% of the variance in the retention data of the indoles. The first PLS component, illustrated by its score plot in Fig. 2, used 23.4% of the variance in descriptor data to explain 50.7% of the variance in retention data. The second PLS component used 4.9% of the variance in descriptor data to explain another 29.4% of the variance in retention data. The score plot (Fig. 2) shows that the indoles are grouped according to their conformations, *anti* and *syn*, thus implying different retention mechanisms.

The retentions of the indoles calculated from the model are listed in Table I and illustrated in Fig. 3. The retention orders of the enantiomers were correctly calculated

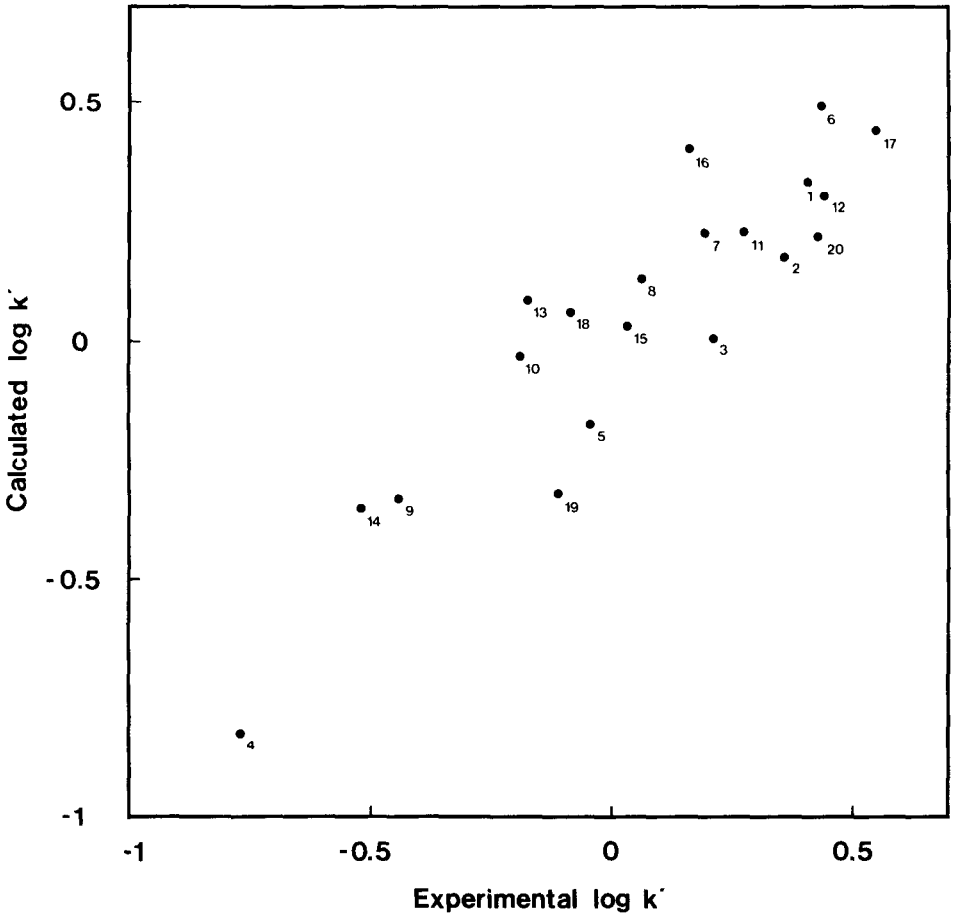

Fig. 3. Plot of predicted log k' against experimental log k' according to Table I. $R = 0.91$. Numbering as in Fig. 2.

TABLE I

EXPERIMENTAL AND PREDICTED RETENTION DATA FOR INDOLES 1 AND 2

Compound			*Experimental values*			*Predicted values*		
No.	*R*	*Conformation*	*Log k'$_S$*	*Log k'$_R$*	*1st eluted isomer*	*Log k'$_S$*	*Log k'$_R$*	*1st eluted isomer*
1a	H	*anti*	0.4048	0.2685	*R*	0.3324	0.2251	*R*
1b	CH_3	*syn*	0.3541	0.4393	*S*	0.1758	0.2968	*S*
1c	CO_2CH_3	*syn*	0.2122	−0.1805	*R*	0.0620	0.0847	*R*
1d	CH_3COOH	*syn*	−0.7696	−0.5229	*S*	−0.8230	−0.3495	*S*
1e	$CH_3C=CH_2$	*syn*	−0.0506	0.0334	*S*	−0.1756	0.0337	*S*
2a	H	*anti*	0.4281	0.1553	*R*	0.4958	0.4034	*R*
2b	CH_3	*syn*	0.1931	0.5465	*S*	0.2221	0.4409	*S*
2c	CO_2CH_3	*syn*	0.0607	−0.0915	*R*	0.1315	0.0599	*R*
2d	CH_3COOH	*syn*	−0.4437	−0.1135	*S*	−0.3288	−0.3148	*S*
2e	$CH_3C=CH_2$	*syn*	−0.1938	0.4249	*S*	−0.0305	0.2176	*S*

in all instances, which indicates the possibility of predicting absolute configurations from retention data on TAC columns. Preliminary analysis of the PLS model shows that the most important descriptors are the coordinates of the substituents attached to the asymmetric carbon, dipole moment, volume and area of the solutes (see Table II). As the PLS model differs in its assumptions from regression, these coefficients (loadings) should not be interpreted as measures of the independent contributions of the variables of retention behaviour, but rather how much (relative) information the variables contain about the retention behaviour.

It seems as if the PLS approach described here may afford an insight into the mechanisms of retention and chiral recognition on TAC columns.

TABLE II

LOADINGS (P_{ka}) OF THE FIVE MOST IMPORTANT VARIABLES IN EACH COMPONENT OF THE PLS MODEL.

A high absolute value of a loading indicates a high contribution to the model. The maximum loading is 1 or −1.

PLS component	*Parameter*	P_{ka}
First (a = 1)	Area of R	−0.2814
	Volume of R	−0.2812
	Molar refractivity of R	−0.2806
	Breadth of R	−0.2762
	Length of R	−0.2685
Second (a = 2)	x Coordinate of the phenyl group attached to chiral carbon	−0.3832
	Total dipole moment	−0.3257
	y Component of dipole moment	0.2522
	x Coordinate of R	0.2429
	y Coordinate of R	0.2237

ACKNOWLEDGEMENTS

Robert E. Carter and Tommy Liljefors are gratefully thanked for maintaining the MIMIC facility at the Chemical Centre, University of Lund.

This work was supported by grants from the Swedish Natural Science Council and the Swedish Board for Technical Development.

REFERENCES

1 G. Hesse and R. Hagel, *Justus Liebigs Ann. Chem.*, (1976) 996.
2 H. Staudinger and W. J. Döhle, *J. Prakt. Chem.*, 161 (1942) 219.
3 H. Koller, K. H. Rimböck and A. Mannschreck, *J. Chromatogr.*, 282 (1983) 89.
4 G. Blaschke, H. P. Kraft and H. Markgraf, *Chem. Ber.*, 116 (1983) 3611.
5 E. Francotte, R. M. Wolf, D. Lohmann and R. Mueller, *J. Chromatogr.*, 347 (1985) 25.
6 Y. Okamoto, M. Kawashima, K. Yamamoto and K. Hatada, *Chem. Lett.*, (1984) 734.
7 G. Lindsten, O. Wennerström and R. Isaksson, *J. Org. Chem.*, 52 (1987) 547.
8 J. Schulze and W. A. König, *J. Chromatogr.*, 355 (1986) 165.
9 R. G. Brereton, *Analyst (London)*, 112 (1987) 1635.
10 B.-K. Chen and Cs. Horváth, *J. Chromatogr.*, 171 (1979) 15–28.
11 S. Wold, C. Albano, W. J. Dunn, III, U. Edlund, K. Esbensen, P. Geladi, S. Hellberg, E. Johansson, W. Lindberg and M. Sjöström, in B. R. Kowalski (Editor), *Proceedings of NATO Advanced Study in Chemometrics, Cosenza, Italy, September 1983*, Reidel, Dordrecht, 1984, pp. 17–96.
12 S. Wold, C. Albano, W. J. Dunn, III, K. Esbensen, P. Geladi, S. Hellberg, W. Lindberg, M. Sjöström, B. Skagerberg, C. Wikström and J. Öhman, paper presented at the *7th International Conference on Computers in Chemical Research and Education, Garmisch-Partenkirchen, June 10–14, 1985*, in press.
13 S. Hellberg, *A Multivariate Approach to QSAR*, University of Umeå, Umeå, 1986.
14 A. Höskuldsson, *J. Chemometr.*, 2 (1988) 211; and references cited therein.
15 I. Nilsson and R. Isaksson, *Acta Chem. Scand., Ser. B*, 39 (1985) 531.
16 I. Nilsson, U. Berg and J. Sandström, *Acta Chem. Scand., Ser. B*, 40 (1986) 625.
17 I. Nilsson, U. Berg and J. Sandström, *Acta Chem. Scand., Ser. B*, 41 (1987) 261.
18 I. Nilsson, U. Berg and J. Sandström, *Acta Chem. Scand., Ser. B*, 38 (1984) 491.
19 N. L. Allinger and J. T. Sprague, *J. Am. Chem. Soc.*, 94 (1972) 5734.
20 T. Liljefors, *J. Mol. Graphics*, 1 (1983) 111.
21 J. K. Seydel and K.-J. Schaper, *Chemische Struktur und Biologische Aktivität von Wirkstoffen*, Verlag Chemie, Weinheim, 1979.