CHROM. 25 434

Chromatographic separation of amide diastereomers: correlation with molecular descriptors

Göran Hansson

Analytical and Marine Chemistry, University of Göteborg, S-412 96 Göteborg (Sweden)

Martin Ahnoff*

Bioanalytical Chemistry, Astra Hässle AB, S-431 83 Mölndal (Sweden)

ABSTRACT

The retention and separation of diastereomeric amides were studied on a reversed-phase chromatographic system. Amides were prepared by coupling chiral carboxylic acids to chiral amines with ethyl chloroformate as activating agent. Both increasing and decreasing α-values (range 0.87–1.39) with increasing eluent strength were observed, and in a few instances the elution order of diastereomers was reversed. Chromatographic data were correlated by partial least-squares (PLS) analysis with different sets of theoretical molecular descriptors, obtained from molecular mechanic calculations. A three-factor PLS model for retention explained 99.2% of the variance of the 80 amides. A six-factor model for separation factors explained 97.3% of the variance of the 40 pairs of amides. Electric potential distribution spectra, and descriptors derived from such spectra, were useful for the prediction of both absolute and relative retention.

INTRODUCTION

Techniques for the separation of enantiomers as diastereomeric derivatives play an important role in both organic synthesis and analysis. The large number of papers published in this area throughout the years clearly indicate that the selection of a proper chiral derivatizing reagent is not a trivial task [1–5], one reason being the difficulty of predicting the separation of the diastereomers. The work reported here and in related papers [6,7] is an attempt to investigate the potential of combining molecular modelling and multivariate analysis for correlating chemical structure with chromatographic retention of diastereomers.

In order to obtain a set of chromatographic data suitable for systematic studies, we chose to prepare diastereomeric amides, which represent all possible combinations of a number of chiral carboxylic acids and chiral amines. In this way, each chiral moiety of any of the molecules will appear in several other combinations. To be able to prepare a relatively large number of different amides, we utilized an analytical scale derivatization technique for carboxylic acids, described by Björkman [8] and applied by others for the separation of non-steroid anti-inflammatory drugs [9,10]. The preparation was automated, using a liquid chromatographic autosampler with premixing facilities, and was carried out directly before injection on to the chromatographic system. All separations were carried out on the same column, which was characterized by test mixtures [11,12] at the beginning and end of the study.

* Corresponding author.

EXPERIMENTAL

Preparation of amides

The following solutions were used for the preparation of diastereomeric amides: *R*- and *S*-enantiomers of chiral carboxylic acids (Table I), 10–35 mmol/l, 50 m*M* triethylamine and 60 m*M* ethyl chloroformate (Sigma, St. Louis, MO, USA), all in dry acetonitrile; and enantiomers of chiral amines (Table I), 0.8–1.0 mmol/l in methanol. Solutions of hydrochloride salts of amines were neutralized with equimolar amounts of triethylamine.

Amides were prepared prior to chromatographic separation using a Model 9090 autosampler (Varian, Sunnyvale, CA, USA) with premixing facilities. The derivatization rection is outlined in Fig. 1. Reagent solutions were kept in 1.2-ml screw-capped vials and derivatization was carried out in vials with 50-μl inserts. Aliquots of 4 μl of carboxylic acid, triethyl amine and ethyl chloroformate solutions were
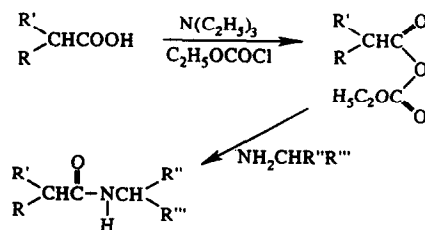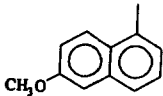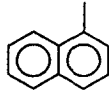

Fig. 1. Reaction scheme for the formation of diastereomeric amides.

mixed and allowed to react before 4 μl of amine solution were added. Finally, 16 μl of 0.25 *M* hydrochloric acid was added to neutralize excess amine reagent and to make the sample more compatible with the chromatographic system. The total time for the derivatization procedure was 16 min and included reaction times, which were not shorter than those reported by Björkman [8] for this type of reaction (0.5 min for formation of the mixed carbonic–carboxylic anhydride, 2 min for the rection with the amine).

TABLE I

CHIRAL CARBOXYLIC ACIDS AND AMIDES USED TO PREPARE DIASTEREOMERIC AMIDES

R, R', R″ and R‴ refer to the structural formula in Fig. 1 for the corresponding amides

	R-	R'-	R″-	R‴-
Carboxylic acids				
2-Phenylbutanoic acid (A)	CH_3CH_2-	C_6H_5-		
O-Acetylmandelic acid (C)	CH_3COO-	C_6H_5-		
α-Methoxyphenylacetic acid (D)	CH_3O-	C_6H_5-		
Naproxen (G)	CH_3	CH₃O⟨naphthyl⟩		
2-Phenylpropionic acid (H)	CH_3	C_6H_5		
Amines				
α-(1-Naphthyl)ethylamine (N)			CH_3	⟨naphthyl⟩
1-Phenylethylamine (O)			CH_3-	C_6H_5-
2-Amino-1-butanol (P)			CH_3CH_2-	$HOCH_2$-
Methioninamide (Q)			NH_2CO-	$CH_3SCH_2CH_2$-
Alaninamide (R)			NH_2CO-	CH_3-
Serinamide (S)			NH_2CO-	$HOCH_2$-
Phenylalaninamide (T)			NH_2CO-	$C_6H_5CH_2$-
Leucinamide (U)			NH_2CO-	$(CH_3)_2CHCH_2$-

Before injection on to the chromatographic column, 4 μl of a solution were added, containing a suitable retention reference substance (a substituted phenone or 2-phenylethanol).

Measurement of chromatographic retention and separation

Samples (25 μl) were separated on a 300 × 4.6 mm I.D. column packed with 5-μm octylsilica particles (Kromasil C_8, 100 Å; Eka Nobel, Bohus, Sweden). The analytical column was thermostated at 30°C and protected by a 36 × 4.6 mm I.D. scavenger column (same packing) and a 15 × 3.2 mm I.D. guard column (RP-8, 7 μm, 300 Å; Applied Biosystems, Santa Clara, CA, USA) positioned before and after the injector, respectively. The mobile phase, delivered by a Varian Model 5500 pump, was a mixture of acetonitrile and phosphate buffer (0.10 M, pH 6.0). UV absorbance was measured at 254 nm. Chromatograms were recorded and processed with a Varian DS-651 data system. Isocratic separations were carried out with three or four different acetonitrile–buffer mixtures for each amide so that the retention [capacity factor $k' = (t_R - t_0)/t_0$] of each amide varied over the range 2–16. Single diastereomers were injected in separate runs to establish the relative elution order, while chromatograms (duplicates) from mixtures of both diastereomers were used for evaluation of separation factors, defined as $\alpha = k'_{SS}/k'_{RS}$ (or $\alpha = k'_{RR}/k'_{SR}$). For those cases where diastereomers were incompletely resolved, i.e., when the valley between peaks was higher than 25% of the mean peak height [corresponding to resolution $R_s = \Delta t_R/(2\sigma_1 + 2\sigma_2) \approx 1.0$], the difference in retention times, Δt_R, was calculated from the height h (μV) at the centre of the unresolved peaks, their total area A (μV s) and the mean band dispersion σ (s) of the single diastereomers, according to the equation

$$\Delta t_R = \sqrt{8}\sigma\sqrt{-\ln(\sqrt{2\pi}\sigma h/A)} \qquad (1)$$

A derivation of eqn. 1 is given in Appendix 1. Band dispersion was calculated as $\sigma_i = A_i/(\sqrt{2\pi}h_i)$. For each amide, the percentage of organic modifier (%CH_3CN) was plotted against log k'. For each pair of diastereomers the separation factor α was plotted against log k' for the

SS (or RR) enantiomer. Slopes and values at $k' = 6.0$ for %CH_3CN and α were calculated by linear regression.

In order to characterize the chromatographic system and to check it for long-term drift in retention, two sets of test substances were used at the start and at the end of the measurement period. One set consisted of 2-phenylethanol, p-cresol, N-methylaniline methyl benzoate, nitrobenzene, toluene and uracil (Fluka, Buchs, Switzerland) and the other acetophenone and homologues up to hexanophenone (Aldrich, Steinheim, Germany).

Calculation of molecular descriptors. Multivariate analysis

Molecular conformations for the amides under investigation were calculated using the SYBYL software (Tripos, St. Louis, MO, USA) as described in a separate paper [7]. A variety of molecular descriptors were derived from atomic coordinates, atomic electric charges and the electric potentials at the accessible surface. A list of the molecular descriptors is given in Table II. Apart from the descriptors derived from molecular conformation, a number of descriptors were used which were not sensitive to changes in configuration and consequently identical for diastereomers. One set (F) of Free–Wilson type descriptors defined the carboxyl and amine residues of the amides. Fourteen descriptors (a subset of M) were derived from calculated atom charges. One set of descriptors (D) made up the electric potential distribution spectrum, described in more detail in ref. 7. The electric potential was calculated (in SYBYL) at 1.4 Å outside the Van der Waal's surface, with a resolution of 30 points, or "dots", per Å^2. The electric potential range was divided into 77 intervals, and each of 77 descriptors indicated the partial surface area with electric potentials within one of the 77 intervals. A set of 14 descriptors (a subset of M) was from statistical derivations of the spectrum. One set of descriptors were related to the size and shape of molecules, while another consisted of the partial energy terms used in energy minimization in SYBYL. Further, six descriptors were derived from atom charges and coordinates; among them

TABLE II

MOLECULAR DESCRIPTORS

Fragment type descriptors (**F**)

Free–Wilson type of descriptors. Five descriptors indicating acid fragment, and eight descriptors indicating amine fragment (13 descriptors). See Table I.

A; C; D; G; H; N; O; P; Q; R; S; T; U.

Electric potential distribution spectra (**D**)

Points are distributed, 30 per Å^2, around the molecule at 1.4 Å from the Van der Waals surface. Electric potentials are calculated at these points and are collected in a histogram (77 descriptors) [7].

General descriptors (**M**)

Energy terms from molecular mechanic calculations (9 descriptors).

$E_{\text{bond stretching}}$; $E_{\text{angle bending}}$; E_{torsion}; $E_{\text{out-of-plane bending}}$; $E_{\text{1–4 Van der Waals}}$; $E_{\text{Van der Waals}}$; $E_{\text{1–4 electrostatic}}$; $E_{\text{electrostatic}}$; E_{total}.

Descriptors derived from a box circumscribing the molecule (11 descriptors) [7].

length (L); breadth (B); depth (D); L/B; B/D; box area (A); box volume (V); V/A; $LV^{-1/3}$; $BV^{-1/3}$; $DV^{-1/3}$.

Other size related descriptors.

Total number of atoms; relative molecular mass (M_r); Van der Waals volume (VdW); VdW/M_r; polarizability [13]; $\Sigma |r_i - r_j|^{-2}$; $\Sigma |r_i - r_j|^2$.

Descriptors derived from the electric potential distribution spectrum of the molecule (14 descriptors) [7].

First moment; second moment; third moment; fourth moment; skewness; kurtosis; the lowest potential dot extracted (Min); the highest potential dot extracted (Max); Max − Min; total number of dots; number of dots at the positive extreme (Σ^+); number of dots close to zero potential (Σ^0); number of points at the negative extreme (Σ^-); $\Sigma^0 - \Sigma^+ - \Sigma^-$.

Descriptors derived from atom point charges, q_i (14 descriptors).

First moment; second moment; third moment; fourth moment; skewness; kurtosis; lowest atom charge (q_{min}); highest atom charge (q_{max}); $q_{\text{max}} - q_{\text{min}}$; $\Sigma |q_i|$; $\Sigma |q_i - q_j|$; $\Sigma (n_i q_i)$, where n_i is the number of lone pairs; $\Sigma (h_i q_i)$, where h_i is the number of hydrogen atoms at oxygen or nitrogen atoms; number of π-electrons.

Charge-related descriptors derived from point charges, q, and coordinates, r, of atoms. The coordinate axes x, y and z are oriented in the molecule's length, breadth and depth direction (10 descriptors).

Dipole $= \Sigma (q_i r_i)$; |**Dipole**|; $H = \Sigma (|q_i| r_i)$; |**H**|; $T^{\text{E}} = \Sigma \Sigma (|q_i - q_j| \cdot |r_i - r_j|^{-2})$ [14]; $\Sigma \Sigma (|q_i - q_j| \cdot |r_i - r_j|^2)$.

Descriptors specially designed for describing the conformation at the amide bond (**O**)

Vectors centred at the middle of the amide bond (13 descriptors).

Lipo $= \Sigma (1 - |q_i|/Q_{\text{max}} r_i)$; **Hydro** $= \Sigma (|q_i|/Q_{\text{max}} r_i)$; |**Lipo**|; |**Hydro**|; **Lipo · Hydro**; **Lipo × Hydro**; $\cos \phi$; $\sin \phi$; ϕ (where Q_{max} are the highest $|q_i|$ of all the molecules in this study and ϕ is the angle between **Lipo** and **Hydro**).

Internal coordinates of the hydrogen at the chiral carbons. One set of the carboxylic acid fragment and one for the amine fragment (6 descriptors).

Bond length HC_{acid} and HC_{amine}; angle $HC_{\text{acid}}C_{\text{amide}}$ and $HC_{\text{amine}}N_{\text{amide}}$; torsion $HC_{\text{acid}}C_{\text{amide}}$; and $HC_{\text{amine}}N_{\text{amide}}C_{\text{amide}}$.

Free–Wilson type of descriptors describing torsion ($HC_{\text{acid}}C_{\text{amide}}N_{\text{amide}}$ and $HC_{\text{amine}}N_{\text{amide}}C_{\text{amide}}$) of the hydrogen at the chiral carbons. One set for the carboxylic fragment, one for the amine fragment and one set for the combination of them (15 descriptors).

Three descriptors for the carboxylic residue: g_C^+ (*gauche*, $+60°$); g_C^- (*gauche*, $-60°$); a_C (*anti*, $+180°$). Three descriptors for the amine residue: g_A^+; g_A^-; a_A. Nine descriptors for the combinations: $g_C^+ g_A^+$; $g_C^+ g_A^-$; $g_C^+ a_A$; $g_C^- g_A^+$; $g_C^- g_A^-$; $g_C^- a_A$; $a_C g_A^+$; $a_C g_A^-$; $a_C a_A$.

---

were the dipole moment and the topological electronic index [14]. A set of descriptors (**O**) described, in different terms, the conformation around the amide bond with the two chiral centers, and was therefore specific for the type of substances studied, in contrast to the other descriptors derived from the calculated molecular conformation.

For all amides, six molecular conformations were calculated in SYBYL, corresponding to the six local energy minima with lowest estimated conformational energy, electrostatic interactions not included. Boltzmann-weighted molecular descriptor values were obtained by calculating the Boltzmann distribution between the six conformations and weighting the corresponding descriptor values with the Boltzmann probability [7] factors to obtain a weighted average. In these calculations, electrostatic interactions were included in the total energies, and values of the relative dielectric constant ε of 2 and 60 were chosen to represent a non-aqueous medium (the stationary phase) and an aqueous medium (the mobile phase), respectively.

Among the descriptors derived from molecular conformation, a few were of Free–Wilson type, i.e., they assume only the values zero or one. Boltzmann weighting turned these descriptors into continuous variables with range 0–1.

Unscrambler, Version 4.00EX (CAMO, Trondheim, Norway) was used for principal component analysis and partial least-squares analysis of descriptor data and chromatographic data. For the correlation of descriptors with chromatographic separation factors α, descriptor values for the SS isomers $(\mathbf{d}_{SS})$ were used together with the difference between SS and RS descriptor values $(\Delta \mathbf{d} = \mathbf{d}_{SS} - \mathbf{d}_{RS})$.

RESULTS AND DISCUSSION

Precision of chromatographic measurements

All chromatographic measurements, 930 injections in total, were carried out with the same column over a period of 97 days. An example of a chromatogram is shown in Fig. 2. The number of theoretical plates, measured for the phenones with 50% acetonitrile, was on average 17 600 at the beginning and 16 200 at the end of measurements, corresponding to an increase in peak width of 4%. The average difference between retention times measured in consecutive runs was 0.2%. After 97 days and 930 injections, the retention k' of the test compounds and phenones had decreased by 5–10%, the larger drift being seen for more retained compounds. For the
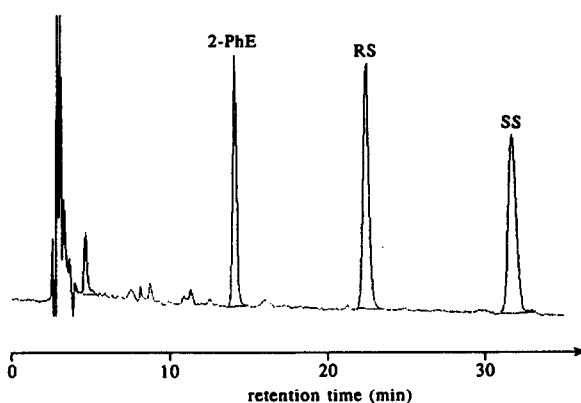

Fig. 2. Chromatographic separation of the diastereomeric amides DU formed by coupling $(R + S)$-α-methoxyphenylacetic acid (D) to L-leucinamide (U) with ethyl chloroformate (see Experimental for details). Column, 300×4.6 mm I.D. Kromasil C$_8$, 5 μm, 100 Å; mobile phase, phosphate buffer (0.1 M, pH 6.0)–acetonitrile (72.6:27.4, v/v); UV detection at 254 nm.

phenones, the corresponding change in acetonitrile concentration giving $k' = 6.00$ was on average -1.26%. Retention indices which were calculated for the test compounds, using the homoguous series of phenones as reference compounds, decreased by only 4–7 units over the period, indicating that the change in selectivity was small. Retention data for the test compounds are listed in Appendix 2.

When retention data (%CH$_3$CN at $k' = 6.00$) were calculated for the amides by linear regression (see Experimental and Fig. 3), the average root mean square (RMS) error was 0.36. The imprecision (standard deviation) of separation factors α for diastereomers, measured in consecutive runs, was on average $2 \cdot 10^{-4}$ for partially or fully resolved peaks. For unresolved peaks with a single apex, the imprecision was $3 \cdot 10^{-4}$. The lowest measured value of $|\alpha - 1|$ was $16 \cdot 10^{-4}$ ($\alpha = 1.0016$). The relative elution order of the diastereomers was determined from separate runs with each diastereomer, where the retention relative to a common phenone reference compound was calculated when the difference in retention times was small. When α-values were fitted to log k', the average RMS error was $2 \cdot 10^{-3}$ (Fig. 3).
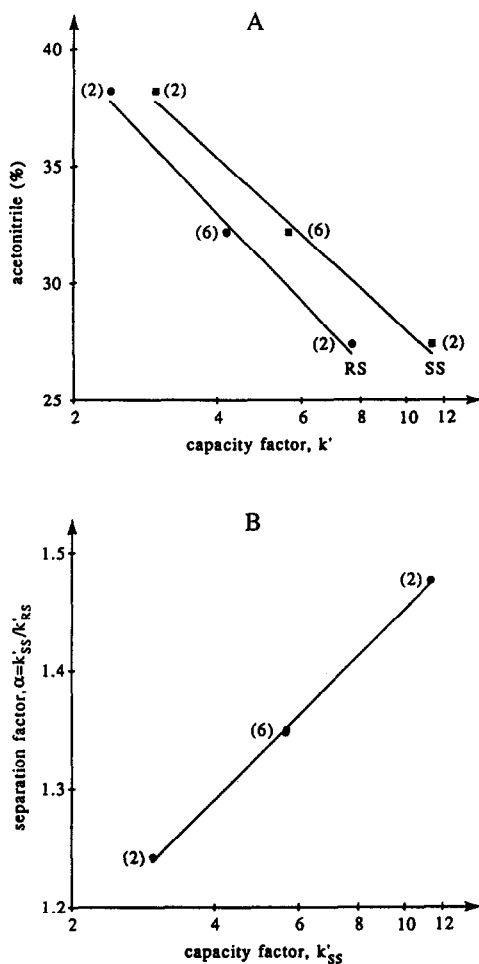
Fig. 3. Measured retention and separation of the diastereomers DU (see Fig. 2). Figures in parentheses denote the number of measurements. (A) Plot of %CH$_3$CN against measured retention (logarithmic scale). Regression lines were used to calculate %CH$_3$CN at $k' = 6.00$. (B) Plot of measured separation against retention of the SS isomer. The regression line was used to calculate separation at $k'_{SS} = 6.00$.

Measured retention and separation

The 80 amides had widely different retentions in the reversed-phase system used, and it was not meaningful to use a single isocratic system for all of them. Although an alternative would be to use gradient elution, we chose a series of isocratic systems, which required a large number of runs, but allowed a study of the absolute and relative retention as a function of the level of organic modifier in the mobile phase. Between

13 and 60% CH$_3$CN was needed to obtain $k' = 6.00$. The separation factor α at $k'_{SS} = 6.00$ ranged between 0.87 and 1.39. Plots of %CH$_3$CN and α against retention are shown in Fig. 3 for the same diastereomers as in Fig. 2. In Fig. 4, plots of α against retention are shown for all 80 amides, grouped according to the identity of the carboxylic acid residue. Both positive and negative changes of α with increasing retention can be seen. A complete listing of chromatographic retention data is given in Appendix 3.

Reversal of elution order of diastereomers

In four cases, the elution order of diastereomers could be reversed by changing the concentration of acetonitrile in the mobile phase. In the plots of α against k'_{SS} in Fig. 4, the regression lines for these amides cross $\alpha = 1$ between $k'_{SS} = 2.0$ and $k'_{SS} = 6.0$. Chromatograms from one pair of amides are shown in Fig. 5. At low retention (higher concentration of acetonitrile) the four SS diastereomers were more retained than their RS isomers, while the opposite was true at high retention (lower concentration of acetonitrile). This clearly shows that the selectivity of a reversed-phase chromatographic system can be significantly changed, also for closely related solutes, by changing the concentration of organic modifier in the mobile phase.

Correlation of retention with sets of molecular descriptors

PLS was used to correlate chromatographic retention (%CH$_3$CN at $k' = 6.00$) with molecular descriptors. In contrast to multiple linear regression, PLS allows the use of a large number of variables, and covariation between variables is not a drawback. This allows the use of similar or closely related descriptors. An optimum number of PLS factors is determined by cross-validation [6].

Table III shows results from calibration experiments with five sets of descriptors and some combinations of these. The optimum number of factors, determined by cross-validation, was in most instances three. A successful calibration should not only have a high explained variance (R^2) and a low root mean square error (RMSE) of calibration, but also a low error of prediction
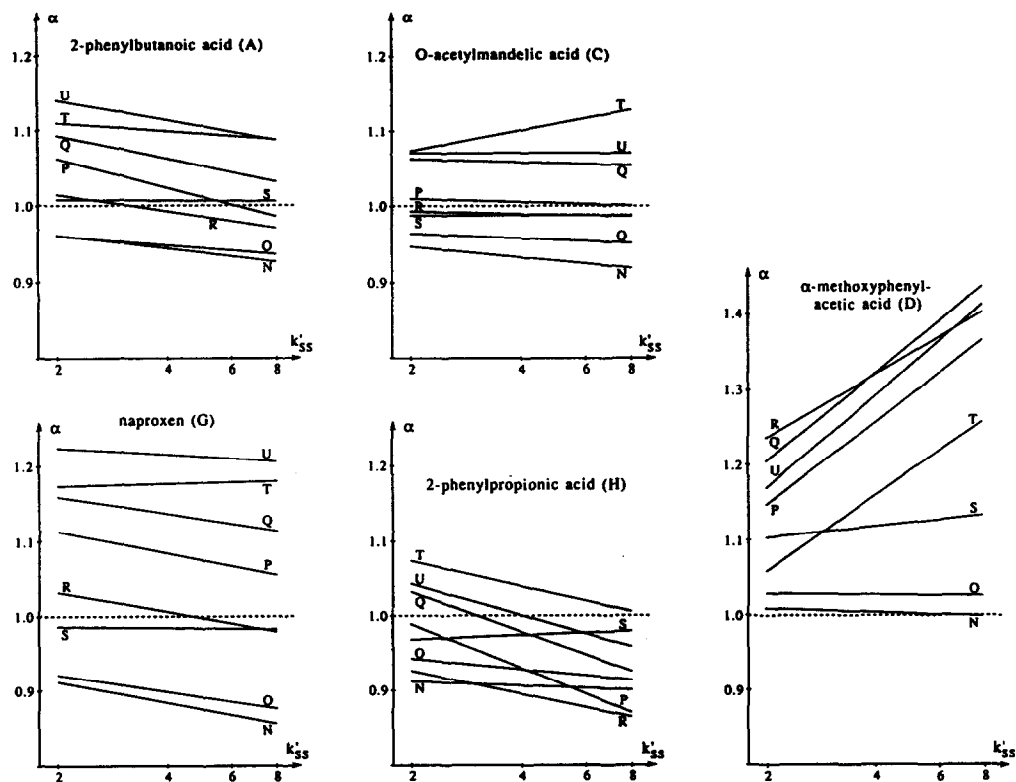
Fig. 4. Separation factors $\alpha = k'_{SS}/k'_{RS}$ plotted against retention k'_{SS} for 40 pair of amides grouped according to the identity of their carboxylic acid residue (see Table I for explanation of symbols).

of objects excluded from calibration. As can be seen from Table III, the thirteen descriptors (F) identifying the carboxyl and amine residues described 99% of the total variance and gave an RMSE of prediction of 2.6 (%CH$_3$CN) of pairwise excluded diastereomeric amides. Descrip-

tors derived from molecular conformations and/ or atomic charges (M) gave an improved prediction. Minor differences were seen between the two different modes of Boltzmann weighting. Potential distribution spectra (D) contain useful information but give higher errors of prediction

Fig. 5. Reversal of elution order by changing eluent composition. The diastereomeric amides from $(R + S)$-2-phenylpropionic acid and L-leucinamide were separated: (a) at 44.2%, (b) at 32.2% and (c) at 27.4% CH$_3$CN. The calculated separation factors $\alpha = k'_{SS}/k'_{RS}$ were 1.078, 0.993 and 0.931, respectively. 2-PhE = 2-phenylethanol.

512

G. Hansson and M. Ahnoff / J. Chromatogr. A 666 (1994) 505–517

TABLE III

CORRELATION BY PLS OF CHROMATOGRAPHIC RETENTION WITH DIFFERENT SETS OF DESCRIPTORS (SEE TABLE II)

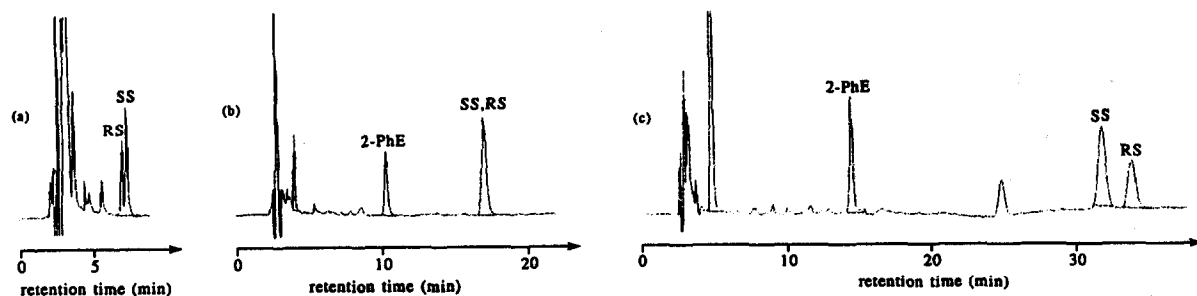Listed are the number of descriptors (n_d), the number of PLS factors (n_t), the explained variance (R^2), adjusted for degrees of freedom. The root mean square error of estimation (RMSEE) of calibration objects is given for calibration with all objects present. The root mean square error of prediction (RMSEP) is given for cross-validation with a single amide or a pair of diastereomers excluded at a time, respectively. Descriptor subscripts 2 and 60, respectively, indicate the relative dielectric constant used for Boltzmann weighting of descriptor values. Further correlation data for descriptors FM_2 are presented in Fig. 6 and Table IV.

X matrix	n_d	n_t	R^2 (%)	RMSEE	RMSEP	
					Single	Pairs
F	13	1	99.0	1.23	1.71	2.58
M_{60}	65	3	98.4	1.58	1.74	1.90
M_2	65	3	98.7	1.42	1.71	1.91
D_{60}	77	5	95.5	2.60	3.55	3.94
D_2	77	5	94.7	2.83	3.80	4.32
FM_{60}	78	3	99.0	1.26	1.45	1.58
FM_2	**78**	**3**	**99.2**	**1.14**	**1.42**	**1.61**
$M_{60}M_2$	130	3	98.5	1.53	1.75	1.89
$FM_{60}M_2$	143	3	98.8	1.34	1.54	1.68
$FM_{60}D_{60}$	155	3	98.3	1.64	2.02	2.15
FM_2D_2	155	3	97.8	1.86	2.22	2.40
$FM_{60}D_{60}M_3D_2$	297	3	97.7	1.89	2.29	2.47

when used alone. The combined descriptors F + M gave better results than each set alone, but further expansion of the set did not result in better prediction. In fact, inclusion of the descriptor set **D** resulted in increased prediction errors for the expanded model, evidently owing to the addition of more noise than information. A plot of calculated *versus* measured retention, using a model with the descriptors $F + M_2$, is shown in Fig. 6.

The model was further tested by cross-validation where all amides, containing a certain carboxylic acid or amine residue, were removed simultaneously from the calibration set and were predicted from the remaining amides (Table IV). When the RMS error of prediction is split into bias and standard deviation, it can be seen that the standard deviation for the predicted objects is of the same order as the calibration error. Some groups, especially amides of naproxen (G) or 2-aminobutanol (P), have significant bias. Such chemical structures evidently must be represented in the calibration set.
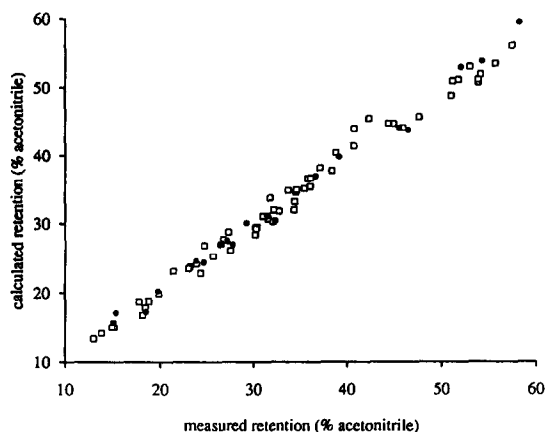
Fig. 6. Correlation between chromatographic retention (%CH_3CN at $k' = 6.0$) and molecular descriptors (sets F and M_2, see Table III). A PLS model with three factors was used. ● = Calibration objects $(n = 20)$; □ = test objects $(n = 60)$. The adjusted explained variance of calibration objects was 99.2%. The RMS error of estimation of calibration objects was 1.15 (%CH_3CN). The RMS error of prediction of test objects was 1.51.

TABLE IV

CORRELATION BY PLS OF CHROMATOGRAPHIC RETENTION WITH MOLECULAR DESCRIPTORS (FROM THE SET F, M_2, SEE TABLES II AND III)

Cross-validation with simultaneous exclusion of all amides containing a certain carboxyl or amine residue. The root mean square error (RMSE, %CH_3CN) of prediction for the excluded amides (a test set) is given together with bias and standard deviation (S.D.). Explained variance, adjusted for degrees of freedom (R^2) and RMSE of the calibration sets, is also listed.

	Calibration objects		Test set objects		
	R^2 (%)	RMSE	RMSE	Bias	S.D.
A	98.8	1.28	1.67	1.42	0.91
C	98.6	1.35	1.04	0.05	1.07
D	99.3	0.95	1.96	0.23	2.00
G	98.8	1.27	5.72	−5.32	2.16
H	98.9	1.18	1.55	0.54	1.50
N	99.0	0.93	3.28	2.51	2.23
O	98.9	1.15	3.43	3.26	1.10
P	98.9	1.24	6.38	−6.25	1.33
Q	99.2	1.06	1.74	−1.21	1.32
R	98.9	1.18	2.45	−1.80	1.75
S	98.5	1.30	1.54	−1.45	0.54
T	98.9	1.28	2.25	1.65	1.61
U	99.1	1.12	2.58	2.40	0.99

Principal component analysis of the 78 descriptors in the set $F + M_2$ showed that 95.5% of the total variance of the descriptor space was covered by sixteen factors, the number of factors suggested by cross-validation with an F-test. The correlation coefficients of descriptors *versus* the three PLS factors were examined for the calibration model with descriptors $F + M_2$. The highest positive correlation (0.96) to the first PLS factor was found for the descriptor Σ^0 derived from the electric potential distribution spectrum. It is the partial surface area near zero potential [between −4.00 and +3.25 kcal/mol (1 kcal = 4.184 kJ)] and can tentatively be described as lipophilic surface area. The highest negative correlation (−0.95) to the first PLS factor was for the descriptors $q2$ and $q4$, which are the second moment (variance) and fourth moment of the atomic charges, respectively. They are different measures of the spreading of atomic charges away from the mean and might be regarded as indicators of "polarity". The dipole moment was among the three descriptors with highest correlation to the second factor

($r = 0.71$). However, when the dipole moment is correlated directly with retention, a (weak) negative correlation is seen, which is what would be expected. An explanation to the positive correlation with the second PLS factor is that the dipole moment is also represented in the first PLS factor. This illustrates that correlations with higher PLS factors may be difficult to interpret.

Correlation of diastereomer separation with molecular descriptors

Correlation of molecular descriptors with chromatographic separation (α) of diastereomers was carried out with a separate PLS model. While the retention model treated the 40 pairs of diastereomers as 80 separate objects, the separation model contained 40 objects. Each object was represented by descriptors for the *SS* isomer (d_{SS}) and descriptors defined as the difference between *SS* and *RS* descriptor values ($\Delta d = d_{SS} - d_{RS}$) [6]. The descriptor matrix with, *e.g.*, descriptors $F + M$ in this way contained 143 variables compared with 78 for the retention model (some descriptors were by definition

identical for *SS* and *RS* diastereomers). Results of correlation experiments with different sets of descriptors are shown in Table V. The figures for explained variance are lower than for the retention models. This is to be expected, as the variance of relative retention of diastereomers is very much smaller than the variance of retention of all 80 amides. The electric potential distribution spectra (descriptor set **D**) turned out to be more valuable for modelling diastereomer separation than for modelling retention. When spectra (Boltzmann weighted with $\varepsilon = 60$) were correlated with α, a six-factor PLS model explained 94% of the total variance, and cross-validation

(with one pair of diastereomers excluded from calibration at a time) gave an RMS error of prediction of α of 0.135. By combining the descriptor sets $\mathbf{F} + \mathbf{O} + \mathbf{M} + \mathbf{D}$ ($\varepsilon = 60$), the explained variance increased to 97%. A plot of calculated *versus* measured separation is shown in Fig. 7. The RMS error of prediction (cross-validation) was 0.076. This descriptor set contained 365 variables. Principal component analysis indicated that the descriptor space could be described by seventeen factors. This is about the same as found for the 78 descriptors $\mathbf{F} + \mathbf{M}$ used for modelling retention. An examination of the six-factor PLS model showed that the de-

TABLE V

CORRELATION BY PLS OF CHROMATOGRAPHIC SEPARATION OF DIASTEREOMERS (α AT $k' = 6.00$) WITH DIFFERENT SETS OF MOLECULAR DESCRIPTORS (SEE TABLE II AND REF. 6)

Listed are root mean square error of estimation (RMSEE) obtained in calibration with all 40 amides, and root mean and median square errors of prediction obtained in cross-validation where one diastereomer pair was excluded from calibration at a time (see also explanation of Table IV). A correlation plot for descriptors $FO_{60}M_{60}D_{60}$ is shown in Fig. 7.

| X matrix	n_d	n_f	R^2 (%)	RMSEE	Root mean square error of prediction	Root median square error of prediction
F	13	1	81.0	0.058	0.083	0.053
$\mathbf{O}_{60}$	68	6	80.5	0.055	0.115	0.079
$\mathbf{O}_2$	68	6	86.7	0.045	0.100	0.066
$\mathbf{M}_{60}$	130	6	87.2	0.045	0.092	0.072
$\mathbf{M}_2$	130	6	90.8	0.038	0.084	0.062
$\mathbf{D}_{60}$	154	6	94.0	0.030	0.135	0.070
$\mathbf{D}_2$	154	6	92.8	0.033	0.112	0.070
$\mathbf{FO}_{60}$	81	6	89.2	0.041	0.096	0.069
$\mathbf{FO}_2$	81	6	91.1	0.037	0.090	0.063
$\mathbf{FM}_{60}$	143	6	92.4	0.034	0.075	0.053
$\mathbf{FM}_2$	143	6	93.0	0.033	0.081	0.053
$\mathbf{FD}_{60}$	167	6	93.1	0.024	0.103	0.048
$\mathbf{FD}_2$	167	6	95.1	0.028	0.095	0.056
$\mathbf{O}_{60}\mathbf{M}_{60}$	198	6	91.0	0.037	0.095	0.060
$\mathbf{O}_2\mathbf{M}_2$	198	6	92.2	0.035	0.090	0.067
$\mathbf{FO}_{60}\mathbf{M}_{60}$	211	6	92.9	0.033	0.086	0.058
$\mathbf{FO}_2\mathbf{M}_2$	211	6	93.6	0.032	0.085	0.062
$\mathbf{FM}_{60}\mathbf{D}_{60}$	297	6	97.3	0.020	0.078	0.043
$\mathbf{FM}_2\mathbf{D}_2$	297	6	96.4	0.024	0.081	0.046
$\mathbf{FO}_{60}\mathbf{M}_{60}\mathbf{D}_{60}$	**365**	**6**	**97.3**	**0.021**	**0.076**	**0.038**
$\mathbf{FO}_2\mathbf{M}_2\mathbf{D}_2$	365	6	96.2	0.024	0.084	0.058
$\mathbf{FO}_{60}\mathbf{O}_2\mathbf{M}_{60}\mathbf{M}_2\mathbf{D}_{60}\mathbf{D}_2$	730	6	96.9	0.022	0.074	0.045
X matrix as in ref. 6:						
Boltzmann weighted ($\varepsilon = \infty$)	194	5	96.5	0.024	0.090	0.056
Global minima	194	5	94.6	0.029	0.127	0.063
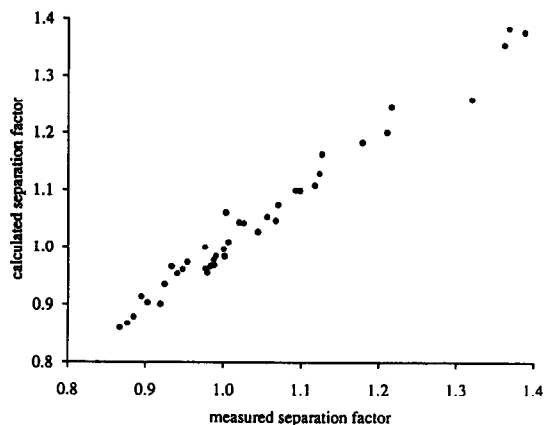
Fig. 7. Correlation between chromatographic separation $\alpha = k'_{SS}/k'_{RS}$ and molecular descriptors $\mathbf{F + O + M + D}$ (Boltzmann weighted, $\varepsilon = 60$) using a six-factor PLS model.

scriptor $\Delta g_C^- g_A^-$ (See Table II) was among those with highest correlation with the first PLS factor ($r = -0.82$) and at the same time was one of the descriptors with highest correlation with α ($r = -0.64$). This descriptor is a measure of the calculated differential probability of finding the SS and RS diastereomers with the hydrogens at both chiral centers in a *gauche* ($-60°$) orientation. An examination of the descriptor $g_C^- g_A^-$ revealed that only the SS isomers assume the *gauche* ($-60°$)–*gauche* ($-60°$) conformation.

CONCLUSIONS

Descriptors derived from calculated molecular conformations can be used for the prediction of the separation of diastereomers. A large number of calibration objects is needed for accurate prediction, which means that a large chromatographic database has to be built. Calibration models require a substantial number of descriptors. However, the inclusion of too many descriptors with large random variation (noise) may be a problem. The electric potential distribution spectrum is a multi-dimensional descriptor which is general, *i.e.*, is not restricted to certain chemical structures. The spectrum, and descriptors derived from it, were found to be useful for predicting absolute and relative retentions of diastereomeric amides.

APPENDIX 1

Calculation of retention time difference $\Delta t_R = |t_{R1} - t_{R2}|$ for two incompletely resolved chromatographic peaks

We assume Gaussian distributions with the same band width σ for both peaks. The total height at time t is

$$h(t) = h_1\, e^{-(t-t_{R1})^2/2\sigma^2} + h_2\, e^{-(t-t_{R2})^2/2\sigma^2} \qquad (1a)$$

The peak height h at time $t_R = (t_{R1} - t_{R2})/2$, *i.e.*, in the middle between peaks 1 and 2, becomes

$$h = (h_1 + h_2)\, e^{-(t_{R1}-t_{R2})^2/8\sigma^2} \qquad (1b)$$

Now, h_1 and h_2 can be replaced by the total peak area A and the mean peak band width σ, using the relationships

$$A_i/h_i = \sqrt{2\pi}\,\sigma_i \qquad (2a)$$

$$A = A_1 + A_2 = \sqrt{2\pi}\,\sigma(h_1 + h_2) \qquad (2b)$$

so that eqn. (1b) becomes

$$h = \frac{A_1 + A_2}{\sqrt{2\pi}\,\sigma} \cdot e^{-(t_{R1}-t_{R2})^2/8\sigma^2} \qquad (3a)$$

Let $\Delta t_R = |t_{R1} - t_{R2}|$ and rearrange so that

$$e^{-\Delta t_R^2/8\sigma^2} = \sqrt{2\pi}\,\sigma h/A \qquad (3b)$$

$$\Delta t_R = \sqrt{8}\,\sigma\sqrt{-\ln(\sqrt{2\pi}\,\sigma h/A)} \qquad (3c)$$

APPENDIX 2

Retention times and retention indices

Retention times (t_R) and retention indices (I) are given for test substances on an octylsilica column (300×4.6 mm I.D. Kromasil C_8, 5 μm, 100 Å), used for collecting data for diastereomeric amides. Mobile phase, 50.2% acetonitrile in phosphate buffer (0.1 M, pH 6.0); flow-rate, 1.00 ml/min; column temperature, 30°C.

Compound	Before study		After study		ΔI
	t_R(min)	I	t_R(min)	I	
Uracil	2.63		2.63		
2-Phenylethanol	5.63	710.6	5.46	706.5	−4.1
p-Cresol	7.10	773.3	6.75	766.7	−6.6
Acetophenone	7.83	800.0	7.59	800.0	
N-Methylaniline	9.51	852.8	8.97	847.2	−5.6
Nitrobenzene	10.43	877.9	9.75	871.1	−6.8
Methyl benzoate	10.87	19889.1	10.18	883.0	−6.1
Propiophenone	11.32	900.0	10.80	900.0	
Butyrophenone	16.15	1000.0	15.21	1000.0	
Toluene	18.21	1031.3	16.65	1024.6	−6.7
Valerophenone	23.68	1100.0	22.00	1100.0	
Hexanophenone	35.63	1200.0	32.69	1200.0	

APPENDIX 3

Chromatographic retention data for 40 pairs of diastereomeric amides

The two-letter code refers to the carboxylic and amine residue of each amide (see Table I); φ is the percentage of acetonitrile giving $k' = 6.00$; slope φ is the linear change of φ with increase in log k'; α is the relative retention, k'_{SS}/k'_{RS}, at $k'_{SS} = 6.00$; slope α is the change in α with increase in log k'_{SS}.

Amide	SS		RS		α	Slope $\alpha \cdot 10^3$
	φ	Slope φ	φ	Slope φ		
AN	57.50	−27.11	58.28	−26.42	0.9339	−56.8
AO	51.07	−28.15	51.79	−27.66	0.9412	−38.3
AP	30.21	−19.66	30.20	−18.62	1.0019	−126.4
AQ	31.81	−16.72	31.50	−16.03	1.0448	−101.6
AR	22.97	−21.16	23.16	−20.50	0.9796	−73.6
AS	19.88	−15.68	19.84	−15.67	1.0066	−1.6
AT	39.11	−20.28	38.34	−19.99	1.0929	−35.6
AU	36.66	−22.71	35.77	−21.95	1.0986	−87.2
CN	51.19	−26.52	52.07	−25.96	0.9248	−46.0
CO	44.38	−25.21	44.89	−24.98	0.9539	−20.4
CP	23.82	−21.69	23.79	−21.56	1.0032	−14.1
CQ	27.21	−18.56	26.77	−18.46	1.0563	−13.0
CR	18.42	−18.16	18.51	−18.07	0.9878	−11.0
CS	15.10	−12.72	15.16	−12.73	0.9882	1.8
CT	34.53	−15.98	33.73	−16.56	1.1173	91.9
CU	31.57	−19.01	31.00	−19.01	1.0709	0.6
DN	53.91	−26.86	53.90	−26.66	1.0005	−17.1
DO	45.88	−32.08	45.52	−32.00	1.0264	−5.3
DP	24.29	−21.24	21.36	−24.28	1.3199	363.5
DQ	27.49	−17.54	24.66	−19.95	1.3883	385.7
DR	18.16	−18.64	15.37	−20.52	1.3678	276.3
DS	13.84	−15.83	13.01	−16.13	1.1265	48.9
DT	34.33	−16.62	32.75	−18.77	1.2164	330.8
DU	32.11	−18.54	29.27	−21.30	1.3619	404.1
GN	58.57	−25.49	60.08	−24.36	0.8675	−92.7

Amide	SS		RS		α	Slope $\alpha \cdot 10^3$
	φ	Slope φ	φ	Slope φ		
GO	53.01	−25.47	54.31	−24.56	0.8852	−74.9
GP	36.05	−22.20	35.44	−21.37	1.0676	−95.3
GQ	37.09	−20.19	36.09	−19.63	1.1236	−74.4
GR	30.31	−18.54	30.38	−17.84	0.9908	−87.5
GS	26.35	−15.07	26.46	−15.04	0.9838	−5.2
GT	42.29	−21.75	40.73	−21.83	1.1788	11.2
GU	40.65	−23.30	38.74	−23.05	1.2107	−28.8
HN	54.15	−34.90	55.70	−34.39	0.9031	−18.8
HO	46.47	−30.86	47.57	−30.18	0.9196	−47.9
HP	24.62	−22.55	25.62	−20.63	0.8952	−197.4
HQ	27.33	−19.70	27.76	−18.22	0.9477	−178.1
HR	17.78	−19.28	18.82	−18.38	0.8773	−102.1
HS	14.93	−13.51	15.06	−13.63	0.9777	19.5
HT	34.59	−22.65	34.41	−21.63	1.0202	−112.8
HU	32.17	−19.03	32.36	−17.87	0.9764	−141.0

REFERENCES

1 W.H. Pirkle and J. Finn, in J.D. Morrison (Editor), *Asymmetric Synthesis, Vol. 1: Analytical Methods*, Academic Press, New York, 1983, Ch. 6.

2 R.W. Souter, *Chromatographic Separations of Stereoisomers*, CRC Press, Boca Raton, FL, 1985.

3 J. Gal, in I.W. Wainer and D.E. Drayer (Editors), *Drug Stereochemistry*, Marcel Dekker, New York, 1988, Ch. 4.

4 M. Ahnoff and S. Einarsson, in W.J. Lough (Editor), *Chiral Liquid Chromatography*, Blackie, London, 1989, Ch. 4.

5 N.R. Srinivas and L.N. Igwemezie, *Biomed. Chromatogr.*, 6 (1992) 1634.

6 G. Hansson and M. Ahnoff, *Chemometr. Intell. Lab. Syst.*, 17 (1992) 223.

7 G. Hansson and M. Ahnoff, in preparation.

8 S. Björkman, *J. Chromatogr.*, 339 (1985) 339.

9 W.T.C. Liang, D.R. Brocks and F. Jamali, *J. Chromatogr.*, 577 (1992) 317.

10 H. Spahn, *J. Chromatogr.*, 423 (1987) 334.

11 R.M. Smith, *Anal. Chem.*, 56 (1984) 256.

12 R.M. Smith, G.A. Murilla and C.M. Burr, *J. Chromatogr.*, 388 (1987) 37.

13 K.J. Miller and J.A. Savchik, *J. Am. Chem. Soc.*, 101 (1979) 7206.

14 K. Osmialowski, J. Halkiewicz and R. Kaliszan, *J. Chromatogr.*, 361 (1986) 63.