Journal of Chromatography, 171 (1979) 15–28

CHROM. 11,536

EVALUATION OF SUBSTITUENT CONTRIBUTIONS TO CHROMATOGRAPHIC RETENTION: QUANTITATIVE STRUCTURE–RETENTION RELATIONSHIPS

BOR-KUAN CHEN and CSABA HORVÁTH

Chemical Engineering Group, Department of Engineering and Applied Science, Yale University, New Haven, Conn. 06520 (U.S.A.)

(Received October 19th, 1978)

SUMMARY

The use of multiple regression analysis with the indicator variables in the statistical formulation of quantitative structure and retention relationships is demonstrated. Retention data of aromatic–aliphatic acids in paper chromatography and those of catecholamine derivatives in reversed-phase chromatography with octadecyl-silica stationary phases and an aqueous eluent were analyzed. Statistical tests showed that the substituent parameters ΔR_M, or the corresponding τ values in column chromatography, can be estimated with high accuracy. Very good agreement was found between the observed and predicted R_M or κ values, the latter expressing the logarithm of retardation (capacity) factor. Data obtained with different octadecyl-silica stationary phases at various temperatures suggest that quantitative structure–retention relationships can be transformed from one reversed phase system to another as long as the eluent composition is the same.

INTRODUCTION

Martin[1] suggested first that a substituent changes the partition coefficient of a substance by a given factor that depends on the nature of the substituent and the two phases employed, but not on the rest of the molecule. Shortly thereafter, Bate-Smith and Westall[2] showed experimentally that a number of compounds follow Martin's extra-thermodynamic rule in paper chromatography. The observed substituent effect on retention is a manifestation of linear free energy relationships (LFER) that are widely used to interpret the effect of structural parameters on chemical rates and equilibria[3]. Since Lederer[4] gave the first account of the significance of LFER in chromatography, the evaluation and prediction of substituent contributions to chromatographic retention have been the subject of extensive study.

On the other hand, a similar approach to correlating molecular structure with biological activity has made substantial advances since Hansch[5] successfully applied LFER to the study of quantitative structure–activity relationships (QSAR). In order to quantify hydrophobic properties in QSAR water–octanol partition coefficients

have most commonly been used, although Leo *et al.*[6] have demonstrated the collinearity of partition coefficients in different water–organic solvent systems in agreement with an earlier suggestion by Collander[7].

In order to exploit the convenience offered by chromatographic techniques, attempts were made to obtain partition coefficients by using liquid–liquid chromatography. Moreover, chromatographic data have directly been used to establish QSAR[8]. High-performance liquid chromatography (HPLC) with bonded hydrocarbonaceous stationary phases has also been suggested as a method to evaluate partition coefficients[9–12]. Indeed, with the sophisticated instrumentation presently available, HPLC allows an easy and precise measurement of chromatographic retention under controlled conditions.

Reversed-phase chromatography with octadecyl-silica as the stationary phase is presently the most popular method in HPLC[13]. It has been shown that in this technique the retention behavior is governed by solvent effects which have a rigorous theoretical basis[14–16]. Advances in understanding the physico-chemical phenomena underlying the retention process in reversed-phase chromatography suggest that the technique may become a precise method for the convenient measurement of hydrophobic properties. Concomitantly, it is expected that the relationship between chemical structure and retention can be treated in a much more fundamental way than in other types of chromatography.

Quantitative structure–retention relationships (QSRR), however, can be generally established in most chromatographic systems on the basis of extra-thermodynamic LFER. The main reason for formulating QSRR in practice is that they can greatly facilitate the prediction of chromatographic retention from the molecular structure of the eluite. A large quantity of chromatographic data is readily processed by the computer, and advanced statistical methods, which have found application in QSAR studies, can be used for the establishment of QSRR.

In this paper, we present a general approach to the evaluation of substituent contributions to chromatographic retention by using multiple linear regression analysis and indicator variables. The application of the method is illustrated for reversed-phase chromatography of catecholamine derivatives with retardation (capacity) factors measured earlier[17] and recently in our laboratory. Retention data obtained by Kuchař *et al.*[18] with aromatic–aliphatic acids in paper chromatography are also analyzed to establish QSRR. It is believed that the method can generally be used for formulation of QSRR and the prediction of retention values under given chromatographic conditions. It will be shown that there is a relationship between QSRR in slightly different reversed-phase chromatographic systems, and consequently retention values measured in one system can be used to predict the retention of chemically similar substances in another.

THEORETICAL

Measures of retention and retention increments

Under ideal conditions, retention in liquid–liquid chromatography is directly related to the partition coefficient of the solute in the bulk eluent–liquid stationary phase system. In paper or thin-layer chromatography, the retention of a substance is measured by its R_F value. In order to use a thermodynamically more meaningful

retention parameter, Bate-Smith and Westall[2] introduced the term R_M given by

$$R_M = \log\left(\frac{1}{R_F} - 1\right) \tag{1}$$

For an "ideal" chromatographic system, the R_M value of a substance is linearly related to the logarithm of its partition coefficient, P:

$$R_M = \log P + \log C \tag{2}$$

where C is a constant for a particular system. In accordance with Martin's proposal[1], the contribution of a given substituent to retention in paper and thin-layer chromatography is represented by the corresponding ΔR_M value. With the octanol–water system, the chromatographically evaluated ΔR_M is equivalent to the corresponding π value of the substituent[5].

The retention in gas or liquid column chromatography is usually expressed in terms of the dimensionless retardation factor, k, which is defined by

$$k = (t_R - t_0)/t_0 \tag{3}$$

where t_R and t_0 are the retention time of the substance under investigation and the hold-up time of an unretained tracer, respectively. The logarithm of the retardation factor, $\log k$, has the same physico-chemical meaning as R_M; both are proportional to the free energy change associated with the chromatographic distribution process[14]. For convenience and simplicity, we shall use the symbol κ for $\log k$. The difference between the κ value of a substance i, κ_i, and that of its parent compound, κ_p, is denoted by $\log r_{i,p}$, and is given by

$$\log r_{i,p} = \kappa_i - \kappa_p = \sum_{j=1}^{m} \tau_{ji} \tag{4}$$

where $r_{i,p}$ is the corresponding relative retention and τ_{ji} is a substituent parameter which measures the change in chromatographic retention upon replacing a hydrogen atom by substituent j. The maximum number of substituent parameters is m for the set of congeners.

It has been shown[17] that the τ_{ji} values for a given substituent in a certain position are the same in different compounds i. Consequently, we can write τ_j instead of τ_{ji}. A given substituent in different positions, however, may have different τ_j values, depending on the particular molecular environment which can change the effect of the substituent on retention, and therefore the number of τ_j may be greater than the number of actual substituents.

Evaluation of quantitative structure–retention relationships

In analyzing solute retention in a given chromatographic system, the κ values of n congeners containing m possible substituents can be expressed by a set of linear equations:

$$\begin{aligned}
\kappa_1 &= \kappa_p + I_{11}\tau_1 + I_{12}\tau_2 + \cdots + I_{1m}\tau_m \\
\kappa_2 &= \kappa_p + I_{21}\tau_1 + I_{22}\tau_2 + \cdots + I_{2m}\tau_m \\
&\cdots \\
\kappa_n &= \kappa_p + I_{n1}\tau_1 + I_{n2}\tau_2 + \cdots + I_{nm}\tau_m
\end{aligned} \tag{5}$$

where the coefficient I_{ij} is the indicator variable, which is set to unity when a τ_j value is assigned to a substituent in compound i, and to zero otherwise. Accordingly, for the parent compound, all indicator variables are zero and for the congeners only those which correspond to a position without substituent.

Eqn. 5 can be rewritten in matrix form and simplified as

$$[I_{ij}]_{n\times m}\,[\tau_j]_{m\times 1} + [\kappa_p]_{n\times 1} = [\kappa_i]_{n\times 1} \tag{6}$$

Eqns. 5 and 6 are also applicable to the analysis of retention data from paper or thin-layer chromatography, in which case κ and τ are replaced by R_M and ΔR_M, respectively.

Subtracting κ_p from each array of κ_1 in eqn. 6, we obtain an expression containing the relative retention as

$$[I_{ij}]_{n\times m}\,[\tau_j]_{m\times 1} = [\kappa_i]_{n\times 1} - [\kappa_p]_{n\times 1} = [\log r_{i,p}]_{n\times 1} \tag{7}$$

If the number of substances is equal to the number of retention increments to be determined, the τ_j values can be obtained by simple matrix operation. In practice, however, the number of congeners should exceed the number of τ_j values in order to have a relatively large sample size for statistical evaluation. By using multiple regression analysis, the matrix τ_j is solved so that the sum of squared deviations between the regressed value $\widehat{\log r_{i,p}}$ and the observed $\log r_{i,p}$ is minimized. This is accomplished[19] when the derivative of the sum of squared derivations with respect to m unknown τ_j values is zero, that is:

$$\frac{\partial}{\partial \tau_j} \sum_{i=1}^{n} (\widehat{\log r_{i,p}} - \log r_{i,p})^2 = 0 \qquad j = 1, \ldots, m \tag{8}$$

This approach results in m linear equations which are solved to yield all unknown τ_j values.

The statistical significance of the τ_j values so obtained is tested by computing a measure of goodness-of-fit. The most useful test is the F test[19], which is based on the calculation of the F ratio by

$$F_{m\,n-m} = \frac{n-m}{m} \cdot \frac{R^2}{1-R^2} \tag{9}$$

where R is the multiple correlation coefficient.

Comparison of the F ratio computed from eqn. 9 with values tabulated for the same degrees of freedom at a certain significance level yields the probability of the distribution tail. The lower is this probability, the more reliable are the τ_j values.

EXPERIMENTAL

An Altex (Berkeley, Calif., U.S.A.) Model 100 HPLC solvent metering system with a Rheodyne (Berkeley, Calif., U.S.A.) Model 7010 sample valve and a Schoeffel (Westwood, N.J., U.S.A.) Model 770 variable-wavelength UV detector at 200 nm was used. Chromatograms were obtained with a Honeywell (Ft. Washington, Pa., U.S.A.) Model Electronik 194 recorder.

Both commercial and home-made columns were used. A Partisil-5 ODS column (25 × 0.46 cm) packed with irregularly shaped 5-μm octadecyl-silica was obtained from Whatman (Clifton, N.J., U.S.A.). Bulk 5-μm Spherisorb supplied by Phase-Sep (Hauppage, N.Y., U.S.A.) was treated with trichlorooctadecylsilane. The resulting spherical octadecyl-silica contained 11.8% of carbon and was packed by the balanced-slurry method into No. 316 stainless-steel tubing (15 × 0.46 cm).

The eluent was 0.1 *M* phosphate buffer, pH 2.1. The temperature of the Partisil ODS column was maintained at 333 K by circulating water through appropriate jacketing from a Model K2R-D (Messgeräte-Werk, Lauda, G.F.R.) constant-temperature bath. The flow-rate of eluent was 2 ml/min and the column inlet pressure was 21.4 MPa. Experiments with the Spherisorb ODS column were carried out at room temperature (296 K) at a flow-rate and inlet pressure of 0.5 ml/min and 2.41 MPa, respectively. The sample compounds were purchased from Aldrich (Milwaukee, Wisc., U.S.A.) or Sigma (St. Louis, Mo., U.S.A.). Retardation factors were evaluated from the chromatograms as customary by using sodium nitrate as the non-sorbed tracer.

Computations were carried out with the IBM 370/158 computer at Yale Computer Center.

RESULTS AND DISCUSSION

QSRR of aromatic–aliphatic acids in paper chromatography

Kuchăr *et al.*[18] have used paper chromatography to investigate R_M values of a large number of aromatic–aliphatic acids. In order to illustrate the present approach to QSRR in paper chromatography, we used their data to evaluate substituent contributions. The R_F values of three groups of compounds derived from phenylacetic acid, cinnamic acid and α-methylcinnamic acid, which were measured in two paper chromatographic systems, were chosen because of the relatively large data base. The substances under investigation and their R_M values are listed in Table I.

In the first chromatographic system, denoted by A1, the retention values of two groups of congeners derived from phenylacetic and cinnamic acid were measured. Accordingly, two sets of linear equations such as eqn. 5 were formulated, one for the 16 phenylacetic acid derivatives and another for the 20 cinnamic acid derivatives. These two sets of equations, however, were put together and solved simultaneously as the substituent contributions are expected to be the same for similar compounds in a given chromatographic system. The ΔR_M values thus evaluated for pertinent substituents are shown together with the statistics of the regression analysis in Table II.

The R_M values predicted on the basis of the statistically evaluated ΔR_M values are consistent with the observed data as seen in Table I with the exception of the dialkoxy derivatives, compounds 36, 37, 67 and 68, whose irregular behavior was also observed by Kuchăr *et al.*[18].

TABLE I

R_M VALUES OBSERVED WITH AROMATIC–ALIPHATIC ACIDS BY KUCHAR *et al.*[18] IN PAPER CHROMATOGRAPHY AND THE VALUES PREDICTED FROM THE PRESENT QSRR

In chromatographic system A1 Whatman No. 4 paper was impregnated with 40% formamide in ethanol containing 5% formic acid and benzene–cyclohexane (1:1) was used as the mobile phase. In system B4 the paper was impregnated with 40% formamide in ethanol containing 5% of ammonium formate and the mobile phase was benzene–cyclohexane (7:3). The numbering of compounds is the same as used by Kuchăr *et al.*[18].

System A1				*System B4*			
Compound		R_M		*Compound*		R_M	
No.	*Substituent*	*Observed*	*Predicted*	*No.*	*Substituent*	*Observed*	*Predicted*
Phenylacetic acid derivatives				*Phenylacetic acid derivatives*			
1	H	1.06	1.060	1	H	1.28	1.280
2	3-Cl	0.72	0.848	2	3-Cl	1.12	1.040
3	4-Cl	0.72	0.735	3	4-Cl	1.00	1.000
4	4-*tert.*-C_4H_9	−0.27	−0.140	4	4-*tert.*-C_4H_9	−0.05	−0.045
5	4-*iso*-C_4H_9	−0.37	−0.265	5	4-*iso*-C_4H_9	−0.21	−0.230
6	4-*iso*-C_3H_7	0.05	0.050	6	4-*iso*-C_3H_7	0.14	0.145
7	4-C_2H_5	0.37	0.370	7	4-C_2H_5	0.50	0.500
8	4-CH_3O	1.19	1.258	8	4-CH_3O	1.19	1.173
9	4-*n*-$C_6H_{13}O$	−0.87	−0.670	9	4-*n*-$C_6H_{13}O$	−0.83	−0.856
10	4-*iso*-C_4H_9O	−0.03	−0.007	10	4-*iso*-C_4H_9O	−0.05	−0.070
11	4-*iso*-C_3H_7O	0.37	0.353	11	4-*iso*-C_3H_7O	0.48	0.456
12	3-Cl,4-*n*-$C_6H_{13}O$	−1.12	−0.881	12	3-Cl,4-*n*-$C_6H_{13}O$	−1.12	−1.094
13	3-Cl,4-*iso*-C_4H_9O	−0.31	−0.220	13	3-Cl,4-*iso*-C_4H_9O	−0.33	−0.310
14	3-Cl,4-*iso*-C_3H_7O	0.21	0.142	14	3-Cl,4-*iso*-C_3H_7O	0.21	0.218
15	3-Cl,4-*n*-C_3H_7O	0.08	0.080	15	3-Cl,4-*n*-C_3H_7O	0.12	0.120
16	3-Cl,4-CH_3O	0.95	1.046	16	3-Cl,4-CH_3O	0.91	0.934
Cinnamic acid derivatives				*α-Methylcinnamic acid derivatives*			
17	H	0.91	0.910	55	H	0.41	0.410
18	3-Cl	0.60	0.698	56	3-Cl	0.0*	0.171
19	4-Cl	0.60	0.585	57	3-Br	−0.14	−0.140
20	3-Br	0.35	0.350	58	4-Br	−0.09	−0.090
21	4-*tert.*-C_4H_9	−0.16	−0.290	59	4-NO_2	1.00	1.000
22	4-*iso*-C_4H_9	−0.31	−0.415	60	4-*tert.*-C_4H_9	−0.91	−0.915
23	4-*iso*-C_3H_7	−0.10	−0.100	61	4-*iso*-C_4H_9	−1.12	−1.100
24	4-*n*-$C_6H_{13}O$	−0.66	−0.819	62	4-*iso*-C_3H_7	−0.72	−0.725
25	4-cyclo-$C_6H_{11}O$	−0.50	−0.500	63	4-*iso*-C_3H_7O	−0.43	−0.414
26	4-*iso*-C_4H_9O	−0.18	−0.160	64	4-CH_2=$CHCH_2O$	−0.21	−0.210
27	4-*iso*-C_3H_7O	0.16	0.203	65	3-CH_3O	0.23	0.230
28	4-CH_2=$CHCH_2O$	0.55	0.596	66	4-CH_3O	0.31	0.303
30	3-Cl,4-*n*-$C_6H_{13}O$	−0.75	−1.030	67	3-CH_3O,4-*n*-$C_6H_{13}O$	−1.12*	−1.906
31	3-Cl,4-*iso*-C_4H_9O	−0.23	−0.368	68	3-CH_3O,4-*iso*-C_3H_7O	−0.02*	−0.594
32	3-Cl,4-*iso*-C_3H_7O	−0.05	−0.010				
33	3-Cl,4-CH_2=$CHCH_2O$	0.43	0.384				
34	3-Cl,4-C_2H_5O	0.60	0.600				
35	3-Cl,4-CH_3O	1.06	0.900				
36	3-CH_3O,4-*n*-$C_6H_{13}O$	−0.14*	−0.319				
37	3-CH_3O,4-*iso*-C_3H_7O	0.63*	0.703				

* Not included in regression analysis.

TABLE II

ΔR_M VALUES FOR SUBSTITUENTS IN PAPER CHROMATOGRAPHY

Substituent	ΔR_M	
	System A1	*System B4*
3-Cl	−0.211	−0.239
4-Cl	−0.325	−0.280
3-Br	−0.560	−0.550
4-Br	—	−0.500
4-NO_2	—	0.590
4-*tert.*-C_4H_9	−1.200	−1.325
4-*iso*-C_4H_9	−1.325	−1.510
4-*iso*-C_3H_7	−1.010	−1.135
4-C_2H_5	−0.690	−0.780
4-C_2H_5O	−0.099	—
3-CH_3O	—	−0.180
4-CH_3O	0.198	−0.107
4-*n*-$C_6H_{13}O$	−1.729	−2.136
4-cyclo-$C_5H_{11}O$	−1.410	—
4-*iso*-C_4H_9O	−1.067	−1.351
4-*iso*-C_3H_7O	−0.707	−0.824
4-*n*-C_3H_7O	−0.769	−0.922
4-CH_2=$CHCH_2O$	−0.314	−0.620
Sample size (*n*)	34	27
Multiple correlation (*R*)	0.995	0.999
Standard error (*s*)	0.139	0.032
F test (*F*)	120.1	1938.6
Probability [*P*(tail)]	0.000	0.000

QSRR of catecholamine derivatives in reversed-phase chromatography on octadecyl-silica with an aqueous eluent

The retention behavior of the aromatic amines, acids and amino acids, the general structures of which are shown in Fig. 1, has been investigated. Their substituents and retardation factors are listed in Tables III and IV. The retardation factors were measured in two reversed-phase systems as described under Experimental and compared with those given by Molnár and Horváth[17]. In all of these studies aqueous 0.1 *M* phosphate buffer (pH 2.1) was used as the eluent. The main differences between the chromatographic systems arose from the siliceous supports used for the preparation of the octadecyl-silica column material and the operating temperature.

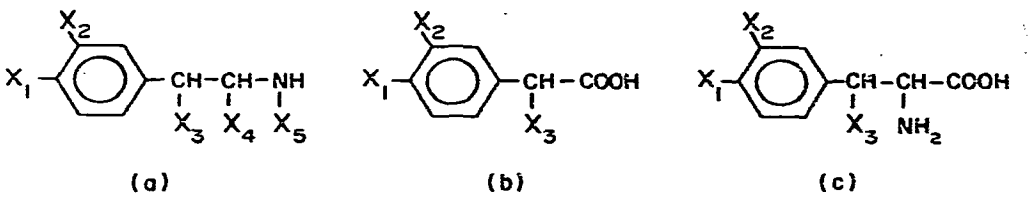

Fig. 1. General structures of the three groups of compounds investigated: (a) amines; (b) acids; (c) amino acids.

TABLE III

RETARDATION FACTORS OF AROMATIC AMINES MEASURED WITH PARTISIL ODS AND SPHERISORB ODS COLUMNS

The general structures and the positions of the substituents are shown in Fig. 1a. The interpretation of the indicator variables is given in the text.

Compound	*Substituent*					*Indicator variable*									*Retardation factor, k*	
	X_1	X_2	X_3	X_4	X_5	I_{11}	I_{12}	I_{13}	I_{14}	I_{15}	I_{16}	I_{17}	I_{18}	I_{19}	*Partisil ODS (333 K)*	*Spherisorb ODS (296 K)*
Norepinephrine	OH	OH	OH	H	H	1	0	0	1	0	1	0	0	0	0.088	0.01
Octopamine	OH	H	OH	H	H	1	0	0	0	0	1	0	0	0	0.22	0.058
Epinephrine	OH	OH	OH	H	CH_3	1	0	0	1	0	1	0	0	1	0.26	0.068
Normetanephrine	OH	OCH_3	OH	H	H	1	0	0	0	1	1	0	0	0	0.50	0.12
Synephrine	OH	H	OH	CH_3	H	1	0	0	0	0	1	0	1	0	0.50	0.12
Dopamine	OH	OH	H	H	H	1	0	0	1	0	0	0	0	0	0.58	0.18
Metanephrine	OH	OCH_3	OH	H	CH_3	1	0	0	0	1	1	0	0	1	1.04	0.30
Tyramine	OH	H	H	H	H	1	0	0	0	0	0	0	0	0	0.99	0.40
Phenylethanolamine	H	H	OH	H	H	0	0	0	0	0	1	0	0	0	1.31	0.42
3-O-Methyldopamine	OH	OCH_3	H	H	H	1	0	0	0	1	0	0	0	0	2.20	0.91
Phenylethylamine*	H	H	H	H	H	0	0	0	0	0	0	0	0	0	3.16	1.10
Ephedrine	H	H	OH	CH_3	CH_3	0	0	0	0	0	1	0	1	1	5.22	2.07

* Parent compound of amines.

TABLE IV

RETARDATION FACTORS OF AROMATIC ACIDS AND AMINO ACIDS MEASURED WITH PARTISIL ODS AND SPHERISORB ODS COLUMNS

The general structures of these compounds and the positions of the substituents are shown in Fig. 1b and c. The interpretation of the indicator variables is given in the text.

Type	*Compound*	*Substituent*			*Indicator variable*							*Retardation factor, k*	
		X_1	X_2	X_3	I_{11}	I_{12}	I_{13}	I_{14}	I_{15}	I_{16}	I_{17}	*Partisil ODS*	*Spherisorb ODS*
Acids	Dihydroxymandelic acid	OH	OH	OH	0	1	0	1	0	0	1	0.47	0.10
	p-Hydroxymandelic acid	OH	H	OH	0	1	0	0	0	0	1	0.90	0.21
	m-Hydroxymandelic acid	H	OH	OH	0	0	1	0	0	0	1	1.56	0.50
	Vanilmandelic acid	OH	OCH_3	OH	0	1	0	0	1	0	1	1.77	0.48
	Dihydroxyphenylacetic acid	OH	OH	H	0	1	0	1	0	0	0	3.96	1.85
	Mandelic acid	H	H	OH	0	0	0	0	0	0	1	4.26	1.67
	p-Hydroxyphenylacetic acid	OH	H	H	0	1	0	0	0	0	0	7.32	4.04
	Homovanillic acid	OH	OCH_3	H	0	1	0	0	1	0	0	15.50	10.62
	Phenylacetic acid*	H	H	H	0	0	0	0	0	0	0	23.96	19.87
Amino acids	Dihydroxyphenylserine	OH	OH	OH	1	0	0	1	0	1	0	0.08	0.013
	Dihydroxyphenylalanine	OH	OH	H	1	0	0	1	0	0	0	0.63	0.23
	Phenylserine	H	H	OH	0	0	0	0	0	1	0	0.86	0.25
	Tyrosine	OH	H	H	1	0	0	0	0	0	0	1.04	0.46
	Phenylalanine**	H	H	H	0	0	0	0	0	0	0	2.59	0.97

* Parent compound of acids.
** Parent compound of amino acids.

Unlike before, in this instance the number of retention increments, τ_j, is greater than that of the substituents. With the compounds under study, a minimum of three different substituents in five positions has to be considered. Because of the different molecular environments in different positions as well as the effect of other groups, *e.g.*, an ionized amino group or phenolic hydroxyl group in the molecule, the retention behavior will be affected[17]. The number of retention increments can be conveniently determined by using multivariate statistical methods of data analysis, *e.g.*, cluster analysis[20]. In a previous study on QSAR of quinazolines[21] the use of such a pre-processing analytical method has been found to simplify greatly the attainment of an adequate correlation equation.

TABLE V

RETENTION INCREMENTS, τ_j, OF SUBSTITUENTS IN COMPOUNDS SHOWN IN FIG. 1

j	*Substituent*	τ_j		
		Partisil ODS	*Spherisorb ODS*	*LiChrosorb ODS*
1	$X_1 = OH^*$	−0.512	−0.489	−0.635
2	$X_1 = OH^{**}$	−0.537	−0.676	−0.803
3	$X_2 = OH$, $X_1 = H$	−0.331	−0.390	−0.522
4	$X_2 = OH$, $X_1 = OH$	−0.292	−0.424	−0.224
5	$X_2 = OCH_3$	0.311	0.324	0.284
6	$X_3 = OH^*$	−0.606	−0.800	−0.586
7	$X_3 = OH^{**}$	−0.856	−1.209	−0.936
8	$X_4 = CH_3$	0.390	0.445	0.399
9	$X_5 = CH_3$	0.361	0.512	0.330

* Ionized amino group in the molecule.
** No ionized amino group in the molecule.

Considering the molecular structures and on the basis of earlier observations[17], we find that nine retention increments, τ_j (j=1, 2, ..., 9), which are shown in Table V, are required for an accurate correlation of the retention with chemical structure. The appropriate τ_j are selected by setting the indicator variables to unity or zero as shown in Tables III and IV. I_{i1} is unity for compound i having OH as substituent X_1 and an ionized amino group in the molecule; if there is no ionized amino group in the molecule, then I_{i2} is unity. With an OH group as substituent X_2 and H or OH as substituent X_1, the indicator variables I_{i3} or I_{i4} are unity, respectively. I_{i5} is unity when a methoxy group is attached to the aromatic ring as X_2. For compounds with an ionized amino group and X_3 = OH, I_{i6} is unity, whereas for those with OH as substituent X_3 but no amino group I_{i7} is unity. I_{i8} indicates the replacement of a hydrogen atom by a CH_3 group in substituent X_4. When substituent X_5 is CH_3, I_{i9} is used. For the acids and amino acids I_{i8} and I_{i9} are zero, and, therefore, they are not shown in Table IV.

In the present analysis, phenylethylamine, phenylacetic acid and phenylalanine served as parent compounds for the amines, acids and amino acids, respectively. The τ_j values are calculated by solving 26 non-homogeneous linear equations (*cf.*, eqn. 5), each representing a substance. The results listed in Table V are statistically reliable, as shown by the data in Table VI.

The τ_j values in Table V establish QSRR for the chromatographic systems. The

TABLE VI

STATISTICS OF THE COMPUTATION OF τ_j VALUES FOR COMPOUNDS LISTED IN TABLES III AND IV

Statistical parameter	*Column used to obtain retardation factors*		
	Partisil ODS	*Spherisorb ODS*	*LiChrosorb ODS*
Sample size (n)	26	26	26
Multiple correlation (R)	0.995	0.992	0.999
Standard error of estimation (s)	0.105	0.178	0.035
$F_{9,17}$ ratio	196	110	2323
Probability [P(tail)]	0.000	0.000	0.000

retardation factor of any congener can be predicted according to eqn. 4 as the sum of the pertinent κ_p and the τ_j values representing the appropriate substituents.

Correspondence of QSRR obtained with slightly different chromatographic systems

According to Collander[7], the logarithms of partition coefficients in systems containing water and different organic solvents are collinear. In reversed-phase chromatography the relationship between κ values obtained with different octadecylsilica phases has not yet been investigated, although the observed collinearity between κ and log P values[9–12] would suggest a proportionality between corresponding κ values obtained with two columns by using the same type of hydrocarbonaceous bonded stationary phase.

In order to compare QSRR obtained with slightly different reversed-phase systems, in this study three sets of κ values were measured with different octadecylsilica stationary phases and the same aqueous eluent at various column temperatures. The data are plotted in Fig. 2, which shows that the κ_i values are collinear. The correlation equations are as follows:

$$\kappa_i^{(2)} = -0.512 + 1.268\,\kappa_i^{(1)} \tag{10}$$

$$\kappa_i^{(3)} = 0.047 + 0.996\,\kappa_i^{(1)} \tag{11}$$

where the superscripts (1), (2) and (3) denote data obtained with the Partisil ODS, Spherisorb ODS and LiChrosorb ODS columns, respectively.

The difference between the numerical coefficients in eqns. 10 and 11 can be explained by the temperature differences between the chromatographic systems investigated. From a study of enthalpy–entropy compensation observed in reversed-phase chromatography under similar conditions[22], we can derive a relationship for the corresponding κ_i values measured with the same chromatographic system at two different temperatures, $T^{(a)}$ and $T^{(b)}$, as follows:

$$\kappa_i^{(a)} = \left[\left(\frac{\beta - T^{(a)}}{\beta - T^{(b)}}\right) \cdot \frac{T^{(b)}}{T^{(a)}}\right] \kappa_i^{(b)} + f(\Delta S) \tag{12}$$

where $f(\Delta S)$ is a function of entropy and β is the compensation temperature, which can be taken as 600 K, the mean of the experimental values. The slopes for $\kappa_i^{(2)} =$

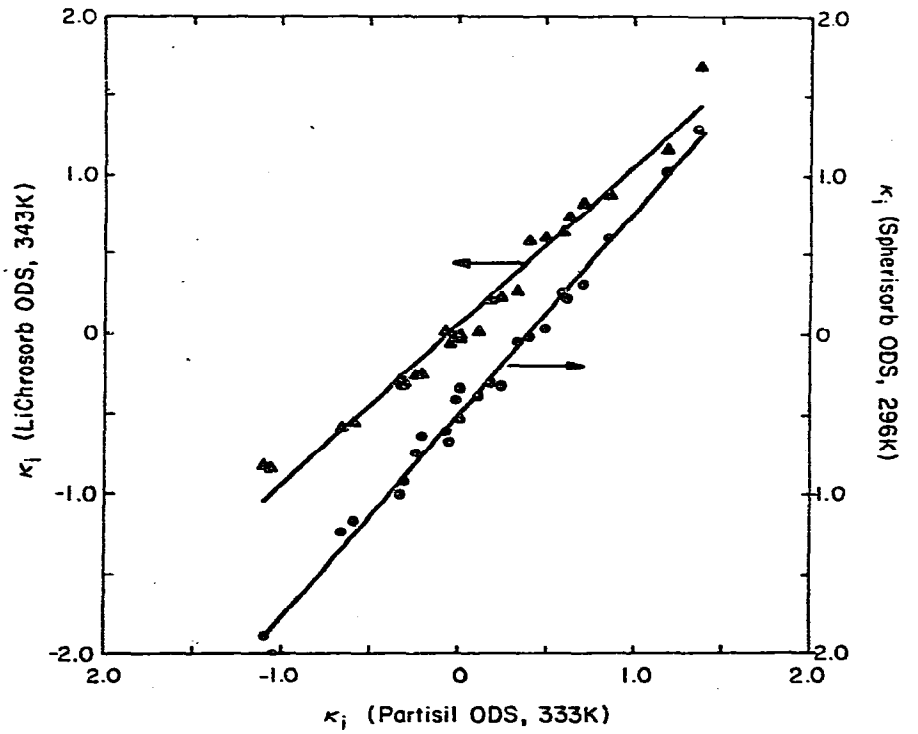

Fig. 2. Correlations between the logarithm of the retardation factors, κ_i, measured with the catecholamine derivatives with different octadecyl-silica columns at different temperatures. In each instance the eluent was 0.1 *M* phosphate buffer, pH 2.1.

$f[\kappa_i^{(1)}]$ and $\kappa_i^{(3)} = f[\kappa_i^{(1)}]$ according to eqn. 12 are 1.28 and 0.94, respectively. These values are commensurable with 1.268 and 0.996, which are the respective slopes in the correlation of κ_i data shown in eqns. 10 and 11.

The results strongly suggest that the variation between the three sets of κ values is essentially due to differences in column temperature and not to dissimilarities between the octadecyl-stationary phases. This finding is in agreement with the predictions of the solvophobic theory[14] for the case of invariant eluent composition. In contradistinction, the theory would not predict such a simple relationship between QSRR evaluated at different mobile phase compositions even by using the same column.

The τ_j values calculated from the corresponding κ_i data obtained with the three different chromatographic systems have also shown satisfactory correlation according to the following equations:

$$\tau_j^{(2)} = -0.008 + 1.272\,\tau_j^{(1)} \quad (13)$$

$$\tau_j^{(3)} = -0.050 + 1.084\,\tau_j^{(1)} \quad (14)$$

In eqns. 13 and 14 the respective slopes are very similar to those in eqns. 10 and 11 and this finding, which is also evident from Fig. 3, suggests that the above equations allow

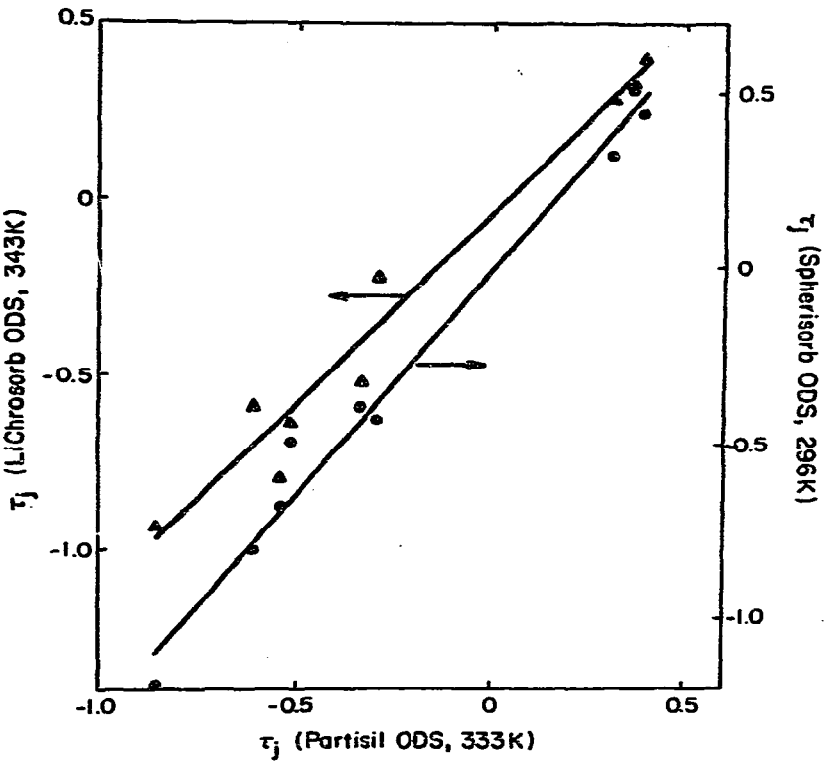

Fig. 3. Correlation between the substituent increments, τ_j, evaluated from retention data measured with the same aqueous eluent on different octadecyl-silica columns.

the transfer of a set of retardation factors from one chromatographic system to another, and thus facilitate the prediction of κ values from structural parameters under different chromatographic conditions. Furthermore, the statistical method proposed here can serve as a diagnostic tool for the study of the different retention behaviors between various chromatographic systems.

An outstanding feature of HPLC is the facility of generating a large amount of precision data. Appropriate statistical methods can not only cope with the task of data analysis, but also augment the scope and potential of this powerful analytical technique.

ACKNOWLEDGMENT

This work was supported by grants Nos. 20993 and CA21948 from the National Institute of Health, U.S. Public Health Service, DHEW.

REFERENCES

1 A. J. P. Martin, *Biochem. Soc. Symp.*, 3 (1949) 4.
2 E. C. Bate-Smith and R. G. Westall, *Biochim. Biophys. Acta*, 4 (1950) 427.
3 J. E. Leffler and E. Grunwald, *Rates and Equilibria of Organic Reactions*, Wiley, New York, 1963.

4 M. Lederer, in J. H. Schulman (Editor), *Proceedings of the 2nd International Congress on Surface Activity, Vol. III, Electrical Phenomena and Solid/Liquid Interface*, Butterworths, London, 1957, pp. 506–515.
5 C. Hansch, *Accounts Chem. Res.*, 2 (1969) 232.
6 A. Leo, C. Hansch and D. Elkins, *Chem. Rev.*, 71 (1971) 533.
7 R. Collander, *Acta Chem. Scand.*, 5 (1951) 774.
8 E. Tomlinson, *J. Chromatogr.*, 113 (1975) 1.
9 W. J. Haggerty and E. A. Murrill, *Res./Develop.*, 25 (1974) 30.
10 J. M. McCall, *J. Med. Chem.*, 18 (1975) 549.
11 R. M. Carlson, R. E. Carlson and H. L. Kopperman, *J. Chromatogr.*, 107 (1975) 219.
12 M. S. Mirrless, S. J. Moulton, C. T. Murphy and P. J. Taylor, *J. Med. Chem.*, 19 (1976) 615.
13 C. Horváth and W. Melander, *J. Chromatogr. Sci.*, 15 (1977) 393.
14 C. Horváth, W. Melander and I. Molnár, *J. Chromatogr.*, 125 (1976) 129.
15 C. Horváth, W. Melander and I. Molnár, *Anal. Chem.*, 49 (1977) 142.
16 C. Horváth, W. Melander, I. Molnár and P. Molnár, *Anal. Chem.*, 49 (1977) 2295.
17 I. Molnár and C. Horváth, *J. Chromatogr.*, 145 (1978) 371.
18 M. Kuchař, V. Rejholec, M. Jelinková and O. Neňeček, *J. Chromatogr.*, 150 (1978) 419.
19 N. R. Draper and R. Smith, *Applied Regression Analysis*, Wiley, New York, 1966.
20 D. F. Morrison, *Multivariate Statistical Methods*, McGraw-Hill, New York, 2nd ed., 1976.
21 B. K. Chen, C. Horváth and J. R. Bertino, *J. Med. Chem.*, in press.
22 W. Melander, D. E. Campbell and C. Horváth, *J. Chromatogr.*, 158 (1978) 215.