ELSEVIER

Journal of Chromatography A, 719 (1996) 45–51

JOURNAL OF CHROMATOGRAPHY A

Prediction of chromatographic properties for a group of natural phenolic derivatives by molecular topology

F.J. García-March*, G.M. Antón-Fos, F. Pérez-Giménez, M.T. Salabert-Salvador, R.A. Cercós-del-Pozo, J.V. de Julián-Ortiz

Unit of Drug Design and Molecular Connectivity Investigation, Department of Physical Chemistry, Faculty of Pharmacy, University of Valencia, 46100 Burjassot, Valencia, Spain

Abstract

A study was made of the relationship between the R_M values obtained by thin-layer chromatography for a group of phenols and connectivity indices proposed by Kier and Hall. By using multivariate regression the corresponding connectivity functions were obtained, which were selected based on their respective statistical parameters. Regression analysis of the connectivity functions showed a correct prediction of the experimental elution sequence for this group of molecules using silica gel stationary phases and mobile phases of different polarity. Random and stability studies of the different prediction models selected were carried out, and good stability and null randomness were obtained in all cases.

1. Introduction

In quantitative structure–activity relationship (QSAR) studies [1–3], molecular connectivity is a topological method capable of describing the structure of a molecule by means of numbers called indices (χ_i), calculated from the graph of the suppressed hydrogens of the molecule under study, and subsequent regression in relation to the experimental values of the physical, chemical and/or biological properties yields a series of functions, called connectivity functions [4]. Molecular topology has been shown to be a very important structural model for describing the chromatographic [5–8] and environmental [9,10] behaviour of chemicals. Topological parameters, such as molecular connectivity indices [4], can be used to quantify these properties.

In recent years, molecular connectivity studies have been used to predict several parameters related to the biological activities of drugs [11]. It was concluded that direct correlation of molecular topology with biological activity is possible [12]. Thus the chromatographic behaviour of drugs in phases of different polarity might contain information of use in describing their pharmacological behaviour, e.g., for barbiturates [13] and neuroleptics [14]. Later studies [15,16] established that chromatographic parameters in a polar stationary phase system correlate better with the valence connectivity indices, whereas Kováts parameters, obtained from the apolar phase interaction, correlate best with the non-valence connectivity terms.

Recent studies have shown that molecular connectivity satisfactorily predicts chromato-

* Corresponding author.

0021-9673/96/$15.00
SSDI 0021-9673(95)00450-5

graphic parameters such as retention times in gas–liquid chromatography [17] and R_F in thin-layer chromatography (TLC) [17,18].

In this work, the relationship between different R_M values in TLC and the connectivity indices of a group of phenols and natural phenolic derivatives was investigated in order to obtain the experimental elution sequence. We chose this group of compounds because of its structural homogeneity (phenolic derivatives). This is necessary for the application of a prediction method based on molecular topology. Most of the compounds and derivatives from natural products and some of them have pharmacological activity as antiseptic, anthelmintic, keratolytic or astringent agents.

2. Theory

Several extensive reviews have been published [11,19–21] giving detailed descriptions of the theory and method of calculation of all valence and non-valence molecular connectivity indices used in this study. The molecular connectivity indices are calculated from the equation for suppressed hydrogens of the molecule, according to the method proposed by Kier and Hall [4], and are defined for a type t subgraph with m ties, represented as ${}^{m}\chi_t$. The molecular connectivity indices are obtained as a result of the sum of the terms corresponding to the different subgraphs:

$$ {}^{m}\chi_t = \sum_{j=1}^{n_m} {}^{m}S_j \qquad (1) $$

where n_m is the number of type t and m order subgraphs and ${}^{m}S_j$ is a quantity calculated for each subgraph from the following expression:

$$ {}^{m}S_j = \left[\prod_{i=1}^{m+1} (\delta_i) \right]^{-\frac{1}{2}} \qquad (2) $$

where j defines a specific subgraph and δ_i is the valence of a vertex, which depends on the type of subgraph, that is, the subgraphs are defined by $m+1$ vertices, m being the order of the subgraph. The vertex valence, δ^{v}, of the unsaturated carbon atoms and the heteroatoms can be calculated using

$$ \delta^{v} = Z^{v} - H \qquad (3) $$

where Z^{v} is the number of valence electrons of the atom and H is the number of hydrogen atoms attached to it. We used three ${}^{m}\chi_t$ indices in our study, ranging from 0 to 4, whose types are path, cluster and path–cluster.

A multiple regression analysis was used to establish the relationship between the chromatographic properties of phenols and natural phenolic derivatives and the connectivity indices, as calculated from the equation

$$ C(\chi) = P = A_0 + \sum_{i=1}^{n} A_i \chi_i \qquad (4) $$

where P is a property and A_0 and A_i represent the regression coefficients of the equation obtained. Once the connectivity function (Eq. 4) has been established, its value for a specific molecule may be predicted. This equation was obtained by multilinear regression with the BMDP 9R program of the biostatistics BMDP (Biomedical Computer Programs) package [21]. To test the quality of the regression equations, the following statistical parameters were used: multiple correlation coefficient (r), standard error of estimates (s.e.), F-Snedecor function values (F), Mallow's CP and Student's t-test (statistical significance), and corresponding cross-validation studies of the selected functions.

Random and stability tests were carried out for the selected equations as follows.

(a) Randomness: the values of the independent variables which intervene in the equation are modified randomly. The value of the property (independent variable) is randomly modified in the same way. After each modification the BMDP 9R program is executed. Subsequently, the correlation coefficient obtained after performing the random modifications is compared with that obtained in the selected equation [22].

(b) Stability (cross-validation): the jacknife method is carried out [23], eliminating n observations by means of a random process and subsequently executing the regression program, repeating the process as many times as necessary

until the observations have been eliminated a minimum of once and a maximum of four times (in the cases where the number of eliminations is greater than one). The correlation coefficients, standard deviations and residuals obtained are subsequently compared with those of the selected equation.

3. Experimental

3.1. Chemicals and reagents

All natural phenolic derivatives and benzidine where obtained from Aldrich (Madrid, Spain), silica gel 60 F_{254} plates from SDS (Peypin, France) and Benzene, dioxane, acetic acid, methanol, hydrocholric acid and sodium nitrite from Panreac (Barcelona, Spain).

3.2. Chromatographic measurements

The different experimental values of R_M were obtained using silica gel 60 F_{254} TLC plates (6 × 12 cm, 0.25 mm layer thickness). Application was effected 1.5 cm from the edge of the plate and the run area was 21 × 21 × 9 cm, previously saturated with 100 ml of the mobile phase for 12 h. The run length was 8 cm in all cases. All of the experiments were performed at 20°C using different mobile phases in each case: (A) benzene–dioxane–acetic acid (90:25:4) and (B) benzene–methanol–acetic acid (45:8:4).

Development was carried out by drying the chromatogram with warm air and subsequently spraying it with diazotized benzidine, consisting of solution 1, 5 g of benzidine dissolved in 14 ml of concentrated hydrochloric acid and diluted with water to 1 l, and solution 2, 10% aqueous sodium nitrite solution. Equal parts of the two solutions were mixed immediately prior to use. After pulverizing, the plates were placed for a few minutes in an oven at 105°C until the stains became visible. Six chromatograms for each molecule and for each of the systems employed were obtained and the mean and standard error of each of the measurements obtained were calculated (Table 1).

4. Results and discussion

The experimental values of R_M and the molecular connectivity indices of the 23 phenols used in this study are given in Table 1.

The equations selected for R_{M_A} and R_{M_B} for the compounds studied were as follows:

$$R_{M_A} = -0.706 + 7.214^3\chi_c^v + 1.144^4\chi_{pc} - 4.975^4\chi_{pc}^v \quad (5)$$

$n = 23$; $r = 0.918$; $s = 0.134$; $F = 34.07$

$$R_{M_B} = -0.488 + 5.736^3\chi_c^v + 0.892^4\chi_{pc} - 4.271^4\chi_{pc}^v \quad (6)$$

$n = 23$; $r = 0.892$; $s = 0.117$; $F = 24.67$

Statistically, all the equations were significant above the 99.9% level. The variables present in all of the equations are significant above the 99.9% level in all cases.

In Table 2 the random study of all the equations is shown. As can be observed, in the case of R_{M_A}, one correlation coefficient >0.7 is obtained on studying both the dependent and independent variables, and therefore the probability of finding a correlation coefficient >0.9 is considerably less than 0.01 in both cases. For R_{M_B}, two correlation coefficients >0.6 are obtained when the dependent variable is studied and three correlation coefficients >0.6 when the independent variable is considered. Therefore, the probability of finding a correlation coefficient >0.9 will be less than 0.02 and 0.03, respectively. All this makes the non-randomness of the selected equations obvious.

A stability study (cross-validation) of the selected equations was carried out by varying the number of eliminations (between 1 and 5) and the number of runs (23 runs for all cases), observing that on increasing the number of eliminations the model because more unstable. This could be explained by the considerable decrease in the number of degrees of freedom. In all cases the stability studies corresponding to three eliminations (23 runs) were chosen, which correspond to approximately 10% of the observations, a value recommended by some work-

Table 1
Experimental chromatographic R_M values and connectivity indices used in the correlations of a group of phenols

Molecule	$^3\chi_c^v$	$^4\chi_{pc}$	$^4\chi_{pc}^v$	R_{M_A}	R_{M_B}
o-Methoxyphenol (guaiacol)	0.123	0.591	0.254	−0.689	−0.410
2-Hydroxybenzaldehyde (salicylaldehyde)	0.148	0.488	0.279	−0.659	−0.689
1,2-Dimethoxybenzene (veratrole)	0.118	0.584	0.302	−0.602	−0.477
Phenol	0.075	0.192	0.086	−0.501	−0.176
3,4-Dimethoxybenzoic acid (veratric acid)	0.247	1.012	0.511	−0.432	−0.368
4-Hydroxy-3-methoxybenzaldehyde (vanillin)	0.220	0.755	0.393	−0.368	−0.250
Benzaldehyde	0.096	0.179	0.150	−0.327	−0.213
1,4-Benzoquinone (quinone)	0.136	0.272	0.157	−0.269	−0.347
2-Hydroxybenzoic acid (salicylic acid)	0.182	0.708	0.333	−0.251	−0.327
p-Hydroxybenzaldehyde	0.171	0.372	0.236	−0.231	−0.105
3-Phenyl-2-propenoic acid (cinnamic acid)	0.149	0.285	0.197	−0.231	−0.269
o-Dihydroxybenzene (pyrocatechol)	0.129	0.622	0.207	−0.140	−0.070
m-Dihydroxybenzene (resorcinol)	0.149	0.359	0.161	−0.105	−0.035
7-Hydroxy-2*H*-1-benzopyran-2-one (umbelliferone)	0.348	1.080	0.650	−0.087	−0.140
p-Hydroxybenzoic acid	0.204	0.635	0.306	−0.087	−0.176
2,4-Dihydroxybenzoic acid (β-resorcyclic acid)	0.257	0.876	0.409	−0.070	−0.035
4-Hydroxy-3-methoxybenzoic acid (vanillic acid)	0.252	1.019	0.463	−0.070	−0.194
3-(4-Hydroxy-3-methoxyphenyl)-2-propenoic acid (ferulic acid)	0.272	0.860	0.439	0.000	−0.140
3-(4-Hydroxyphenyl)-2-propenoic acid (*p*-coumaric acid)	0.223	0.477	0.283	0.017	−0.035
1,2,3-Trihydroxybenzene (pyrogallol)	0.185	0.994	0.311	0.327	0.087
3,4-Dihydroxybenzaldehyde (protocatechualdehyde)	0.225	0.783	0.345	0.327	0.087
3,4-Dihydroxybenzoic acid (protocatechuic acid)	0.258	1.047	0.416	0.327	0.194
3,4,5-Trihydroxybenzoic acid (gallic acid)	0.314	1.401	0.409	0.659	0.525

Table 2
Correlation coefficients computed form random number variables for a three-variable model of R_{M_A} and R_{M_B} data for phenols

Range of *r*	Modification variable			
	Independent (100 runs)		Dependent (100 runs)	
	Number of values		Number of values	
	R_{M_A}	R_{M_B}	R_{M_A}	R_{M_B}
<0.1	1	1	3	0
0.1–0.2	14	15	10	10
0.2–0.3	18	23	26	29
0.3–0.4	32	34	39	27
0.4–0.5	26	17	14	21
0.5–0.6	7	7	3	11
0.6–0.7	1	3	4	2
0.7–0.8	1	0	1	0
0.8–0.9	0	0	0	0
>0.9	0	0	0	0

ers [11] (Tables 3 and 4). The comparison of the results between the values obtained for the model with three eliminations with those of the selected equation shows that in all cases the selected equation is stable, as low standard deviations are obtained for each of the coefficients of the independent variables present in the equations. Analysis of the residual obtained for the three eliminators model and for the selected equation shows minimum discrepancies for both the means and their standard deviations, thus reinforcing the predictive quality of the selected model.

The index $^3\chi_c^v$, which appears in both equations, increases with increasing number of ramifications of the benzene ring and/or the presence of carbonyl groups, and hence with the presence of non-σ-electrons. In both cases we generally observed a greater value of the property as the index value increases.

This index is therefore related to the polarity of the molecule and to the solvation effect,

Table 3
Statistical stability test on the regression three-variable model R_{M_A} data for phenols

Parameter	Original model (no deletions)		Three deflections per run (23 runs)	
	Regression value	Standard deviation	Regression value	Standard deviation
Correlation coefficient	0.918		0.918	0.018
Standard deviation	0.134		0.132	0.009
Coefficient of $^3\chi_c^v$	7.214	1.121	7.272	1.174
Coefficient of $^4\chi_{pc}$	1.144	0.197	1.156	0.219
Coefficient of $^4\chi_{pc}^v$	−4.975	0.650	−5.036	0.261
Constant	−0.706	0.090	−0.706	0.103
Average residual	0.098	0.015	0.100	0.011
Residuals less than one standard deviation (%)	73.91		69.38	
Residuals between one and two standard deviations (%)	26.09		29.49	
Residuals greater than two standard deviations (%)	0		1.13	

increasing as the latter effect grows. As a result, when the solvatation effect increases, a corresponding increase in R_M is observed, as in this case the eluents involved have low polarity. In confirming this effect, a slightly lower R_M value is generally seen for the more polar eluent (B), as reflected by comparing the coefficients affecting the index in both equations.

The index $^4\chi_{pc}$ increases with increasing number of adjacent substituent pairs in the benzene ring; this leads to an increase in the property, as the equation is influenced by a positive sign coefficient. The appearance of this index enables us to distinguish between structural isomers possessing different index value (*ortho* different to both *meta* and *para*), assigning differences of the same sign to evalate the property.

The presence of index $^4\chi_{pc}$ of opposite sign to $^4\chi_{pc}^v$ reflects the importance of the polar character of the eluent in the value of the property under study; this is manifested by the influence of the non-σ electrons on the property, i.e., the influence of the number of unsaturations and/or the number of heteroatoms present in the structure.

A comparison between the experimental and

Table 4
Statistical stability test on the regression three-variable model R_{M_B} data for phenols

Parameter	Original model (no deletions)		Three deflections per run (23 runs)	
	Regression value	Standard deviation	Regression value	Standard deviation
Correlation coefficient	0.892		0.895	0.028
Standard deviation	0.117		0.112	0.009
Coefficient of $^3\chi_c^v$	5.736	0.981	5.763	0.995
Coefficient of $^4\chi_{pc}$	0.892	0.172	0.890	0.193
Coefficient of $^4\chi_{pc}^v$	−4.271	0.568	−4.250	0.593
Constant	−0.488	0.079	−0.505	0.087
Average residual	0.085	0.014	0.086	0.015
Residuals less than one standard deviation (%)	82.61		81.96	
Residuals between one and two standard deviations (%)	13.04		13.80	
Residuals greater than two standard deviations	4.35		4.24	

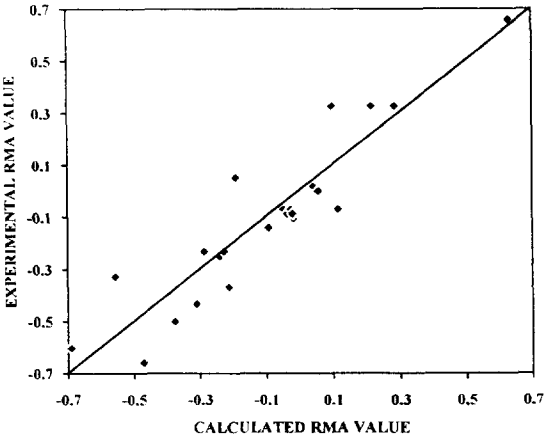

Fig. 1. Correlation between experimental and calculated R_{M_A} values of 23 phenols (Eq. 5).

theoretical values for the studied properties (Eqs. 5 and 6) is shown in Figs. 1 and 2.

The obtaining of a connectivity function with a high degree of correlation between experimental and theoretical values for this group of compounds allows us to obtain this elution sequence and to predict the behaviour of any derivative with structural homogeneity with them.

5. Conclusion

The molecular connectivity method for predicting and interpreting different R_M values in TLC was applied and statistical studies of stability (cross-validation using the jacknife method) and randomness confirmed that the predictive model is adequate for the phenol series studied.

In order to predict correctly the experimental elution sequence in the group of molecules, a three-variable model was necessary, in which the simultaneous appearance of indices ${}^3\chi_c^v$ and ${}^4\chi_{pc}$ reveals the importance of polarity and the solvation effect in the property studied, allowing the use of such equations in predicting the value of the property; structural isomers may even be differentiated in other phenol derivatives not included in the series.

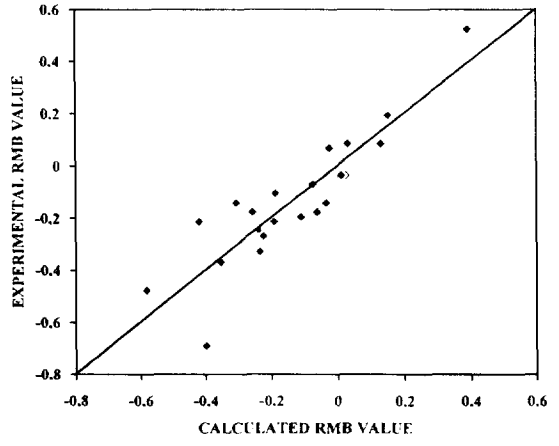

Fig. 2. Correlation between experimental and calculated R_{M_B} values of 23 phenols (Eq. 6).

Acknowledgement

The authors thank the CICYT, SAF92-0684 (Spanish Ministry of Science and Education), for financial support.

References

[1] P.N. Graig, Drug Inf. J., 18 (1984) 123.
[2] P. Gund, J.D. Andose, J.B. Rhodes and G.M. Smith, Science, 208 (1980) 1425.
[3] D. Hadzi and B. Jerman-Blazic, QSAR and Drug Design in Toxicology, Elsevier, Amsterdam, 1987.
[4] L.B. Kier and L.H. Hall, Molecular Connectivity in Chemistry and Drug Research, Academic Press, New York, 1976.
[5] A. Sabljic, J. Chromatogr., 314 (1984) 1.
[6] M. Gassiot-Matas and G. Firpo-Panies, J. Chromatogr., 187 (1980) 1.
[7] Gy. Szász, O. Papp, J. Vámos, K. Hanko-Novák and L.B. Kier, J. Chromatogr., 269 (1983) 91.
[8] J.K. Haken and I.O.O. Korhonen, J. Chromatogr., 265 (1983) 323.
[9] A. Sabljic and M. Protic, Bull. Environ. Contam. Toxicol., 28 (1982) 162.
[10] R. Koch, Toxicol. Environ. Chem., 6 (1983) 87.
[11] L.B. Kier and L.H. Hall, Molecular Connectivity in Structure–Activity Analysis, Research Studies Press, Letchworth, 1986.
[12] G.M. Antón-Fos, R. García-Doménech, F. Pérez-Giménez, J.E. Peris-Ribera, F.J. García-March and M.T. Salabert-Salvador, Arzneim.-Forsch., 44 (1994) 821.
[13] R. Kaliszan, J. Chromatogr., 220 (1981) 71.
[14] L. Buydens, D.L. Massart and P. Geerlings, J. Chromatogr. Sci., 23 (1985) 304.

[15] J.S. Millership and A.D. Woolfson, J. Pharm. Pharmacol., 32 (1980) 610.
[16] Gy. Szász, O. Papp, J. Vámos. K. Hankó-Novák and L.B. Kier, J. Chromatogr., 269 (1983) 91.
[17] R.M. Soler, F.J. García, G.M. Antón, R. García, F. Pérez and J. Gálvez, J. Chromatogr., 607 (1992) 91.
[18] G.M. Antón-Fos, F.J. García-March, F. Pérez-Giménez, M.T. Salabert-Salvador and R.A. Cercós-del-Pozo, J. Chromatogr., 672 (1994) 203.
[19] R. García, J. Gálvez, R. Moliner and F.J. García, Drug Invest., 3 (1991) 344.
[20] N. Trinasjstic, Chemical Graph Theory, CRC Press, Boca Raton, Fl, 1983.
[21] W.J. Dixon, BMD Manual: Biomedical Computer Programs, University of California Press, Los Angeles, 1982.
[22] J. Topliss and R. Costello, J. Med. Chem., 15 (1972) 1066.
[23] H.L. Gray and W.R. Shucany, The Generalized Jackknife Statistic, Marcel Dekker, New York, 1972.