Calculation of Retention Times of Anthocyanins with Orthogonalized Topological Indices

Dragan Amić and Dušanka Davidović-Amić

Faculty of Agriculture, The University of Osijek, P.O. Box 117, HR-54001 Osijek, The Republic of Croatia

Nenad Trinajstić*

The Rugjer Bošković Institute, P.O. Box 1016, HR-41001 Zagreb, The Republic of Croatia

Received September 20, 1994[⊗]

The HPLC retention times (RT) of 12 anthocyanidin malonylglucosides, which appear in flowers of *Hibiscus syriacus*, were calculated using several structure−property models based on three different types of orthogonalized topological indices, that is, the path numbers, the connectivity indices, and the Harary indices. The best agreement between the experimental and calculated values is achieved with a model with path numbers. This model is recommended to experimental researchers which are interested in quick assessment of the retention times of their newly found anthocyanin derivatives.

1. INTRODUCTION

Anthocyanins, the glycosides of anthocyanidins, are the most important group of water soluble plant pigments, largely responsible for most of the red, blue, and purple colors of flowers, fruits, and vegetables.[1] All anthocyanins have the basic flavylium (2-phenylbenzopyrylium) structure. Differences between the individual anthocyanins occur in the pattern of hydroxylation or methylation and the number and type of sugar(s) or acylated sugar(s) attached to the flavylium nuclei. Six most frequently naturally occurring anthocyanidins, pelargonidin, cyanidin, peonidin, delphinidin, petunidin, and malvidin, differ only by the number and position of hydroxyl and/or methoxyl groups in the B-ring.

Anthocyanins exhibit intense coloration in acidic pH range, between 1 and 3, and readily undergo reversible structural transformations. In slightly acidic to neutral aqueous solutions, i.e., in pH conditions close to those prevailing in their natural medium, anthocyanins exist under colored (flavylium cation and quinonoidal bases) and colorless (hemiacetal and chalcone) forms, in chemical equilibrium.[2]

As no harmfulness has been established, anthocyanins have considerable potential as food additives, although their use is restricted to those foods and beverages whose pH is below 4. A major interest of the food industry in these pigments is their use as natural colorants to replace synthetic red dyes. The most abundant anthocyanin extracted used for this purpose is derived from grape pomace.[3]

Anthocyanins have also known pharmacological properties and are used for therapeutic purposes.[4,5] They have antiinflammatory and antiulcer activity and provide protection against UV radiation. Many pharmaceutical products are now in use containing anthocyanins.

Qualitative and quantitative analyses of unknown mixtures of anthocyanins are not an easy task. The basic anthocyanidin structures of most of these compounds differ only a little in molecular weights and substituents, making separations of anthocyanin mixtures difficult. Presently, the most efficient approach for analyzing mixtures of anthocyanins is the multistep method of quantification, separation, and isolation by HPLC and peak identification by fast atom bombardment (FAB) MS and high-field NMR.[6]

In this report we reinvestigate the relationship between HPLC retention indices and the structural properties of anthocyanins using quantitative structure−chromatographic retention relationship (QSCRR) with orthogonalized topological indices. QSCRR represents a convenient model for studying the correlations between chromatographic and structural properties of molecules.[7]

TOPOLOGICAL INDICES

Different models for predicting the retention times of studied anthocyanins can be achieved by considering several types of topological indices. We used in this work the path numbers p_ℓ ($\ell = 1,2,3$), connectivity indices ${}^{\ell}\chi$ ($\ell = 1, 2, 3$), and the Harary indices H. The path number p_ℓ represents the count of paths of length ℓ in a structure.[8] The connectivity indices are defined in the standard way[9]

$$ {}^{\ell}\chi = \sum_{\text{paths}} [D(i)D(j),...,D(\ell+1)]^{-1/2} \qquad (1) $$

where $D(i)D(j),...,D(\ell+1)$ are valencies of the vertices $i,j,...\ell+1$ in the path of length ℓ. The Harary index H is defined as[10,11]

$$ H = (1/2) \sum_{i=1}^{N} \sum_{j=1}^{N} (\mathbf{D}^{r})_{ij} \qquad (2) $$

where $\mathbf{D}^{r}$ is the reciprocal distance matrix[10,12,13]

$$ (\mathbf{D}^{r})_{ij} = 1/(\mathbf{D})_{ij} \qquad i \neq j \qquad (3) $$

A variant of the Harary index H' (H'') is based on the matrix containing as elements squares (cubes) of reciprocal distance matrix.

Protocol for construction of orthogonal descriptors was detailed elsewhere.[14]

RESULTS AND DISCUSSION

Our initial model[15] was a very simple relationship between experimental HPLC retention times (RTs) and the Wiener

[⊗] Abstract published in *Advance ACS Abstracts,* December 15, 1994.

0095-2338/95/1635-0136$09.00/0

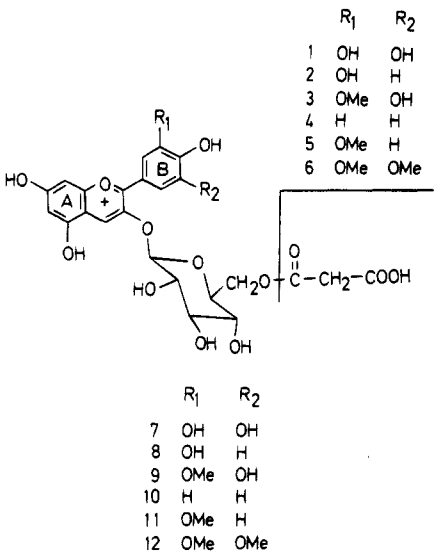

Figure 1. Structural formulas of anthocyanins studied.

Table 1. Experimental Retention Times (RT) of Anthocyanins **1**−**12**

pigment	RT (min)
delphinidin 3-*O*-glucoside (**1**)	8.62
cyanidin 3-*O*-glucoside (**2**)	11.29
petunidin 3-*O*-glucoside (**3**)	13.48
pelargonidin 3-*O*-glucoside (**4**)	14.47
peonidin 3-*O*-glucoside (**5**)	16.21
malvidin 3-*O*-glucoside (**6**)	18.53
delphinidin 3-*O*-malonylglucoside (**7**)	17.15
cyanidin 3-*O*-malonylglucoside (**8**)	20.73
petunidin 3-*O*-malonylglucoside (**9**)	22.74
pelargonidin 3-*O*-malonylglucoside (**10**)	24.14
peonidin 3-*O*-malonylglucoside (**11**)	25.97
malvidin 3-*O*-malonylglucoside (**12**)	27.08

number (W), polarity number (p), and the number of OH groups (n_{OH}) in the flavylium structure. We showed that the retention times are proportional to the Wiener number and inversely proportional to the polarity number and the number of OH groups

$$RT = a\,[W/p\,n_{OH}] + b \qquad (4)$$

The number of OH groups have the greatest influence on retention times of anthocyanins, that is, as the number of hydroxyl groups in the B-ring increases, the retention times of the flavyliums decreases. A reasonable rationalization of the above observation appears to be related to the formation of the hydrogen bonds between the flavylium salts and the polar eluent, which can greatly influence retention times and separability. Therefore, the number of hydrogen atoms bonded to oxygen atoms is a possible parameter for describing hydrogen bonding. Consequently, the more OH groups in the flavylium structure, the stronger the hydrogen bonding with polar eluent and the shorter the retention time.

We tested the validity of model 4 for 12 anthocyanidin malonylglucosides which appear in flowers of *Hibiscus syriacus*. Structures of these pigments are shown in Figure 1. Table 1 gives their experimental HPLC retention times.[16] The statistical characteristics of the above model are as follows: $n = 12$, $r = 0.9848$, $s = 1.07$, $F = 321$. This model gives a rather high value of the standard error s and the wrong order of elution for only one compound.

After this initial work we tested a few more models. The simplest model that we could think of is

$$RT = a[n_{OH}/n] + b \qquad (5)$$

The quantity n_{OH}/n is a quotient between the number of OH groups and the number of atoms of a considered anthocyanin. We already pointed out the importance of the OH groups for the retention times of anthocyanins. The quotient n_{OH}/n scales their influence by considering the size of the molecule. Surprisingly, this model has better statistical characteristic than model 4, that is $n = 12$, $r = 0.997$, $s = 0.48$, and $F = 1621$. Additionally, we tested the validity of model 5 for different types of falvylium salts, anthocyanidins, anthocyanin 3-glucosides, and anthocyanin 3,5-diglucosides, compounds studied by us with model 4.[15] The statistical characteristics for correlations were of poor quality; associated standard errors were too high.

Further we applied multiple regression analysis using orthogonal combinations of molecular descriptors. As molecular descriptors we have adopted the path numbers p_1, p_2, p_3, connectivity indices ${}^1\chi$, ${}^2\chi$, ${}^3\chi$ and Harary indices H, H', H''. In Table 2 we give numerical values of these indices for 12 studied anthocyanins. All indices were considered in their nonorthogonalized and orthogonalized form. Construction of orthogonalized descriptors has been outlined recently. We emphasize here some significant features of orthogonalization procedure.[17,18] The standard error s and the coefficient of correlation r for the correlation do not change with the orthogonalization, i.e., they are not affected whether a basis of orthogonal or nonorthogonal descriptors is used. Also note that the coefficients of orthogonalized regression equations occur as "diagonal" entries in nonorthogonal equations and, therefore, can be extracted without actually constructing orthogonal descriptors.

In Tables 3−5 we report the regression equations for retention indices of 12 anthocyanins with the path numbers (Table 3), the connectivity indices (Table 4), and the Harary indices (Table 5).

Comparison of these three tables shows that the model with three path numbers is superior to the other models. The regression equations with the nonorthogonalized and orthogonalized path numbers p_1−p_3 are as follows

(i) Nonorthogonalized path numbers p_1−p_3

$$RT = -6.8001(\pm 1.4832) + 12.5025(\pm 0.3439)p_1 - 8.7775(\pm 0.2965)p_2 + 0.6295(\pm 0.0770)p_3$$

(ii) Orthogonalized path numbers ${}^1\Pi$−${}^3\Pi$

$$RT = -37.6432(\pm 1.1277) + 1.4362(\pm 0.0288){}^1\Pi - 7.1807(\pm 0.2231){}^2\Pi + 0.6295(\pm 0.0770){}^3\Pi$$

A plot of $(RT)_{exp}$ versus $(RT)_{calc}$ using the model with orthogonalized path numbers p_1−p_3 is given in Figure 2.

Qualitative analysis of obtained results based on the path numbers allowed some insight into the structural factors which dominate the chromatographic process. The paths of length one, p_1, represent the number of bonds in molecule. It is reasonable to predict that this molecular descriptor parallels molecular size or molecular volume. It is well-known that chromatographic retention times are dependent on molecular size. The paths of length two, p_2, is the next natural variable that discriminates among similar structures. Some types of anthocyanins differs only by a few bonds from each other. It was Wiener[19] who already in 1947 introduced

Table 2. Path Numbers (p), Connectivity Indices (χ), and Harary Indices (H) of Anthocyanin-3-O-glucosides

compd no.	p_1	p_2	p_3	$^1\chi$	$^2\chi$	$^3\chi$	H	H'	H''
1	36	54	62	11.6201	9.9599	7.6915	223.927	242.345	653.769
2	35	52	59	11.5286	9.8109	7.6362	212.376	229.643	618.050
3	37	55	64	12.0059	10.1667	7.9717	232.493	246.244	656.946
4	34	50	56	11.4370	9.6717	7.5559	201.179	217.065	582.375
5	36	53	61	11.9143	10.0177	7.9147	220.664	233.465	621.205
6	38	56	66	12.3916	10.3735	8.2520	241.291	250.197	660.134
7	42	62	69	13.3744	11.2305	8.3520	275.474	295.748	798.776
8	41	60	66	13.2828	11.0816	8.2967	263.338	282.988	763.050
9	43	63	71	13.8292	11.3689	8.6830	284.587	299.697	801.957
10	40	58	63	13.1913	10.9424	8.2164	251.554	270.353	727.369
11	42	61	68	13.6686	11.2884	8.5752	272.171	286.860	766.210
12	44	64	73	14.1459	11.6442	8.9125	293.931	303.700	805.150

Table 3. Regression Equations for Retention Indices with the Path Numbers p_1–p_3

(a) Nonorthogonalized

statistical data			regression eq coeffs			
F	r	s	p_1	p_2	p_3	const
22.36	0.8313	3.44	1.436			−37.643
470.38	0.9895	0.89	11.198	−7.181		−6.639
4484.00	0.9989	0.29	12.503	−8.777	0.630	−6.800

(b) Orthogonalized

statistical data			regression eq coeffs			
F	r	s	$^1\Pi$	$^2\Pi$	$^3\Pi$	const
22.36	0.8313	3.44	1.436			−37.643
470.38	0.9895	0.89	1.436	−7.181		−37.643
4484.00	0.9989	0.29	1.436	−7.181	0.630	−37.643

Table 4. Regression Equations for Retention Indices with the Connectivity Indices $^1\chi$–$^3\chi$

(a) Nonorthogonalized

statistical data			regression eq coeffs			
F	r	s	$^1\chi$	$^2\chi$	$^3\chi$	const
39.76	0.8939	2.77	5.350			−49.569
161.03	0.9703	1.49	31.113	−36.472		10.943
168.54	0.9716	1.46	33.301	−37.941	−2.730	21.080

(b) Orthogonalized

statistical data			regression eq coeffs			
F	r	s	$^1\Omega$	$^2\Omega$	$^3\Omega$	const
39.76	0.8939	2.77	5.350			−49.569
161.00	0.9703	1.49	5.350	−36.478		−49.569
168.52	0.9716	1.46	5.350	−36.478	−2.731	−49.569

use of the number of paths of length three, p_3, as a potentially useful molecular descriptor for studying molecular properties. The number of pairs of atoms separated by three bonds was originally termed by Wiener as a polarity number, although some, e.g., Platt,[20] suggested that the paths of length three are more properly related to steric aspects of a molecule. Experimental results have shown that the polarity of the anthocyanins is the most important factor affecting the HPLC retention times.[21]

It was also interesting to see how a model with path numbers works in the case of 10 anthocyanin 3-glucosides and 10 anthocyanin 3,5-diglucosides recently studied by us with model 4.[15] Reproduction of experimental elution sequences were successful, and in the case of anthocyanin 3,5-diglucosides statistical characteristics were even better than with model 4. Regression equations and statistical characteristics using orthogonalized path numbers are as

Table 5. Regression Equations for Retention Indices with the Harary Indices H–H''

(a) Nonorthogonalized

statistical data			regression eq coeffs			
F	r	s	H	H'	H''	const
17.65	0.7990	3.72	0.155			−20.117
21.28	0.8248	3.50	0.480	−0.328		−14.280
54.16	0.9188	2.44	−5.934	18.086	−4.496	−103.706

(b) Orthogonalized

statistical data			regression eq coeffs			
F	r	s	$^1\sigma$	$^2\sigma$	$^3\sigma$	const
17.65	0.7990	3.72	0.155			−20.117
21.28	0.8248	3.50	0.155	−0.328		−20.177
54.17	0.9188	2.44	0.155	−0.328	−4.514	−20.177

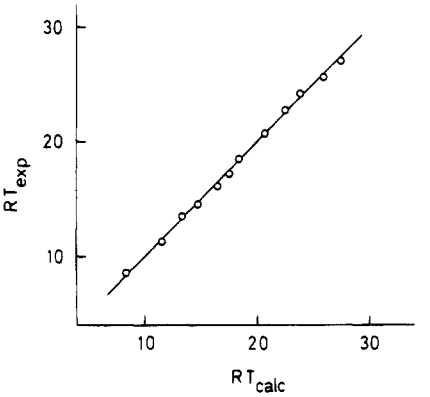

Figure 2. Plot of RT_{exp} vs RT_{calc} for 12 anthocyanins.

follows (values in parentheses for r, s, and F are obtained using model 4):

(i) Anthocyanidin 3-glucosides

$$RT = -160.6269(\pm 16.9740) + 5.6796(\pm 0.3963)^1\Pi - 30.4143(\pm 7.0948)^2\Pi + 2.9277(\pm 2.2159)^3\Pi$$

$$r = 0.9870(0.9946) \quad s = 6.59(4.25) \quad F = 302(739)$$

(ii) Anthocyanidin 3,5-glucosides

$$RT = -135.6655(\pm 4.8334) + 3.4547(\pm 0.0883)^1\Pi - 15.0762(\pm 1.5809)^2\Pi + 0.4979(\pm 0.4938)^3\Pi$$

$$r = 0.9982(0.9940) \quad s = 1.47(2.66) \quad F = 2164(656)$$

Thus, the results discussed indicate that the model with path numbers represent improvement over the model 4. The

path model also reproduces the experimental retention sequence of all studied types of anthocyanins.

3. CONCLUDING REMARKS

To conclude we point out that the model with three path numbers reproduce satisfactorily the retention times of studied anthocyanins. It appears that for practical purposes the path model is quite satisfactory for experimental chemists. It is easy to calculate and offers quick assessment of the retention times of anthocyanin derivatives.

ACKNOWLEDGMENT

This work was supported in part by Grant No. 1-07-159 awarded by the Ministry of Science and Technology of the Republic of Croatia.

REFERENCES AND NOTES

(1) Harborne, J. B. *Comparative Biochemistry of the Flavonoids*; Academic Press: New York, 1967.
(2) Brouillard, R. *Chemical Structure of Anthocyanins*. In *Anthocyanins as Food Colors*; Markakis, P., Ed.; Academic Press: New York, 1982.
(3) Markakis, P. *Anthocyanins as Food Additives*. In *Anthocyanins as Food Colors*; Markakis, P., Ed.; Academic Press: New York, 1982.
(4) Ferrell, J. E., Jr.; Chang Sing, P. D. G.; Loew, G.; King, R.; Mansour, J. M.; Mansour, T. E. Structure/Activity Studies of Flavonoids as Inhibitors of Cyclic AMP Phosphodiesterase and Relationship to Quantum Chemical Indices. *Mol. Pharmacol.* **1979**, *16*, 556−568.
(5) Baj, A.; Bombardelli, E.; Gabetta, B.; Martinelli, E. M. Qualitative and Quantitative Evaluation of *Vaccinium Myrtillus* Anthocyanins by High-Resolution Gas Chromatography and High-Performance Liquid Chromatography. *J. Chromatogr.* **1983**, *279*, 365−372.
(6) Mazza, G.; Miniati, E. *Anthocyanins in Fruits, Vegetables and Grains*; CRC Press: Boca Raton, FL, 1993.
(7) Kaliszan, R. *Quantitative Structure−Chromatographic Retention Relationships*; Wiley: New York, 1988.
(8) Trinajstić, N. *Chemical Graph Theory*, 2nd revised ed.; CRC Press: Boca Raton, FL, 1992.
(9) Reference 8, p 228.
(10) Plavšić, D.; Nikolić, S.; Trinajstić, N.; Mihalić, Z. On the Harary Index for the Characterization of Chemical Graphs. *J. Math. Chem.* **1993**, *12*, 235−250.
(11) Mihalić, Z.; Trinajstić, N. A Graph-Theoretical Approach to Structure-Property Relationships. *J. Chem. Educ.* **1992**, *62*, 701−712.
(12) Balaban, T.-S.; Filip, P. A.; Ivanciuc, O. Computer Generation of Acyclic Graphs Based on Local Vertex Invariants and Topological Indices. Derived Canonical Labelling and Coding of Trees of Alkanes. *J. Math. Chem.* **1992**, *11*, 79−105.
(13) Balaban, T. A. Local versus Global (i.e. Atomic versus Molecular) Numerical Modeling of Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 398−402.
(14) Randić, M.; Trinajstić, N. Isomeric Variations in Alkanes: Boiling Points of Nonanes. *New J. Chem.* **1994**, *18*, 179−189.
(15) Amić, D.; Davidović-Amić, D.; Trinajstić, N. Application of Topological Indices to Chromatographic Data. Calculation of the Retention Indices of Anthocyanins. *J. Chromatogr.* **1993**, *A 653*, 115−121.
(16) Kim, J. H.; Nonaka, G.-I.; Fujieda, K.; Uemoto, S. Anthocyanidin Malonylglucosides in Flowers of *Hibiscus Syriacus*. *Phytochemistry* **1989**, *28*, 1503−1506. Details of HPLC analysis were fully outlined in this paper. For additional information, see: Wulf, L. W.; Nagel, C. W. High-Pressure Liquid Chromatographic Separation of Anthocyanins of *Vitis Vinifera*. *Am. J. Enol. Vitic.* **1978**, *29*, 42−49.
(17) Randić, M. Orthogonal Molecular Descriptors. *New J. Chem.* **1991**, *15*, 517−525.
(18) Randić, M. Resolution of Ambiguities in Structure−Property Studies by Use of Orthogonal Descriptors. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 311−320.
(19) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17−20. Correlation of Heats of Isomerization and Differences in Heats of Vaporization of Isomers Among the Paraffin Hydrocarbons. *J. Am. Chem. Soc.* **1947**, *69*, 2636−2638.
(20) Platt, J. R. Prediction of Isomeric Differences in Paraffin Properties. *J. Phys. Chem.* **1952**, *56*, 328−336.
(21) Williams, M.; Hrazdina, G.; Wilkinson, M. M.; Sweeny, J. G.; Iacobucci, G. A. High-Pressure Liquid Chromatographic Separation of 3-Glucosides, 3,5-Diglucosides, 3-(6-O-*p*-Coumaryl)glucosides and 3-(6-O-*p*-Coumarylglucoside)-5-Glucosides of Anthocyanidins. *J. Chromatogr.* **1978**, *155*, 389−398.

CI9401074