Prediction of p*K* Values, Half-Lives, and Electronic Spectra of Flavylium Salts from Molecular Structure

Dragan Amić,* Dušanka Davidović-Amić, and Drago Bešlo

Faculty of Agriculture, The Josip Juraj Strossmayer University, P.O. Box 719,
HR-31001 Osijek, The Republic of Croatia

Bono Lučić and Nenad Trinajstić

The Rugjer Bošković Institute, P.O. Box 1016, HR-10001 Zagreb, The Republic of Croatia

Simple QSAR/QSPR models have been developed which enable the prediction of the p*K* values and half-lives of flavylium salts as a function of their molecular descriptors. On the basis of these results, as well as calculated UV−vis spectra, flavylium salt model structures are proposed, which are expected to be useful as potential food dyes.

INTRODUCTION

The anthocyanins are glycosides of flavylium or 2-phenylbenzopyrylium salts.[1] These are water-soluble plant pigments and are responsible for most of the orange, red, and blue colors of flowers and fruits. Apparently harmless to health, anthocyanins have considerable potential in the food industry as safe and effective food colorants.[2] There has also been current medicinal interest in anthocyanins as biologically active substances.[3] At present, much work is being carried out on the use of natural flavylium salts as food-coloring materials.[4] This use, due to their structural transformations,[5] is rather restricted. In very acidic solutions, the flavylium salts exist primarily in the form of the red (R_3 = OR′) or yellow (R_3 = H) flavylium cation AH$^+$. As the pH increases (ranging from 3 to 6) competition between water addition and proton-transfer reactions occurs: flavylium cation AH$^+$ is rapidly hydrated to colorless pseudobase B which is in a fast cycle−chain tautomeric equilibrium with the colorless chalcone C, and/or a rapid proton loss occurs to yield the red or blue quinoidal base A. The pseudobase B and the quinoidal base A are the thermodynamic and kinetic products, respectively. Equilibrium between the quinoidal base A and the pseudobase B occurs exclusively via the flavylium cation AH$^+$.[6] Figure 1 illustrates structural transformations of flavylium salts with pH in acidic aqueous solutions.

Most natural anthocyanins tend to exist in acidic solutions (pH range 2.5−6.0 for most food systems) predominantly as the colorless forms B and C. Therefore, it is of interest for food application purposes to find anthocyanins or related flavylium salts that would exist, under those pH conditions, largely in the colored forms AH$^+$ and A.

Three decades ago it was suggested that modification in the structure may result in synthetic flavylium salts with potential use as color additives for fruit drinks and juices.[7]
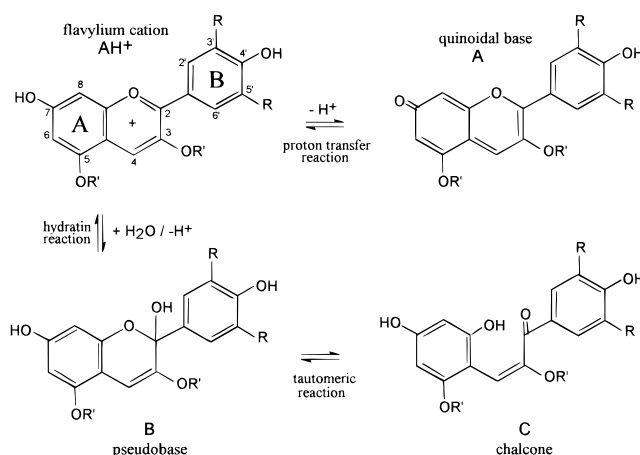
Figure 1. Structural transformations of flavylium salts. R = H, OH, OCH$_3$; R′ = H, sugar.

Attempts to improve the stability of cationic form AH$^+$ as well as quinoidal form A resulted in the following findings.[8] The pigment molecules get stabilized by hydroxylation at the 5 and 7 positions in the benzopyrylium ring, and labilized by the same modification at the 3 position. The stability of the pigment molecules increased with increasing methoxylation in the B ring. Flavylium salts containing a 4-methyl or 4-phenyl group are very stable in acidic and neutral media.[9] By varying the substitution pattern of the flavylium core, salts can be prepared that in nearly neutral solutions can exist primarily in the colored quinoidal base form A.[10] The substitution of a carboxy group at the 4 position gives exceptionally stable salts.[11]

Despite the considerable work performed dealing with natural anthocyanin-type colorants as well as related compounds based on flavylium structure, the use of developed products is very limited.[12] The aim of this work was to evaluate simple QSAR/QSPR models which enable the prediction of the p*K* values and half-lives of flavylium salts as a function of molecular descriptors. On the basis of these results and calculated UV−vis spectra, flavylium salt model

* To whom correspondence should be addressed. E-mail: damic@suncokret.pfos.hr.

968 *J. Chem. Inf. Comput. Sci., Vol. 39, No. 6, 1999*

AMIĆ ET AL.

compounds could be proposed that would exhibit stability and coloration suitable for food-coloring purposes.

METHODS

Molecular Descriptors and Data Sets. To develop the QSAR/QSPR for flavylium salts, a variety of molecular descriptors are used to encode the effect of electronic and topological properties on pK and half-life. The electronic descriptors include the absolute hardness[13] η, the energy of the highest occupied molecular orbital E_{HOMO}, the energy of the lowest unoccupied molecular orbital E_{LUMO}, the dipole moment DM, the π charge on the particular C atom in the flavylium core Q_r, the sum of the π charges located in various fragments of the benzopyrylium or phenyl part of the flavylium chromophore, and superdelocalizability indices[14] S_r^N. The topological descriptors that were employed include the distance indices.[15] In addition, various indicator variables I_i were used. Indicator variables express the contribution of certain substituents on the magnitude of the studied property. Typically, such variables flag specific chemical structural features by assigning them a value of 1 for molecules having the feature and 0 otherwise.

To obtain the models for the pK and half-life of flavylium salts, we used 40 and 10 compounds, respectively. For both data sets 39 descriptors were computed and used in the modeling. In addition, nonlinear descriptors were also taken into account through the cross-product terms of the initial descriptors. In such a way 780 nonlinear terms were produced (by calculating the cross-products of k descriptors, $(k + 1)\cdot k/2$ nonlinear descriptors are obtained). Then, 39 initial descriptors were added to nonlinear terms, which gave a new data set containing 819 descriptors.

Generation of Models. The models obtained in this study are nonlinear multiregression (MR) models. The procedure used for generation of the models was described in our previous papers.[16,17] Here, we summarize that and explain the differences.

From data sets containing 819 descriptors the best possible MR models (according to the highest fitted and leave-one-out cross-validated correlation coefficients and standard errors) containing four (pK modeling) and two descriptors (half-life modeling) were selected. In the case of half-life modeling this gave the final models. But, in the case of pK modeling we had 40 compounds, and we decided to include up to eight descriptors in the final models. This was done by application of the "i by i" stepwise selection procedure, which is detailed in refs 16 and 17. Using this procedure, four descriptors were selected and added to four previously selected descriptors, which gave us the best stepwise MR models for pK containing eight descriptors. In addition, we generated not only the best model but also several top models: four top models for half-life and eight for pK modeling. Prediction of the pK and half-life for a new compound was calculated as an average of predicted values obtained using the top models.

The computations were done on a Hewlett-Packard 9000/E55 (PARISC 7100LC processor, 100 MHz) and a SUN Enterprise 3000 (UltraSPARC processor, 250 MHz) in multiuser mode. The FORTRAN 77 computer programs were developed for the selection of (1) the best possible multi-

Table 1. Parameters Used in the PPP Calculations

atom, μ	I_μ/eV	A_μ/eV	$\gamma_{\mu\mu}$/eV	$\beta_{C\mu}$/eV	n_π	$l_{C\mu}$/Å	ref
C	11.16	0.03	11.13	−2.349	1	1.390	20
O$^+$	34.95	15.30	19.65	−2.800	1	1.360	21
=O	17.28	2.70	14.58	−3.070	1	1.240	22
−O−	35.76	17.70	18.06	−2.270	2	1.360	22
OH	26.14	7.32	18.82	−1.950	2	1.360	21
OCH$_3$	33.00	11.47	21.53	−1.808	2	1.370	20
CH$_3$	28.43	12.99	15.44	−1.673	2	1.500	23
OH from COOH	25.43	10.85	14.58	−2.550	2	1.360	22

regression models, and (2) the best stepwise multiregression models ("the best" means the best according to the highest fitted and cross-validated correlation coefficient and the lowest standard error).

Validation of Models. We have assessed the predictive ability of the models by fit and leave-one-out and leave-20%-out cross-validation parameters. The fit ability is judged by the correlation coefficient (R), the root-mean-square error (S), and the F-test. For each descriptor involved in the MR equation the corresponding regression coefficient and the standard error of regression coefficient are also given, so that the t-test value can be easily calculated (as their ratio). The predictive quality is measured by the corresponding leave-one-out (R_{cv}, S_{cv}) and leave-20%-out ($R(20\%)_{cv}$, $S(20\%)_{cv}$) cross-validated correlation coefficients and root-mean-square errors. In the case of leave-20%-out cross-validation 20 independent runs with randomized groups were performed, and the average values of the $R(20\%)_{cv}$ and $S(20\%)_{cv}$ were reported.

The chance correlation risk was tested by scrambling of y-variables (pK or half-life) for each selected model. This was done in such a way that y-values were reordered in a random manner, and after that, the chance correlation coefficient was calculated. For each selected model 2000 different randomization runs were performed, and the percentages of chance correlations with $R^2 \geq 0.5$ (for pK models) and with $R^2 \geq 0.8$ (for half-life models) were determined, as well as maximum (max$R$) and minimum (min$R$) chance correlation coefficient values. For the models having the maximum chance correlation coefficient (max$R$) the corresponding cross-validated correlation coefficient was also reported.

Theoretical Calculations of Electronic Absorption Spectra. All calculations were carried out using the standard version of the PPP (Pariser−Parr−Pople) method.[18] Idealized, planar geometries have been assumed, with the rings taken as regular hexagons. All C−C bonds within the rings were set equal to 1.39 Å. All valence angles have been taken to be equal 120°. We have used a consistent set of valence state energy parameters presented in Table 1. These parameters have been developed and tested in calculations of various flavylium structures.[19] In Table 1, I_μ and A_μ are the ionization potential and electron affinity of atom μ in the atomic valence state, respectively. The monocentric electronic repulsion integrals and the core resonance integrals between nearest neighbors are designated as $\gamma_{\mu\mu}$ and $\beta_{C\mu}$, respectively. n_π is the number of electrons contributed to the π-system and $l_{C\mu}$ (Å) is the distance between atoms C and μ. The bicentric electronic repulsion integrals were calculated using the Mataga−Nishimoto formula.[24]

FLAVYLIUM SALT MOLECULAR STRUCTURE

J. Chem. Inf. Comput. Sci., Vol. 39, No. 6, 1999 **969**

Table 2. List of Flavylium Salts, the Experimental and Calculated pK Values, and References

compd no.	substitution pattern								exptl pK ± error	calcd pK^a	ref
	3	4	5	7	8	3′	4′	5′			
1	OH	H	H	OH	H	OCH$_3$	OH	H	1.39 ± 0.06	1.53	8
2	OH	H	OH	OH	H	OCH$_3$	OH	H	1.82 ± 0.04	2.12	8
3	OCH$_3$	H	OCH$_3$	OCH$_3$	H	OCH$_3$	OCH$_3$	H	2.31 ± 0.02	2.43	25
4	OCH$_3$	H	OCH$_3$	OCH$_3$	H	OCH$_3$	OH	OCH$_3$	2.33 ± 0.01	2.46	25
5	OCH$_3$	H	OCH$_3$	OH	H	OCH$_3$	OH	OCH$_3$	2.54 ± 0.01	2.51	25
6	OCH$_3$	H	OCH$_3$	OH	H	OCH$_3$	OCH$_3$	H	2.55 ± 0.02	2.47	25
7	H	H	H	OH	H	OCH$_3$	OCH$_3$	OCH$_3$	2.56 ± 0.07	2.86	8
8	H	H	H	H	H	H	H	H	2.60 ± na	2.69	26
9	OH	H	OH	OH	H	H	OH	H	2.63 ± 0.14	2.45	27
10	OCH$_3$	H	OCH$_3$	OCH$_3$	H	OCH$_3$	OH	H	2.63 ± 0.03	2.70	25
11	OCH$_3$	H	OCH$_3$	OCH$_3$	H	OH	OH	H	2.64 ± 0.02	2.66	25
12	H	H	H	OH	H	H	H	H	2.70 ± 0.04	2.53	8
13	OCH$_3$	H	OCH$_3$	OH	H	OH	OH	H	2.86 ± 0.03	2.70	25
14	OCH$_3$	H	OCH$_3$	OCH$_3$	H	H	OH	H	2.89 ± 0.02	2.96	25
15	H	H	H	OH	H	OCH$_3$	OCH$_3$	H	2.98 ± 0.06	2.97	8
16	OCH$_3$	H	OCH$_3$	OH	H	OCH$_3$	OH	H	2.99 ± 0.04	2.74	25
17	H	H	H	OH	H	OH	OH	OH	3.07 ± 0.08	2.97	8
18	H	H	H	OH	H	H	OCH$_3$	H	3.14 ± 0.05	3.11	8
19	H	H	H	OH	H	OH	OH	H	3.18 ± 0.07	3.19	8
20	H	H	H	OH	H	OCH$_3$	OH	H	3.20 ± 0.04	3.28	8
21	OCH$_3$	H	OCH$_3$	OH	H	H	OH	H	3.21 ± 0.05	2.98	25
22	H	H	H	OH	H	H	OH	H	3.30 ± 0.02	3.47	8
23	H	Ph	CH$_3$	OH	H	H	H	H	3.32 ± 0.17	3.91	28
24	H	H	H	H	OCH$_3$	H	OCH$_3$	H	3.65 ± 0.08	3.67	29
25	H	H	OCH$_3$	OH	H	H	OH	H	3.92 ± 0.12	3.72	30
26	H	H	OH	OH	H	OH	OCH$_3$	H	3.96 ± 0.11	3.87	30
27	H	H	H	H	H	OCH$_3$	OH	H	4.09 ± 0.09	4.08	8
28	H	CH$_3$	OH	OH	H	H	OCH$_3$	H	4.10 ± 0.03	3.95	31
29	H	H	OH	OH	H	OCH$_3$	OH	H	4.17 ± 0.12	4.11	8
30	H	H	OH	OH	H	H	OH	H	4.23 ± 0.05	4.33	30
31	H	CH$_3$	H	OH	OH	H	H	H	4.25 ± 0.04	4.20	32
32	H	Ph	H	OH	H	H	H	H	4.27 ± 0.06	3.74	33
33	H	H	H	H	H	H	OH	H	4.33 ± 0.02	4.32	8
34	H	CH$_3$	OH	OH	H	H	H	H	4.38 ± 0.10	4.43	32
35	H	Ph	H	OH	H	H	OCH$_3$	H	4.44 ± 0.09	4.18	28
36	H	CH$_3$	H	OH	H	H	H	H	4.44 ± 0.03	4.49	33
37	H	Ph	H	OH	H	H	OH	H	4.49 ± 0.03	4.45	28
38	H	COOH	H	OH	H	H	OCH$_3$	H	4.78 ± 0.10	5.01	34
39	H	CH$_3$	H	OH	H	H	OCH$_3$	H	4.80 ± 0.03	4.75	31
40	H	CH$_3$	H	OH	H	H	OH	H	4.84 ± 0.03	4.99	28

^{a} Calculated pK by model 1.

RESULTS AND DISCUSSION

Development of a pK Model for Flavylium Salts. The development of a model describing the pK of flavylium salts was our first goal since the coloration is very sensitive to changes in pH. The set of 40 flavylium compounds used in this study was taken from the literature. The data in this set were gathered from many different sources.[8,25−34] The structures of flavylium salts, their experimental (including experimental errors) and calculated pK values, and the reference from which the values were taken are listed in Table 2. The position of the overall transformation AH$^+$ = (A + B + C) + H$^+$ depicted in Figure 1 is characterized by the overall acidity constant K, i.e., the corresponding pK value.

pK is a measure of the stability of the flavylium cation AH$^+$ as a function of pH. A total of 819 descriptors for each flavylium derivative were generated (39 initial descriptors and their cross-products). Though we examined other possibilities, we found that eight-descriptor models led to the most accurate models. Eight top models were selected, and they are given in Table 3. The top models were obtained using the i by i stepwise selection as described in the previous section. According to the leave-20%-out cross-validated

parameters and chance correlation parameters, one can see that the models are stable, having a low level of chance correlation risk. Comparison of experimental pK errors given in Table 2 and standard errors of estimate for the models from Table 3 indicates that the models are less accurate. The best models that were generated involve sixteen descriptors: seven electronic descriptors (E_{HOMO}, $Q_{2,4}$, Q_4, Q_{fl}, Q_{ph}, S_2^N, S_4^N), three topological descriptors (H, H', MTI), and six indicator variables that encode particular structural features of flavylium salts responsible for their acidities. The electronic descriptors (π charges Q_r, superdelocalizabilities S_r^N) describe the ability of flavylium salts to undergo hydration reactions. Because E_{HOMO} is directly related to the ionization potential, one assumes that it is related to the ease of protonation. The topological indexes were the Harary index H, the variant of the Harary index H', and the molecular topological index MTI. They describe the size and shape of the molecules. The indicator variables are combinations of electronic and steric features of the molecules describing the ease of deprotonation reactions and/or hydration reactions.

Figure 2 shows a plot of pK values predicted using model 1 from Table 3 vs measured values.

Table 3. The Eight Best Stepwise Nonlinear Eight-Descriptor Multiregression Models for Calculating pK Values
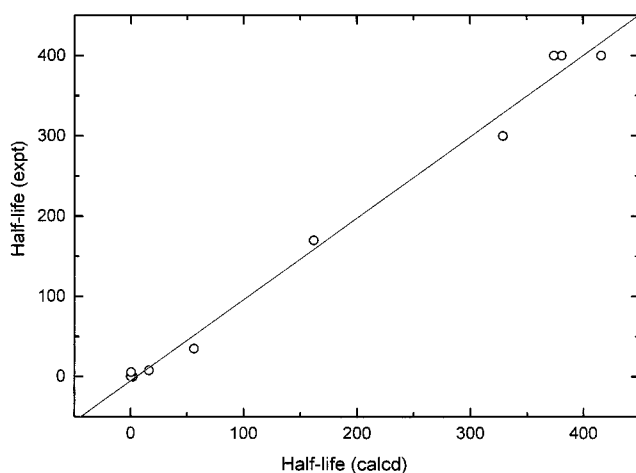
$N = 40$; the total number of descriptors screened, 819

Model 1

$$p K = 7.46(\pm 0.56) - 9.8(\pm 1.4)d2 \cdot d14 + 2.95(\pm 0.23)d1 \cdot d13 + 0.62(\pm 0.16)d1 \cdot d16 + 0.729(\pm 0.061)d4 \cdot d7 - 2.37(\pm 0.19)d5 \cdot d6 + 8.0(\pm 1.2)d6 \cdot d14 + 0.0066(\pm 0.0018)d8 \cdot d11 - 1.12(\pm 0.19)d11 \cdot d15$$

$R = 0.9775$, $S = 0.185$, $R_{cv} = 0.9591$, $S_{cv} = 0.248$, $F = 83$
leave-20%-out cross-correlation, 20 runs: av$R(20\%)_{cv} = 0.956$, av$S(20\%)_{cv} = 0.256$
random correlation:[a] max$R = 0.773$ ($R_{cv} = 0.582$), min$R = 0.134$, $p(R^2 \geq 0.5) = 0.45\%$[b,c]

Model 2

$$p K = 7.12(\pm 0.54) - 9.8(\pm 1.5)d2 \cdot d14 + 3.06(\pm 0.22)d1 \cdot d13 + 0.61 \pm 0.16)d1 \cdot d16 + 0.702(\pm 0.065)d4 \cdot d7 - 2.27(\pm 0.20)d5 \cdot d6 + 7.9(\pm 1.2)d6 \cdot d14 + 0.0056(\pm 0.0016)d9 \cdot d11 - 1.11(\pm 0.20)d11 \cdot d15$$

$R = 0.9768$, $S = 0.187$, $R_{cv} = 0.9585$, $S_{cv} = 0.250$, $F = 81$
leave-20%-out cross-correlation, 20 runs: av$R(20\%)_{cv} = 0.954$, av$S(20\%)_{cv} = 0.261$
random correlation:[a] max$R = 0.748$ ($R_{cv} = 0.565$), min$R = 0.116$, $p(R^2 \geq 0.5) = 0.3\%$[c]

Model 3

$$p K = 7.37(\pm 0.56) - 9.8(\pm 1.5)d2 \cdot d14 + 0.89(\pm 0.27)d3 \cdot d12 + 3.15(\pm 0.21)d1 \cdot d13 + 0.57(\pm 0.17)d1 \cdot d16 + 0.698(\pm 0.066)d4 \cdot d7 - 2.28(\pm 0.20)d5 \cdot d6 + 8.0(\pm 1.2)d6 \cdot d14 - 1.07(\pm 0.19)d11 \cdot d15$$

$R = 0.9766$, $S = 0.188$, $R_{cv} = 0.9582$, $S_{cv} = 0.251$, $F = 80$
leave-20%-out cross-correlation, 20 runs: av$R(20\%)_{cv} = 0.940$, av$S(20\%)_{cv} = 0.286$
random correlation:[a] max$R = 0.749$ ($R_{cv} = 0.567$), min$R = 0.164$, $p(R^2 \geq 0.5) = 0.35\%$[c]

Model 4

$$p K = 7.51(\pm 0.57) - 6.10(\pm 0.92)d2 \cdot d14 + 2.99(\pm 0.24)d1 \cdot d13 + 0.60(\pm 0.16)d1 \cdot d16 + 0.729(\pm 0.063)d4 \cdot d7 - 2.37(\pm 0.19)d5 \cdot d6 + 4.93(\pm 0.75)d7 \cdot d14 + 0.0065(\pm 0.0019)d8 \cdot d11 - 1.11(\pm 0.20)d11 \cdot d15$$

$R = 0.9765$, $S = 0.189$, $R_{cv} = 0.9581$, $S_{cv} = 0.251$, $F = 80$
leave-20%-out cross-correlation, 20 runs: av$R(20\%)_{cv} = 0.952$, av$S(20\%)_{cv} = 0.267$
random correlation:[a] max$R = 0.791$ ($R_{cv} = 0.620$), min$R = 0.105$, $p(R^2 \geq 0.5) = 0.36\%$[c]

Model 5

$$p K = 7.80(\pm 0.62) - 9.9(\pm 1.5)d2 \cdot d14 + 2.97(\pm 0.24)d1 \cdot d13 + 0.69(\pm 0.16)d1 \cdot d16 + 0.7810(\pm 0.0601)d4 \cdot d7 - 2.5361(\pm 0.1911)d5 \cdot d6 + 8.0(\pm 1.2)d6 \cdot d14 + 0.00023(\pm 0.00007)d10 \cdot d11 - 1.00(\pm 0.18)d11 \cdot d15$$

$R = 0.9764$, $S = 0.189$, $R_{cv} = 0.9553$, $S_{cv} = 0.259$, $F = 79$
leave-20%-out cross-correlation, 20 runs: av$R(20\%)_{cv} = 0.955$, av$S(20\%)_{cv} = 0.257$
random correlation:[a] max$R = 0.723$ ($R_{cv} = 0.511$), min$R = 0.153$, $p(R^2 \geq 0.5) = 0.2\%$[c]

Model 6

$$p K = 7.45(\pm 0.57) - 6.15(\pm 0.93)d2 \cdot d14 + 0.92(\pm 0.27)d3 \cdot d12 + 3.18(\pm 0.21)d1 \cdot d13 + 0.54(\pm 0.17)d1 \cdot d16 + 0.694(\pm 0.067)d4 \cdot d7 - 2.27(\pm 0.20)d5 \cdot d6 + 4.98(\pm 0.75)d7 \cdot d14 - 1.09(\pm 0.19)d11 \cdot d15$$

$R = 0.9763$, $S = 0.189$, $R_{cv} = 0.9588$, $S_{cv} = 0.249$, $F = 79$
leave-20%-out cross-correlation, 20 runs: av$R(20\%)_{cv} = 0.951$, av$S(20\%)_{cv} = 0.268$
random correlation:[a] max$R = 0.819$ ($R_{cv} = 0.690$), min$R = 0.115$, $p(R^2 \geq 0.5) = 0.2\%$

Model 7

$$p K = 7.40(\pm 0.57) - 6.35(\pm 0.93)d2 \cdot d14 + 3.16(\pm 0.22)d1 \cdot d13 + 0.57(\pm 0.16)d1 \cdot d16 + 0.699(\pm 0.066)d4 \cdot d7 - 2.28(\pm 0.20)d5 \cdot d6 + 0.303(\pm 0.090)d6 \cdot d11 + 5.15(\pm 0.75)d7 \cdot d14 - 1.02(\pm 0.18)d11 \cdot d15$$

$R = 0.9762$, $S = 0.190$, $R_{cv} = 0.9587$, $S_{cv} = 0.249$, $F = 78$
leave-20%-out cross-correlation, 20 runs: av$R(20\%)_{cv} = 0.952$, av$S(20\%)_{cv} = 0.266$
random correlation:[a] max$R = 0.744$ ($R_{cv} = 0.567$), min$R = 0.172$, $p(R^2 \geq 0.5) = 0.05\%$[c]

Model 8

$$p K = 7.18(\pm 0.55) - 6.10(\pm 0.93)d2 \cdot d14 + 3.09(\pm 0.23)d1 \cdot d13 + 0.59(\pm 0.16)d1 \cdot d16 + 0.701(\pm 0.066)d4 \cdot d7 - 2.27(\pm 0.20)d5 \cdot d6 + 4.93(\pm 0.75)d7 \cdot d14 + 0.0056(\pm 0.0017)d9 \cdot d11 - 1.11(\pm 0.20)d11 \cdot d15$$

$R = 0.9761$, $S = 0.190$, $R_{cv} = 0.9579$, $S_{cv} = 0.251$, $F = 78$
leave-20%-out cross-correlation, 20 runs: av$R(20\%)_{cv} = 0.953$, av$S(20\%)_{cv} = 0.263$
random correlation:[a] max$R = 0.738$ ($R_{cv} = 0.528$), min$R = 0.139$, $p(R^2 \geq 0.5) = 0.3\%$[c]

	descriptor	explanation
$d1$	E_{HOMO}	energy of the highest occupied molecular orbital
$d2$	$Q_{2,4}$	sum of charges on carbon atoms C-2 and C-4
$d3$	Q_4	charge on carbon atom C-4
$d4$	Q_{fl}	sum of charges in the flavylium core
$d5$	Q_{ph}	sum of charges in the phenyl moiety
$d6$	S_2^N	superdelocalizability for the nucleophilic attack at position C-2
$d7$	S_4^N	superdelocalizability for the nucleophilic attack at position C-4
$d8$	H	Harary index
$d9$	H'	variant of the Harary index
$d10$	MTI	molecular topological index
$d11$	I	indicator variable: value of 1 for 7-OH or 4′-OH salts (without 3-R); elsewhere 0
$d12$	I_3	indicator variable: value of 1 for salts with a substituent at position C-3; elsewhere 0
$d13$	I_4	indicator variable: value of 1 for salts with a substituent at position C-4; elsewhere 0
$d14$	I_5	indicator variable: value of 1 for salts with a substituent at position C-5; elsewhere 0
$d15$	I_7	indicator variable: value of 1 for salts with a substituent at position C-7; elsewhere 0
$d16$	$I_{4'}$	indicator variable: value of 1 for salts with a substituent at position C-4′; elsewhere 0

[a] For each model 2000 runs with randomized pK values were performed. An exception is model 4 for which 5000 runs were performed. max$R$ and min$R$ are maximum and minimum chance correlation coefficient values, respectively. [b] $p(R^2 \geq 0.5)$ is the probability of encountering a chance correlation at the level $R^2 \geq 0.5$. [c] For this model $p(R^2 \geq 0.6) = 0\%$, means the probability of encountering a chance correlation at the level $R^2 \geq 0.6$ is zero.

Flavylium Salt Molecular Structure

J. Chem. Inf. Comput. Sci., Vol. 39, No. 6, 1999 **971**
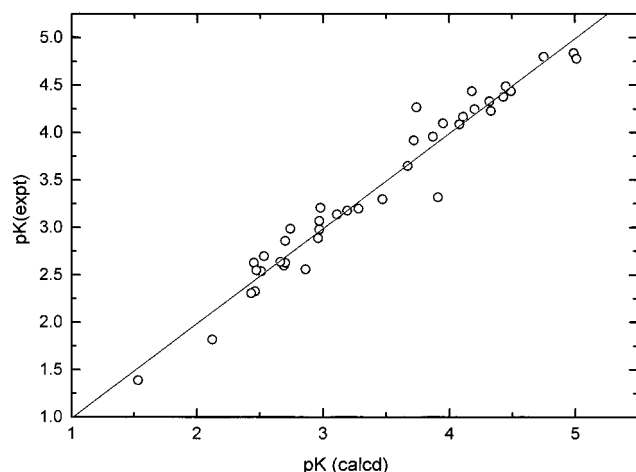
Figure 2. A comparison of computed (using model 1 from Table 3) and experimental pK values of 40 flavylium salts.

Figure 3. A comparison of computed (using model 1 from Table 4) and experimental half-lives of 10 flavylium salts.

Table 4. The Best Possible Nonlinear Two-Descriptor Multiregression Models for Calculating Half-Life

$N = 10$; the total number of descriptors screened, 779

Model 1
half-life = $-2737.3(\pm320.7) - 4480.3(\pm517.4)d1{\cdot}d3 + 129.2(\pm5.9)d3{\cdot}d6$
$R = 0.9950$, $S = 17.47$, $R_{cv} = 0.9887$, $S_{cv} = 26.28$, $F = 346$
leave-20%-out cross-correlation, 20 runs: av$R(20\%)_{cv} = 0.9848$, av$S(20\%)_{cv} = 29.37$
random correlation:[a] max$R = 0.950$ ($R_{cv} = 0.902$), min$R = 0.032$, $p(R^2 \geq 0.8) = 0.8\%$[b]

Model 2
half-life = $-2736.6(\pm345.8) - 4479.3(\pm558.0)d1{\cdot}d3 + 134.2(\pm6.6)d2{\cdot}d6$
$R = 0.9942$, $S = 18.84$, $R_{cv} = 0.9868$, $S_{cv} = 28.44$, $F = 297$
leave-20%-out cross-correlation, 20 runs: av$R(20\%)_{cv} = 0.9816$, av$S(20\%)_{cv} = 32.46$
random correlation:[a] max$R = 0.952$ ($R_{cv} = 0.905$), min$R = 0.101$, $p(R^2 \geq 0.8) = 0.6\%$[b]

Model 3
half-life = $-2400.6(\pm420.2) + 307.0(\pm18.3)d6 - 3936.8(\pm678.5)d1{\cdot}d3$
$R = 0.9916$, $S = 22.61$, $R_{cv} = 0.9763$, $S_{cv} = 38.43$, $F = 205$
leave-20%-out cross-correlation, 20 runs: av$R(20\%)_{cv} = 0.9675$, av$S(20\%)_{cv} = 43.12$
random correlation:[a] max$R = 0.955$ ($R_{cv} = 0.906$), min$R = 0.101$, $p(R^2 \geq 0.8) = 0.6\%$[b]

Model 4
half-life = $-2663.8(\pm572.4) + 448.3(\pm95.0)d4{\cdot}d5 + 303.6(\pm21.8)d6$
$R = 0.9883$, $S = 26.65$, $R_{cv} = 0.9693$, $S_{cv} = 43.77$, $F = 147$
leave-20%-out cross-correlation, 20 runs: av$R(20\%)_{cv} = 0.9604$, av$S(20\%)_{cv} = 47.54$
random correlation:[a] max$R = 0.955$ ($R_{cv} = 0.891$), min$R = 0.102$, $p(R^2 \geq 0.8) = 0.65\%$[b]

descriptor		explanation
$d1$	E_{LUMO}	energy of the lowest unoccupied molecular orbital
$d2$	S_2^N	superdelocalizability for the nucleophilic attack at position C-2
$d3$	S_4^N	superdelocalizability for the nucleophilic attack at position C-4
$d4$	Q_3	charge on carbon atom C-3
$d5$	Q_{ph}	sum of charges in the phenyl moiety
$d6$	I	indicator variable: value of 1 for salts with an ionizable OH group (without 3-R); elsewhere 0

[a] For each model 2000 runs with randomized half-life values were performed; max$R$ and min$R$ are maximum an minimum chance correlation coefficient values, respectively. [b] $p(R^2 \geq 0.8)$, probability of encountering a chance correlation at the level $R^2 \geq 0.8$.

Half-Life of Flavylium Salts. Coloration of slightly acidic and neutral solutions of flavylium salts is the prerequisite for considering flavylium salts as food colors. A high pK value implies coloration of acidic solutions (predominantly by the AH$^+$ form) as well as nearly neutral solutions (predominantly by the A form). Flavylium salt containing media, however, are susceptible to color deterioration during processing and storage caused by light, heat, and additional factors such as ascorbic acid/O$_2$, and so on.[4] Despite the fact that the stability of anthocyanin-type colorants in model systems was reported by many researchers,[8,11,35] evaluation of QSAR is a very difficult task. Namely, in all of these reports the numbers of studied compounds and data points are very limited and not suitable for QSAR. To the best of our knowledge, only the report by Iacobucci and Sweeny[36] can be used for QSPR modeling.

Measurement of flavylium salt color stability can be carried out by following spectrophotometrically the changes of absorbance at the λ_{max} of buffered solutions. The resistance of the color to fading is a measure of its stability. Such defined stability reflects the overall changes illustrated in Figure 1. Using experimental values taken from ref 36, we have developed QSPR models that describe the variation of the half-life (50% reduction in absorbance at λ_{max}). A total of 819 descriptors were generated for each flavylium salt (39 initial descriptors and their cross-products). The best top

models containing two descriptors were selected and are presented in Table 4. The values of leave-20%-out cross-validated, as well as of the chance correlation parameters, indicate the model's stability. Used descriptors are related to the ability of the cationic form AH$^+$ to undergo hydration reactions which result in discoloration of flavylium salt solutions.
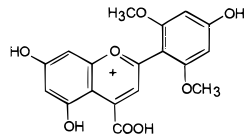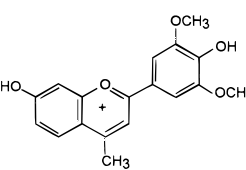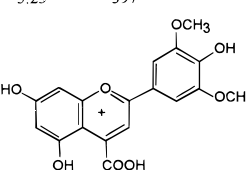
Experimental values and calculated values of half-life using model 1 from Table 4 are presented in Table 5. In Figure 3 we plot the computed half-life from model 1 and the experimental half-life for the studied flavylium salts.

Prediction of Favorable Flavylium Salts. Using developed models for evaluating pK and half-life of flavylium salts, we are able to perform calculations for many postulated structures. Postulated structures are structures in which the flavylium chromophoric system bears various numbers of

972 *J. Chem. Inf. Comput. Sci., Vol. 39, No. 6, 1999*

AMIĆ ET AL.

Table 5. Effect of Structure on the Half-Life of the Cationic Form of Flavylium Salts

flavylium salt	half-life (days) exptl[a]	half-life (days) calcd[b]
3-OH-5,7,4′,5′-tetraOCH$_3$	0.04	1.69
3,5,7,4′,5′-pentaOH	0.5	0.00
3,5,7,4′,5′-pentaOCH$_3$	6	0.33
5,7,4′,5′-tetraOCH$_3$	170	162
5,7,4′-triOH	400	416
7,4′-diOH	400	381
4′-OH	400	374
4′-OCH$_3$	35	56
7-OH	300	329
7-OCH$_3$	8	16

[a] Experimental data of the half-life were taken from ref 36. [b] Calculated using model 1 from Table 4.

Table 6. Proposed Stable Flavylium Model Compounds: pK Values, Half-Life (days), Absorption Maxima in UV−Vis Spectra λ_{max} (nm), and Corresponding Oscillator Strengths (f)

pK value[a]	Half-life[b] (days)	Electronic spectra[c]					
		AH$^+$ form λ_{max} nm	f	A-7 form λ_{max} nm	f	A-4′ form λ_{max} nm	f
4.94	488	493.8	1.032	533.1	1.260	523.5	1.634
		415.3	0.177	508.6	0.074	441.8	0.014
		343.2	0.107	356.3	0.124	387.0	0.069
4.29	420	452.2	1.214	514.7	1.319	509.4	1.895
		353.7	0.038	373.9	0.091	358.5	0.017
		322.4	0.089	338.3	0.057	330.1	0.046
5.23	397	497.8	1.031	530.5	1.228	545.0	1.820
		422.7	0.231	513.3	0.133	391.8	0.062
		354.8	0.116	376.4	0.113	366.4	0.035

[a] Mean value using eight models from Table 3. [b] Mean value using four models from Table 4. [c] Calculated by the PPP method.

OH and/or OCH$_3$ substituents in various positions, with position C-4 exclusively occupied by an H, CH$_3$, Ph, or COOH substituent. Among postulated structures we chose structures with a high pK value and long half-life. For these structures we have calculated electronic absorption spectra for the cationic form AH$^+$ and anhydrobase forms A-7 and A-4′. Finally, we selected favorable structures of flavylium salts, i.e., compounds with a high pK value, long half-life, and intense orange-red coloration. Table 6 presents the representatives of favorable model compounds. pK and half-life values for these compounds were calculated as an average of the values predicted by the top models from Tables 3 and 4, respectively.

The favorable structures shown in Table 6 are expected to exhibit stable orange-red coloration in slightly acidic and neutral solutions. These qualities make them possible candidates as food colorants.

CONCLUSION

In this paper, the pK value, half-life, and electronic spectra of many postulated flavylium salts were calculated. Proposed model flavylium salt compounds should give rise to stable solutions that are colored, regardless of the pH. Future synthetic efforts directed toward the modification of natural flavylium salts with the goal of improving their stability in slightly acidic and neutral solutions are expected to be guided by QSAR/QSPR models. It remains to be seen whether there will be application of proposed flavylium salts as food colorants.

REFERENCES AND NOTES

(1) Harborne, J. B. *Comparative Biochemistry of the Flavonoids*; Academic Press: London and New York, 1967.

(2) Timberlake, C. F.; Bridle, P. Anthocyanins. In *Developments in Food Colours-1*; Walford, J., Ed.; Applied Science Publishers Ltd.: London, 1980; pp 115−150.

(3) (a) Harborne, J. B.; Grayer, R. J. The Anthocyanins. In *The Flavonoids: Advances in Research Since 1980*; Harborne, J. B., Ed.; Chapman and Hall: London, 1988; pp 1−20. (b) Tamura, H.; Yamagami, A. Antioxidative Activity of Monoacylated Anthocyanins Isolated from Muscant Bailey A Grape. *J. Agric. Food Chem.* **1994**, *42*, 1612−1615. (c) Costantino, L.; Rastelli, G.; Albasini, A. Anthocyanins as Inhibitors of Xanthine Oxidase. *Pharmazie* **1995**, *50*, 573−574.

(4) (a) Markakis, P. Anthocyanins as Food Additives. In *Anthocyanins as Food Colors*; Markakis, P., Ed.; Academic Press: New York, 1982; pp 245−253. (b) Mazza, G.; Brouillard, R. Recent Developments in the Stabilization of Anthocyanins in Food Products. *Food Chem.* **1987**, *25*, 207−225. (c) Timberlake, C. F.; Henry, B. S. Anthocyanins as Natural Food Colorants. In *Plant Flavonoids in Biology and Medicine II: Biochemical, Cellular, and Medicinal Properties*; Cody, V., Middleton, E., Jr., Harborne, J. B., Beretz, A., Eds.; Alan R. Liss: New York; 1988; pp 107−121. (d) Francis, F. J. Food Colorants: Anthocyanins. *Crit. Rev. Food Sci. Nutr.* **1989**, *28*, 273−314.

(5) Brouillard, R. Chemical Structure of Anthocyanins. In *Anthocyanins as Food Colors*; Markakis, P., Ed.; Academic Press: New York, 1982; pp 1−40.

(6) (a) Brouillard, R.; Dubois, J.-E. Mechanism of the Structural Transformations of Anthocyanins in Acidic Media. *J. Am. Chem. Soc.* **1977**, *99*, 1359−1364. (b) Brouillard, R.; Delaporte, B. Chemistry of Anthocyanin Pigments. 2. Kinetic and Thermodynamic Study of Proton Transfer, Hydration, and Tautomeric Reactions of Malvidin 3-Glucoside. *J. Am. Chem. Soc.* **1977**, *99*, 8461−8468. (c) Brouillard, R.; Delaporte, B.; Dubois, J.-E. Chemistry of Anthocyanin Pigments. 3. Relaxation Amplitudes in pH-Jump Experiments. *J. Am. Chem. Soc.* **1978**, *100*, 6202−6205.

(7) (a) Jurd, L. Some Benzopyrylium Compounds Potentially Useful as Color Additives for Fruit Drinks and Juices. *Food Technol.* **1964**, *18*, 157−159. (b) Jurd, L. Methods of Using Flavylium Compounds for Food Coloring. U.S. Patent No. 3,266,903, 1966.

(8) Ohta, H.; Akuta, S.; Osajima; Y. Stability of Anthocyanin Pigments and Related Compounds in Acidic Solutions. *Nippon Shokuhin Kogyo Gakkaishi* **1980**, *27*, 81−85.

(9) (a) Timberlake, C. F.; Bridle, P. Flavylium Salts Resistant to Sulphur Dioxide. *Chem. Ind. (London)* **1968**, 1489. (b) Timberlake, C. F. Improvements Relating to the Colouring of Foodstuffs. British Patent No. 5477/68, Feb 2, 1968.

Flavylium Salt Molecular Structure

J. Chem. Inf. Comput. Sci., Vol. 39, No. 6, 1999 **973**

(10) Sweeny, J. G.; Iacobucci, G. A.; Brusick, D.; Jagannath, D. R. Structure−Activity Relationships in the Mutagenicity of Quinone Methides of 7-Hydroxyflavylium Salts for *Salmonella typhimurium*. *Mutat. Res.* **1981**, *82*, 275−283.

(11) Sweeny, J. G.; Iacobucci, G. A. Effect of Substitution on the Stability of 3-Deoxyanthocyanidins in Aqueous Solutions. *J. Agric. Food Chem.* **1983**, *31*, 531−533.

(12) Bridle, P.; Timberlake, C. F. Anthocyanins as Natural Food Colours− Selected Aspects. *Food Chem.* **1997**, *58*, 103−109.

(13) Zhou, Z.; Parr, R. G. New Measures of Aromaticity: Absolute Hardness and Relative Hardness. *J. Am. Chem. Soc.* **1989**, *111*, 7371−7379.

(14) Fukui, K. *Theory of Orientation and Stereoselection*, Springer: Berlin, 1975; p 34.

(15) (a) Wiener, H. Prediction of Isomeric Differences in Paraffin Properties. *J. Am. Chem. Soc.* **1947**, *69*, 17−20. (b) Schultz, H. P. Topological Organic Chemistry. 1. Graph Theory and Topological Indices of Alkanes. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 227−228. (c) Mihalić, Z.; Nikolić, S.; Trinajstić, N. Comparative Study of the Distance Matrix. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 28−37. (d) 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. (e) Nikolić, S.; Trinajstić, N.; Mihalić, Z. Molecular Topological Index: An Extension to Heterosystems. *J. Math. Chem.* **1993**, *12*, 251− 264.

(16) Lučić, B.; Trinajstić, N. Multivariate Regression Outperforms Several Robust Architectures of Neural Networks in QSAR Modeling. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 121−132.

(17) Lučić, B.; Trinajstić, N.; Sild, S.; Karelson, M.; Katritzky, A. R. A New Efficient Approach for Variable Selection Based on Multiregression: Prediction of Gas Chromatographic Retention Times and Response Factors. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 610−621.

(18) Herman, A.; Blonski, S. Extended Pariser−Parr−Pople-Sandorfy C Method. Software package SAA-012-S88. *Spectrochim. Acta* **1988**, *44A*, Software Survey Section.

(19) Davidović-Amić, D.; Amić, D.; Trinajstić, N. 4-Carboxyflavylium Salts: Stable Red Dyes? *Croat. Chem. Acta* **1994**, *67*, 163−170.

(20) Ilavsky, D.; Kuthan, J. Electronic Spectra of 2- and 4-Substituted 3,5-Dicyano-6-Methylpyridines. *Collect. Czech. Chem. Commun.* **1981**, *46*, 2687−2695.

(21) Kurtin, W. E.; Song, P.-S. Electronic Structures and Spectra of Some Natural Products of Theoretical Interest-I. Molecular Orbital Studies of Anthocyanidins. *Tetrahedron* **1968**, *24*, 2255−2267.

(22) Glier, Chr.; Dietz, F.; Scholz, M.; Fischer, G. MO−Berechnungen an Heterocyclen-VII. Tautomerieprobleme an substituierten 1,3,8-Triazaindolizinen. *Tetrahedron* **1972**, *28*, 5789−5796.

(23) Gey, E.; Tino, J. Zur quantenchemischen Berechnung der Spindichteverteilung in parasubstituierten Benzylradikalen in PPP- und INDO-UHF-Näherung mit und ohne Spinprojektion. *Croat. Chem. Acta* **1978**, *51*, 11−17.

(24) Mataga, N.; Nishimoto, K. Electronic Structure and Spectra of Nitrogen Heterocycles. *Z. Phys. Chem. (Neue Folge)* **1957**, *13*, 140−157.

(25) Abe, K.; Sakaino, Y.; Kakinuma, J.-i.; Kakisawa, H. Structures of Anthocyanin Anhydro Bases. *Nippon Kagaku Kaishi* **1977**, 1197− 1204.

(26) Sperling, W.; Werner, F. C.; Kuhn, H. Vorgänge in Lösungen von Anthocyanidinen bei pH-Änderung und bei Belichtung. *Ber. Bunsenges. Phys. Chem.* **1966**, *70*, 530−544.

(27) Baranac, J. M. Spectrochemical Investigation of Flavylium Chromophoric System and its Transformations in Different Media. Ph.D. Thesis, University of Belgrade, 1973.

(28) Baranac, J.; Amić, D.; Vukadinović, V. Spectrophotometric Study of the Influence of Individual Substituted Positions on Flavylium Chromophore Stability. *J. Agric. Food Chem.* **1990**, *38*, 932−936.

(29) Ohta, H.; Watanabe, H.; Yamada, K.; Akuta, S.; Osajima; Y. Preparations of Flavylium Compounds and their Spectral Properties. *Nippon Shokuhin Kogyo Gakkaishi* **1980**, *27*, 75−80.

(30) Baranac, J. M.; Amić, D. S. Structural Transformations of Apigeninidin-Type Flavylium Salts. *J. Agric. Food Chem.* **1990**, *38*, 2111− 2115.

(31) Brouillard, R.; Iacobucci, G. A.; Sweeny, J. G. Chemistry of Anthocyanin Pigments. 9. UV−Visible Spectrophotometric Determination of the Acidity Constants of Apigeninidin and Three Related 3-Deoxyflavylium Salts. *J. Am. Chem. Soc.* **1982**, *104*, 7585−7590.

(32) Baranac, J. M.; Amić, D. S. A Spectroscopic Investigation of the Influence of Substituents in the Flavylium Chromophore A-Ring. II. *J. Serb. Chem. Soc.* **1989**, *54*, 259−264.

(33) Baranac, J. M.; Amić, D. S. Spectrochemical Investigations of the Influence of pH on the Mechanism of Structural Transformations of the 4-Substituted Flavylium Chromophore. *J. Serb. Chem. Soc.* **1985**, *50*, 299−305.

(34) Mazza, G.; Brouillard, R. Color Stability and Structural Transformations of Cyanidin 3,5-Diglucoside and Four 3-Deoxyanthocyanins in Aqueous Solutions. *J. Agric. Food Chem.* **1987**, *35*, 422−426.

(35) See, for example: (a) Palamidis, N.; Markakis, P. Stability of Grape Anthocyanin in a Carbonated Beverage. *J. Food Sci.* **1975**, *40*, 1047− 1049. (b) Shi, Z.; Lin, M.; Francis, F. J. Stability of Anthocyanins from *Tradescania pallida. J. Food Sci.* **1992**, *57*, 758−760. (c) Dao, L. T.; Takeoka, G. R.; Edwards, R. H.; Berrios, J. D. J. Improved Method for the Stabilization of Anthocyanidins. *J. Agric. Food Chem.* **1998**, *46*, 3564−3569.

(36) Iacobucci, G. A.; Sweeny, J. G. The Chemistry of Anthocyanins, Anthocyanidins and Related Flavylium Salts. *Tetrahedron* **1983**, *39*, 3005−3038.