

# On electronegativity and chemical hardness relationships with aromaticity

Laszlo Tarko · Mihai V. Putz

Received: 11 June 2009 / Accepted: 31 July 2009 / Published online: 25 August 2009  
© Springer Science+Business Media, LLC 2009

**Abstract** Quantitative structure aromaticity relationship studies on various aromatic, non-aromatic and anti-aromatic molecules are presented aiming for checking the predictor quality of finite-difference electronegativity and chemical hardness descriptors in aromaticity computations. The results show that the “aromaticity of peripheral topological path” may be well described by superior finite difference schemes of electronegativity and chemical hardness indices in certain calibrating conditions.

**Keywords** QSArR · TOPAZ algorithm · Finite difference schemes · PRECLAV software

## 1 Introduction

Quantitative structure-aromaticity relationships (QSArR) studies are not usual in QSPR practice, because of lack of physical–chemical quantitative data regarding to aromaticity, because many aromaticity descriptors are used when the dependent property is not aromaticity [1–8] and because there are specific non-statistical algorithms used in aromaticity computations [5,9]. The recent progress in chemical reactivity modeling by the electronegativity and chemical hardness principles [10,11] highly

---

L. Tarko  
Center of Organic Chemistry “C. D. Nenitescu” of Romanian Academy, 202B Spl. Independenței,  
PO Box 35-108, 060023 Bucharest - 6th Sector, Romania  
e-mail: ltarko@cco.ro

M. V. Putz (✉)  
Laboratory of Computational and Structural Physical Chemistry, Chemistry Department,  
West University of Timișoara, Str. Pestalozzi No. 16, 300115 Timișoara, Romania  
e-mail: mvputz@cbg.uvt.ro

motivates such type of analysis to be undertaken. If successfully, it may provide both conceptual new insight on the aromaticity concept as well as generalization of the chemical reactivity theory and principles grounded on structural indices as electronegativity and chemical hardness that have been intensively considered in the last decades within density functional theory and quantum chemistry [12, 13]. With this intend the present work steps aside by reporting few QSArR relevant checks on a typical set of organic compounds with the aid of complex finite difference schemes for electronegativity and chemical hardness computations among the large classes of descriptors available with dedicated QSPR software.

## 2 Method

The calibration set includes the molecules in Table 1.

**Table 1** Aromaticity values for common organic molecules by means of the topological paths and aromatic zones (TOPAZ) methodology [5], along their QSArR relevant electronegativity and chemical hardness values (in electron-volts eV), computed upon the Eqs. (3) and (4) and the schemes of Table 2

No.	Molecule	$A_{TOPAZ}$	$A^a$	$\chi_{4C}$	$\eta_{10P}$	$\chi_{SLR}$	$\eta_{SLR}$
1	Benzene	999.2	999.6	4.57	4.11	2.73	5.67
2	Naphthalene	616.2	808.1	4.47	3.29	2.62	4.19
3	Anthracene	585.0	792.5	4.46	2.88	2.61	3.78
4	Phenanthrene	520.3	760.15	4.51	3.18	2.66	3.96
5	Pyrene	561.2	780.6	4.51	3.18	2.66	3.96
6	Naphthacene	579.4	789.7	4.50	2.85	2.65	3.69
7	Benz[ <i>a</i> ]anthracene	522.4	761.2	4.46	2.61	2.53	3.53
8	Chrysene	468.8	734.4	4.51	3.00	2.67	3.78
9	Triphenylene	474.9	737.45	4.55	3.22	2.68	4.13
10	Perylene	556.3	778.15	4.48	2.70	2.61	3.68
11	Benzo[ <i>e</i> ]pyrene	501.6	750.8	4.53	2.87	2.67	3.65
12	Benzo[ <i>a</i> ]pyrene	517.9	758.95	4.50	2.69	2.65	3.51
13	Pentacene	578.7	789.35	4.50	2.40	2.65	3.25
14	Benzo[ <i>a</i> ]naphthacene	531.4	765.7	4.51	2.64	2.66	3.46
15	Dibenz[ <i>a, h</i> ]anthracene	482.0	741.0	4.52	2.90	2.68	3.63
16	Benzo[ <i>b</i> ]chrysene	474.7	737.35	4.51	2.78	2.67	3.58
17	Picene	444.8	722.4	4.53	2.96	2.68	3.67
18	Benzo[ <i>ghi</i> ]perylene	502.0	751.0	4.54	2.72	2.68	3.49
19	Anthanthrene	572.1	786.05	4.51	2.50	2.65	3.35
20	Naphtho[2,1,8- <i>gra</i> ]naphthacene	519.0	759.5	4.51	2.54	2.66	3.34
21	Benzo[ <i>a</i> ]perylene	531.6	765.8	4.51	2.50	2.71	3.46
22	Benzo[ <i>b</i> ]perylene	514.0	757.0	4.51	2.65	2.65	3.50
23	Coronene	471.8	735.9	4.69	2.83	2.94	3.63
24	Zethrene	581.9	790.95	4.48	2.41	2.63	3.35
25	Benzo[ <i>a</i> ]pentacene	476.7	738.35	4.50	2.44	2.61	3.30

**Table 1** continued

No.	Molecule	$A_{\text{TOPAZ}}$	$A^a$	$\chi_{4C}$	$\eta_{10P}$	$\chi_{\text{SLR}}$	$\eta_{\text{SLR}}$
26	Dibenzo[ <i>b, k</i> ]chrysene	476.2	738.1	4.53	2.64	2.74	3.43
27	Naphtho[2,3- <i>g</i> ]chrysene	451.4	725.7	4.56	2.79	2.72	3.53
28	Naphtho[8,1,2- <i>bcd</i> ]perylene	526.2	763.1	4.52	2.50	2.66	3.35
29	Dibenzo[ <i>cd, lm</i> ]perylene	502.8	751.4	4.52	2.44	2.66	3.26
30	Dibenzo[ <i>a, f</i> ]perylene	533.3	766.65	4.51	2.28	2.66	3.22
31	Phenanthro[1,10,9,8- <i>opqra</i> ]perylene	565.1	782.55	4.52	2.22	2.66	3.15
32	Dibenzo[ <i>de, op</i> ]pentacene	602.0	801.0	4.49	2.20	2.64	3.04
33	Dibenzo[ <i>a, l</i> ]pentacene	511.5	755.75	4.53	2.48	2.70	3.29
34	Benzo[2,1- <i>a:3,4, -a'</i> ]dianthracene	466.0	733.0	4.48	2.76	2.64	3.54
35	Naphtho[2,1,8- <i>yzal</i> ]hexacene	532.5	766.25	4.54	2.26	2.69	2.98
36	Cyclobutadiene	-993.5 <sup>b</sup>	3.25	4.15	3.53	2.33	5.00
37	Cyclopentadiene	0.000 <sup>c</sup>	500.0	4.29	3.81	2.55	5.15
38	Pyrrol	539.2	769.6	3.64	3.92	2.00	5.04
39	Furan	481.7	740.85	4.27	3.95	2.38	5.17
40	Thiophene	571.3	785.65	4.52	3.68	2.67	4.66
41	Pyrazole	509.4	754.7	4.05	4.00	2.38	5.25
42	Pyridine	831.4	915.7	4.85	3.91	2.92	4.92
43	Pyrimidine	717.2	858.6	5.33	4.03	3.13	5.11
44	Pyrazine	790.9	895.45	5.30	3.83	3.24	4.78
45	Pyridazine	820.3	910.15	5.70	4.12	3.89	5.42
46	Cycloheptatriene	0.000 <sup>c</sup>	500	4.27	3.53	2.50	4.67
47	Pentalene	-977.1	11.45	4.87	2.98	3.10	4.17
48	1,4-dihydro-pyrrolo[3,2- <i>b</i> ] pyrrol	313.0	656.5	3.50	3.46	2.13	4.47
49	4H-Furo[3,2- <i>b</i> ]pyrrol	288.9	644.45	3.94	3.47	2.36	4.54
50	Benzocyclobutadiene	-492.2 <sup>b</sup>	253.9	4.34	3.25	2.47	4.33
51	Indole	458.6	729.3	4.07	3.39	2.42	4.28
52	Azulene	441.7	720.85	4.40	2.80	2.55	3.65
53	Quinazoline	557.0	778.5	5.10	3.40	3.03	4.30

<sup>a</sup> Computed upon Eq. (2)<sup>b</sup> Anti-aromatic peripheral topological path<sup>c</sup> Non-aromaticity of peripheral topological path because of the discontinuous conjugation

In poly(hetero)cyclic compounds the aromaticity of different chemical cycles can be very different, for instance in phenanthrene, anthracene, azulene or indole. Therefore, to avoid any confusion in computations presented here, the dependent property was “aromaticity of the peripheral topological path”. The value of dependent property was computed using Topological Paths and Aromatic Zones (TOPAZ) algorithm [5]. This algorithm uses the bond orders value of chemical bonds, computed using the semi-empirical quantum method PM6 [14]. According to TOPAZ formulas, the aromaticity of anti-aromatics is within  $[-1000, -350]$  range, the aromaticity of non-aromatics is within  $(-350, 0]$  range and the aromaticity of aromatics is within  $(0, 1000]$  range. The

translation of computed aromaticity values within  $[-1, 1]$  or  $[-1000, 1000]$  range is usual. However, in QSPR computations, usage of negative or close to zero values can be difficult. The coefficient of variation CV, computed according to formula (1) for descriptors and dependent property, is meaningless if the denominator is negative or close to zero,

$$CV = 100 \frac{\sigma}{V_m}, \quad (1)$$

where  $\sigma$  is standard deviation and  $V_m$  is average value. Consequently, we use the normalized values of aromaticity, according to formula (2), placed within  $[0, 1000]$  range.

$$A = \frac{A_{\text{TOPAZ}} + 1000}{2} \quad (2)$$

The normalized aromaticity value of peripheral topological paths is presented in the fourth column of Table 1. The minimum energy geometry, for each molecule in Table 1, was obtained by the PCModel v. 9.0 software [15], using MMX force field [16]. Then the quantum mechanics program MOPAC v. 8.137W [17], using PM6 method [14], optimized the geometry more rigorously. In MOPAC analysis we used the keyword string “pm6 pulay gnorm = 0.2 shift = 50 geo-ok camp-king bonds vectors”.

In the next step, the programs MOPAC, DRAGON v. 5.4 [18] and PRECLAV v. 0809 [19–22] computed, for each molecule, the value for almost 1700 molecular descriptors. All these software related descriptors were challenged against the various compact finite difference (CFD) formulas for electronegativity and chemical hardness indices,

$$\begin{aligned} \chi_{\text{CFD}} = & - \left[ a_1 (1 - \alpha_1) + \frac{1}{2} b_1 + \frac{1}{3} c_1 \right] \frac{E_{\text{HOMO}(1)} + E_{\text{LUMO}(1)}}{2} \\ & - \left[ b_1 + \frac{2}{3} c_1 - 2a_1 (\alpha_1 + \beta_1) \right] \frac{E_{\text{HOMO}(2)} + E_{\text{LUMO}(2)}}{4} \\ & - \left[ c_1 - 3a_1 \beta_1 \right] \frac{E_{\text{HOMO}(3)} + E_{\text{LUMO}(3)}}{6}, \end{aligned} \quad (3)$$

$$\begin{aligned} \eta_{\text{CFD}} = & \left[ a_2 (1 - \alpha_2 + 2\beta_2) + \frac{1}{4} b_2 + \frac{1}{9} c_2 \right] \frac{E_{\text{LUMO}(1)} - E_{\text{HOMO}(1)}}{2} \\ & + \left[ \frac{1}{2} b_2 + \frac{2}{9} c_2 + 2a_2 (\beta_2 - \alpha_2) \right] \frac{E_{\text{LUMO}(2)} - E_{\text{HOMO}(2)}}{4} \\ & + \left[ \frac{1}{3} c_2 - 3a_2 \beta_2 \right] \frac{E_{\text{LUMO}(3)} - E_{\text{HOMO}(3)}}{6}, \end{aligned} \quad (4)$$

involving the first three HOMO and LUMO ionization and affinity energy levels [23], respectively, in Koopmans or frozen core approximation [24], whose parameters are given in the Table 2 for each scheme of computation [25].

The statistical computations were conducted using the specific formulas and procedures of PRECLAV program. We obtain, in each QSPR study, tens thousand QSPR type (5) multilinear equations:

**Table 2** Numerical parameters for the compact finite second (2C)-, fourth (4C)- and sixth (6C)-order central differences; standard Padé (SP) schemes; sixth (6T)- and eighth (8T)-order tridiagonal schemes; eighth (8P)- and tenth (10P)-order pentadiagonal schemes up to spectral-like resolution (SLR) schemes for the electronegativity and chemical hardness Eqs. (3) and (4) [23,25]

Scheme	Electronegativity					Chemical hardness				
	$a_1$	$b_1$	$c_1$	$\alpha_1$	$\beta_1$	$a_2$	$b_2$	$c_2$	$\alpha_2$	$\beta_2$
2C	1	0	0	0	0	1	0	0	0	0
4C	$\frac{4}{3}$	$-\frac{1}{3}$	0	0	0	$\frac{4}{3}$	$-\frac{1}{3}$	0	0	0
6C	$\frac{3}{2}$	$-\frac{3}{5}$	$\frac{1}{10}$	0	0	$\frac{12}{11}$	$\frac{3}{11}$	0	$\frac{2}{11}$	0
SP	$\frac{5}{3}$	$\frac{1}{3}$	0	1	0	$\frac{6}{5}$	0	0	$\frac{1}{10}$	0
6T	$\frac{14}{9}$	$\frac{1}{9}$	0	1	0	3	$-\frac{3}{5}$	1	0	0
8T	$\frac{19}{12}$	$\frac{1}{6}$	0	3	0	$\frac{147}{2}$	$\frac{51}{95}$	$\frac{23}{760}$	$\frac{9}{38}$	0
8P	$\frac{40}{27}$	$\frac{25}{54}$	0	4	$\frac{1}{36}$	$\frac{320}{393}$	$\frac{310}{393}$	0	$\frac{344}{1179}$	$\frac{23}{43}$
10P	$\frac{17}{12}$	$\frac{150}{994}$	$\frac{100}{0.038}$	2	$\frac{20}{0.09}$	$\frac{1065}{0.216}$	$\frac{1038}{0.216}$	$\frac{79}{0.177}$	$\frac{899}{0.502}$	$\frac{1798}{0.056}$
SLR	1.303	0.994	0.038	0.577	0.09	0.216	1.723	0.177	0.502	0.056

$$A = C_0 + \sum_{i=1}^k C_i \cdot D_i \quad (5)$$

where  $A$  is aromaticity of peripheral topological path (normalized value),  $C_0$  is intercept,  $C_i$  are weighting factors,  $D_i$  are (values of) descriptors, and  $k$  is number of descriptors.

PRECLAV program uses an “internal” cross-validation procedure (LOO method). We are not using here any “external” validation procedure, because the goal of our computations was not the validation of used statistical procedure(s).

### 3 Results and discussions

#### 3.1 QSArR study #1

Calibration set is made by 48 aromatic molecules of Table 1 from where the non-aromatic molecules **37** and **46** and anti-aromatic molecules **36**, **47** and **50** were eliminated; as descriptors the difference electronegativity and chemical hardness indices of Eqs. (3) and (4) and Table 2 were considered. The outlier molecules were identified nos. **1**, **38** and **42** of Table 1. The resulted QSArR equation, computed in absence of outliers, provides the results as

$$C_0 = 251.2821 \quad C_1 = 113.1341$$

$$D_1 \text{ is } \chi_{4C}$$

whereas the quality of correlation is described by the statistical indices

$$s = 27.0, \quad r^2 = 0.6624, \quad F = 42.2, \quad r_{CV}^2 = 0.6258$$

providing, nevertheless, a modest prediction power for computed QSArR in terms of the fourth order central difference electronegativity,  $\chi_{4C}$ , see eq. (3) with (4C) parameters of Table 2.

#### 3.2 QSArR study #2

Now, the calibration set contains all molecules of Table 1. The molecules: **1**, **37**, **42**, **46**, **47**, **49** and **50** were identified as outliers. The descriptors used are again the various finite difference forms of electronegativity and chemical hardness of Eqs. (3) and (4) with the parameters of Table 2. In these conditions, the QSArR equation, computed in absence of outliers, has the specialization:

$$C_0 = 373.7334 \quad C_1 = 161.667 \quad C_2 = 106.8799 \quad C_3 = -174.3477$$

$$D_1 \text{ is } \eta_{SLR} \quad D_2 \text{ is } \chi_{SLR} \quad D_3 \text{ is } \eta_{10P}$$

in terms of electronegativity and chemical hardness indices computed upon the spectral-like resolution ( $\chi_{SLR}$ ,  $\eta_{SLR}$ ) and tenth order pentadiagonal ( $\eta_{10P}$ ) schemes, see eqs. (3) and (4) with (SLR) and (10P) parameters of Table 2, respectively, while displaying the quality factors:

$$s = 26.9, \quad r^2 = 0.9499, \quad F = 272.0, \quad r_{CV}^2 = 0.9224$$

Worth noting the minimum correlation predictor / aromaticity found for  $D_3$  as  $r^2 = 0.0025$ , while the maximum inter-correlation of predictors was identified among the pair  $D_1 / D_3$  with the value  $r^2 = 0.3797$ . Overall, the utility of predictors in prediction of aromaticity is found as very high for chemical hardness indices within spectral-like resolution ( $D_1$ ) and within superior finite difference scheme of computation (the 10th order pentagonal  $D_3$ ), and with moderate results for spectral like resolution electro-negativity ( $D_2$ ).

### 3.3 QSArR study #3

Here we select again the calibration set of the study #1 above in conjunction with descriptors of electronegativity and chemical hardness as above, however supplemented with those provided by PRECLAV and DRAGON programs. In these conditions, the identified outliers are the molecules nos. **1** and **52** in Table 1, while the best correlation was found (in absence of outliers) with

$C_0 =$ 1647.2013	$C_1 =$ 418.5093	$C_2 =$ − 47.4139	$C_3 =$ − 78.3529	$C_4 =$ 0.1198
	$D_1$ is molecular volume/ molecular mass ratio	$D_2$ is spectral moment 09 from edge adjacency matrix weighted by edge degrees [26]	$D_3$ is 10+ $E_{LUMO}$ sum	$D_4$ is maximum aromaticity of aromatic chemical bonds

Note that the computed QSArR model does not include finite difference electro-negativity or chemical hardness descriptors, yet with a superior quality respecting the correspondent study #1:

$$s = 16.8, \quad r^2 = 0.8899, \quad F = 84.9, \quad r_{CV}^2 = 0.8511 \quad (6)$$

Nevertheless, the minimum correlation predictor / aromaticity was found for descriptor  $D_4$  as  $r^2 = 0.0046$ , whereas the maximum inter-correlation of predictors was established for the pair  $D_2 / D_3$  with  $r^2 = 0.5877$ . In conclusion, this model prescribes the utility of predictors in prediction of aromaticity as being very high for  $D_2$  and  $D_3$ , moderate for  $D_1$ , very low for  $D_4$ .

### 3.4 QSArR study #4

Finally, one considers again the calibration set of study #2 with the descriptors of the study #3. The outliers are identified as being molecules nos. **47** and **52** in Table 1; QSArR equation computed in absence of outliers takes the form:

$C_0 = -83.4478$	$C_1 = 0.4402$	$C_2 = 77.3705$	$C_3 = 712.2473$
	$D_1$ is maximum aromaticity of aromatic chemical bonds	$D_2$ is $\eta_{SLR}$	$D_3$ is D total accessibility index weighted by atomic masses [27]

With this model one record the mix influence of chemical hardness with spectral-like resolution descriptor  $\eta_{SLR}$  index with other structural descriptors, featuring the significant quality:

$$s = 50.5, \quad r^2 = 0.8851, \quad F = 123.3, \quad r_{CV}^2 = 0.8495$$

Since, the minimum correlation predictor / aromaticity was found for  $D_3$  descriptor as  $r^2 = 0.0899$ , while the maximum inter-correlation of predictors belongs to the descriptors' pair  $D_1/D_3$  as  $r^2 = 0.0999$  there can be assessed that the utility of predictors in prediction of aromaticity is very high for  $D_1$ , moderate for  $D_2$  (chemical hardness in spectral-like resolution scheme), and very low for  $D_3$ .

## 4 Conclusions

With the continuous interest in assessing the aromaticity concept an inherent molecular definition the so called QSArR multivariate equation between aromaticity activity and various structural parameters is here advanced questing upon the statistical influence the popular chemical reactivity indices such as electronegativity and chemical hardness have on it. While considering common organic compounds, thousands of software molecular descriptors and the compact finite difference formulas (up to those of spectral-like resolution) for electronegativity and chemical hardness, there was found that for some calibration sets, the superior finite difference electronegativity and chemical hardness schemes well correlate with aromaticity, especially when their regressions exclude the appearance of other molecular descriptors (see the study # 2 above). Such results may be useful when likely to implement electronegativity and chemical hardness schemes in modeling chemical reactivity for aromatic compounds. Yet, although QSArR methodology cannot replace the specific non-statistical algorithms used in aromaticity computations, it is useful for checking of predictor quality of certain descriptors in aromaticity computations. Further studies are necessary in order to better understand the correlation information contained in the aromaticity, with special focus on those descriptors related with electronic density distribution and their energetic functionals.



## References

1. B.T. Stepien, M.K. Cyranski, T.M. Krygowski, *Chem. Phys. Lett.* **350**, 537 (2001)
2. M. Thomsen, S. Döbel, P. Lassen, L. Carlsen, B.B. Mogensen, P.E. Hansen, *Chemosphere* **49**, 1317 (2002)
3. L. Caruso, G. Musumarra, A.R. Katritzky, *QSARs* **12**, 146 (1993)
4. R. Kiralj, Y. Takahata, M.M.C. Ferreira, *QSARs* **22**, 430 (2003)
5. L. Tarko, *ARKIVOC XI* 24 (2008)
6. F. Zheng, G. Zheng, G. Deaciuc, C.-G. Zhan, L.P. Dwoskin, P.A. Crooks, *Bioorg. Med. Chem.* **15**, 2975 (2007)
7. D. Ajloo, A.A. Saboury, N. Haghi-Asli, G. Ataei-Jafari, A.A. Moosavi-Movahedi, M. Ahmadi, K. Mahnam, S. Namaki, *J. Enz. Inh. Med. Chem.* **22**, 395 (2007)
8. G. Carrera, G. Aires-de Sousa, *Green Chem.* **7**, 20 (2005)
9. T.M. Krygowski, *J. Chem. Inf. Comput. Sci.* **33**, 70 (1993)
10. W. Kohn, A.D. Becke, R.G. Parr, *J. Phys. Chem.* **100**, 12974 (1996)
11. M.V. Putz, *Absolute and Chemical Electronegativity and Hardness* (Nova Publishers Inc, New York, 2008)
12. R.G. Parr, W. Yang, *Density Functional Theory of Atoms and Molecules* (Oxford University Press, New York, 1989)
13. M.V. Putz, *Int. J. Quantum Chem.* **106**, 361 (2006)
14. J.J.P. Stewart, *J. Mol. Model* **13**, 1173 (2007)
15. PCModel is available from Serena Software, Box 3076, Bloomington, IN 47402-3076, USA, Internet page <http://www.serenasoft.com/>
16. N.L. Alliger, *J. Am. Chem. Soc.* **99**, 8127 (1977)
17. The last version of MOPAC is available from Internet page <http://www.openmopac.net/>
18. The last version of DRAGON is available from Talete srl., via V. Pisani, 13-20124, Milano, Italy, Internet page <http://www.talete.mi.it>
19. The last version of PRECLAV is available from Center of Organic Chemistry - Bucharest, E-mail: ltarko@cco.ro or pfilip@cco.ro
20. L. Tarko, *Rev. Chim. (București)* **56**, 639 (2005)
21. L. Tarko, I. Lupescu, D. Groposila, Constantinescu, *ARKIVOC X*, 254 (2005)
22. L. Tarko, C.E. Stecoza, C. Ilie, M.C. Chifiriuc, *Rev. Chim. (București)* **60**, 476 (2009)
23. M.V. Putz, N. Russo, E. Sicilia, *J. Comput. Chem.* **25**, 994 (2004)
24. T. Koopmans, *Physica* **1**, 104 (1934)
25. S.G. Rubin, P.K. Khosla, *J. Comp. Phys.* **24**, 217 (1977)
26. E. Estrada, *J. Chem. Inf. Comput. Sci.* **36**, 844 (1996)
27. R. Todeschini, M. Lasagni, E.N. Marengo, *J. Chemom.* **8**, 263 (1994)