Inter*net* Electronic Journal of Molecular Design

March 2002, Volume 1, Number 3, Pages 108–114

Editor: Ovidiu Ivanciuc

Special issue dedicated to Professor Alexandru T. Balaban on the occasion of the 70th birthday
Part 3

Guest Editor: Mircea V. Diudea

QSAR Modeling of Mutagenicity Based on Graphs of Atomic Orbitals

Andrey A. Toropov and Alla P. Toropova

Vostok Holding Innovation Company, 700000, Sidik Asimov 4–th street 15, Tashkent, Uzbekistan

Received: August 24, 2001; Revised: November 26, 2001; Accepted: December 14, 2001; Published: March 31, 2002

Citation of the article:
A. A. Toropov and A. P. Toropova, QSAR Modeling of Mutagenicity Based on Graphs of Atomic Orbitals, *Internet Electron. J. Mol. Des.* **2002**, *1*, 108–114, http://www.biochempress.com.

Inter*net* Electronic Journal of Molecular Design

BioChem Press
http://www.biochempress.com

QSAR Modeling of Mutagenicity Based on Graphs of Atomic Orbitals[#]

Andrey A. Toropov* and Alla P. Toropova

Vostok Holding Innovation Company, 700000, Sidik Asimov 4–th street 15, Tashkent, Uzbekistan

Received: August 24, 2001; Revised: November 26, 2001; Accepted: December 14, 2001; Published: March 31, 2002

Internet Electron. J. Mol. Des. **2002**, *1* (3), 108–114

Abstract

The graph of atomic orbitals (*GAO*) has been used to represent molecular structures. Rules by which the Labeled Hydrogen–Filled Graphs (*LHFG*) were converted into the *GAO* are described. The *GAO* is an attempt at taking into account the structures of atoms (*i.e*., atomic orbitals such as $1s^1$, $2p^2$, $3d^{10}$, etc) for QSPR/QSAR studies. Optimization of correlation weights of local graph invariants (*OCWLI*) of the *LHFG* and the *GAO* have been used to model mutagenicity. As local graph invariants we have used the presence of different kinds of chemical elements and the presence of different values of the vertex degrees in the *LHFG*. In the case of the *GAO* as local invariants the presence of different kinds of the atomic orbitals and the presence of different values of the vertex degrees in the *GAO* have been used. Statistical characteristics of such models based on the *OCWLI* of *GAO* are better than those based on the *OCWLI* of the *LHFG*.

Keywords. QSAR; mutagenicity; optimization of correlation weights; graph of atomic orbitals.

1 INTRODUCTION

Quantitative structure–property/activity relationships (QSPR/QSAR) are important aspects of modern theoretical chemistry [1–13]. In recent studies Basak and Grunwald [14–16] used similarity methods to estimate mutagenicity of chemicals. As alternative of the similarity based mutagenicity modeling, the optimization of correlation weights of local graph invariants (*OCWLI*) may be used [17].

The present study is aimed at estimating the predictive ability of the *OCWLI* for modeling mutagenicity of heteroaromatic amines. The 73 heteroaromatic amines used to study the mutagenic potency were taken from Ref. [16]. The mutagenic activities of these compounds in *S. typhimurium* TA100 + S9 microsomal preparation are expressed in log of revertant per nonamole, ln *R*.

[#] Dedicated on the occasion of the 70th birthday to Professor Alexandru T. Balaban.

* Correspondence author; Sergeli 8A, Home 4, Room 6, 700085, Tashkent, Uzbekistan; E–mail: aator@saturn.silk.org.

2 MATERIALS AND METHODS

A labeled hydrogen–filled graph (*LHFG*) is the basis of the QSPR/QSAR studies. The vertices of such graphs represent atoms and the edges represent covalent bonds [1–17]. The *LHFG* has no information on the structure of atoms (*i.e.*, atomic orbitals such as $1s^1$, $2p^2$, $3d^{10}$, etc). Recently, the graph of atomic orbitals (*GAO*) has been suggested as an alternative to the *LHFG* [17]. The *GAO* is an attempt to take into account the structure of atoms in QSPR/QSAR analyses. The *GAO* are constructed on the basis of the *LHFG*. The conversion of the *LHFG* into the corresponding *GAO* may be carried out by the following scheme [17]:

1. Each vertex of the *LHFG* is replaced by a group of atomic orbitals (AOs) . Such groups of AOs on all atoms under consideration are listed in Table 1.

2. The element a_{ij} of adjacency matrix of the *GAO* is defined as 1 if the *i*–th and *j*–th vertices of the *GAO* fall in groups of different atoms from *LHFG* and these atoms have joint edge in *LHFG*; otherwise, $a_{ij} = 0$.

Table 1. Groups of Atomic Orbitals for Various Atoms

Atom	Atomic Orbitals
H	$1s^1$
C	$1s^2$, $2s^2$, $2p^2$
N	$1s^2$, $2s^2$, $2p^3$
O	$1s^2$, $2s^2$, $2p^4$
F	$1s^2$, $2s^2$, $2p^5$
S	$1s^2$, $2s^2$, $2p^6$, $3s^2$, $3p^4$
Cl	$1s^2$, $2s^2$, $2p^6$, $3s^2$, $3p^5$
Br	$1s^2$, $2s^2$, $2p^6$, $3s^2$, $3p^6$, $3d^{10}$, $4s^2$, $4p^5$

The *OCWLI* may be carried out by means of the following scheme. First of all, a descriptor calculated with the correlation weights of local graph invariants must be defined. For example one can use the following descriptor:

$$ {}^0X_{CW}(LHFG) = \sum_{i=1}^{n} CW(A_i) \times CW(VD_i) \tag{1} $$

where $CW(A_i)$ are the correlation weights *CW* in the presence of the atom that is represented by the *i*–th vertex of the *LHFG*, $CW(VD_i)$ is the *CW* of vertex degree, 1 to 4, of the *i*–th vertex of *LHFG*, and *n* is total number of vertices in the *LHFG*. Also one can use as the basis of the *OCWLI* the descriptor calculated with the *GAO*

$$ {}^0X_{CW}(GAO) = \sum_{i=1}^{n} CW(AO_i) \times CW(VD_i) \tag{2} $$

where $CW(AO_i)$ are *CW* in the presence of the AO that is represented by the *i*–th vertex of *GAO*, $CW(VD_i)$ the *CW* of the vertex degree, 1 to 4, of *i*–th vertex of *GAO*, and *n* is total number of vertices in the *GAO*.

The values of descriptors in Eqs. (1) and (2) are functions of the *CW*s. The value of the correlation coefficient between the descriptors of Eqs. (1) and (2) and property/activity of interest is a function of the above–mentioned *CW*s. The algorithm is described by the following steps:

1. Using the Monte Carlo method [17] the values of the *CW*s are calculated that produce as large as possible values for the correlation coefficient between the values of property/activity of interest and the values of the descriptors of Eqs. (1) and (2). The start values of the *CW*s are defined as 1.0.

2. Using the least squares method the following equation is obtained:

$$PA = C_0 + C_1 \times {}^0X_{CW}(G) \tag{3}$$

where *PA* is the property/activity of interest, ${}^0X_{CW}(G)$ is the descriptor of Eqs. (1) and (2), and *G* is the *LHFG* or the *GAO*.

3. The predictive ability of Eq. (3) may be validated with compounds of a test set.

3 RESULTS AND DISCUSSION

Statistical characteristics of the mutagenicity models obtained by means of three probes of the *OCWLI* based on the *LHFG* and three probes of the *OCWLI* based on the *GAO* are presented in Table 2. From the Table 2 one can see that the statistical characteristics of these models for each probe of the *OCWLI* are practically identical. Also one can see that the models based on the *GAO* are better than those based on the *LHFG*.

Table 2. Statistical Characteristics of the *OCWLI* Models of Mutagenicity

		Training set, n = 36			Test set, n = 37		
C_1	C_0	*r*	*s*	F	*r*	*s*	F
Models based on the *LHFG*							
0.200	–1.240	0.8029	0.861	62	0.8856	0.641	127
0.155	–1.039	0.8020	0.862	61	0.8959	0.616	142
0.249	–1.058	0.8021	0.862	61	0.8957	0.616	142
Models based on the *GAO*							
0.712	–0.963	0.8233	0.819	72	0.8972	0.610	145
0.669	–0.979	0.8225	0.821	71	0.8986	0.609	147
0.620	–0.979	0.8231	0.820	71	0.8979	0.611	146

The values of the *CW*s for the three *OCWLI* probes of local invariants of the *LHFG* that produce the largest possible value of correlation coefficient between the ln *R* and ${}^0X_{CW}(LHFG)$ are listed in Table 3.

The values of the CWs for the three *OCWLI* probes of the *GAO* local invariants for the three probes of the *OCWLI* that produce the largest possible value of correlation coefficient between the ln *R* and ${}^0X_{CW}(GAO)$ are listed in Table 4. The statistical characteristics of the mutagenicity model that is the result of the first probe of the *OCWLI* based on the *GAO* are the best. The model is calculated as:

$$\ln R = 0.712\ {}^{0}X_{CW}(GAO) - 0.963$$
$$\text{Training set } n = 36,\ r = 0.8233,\ s = 0.819,\ F = 72 \qquad (4)$$
$$\text{Test set } n = 37,\ r = 0.8972,\ s = 0.610,\ F = 145$$

Table 3. Correlation Weights (*CW*s) of the ${}^{0}X_{CW}(LHFG)$

Local invariants	*CW*s of Probe 1	*CW*s of Probe 2	*CW*s of Probe 3
Atoms (A_i)			
H	–0.550	–0.600	–0.438
C	2.257	2.866	1.303
N	–5.310	–6.571	–3.097
O	6.982	–19.069	–9.367
F	–2.712	–2.874	–2.205
S	–6.853	16.145	7.860
Cl	0.470	0.156	0.150
Br	2.291	1.921	1.502
LHFD vertex degree (VD_i)			
1	1.369	1.768	1.429
2	–0.358	0.186	0.238
3	0.792	0.795	1.079
4	0.937	0.962	1.272

Table 4. Correlation Weights (*CW*s) of the ${}^{0}X_{CW}(GAO)$

Local invariants	*CW*s of probe 1	*CW*s of probe 2	*CW*s of probe 3
Atomic orbitals (AO_i)			
$1s^1$	–2.001	–1.670	–1.430
$1s^2$	–1.068	–1.123	–1.193
$2s^2$	–0.480	–0.836	–0.698
$2p^2$	3.799	3.848	3.943
$2p^3$	1.377	1.800	1.770
$2p^4$	1.296	1.715	1.657
$2p^5$	–0.238	0.318	0.565
$2p^6$	0.572	0.936	0.979
$3s^2$	0.729	0.475	0.678
$3p^4$	0.852	0.812	0.600
$3p^5$	0.525	0.286	0.164
$3p^6$	–0.348	0.131	0.316
$3d^{10}$	–0.457	0.238	–0.112
$4s^2$	0.206	0.114	0.331
$4p^5$	0.637	0.150	0.128
GAO vertex degree (VD_i)			
3	0.762	0.875	1.180
4	3.064	4.032	4.028
5	–9.683	–9.060	–14.667
6	1.915	2.175	2.321
7	0.752	0.872	0.918
8	1.448	1.668	1.762
9	0.269	0.337	0.338
10	0.260	0.242	0.297
11	0.136	0.421	0.338
14	0.587	0.623	0.412

Table 5 lists the values of ${}^{0}X_{CW}(GAO)$, the experimental [16] and calculated with Eq.(4) values of mutagenicity of the 73 heteroaromatic amines considered. It is to be noted that the model reported in Ref. [16] is based on the octanol/water partition coefficient values and the electronic descriptors E_{LUMO} and E_{HOMO}. We have to consider that the statistical characteristics of the model of

Eq. (4) on all compounds under consideration are better (n = 73, r = 0.857, s = 0.72) than those reported in ref. [16], namely n = 73, r = 0.84, s = 0.75.

Table 5. Lists of the Training Set and Test Set, Experimental [16] and Calculated with Eq. (4) Values of Mutagenicity for the Heteroaromatic Amines Under Consideration

No	Compound	$^0X_{CW}(GAO)$	log R		
			Exp	Calc	Res
	Training Set				
1	1–Aminofluoranthene	4.357	2.34	2.14	0.20
2	4–Methyl–2–bromoaniline	1.220	–0.64	–0.09	–0.55
3	2–Ethyl–4–chloroaniline	0.786	0.08	–0.40	0.48
4	4–Cyclohexylaniline	0.599	–0.14	–0.54	0.40
5	4,4'–Ethylenebis(aniline)	–0.225	–1.51	–1.12	–0.39
6	2,4,5–Trimethylaniline	0.573	–0.26	–0.56	0.30
7	2,4–Diamino–n–butylbenzene	–0.101	–0.84	–1.03	0.19
8	7–Aminofluoranthene	4.357	2.76	2.14	0.62
9	4–Methoxy–2–methylaniline	–0.083	–2.10	–1.02	–1.08
10	2–Aminobiphenyl	1.935	–0.51	0.41	–0.92
11	2,6–Dichloro–1,4–phenylenediamine	–0.205	–1.12	–1.11	–0.01
12	4–Aminophenyldisulfide	2.132	0.54	0.55	–0.01
13	2–Aminocarbazole	1.156	–0.56	–0.14	–0.42
14	1–Aminofluorene	3.020	–0.04	1.19	–1.23
15	2–aminoanthracene	3.146	2.76	1.28	1.48
16	2–Amino–3–methylnaphthalene	1.773	1.09	0.30	0.79
17	4,4'–Methylene–bis–(*o*–fluoroaniline)	–0.658	–1.16	–1.43	0.27
18	3–Methoxy–4–methylaniline	–0.083	–0.81	–1.02	0.21
19	2–Chloroaniline	0.402	–2.05	–0.68	–1.37
20	4–Phenoxyaniline	1.452	0.63	0.07	0.56
21	2–Amino–4–chlorophenol	–1.458	–2.00	–2.00	0.00
22	6–Aminochrysene	4.693	2.41	2.38	0.03
23	2–Methyl–4–bromoaniline	1.220	0.46	–0.09	0.55
24	4,4'–Methylenebis(*o*–ethylaniline)	1.956	–0.55	0.43	–0.98
25	4–Ethoxyaniline	–0.047	–0.61	–1.00	0.39
26	5–Aminoquinoline	1.103	0.14	–0.18	0.32
27	2–Methyl–4–chloroaniline	0.576	0.38	–0.55	0.93
28	1–Aminonaphthalene	1.599	–1.00	0.18	–1.18
29	2,4–Dimethylaniline	0.400	–0.23	–0.68	0.45
30	2,4–Difluoroaniline	–1.795	–2.52	–2.24	–0.28
31	3,4'–Diaminobiphenyl	0.979	0.65	–0.27	0.92
32	3–Aminophenanthrene	3.146	2.66	1.28	1.38
33	2–Aminophenanthrene	3.146	2.74	1.28	1.46
34	1–Aminoanthracene	3.146	0.36	1.28	–0.92
35	1–Aminopyrene	4.357	1.05	2.14	–1.09
36	9–Aminoanthracene	3.146	–0.24	1.28	–1.52
	Test Set				
1	2,3–Dimethylaniline	0.400	–1.36	–0.68	–0.68
2	2,5–Dimethylaniline	0.400	–1.43	–0.68	–0.75
3	4–Chloro–1,2–phenylenediamine	–0.554	–1.44	–1.36	–0.08
4	4–Aminophenylsulfide	1.538	0.48	0.13	0.35
5	4–Aminopyrene	4.357	2.69	2.14	0.55
6	2–Amino–4–methylphenol	–1.634	–1.68	–2.13	0.45
7	2–Aminofluorene	3.020	0.78	1.19	–0.41
8	Benzidine	0.979	–0.66	–0.27	–0.39
9	8–Aminoquinoline	1.103	–0.34	–0.18	–0.16
10	3,4–Dimethylaniline	0.400	–1.08	–0.68	–0.40
11	3–Aminofluorene	3.020	0.10	1.19	–1.09
12	4–Methyl–2–chloroaniline	0.576	–0.40	–0.55	0.15

Table 5. (Continued)

No	Compound	${}^0X_{CW}(GAO)$	log *R*		
			Exp	Calc	Res
13	4–Aminofluorene	3.020	0.64	1.19	–0.55
14	4–Chloroaniline	0.402	–1.51	–0.68	–0.83
15	8–Aminofluoranthene	4.357	1.98	2.14	–0.16
16	2–Aminopyrene	4.357	2.58	2.14	0.44
17	2–aminonaphthalene	1.599	0.39	0.18	0.21
18	6–Aminoquinoline	1.103	–1.22	–0.18	–1.04
19	2–Amino–1–methylnaphthalene	1.773	0.84	0.30	0.54
20	4–Amino–3–methylbiphenyl	2.109	1.12	0.54	0.58
21	2–Methoxy–5–methylaniline	–0.083	–1.85	–1.02	–0.83
22	1–Aminocarbazole	1.156	–0.25	–0.14	–0.11
23	1–Aminophenanthrene	3.146	1.79	1.28	0.51
24	4–Amino–3–methylbiphenyl	2.109	0.09	0.54	–0.45
25	3–Aminocarbazole	1.156	–0.11	–0.14	0.03
26	3–Aminofluoranthene	4.357	2.25	2.14	0.11
27	4–Aminobiphenyl	1.935	0.85	0.41	0.44
28	3,3'–Dichlorobenzidine	1.678	0.66	0.23	0.43
29	3,3'–Dimethoxybenzidine	0.361	–0.85	–0.71	–0.14
30	4–Aminocarbazole	1.156	–0.47	–0.14	–0.33
31	2–Aminofluoranthene	4.357	2.87	2.14	0.73
32	3–Aminoquinoline	1.103	0.07	–0.18	0.25
33	1–Amino–2–methylnaphthalene	1.773	–0.37	0.30	–0.67
34	4–Aminophenanthrene	3.146	–0.11	1.28	–1.39
35	4–Aminophenylether	0.496	–0.27	–0.61	0.34
36	4,4'–Methylenedianiline	1.188	–0.15	–0.12	–0.03
37	9–Aminophenanthrene	3.146	2.79	1.28	1.51

4 CONCLUSIONS

The present study indicates that the optimized molecular descriptor ${}^0X_{CW}(GAO)$ shows reasonable correlation between the chemical structure and the mutagenicity of heteroaromatic amines, *i.e*. the *GAO* may be used as alternative to *LHFG* in the QSAR modeling of the mutagenicity.

5 REFERENCES

[1] M. V. Diudea (Ed.), *QSPR/QSAR Studies by Molecular Descriptors*, Nova Science, Huntington, N.Y., 2001.
[2] O. Ivanciuc and A. T. Balaban, Graph Theory in Chemistry; in: *The Encyclopedia of Computational Chemistry*, Eds. P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, and P. R. Schreiner, John Wiley & Sons, Chichester, 1998, pp. 1169–1190.
[3] O. Ivanciuc and A. T. Balaban, The Graph Description of Chemical Structures; in: *Topological Indices and Related Descriptors in QSAR and QSPR*, Eds. J. Devillers and A. T. Balaban, Gordon and Breach Science Publishers, Amsterdam, 1999, pp 59–167.
[4] O. Ivanciuc, T. Ivanciuc, and A. T. Balaban, Vertex– and Edge–Weighted Molecular Graphs and Derived Structural Descriptors; in: *Topological Indices and Related Descriptors in QSAR and QSPR*, Eds. J. Devillers and A. T. Balaban, Gordon and Breach Science Publishers, Amsterdam, 1999, pp 169–220.
[5] O. Ivanciuc, T. Ivanciuc, D. Cabrol–Bass, and A. T. Balaban, Comparison of Weighting Schemes for Molecular Graph Descriptors: Application in Quantitative Structure–Retention Relationship Models for Alkylphenols in Gas–Liquid Chromatography, *J. Chem. Inf. Comput. Sci*. **2000**, *40*, 732–743.
[6] A. T. Balaban, Chemical Graphs. Part 48. Topological Index J for Heteroatom–Containing Molecules Taking Into Account Periodicities of Element Properties, *MATCH* (*Commun. Math. Chem*.) **1986**, *21*, 115–122.
[7] O. Ivanciuc, T. Ivanciuc, D. J. Klein, W. A. Seitz, and A. T. Balaban, Wiener Index Extension by Counting

Even/Odd Graph Distances, *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 536–549.
[8] O. Ivanciuc, T. Ivanciuc, D. J. Klein, W. A. Seitz, and A. T. Balaban, Quantitative Structure–Retention Relationships for Gas Chromatographic Retention Indices of Alkylbenzenes with Molecular Graph Descriptors, *SAR QSAR Environ. Res*. **2001**, *11*, 419–452.
[9] A. T. Balaban, D. Mills, O. Ivanciuc, and S. C. Basak, Reverse Wiener Indices, *Croat. Chem. Acta* **2000**, *73*, 923–941.
[10] O. Ivanciuc, T. Ivanciuc, and A. T. Balaban, The Complementary Distance Matrix, a New Molecular Graph Metric, *ACH – Models Chem*. **2000**, *137*, 57–82.
[11] A. T. Balaban and T.–S. Balaban, New Vertex Invariants and Topological Indices of Chemical Graphs Based on Information on Distances, *J. Math. Chem*. **1991**, *8*, 383–397.
[12] A. T. Balaban and M. V. Diudea, Real Number Vertex Invariants: Regressive Distance Sums and Related Topological Indices, *J. Chem. Inf. Comput. Sci*. **1993**, *33*, 421–428.
[13] A. T. Balaban, Highly Discriminating Distance–Based Topological Index, *Chem. Phys. Lett*. **1982**, *89*, 399–404.
[14] S. C. Basak and G. D. Grunwald, Tolerance Space and Molecular Similarity, *SAR QSAR Environ. Res*. **1995**, *3*, 265–217.
[15] S. C. Basak and G. D. Grunwald, Molecular Similarity and Estimation of Molecular Properties, *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 366–372.
[16] S. C. Basak and G. D. Grunwald, Prediciting Mutagenicity of Chemicals using Topological and Quantum Chemical Parameters: a Similarity based Study, *Chemosphere* **1995**, *31*, 2529–2546.
[17] A. A. Toropov and A. P. Toropova, Prediction of Heteroaromatic Amine Mutagenicity by Means of Correlation Weighting of Atomic Orbital Graph of Local Invariants, *J. Mol. Struct*. (*Theochem*) **2001**, *538*, 287–293.