

Computation of Parameters for Generation of the General a_N -index (GAI) and Applications of GAI

Lu Xu*, Fei Zhong and Chang-Yu Hu

(Changchun Institute of Applied Chemistry, Academia Sinica,
Changchun 130022, People's Republic of China)

Abstract

The general a_N -index suggested by us is based on quantum theory. In this paper we give that a complete set of parameters for the general a_N -index derivation calculated recently by us, and some key steps for computation of those parameters which were not given in our previous papers. Examples of application are also given.

INTRODUCTION

The essential assumption for quantitative structure-activity relationship (QSAR) or quantitative structure-property relationship (QSPR) studies is that the activities or properties depend on the structure of compounds. Thus, if we have methods to characterize appropriately the structures of compounds, the activities or properties of compounds can be quantitatively predicted in terms of the characterization. For this purpose, a variety of methods has been suggested, in which the topological index technique is an active area. The topological index is a numeric quantity that is mathematically derived from the structure graph of a molecule. Many topological indices, such as the Wiener number¹, the Hosoya index², the Randic connectivity index³ and its extension⁴, Balaban averaged distance sum connectivity⁵, have been proposed to convert chemical structures into numerical values.

In 1983, Yang and Jiang⁶, suggested a new topological index a_N , based on quantum theory which performs well but can only be used for saturated hydrocarbons. For this reason, the a_N -index was extended, called the general a_N -index(GAI), and applied to a variety of phosphorous-containing compounds with satisfactory results⁷. In this paper, we will give: (1) a complete set of parameters for the general a_N -index derivation calculated recently by us. These parameters enable the GAI to be used for nearly all organic compounds including heteroatom-containing molecules as well as cis/trans isomers. (2) some key steps for computation of those parameters which were not given in our previous papers.

CALCULATION OF THE PARAMETERS

For the GAI computation, the following assumptions are made: only the hydrogen-depleted structure, i.e., the molecular skeleton is considered; the interaction between the atoms of a molecule is described by using hybrid orbitals, and only the linked atoms are considered. The types of interaction between the orbitals can be distinguished as (a), (b), (c), (c)', (d), and (d)'^{8,9}, and the relative magnitudes corresponding to these types are denoted respectively by b, c, c', d, and d'. For saturated hydrocarbons, c'=c, d'=d. If a molecule contains heteroatom(s) and / or multiple bonds, e.g., when all the hybrid orbitals to the linked atoms are sp^2 , then c' will be not equal to c, and d' will be not equal to d.

The main procedure for computing the GAIs includes three steps:

(1) The calculation of the parameters b, c, c', d, and d' by means of a quantum chemical technique, e.g., CNDO/2.

(2) The construction of a molecule skeleton graph and an orbital interaction graph of linked atoms (OIGLA) in this molecule.

(3) The formation of an orbital interaction matrix of linked atoms (OIMLA). The general a_N -index is defined as the absolute value of the constant item of a molecular eigenpolynomial. It is equivalent to the computation of the determinant of OIMLA, i.e., the $GAI = |\det(OIMLA)|$.

As an example, the skeleton of the example molecule and its OIGLA as well as OIMLA are shown in Fig 1. The values of δ_i , b_{ij} , c_{ij} and d_{ij} in Fig 1 are given by

$$\begin{aligned} \delta_i &= (\alpha_i - \alpha) / \beta, \quad b_{ij} = \beta_{ij} (b) / \beta \\ c_{ij} &= \beta_{ij} (c) / \beta, \quad d_{ij} = \beta_{ij} (d) / \beta \end{aligned} \quad (1)$$

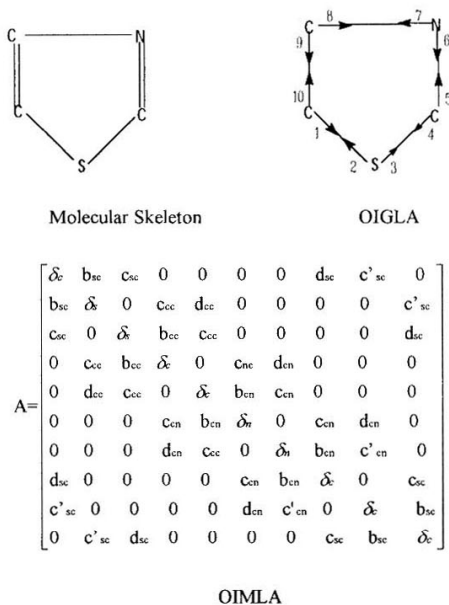


Fig 1. Example of molecular skeleton, and the corresponding OIGLA as well as OIMLA

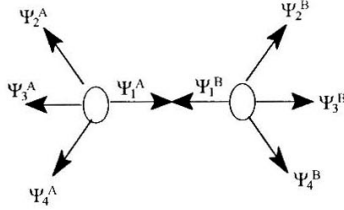
where α_i stands for the energy of the i -th hybrid orbital ϕ_i ; β_{ij} is the interaction energy of ϕ_i, ϕ_j using standard bond lengths and bond angles; α represents the energy of an sp^3 -hybrid orbital and β denotes the interaction energy of sp^3 - sp^3 within single C-C bond length. Evidently, the function for c'_{ij} and the function for d'_{ij} are the same as c_{ij} and d_{ij} , respectively. The Wolfsberg-Helmholz expression for β_{ij} is

$$\beta_{ij} = k (\alpha_i + \alpha_j) s_{ij} / 2 \quad (2)$$

where s_{ij} is the overlap integral of ϕ_i and ϕ_j , obtained using a quantum chemical technique, such as the semiempirical CNDO/2 as mentioned above; $k=1$, when the element belongs to the first or second family in the period table of elements; $k=2$, when the element belongs to the third family. In the following section, we will take orbitals sp^3 - sp^3 of the C-C bond as an example to illustrate the calculation of those parameters. Similarly, we can calculate the other hybrid orbital types.

All the computations are based on the right hand coordinate system using the theory of valence bond.

The wave function of sp^3-sp^3 is represented as



$$\Psi_1^A = 1/2\Psi_s^A - \sqrt{3}/2\Psi_x^A$$

$$\Psi_2^A = 1/2\Psi_s^A - \sqrt{6}/3\Psi_x^A + \sqrt{3}/6\Psi_z^A$$

$$\Psi_3^A = 1/2\Psi_s^A + \sqrt{6}/6\Psi_x^A - \sqrt{2}/2\Psi_y^A + \sqrt{3}/6\Psi_z^A$$

$$\Psi_4^A = 1/2\Psi_s^A + \sqrt{6}/6\Psi_x^A + \sqrt{2}/2\Psi_y^A + \sqrt{3}/6\Psi_z^A$$

$$\Psi_1^B = 1/2\Psi_s^B + \sqrt{3}/2\Psi_x^B$$

$$\Psi_2^B = 1/2\Psi_s^B + \sqrt{6}/3\Psi_x^B - \sqrt{3}/6\Psi_z^B$$

$$\Psi_3^B = 1/2\Psi_s^B - \sqrt{6}/6\Psi_x^B + \sqrt{2}/2\Psi_y^B - \sqrt{3}/6\Psi_z^B$$

$$\Psi_4^B = 1/2\Psi_s^B - \sqrt{6}/6\Psi_x^B - \sqrt{2}/2\Psi_y^B - \sqrt{3}/6\Psi_z^B$$

Thus, the overlap integrals of ϕ_i and ϕ_j are (Ψ_1^A, Ψ_1^B) , (Ψ_1^A, Ψ_2^B) , (Ψ_1^A, Ψ_3^B) , (Ψ_1^A, Ψ_4^B) , (Ψ_2^A, Ψ_1^B) , (Ψ_2^A, Ψ_2^B) , (Ψ_2^A, Ψ_3^B) , (Ψ_2^A, Ψ_4^B) , ..., (Ψ_4^A, Ψ_4^B) . For example,

$$(\Psi_1^A, \Psi_1^B) = 1/4\Psi_s^A\Psi_s^B + \sqrt{3}/4\Psi_x^A\Psi_x^B - \sqrt{3}/4\Psi_z^A\Psi_z^B - 3/4\Psi_x^A\Psi_z^B$$

$$(\Psi_1^A, \Psi_2^B) = 1/4\Psi_s^A\Psi_s^B - \sqrt{3}/4\Psi_x^A\Psi_x^B + \sqrt{6}/6\Psi_s^A\Psi_x^B - \sqrt{2}/2\Psi_z^A\Psi_x^B - \sqrt{3}/12\Psi_s^A\Psi_z^B + 1/4\Psi_z^A\Psi_z^B$$

The overlap integral values of hybrid sp^3-sp^3 obtained using the CNDO/2 program are shown in Table I. Among the hybrid orbitals $sp^3(A)-sp^3(B)$, we have

$$b \propto (\Psi_1^A, \Psi_1^B)$$

$$c \propto (\Psi_1^A, \Psi_4^B)$$

$$d \propto [(\Psi_2^A, \Psi_2^B) + \Psi_3^B + \Psi_4^B]/3$$

Table I Values of sp^3-sp^3 overall integral of C-C bond

	φ_s^A	φ_x^A	φ_y^A	φ_z^A
φ_s^B	0.3454	0	0	-0.3689
φ_x^B	0	0.1957	0	0
φ_y^B	0	0	0.1957	0
φ_z^B	0.3689	0	0	-0.3299

Table II Energies of the atomic orbitals

orbital	C	N	O	P	S	F	Cl	Br	I
sp^3	0	0.58	1.51	0.30	0.55	2.38	1.17	0.81	0.92
sp^2	0.03	0.60	1.53	0.32	0.56	2.40	1.18	0.82	0.925
sp	0.08	0.63	1.56	0.34	0.57	2.42	1.19	0.83	0.935

So that from Table I we obtain

$$(\Psi_1^A, \Psi_1^B) = 0.653$$

$$(\Psi_1^A, \Psi_2^B) = (\Psi_1^A, \Psi_3^B) = (\Psi_1^A, \Psi_4^B) = 0.110$$

$$((\Psi_2^A, \Psi_2^B) + \Psi_3^B + \Psi_4^B) / 3 = 0.00730$$

and then

$$\beta = K(\alpha_{sp^3} + \alpha_{sp^3})(\Psi_1^A, \Psi_1^B) / 2 = 28.614$$

$$\beta_{sp^3}(b) = K(\alpha_{sp^3} + \alpha_{sp^3})(\Psi_1^A, \Psi_1^B) / 2 = 28.614$$

$$\beta_{sp^3}(c) = K(\alpha_{sp^3} + \alpha_{sp^3})((\Psi_2^A, \Psi_2^B) + \Psi_3^B + \Psi_4^B) / 6 = 0.320$$

where, $K=1$, because element carbon is in the second family of the periodic table of elements ; α_i are taken from Table II. Putting the above results into eqs.1 and eqs.2, we obtain

$$\delta_c = (\alpha_{sp^3} \cdot \alpha_{sp^3}) = 0$$

$$b = 28.614 / 28.614 = 1$$

$$c = c' = 4.820 / 28.614 = 0.168$$

$$d = d' = 0.320 / 28.614 = 0.011$$

Similarly, we have calculated the other types of interactions between the orbitals . All the results are displayed in Table III. The bond lengths for above computation are shown in Table IV.

Similarly, we have calculated the other types of interactions between the orbitals . All the results are displayed in Table III. The bond lengths for above computation are shown in Table IV.

Table III Weights of orbital interactions

hybrid orbital	A	B	b	c	c'	d	d'
sp ³ (A)-sp ³ (B)	C	C	1.00000	0.167015	0.167015	0.110317	0.110317
	C	N	1.098089	0.1360393	0.2084177	0.0106463	0.0106463
	C	O	1.251406	0.1181717	0.2663458	0.0108371	0.0108371
	C	P	0.7881041	0.159038	0.0814103	0.0073915	0.0073915
	C	S	0.8275154	0.134615	0.0950946	0.0071154	0.0071154
	C	F	1.3522173	0.1010256	0.321407	0.0107599	0.0107599
	C	Cl	0.9230277	0.1214806	0.1160519	0.0072234	0.0072234
	C	Br	0.8793208	0.1455204	0.1455204	0.0069516	0.0069516
	N	N	1.1427859	0.1556095	0.1556095	0.0091623	0.0091623
	N	P	0.9741623	0.3021858	0.1057364	0.0132359	0.0132359
	O	O	1.4921095	0.1871053	0.1871053	0.0104959	0.0104959
	O	P	0.861355	0.2245476	0.0464235	0.0060938	0.0060938
	O	S	1.1602628	0.3182999	0.0991994	0.0120407	0.0120407
	P	P	0.8949119	0.1143954	0.1143954	0.0073935	0.0073935
	P	S	1.0783792	0.1435726	0.1990457	0.0121206	0.0121206
	S	S	0.9129758	0.1042098	0.1042098	0.0063466	0.0063466
F	F	1.1091131	0.1021734	0.1021734	0.0047381	0.0047381	
Cl	Cl	1.0150223	0.1016619	0.1016619	0.0057625	0.0057625	
sp ³ (A)-sp ² (B)	C	C	1.0390664	0.1432331	0.0643796	0.0027684	0.0027684
	C	N	1.1258503	0.0963783	0.092435	-0.000506	-0.000506
	C	O	1.3661387	0.0834898	0.1560766	-0.0037309	-0.0037309
	O	N	1.6606858	0.2709208	0.0512497	0.006301	0.006301
	N	C	1.2910668	0.2596766	0.0585939	0.0072826	0.0072826
	O	C	1.4593972	0.3172823	0.014898	0.0082283	0.0082283
	S	C	0.9011421	0.076747	0.0628515	0.0003527	0.0003527
Cl	C	1.0173658	0.0999526	0.0415764	0.0014762	0.0014762	

Table III (Continued)

sp ³ (A)-sp(B)	C	C	1.1347908	0.0497591	0.1967208	-0.0282162	-0.0282162
	P	O	1.0235146	-0.1444748	0.3106191	-0.0650107	-0.0650107
sp ² (A)-sp ² (B)	C	C	1.2091428	0.2186375	0.2186375	-0.2184651	0.053383
	C	N	1.3705965	0.1733069	0.2914545	-0.2346396	0.2181925
	C	O	1.6044153	0.1382156	0.3869338	-0.2614253	0.2384231
	C	S	0.9391159	0.1393451	0.0805768	-0.1309541	0.121631
	O	O	1.823674	0.1932094	0.1932094	-0.2342501	0.2147674
	O	N	1.7564684	0.2975987	0.1783341	-0.2598936	0.2391699
	N	N	1.560686	0.2410216	0.2410216	-0.2535538	0.2361372
	S	S	1.1077123	0.1084256	0.1084256	-0.1457436	0.1375659
sp ² (A)-sp(B)	C	C	1.1449966	0.03402	0.1568071	-0.0359031	-0.0359031
	C	S	1.0719438	0.0926412	0.113081	-0.026965	-0.026965
	S	C	1.0645241	-0.0144531	0.1906606	-0.040354	-0.040354
sp(A)-sp(B)	C	C	1.4179021	0.1795831	0.179583	-0.1161903	-0.1161903
	C	N	1.5868528	0.060829	0.2547209	-0.1181101	-0.1181101
	N	N	1.795723	0.1366176	0.1366176	-0.1184935	-0.1184935

EXAMPLE OF APPLICATION OF THE GAI

1. Distinguishing of structural isomers

At present, over one hundred of topological indices have been suggested. Most of them are generated based on the topological features of molecules, such as atom type, bond type as well as the immediate bonding environment of each atom, etc., having neither a quantum-mechanical nor any other theoretical foundation. Thus, those indices usually can not distinguish cis/trans isomers. The general a_N -index is a quantum-theory-based topological number, so that a prominent characteristic of this index is that it can distinguish between geometric isomers since the d_{ij} and d_{ij}' parameters which are different from sp^2 -hybrid orbitals, are considered when calculating the general a_N -index. For instance, the cis and trans forms of an olefin can be distinguished.

Table IV List of bond length

sp ³ (A)-sp ³ (B)	C-C:1.54	C-N: 1.479	C-O:1.43	C-P:1.87	C-S:1.81
	C-Cl:1.77	C-F:1.38	C-Br:1.93	C-I:2.14	
	N-N:1.46	N-P:1.49			
	O-O:1.32	O-P:1.76	O-S:1.43	O-F:1.415	O-Cl:1.546
	O-Br:1.65	O-I:1.83			
	P-P:2.21	P-S:1.86			
	S-S:2.07				
	F-F:1.42	Cl-Cl:1.99	Br-Br:2.28	I-I:2.67	
sp ² (A)-sp ² (B)	C-C:1.34	C-N:1.27	C-O:1.22	C-S:1.71	
	N-N:1.20				
	O-O:1.21	O-P:1.55			
	S-S:1.88				
sp(A)-sp(B)	C-C:1.53	C-N:1.158			
	N-N:1.10				
sp ³ (A)-sp ² (B)	C-C:1.53	C-N:1.479	C-O:1.36		
	O-N:1.21				
sp ³ (A)-sp(B)	C-C:1.46				
sp ² (A)-sp ³ (B)	C-C:1.53	C-N:1.322	C-O:1.312	C-S:1.73	C-Cl:1.69
sp ² (A)-sp(B)	C-C:1.49	C-S:1.558			

In Table V, the general a_N -indices of 22 geometric isomers of olefins and boiling points are given. This Table shows that the general a_N -indices decrease with increasing the size of the molecule of both the cis and trans forms. The correlation between the GAI, GAI^2 , number of C atom, N, and boiling points of some olefins have been examined and the results are as follows.

$$bp(^{\circ}C) = 14.0103-259.4834GAI+109.0879GAI^2+32.9619N$$

$$R = 0.9968, \quad F = 933, \quad S = 3.3, \quad n = 22$$

The calculated boiling points are also shown in Table V, where R denotes correlation coefficient; F represents F-test value, S stands for standard derivation; n is the number of samples.

Table V. The GAI and Boiling Points of 22 Olefins

Compound	GAI	GAI ²	N	bp(^o C)(exp)	bp(^o C)(cal)
1. cis-butene-2	1.5264	2.3298	4	3.72	3.86
2. trans-butene-2	1.5479	2.3961	4	0.88	5.90
3. cis-pentene-2	1.4943	2.2329	5	36.90	34.50
4. trans-pentene-2	1.5156	2.2969	5	36.40	36.24
5. cis-4-Me-pentene-2	1.4907	1.9339	6	36.30	62.02
6. trans-4-Me-pentene-2	1.4106	1.9897	6	58.60	62.96
7. cis-hexene-3	1.4629	2.1401	6	66.40	65.63
8. trans-hexene-3	1.4839	2.2018	6	67.10	66.95
9. cis-3-Me-pentene-2	1.4826	2.1980	6	70.40	66.09
10. trans-3-Me-pentene-2	1.5205	62.3119	6	67.60	64.03
11. cis-hexene-2	1.4116	61.9925	6	68.80	69.23
12. trans-hexene-2	1.4317	62.0497	6	67.80	63.02
13. cis-heptene-3	1.3819	71.9374	7	95.70	94.52
14. trans-heptene-3	1.4017	71.9648	7	96.00	95.49
15. cis-heptene-2	0.9247	0.8554	7	98.50	98.07
16. trans-heptene-2	1.3516	1.8267	7	98.00	93.34
17. cis-octene-3	1.3046	1.70267	8	112.90	124.97
18. trans-octene-3	1.3233	1.7501	8	113.30	125.24
19. cis-octene-4	1.3504	1.7041	8	112.50	124.75
20. trans-octene-4	1.3241	1.7533	8	121.90	125.35
21. cis-octene-2	1.2580	1.5826	8	125.60	123.96
22. trans-octene-2	1.2796	1.6279	8	125.00	124.22

2. Relationship between the logP of 25 barbiturates and the GAI

Barbiturates were thought to be nonspecific narcotic agents principally because logP (P= partition coefficient in octanol-water) correlates very well with their biological potency^[12].

Other studies^[10] show a dependence of the action of barbiturates upon chemical structure. Therefore, it was of interest to carry out a correlation analysis of logP and topological parameter. Correlations between the topological index GAI and logP for barbiturates have been revealed by this study.

The topological index GAI and logP of 25 barbiturates acid derivatives with structure I and II are listed in Table VI and Table VII, respectively.

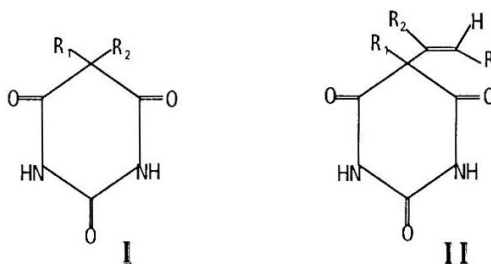


Table VI logP and Topological Index GAI for Barbiturates with Structure I

	R ₁	R ₂	GAI ⁻¹	N	log P	
					exptl.	calc.
1	methyl	1-methyl, 1-propenyl	0.0191904	5	0.65	0.65
2	ethyl	1-methyl, 1-propenyl	0.0203391	6	1.15	1.14
3	propyl	1-methyl, 1-propenyl	0.0215451	7	1.65	1.63
4	allyl	1-methyl, 1-propenyl	0.0228224	8	2.15	2.12
5	methyl	1-methylvinyl	0.0097828	4	0.15	0.14
6	ethyl	1-methylvinyl	0.0103628	5	0.65	0.63
7	propyl	1-methylvinyl	0.0109772	6	1.15	1.12
8	butyl	1-methylvinyl	0.0116279	7	1.65	1.61
9	iso-butyl	1-methylvinyl	0.0121162	7	1.45	1.61
10	amyl	1-methylvinyl	0.0123173	8	2.15	2.10
11	iso-amyl	1-methylvinyl	0.0128345	8	1.95	2.10

Table VII logP and Topological Index GAI for Barbiturates with Structure II

	R ₁	R ₂	R ₃	GAI ⁻¹	N	log P	
						exptl.	calc.
12	methyl	ethyl	methyl	0.0098871	6	1.15	1.12
13	ethyl	ethyl	methyl	0.0104733	7	1.65	1.61
14	propyl	ethyl	methyl	0.0110942	8	2.15	2.09
15	isopropyl	ethyl	methyl	0.0115601	8	1.95	2.09
16	methyl	methyl	ethyl	0.0098885	6	1.15	1.12
17	ethyl	methyl	ethyl	0.0104748	7	1.65	1.61
18	propyl	methyl	ethyl	0.0110958	8	2.15	2.09
19	isopropyl	methyl	ethyl	0.0115617	8	1.95	2.09
20	methyl	propyl	methyl	0.0104666	7	1.65	1.61
21	ethyl	propyl	methyl	0.0110871	8	2.15	2.09
22	methyl	isopropyl	methyl	0.0106244	7	1.45	1.61
23	methyl	butyl	methyl	0.0110870	8	2.15	2.09
24	propyl	butyl	methyl	0.0117443	9	2.65	2.58
25	ethyl	ethyl	propyl	0.0113273	9	2.65	2.58

The regression equation, correlation coefficient, F-test value, and standard deviation are

$$\log P = -1.8255 + 2.4771 \times \text{GAI}^{-1} + 0.4865 \times N$$

$$R=0.9917, \quad F=653, \quad S=0.08, \quad n=25$$

where N is the number of nonhydrogen atoms in the substituted part in a barbiturate skeleton.

The calculated logP values based on above equation are also shown in Table VI and Table VII, respectively.

ACKNOWLEDGMENT

The financial support of the National Natural Science of Foundation of China is gratefully acknowledged.

REFERENCES

- [1] Wiener.H., J. Am.Chem.Soc. 1947,69,17-20 .
- [2] Hosoya,H.,Bull.Chem.Soc.Jpn,1971,44,2332-2335.
- [3] Randic,M., J.Am.Chem.Soc.1975,97,6609-66155 .
- [4] Kier,L.B.;Hall,L.N. Molecular Connectivity in Structure-Activity Analysis;
Academic;NewYork,1986..
- [5] Balaban,A.T., Graphs. Pure Appl.Chem.1983,55,199.
- [6] Yang,J.A.;Jiang,Y,S.,Acta.Chim.Sinica.1983,55,884.
- [7] Xu,L.;Wang,H.Y.;Su,Q., Comput.Chem.1992,16,187.
- [8] Gutman.I.;Xu.L.; J.Serb.Chem.Soc.1992,57,11.
- [9] Xu,L., J.Serb.Chem.Soc.1992,57,485.
- [10] Basak,S.C.; Mousrud,L.J.;Frane,C.M.;Magnuson.V.R.A , Acta Pharm.Jugvsl.,1986,36,81.
- [11] Basak,S.C.;Harriss,D.K.;Magnuson,V.R., J.Pharm.Sci.1984,73,429.
- [12] Hansch,C.;Anderson,S.M.,J. Med. Chem.,1967,10,745.