

# The Energy of Chemical Graphs Related to Nanostructures in the Case of Alternate Polycyclic Conjugated Hydrocarbons and its computation

Dr. Rushina Natu

P.M.B. Gujarati Science College, Indore

**Abstract :** The structure factor represents the gross part of Energy which is identified with the help of graph theory. The total energy determined in the case of benzenoid structures is more than 99.5%. The present study is aimed to use well known. Topological index, Padmakar Ivan index for modeling of alternate polycyclic conjugated hydrocarbons.

**Keywords:** Nanostructures, Conjugated Hydrocarbons, Cioslowski's formula

## INTRODUCTION

The graph energy is a quantity closely related to total  $\pi$ -electron energy<sup>1-5</sup>. As usual, E is expressed in units of the carbon-carbon resonance integral  $\beta$ . The question of the dependence of E on the structure of the respective conjugated molecule was first addressed by Charles A. Coulson in 1940<sup>6</sup> and eventually because the subject of a large number of investigations. Although a complete solution of the problem is still not in sight, the structure factor represent the gross part of E are now identified. These are n, the number of vertices, and m, the number of edges of the molecular graph. In the case of Benzenoids hydrocarbons they determine somewhat more than 99.5% of E<sup>7, 8</sup>. The main factor responsible for the energy differences between \* benzenoid systems is recognized to be the number of Kekule structures (K). The seemingly simple question of how E depends on K resulted in three different and mutually contradicting approaches.

Based on the fact that  $E(G_1 \cup G_2) = E(G_1) + E(G_2)$  and  $K(G_1 \cup G_2) = K(G_1) \cdot K(G_2)$ , a logarithmic dependence of E on K was anticipated<sup>8-11</sup>. Here  $G_1 \cup G_2$  denotes the graph composed of disconnected components  $G_1$  and  $G_2$ . Based on certain assumptions about the distribution of the graph Eigen value Cioslowski at the following formula<sup>12</sup>:

$$E = \sqrt{2mn} F(2m/n)^{\frac{-1}{2}} K^{\frac{2}{n}} \quad (1)$$

Where F is a certain function whose actual form could not be determined<sup>13-16</sup>. It was shown<sup>17</sup> that Cioslowski's formula requires that the coefficients of the characteristic polynomial are not mutually correlated, a condition that seems not to be fulfilled in a series of isomeric benzenoid systems.

The observation that in case of benzenoid hydrocarbon E is a linear function of K was first made by Hall<sup>18</sup> and was later put in a following quantitative form:

$$E = A_n + B_m + CKD^{m-n} \quad (2)$$

where A, B, C and D are empirical parameters, determined by least-square fitting.

Irrespective of the form of the function F, equation (1) implies a non-linear dependence of E on K. This makes it possible to decide which of the approximations equation (1) and (2) is more appropriate. Empirical and statistical examination of the E values of series of isomeric benzenoid hydrocarbon revealed that the linear formula (2) is more suitable than any of the Cioslowski's type expressions. In spite of numerous efforts along these lines no satisfactory explanation for the success of (2) has been offered so far. The present study is aimed at filling this gap in that we have used the well-known topological index viz., Padmakar-Ivan (PI) index<sup>19</sup> for modelling \* of alternate polycyclic conjugated hydrocarbons.

## METHODOLOGY USED

Topological indices are numbers associated with molecular graphs for the purpose of allowing quantitative structure-activity / property / toxicity relationship. The first topological index viz., Wiener index, W, was introduced by Harold Wiener in 1947. Ivan Gutman defined cyclic version of Wiener index W, and assumed it Szeged index Sz. For trees  $W = Sz$ , but for cyclic graphs the two indices are different. Consequently, Padmakar Khadikar introduced a new topological index that he called the Padmakar-Ivan index, PI. This PI index is different for acyclic as well as cyclic graphs. Subsequently, Khadikar introduced yet another index called Sadhana index, Sd, which is applicable to cyclic graphs only. In addition, there are some other types of indices based on the concept of connectivity, chief among than being Randic \* and \* Hall valence \* indices. However, the methodology used by us is chiefly based on Wiener and Wiener type indices: W, Sz, PI and Sd. The above mentioned indices are calculated using the following expressions:

### (i) Wiener Index (W)

The Wiener index of a graph G is the half sum of distances over all its vertex pairs (i, j):

$$W = W(G) = \frac{1}{2} \sum_{(i,j)} d(i,j) \quad (3)$$

**(ii) Padmakar – Ivan Index (PI)**

The Padmakar – Ivan, PI index of a graph G is defined as:

$$PI = PI(G) = \sum [n_{eu}(e|G) + n_{ev}(e|G)] \quad (4)$$

where  $n_{eu}(e|G)$  is the number of edges of G lying closer to u than v,  $n_{ev}(e|G)$  is the number of edges of G lying closer to v than u and summation goes over all edges of G. The edges equidistant from both u and v are not taken into account for the calculation of PI index.

**(iii) Sadhana Index (Sd)**

The Sadhana index, Sd, for counting \* strips in G was defined by Khadikar as

$$Sd = Sd(G) = \sum_c m(G,c) (|E(G)| - c) \quad (5)$$

where  $m(G, c)$  is the number of strips of length c.

**(iv) Szeged Index (Sz)**

The Szeged index, Sz, of a graph G is defined as:

$$Sz = Sz(G) = \sum_{e \in E(G)} n_1(e|G) \cdot n_2(e|G) \quad (6)$$

where  $n_1$  and  $n_2$  are the number of vertices lying closer to the ends of an edge e. The vertices equidistant from both the ends of an edge are not counted for the calculation of Sz. The summation goes over all the edges of a graph G.

**RESULTS AND DISCUSSIONS**

The structural details and the parameters used are given in **Table 1**, while the variable selection for multiple regression are shown in **Table 2**. The **Table 3** records statistically significant models for modelling the energy. Ridge regression data are given in **Table 4** and **Table 5**. The comparison of calculated and observed energy is made in **Table 6**. The results are discussed below.

**One variable modeling of energy**

The details of one-variable modeling of E is given below:

Parameter used	R	R2	R2A	CV
W	0.9748	0.9430	0.0486	199.673
Sz	0.9320	0.9258	0.0554	150.789
PI	0.9696	0.9669	0.0370	351.351
Sd	0.8331	0.8179	0.0868	54.916
K	0.8727	0.8612	0.0758	75.439
CSC	0.4700	0.4218	0.1547	9.755
OMO	0.0228	0.0000	0.2101	0.357

**Comment on adjustable  $R^2$  ( $R^2_A$ )**

The adjustable- $R^2$  ( $R^2_A$ ) takes into account of advancement of  $R^2$  and is given by following expression:

$$R^2_A = 1 - (1 - R^2) (n - 1 / n - k - 1) \quad (18)$$

If a variable is added that does not contribute its fair share, then  $R^2_A$  will actually decline. This parameter  $R^2_A$  is particularly important when the number of independent variables is larger relative to the sample size.  $R^2$  may appear artificially high if the number of variables is high compared to the sample size. In fact,  $R^2$  will always increase when an independent variable is added, while  $R^2_A$  will decrease if the added variable does not reduce the unexplained variation enough to affect the loss of degrees of freedom.

**Problem of Co-linearity and Randic recommendations**

The problem of co-linearity can be resolved in two different ways:

(i) applying pure statistics and forgetting the possible physical significances of the parameter involved in the model

or

(ii) do not entirely depend on the statistics and use Randic recommendations.

The first approach uses the results obtained from (i) correlation matrix; (ii) Ridge statistics, (iii)  $\lambda$ -statistics.

**(i) Correlation matrix**

In order to investigate co-linearity problem in the proposed models we have to first obtained correlation matrix for the best model in modeling energy. Fortunately, we obtain a best model (**Table 3**) which contains *PI*, *HOMO* and *K* as the correlating parameters. This situation has an additional advantage that using such models containing common correlating parameters we can study relative potential of these indices in modeling the referred three activities. It is worthy to mention that the correlation matrix is very useful for determining which independent variables are likely to help explain variation in the dependent variables. Here we look the correlation close to  $\pm 1.0$  since that indicates changes in the independent variables are linearly related to changes in the dependent variables. We can also use correlation matrix to determine the extent to which independent variables are correlated with one another i.e. their inter-correlated ness or auto-correlation. This can be useful in determining if certain independent variables are redundant and not needed in the model. In practice every term in the correlation matrix  $> 0.4$  can be taken as being suspicious due to co linearity.

A perusal of Table shows the following:

(a) in case of equation (1) *W*,  $\chi$ , and *Sz* are highly correlated. Similarly *J*hetm is highly correlated with *J*hete. Thus, this model expressed by equation (1) suffers from co-linearity-defect;

(b) Correlation matrix involving quantum-theoretical descriptors indicates [equation (11)] that none of the

parameters used exhibit any co-linearity. That is, all the models using quantum-theoretical descriptors only will be free from co-linearity defect. Thus, the model expressed by equation again is free from such defect;

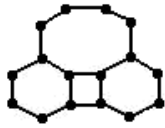
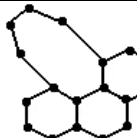
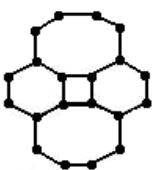
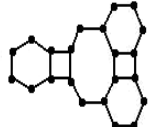
(c) Finally equation considered the combination of topological and quantum-theoretical descriptors its correlation matrix shows that topological indices  $W$ ,  ${}^1\chi$  and  $S_z$  are highly linearly correlated. Same is the case with  $J_{hetm}$  and  $J_{hete}$ . Thus, like equation (1) this model also suffers from co-linearity defect.

Finally, we will use Randic recommendation for making final conclusion.

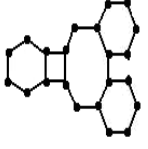
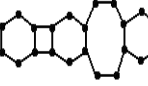
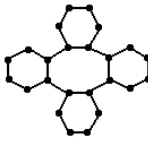
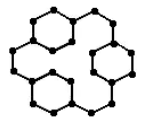
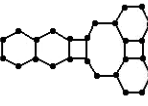
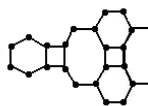
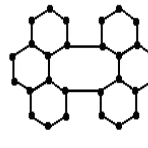
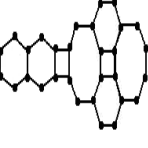
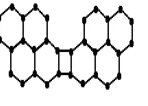
#### (ii) Randic recommendations

Randic stated that if a descriptor strongly correlates with another descriptor already used in a regression, such a descriptor in most studies should be discarded. For example  ${}^1\chi$  and  ${}^2\chi$ ,  ${}^1\chi$  often strongly correlate and in many structure-property-activity studies  ${}^2\chi$  has been discarded. This is not theoretically justified and despite the widespread practice should be stopped. Although two highly correlated descriptors overall depict the same features of molecular structure, it is important to recognize that even highly interrelated descriptors differ in some other structural traits. The difference between them may be relatively small but nevertheless very important for structure-property regression.

**Table 1. Structures and parameters.**

S. No.	Structure	K	C Sc	$E_\pi$	HO MO	W	K0	K1	K2	KV0	KV1	KV2	$S_z$	PI	S d
1		6	2	21.77 65	0.209 0	37 0	10.53 52	7.949 5	6.763 9	10.53 52	7.949 5	6.763 9	10 44	30 6	11 4
2		6	4	24.78 86	0.265 7	50 1	11.94 94	8.932 7	7.918 6	11.94 94	8.932 7	7.918 6	13 57	38 4	14 7
3		9	1	27.03 68	0.110 1	66 4	13.10 41	9.932 7	8.688 4	13.10 41	9.932 7	8.688 4	20 64	50 0	16 8
4		1 3	5	30.59 35	0.271 7	90 1	14.25 88	10.91 58	9.505 9	14.25 88	10.91 58	9.505 9	25 97	64 6	24 3

The criteria for inclusion or exclusion of descriptors should not be based on parallelism between descriptors even if overwhelming, but should be based on whether the part in which two descriptors disagree is or is not relevant for the characterization of the property considered. If the part in which the second descriptor differ from the first, regardless of how small it is, is relevant for the property under consideration, then the descriptor should be included. Randic [59,60] further stated that the selection of descriptors to be used in structure-property-activity studies should not be delegated solely to computers, although statistical criteria will continue to be useful for preliminary screening of descriptors taken from a large pool. Often in an automated selection of descriptors, a descriptor will be discarded because it is highly correlated with another descriptor already selected. But what is important is not whether two descriptors parallel one another; i.e., duplicates much of the same structural information, but whether they are complementary in those parts that are important for structure-property-activity correlations. Hence, the residual of the correlation between two descriptors should be examined and kept or discarded depending on how well it can improve the correlation based on already selected descriptors.

5		1 1	7	30.44 24	0.314 2	91 9	14.51 83	10.91 58	9.443 2	14.51 83	10.91 58	9.443 2	23 20	61 0	26 0
6		1 2	4	30.20 12	0.180 8	10 15	14.51 83	10.89 9	9.505 9	14.51 83	10.89 9	9.505 9	25 14	60 0	26 0
7		1 7	15	33.54 39	0.445 0	10 80	15.93 25	11.93 27	10.44 87	15.93 25	11.93 27	10.44 87	27 36	71 2	30 8
8		9	9	32.72 23	0.456 8	13 14	16.19 2	11.84 85	10.39 23	16.19 2	11.84 85	10.39 23	38 70	64 8	21 6
9		1 9	7	36.23 00	0.228 0	14 79	16.82 77	12.88 21	11.49 74	16.82 77	12.88 21	11.49 74	41 94	92 2	35 2
10		2 1	3	35.82 58	0.158 6	14 11	16.82 77	12.89 9	11.43 04	16.82 77	12.89 9	11.43 04	43 94	91 6	32 0
11		2 6	24	39.69 16	0.468 9	15 44	18.24 19	13.89 9	12.88 35	18.24 19	13.89 9	12.88 35	47 24	10 52	25 2
12		3 0	4	41.50 81	0.136 7	21 73	19.39 66	14.86 53	13.42 19	19.39 66	14.86 53	13.42 19	66 07	12 40	44 4
13		4 5	27	45.59 41	0.260 7	25 30	20.55 13	15.86 53	14.93 77	20.55 13	15.86 53	14.93 77	73 26	14 72	48 0

**Table 2. Variable selection regression report**

Code	Variable		Model Size	R <sup>2</sup>	Coded Variables	Model
	C5	E <sub>π</sub>				
A	C3	K	1	0.969643	F	<b>C15</b>
B	C4	CSC	2	0.985594	CF	<b>C6, C15</b>
C	C6	HOMO	3	0.993175	ACF	<b>C3, C6, C15</b>
D	C7	W	4	0.995170	ABFG	<b>C3, C4, C15, C16</b>
E	C14	Sz	5	0.995331	ABCDFG	<b>C3, C4, C6, C15, C16</b>
F	C15	PI	6	0.995391	ABCDFG	<b>C3, C4, C6, C7, C15, C16</b>
G	C16	Sd	7	0.995391	ABCDEFG	<b>C3, C4, C6, C7, C14, C15, C16</b>

**Table 3. Multiple regressions: Statistically significant models for modeling E<sub>π</sub>**

**Model 1: C15**

$$E_{\pi} = 17.9339 + 0.0197 (\pm 0.0010) \text{ PI}$$

$$R^2 = 0.9696, R^2A = 0.9696, CV = 0.0370, F = 351.351$$

**Model 2: C6, C15**

$$E_{\pi} = 16.0908 + 0.0196 (\pm 0.0008) \text{ PI} + 7.0149 (\pm 2.1081) \text{ HOMO}$$

$$R^2 = 0.9856, R^2A = 0.9827, CV = 0.0268, F = 342.078$$

**Model 3: C3, C6, C15**

$$E_{\pi} = 14.5133 + 0.0269 (\pm 0.0024) \text{ PI} - 0.2274 (\pm 0.0719) \text{ K} + 6.4303 (\pm 1.5406) \text{ HOMO}$$

$$R^2 = 0.9932, R^2A = 0.9909, CV = 0.0194, F = 436.571$$

**Model 4: C3, C4, C15, C16**

$$E_{\pi} = 14.8737 + 0.0279 (\pm 0.0025) \text{ PI} + 0.0096 (\pm 0.0045) \text{ Sd} - 0.4250 (\pm 0.0715) \text{ K} + 0.1641 (\pm 0.0327) \text{ CSC}$$

$$R^2 = 0.9952, R^2A = 0.9928, CV = 0.0173, F = 416.061$$

**Model 5: C3, C4, C6, C15, C16**

$$E_{\pi} = 14.6270 + 0.0269 (\pm 0.0033) \text{ PI} + 0.0088 (\pm 0.0051) \text{ Sd} - 0.3656 (\pm 0.1422) \text{ K} + 0.1187 (\pm 0.0985) \text{ CSC} + 2.1469 (\pm 4.3648) \text{ HOMO}$$

$$R^2 = 0.9953, R^2A = 0.9920, CV = 0.0182, F = 298.461$$

**Table 4. Ridge Regression Reports Correlation Matrix**

	C3( E )	C4( CSC )	C15 ( PI )	C16 ( Sd )	C5
C3	1.000000	0.712995	0.972673	0.891429	0.934207
C4	0.712995	1.000000	0.653412	0.477148	0.685568
C15	0.972673	0.653412	1.000000	0.924817	0.984704
C16	0.891429	0.477148	0.924817	1.000000	0.912754
C5	0.934207	0.685568	0.984704	0.912754	1.000000

**Table 5. Values of VIF, Tolerance, Eigen Values and Condition Number**

Independent Variable	VIF	Tolerance	igen Value	Condition Number
C3	23.3990	0.0427	3.345782	1.00
C4	2.7109	0.3689	0.564565	5.93
C15	26.5860	0.0376	0.067161	49.82
C16	8.7248	0.1146	0.022492	148.76

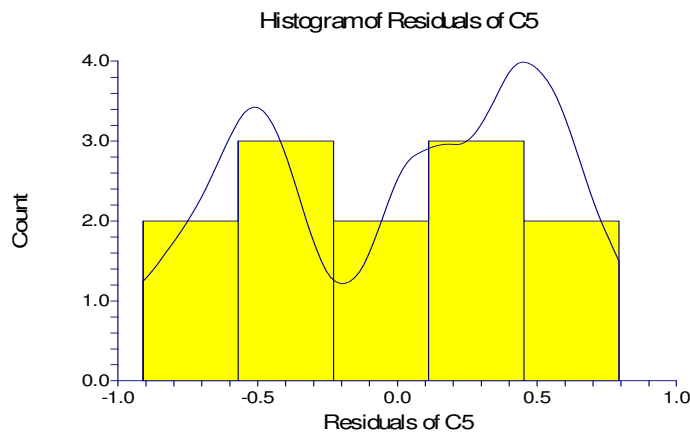
Since some VIF's are greater than 10, multi-collinearity is a problem.  
Some Condition Numbers greater than 100. Multi-collinearity is a MILD problem.

**Table 6. Correlation of actual (observed) and predicted (calculated)  $E_{\pi}$  values.**

	Actual	Predicted	Residue
1	21.777	22.283	-0.506
2	24.789	25.104	-0.315
3	27.037	26.775	0.262
4	30.594	30.524	0.070
5	30.442	30.860	-0.418
6	30.201	29.664	0.537
7	33.544	32.929	0.615
8	32.722	32.677	0.045
9	36.230	37.048	-0.818
10	35.826	35.068	0.758
11	39.692	39.533	0.159
12	41.508	41.636	-0.128
13	45.594	45.854	-0.260

## CONCLUSION

Padmakar Ivan index shows that linear formula is more suitable. The gaps are filled well the above index. The variation inflation factor plot predicts there is decline in C<sub>15</sub>, C<sub>16</sub> than is C<sub>4</sub> & C<sub>3</sub> model hence the predicted & actual values have small difference obtained as an residue. Hence shown in the graphs & in the values determined.



### Bibliography :

- [1]. Diudea, M.V.; Florescu, M.S.; Khadikar, P.V. **Molecular Topology and Its Applications**, EFICON, Bucharest, **2006**, 412 pp.
- [2]. C. Hansch, A. Leo, In Exploring QSAR : Fundamental and application in chemistry and biology, S.R. Heller, Ed. American Chemical Society: Washington, DC. (1995).
- [3]. M.Karleson, Molecular descriptors in QSAR/QSPR, Wiley-Interscience (2000).
- [4]. M.V. Diudea, M.S. Florescu, P.V. Khadikar, Molecular topology and its applications, EFICON Press: Bucharest. (2006)
- [5]. C. Hansch, B.R. Telzer, L. Zhang, Crit. Rev. Toxicol. **25**, 67-89. (1995).
- [6]. H. Branes, Mamalian toxicology - property prediction and QSAR technique, Peter Fisk Associates Whitestable : Kent. (2004)
- [7]. M.V. Diudea, (Ed) QSPR/QSAR Studies by Molecular Descriptors, Nova Science, (2000)
- [8]. J. Devillers, A.T. Balaban, (Eds) Topological Indices and Related Descriptors in QSAR and QSPR, Gordon & Breach Science Pub., Amsterdam (The Netherlands), (1999).
- [9]. M.V. Diudea, (Ed) QSPR/QSAR Studies by Molecular Descriptors, Nova Science (2000)
- [10]. D. Janezic, Nikolic, T.N. Sonja, **Graph theoretical matrices in chemistry**, University of Kragujevac, Faculty of Science, Kragujevac, 205 (2007).
- [11]. Wiener, H; J.Am. Chem. Soc., **69**. 17, (1947).
- [12]. Hosoya, H.;
- [13]. Khadikar, P.V.; Deshpande, N.V.; Kale, P.P.; Dabrynin, A.; Gutman, I.; Domotor, G. The Szeged Index and an Analogy with the Wiener Index, J.Chem. Inf. Comput. Sci., **1995**, 35, 545-550.
- [14]. Gutman, I.; Khadikar, P.V.; Rajput, P.V.; Karmarkar, S., The Szeged Index of Polyacenes, J.Serb. Chem. Soc., **1995**, 60, 759-764.
- [15]. Khadikar, P.V.; Karmarkar, S.; Joshi, S.; Gutman, I., Estimation of Protonation Constants of Salicylhydroxamic Acids by Means of the Wiener Topological Index, J.Serb. Chem. Soc., **1996**, 61, 89-95.
- [16]. Gutman, I.; Popovic, L.; Khadikar, P.V. Karmarkar, S.; Joshi, S.; Mandloi, M., Relations between Wiener and Szeged Indices of Monocyclic Molecules, Commun. Math. Comput. Chem. (MATCH). **1997**. 35, 91-103.
- [17]. Gutman, I.; Khadikar, P.V. Karmarkar, J.; Wiener and Szeged Indices of Benzenoid Hydrocarbons Containing a Linear Polyacene Fragment, Commun. Math. Comput. Chem. (MATCH). **1997**. 35, 1051-116.
- [18]. Gutman, I. Personal Communication, 1994.
- [19]. Khadikar, P.V., On a Novel Structural Descriptor PI, Nat. Acad. Sci. Lett. **2000**, 23, 113-118.
- [20]. Khadikar P.V.; Karmarkar, S.; Agrawal, V.K., PI Index of Polyacenes and its Use in Developing QSPR, Nat. Acad. Sci. Lett. **2000**, 23, 124-128.