MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **56** (2006) 131-139

ISSN 0340 - 6253

New Structural Parameters and Permanents of Adjacency Matrices of Fullerenes

Shan Jiang Heng Liang Fengshan Bai[*†]

Department of Mathematical Sciences,
Tsinghua University, Beijing, 100084, P.R.CHINA.

(Received November 25, 2004)

Abstract

The permanent of the adjacency matrix of a fullerene C_n is related to structural parameters involving the presence of contiguous pentagons p, q and r, so that the permanent of the adjacency matrix can be predicted through those structural parameters. This is more meaningful for n large, since the computation of permanent is proved to be a $\#P$-complete problem in counting. However structural parameters p, q and r become all zero for isolated-pentagon isomers when n is large (say $n \geq 60$). New structural parameters u, v and w are introduced in this paper, which count two or three contiguous hexagons. The stepwise regression model is used to predict the permanents of adjacency matrices of fullerenes. Computational results show that our new models are simple, reliable and efficient.

1 Introduction

A fullerene C_n is a polyhedral carbon cage with n atoms arranged in 12 pentagonal and $(n-20)/2$ hexagonal rings. It is well known that the stability of the fullerene C_n is related to the adjoining of pentagons and hexagons[1, 2, 3, 4]. For example, a structure in which a pentagon is completely surrounded by hexagons is believed to be more stable[1, 5, 6].

The adjacency matrix $A = (a_{i,j})$ for a graph with n vertexes is an $n \times n$ (0-1)-matrix. The value of each $a_{i,j}$ is taken to be 1, if there exists an edge between the vertexes i and j; and 0 otherwise. The permanent of the adjacency matrix and Kekulé structure count are both important and widely

[*]E-mail address: fbai@math.tsinghua.edu.cn

[†]Supported by National Science Foundation of China G10228101 and National Key Basic Research Special Fund G1998020306.

used metrics for fullerenes[7, 8]. They are strongly associated with their stability. Efficient algorithms are developed for computing the Kekulé structure count[3, 9, 10]. The Kekulé structure counts for all 1812 distinct fullerene isomers of C_{60} have been reported[2]. However, computing the permanent of the adjacency matrix is a hard task[11, 12, 13, 14].

The permanent of an $n \times n$ matrix $A = [a_{ij}]$ is defined as

$$per(A) = \sum_{\sigma \in \Lambda_n} \prod_{i=1}^{n} a_{i\sigma(i)},$$

where Λ_n denotes the set of all possible permutations of $\{1, 2, ..., n\}$. The definition of $per(A)$ looks similar to the determinant of matrix $det(A)$. We all know that $det(A)$ can be calculated in $O(n^3)$ time. However, there is no efficient algorithm on $per(A)$ for general matrices. The computation of permanent is proved to be a $\#P$-complete problem in counting [15], which is no easier than an NP-complete problem in combinatorial optimization. Therefore it is interesting to find the correlation between permanent and quantitative structure-property relationship(QSPR) through representative smaller fullerenes. Cash and Torrens give individual models that fit pentagonal parameters[13, 16, 17]. However, their results are restricted only to smaller fullerenes ($n < 60$), because all pentagonal parameters p, q and r vanish for isolated-pentagon isomers. In order to predicting large fullerenes ($n > 60$), hexagonal parameters should be taken into account.

In the next section, structural parameters are discussed and some new parameters are introduced. Computational methods used in this paper, are presented in section 3. Calculation results with new structural parameters are given in section 4, and some discussions are also made.

2 Structural Parameters

The structural features involving adjacent pentagons are encoded by the parameters p, q and r. Figure 2.1 shows the local structures that contribute to the values of p, q and r respectively.

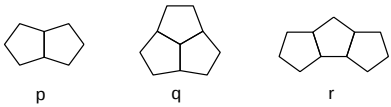

Figure2.1: Substructures that contribute to the p, q and r counts

The parameters p and q enumerate the number of edges common to two pentagons and the number of vertices common to three pentagons respectively[18]. The parameter r counts the number of pairs of nonadjacent pentagon edges shared with two other pentagons [13].

Cash uses a group of 27 fullerenes to correlate $\ln[per(A)]$ with structural parameters p, q and r [13]. After removing the outliers, a linear model is given by (1).

$$ln[per(A)] = 18.8884 - (0.5183 \pm 0.1572)p + (0.3482 \pm 0.1039)q - (0.0482 \pm 0.0361)r, \quad (1)$$

with $R = 0.909, s = 0.748$, where R is correlation coefficient, and s is the standard error of the estimation.

Torrens uses a group of 29 fullerenes to correlate $\ln[per(A)]$ with the structural parameters p, q, r [16, 17]. A linear model is obtained as (2).

$$ln[per(A)] = 20.2 - 0.660p + 0.383q, \quad (2)$$

with $R = 0.949, s = 0.757, F = 118.5$ and $MAPE = 4.05\%, AEV = 0.0988$. Here $MAPE$ is the mean absolute percentage error and AEV is the approximation error variance. By viewing the values of R and s, one can see that (2) is a better model, comparing with (1).

By considering interactive effects of structural parameters p, q and r [16, 17], a nonlinear model is given by Torrens as (3).

$$ln[per(A)] = 20.0 - 0.666p + 0.616q - 0.00850pq, \quad (3)$$

with $MAPE = 3.90\%; AEV = 0.0871$.

The results of Torrens improve that of Cash. However such an improvement is still limited, since the structural parameters p, q and r all vanish for isolated-pentagon isomers.

We consider the structure of adjacent hexagons and introduce three new structural parameters u, v and w accordingly. The u and v enumerate, respectively, the number of edges common to two hexagons and the number of vertices common to three hexagons. The variable w counts the number of pairs of nonadjacent hexagon edges shared with two other hexagons. Thus, v and w complement each other by counting the two possible arrangements of three contiguous hexagons.

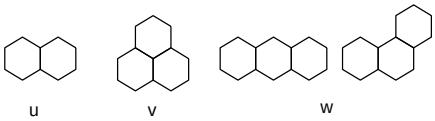

Figure2.2: Substructures that contribute to the u, v and w counts

Table 2.1: Values of Parameters and Counts for Fullerenes

Fullerene	K	$per(A)$	$\frac{\ln[per(A)]}{\ln(K)}$	p	q	r	u	v	w
$C_{20}(I_h)$	36	1392	2.0199	30	20	30	0	0	0
$C_{24}(D_{6d})$	54	4692	2.1192	24	12	36	0	0	0
$C_{26}(D_{3h})$	63	8553	2.1853	21	8	30	0	0	0
$C_{28}(T_d)$	75	15705	2.2378	18	4	24	0	0	0
$C_{28}(D_2)$	90	16196	2.1540	20	8	24	2	0	0
$C_{30}(C_{2v})I$	107	29621	2.2034	17	4	20	2	0	0
$C_{30}(C_{2v})II$	117	30053	2.1651	18	6	20	3	0	1
$C_{30}(D_{5h})$	151	31945	2.0672	20	10	20	5	0	5
$C_{32}(D_3)$	144	55140	2.1968	15	2	18	3	0	0
$C_{32}(C_2)I$	151	55705	2.1780	16	4	16	4	0	2
$C_{32}(C_2)II$	168	57092	2.1375	17	6	16	5	0	4
$C_{32}(D_2)$	184	58384	2.1045	18	8	15	6	0	6
$C_{34}(C_{3v})$	195	103665	2.1902	15	3	15	6	1	3
$C_{34}(C_s)$	196	104484	2.1896	15	3	16	6	1	2
$C_{34}(C_2)I$	204	103544	2.1714	14	2	14	5	0	3
$C_{34}(C_2)II$	212	107720	2.1632	17	6	16	8	2	5
$C_{36}(D_{6h})$	272	192528	2.1706	12	0	12	6	0	6
$C_{36}(D_{2d})$	288	192720	2.1489	12	0	12	6	0	4
$C_{36}(C_{2v})$	312	197340	2.1231	13	2	10	7	0	6
$C_{36}(D_{3h})$	364	207924	2.0764	15	6	6	9	0	12
$C_{38}(C_{2v})$	360	366820	2.1768	14	2	14	11	4	6
$C_{38}(C_{3v})$	378	363300	2.1572	12	1	9	9	1	6
$C_{38}(D_{3h})$	456	411768	2.1116	18	8	18	15	6	12
$C_{40}(D_{5d})I$	562	691092	2.1237	10	0	10	10	0	10
$C_{40}(T_d)$	576	704640	2.1185	12	4	0	12	0	18
$C_{40}(D_{5d})II$	701	803177	2.0750	20	10	20	20	10	30
$C_{44}(T)$	864	2478744	2.1775	12	4	0	18	4	24
$C_{44}(D_{3h})$	960	2436480	2.1416	9	2	0	15	0	24
$C_{60}(I_h)$	12500	395974320	2.0986	0	0	0	30	0	60
$C_{70}(D_{5h})$	52168	9193937544	2.1121	0	0	0	45	10	
$C_{80}(I_h)$	140625	189275868081	2.1906	0	0	0	60	20	
$C_{140}(I_h)$	2167239697			0	0	0	150	80	
$C_{180}(I_h)$	1389029765625			0	0	0	210	120	
$C_{240}(I_h)$	21587074966666816			0	0	0	300	180	

Values of p, q, r, u, v and w for some fullerenes are listed in Table 2.1. One can see clearly that p, q and r are all zero, as $n \geq 60$. The exact values of Kekulé structure count K and permanent $per(A)$ are also given, if possible, for later reference[19].

Note that chemists argue that the ratio

$$\frac{\ln[per(A)]}{\ln(K)}$$

should be bigger than, but not too far away from 2 for fullerenes and structurally similar systems[13, 14]. Hence those ratios are also included in Table 2.1.

3 Computational Method

Regression analysis is a statistical method for predicting values of one or more response (dependent) variables from a collection of predictor (independent) variable values. It can also be used for assessing the effects of the predictor variables on the responses.

The classical linear regression model: Let $x_1, x_2, \cdots, x_r$ be r predictor variables thought to be related to a response variable Y. The classical linear regression model states that Y is composed of a mean, which depends in a continuous manner on the x_i's, and a random error ϵ, which acounts for measurement error and the effects of other variables not explicitly considered in the model. The values of the predictor variables recorded from the experiment or set by the investigator are treated as *fixed*. The error (and hence the response) is viewed as a random variable whose behavior is characterized by a set of distributional assumptions.

Specifically, the linear regression model with a single response takes the form

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_r x_r + \epsilon$$

$$\{[response] = [mean(depending\ on\ x_1 \cdots] + [error]\}$$

The term 'linear' refers to the fact that the mean is a linear function of the unknown parameters $\beta_0, \beta_1 \cdots, \beta_r$. The predictor variables may or may not enter the model as first-order terms.

Stepwise regression: Stepwise selection begins with no candidate effects in the model and then systematically adds effects that are significantly associated with the target. However, after an effect is added to the model, stepwise may remove any effect already in the model that is not significantly associated with the target variable anymore.

In modeling it is essential to determine the complexity of the model to avoid overfitting. The predictive capability of the resulting model depends on the quality of the data(the more and better the data available, the more accurate prediction is possible) and on the number k of significant latent variables necessary. In stepwise selection several practical and reliable criteria are used for testing this significance,

such as Akaike's Information Criterion(AIC)[20, 21], Bayesian information criterion(BIC)[22] and cross-validation[23].

Based on the observations of the preceding section an information criterion AIC of θ is defined by

$$AIC(\hat{\theta}) = (-2)log(maximun\ likelihood) + 2k$$

where k is the number of independently adjusted parameters to get $\hat{\theta}$.

An asymptotic approximation of the integrated likelihood, valid under regularity conditions, has been proposed by Schwarz[22]

$$log f(x|m, K) \approx log(x|m, K, \hat{\theta}) - \frac{v_{m,K}}{2} log(n)$$

where $\hat{\theta} = \arg\max_\theta f(x|m, K, \theta)$ is the ML estimate of θ, and $v_{m,K}$ is the number of free parameters in the model m with K components. It leads to minimize the so-called BIC criterion

$$BIC = (-2)L_{m,K} + v_{m,K} \ln n$$

where $L_{m,K} = log(x|m, K, \hat{\theta})$ is the maximum log-likelihood for m and K.

4 Calculation Results

Using the stepwise regression, the best correlation of $\ln[per(A)]$ with parameters p, q, r, u, v and w is obtained as (4).

$$ln[per(A)] = 13.27932 - 0.20102p + 0.21662u, \tag{4}$$

with $n = 29, R = 1.0000, s = 0.0163778055$ and $F = 283000$. The quantity R gives the proportion of the total variation in the y's "explained" by, or attributable to, the predictor variables p, q, r, u, v and w. Besides the great improvement in R, the standard error of the estimate(s) for (4) is much less than those for (1)-(3). Furthermore that R is 1.0000 demonstrates that these two parameters (p and u) can almost give full information of $\ln[per(A)]$, as shown by Figure 4.1.

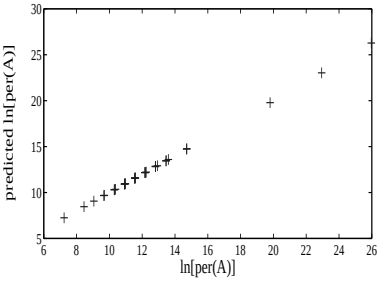

Figure4.1: The relations between $\ln per(A)$ and predicted $\ln per(A)$

Note that the same equation (4) is obtained by using the three different methods(AIC, BIC, Cross-validation) for testing its significance. It gives $AIC = -235.652814$, $BIC = -233.1934$ and $MAPE = 0.09\%$; $AEV = 0.00025$. Furthermore the standard errors of the estimations of the intercept, the parameter p and u are 0.01612, 0.00078453 and 0.00061269 respectively, which are all very small. From the procedure of the stepwise regression, the contribution of u is the most important factor, and that of p is the next.

Computational results of the estimated permanents, in particular for some large fullerenes, are given in Table 4.1.

Table 4.1: Estimation values of permanent for fullerenes

Fullerene	$\ln per(A)$			
	linear fit	exact	ref.13	ref.16
$C_{32}(D_{3d})$	10.9607	10.9794		
$C_{44}(C_2)$	14.8287	14.8065		
$C_{50}(D_{3h})$	16.8094	16.8969		
$C_{70}(D_{5h})$	23.0271	22.9418	18.8884	20.2000
$C_{80}(I_h)$	26.2764	25.9665	18.8884	20.2000
$C_{140}(I_h)$	45.7723			
$C_{180}(I_h)$	58.7695			
$C_{240}(I_h)$	78.2653			

linear fit: computed with new parameters;

exact: the exact values that computed with the method in [19];

ref.13: estimations by Cash [13];

ref.16: estimations by Torrens [16].

Comparing results in Table 4.1, one can see that our model with new parameters is simple, reliable and efficient at least for cases of $n \leq 80$. When n becomes even larger, it is too hard to calculate the exact values of permanents of adjacency matrices of fullerenes.

Table 4.2: Estimation values of the ratio for fullerenes

Fullerene	$\ln[per(A)]/\ln K$			
	exact	ref.13	ref.16	estimation
$C_{70}(D_{5h})$	2.1121	1.6895	2.0868	2.1199
$C_{80}(I_h)$	2.1906	1.6895	2.0868	2.2167
$C_{140}(I_h)$				2.1293
$C_{180}(I_h)$				2.1019
$C_{240}(I_h)$				2.0809

exact: computed by taking quotient of exact values of $\ln per(A)$ from Table 4.1 and that of K from Table 2.1;

ref.13: estimations by Cash [13];

ref.16: estimations by Torrens [16];

estimation: computed by taking quotient of estimated values of $\ln per(A)$ from Table 4.1 and exact values of K.

Results in Table 4.2 give the ratio of estimated values of $\ln[per(A)]$ and the exact values of $\ln(K)$. Those give very strong evidences, which show that our model with new parameters are quite promising for large fullerenes.

Considering the structure of adjacent hexagons, three new parameters u, v and w are introduced in this paper. A simple linear correlation is proved to be a good model for the permanent of the adjacency matrix of fullerene, and q,r,v,w are redundant information. Thus p and u contain the essential characters of the permanent for fullerene structure. The method proposed in this paper allows rapid estimation of $per(A)$ for large enough fullerenes.

References

[1] Kroto, H. W. The stability of the fullerenes C_n, with n=24,28,32,36,50,60 and 70. *Nature (London)* **1987**, *329*, 529-531.

[2] Austin, S. J.; Fowler, P. W.; Hansen, P.; Manolopoulos, D. E.; Zheng, M. Fullerene isomers of C_{60}. Kekulé counts versus stability. *Chem. Phys. Lett.* **1994**, *228*, 478-484.

[3] Schmalz, T. G.; Seitz, W. A.; Klein, D. J.; Hite, G. E. C_{60} carbon cages. *Chem. Phys. Lett.* **1986**, *130*, 203-207.

[4] Schmalz, T. G.; Seitz, W. A.; Klein, D. J.; Hite, G. E. Elemental carbon cages. *J. Amer. Chem. Soc.* **1988**, *110*, 1113-1127.

[5] Barth, W. E.; Lawton, R. G. Synthesis of corannulene. *J. Amer. Chem. Soc.* **1971**, *93*, 1730-1745.

[6] Manolopoulos, D. E. Comment on "Favourable structures for higher fullerenes". *Chem. Phys. Lett.* **1992**, *192*, 330.

[7] Trinajstić, N. *Chemical graph theory;* 2nd ed.; CRC Press: Boca Raton, FL, 1992.

[8] Cvetković, D. M.; Gutman, I.; Trinajstić, N. Kekulé structures and topology. *Chem. Phys. Lett.* **1972**, *16*, 614-616.

[9] Kasteleyn, P. W. Graph theory and crystal physics, In *Graph Theory and Theoretical Physics*; Harary, F., Ed.; Academic Press, New York, 1967; Chapter 2.

[10] Klein, D. J.; Liu, X. Many-body conjugated-circuit computations. *J. Comput. Chem.* **1991**, *12*, 1260-1264.

[11] Cash, G. G. A fast computer algorithm for finding the permanent of adjacency matrices. *J. Math. Chem.* **1995**, *18*, 115-119.

[12] Cash, G. G. A simple means of computing the Kekule structure count for toroidal polyhex fullerenes. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 58-61.

[13] Cash.G.G., Permanents of adjacency matrices of fullerenes. *Polycyclic Arom. Compd.* **1997**, *12*, 61-69.

[14] Gutman, I., Permanents of adjacency matrices and their dependence on molecular structure. *Polycyclic Arom. Compd.* **1998**, *12*, 281-287.

[15] Valliant, L. The complexity of computing the permanent. *Theoret. Comput. Sci.* **1979**, *8*, 189-201.

[16] Torrens, F. Computing the permanent of the adjacency matrix for fullerenes. *Internet Electron. J. Mol. Des.* **2002**, *1*, 351-359.

[17] Torrens, F. Table of periodic properties of fullerenes based on structural parameters. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 60-67.

[18] Liu, X.; Klein, D. J.; Schmalz, T. G.; Seitz, W. A. Generation of carbon-cage polyhedra. *J. Comput. Chem.* **1991**, *12*, 1252-1259.

[19] Liang, H.; Bai, F. A partially structure-preserving algorihm for the permanents of adjacency matrices of fullerenes. *Comput. Phys. Commun.* **2004**, *163*, 79-84.

[20] Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **1974**, *19*, 716-723.

[21] Akaike, H. Stochastic theory of minimal realization. *IEEE Trans. Automat. Contr.* **1974**, *19*, 667-6743.

[22] Schwartz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *5*, 461-464.

[23] Besalú, E. Fast computation of cross-validated properties in full linear leave-many-out procedures. *J. Math. Chem.* **2001**, *29*, 191-203.