Journal of Mathematical Chemistry 6(1991)41–50

GENERALIZATION OF CYVIN'S FORMULA

I. GUTMAN*
Department of Chemistry, Vanderbilt University, Nashville, TN 37325, USA

and

D.J. KLEIN
Department of Marine Sciences, Texas A&M University at Galveston, Galveston, TX 77553-1675, USA

Received 1 February 1990

Abstract

A formula, recently discovered by Cyvin, relates the numbers of Kekulé structures of unbranched catacondensed benzenoid molecules. It is shown that Cyvin's formula is a special case of a much more general result. The determinants occurring in Cyvin's formula as well as in our generalized version thereof can be interpreted in terms of correlation functions for the choices of pairs of double or pairs of single bonds in the Kekulé structures of the respective conjugated system.

1. Introduction

Some time ago, Cyvin [1] reported an identity for the number of Kekulé structures of unbranched catacondensed benzenoid hydrocarbons. This formula was then applied [1–5] in the study of the Kekulé structures of benzenoid and coronoid molecules.

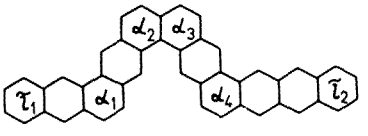
In order to formulate Cyvin's identity, denote by U_0 an unbranched catacondensed benzenoid system [6] possessing k kinks, $k \geq 2$. Label the kinks of U_0 consecutively by $\alpha_1, \alpha_2, \ldots, \alpha_k$ and the terminal hexagons closest to α_1 and α_k by τ_1 and τ_2, respectively. Then, U_1 will denote the fragment obtained by deleting from U_0 the first linear segment, i.e. τ_1, α_1 and the hexagons lying between them. Further, U_2 is the fragment obtained by deleting from U_0 the last linear segment, i.e. α_k, τ_2 and the hexagons lying between them. Finally, let U_3 be the fragment obtained by deleting from U_0 both its first and its last linear segment.

*Permanent address: Faculty of Science, University of Kragujevac, P.O. Box 60, YU-34000 Kragujevac, Yugoslavia.

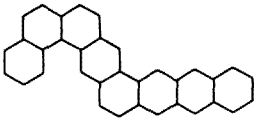
Then, Cyvin's formula reads [1]

$$\begin{vmatrix} K\{U_0\} & K\{U_1\} \\ K\{U_2\} & K\{U_3\} \end{vmatrix} = -(-1)^k, \tag{1}$$

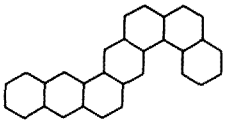
where $K\{G\}$ stands for the number of Kekulé structures of G. Formula (1) is illustrated by the example below:



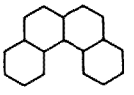
U_0; $k = 4$; $K\{U_0\} = 118$



U_1; $K\{U_1\} = 35$



U_2; $K\{U_2\} = 27$



U_3; $K\{U_3\} = 8$

Indeed,

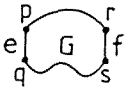
$$\begin{vmatrix} 118 & 35 \\ 27 & 8 \end{vmatrix} = -1 = -(-1)^4.$$

Although the aim of the present work is to provide a proof of Cyvin's identity and a far-reaching generalization thereof, we wish to point out its importance in the calculation of the number of Kekulé structures. By means of formula (1), Cyvin was able to obtain combinatorial expressions for the K-values of a variety of benzenoid (both cata- and pericondensed) and coronoid molecules [1–5] which could hardly be deduced without the use of (1). The importance of the Kekulé structure count in the theory of benzenoid and coronoid molecules need not be emphasized here [6].

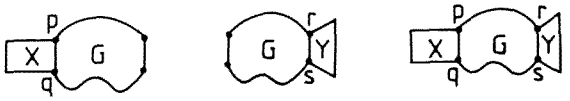
In the present paper, we show that eq. (1) can be considered as a special case of a much more general result which holds for the numbers of Kekulé structures of conjugated molecules whose molecular graphs have the form $X:G:Y$, where the "colon" symbol : indicates that the two graphs it separates are to be adjoined (or fused) along a single edge. More precisely, we deduce statements about the number of perfect matchings of any graph of the type $X:G:Y$ as well as statements which hold for any graph. (Recall that if H is a graph, then $K\{H\}$ is the number of perfect matchings of

H. A perfect matching of *H* is a set of independent, i.e. mutually non-incident edges of *H*, which cover all vertices of *H* [7].)

Let *G* be a graph possessing at least two independent edges (and thus at least four vertices). Denote these edges by *e* and *f* and the corresponding endpoints by *p*, *q* and *r*, *s*, respectively.



Let *X* be another graph with an edge *e'*. Then, *X*:*G* is the graph obtained by identifying the edge *e'* of *X* with the edge *e* of *G*. The graphs *G*:*Y* and *X*:*G*:*Y* are defined analogously:



X:G G:Y X:G:Y

It is clear that the unbranched catacondensed system U_0 is of the form $X:G:Y$, where *X* and *Y* are linear polyacene fragments (the first and the last linear segment) and *G* is another unbranched catacondensed benzenoid system. Then, of course, $U_1 = G:Y$, $U_2 = X:G$ and $U_3 = G$. Hence, the determinant on the l.h.s. of (1) is a special case of a topological function $D(G, X, Y)$, defined as

$$D(G,X,Y) = \begin{vmatrix} K\{X:G:Y\} & K\{G:Y\} \\ K\{X:G\} & K\{G\} \end{vmatrix}, \tag{2a}$$

i.e.

$$D(G,X,Y) = K\{X:G:Y\}K\{G\} - K\{X:G\}K\{G:Y\}. \tag{2b}$$

In the subsequent section, we establish some general properties of this function. Before doing this, however, we wish to emphasize once again that our discussion applies to all (molecular) graphs, both cata- and pericondensed benzenoids and nonbenzenoids. As a matter of fact, the structures of the graphs *G*, *X* and *Y* are fully arbitrary with the only restriction that *G* must possess two independent edges.

2. Identities for $D(G, X, Y)$

THEOREM 1

If *G*, *X* and *Y* are graphs with even numbers of vertices, then the function $D(G, X, Y)$, defined via (2), satisfies the relation

$$D(G,X,Y) = K\{X-e'\}K\{Y-f'\}\begin{vmatrix} K\{G\} & K\{G-e\} \\ K\{G-f\} & K\{G-e-f\} \end{vmatrix}. \tag{3}$$

Proof

Let A and B be graphs with an even number of vertices. Let a be an edge of A and b be an edge of B. Denote by $A:B$ the graph obtained by identifying the edges a and b. It has been determined elsewhere [8,9] that

$$K\{A:B\} = K\{A\}K\{B\} - K\{A-a\}K\{B-b\}. \tag{4}$$

Applying (4) to $X:G$, $G:Y$ and $X:G:Y$, we obtain in a straightforward manner:

$$K\{X:G\} = K\{X\}K\{G\} - K\{X-e'\}K\{G-e\},$$

$$K\{G:Y\} = K\{Y\}K\{G\} - K\{Y-f'\}K\{G-f\},$$

$$K\{X:G:Y\} = K\{X\}K\{Y\}K\{G\} - K\{X-e'\}K\{Y\}K\{G-e\}$$
$$- K\{Y\}K\{Y-f'\}K\{G-f\} + K\{X-e'\}K\{Y-f'\}K\{G-e-f\}.$$

Substituting these three equalities into the r.h.s. of (3), we arrive at the r.h.s. of (2) after appropriate algebraic manipulations. □

In a trivial manner, theorem 1 holds also if

(i) either X or Y have odd numbers of vertices, or

(ii) G has an odd number of vertices and both X and Y have even numbers of vertices.

Then, namely, both sides of (3) are equal to zero due to the fact that graphs with odd numbers of vertices cannot have perfect matchings. If, on the other hand,

(iii) both X and Y have odd numbers of vertices,

then relation (3) is not obeyed. This is because (4) is not applicable in such a case.

THEOREM 2

Let G be a graph and e and f its two independent edges. The endpoints of e are the vertices p and q; the endpoints of f are the vertices r and s. Then,

$$\begin{vmatrix} K\{G\} & K\{G-e\} \\ K\{G-f\} & K\{G-e-f\} \end{vmatrix} = \begin{vmatrix} K\{G\} & K\{G-p-q\} \\ K\{G-r-s\} & K\{G-p-q-r-s\} \end{vmatrix}. \tag{5}$$

Proof

If G does not possess perfect matchings, then all K-numbers appearing in (5) are equal to zero and thus (5) is obeyed in a trivial manner.

If $K\{G\} \neq 0$, then it is easy to verify (5) by direct calculation, using the facts that

$$K\{G\} = K\{G-e\} + K\{G-p-q\} = K\{G-f\} + K\{G-r-s\}, \tag{6a}$$

$$K\{G-e-f\} - K\{G-p-q-r-s\} = K\{G-e\} - K\{G-r-s\}$$
$$= K\{G-f\} - K\{G-p-q\}. \tag{6b}$$

□

COROLLARY 2.1

$$D(G,X,Y) = K\{X-e'\}K\{Y-f'\}\begin{vmatrix} K\{G\} & K\{G-p-q\} \\ K\{G-r-s\} & K\{G-p-q-r-s\} \end{vmatrix}.$$

It is possible to give an alternative interpretation and proof of theorem 2.

Let G be a graph possessing perfect matchings and let e and f be edges of G. Define a variable quantity ε, such that $\varepsilon = 1$ if e belongs to the perfect matching and $\varepsilon = 0$ if e does not belong to the perfect matching. Let another variable φ be associated with the edge f in a fully analogous manner: $\varphi = 1$ if f belongs to the perfect matching and $\varphi = 0$ otherwise. The expectation values of ε and φ are denoted by $\langle\varepsilon\rangle$ and $\langle\varphi\rangle$, respectively. Evidently, $\langle\varepsilon\rangle = K\{G-p-q\}/K\{G\}$, $\langle\varphi\rangle = K\{G-r-s\}/K\{G\}$. These expectation values measure the probability that e and f belong to a perfect matching of G.

The probability that e and f simultaneously belong to a perfect matching of G is given by $\langle\varepsilon\varphi\rangle = K\{G-p-q-r-s\}/K\{G\}$. The respective correlation function is then

$$\Gamma(e,f) = \langle(\varepsilon - \langle\varepsilon\rangle)(\varphi - \langle\varphi\rangle)\rangle.$$

It follows immediately that

$$\Gamma(e,f) = \langle\varepsilon\varphi\rangle - \langle\varepsilon\rangle\langle\varphi\rangle,$$

and therefore

$$\Gamma(e,f) = K\{G\}^{-2}\begin{vmatrix} K\{G\} & K\{G-p-q\} \\ K\{G-r-s\} & K\{G-p-q-r-s\} \end{vmatrix}.$$

Clearly, $\Gamma(e,f)$ measures the influence which the presence of the edge e in a perfect matching of G has on the presence of the edge f in the same perfect matching (or vice versa).

Denote by $\Gamma(\bar{e},\bar{f})$ the correlation function measuring the influence of the absence of the edge e from a perfect matching of G on the absence of the edge f from the same perfect matching. Then,

$$\Gamma(\bar{e},\bar{f}) = \langle(\bar{\varepsilon}-\langle\bar{\varepsilon}\rangle)(\bar{\varphi}-\langle\bar{\varphi}\rangle)\rangle.$$

On the other hand, $\langle\bar{\varepsilon}\rangle = K\{G-e\}/K\{G\}$, $\langle\bar{\varphi}\rangle = K\{G-f\}/K\{G\}$ and $\langle\bar{\varepsilon}\bar{\varphi}\rangle = K\{G-e-f\}/K\{G\}$. From this, it follows that

$$\Gamma(\bar{e},\bar{f}) = K\{G\}^{-2}\begin{vmatrix} K\{G\} & K\{G-e\} \\ k\{G-f\} & K\{G-e-f\} \end{vmatrix}.$$

Thence, theorem 2 is seen to be equivalent to

$$\Gamma(e,f) = \Gamma(\bar{e},\bar{f}). \tag{7}$$

This fact, on the other hand, is readily proved using the substitution $\bar{\varepsilon} = 1 - \varepsilon$ and $\bar{\varphi} = 1 - \varphi$.

Relation (7) may easily be generalized. For three disjoint edges e,f,g with associated variables $\varepsilon, \varphi, \gamma$, one readily has

$$\langle(\varepsilon-\langle\varepsilon\rangle)(\varphi-\langle\varphi\rangle)(\gamma-\langle\gamma\rangle)\rangle = -\langle(\bar{\varepsilon}-\langle\bar{\varepsilon}\rangle)(\bar{\varphi}-\langle\bar{\varphi}\rangle)(\bar{\gamma}-\langle\bar{\gamma}\rangle)\rangle.$$

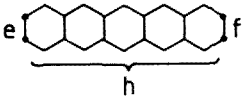
This, of course, relates various Kekulé structure counts. Further generalization to an arbitrary number of disjoint edges is possible, as shown in appendix 1. A third possible direction in which the present results can be extended is put forward in appendix 2.

3. Cyvin's formula follows from theorem 1

In order to deduce Cyvin's formula, we first point out two special cases of theorem 1.

COROLLARY 1.1

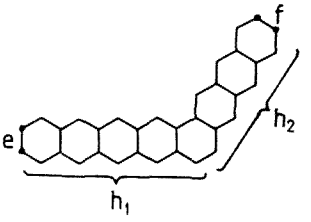
Let G in formula (3) be the linear polyacene with h hexagons:



Then, $K\{G\} = h + 1$, $K\{G - e\} = K\{G - f\} = h$, $K\{G - e - f\} = h - 1$ and the r.h.s. of (3) is equal to -1, irrespective of the value of the parameter h.

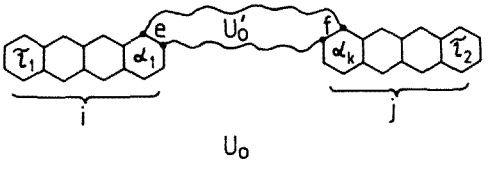
COROLLARY 1.2

Let G in formula (3) be the phene with $h_1 + h_2 - 1$ hexagons:



Then, $K\{G\} = h_1 h_2 + 1$, $K\{G - e\} = (h_1 - 1)h_2 + 1$, $K\{G - f\} = h_1(h_2 - 1) + 1$, $K\{G - e - f\} = (h_1 - 1)(h_2 - 1) + 1$ and the r.h.s. of (3) is equal to $+1$, irrespective of the values of the parameters h_1 and h_2.

Every unbranched catacondensed molecular graph U_0 with at least two kinks can be presented in the form $L_i : U_0' : L_j$, where U_0' is another unbranched catacondensed system and L_i, L_j denote linear polyacene fragments with i and j hexagons, $i, j \geq 2$:



Applying theorem 1 to $L_i : U_0' : L_j$, we obtain

$$\begin{vmatrix} K\{U_0\} & K\{U_1\} \\ K\{U_2\} & K\{U_3\} \end{vmatrix} = \begin{vmatrix} K\{U_0'\} & K\{U_1'\} \\ K\{U_2'\} & K\{U_3'\} \end{vmatrix}, \tag{8}$$

because $K\{L_i - e\} = K\{L_j - f\} = 1$. In eq. (8), the symbols U_1', U_2', U_3' are related to U_0' in the same manner as U_1, U_2, U_3 are related to U_0.

Thus, the determinant on the l.h.s. of (1) is equal to an analogous determinant for a smaller unbranched catacondensed system U_0'. It is easy to see that U_0' has two kinks less than U_0.

We now have to distinguish between three cases.

CASE 1

U_0' has no kinks, i.e. U_0' is a linear polyacene. Then k, the number of kinks in U_0, is even. By corollary 1.1, the r.h.s. of (8) is then equal to -1. Therefore, identity (1) holds.

CASE 2

U_0' has one kink, i.e. U_0' is a phene. Then k is an odd number. By corollary 1.2, the r.h.s. of (8) is now equal to $+1$. Again, identity (1) holds.

CASE 3

U_0' has more than one kink. Then the above described reduction procedure (which decreases the number of kinks by two) has to be repeated until we obtain an unbranched catacondensed system with less than two kinks. Then, either case 1 or case 2 will be applicable.

This completes the deduction of Cyvin's formula (1) from theorem 1.

Appendix 1

Theorem 1 is, in fact, a special case of a somewhat more general formula.

In theorem 1, it was assumed that the graph G has two independent edges e and f. Suppose now that G possesses $2n$ independent edges $e_1, e_2, \ldots, e_n, f_1, f_2, \ldots, f_n$. (Note that the graph G may possess other independent edges as well.)

Let $X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_n$ be $2n$ distinct graphs, each possessing at least one edge but otherwise being arbitrary. Let e_i' and f_i' denote an edge of X_i and Y_i, respectively, $i = 1, \ldots, n$.

Denote by $\mathcal{X}$ and $\mathcal{Y}$ the n-element sets $\{X_1, X_2, \ldots, X_n\}$ and $\{Y_1, Y_2, \ldots, Y_n\}$, respectively. Denote by $\mathcal{X}_i$ and $\mathcal{Y}_i$, $i = 1, 2, \ldots, 2^n$, the subsets of $\mathcal{X}$ and $\mathcal{Y}$, respectively, and label them so that $X_j \in \mathcal{X}_i \Leftrightarrow Y_j \in \mathcal{Y}_i$. It is convenient to choose $\mathcal{X}_1 = \mathcal{X}$ and $\mathcal{Y}_1 = \mathcal{Y}$.

The graph $\mathcal{X}_i : G : \mathcal{Y}_j$ is constructed in the following way. If $X_k \in \mathcal{X}_i$, then the edge e_k' of X_k is identified with the edge e_k of G. If $Y_k \in \mathcal{Y}_j$, then the edge f_k' of Y_k is identified with the edge f_k of G. This procedure is repeated for all elements of $\mathcal{X}_i$ and $\mathcal{Y}_j$. In the special case when both $\mathcal{X}_i$ and $\mathcal{Y}_j$ are empty sets, $\mathcal{X}_i : G : \mathcal{Y}_j = G$.

Denote by $\mathcal{E}$ and $\mathcal{F}$ the n-element sets $\{e_1, e_2, \ldots, e_n\}$ and $\{f_1, f_2, \ldots, f_n\}$, respectively. Let $\mathcal{E}_i$ and $\mathcal{F}_i$, $i = 1, 2, \ldots, 2^n$, be the subsets of $\mathcal{E}$ and $\mathcal{F}$, respectively, and let them be labelled so that $e_i \in \mathcal{E}_j \Leftrightarrow X_i \notin \mathcal{X}_j$, $f_i \in \mathcal{F}_j \Leftrightarrow Y_i \notin \mathcal{Y}_j$. With this convention, we have $\mathcal{E}_1 = \mathcal{F}_1 = \varnothing$. Then the following identity is obeyed for all values of n:

$$\det ||K\{\mathcal{X}_i : G : \mathcal{Y}_j\}|| = \prod_{k=1}^{n} K\{X_k - e_k'\} K\{Y_k - f_k'\} \det ||K\{G - \mathcal{E}_i - \mathcal{F}_j\}||. \quad (9)$$

Note that the matrices which appear in (9) are of order 2^n. Theorem 1 is just the special case of (9) for $n = 1$.

Formula (9) can be deduced in an analogous manner as theorem 1, namely by a repeated application of eq. (4). The actual procedure is quite lengthy and will not be presented here.

Also, theorem 2 can be generalized. In order to do this, denote by $\mathcal{E}_i^*$ and $\mathcal{F}_i^*$ the sets of vertices of G which are the endpoints of the elements of $\mathcal{E}_i$ and $\mathcal{F}_i$, respectively. Then the following relation holds:

$$\det \| K\{G - \mathcal{E}_i - \mathcal{F}_j\} \| = \det \| K\{G - \mathcal{E}_i^* - \mathcal{F}_j^*\} \|. \tag{10}$$

Evidently, for $n = 1$, eq. (10) reduces to (5).

The proof of (10) is elementary: The matrix $\| K\{G - \mathcal{E}_i - \mathcal{F}_j\} \|$ is transformed into the matrix $\| K\{G - \mathcal{E}_i^* - \mathcal{F}_j^*\} \|$ by means of operations which maintain the value of the determinant (subtraction of one row from another, etc.), and by appropriate application of relations of the type (6).

Appendix 2

From the discussion in appendix 1, it is clear that both the identities (5) and (10) are stipulated by the existence of a recursion relation of the type (6). This means that for the validity of relations of the type (5) or (10), it is sufficient that a certain topological function $T\{G\}$ conforms to recurrence relations analogous to (6), viz.,

$$T\{G\} = T\{G - e\} + T\{G - p - q\} = T\{G - f\} + T\{G - r - s\},$$

$$T\{G - e - f\} - T\{G - p - q - r - s\} = T\{G - e\} - T\{G - r - s\}$$
$$= T\{G - f\} - T\{G - p - q\}.$$

One such topological function is the matching polynomial [10], if it is defined as [11]

$$\alpha^+\{G\} = \sum_{k=0}^{[n/2]} m(G,k) x^{n-2k},$$

where $m(G, k)$ stands for the number of k-matchings of the graph G. Therefore, formulas (5) and (10) remain valid if each symbol K is interchanged by a symbol α^+. On the other hand, after such a substitution, formulas (1) and (3) are no longer true. This is because there exists no relation for the matching polynomial, analogous to (4), i.e. in general [12]

$$\alpha^+\{A : B\} \neq \alpha^+\{A\}\alpha^+\{B\} - \alpha^+\{A - a\}\alpha^+\{B - b\}.$$

Recall that for $x = 0$, $\alpha^+\{G\}$ is equal to $K\{G\}$. Note also that for $x = 1$, $\alpha^+\{G\}$ is the Hosoya index [13], i.e. the total number of matchings of the graph G. Hence, results analogous to (5) and (10) hold for the Hosoya index as well.

As a final remark, we mention that formulas (5) and (10) remain valid also if the symbols K are interchanged into α^-, where $\alpha^-\{G\}$ is the usual form of the matching polynomial [10,11]:

$$\alpha^-\{G\} = \sum_{k=0}^{[n/2]} (-1)^k m(G,k) x^{n-2k}.$$

Acknowledgement

One of the authors (I.G.) gratefully acknowledges financial support from the Fulbright Program. The other author (D.J.K.) acknowledges the support of the Welch Foundation of Houston, Texas.

References

[1] S.J. Cyvin, Chem. Phys. Lett. 147(1988)384.
[2] S.J. Cyvin, J. Brunvoll and B.N. Cyvin, Comput. Math. Appl. 17(1989)355.
[3] S.J. Cyvin, Monatsh. Chem. 120(1989)243.
[4] S.J. Cyvin, J. Brunvoll, B.N. Cyvin, R. Tošić and M. Kovačević, J. Mol. Struct. (THEOCHEM) 200(1989)261.
[5] S.J. Cyvin, B.N. Cyvin, J. Brunvoll and I. Gutman, Monatsh. Chem., to be published.
[6] For details of the topological properties of benzenoid molecules, see: I. Gutman and S.J. Cyvin, *Introduction to the Theory of Benzenoid Hydrocarbons* (Springer-Verlag, Berlin, 1989). In this book, also a full-length presentation of the chemical applications of K-numbers can be found.
[7] For details on perfect matchings, see: L. Lovász and M.D. Plummer, *Matching Theory* (North-Holland, Amsterdam, 1986).
[8] S.J. Cyvin, B.N. Cyvin and I. Gutman, Z. Naturforsch. 42a(1987)181.
[9] S.J. Cyvin and I. Gutman, *Kekulé Structures in Benzenoid Hydrocarbons* (Springer-Verlag, Berlin, 1988), p. 90, eq. (7.20).
[10] C.D. Godsil and I. Gutman, J. Graph Theory 5(1981)137.
[11] V.R. Rosenfeld and I. Gutman, MATCH 24(1989)121.
[12] I. Gutman and S.J. Cyvin, MATCH 23(1988)83.
[13] H. Hosoya, Bull. Chem. Soc. Japan 44(1971)2332.