Tetrahedron. Vol. 31, pp. 147 to 152. Pergamon Press 1975. Printed in Great Britain

GRAPH THEORY AND MOLECULAR ORBITALS—XI

AROMATIC SUBSTITUTION[1]

C. F. WILCOX, JR.[a]
Cornell University, Department of Chemistry, Ithaca, N.Y. 14850, U.S.A.

and

I. GUTMAN[a] and N. TRINAJSTIĆ*
[a]Institute "Rugjer Bošković", P.O.B. 1016, 41001 Zagreb, and *Chemistry Department, Faculty of Science and Mathematics, University of Zagreb, Zagreb, Croatia, Yugoslavia

(*Received in U.K.* 17 *June* 1974; *Accepted for publication* 8 *August* 1974)

Abstract—It is demonstrated that a previously developed topological expression for the total π-electron energy of aromatic molecules provides a good qualitative account of localization energies. The logarithm of the ratio of the algebraic structure counts for the ground state and the localized reaction intermediate is the principle energy determining factor. This expression fails for hydrocarbons with unusually small HOMO-LUMO separations. An alternative topological expression that includes a correction for this situation provides an excellent *non-empirical* explanation for the successful empirical use of Dewar reactivity numbers in correlating exact localization energies. The present analysis provides an explanation for the success of the traditional resonance structure counting technique in predicting relative rates of aromatic substitution. The analysis develop applied only to *alternant* hydrocarbons.

1. INTRODUCTION

In part X[1] of this series[2] the validity of resonance theory was analysed with emphasis on the role of Kekulé structures in determining the thermodynamic stability of conjugated molecules. One of the conclusions obtained was that Kekulé structures should not be simply summed, but their parity should also be taken into account. Hence, if the ASC = algebraic structure count ($K^+ - K^-$) is considered instead of the ordinary structure count ($K^+ + K^-$), resonance theory can be used with equal quantitative success for benzenoid and nonbenzenoid alternant systems. In the present work we would like to extend these arguments to the study of substitution reactions which may occur on an sp^2-C atom of the conjugated molecule. It seems worth mentioning here that in all previous work[2] on the application of graph theory to the chemistry of conjugated systems[3] only ground state properties have been investigated, so that until now it has been possible to give only indirect predictions of reactivity.

The approach used in the previous part and employed here was to develop a simple quantitative expression for the Hückel π-energy in terms of the graph-theoretical parameters N (the number of vertices in the π-network), ν (the number of edges) and ASC. Although the equation given in part X reproduced π-energies to within a few percent, the application of it to calculating the small energy differences that are associated with localizing a π-electron during an electrophilic reaction is a much more demanding test. As will be seen, the previous equation is only partially successful. Analysis of the source of the error leads to an improved non-empirical expression. It will be shown that this expression accounts quantitatively for the success of the empirical applications of Dewar reactivity numbers.

2. A TOPOLOGICAL FORMULA FOR TOTAL π-ELECTRON ENERGY

In this section the essential results derived in Part X will be outlined and recast in a form more suitable for the present analysis. The notation is the same as before and will not be defined in detail.

Following Part X, if there are n zeros in the spectrum of the molecular graph (n > 0), a quantity *asc* is defined as

$$(asc)^2 = \prod_{j=1}^{N}{}' |x_j| \tag{1}$$

where Π' means multiplication over all non-zero x_j's. Since[4]

$$(ASC)^2 = \prod_{j=1}^{N} |x_j| \tag{2}$$

asc is a generalization of the notion of algebraic structure count. For the case where there are no zeros in the graph spectrum (n = 0), of course,

$$ASC = asc \tag{3}$$

---

[a]This work was carried out at the Rugjer Bošković Institute, Zagreb, Croatia, Yugoslavia.

If $n = 1$

$$(asc)^2 = \sum_{p=1}^{N} ((\mathrm{ASC})_p)^2 \tag{4}$$

where $(\mathrm{ASC})_p$ is the algebraic structure count of the graph G-p. Expressions for *asc* can be obtained[1] also for $n > 1$, but they will not be used here because graphs of chemically relevant hydrocarbons have as a rule $n = 0$ when the number of carbon atoms, N, is even and $n = 1$ when N is odd. Note that *asc* is defined in such a way that it is always positive.

In Part X the following approximate formula is derived:

$$2\ln(asc) = \sum_{t=0}^{m} (-1)^{t+1} u_t W'_t \tag{5}$$

where

$$W'_t = \sum_{j=1}^{N}{}' |x_j|^t \tag{6}$$

and u_t's are *positive* constants, which have been determined previously[1] by a least squares method. Σ' means summation over all non-zero x_j's. If

$$W_t = \sum_{j=1}^{N} |x_j|^t \tag{7}$$

it is easily seen that

$$W'_t = W_t \tag{8'}$$

for $t = 1, 2, \ldots$, and

$$W'_0 = W_0 - n \tag{8''}$$

Since

$$\begin{aligned} W_0 &= N \\ W_1 &= E_\pi \\ W_2 &= 2\nu \\ &\text{etc.} \end{aligned} \tag{9}$$

Eq. (5) can be written as

$$E_\pi = \frac{u_0}{u_1}(N - n) + \frac{2}{u_1}\ln(asc) + \frac{2u_2}{u_1}\nu + \text{higher order terms} \tag{10}$$

Let us denote u_0/u_1, $2/u_1$ and $2u_2/u_1$ by A, B and C, respectively and temporarily neglect the "higher order terms". This finally gives a topological formula for total π-electron energy

$$E_\pi = A(N - n) + B\ln(asc) + C\nu \tag{11}$$

The previous values[1] of u_0, u_1 and u_2 cannot be used in Eq (11) because they were evaluated for an equation that included higher order terms which are neglected here. Values of A, B and C derived by a least-squares fitting of Eq (11) to an arbitrarily selected group of 10 benzenoid hydrocarbons and 20 hydrocarbon ions derived from them are 0·913, 0·765 and 0·347, respectively. These constants reproduce the π-elergies with a standard deviation of 0·12 β. The previous four-parameter equation, fitting a different but related set of molecules gave a standard deviation of 0·096 β. Because the present set of parameters fits simultaneously both hydrocarbons and hydrocarbon ions, it would be expected that they should reproduce differences in energy with a precision of 0·12 β.

3. A TOPOLOGICAL FORMULA FOR LOCALIZATION ENERGY

It is well known[5] that localization energies calculated using the HMO theory correlate empirically very well with the reactivity towards aromatic substitution at a particular position in conjugated molecule. In graph-theoretical language the σ-complex generated in the course of a substitution reaction is represented by a graph obtained after the deletion of the appropriate vertex from the molecular graph. For example, the molecule **1** and its two possible σ-complexes **2** and **3** are represented by graphs **4–6**, respectively.

Note that the graphs **5** and **6** have one vertex and two edges less than the graph **4**. Note also that $n(\mathrm{IV}) = 0$ and $n(\mathrm{V}) = n(\mathrm{VI}) = 1$.

Now, if one considers a substitution reaction on the atom labelled by p, the localization energy, L_p, is

$$L_p = E_\pi(G) - E_\pi(G_p) \tag{12}$$

where G is the molecular graph and $G_p = G - p$ is the graph obtained by deletion of the vertex p from G, and, of course, all the edges incident on it. Application of Eq (11) gives

$$E_\pi(G) = AN + B\ln[\mathrm{ASC}(G)] + C\nu \tag{13'}$$

$$E_\pi(G_p) = A[(N-1) - 1] + B\ln[asc(G_p)] + C(\nu - 2) \tag{13''}$$

where it is assumed that $n(G) = 0$ and therefore $n(G_p) = 1$. Hence,

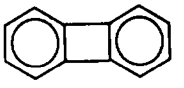
$$L_p = 2(A + C) + B\ln[\mathrm{ASC}(G)/asc(G_p)] \tag{14}$$

where $2(A + C)$ and B are constants (but see discussion later) and thus the localization energy for the position p is determined by a unique topological parameter.

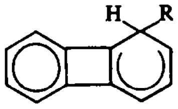
$$D_p = \mathrm{ASC}(G)/asc(G_p) \tag{15}$$

In the subsequent section it will be shown that the quantity $\bar{D}_p = 2D_p$ is identical with an index introduced by Dewar using a completely different and independent line of reasoning.

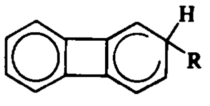
The following discussion should elucidate the chemical meaning of D_p and its close relation to resonance theory.



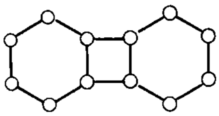
1



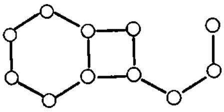
2



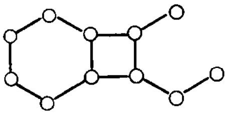
3



4



5



6

The relative reactivity of two atoms p and q (not necessarily of the same molecule) is determined by the difference in the corresponding localization energies

$$\Delta L_{pq} = L_p - L_q = B \ln D_p/D_q \tag{16}$$

Moreover it it is assumed that this energy can be identified with the difference in free energy of activation,[6]

$$k_p/k_q \approx \exp(-\Delta L_{pq}/RT) \tag{17}$$

where k_p and k_q are the rate constants for an aromatic substitution reaction on the atoms p and q, respectively, and T is the absolute temperature. By substituting Eqs (16) back into (17) one obtains finally

$$k_p/k_q = (D_q/D_p)^{\rho} \tag{18'}$$

or

$$\ln(k_p/k_q) = \rho \ln(D_q/D_p) \tag{18''}$$

The form of Eqs (18) strikingly resembles the Hammett equation.[7] However, $\rho = B/RT$ is here a positive constant (for a given temperature) which should not depend on the type of substitution reaction. To obtain the Hammett equation it is necessary to make the additional assumption that linear free energy relationship holds such that different reactions reflect different (but constant) degrees of approach to the fully localized intermediate.

Another property of Eqs (18) seems also to be of particular importance. Namely, if two positions of the same molecule are compared,

$$D_p/D_q = asc(G_q)/asc(G_p) \tag{19}$$

it follows that *the position with larger* asc *value will be more reactive.* By inspection of Eq (4) one sees immediately that *asc* is a quantity closely analogous to the number of resonance forms of the corresponding δ-complex. If all resonance forms are of the same parity, *the more reactive position is the one which has a larger number of resonance structures in the corresponding δ-complex.* However, in the general case the parity of the resonance forms is to be considered and "the number of resonance forms" should be replaced by "the algebraic count of the resonance forms". Therefore, a graph-theoretical justification of the well known resonance theory technique (which is in fact a postulate)[8] for predicting substituent orientation is obtained. Modifications of the traditional theory are necessary when all resonance forms are not of the same parity. The use of algebraic structure count instead of simple structure count enables the extension of the simple resonance treatment of aromatic substitution to nonbenzenoid systems, or, what is equivalent, the traditional resonance treatment is justified only for benzenoid molecules.

Applications to some molecules containing cyclobutadiene rings will be given in the next section.

4. ON THE DEWAR NUMBER

The third important property of Eqs (14), (16) and (18) which we would like to point out is their relationship to the Dewar number $\tilde{D}$. The Dewar number[5] is in fact the approximate localization energy calculated by a simplified version of the perturbational molecular orbital method, and can be evaluated from the coefficients of the non-bonding molecular orbital (NBMO) of the corresponding hydrocarbon radical. Note that the graph which represents this radical was labeled by G_p. The unnormalized) NBMO can be simply calcuated by a method of Longuet-Higgins.[9]

Let this NBMO have a coefficient C_r on the vertex r, and let r and s be the vertices which are in G adjacent to the vertex p. Then the Dewar number (in β units) is defined as[5]

$$\tilde{D}_p = 2(C_r + C_s)/\sum_{r=1}^{N-1} C_r^2 \tag{20}$$

It has been shown by Dewar and Longuet-Higgins[10] and recently elaborated in more detail by Herndon[11] that

$$C_r = \pm ASC(G_{pr}) \tag{21}$$

where $G_{pr} = G_p - r$. Therefore, by comparison of Eqs (4) and (21)

$$\sum_{r=1}^{N-1} C_r^2 = [asc(G_p)]^2 \tag{22}$$

Besides[11]

$$C_r + C_s = ASC(G) \tag{23}$$

and this gives simply

$$\tilde{D}_p = 2D_p \qquad (24)$$

In other words, our graph-theoretical arguments suggest *localization energy is a linear function of the logarithm of Dewar number* (Eq 14). Moreover, in Eqs (16) and (18), which determine the relative reactivities of two C atoms, the Dewar number can be written instead of D.

In Fig 1 are plotted the exact HMO localization energies as a function of the corresponding Dewar numbers. On this plot is drawn both the empirical linear correlation line discussed by Streitwieser[5] (Curve I) and the present logarithmic function (Curve II). For the molecules normally encountered in aromatic substitution studies the Dewar numbers fall in the range of 1·0 to 2·5. As can be seen, the two correlation lines are effectively equivalent in this range, particularly if it is considered that the logarithmic function was not obtained by fitting the localization energies directly. Clearly, the logarithmic function form is flexible enough to fit the localization energies in the range of 1·0 to 2·5 with much greater precision if that were desired.

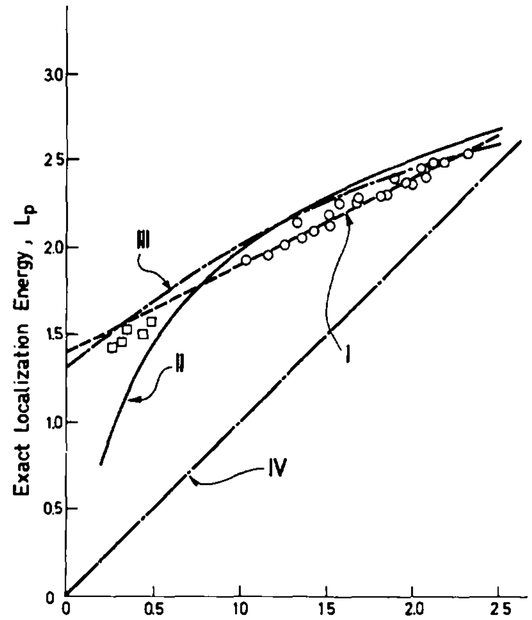


Fig 1. Exact localization energies (in β units) *vs* several functions of the Dewar reactivity numbers. Curve I-Streitwieser empirical linear relationship, curve II-logarithmic relationship (see Eq. (11)), curve III-non-empirical graph theory, and curve IV-non-empirical first order perturbation theory. Circled points correspond to conventional benzenoid molecules; squares correspond to cyclobutadienoid molecules 7–9.

The two lines diverge markedly for smaller values of $\tilde{D}$. To explore this region we have calculated the Dewar numbers and localization energies for indicated positions in the molecules 7–9, which contain formal cyclobutadiene rings with the possibility for introducing Kekulé structures of opposite parity. These molecules have the interesting property of each having an ASC = 1 and therefore both small localization energies and Dewar numbers. When those points are added to the graph on Fig 1 it becomes apparent that although the best correlation line contains some curvature, it is no where near as much as the logarithmic relationship requires.

In restrospect, it is obvious that Eq (14) would have to fail at some point since it is easy to design molecules that have $\tilde{D}$-values arbitrarily close to zero and these would give, according to Eq (14), large negative localization energies—a physical absurdity. In the next section a possible resolution of this difficulty is considered.

5. INCLUSION OF THE HIGHER ORDER AND OTHER TERMS

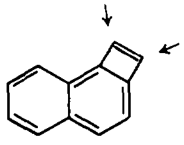
In Eq (10) terms in $|x|^3$ and higher were omitted. However, these can not cause the above mentioned difficulty since they vanish as x goes to zero. It is convinient to add a term like $1/x^2$. In this section we consider the effect of including such terms. For completeness we also include here the $|x|^3$ term since a convenient topological expression is available[1] for W_3. Also, as will be brought out shortly, inclusion of the $|x|^3$ term permits the constants in the final topological expression to be evaluated non-empirically in a way that satisfactorily accounts for the curvature in Fig 1.

Consider the following finite series approximation of $|x|$:

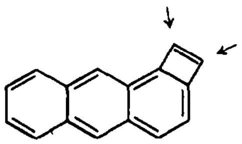
$$|x| = A_L \ln|x| + A_{-2}/x^2 + A_0 + A_2x^2 + A_3|x|^3 \qquad (25)$$

Following the logic developed in Part X, for an alternant hydrocarbon

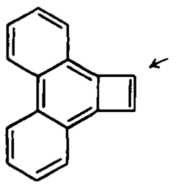
$$E_\pi = A_L \ln(\mathrm{ASC})^2 + A_{-2}\sum\frac{1}{x^2} + A_0N + 2A_2\nu + A_3\left[\frac{2E_\pi}{N}\left(3\nu - \frac{E_\pi^{\,2}}{N}\right)\right] \qquad (26')$$



7



8



9

Similarly for the odd AH derived by localization of one π-electron one has

$$E'_\pi = A_L \ln(asc)^2 + A_{-2}\sum{}' \frac{1}{x^2} + A_0(N-2)$$
$$+ 2A_2(\nu - 2) + A_3\left[\frac{2E'_\pi}{N-1}\left(3(\nu-2) - \frac{E'^2_\pi}{N-1}\right)\right] \quad (26'')$$

where the summation Σ' does not include $x_i = 0$. The difference between the two equations is the required localization energy L_p.

$$L_p = A_L \ln D_p^2 + A_{-2}\left[\sum 1/x^2 - \sum{}' 1/x^2\right] + 2A_0$$
$$+ 4A_2 + A_3\begin{bmatrix}\text{difference}\\ \text{in } W_3 \text{ terms}\end{bmatrix} \quad (27)$$

The first sum over $1/x^2$ contains one more term than the second sum. This extra level arises from the interaction of the NBMO of the odd AH with an isolated zero energy π-center.[5] These split to give the HOMO and LUMO of the even AH. With quite adequate precision (which improves for smaller x) the value of x_{HOMO} can be replaced by $D_p = \tilde{D}_p/2$. Insertion of this approximation into (27) gives:

$$L_p = 2A_L \ln D_p + 2A_{-2}/D_p^2 + 2A_0$$
$$+ 4A_2 + A_3\begin{bmatrix}\text{difference}\\ \text{in } W_3 \text{ terms}\end{bmatrix} \quad (28)$$

Now, Eq (25) can be rewritten as:

$$A_L \ln|x| + A_{-2}/x^2 + A_0 = |x| - A_2x^2 - A_3|x|^3 \quad (29)$$

Substitution of (29) into (28) for $x = D_p$ gives

$$L_p = 2D_p - 2A_2D_p^2 - 2A_3D_p^3 + 4A_2$$
$$+ A_3\begin{bmatrix}\text{difference}\\ \text{in } W_3 \text{ terms}\end{bmatrix} \quad (30)$$

Eq (30) can be further simplified by making the following approximations (see the Appendix)

$$\nu/N = 1{\cdot}11$$

$$E_\pi/N = 1{\cdot}35$$

Although these approximations may be in error for typical molecules by as much as ±10%, it turns out that the constant A_3 in (30) is sufficiently small so that no significant error is introduced in this way. With these approximations the terms of (26) originated from summing over $|x|^3$ become

$$\frac{2E_\pi}{N}\left(3\nu - \frac{E_\pi^2}{N}\right) - \frac{2E'_\pi}{N-1}\left(3(\nu-2) - \frac{E'^2_\pi}{N-1}\right)$$
$$= 3{\cdot}04(E_\pi - E'_\pi) \quad (31)$$

Thus (30) can be rewritten as

$$L_p = \frac{1}{1 - 3{\cdot}04A_3}[4A_2 + 2D_p$$
$$- 2A_2D_p^2 - 2A_3D_p^3] \quad (32)$$

Note that in (32) the localization energy as a function of the Dewar number depends only on the two parameters A_2 and A_3 appearing in Eq (25).

It should be clear from the form of this result that the exact localization energies could be fit rather well by appropriate (least squares) selection of A_2 and A_3. Instead, we have taken a different approach. We have evaluated the parameters A_L, A_{-2}, A_0, A_2 and A_3 in (25) by least-squares fitting a uniform distribution of x's over the range of 0·1 to 3·0. The upper limit corresponds to the maximum possible orbital energy level of a π system;[12] the lower limit corresponds to the smallest energy level encountered in Fig 1. The resulting equation for this range is

$$|x| = 0{\cdot}3578 \ln|x| + 0{\cdot}002608/x^2 + 0{\cdot}6652$$
$$+ 0{\cdot}3879x^2 - 0{\cdot}0581x^3 \quad (33)$$

Eq (33) yields the line marked III in Fig 1.

Note that Eq (32) and the parametrization *via* Eq (33) are non-empirical. (Actually only the bracketed term is strictly non-empirical. However, as noted in Appendix the denominator shows such a small dependence on the structure of the hydrocarbon that it can be thought of as a characteristic constant). The source of the terms if purely topological and the parameters are derived from fitting a well defined mathematical function over a logically defendible range. The Streitwieser relationship is empirical, although intuitively reasonable. An application of 1st order perturbation theory[5] would require that the delocalization energy be exactly equal to $\tilde{D}_p$. As can be seen from Fig 1 a poor accounting for the variation in localization energy is thus obtained.

One might ask how the present result would change if very small x values were encountered. In principle Eq (25) could be extended to include higher powers of $1/x^2$. Since these, even more than the $1/x^2$ term will be dominated by x_{HOMO}, the same substitution used in (28) can be made. This will lead to the incorporation of these into the right hand side of Eq (29) and thus *one is led again to exactly Eq (32). Thus for arbitrarily small x Eq (32) remains valid.*

Finally, we give a word about the scatter present in Fig 1. Part of this is a consequence of the imprecision in the assumed edge/vertex ratio and the "atom energy density" (E_π/N) used in simplifying Eq (31). Another part is undoubtedly due to the neglect of W_4 which introduces energy differences due to differences in branching. As we will discuss in detail elsewhere,[13] branching introduces changes in π-energy of about 0·1 β/branch. This is about the range of scatter present. This argument for the probable importance of branching and the ultimate need to include the W_4 term suggest that the small but non-zero contribution of the W_3 term we have found is *not just an artifact* of our analysis. Because of the scatter in the points of Fig 1 it is difficult convincingly to demonstrate this more directly.

In summary, Eq (32) provides a non-empirical explanation for the success of the empirical Streitwieser correlation of Dewar reactivity numbers with exact localization energies. In particular, this correlation is not linear. Because the Dewar reactivity numbers are proportional to the ratio of the algebraic structure counts for the ground state and the reaction intermediate, Eq (3) explains why the traditional resonance method accounts for the relative reactivities of benzenoid hydrocarbons. When the parity of the resonance structures is also taken into account, aromatic structures containing 4m-membered rings can be treated by the same equation. Note that there are conjugated hydrocarbons containing 4m-membered rings but with zero eigenvalues in the spectrum of their molecular graphs (e.g. 4m-annulenes). Our discussions do not apply to them. Fortunately, these systems are necessarily not aromatic.[16].

Acknowledgement—C. F. W. Jr. would like to acknowledge support from the United States and Yugoslav National Academies of Sciences for his stay at the Institute Rugjer Bošković.

REFERENCES

[1] Part X: I. Gutman, N. Trinajstić and C. F. Wilcox, Jr. *Tetrahedron* **31**, 143 (1975)

[2] For the preceding papers of the series *Graph Theory and Molecular Orbitals* see Ref 1

[3] For review see: I. Gutman and N. Trinajstić, *Topics Curr. Chem.* **42**, 49 (1973)

[4] C. F. Wilcox, Jr., *Tetrahedron Letters* 795 (1968)

[5] A. Streitwieser, *Molecular Orbital Theory for Organic Chemists* Wiley, New York, (1961); M. J. S. Dewar, *The Molecular Orbital Theory of Organic Chemistry*. McGraw-Hill, New York, (1969)

[6] M. J. S. Dewar and R. J. Sampson, *J. Chem. Soc.* 2789 (1956); M. J. S. Dewar, T. Mole and E. W. T. Warford, *Ibid.* 3581 (1956)

[7] J. E. Leffler and E. Grunwald, *Rates and Equilibria of Organic Reactions.* Wiley, New York (1963)

[8] G. W. Wheland, *The Theory of Resonance.* Wiley, New York, (1953)

[9] H. C. Longuet-Higgins, *J. Chem. Phys.* **18**, 265 (1950)

[10] M. S. J. Dewar and H. C. Longuet-Higgins, *Proc. Roy. Soc.* **A214**, 482 (1952)

[11] W. C. Herndon, *Tetrahedron* **29**, 3 (1973); *J. Chem. Educ.* **51**, 10 (1974)

[12] C. A. Coulson, *Proc. Cambridge Phil. Soc.* **46**, 202 (1949)

[13] I. Gutman, B. Rusić, N. Trinajstić and C. F. Wilcox, Jr., *J. Chem. Phys*, in press

[14] F. Harary, *Graph Theory.* Addison-Wesley, Reading, Mass. 1971 (second printing) p. 39

[15] C. A. Coulson and G. S. Rushbrooke, *Proc. Roy. Soc. Edinburgh* **A62**, 350 (1948)

[16] I. Gutman and N. Trinajstić, *Scientia Yug.*, in press

APPENDIX

It is well known[14] that the number of edges, ν, in a graph is related to the number of vertices, N, by the equation $\nu = N + R - 1$, where R is the number of rings. For molecules such as anthracene containing three rings in ground state (two rings in the intermediate) and 14 vertices, the averaged value of ν/N over ground and intermediate states is 1·11. For two rings the average is 1·05 and for an infinite polyacene the average is 1·25. For a set of nine tricyclic molecules the averaged value of E_π/N (for both ground and intermediate states) is 1·345 ($\delta = 0{\cdot}009$); for the infinite linear polyacene the average is[15] 1·403. We have chosen the three ring case as representative of the typical molecule of interest to the organic chemist.

Because of the minus sign in the expression for W_3, the variations in ν/N and E_π/N largely cancel. Thus for the two ring case the coefficient at A_3 in Eq (32) becomes 2·86 and for the infinite linear polyacene it is 3·56. When the value of A_3 is inserted, the denominator becomes 1·166 for $R = 2$, 1·177 for $R = 3$ and 1·207 for $R = \infty$, showing that the structure of the molecular graph has hardly any influence on the denominator in Eq (32).