troscopy of Neutral Indium Phosphide Clusters. *J. Chem. Phys.* **1990**, *92*, 4759.
(36) Balasubramanian, K. A Method for Nuclear Spin Statistics in Molecular Spectroscopy. *J. Chem. Phys.* **1981**, *74*, 6824.
(37) Balasubramanian, K. Generating Functions for the Nuclear Spin Statistics of Non-Rigid Molecules. *J. Chem. Phys.* **1981**, *75*, 4572.
(38) King, R. B. Chemical Applications of Topology and Group Theory. XXII: Lowest Degree Chirality Polynomial for Regular Polyhedra. *J. Math. Chem.* **1987**, *1*, 45.
(39) King, R. B. Graph Theory and Topology in Chemistry. *Stud. Phys. Theor. Chem.* **1987**, *51*, 325.
(40) Meier, U.; Peyerimhoff, S. D.; Grein, F. Ab initio MRD-CI Study of $GaAs^-$, $GaAs_2^{(\pm)}$, $Ga_2As_2^{(\pm)}$, $As_2^{(\pm)}$, and As_4 Clusters. *Chem. Phys.* **1989**, *90*, 4727.
(41) Harter, W. G.; Weeks, D. E. Rotation-Vibration Spectra of Icosahedral Molecules. I. Icosahedral Symmetry Analysis and Fine Structure. *J. Chem. Phys.* **1989**, *90*, 4727.
(42) Balasubramanian, K.; Pitzer, K. S.; Strauss, H. L. Nuclear Spin Statistics of Cubane and Icosahedral Borohydride Ions. *J. Mol. Spectrosc.* **1982**, *93*, 447.

Chemical Graph-Theoretical Cluster Expansion and Diamagnetic Susceptibility†

T. G. SCHMALZ, D. J. KLEIN,* and B. L. SANDLEBACK

Texas A&M University at Galveston, Galveston, Texas 77553-1675

Received September 26, 1991

A general computationally amenable chemical graph-theoretical cluster expansion method is described and illustrated in application to the treatment of magnetic susceptibilities. The "additive" cluster expansion in terms of molecular fragments is found to entail certain linear "near-dependences" of graph-theoretical invariants, again as illustrated for magnetic susceptibilities. The general implementation and efficacy of the method are commented upon.

1. CHEMISTRY AND GRAPH THEORY

Chemical graph theory[1] offers a mathematical framework for classical chemical-bonding ideas, here illustrated in application to the description of the magnetic susceptibilities of organic molecules comprised from C, H, and O atoms. We view *chemical graphs* to be mathematical graph-theoretical precepts. A graph is specified in terms of two sets: first, a set of *sites* (or vertices), here identifying atoms in a molecule; and second, a set of edges, each often viewed as site pairs and here identified as chemical bonds. We do however (ultimately) suppress H atoms, assuming that whatever apparent deficit there is in valence (of C or O) is understood to be thus augmented by the requisite number of H atoms. (That is, we avoid radicals.) Double bonds are identified as such. A *subgraph* G' of a graph G is simply a graph whose site and edge sets are subsets of G; this relation is denoted $G' \subset G$.

Graphs as defined here, and elsewhere,[1] are *labeled*, and one often removes the individual site labels to speak of isomorphism classes of graphs. Of course for chemical purposes, the labels should generally not be completely removed so that one still distinguishes atoms of different elements, and we speak of *chemical* isomorphism classes of graphs. Hence one arrives at the "valence structures" of classical chemistry. For instance, for hydroxylamine the three graphs of Figure 1a are chemically isomorphic and form one chemical isomorphism class. The three graphs of Figure 1b are the same (graphs) as those of Figure 1a, and together form the same chemical isomorphism class, denoted in Figure 1c. The four different (connected) two-atom subgraphs of the first graph in Figure 1a (or Figure 1b) are shown in Figure 1d, where the first two form a chemical isomorphism class of subgraphs.

Molecular properties often may be illuminatingly expressed as derived from the various component fragments of a molecule via a formal[2] chemical graph-theoretical cluster expansion. This expansion in an "additive" format[2-5] is described in Section 2, when a molecular property is expressed as a sum of graph-theoretical invariants (these here being counts of certain types of molecular fragment subgraphs). In Section 3 we go on to display examples of linear "near-dependences" among the graph invariants. This feature, already addressed by Essam et al.,[5] is generally derived for the type of cluster expansion discussed here. The ideas (of Sections 2 and 3) are illustrated in Section 4 for the case of magnetic susceptibilities, thereby extending classical Pascal constant ideas,[6] and also recasting "Hameka" magnetic-susceptibility expansions[7,8] into what we believe is a more useful form, as also earlier suggested in a more-limited context by Randić,[9] though more generally in unpublished works.

† Research supported by the Welch Foundation of Houston, TX.

2. CLUSTER EXPANSION

A molecular property $X(\Gamma)$ for a molecule with graph Γ is to be expressed in terms of contributions $x(G)$ for connected subgraphs $G \subset \Gamma$. The general (additive) *cluster expansion* then is[2-4]

$$X(\Gamma) = \sum_{G \in C(\Gamma)} x(G) \tag{2.1}$$

where $C(\Gamma)$ is a set of connected subgraphs of Γ and $x(G)$ are fragment contributions. If we denote the number of subgraphs of Γ chemically isomorphic to G by $n_\Gamma[G]$, then eq 2.1 may be recast as

$$X(\Gamma) = \sum_{[G] \in C[\Gamma]} x(G) n_\Gamma[G] \tag{2.2}$$

where now the sum goes only over chemical isomorphism classes. For instance, hydroxylamine of Figure 1 has

$$n_\Gamma[\mathrm{H}] = 3 \quad n_\Gamma[\mathrm{N}] = n_\Gamma[\mathrm{O}] = 1 \quad n_\Gamma[\mathrm{H\text{-}N}] = 2$$
$$n_\Gamma[\mathrm{N\text{-}O}] = n_\Gamma[\mathrm{O\text{-}H}] = 1 \quad n_\Gamma[\mathrm{H\text{-}N\text{-}O}] = 2 \tag{2.3}$$
$$n_\Gamma[\mathrm{H\text{-}N\text{-}H}] = 1$$
$$n_\Gamma[\mathrm{H\text{-}N\text{-}O\text{-}H}] = 2 \quad n_\Gamma[\mathrm{H_2N\text{-}O}] = 1 \quad n_\Gamma[\Gamma] = 1$$

Of course, the full-cluster expansion of eqs 2.1 or 2.2 is exact. Approximations arise with the truncation of these expansions, say at subgraphs of a limited size. One simple *size measure* (of several possibilities[2]) is the number $|G|$ of vertices in G. Such truncated expansions are common, e.g., for magnetic susceptibilities with atomic contributions (and

0095-2338/92/1632-0054$03.00/0

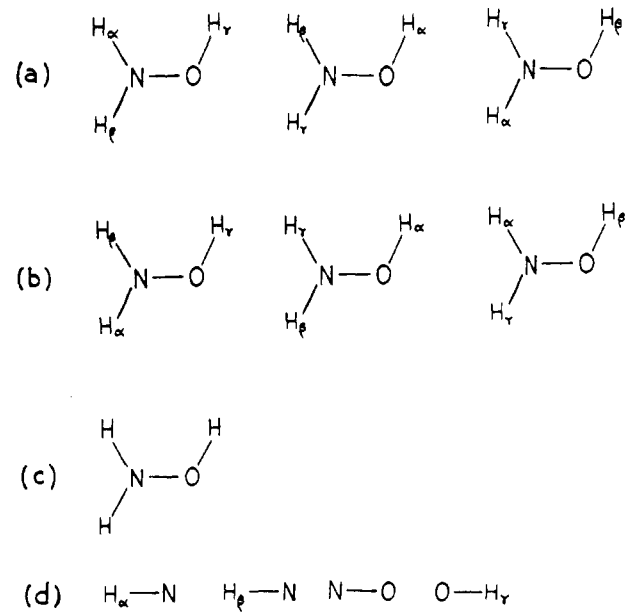

Figure 1. (a) and (b) display in two different ways the three labelings for the hydroxylamine graph, while its "unlabeled" form is given in (c). Finally, (d) displays the various (labeled) 2-atom subgraphs of the first of the graphs in (a) or (b).

sizes limited to $|G| = 1$) or for bond-energy expansions of heats of atomization (and sizes limited to $|G| = 2$). A further choice in making a cluster expansion comes in specifying the set $C(\Gamma)$. For instance, the subgraphs G of $C(\Gamma)$ may be required to not only be connected but also *induced*, i.e., to be such that every pair of vertices of G which are adjacent in Γ are also adjacent in G. Indeed, this is the choice taken here.

For standard organic (nonradical) species, a further simplification is possible. This entails the use of *hydrogen-deleted* graphs, with all H atoms omitted from our graphs, any remnant valences being understood to be satisfied by suppressed-H atoms. Indeed this is a standard approach taken in organic chemistry. Here we imagine the cluster expansions of eqs 2.1 and 2.2 are in terms of such graphs, and corresponding isomorphism classes. That this reduction to H-deleted graphs entails no approximation is indicated in Appendix A. The complete cluster expansion for hydroxylamine now takes the form

$$X(\mathrm{NO}) = x(\mathrm{N}) + x(\mathrm{O}) + x(\mathrm{NO}) \qquad (2.4)$$

For larger molecules the expansion is but slightly less trivial, at least so long as the sizes $|G|$ are not too large.

There are different manners by which cluster expansion may be used. One approach is to seek a size truncation with a few parameters $x(G)$ that are then determined by a least-squares fitting of the truncated expansions to measured (or otherwise known) values $X(\Gamma)$ for a data set of species Γ. Another approach is via "inversion" using only those presumed known $X(\Gamma)$ for Γ up to the limiting size for $|G|$.

3. LINEAR NEAR-DEPENDENCES

With the H-deleted graphs of Section 2 and the restriction to standard organic species, the full-cluster expansion of eq 2.2 is exactly complete. That is the $x(G)$ are all unique if a complete data set is used, so we understand that there must be no linear dependences among the $n_\Gamma[G]$ [otherwise the linear dependence would allow shifting of effects between different $x(G)$].

But in fact for each possible connected subgraph γ there is exactly one linear relation that is true for all Γ other than

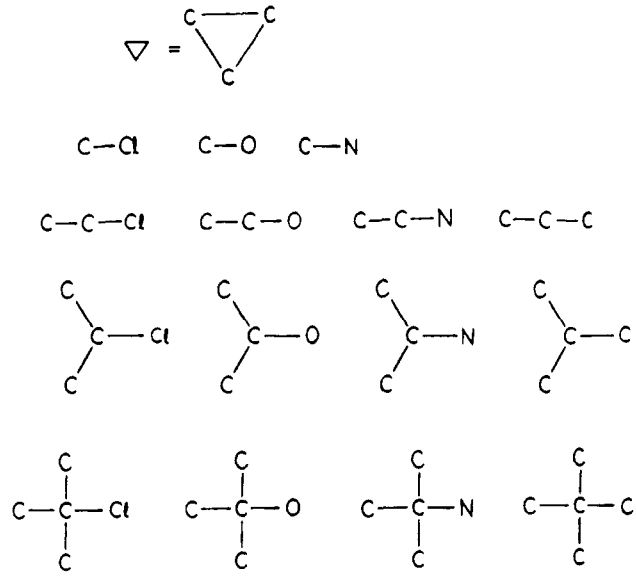

Figure 2. First line indicates the abbreviation for a (cyclopropane) 3-ring, while the remaining lines indicate the graphs $[G]$ in the summation of eq 3.1.

$\Gamma = \gamma$. Examples of such linear near dependences (for $|\gamma| = 1$) are

$$n_\Gamma[\mathrm{Cl}] - n_\Gamma[\mathrm{C\text{-}Cl}] = 0 \quad ,\Gamma \neq \mathrm{Cl}$$

$$n_\Gamma[\mathrm{O}] - n_\Gamma[\mathrm{C\text{-}O}] + n_\Gamma[\mathrm{C\text{-}O\text{-}C}] = 0 \quad ,\Gamma \neq \mathrm{O}$$

$$n_\Gamma[\mathrm{N}] - n_\Gamma[\mathrm{C\text{-}N}] + n_\Gamma[\mathrm{C\text{-}N\text{-}C}] + n_\Gamma[\mathrm{C\text{-}N\text{-}C\text{-}C}] = 0 \quad ,\Gamma \neq \mathrm{N}$$

$$n_\Gamma[\mathrm{C}] - 2n_\Gamma[\mathrm{C\text{-}C}] - 2.2n_\Gamma[\mathrm{C{=}C}] + 3n_\Gamma[\nabla] - \sum(-1)^{|G|}n_\Gamma[G] = 0 \quad ,\Gamma \neq \mathrm{C} \qquad (3.1)$$

where ∇ denotes the first graph of Figure 2 and the G summation is over the remaining (chemical) graphs of Figure 2. In fact, the near-linear dependences for γ = Cl, O, N, or C take the specific forms of equations (see eq 3.1) only if we restrict the parent Γ to involve the following:

(i) The only heteroatoms are Cl, O, and N.

(ii) The only multiple bonds are C=C.

(iii) The only triangular rings are all C.

(iv) There is no more than one of the exceptional structures of (i), (ii), or (iii) attached to any atom and its neighbors (in Γ).

If these restrictions do not apply, all that happens is that there are more terms in the near-linear dependences of eq 3.1. The first of these is especially evident since with our restrictions of (i) and (iv) in play we consider only Γ with the heteroatom Cl attached to C (or H), where it is apparent that the number of Cl atoms is the same as the number of Cl-C bonds (if $\Gamma \neq$ Cl). The other relations of eq 3.1 may be viewed to be of a similar nature, though less trivial. If Γ is an exceptional graph (Cl, O, N, or C corresponding to molecular species HCl, H_2O, NH_3, or CH_4), the right-hand side of the corresponding equation in eq 2.1 may be verified to be 1 rather than 0.

A formal expression for *all* the linear near-dependences may be given, without the restrictions of (i), (ii), (iii), or (iv) of the preceding paragraph. We make use of the extension set $\epsilon[\gamma]$ of valence-structure graphs $[G]$ such that γ can be embedded as a subgraph in G in such a way that every vertex of G either is in γ or is adjacent to one in γ. The number of times γ is so contained in such a G of $\epsilon[\gamma]$ is denoted by $m_G[\gamma]$. Then the linear near-dependences all are (without restriction)

$$\sum_{[G] \in \epsilon[\gamma]} (-1)^{|\gamma|-|G|} m_G[\gamma] n_\Gamma[G] = \delta(\gamma,\Gamma) \qquad (3.2)$$

The sum here is finite since only a finite number of atoms of

a finite number of types connected by a finite number of bonds can be appended to any site in γ. With a restriction to single-site γ and $\Gamma \neq \gamma$ satisfying relations (i)–(iv), eq 3.2 reproduces those of eq 3.1. The general relations of eq 3.2 are established in Appendix B. The existence of such relations was earlier vaguely hinted at by Smolenskii,[3] illustrated by Gordon and Kennedy,[4] and established in a related different circumstance by Essam et al.[5]

The near-dependences revealed in this section have (we believe) implications for the practical application of cluster expansions. The "complications" arise when one uses G of sufficient size that all are included both in an extension $\epsilon[\gamma]$ and as a subgraph in some Γ in the considered data set D. If the Möbius inversion approach is used, then all is fine. But let us consider the common circumstance where a least-squares procedure is used. First, if $\gamma \notin D$, then one has a linear dependence (for D) and one of the $x(G)$ for $G \in \epsilon[\gamma]$, say $x(\gamma)$, could in principle be omitted. But even if $\gamma \in D$ with no strict linear dependence, there are an infinite number of conceivable molecules, all but one of which (namely γ), lie in a lesser dimensional subspace of the full parameter space. Hence some special treatment of the parameter $x(\gamma)$ determining $X(\gamma)$ seems plausible. Indeed, we surmise that if $x(\gamma)$ is chosen so as to exactly fit $X(\gamma)$, then the later $x(G)$, i.e., those $x(G)$ with $G \supset \gamma$, should show a smoother pattern of convergence, say toward smaller values of $x(G)$ for larger G, at least for "well-behaved" properties. Indeed even if $\gamma \notin D$, we advise requiring $x(\gamma)$ to be determined so as to fit a best estimate of $X(\gamma)$, say as taken from some quantum chemical computational or crude experimental estimate. The remaining $x(G)$ would be chosen by least-squares.

4. MAGNETIC SUSCEPTIBILITIES

Magnetic susceptibilities of stable organic species have long been observed[6] to be approximately fit by a sum of atomic contributions—so-called Pascal's constants. A refined (quantum chemically motivated) expansion was developed by Hameka[7] and then later by Flygare and co-workers.[8] This refined expansion, however, was not in terms of the $n_\Gamma[G]$ but rather was in terms of combinations of such invariants based on H-included graphs. These combinations often were originally identified in terms of combinations arising on taking differences for molecular species in homologous series. Beyond counting the various structural fragments in a molecule as in our graph-theoretical expansion, the earlier advocated[7-9] use of the Hameka approach entails appropriately combining them together to form the Hameka invariants for each molecule, this procedure evidently being attempted by intelligent inspection after practice with a sufficient number of examples. The chemical graph-theoretical cluster expansion recommended here, however, entails just the standard and intuitively appealing fragment invariants $n_\Gamma[G]$. Indeed this view has been earlier recommended by Randić,[9] in a more limited application. Making this expansion up through $|G|$ = 3-site fragments, we obtain the $x(\gamma)$ values of Table I via a least-squares fitting to the 108 magnetic susceptibilities of C-, H-, and O-based organics listed by Burnham et al.,[8] except for H_2O, which they did not include. We have followed the recommended approach of the preceding section where H_2O is fit exactly with the single-O-atom parameter x(O) and the remaining eight parameters are determined by least-squares. This occurs because there is but a single linear near-dependence (with our chosen species and fragments G)

$$n_\Gamma[\mathrm{O}] - n_\Gamma[\mathrm{C\text{-}O}] + n_\Gamma[\mathrm{C\text{-}O\text{-}C}] + n_\Gamma[\mathrm{O\text{-}C\text{-}O}] = \delta(\mathrm{O}, \Gamma) \tag{4.1}$$

This relation extends slightly the second one of eq 3.1, since we are presently allowing Γ with O-C-O fragments, relaxing the restrictions (i)–(iv). The standard deviation from the experimental magnetic susceptibilities, which range from 21 to 101 mu, is σ = 0.8 mu, where mu abbreviates the "magnetic unit" of -10^{-6} erg/G^2 mol.

Table I. Susceptibility Expansion Parameters

γ	$x(\gamma)$ (in -10^6erg/G^{2}mol)	Hameka-type parameters
C	17.62	$\chi_{\tilde{A}} + \chi_{\tilde{C}}$
O	12.96	$\chi_{\omega\tilde{O}}$
C-C	−6.66	$-\chi_{\tilde{A}} + \chi_{\tilde{B}} - 2\chi_{\tilde{C}}$
C-O	−8.85	$\chi_{\tilde{F}O} - \chi_{\tilde{G}O} - \chi_{\omega\tilde{O}}$
C=C	−14.71	$-\chi_{\tilde{A}} - \chi_{\tilde{B}} - 2\chi_{\tilde{C}} + \chi_{\tilde{D}} + 2\chi_{\tilde{E}}$
C=O	−18.26	$-\chi_{\tilde{B}} + \chi_{H\tilde{O}} + 2\chi_{I\tilde{O}} - \chi_{\omega\tilde{O}}$
C-C-C	0.52	$\chi_{\tilde{C}}$
C-C-O	0.73	$\chi_{G\tilde{O}}$
C-O-C	1.21	$-\chi_{\tilde{A}} + \chi_{\tilde{B}} + 2\chi_{F\tilde{O}} + \chi_{F\tilde{O}} + \chi_{K\tilde{O}}$
O-C-O	1.60	$-\chi_{\tilde{C}} + 2\chi_{G\tilde{O}} + \chi_{L\tilde{O}}$
C-C=C	0.12	$\chi_{\tilde{C}} - \chi_{\tilde{E}}$
O-C=C	1.06	$-\chi_{\tilde{E}} + \chi_{G\tilde{O}} + \chi_{K\tilde{O}}$
C-C=O	−0.25	$-\chi_{I\tilde{O}}$
O-C=O	4.69	$-\chi_{I\tilde{O}} + \chi_{M\tilde{O}}$

Notably the fit obtained by our procedure yields exactly the same $X(\Gamma)$ estimates as made earlier by Burnham et al.[8] (except for H_2O, which was omitted from their list). This is because there is a linear transformation from the Hameka invariants to our present ones, as also indicated in Table I. (The relation for n_Γ[C=C] corrects that of ref 9.) The occurrence of one more parameter in our approach occurs through inclusion of H_2O, whose susceptibility is not representable in terms of the Hameka invariants earlier used.

Finally in the current chemical graph-theoretical approach, it may be observed that the inclusion of x(O) with a value appropriate for H_2O aids in the appearance of convergence. That is, if H_2O were omitted along with n_Γ[O] as linearly dependent on other higher retained $n_\Gamma[G]$, then the large value 12.96 mu for x(O) would be spread out over other "higher" $x(G)$, making them inordinately large. Thus, even if H_2O were not included in the data set, a reasonable approach would be to constrain x(O) so as to fit a rough "estimate" for the susceptibility of H_2O (say as obtained from a crude quantum chemical computation). The 3-atom contribution for [γ] being [O—C=O] is evidently rather large, to distinguish the notable chemical (and physical) difference between carboxylic acids on one hand and aldehydes or ketones on the other hand.

APPENDIX A. HYDROGEN DELETION

Here we explicitly display the rationale for H-deletion from graphs, in the circumstance that each type of atom A is to satisfy a fixed valence v_A. Then the total number of bonds radiating away from A atoms is $n_\Gamma[\mathrm{A}]v_A$, where those bonds to H atoms are included and where any bonds between A atoms are counted twice. But this total number can also be resolved into contributions from different types of bonds between different pairs of atoms (one of which is A), and including H atoms

$$n_\Gamma[\mathrm{A}]v_A = \sum_B n_\Gamma[\mathrm{A\text{-}B}]n_A[\mathrm{A\text{-}B}] + 2\sum_B n_\Gamma[\mathrm{A{=}B}]n_A[\mathrm{A{=}B}] \tag{A.1}$$

where n_A[A-B] = n_A[A=B] = 1 unless B is A in which case n_A[A-A] = n_A[A=A] = 2. The second B-sum in eq A.1 of course accounts for double bonds, and if triple bonds were to be treated, a third B-sum (with a factor of 3) would be added onto the right-hand side of eq A.1. Now eq A.1 provides a strict linear dependence for the full graph including H. Hence, for A other than H, eq A.1 may be used to eliminate the n_Γ[A-H] in favor of non-H-containing pair counts, as well as the A-count n_Γ[A]. For A being H (and Γ not H_2), eq A.1 may be used to eliminate the H-count in terms of H-B-counts, which are already eliminatable as noted in the preceding sentence. Evidently the case of Γ being H_2 is not treated

(unless somehow we wish to identify H_2 with the empty graph, whence one would presumably add a constant into our formal cluster expansions).

The elimination of H from larger fragments follows a similar development. Let G_A be a (connected induced) subgraph containing an A atom at a given position, and let i be the number of bonds this special A has internal to G_A. Then counting the additional bonds radiating from such an A in G_A in two different ways (as before), we find

$$n_\Gamma[G_A](v_A - i)n_A[G_A] = \sum_B n_\Gamma[G_{AB}]n_{G_A}[G_{AB}] + 2\sum_B n_\Gamma[G_{A:B}]n_{G_A}[G_{A:B}] \quad (A.2)$$

where G_{AB} and $G_{A:B}$ are (induced) graphs obtained from G_A by joining a B atom with a single or double bond. For the case that there are no H atoms in the fragment G_A, eq A.2 can be used to eliminate fragments G_{AH}. For the case that there is one H atom in G_A, eq A.2 can be used to eliminate G_{AH} in terms of G_A, which in turn are eliminated as indicated in the preceding sentence. The idea is similar if G_A contains two (or more) H atoms.

APPENDIX B. LINEAR NEAR-DEPENDENCE

The establishment of the general relation eq 3.2 most quickly follows through the use of known Möbius function theory.[10] This theory derives from the inversion of eq 3.1 to give

$$x(G) = \sum_{\gamma \in C(G)} \mu(\gamma, G)X(\gamma) \quad (B.1)$$

where $\mu(\gamma, G)$ is the *Möbius function* associated to the subgraph partial-ordering relation defined on $C(\Gamma)$. For this "poset" it turns out that[11]

$$\mu(\gamma, G) = \begin{cases} (-1)^{|G| - |\gamma|}, & G \in \varepsilon(\gamma) \\ 0, \text{ otherwise} \end{cases} \quad (B.2)$$

Now also, as is suggested upon substitution of eq B.1 into eq 2.1

$$\sum_{G \in C(\gamma)} \mu(\gamma, G) = \delta(\gamma, \Gamma) \quad (B.3)$$

But the G-sum over labeled graphs here may be broken into two parts; first, one over all "unlabeled" graphs (i.e., all chemical isomorphism classes), and second, one over all possible labelings available for the unlabeled graph, so that

$$\sum_{[G] \in C[\Gamma]} \sum_{G' \in [G]} \mu(\gamma, G') = \delta(\gamma, \Gamma) \quad (B.4)$$

Yet we also may introduce a sum over all $\gamma' \in [\gamma]$, with the right-hand side of eq B.4 remaining unchanged since it is nonzero only in the case that there is but a single term in the sum, so that

$$\sum_{[G] \in C[\Gamma]} \sum_{\gamma' \in [\gamma]} \sum_{G' \in [G]} \mu(\gamma', G') = \delta(\gamma, \Gamma) \quad (B.5)$$

Now the number of $\gamma' \in [\gamma]$ for which $\mu(\gamma', G')$ at a given G' takes a (fixed nonzero) value is just $m_{G'}[\gamma]$ of Section 3, so

$$\sum_{[G] \in C[\Gamma]} \sum_{G' \in [G]} m_{G'}[\gamma]\mu(\gamma', G') = \delta(\gamma, \Gamma) \quad (B.6)$$

Finally utilizing eq B.2 with the elimination of the G'-sum, we obtain the desired result of eq 3.2.

REFERENCES AND NOTES

(1) See, e.g., Trinajstić, N. *Chemical Graph Theory*; Chemical Rubber Co.: Boca Raton, FL, 1983.
(2) Klein, D. J. *Int. J. Quantum Chem.* **1986**, *S20*, 153-173.
(3) Smolenskii, E. A. *Russ. J. Phys. Chem. (Engl. Transl.)* **1964**, *35*, 700-702.
(4) Gordon, M.; Kennedy, J. W. *J. Chem. Soc., Faraday Trans. 2* **1973**, *69*, 484-504.
(5) Essam, J. W.; Kennedy, J. W.; Gordon, M.; Whittle, P. *J. Chem. Soc., Faraday Trans. 2* **1977**, *73*, 1289-1307.
(6) See, e.g., Sellwood, P. W. *Magnetochemistry*; Interscience Pub.: New York, 1956.
(7) Hameka, H. F. *J. Chem. Phys.* **1961**, *34*, 1996. Haley, L. V.; Hameka, H. F. *J. Am. Chem. Soc.* **1974**, *96*, 2020-2024. O'Sullivan, P. S.; Hameka, H. F. *J. Am. Chem. Soc.* **1970**, *92*, 25-32.
(8) Burnham, A. K.; Lee, J.; Schmalz, T. G.; Beak, P.; Flygare, W. H. *J. Am. Chem. Soc.* **1977**, *99*, 1836-1844.
(9) Randić, M. *Chem. Phys. Lett.* **1978**, *53*, 602-605, and private communication.
(10) See, e.g., Berge, C. *Principles of Combinatorics*; Academic Press: New York, 1971; Chapter 3.
(11) Domb, C. In *Phase Transitions and Critical Phenomena*; Domb, C., Green, M. S., Eds.; Academic Press: New York, 1974; pp 1-94.

Representation of Molecular Graphs by Basic Graphs†

MILAN RANDIĆ

Department of Mathematics and Computer Science, Drake University, Des Moines, Iowa 50311

Received August 27, 1991

We consider the problem of the analytical representation of graphs, molecular graphs in particular, based on a sufficiently broad selection of invariants that permits analogies with analytical representations of vectors. Desirable features of basis invariants are discussed and illustrated on paths and matching paths. It appears that they form a sufficiently broad basis for representation of graphs and for discussion of their differences and similarities. The approach is illustrated on smaller alkanes. Next we consider an ad hoc set of several pairs of graphs which have the same count of selected invariants and, thus, may be suspected as counterexamples to the uniqueness of the prime codes. Finally, similarity for a set of monocyclic monoterpenes is considered. Factors influencing a measure of similarity are discussed. Normalization of similarity matrices is proposed when comparisons are based on the representation of graphs using a different number of components. The notion of an overall or global similarity based on all components of a sufficiently comprehensive set of invariants is suggested.

INTRODUCTION

Mathematical modeling in chemistry, physics, and biology often starts with graphs as the objects, the examination of which may answer the questions of interest in such studies. In contrast to computational approaches to chemistry, physics, and biology, conveyed by adopting a suitable vector basis, such as illustrated by the use of selected basis functions in molecular orbitals calculations, in chemical graph theory there are no

† Dedicated to Professor Manfred Eigen (Nobel Laureate in Chemistry, 1967).

0095-2338/92/1632-0057$03.00/0