All acyclic hydrocarbons: formula periodic table and property overlap plots *via* chemical combinatorics

Laimutis Bytautas,† Douglas J. Klein* and Thomas G. Schmalz

Theoretical Chemical Physics Group, Department of Oceanography, Texas A&M University at Galveston, Galveston, TX 77553-1675, USA

Received (in Gainesville, FL, USA) 25th August 1999, Revised manuscript received 30th November 1999, Accepted 28th January 2000

It is sought to extend the characterization of isomer classes by using combinatoric techniques long used for isomer enumerations. To this end a general family of isomers (which is challenging to deal with even at the level of conventional enumeration) is considered, namely that of all acyclic hydrocarbons C_nH_{2m} with classical valence structures. This set of hydrocarbons is partitioned into structural-isomer classes, and these in turn are presented in the form of a "formula periodic table" with n and m identifying rows and columns, respectively. Then various isomer-class-averaged characteristics (beyond isomer enumerations) are sought to be entered into the associated positions in this periodic table. Such characteristics include heat of formation, magnetic susceptibility and index of refraction, each of which may be estimated *via* substructural "cluster expansion". These cluster-expansion-estimated property averages as well as associated standard deviations are developed by way of a generating-function chemical combinatorics for isomers of up to $n = 25$ carbons and $2m = 52$ hydrogens. These results then are incorporated in property "overlap plots" to reveal overall isomer-class property trends in the periodic table for this set of $\approx 10^{13}$ structural isomers distributed over ≈ 360 isomer classes.

1. Introduction

Chemical isomerism has a rich history summarized in ref. 1 and it still represents a concept of crucial importance in modern chemistry. In general, the concept of isomer applies to structural and stereo-isomers, as well as substitutional isomers based on a given molecular skeleton. As early as 1797 Alexander von Humboldt[2] reached the fundamental conclusion that some substances with the same overall chemical composition exhibit different properties. A crucial point in the understanding of chemical isomerism was reached with the introduction of the concept of structural isomerism by Crum Brown[3] and this provided an essential piece of evidence for the validity of classical molecular structural formulas. A classical question concerns the generation of possible structural formulas, or an even simpler question concerns just their numbers, which can turn out to be quite large. Indeed, several different techniques to make such enumerations have been developed over the years, with a renowned systematic effort by A. Cayley (around 1874) for the enumeration of alkane isomers;[4] another notable effort was by the group of Henze and Blair[5] in Texas in the early 1930s. A major advance in chemical isomer combinatorics was provided by G. Polya's establishment and illustration[6] of a suite of fundamental theorems in *ca.* 1936; this theory still remains a foundational tool for chemical enumerations, as attested to in a few books.[7–9] In fact, though motivated by the chemical isomer-enumeration problem, Polya's theory is relevant for all kinds of combinatorial enumerations, as discussed in many combinatorics texts. But because of other possible chemical applications special note might be made first of a book[10] concerning the general utility of graph theory for a diversity of mathematico-combinatorial enumerations; secondly, an especially interesting review is that of Balasubramanian[11] concerning enumerations of various quantum chemical and spectroscopic quantities. Furthermore, the existence of different isomers corresponding to the same chemical formula is closely related to the chemical processes on potential energy surfaces whereon isomers are interconverted, and some aspects of this are mentioned in Fujita's book.[8] Since Polya's work there are perhaps more than 100 papers concerning the use of chemical-combinatoric generating-function enumerations, with many of these referenced in the cited reviews[7–11] as well as in a few more recent articles.[12–14] Additional history of pre-Polya enumerations can be found in refs. 1, 9, 12 and 15.

The present work continues the systematic combinatoric treatment of chemical isomerism, but now extensions are sought beyond enumeration, to enable further characterization of isomer classes, including distinctions between different classes and variations within classes. Indeed, we have previously illustrated[12] extensions of the combinatoric theory in application to the classic case of alkanes. Here we focus on the characterization of isomer classes for the more diverse circumstance of all acyclic hydrocarbons (which have a classical valence structure). In general, such a characterization can be given in terms of: the chemical formula; the (structural or geometric) isomer counts; isomer-class averages for various structural graph invariants; and isomer-class averages (along perhaps with standard deviations) for different physico-chemical properties. We have already begun[13,14] the more formal mathematical (generating function) aspects of the characterization of this (rather extensive) class of compound and here we emphasize the isomer-class averages and standard deviations for a few representative molecular properties, as estimated *via* substructural cluster expansions. That is, the considered chemical properties are taken as linear combinations of suitable sets of graph invariants, such as is often

† On leave from: Institute of Theoretical Physics and Astronomy, Gostauto 12, Vilnius 2600, Lithuania.
Present address: Chemistry Department, Iowa State University, Ames, IA 55010, USA.

done in developing "quantitative structure-property relations" (QSPR). Then granted such an expansion, isomer-class property averages are obtained as the corresponding linear combinations of the averaged graph invariants, which are collectively computed *via* combinatorial "generating-function" techniques (without explicitly computing the properties for each individual isomer). Though these techniques are extensions of those used for the now standard isomer enumerations, even the enumerations have typically been applied to simpler isomer families than that of all acyclic hydrocarbons. Most typically, earlier treatments were for substitutional isomers with reference to a fixed skeleton, or for the alkanes or for alkane-like problems (*e.g.*, for the so-called "polyenoids" with a valence restriction of 3 rather than 4); again many references to such enumerations can be found in refs. 7, 8, 9, 11–14. There is one earlier enumerative treatment[16] of all acyclics, but it differs from our current work in that this earlier work was presented in a different tabular format, was more limited in the range of isomer classes considered, and was more limited in the isomer-class characteristics considered.

The results of our characterization are presented in terms of *a formula periodic table* for acyclic hydrocarbons with classical valence structures. This periodic table is a 2-dimensional array with the position in the nth row and mth column associated to the isomer class of acyclics with formula C_nH_{2m}. Generally such a table can provide a sort of "bird's-eye view", as illustrated in Table 1, where we display enumerations, which (for convenience of display) are rounded off from our earlier exact results.[13] An entry vEp in Table 1 denotes a value $v \times 10^{+p}$. In the general acyclic hydrocarbon "periodic table" (whether displaying enumerations or other properties) the alkanes occur along the upper right diagonal with $m = n + 1$. This alkane diagonal has been repeatedly studied earlier, with a tabulation up to $n = 100$ by Trinajstìc *et al.*[9] The next diagonal with $n = m$ gives counts for the alkene (structural) isomers, while the third diagonal with $m = n - 1$ gives counts for the combination of alkadienes and alkynes. Successive diagonals identify increasing *degrees of unsaturation* $u \equiv n + 1 - m$. The isomer counts down any one of these diagonals have been shown[13] to increase exponentially in n, with the same asymptotic growth factor for every diagonal. Down vertical columns the increase is non-exponential (in n), being $\approx n^{4(m-1)}$ down the mth column. Replacement of enumerations by average values for various properties in our acyclic hydrocarbon table should neatly reveal various further property trends. It may be noted that our formula periodic table of acyclics is conceptually analogous to Dias'[17] formula periodic table of polycyclic aromatic hydrocarbon isomer classes.

A primary question concerns the methods by which isomer-class averages can actually be achieved, since the number of isomers C_nH_{2m} increases dramatically with increasing n and m, there being over 10^{12} isomers in several of the classes appearing in the portion of the table explicitly displayed in Table 1. An efficient way to obtain the property estimates is based on chemical combinatorics, in combination with "cluster-expansion" techniques[18–21] frequently utilized in chemistry. In particular, our earlier article[14] presented a detailed formalism for computing isomer-class average values for a suite of 10 basic "atom-type" graph invariants, and a few consequently

Table 1 Structural isomer counts for isomer classes C_nH_{2m}

$n\backslash 2m$	2	4	6	8	10	12	14	16	18	20	22	24	26	28
2	1.0E0	1.0E0	1.0E0											
3	0	2.0E0	1.0E0	1.0E0										
4	4.0E0	2.0E0	4.0E0	3.0E0	2.0E0									
5	0	4.0E0	6.0E0	9.0E0	5.0E0	3.0E0								
6	1.0E0	5.0E0	1.5E1	2.2E1	2.3E1	1.3E1	5.0E0							
7	0	8.0E0	2.4E2	5.5E1	6.6E1	5.8E1	2.7E1	9.0E0						
8	1.0E0	9.0E0	4.8E2	1.2E2	2.0E2	2.1E2	1.5E2	6.6E1	1.8E1					
9	0	1.4E1	7.6E2	2.6E2	5.1E2	7.0E2	6.3E2	4.0E2	1.5E2	3.5E1				
10	1.0E0	1.7E1	1.3E2	5.2E2	1.3E3	2.1E3	2.4E3	1.9E3	1.1E3	3.8E2	7.5E1			
11	0	2.3E1	2.1E2	1.0E3	2.9E3	5.8E3	8.0E3	8.0E3	5.7E3	2.9E3	9.1E2	1.6E2		
12	1.0E0	2.7E1	3.3E2	1.9E3	6.5E3	1.5E4	2.5E4	3.0E4	2.6E4	1.7E4	7.8E3	2.3E3	3.5E2	
13	0	3.6E1	5.0E2	3.4E3	1.4E4	3.7E4	7.1E4	1.0E5	1.1E5	8.5E4	5.0E4	2.1E4	5.7E3	8.0E2
14	1.0E0	4.3E1	7.6E2	6.0E3	2.8E4	8.7E4	1.9E5	3.2E5	3.9E5	3.8E5	2.7E5	1.5E5	5.7E4	1.4E4
15	0	5.4E1	1.1E3	1.0E4	5.5E4	2.0E5	4.9E5	9.3E5	1.3E6	1.5E6	1.3E6	8.7E5	4.4E5	1.6E5
16	1.0E0	6.3E1	1.6E3	1.7E4	1.1E5	4.2E5	1.2E6	2.6E6	4.3E6	5.5E6	5.6E6	4.4E6	2.7E6	1.3E6
17	0	7.8E1	2.2E3	2.8E4	2.0E5	8.9E5	2.9E6	6.9E6	1.3E7	1.9E7	2.2E7	2.0E7	1.5E7	8.5E6
18	1.0E0	9.2E1	3.2E3	4.5E4	3.5E5	1.8E6	6.5E6	1.8E7	3.7E7	6.0E7	7.9E7	8.4E7	7.2E7	4.9E7
19	0	1.1E2	4.3E3	7.1E4	6.3E5	3.6E6	1.4E7	4.3E7	1.0E8	1.8E8	2.7E8	3.3E8	3.2E8	2.5E8
20	1.0E0	1.3E2	5.9E3	1.1E5	1.1E6	6.9E6	3.1E7	1.0E8	2.6E8	5.3E8	8.8E8	1.2E9	1.3E9	1.2E9
21	0	1.5E2	8.0E3	1.7E5	1.8E6	1.3E7	6.4E7	2.3E8	6.6E8	1.5E9	2.7E9	4.1E9	5.0E9	5.1E9
22	1.0E0	1.7E2	1.1E4	2.5E5	3.1E6	2.4E7	1.3E8	5.2E8	1.6E9	4.0E9	8.0E9	1.3E10	1.8E10	2.1E10
23	0	2.0E2	1.4E4	3.7E5	5.0E6	4.3E7	2.6E8	1.1E9	3.8E9	1.0E10	2.3E10	4.1E10	6.2E10	7.8E10
24	1.0E0	2.3E2	1.8E4	5.4E5	8.2E6	7.7E7	5.0E8	2.4E9	8.9E9	2.6E10	6.2E10	1.2E11	2.0E11	2.8E11
25	0	2.6E2	2.4E4	7.8E5	1.3E7	1.3E8	9.5E8	5.0E9	2.0E10	6.4E10	1.7E11	3.6E11	6.4E11	9.7E11

$n\backslash 2m$	30	32	34	36	38	40	42	44	46	48	50	52		
14	1.9E3													
15	3.7E4	4.3E3												
16	4.3E5	9.4E4	1.0E4											
17	3.7E6	1.2E6	2.4E5	2.5E4										
18	2.6E7	1.1E7	3.2E6	6.2E5	6.1E4									
19	1.6E8	8.2E7	3.2E7	8.8E6	1.6E6	1.5E5								
20	8.8E8	5.3E8	2.5E8	9.1E7	2.4E7	4.2E6	3.7E5							
21	4.3E9	3.0E9	1.7E9	7.6E8	2.6E8	6.7E7	1.1E7	9.1E5						
22	2.0E10	1.6E10	1.0E10	5.4E9	2.3E9	7.7E8	1.8E8	2.9E7	2.3E6					
23	8.3E10	7.4E10	5.5E10	3.4E10	1.7E10	7.0E9	2.2E9	5.1E8	7.7E7	5.7E6				
24	3.3E11	3.3E11	2.8E11	1.9E11	1.1E11	5.5E10	2.1E10	6.4E9	1.4E9	2.0E8	1.4E7			
25	1.2E12	1.4E12	1.3E12	1.0E12	6.7E11	3.8E11	1.7E11	6.4E10	1.8E10	3.9E9	5.4E8	3.7E7		

derived graph invariants with some chemical meaning were examined. The invariants considered included: numbers of C atoms with different types of bonding patterns to C atoms (there being a total of 10 such types); graph-theoretic diameter; degree of branching; counts of bonds of different multiplicities; and counts of sites of different hybridizations. The final resultant generating functions involved in this process and further used here are briefly reviewed in section 2. Here these results are used to estimate isomer-class averages and standard deviations for more conventional physico-chemical properties: heat of formation; magnetic susceptibility; index of refraction. The manner in which the combinatoric techniques are combined with sub-structural cluster expansions for these properties is discussed in section 3. The numerical results give an overview of the degree of distinction between isomer classes as afforded by the studied physico chemical properties. The property trends and overlaps between different isomer classes are neatly considered in a pictorial manner in our periodic table, as described in section 4. In addition to the consequent "overlap diagrams" for our three molecular properties, similar diagrams are given for two other properties that are often instead termed "structural invariants", namely "branching degree" and "diameter". Overall our periodic table and related overlap diagrams seem to be useful tools for the consideration of the general class of acyclic hydrocarbons. They facilitate the comparison of molecular properties for different classes as well as the estimation of the overlap of properties between classes.

2. Preliminaries and generating functions for graph invariants

The characterization of the various structures and their isomer classes is here developed in terms of a suite of graph invariants. Though not a necessary presumption of our generating-function techniques, we here restrict consideration to *structural* isomers (not distinguishing mirror image or even *cis* and *trans* structures). This restriction is made because this is all that is relevant in the simpler cluster expansions used in section 3 and it is a simpler case to treat. The molecular structures are represented in a "hydrogen-deleted" form where only the carbon atoms and the C–C, C=C and C≡C bonds are explicitly considered. Contraction of the usual symbols C in such formulas leads to what we here call *graphs*, though often in the literature they would be termed "multi-graphs" (as multiple bonds are allowed). For acyclic hydrocarbons the resulting graphical structures are termed *trees* (though sometimes they might be termed "multi-trees"). See, for example Fig. 1. Labels for the different C atom types involve a code (*a*, *b*, *c*) where *a*, *b* and *c* are the respective numbers of single, double and triple CC bonds attached to a given carbon center. For example, the octene isomer in Fig. 1 contains four atoms of the (1,0,0) type and one each of atoms of types (2,0,0), (4,0,0), (0,1,0) and (2,1,0). The full list of 10 codes is given in Fig. 2. Frequently, we further abbreviate one of these codes just to a single symbol, as ξ or ζ.

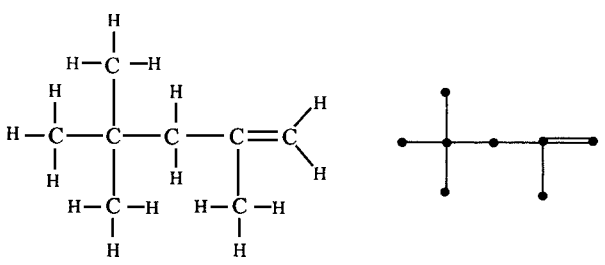

Fig. 1 Example hydrocarbon structure (2,4,4-trimethyl-pent-1-ene), and associated H-deleted graph.

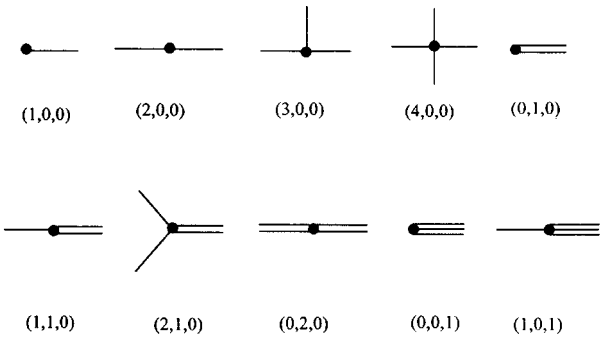

Fig. 2 The 10 types of (C atom) sites.

Here we use three general kinds of generating functions, which result from a detailed development described elsewhere.[14] These involve the dummy variables s and t (which are respectively associated with the counts of H and C atoms). There is just one *zeroth-moment* kind of generating function:

$$P(t, s) \equiv \sum_{n, m} \#_{n, m}\, t^n s^{2m} \tag{1}$$

where $\#_{n, m}$ is the number of structural isomers for class (n, m) (of isomers C_nH_{2m}). The associated counts $\#_{n, m}$ have been given previously (up through the range of Table 1 here) in ref. 13, where their asymptotic behavior is also developed. There are 10 *first-moment* generating functions:

$$P^{(\xi)}(t, s) \equiv \sum_{n, m} \#^{(\xi)}_{n, m}\, t^n s^{2m} \tag{2}$$

where $\#^{(\xi)}_{n, m}$ gives the total count of type-ξ atoms per isomer as summed over all isomers of class (n, m). There are 55 of the *second-moment* generating functions:

$$P^{(\xi, \zeta)}(t, s) \equiv \sum_{n, m} \#^{(\xi, \zeta)}_{n, m}\, t^n s^{2m} \tag{3}$$

where the $\#^{(\xi, \zeta)}_{n, m}$ consist of products of per-isomer numbers of atoms of type ξ and ζ as summed over all isomers of class (n, m). These 66 different generating functions (for 0*th*, 1*st* and 2*nd* moments) are discussed in ref. 14 and in reality are only partially computed, retaining terms only up through the chosen values of n and m (here taken as 25 and 26).

The utility of these generating functions is that their coefficients lead to desired average values for different graph invariants. The numbers n_ξ of C atoms of type ξ per isomer as averaged over all C_nH_{2m} isomers (of fixed n, m) is given by:

$$\langle n_\xi \rangle_{n, m} = \#^{(\xi)}_{n, m} / \#_{n, m} \tag{4}$$

while the product of numbers n_ξ and n_ζ averaged over all C_nH_{2m} isomers is given by:

$$\langle n_\xi n_\zeta \rangle_{n, m} = \#^{(\xi, \zeta)}_{n, m} / \#_{n, m} \tag{5}$$

Finally, different covariances are given as:

$$\begin{aligned} \mathrm{COV}_{n, m}(\xi, \zeta) &= \langle (n_\xi - \langle n_\xi \rangle)(n_\zeta - \langle n_\zeta \rangle) \rangle_{n, m} \\ &= \langle n_\xi n_\zeta \rangle_{n, m} - \langle n_\xi \rangle_{n, m} \langle n_\zeta \rangle_{n, m} \end{aligned} \tag{6}$$

There are two linear dependences amongst the different $\langle n_\xi \rangle_{n, m}$ (ξ ranging) because the total number of bonds is either $n - 1$, if no attention is paid to bond multiplicity, or $2n - m + 1$, if attention is paid to bond multiplicity. As a consequence there are three linear dependences amongst the $\langle n_\xi n_\zeta \rangle_{n, m}$ (not counting the trivial identities $\langle n_\xi n_\zeta \rangle_{n, m} = \langle n_\zeta n_\xi \rangle_{n, m}$). But these dependences play only a small role in the combinatoric manipulations to evaluate these expectations, though these relations may be used as a check on the computations, as we have done in ref. 14.

A major point of this section is that generating-function methods, though not elaborated on here, are capable of being utilized to obtain averages for a variety of graph invariants, especially if the invariants are "local". But it may be mentioned that the condition of locality of graph invariants can seemingly be relaxed since in ref. 12 we have utilized such techniques to compute alkane isomer-class averages for a graph invariant (the "Wiener" index) often considered as being non-local. The generating-function techniques utilized can be viewed as analogous to statistical mechanical techniques for (especially discrete state) partition functions. Here the expectations of the 65 invariants given by eqn. (4) and (6) are used to deal with conventional molecular properties in the following section.

3. Cluster expansions for molecular properties

The present approach to estimate physico-chemical properties is based on the so-called[18–21] *cluster-expansion* method, which for particular "additive" properties is often referred to as a "group-additivity" or "additivity" or "group-function" or "group-increment" or "group-contribution" scheme. In dealing with properties such additivity schemes expand the property as a sum of hypothetical local contributions to the property, where the local property may be for an atom, an atom type (paying attention to the bonding pattern around the atom), a bond type, or even a larger local neighborhood. The general idea of expressing properties in terms of a combination of local contributions goes back to at least the first part of the 1800s. For example, by 1855 H. Kopp[22] summarized his earlier work (dating back to 1839), added some new measurements, and presented his current resolution of molecular volumes into local contributions (which were then for atoms, though the same atom in different functional groups might be distinguished). Kopp also worked extensively on similar approaches for boiling points and specific heats. The introduction in the 1860s of structural formulas engendered possibilities for more systematic extension of the size of contributing local sub-structures. Many investigators throughout the second half of the 1800s studied variations of a great variety of properties along homologous sequences of compounds, attempting to make a resolution into various types of local contributions. Much of this early work is rather thoroughly described in Smiles book[23] of 1910. There has been a great deal of more recent work for a variety of properties as indicated in a standard handbook;[24] for particular properties such as the heat of formation there are even several books or reviews.[25,26] Thus, the cluster expansion references[18–21] simply provide a modern general formalism, which is very widely applicable (even to statistical-mechanical partition functions, molecular wave functions, or model Hamiltonians) and that we choose to utilize in the context of the properties considered here, which include: heats of formation ΔH_f (at $T = 298$ K) following Cohen and Benson;[26] molar refractive indices $Mn_D^{(20)}$ (at 293 K for the sodium D-line) following Vogel *et al.*;[27] and (dia-) magnetic susceptibilities χ_m following Schmalz *et al.*[28]

Now for any one of these properties X the "additive" form of the cluster expansion may be described. Basically for the value X(G) for a molecular structure G (here viewed as a graph) we presume an expansion

$$X(G) = \sum_{\rho} X^{\tau}(\rho) \cdot n_{\rho}(G) \qquad (7)$$

where the sum is over "local" sub-structures (here always some sort of connected subgraph), n_ρ(G) is the number of sub-structures of type ρ occurring in G, and $X^\tau(\rho)$ is a property-dependent (but G-independent) expansion parameter (which may be viewed as the value at ρ of a transform of X). The ρ sum would presumably, if extended to sufficiently large sub-structures, yield arbitrarily accurate results.

In the present case reasonably accurate results seem attainable with the limitation of the sub-structures ρ to just different "types" ξ of C atoms in the H-deleted structures G. Here these *types* simply refer to the bonding patterns to adjacent C atoms as discussed in the previous section (and in Fig. 2). The contributions $\Delta H_f^\tau(\rho)$, $Mn_D^{(20)\tau}(\rho)$, and $\chi_m^\tau(\rho)$ we utilize are given in Table 2. This tabulation for the $\Delta H_f^\tau(\rho)$ correlates closely with that of Cohen and Benson.[26] The tabulated contributions $Mn_D^{(20)\tau}(\rho)$ are not exactly of the same form as in Vogel *et al.*,[27] where the full graph including H atoms is imagined and the only sub-structural contributions are for atoms (C and H), double (=) and triple (≡) CC bonds. But evidently, for all acyclic hydrocarbons the numbers of these 4 substructures are given in terms of the atom-type counts of section 2, thusly

$$\begin{aligned}
n_C &= n_{=C} + n_{=C-} + n_{=C\langle} + n_{C-} + n_{-C-} + n_{-C\langle} + n_{\rangle C\langle} \\
&\quad + n_{=C=} + n_{\equiv C} + n_{\equiv C-} \\
n_H &= 2\, n_{=C} + n_{=C-} + 3\, n_{C-} + 2n_{-C-} + n_{-C\langle} + n_{\equiv C} \\
n_= &= (n_{=C} + n_{=C-} + n_{=C\langle} + 2\, n_{=C=})/2 \\
n_\equiv &= (n_{\equiv C} + n_{\equiv C-})/2
\end{aligned} \qquad (8)$$

where we have suppressed the dependence of the n_ξ(G) on the structure G in as much as these relations are independent of the common argument G. Then the current cluster-expansion coefficients may be expressed in terms of those of Vogel *et al.*,[27] thusly

$$\begin{aligned}
Mn_D^{(20)\tau}(-C) &= Mn_D^{(20)\tau}(C) + 3\, Mn_D^{(20)\tau}(H) \\
Mn_D^{(20)\tau}(=C) &= Mn_D^{(20)\tau}(C) + 2\, Mn_D^{(20)\tau}(H) + Mn_D^{(20)\tau}(=)/2 \\
Mn_D^{(20)\tau}(\equiv C) &= Mn_D^{(20)\tau}(C) + Mn_D^{(20)\tau}(H) + Mn_D^{(20)\tau}(\equiv)/2 \\
Mn_D^{(20)\tau}(-C-) &= Mn_D^{(20)\tau}(C) + 2\, Mn_D^{(20)\tau}(H) \\
Mn_D^{(20)\tau}(=C-) &= Mn_D^{(20)\tau}(C) + Mn_D^{(20)\tau}(H) + Mn_D^{(20)\tau}(=)/2 \\
Mn_D^{(20)\tau}(\equiv C-) &= Mn_D^{(20)\tau}(C) + Mn_D^{(20)\tau}(\equiv)/2 \\
Mn_D^{(20)\tau}(=C=) &= Mn_D^{(20)\tau}(C) + Mn_D^{(20)\tau}(=) \\
Mn_D^{(20)\tau}(=C\langle) &= Mn_D^{(20)\tau}(C) + Mn_D^{(20)\tau}(=)/2 \\
Mn_D^{(20)\tau}(-C\langle) &= Mn_D^{(20)\tau}(C) + Mn_D^{(20)\tau}(H) \\
Mn_D^{(20)\tau}(\rangle C\langle) &= Mn_D^{(20)\tau}(C)
\end{aligned} \qquad (9)$$

To determine the refractive indices $n_D^{(20)}$ one divides $Mn_D^{(20)}$ by the molecular weight of the compound (C_nH_{2m}). For magnetic susceptibilities a somewhat similar modification of the sub-

Table 2 Cluster-expansion coefficients[a] for various properties and C atom types

Type	$Mn_D^{(20)\tau}$	ΔH_f^τ	χ_m^τ
(C–)	18.03	−10.0	14.29
(=C)	17.555	6.27	10.265
(≡C)	16.87	27.1	10.78
(–C–)	20.59	−5.0	11.48
(=C–)	20.115	8.55	7.055
(≡C–)	19.43	27.3	6.97
(=C=)	19.64	34.0	6.39
(=C⟨)	22.675	10.19	4.365
(–C⟨)	23.15	−2.40	9.19
(⟩C⟨)	25.71	−0.10	7.42

[a] $Mn_D^{(20)\tau}$ are in g mol^{-1}, ΔH_f^τ in kcal mol^{-1}, χ_m^τ in m.u.≡-10^{-6} erg G^{-2} mol^{-1}.

graph form used in Schmalz *et al.*[28] is needed, and is given by:

$$\chi_m^\tau(\text{–C}) = \chi_m^\tau(\bullet) + \chi_m^\tau(\bullet\text{–}\bullet)/2$$
$$\chi_m^\tau(\text{=C}) = \chi_m^\tau(\bullet) + \chi_m^\tau(\bullet\text{=}\bullet)/2$$
$$\chi_m^\tau(\equiv\text{C}) = \chi_m^\tau(\bullet) + \chi_m^\tau(\bullet\equiv\bullet)/2$$
$$\chi_m^\tau(\text{–C–}) = \chi_m^\tau(\bullet) + \chi_m^\tau(\bullet\text{–}\bullet) + \chi_m^\tau(\bullet\text{–}\bullet\text{–}\bullet)$$
$$\chi_m^\tau(\text{=C–}) = \chi_m^\tau(\bullet) + [\chi_m^\tau(\bullet\text{–}\bullet) + \chi_m^\tau(\bullet\text{=}\bullet)]/2 + \chi_m^\tau(\bullet\text{=}\bullet\text{–}\bullet)$$
$$\chi_m^\tau(\equiv\text{C–}) = \chi_m^\tau(\bullet) + [\chi_m^\tau(\bullet\text{–}\bullet) + \chi_m^\tau(\bullet\equiv\bullet)]/2 + \chi_m^\tau(\bullet\equiv\bullet\text{–}\bullet)$$
$$\chi_m^\tau(\text{=C=}) = \chi_m^\tau(\bullet) + \chi_m^\tau(\bullet\text{=}\bullet) + \chi_m^\tau(\bullet\text{=}\bullet\text{=}\bullet)$$
$$\chi_m^\tau(\text{=C}\langle) = \chi_m^\tau(\bullet) + \chi_m^\tau(\bullet\text{–}\bullet) + \chi_m^\tau(\bullet\text{=}\bullet)/2 + \chi_m^\tau(\bullet\text{–}\bullet\text{–}\bullet) + 2\,\chi_m^\tau(\bullet\text{=}\bullet\text{–}\bullet)$$
$$\chi_m^\tau(\text{–C}\langle) = \chi_m^\tau(\bullet) + 3\,\chi_m^\tau(\bullet\text{–}\bullet)/2 + 3\,\chi_m^\tau(\bullet\text{–}\bullet\text{–}\bullet)$$
$$\chi_m^\tau(\rangle\text{C}\langle) = \chi_m^\tau(\bullet) + 2\,\chi_m^\tau(\bullet\text{–}\bullet) + 6\,\chi_m^\tau(\bullet\text{–}\bullet\text{–}\bullet) \qquad (10)$$

But the data of Schmalz *et al.*[28] is not complete for our present purposes in that it lacks contributions for triple bonds as well as for cumulated double bonds (as in =C=). These missing values can be estimated from otherwise known values for ethyne, propyne and allene, using the Möbius inversion of refs. 19 and 20 whereupon

$$\chi_m^\tau(\bullet\equiv\bullet) = \chi_m(\text{ethyne}) - 2\,\chi_m^\tau(\bullet)$$
$$\chi_m^\tau(\bullet\equiv\bullet\text{–}\bullet) = \chi_m(\text{propyne}) - 3\,\chi_m^\tau(\bullet) - \chi_m^\tau(\bullet\text{–}\bullet) - \chi_m^\tau(\bullet\equiv\bullet)$$
$$\chi_m^\tau(\bullet\text{=}\bullet\text{=}\bullet) = \chi_m(\text{allene}) - 3\,\chi_m^\tau(\bullet) - 2\,\chi_m^\tau(\bullet\text{=}\bullet) \qquad (11)$$

These three values are found in the theoretical paper of Ruud *et al.*,[29] while ethyne and allene values are found in the experimental paper of Barter *et al.*,[30] though these two sources do not agree with one another. Neither of these papers agrees with more recent experimental results of Oldenziel and Trappeniers[31] on methane and ethene: these more recent experimental results are intermediate between the theoretical results and the earlier experimental results. Ruud *et al.*[29] describe a rescaling of their values to attain a higher degree of agreement with these various experimental data; it is these rescaled values for ethyne, propyne and allene that we use to generate the values of Table 2. Use of the (unscaled) *ab initio* values leads to slightly modified values [$\chi_m^\tau(\equiv\text{C}) = 11.64$, $\chi_m^\tau(\equiv\text{C–}) = 8.67$ and $\chi_m^\tau(\text{=C=}) = 8.55$]. Typically the present cluster expansion for susceptibilities is[28] notably more accurate in comparison to experiment than Pascal's classical rules.[32]

To obtain mean properties for isomer classes of acyclic hydrocarbons the various isomer-class expectations of section 3 are used in conjunction with the present cluster expansions. For a property X with values given as in eqn. (7) an averaging of such expressions (with the limitation of the substructures (to just the 10 described in section 2 and Fig. 2) leads to

$$\langle \text{X} \rangle_{n,\,m} = \sum_{\xi}^{10} \text{X}^\tau(\xi) \cdot \langle n_\xi \rangle_{n,\,m} \qquad (12)$$

These ideas can be extended to obtain standard deviations for the molecular property X. In particular, for the isomer class (n, m) the associated standard deviation $\sigma_{n,\,m}(\text{X})$ for the cluster-expanded property X is

$$\sigma_{n,\,m}(\text{X})^2 = \langle (\text{X} - \langle \text{X} \rangle)^2 \rangle_{n,\,m} = \langle \text{X}^2 \rangle_{n,\,m} - \langle \text{X} \rangle_{n,\,m}^2$$
$$= \left\langle \left[\sum_{\xi} \text{X}^\tau(\xi) \cdot (n_\xi - \langle n_\xi \rangle_{n,\,m}) \right]^2 \right\rangle_{n,\,m}$$
$$= \sum_{\xi,\,\zeta} \text{X}^\tau(\xi)\text{X}^\tau(\zeta) \cdot \text{COV}_{n,\,m}(\xi,\,\zeta) \qquad (13)$$

Thence means and standard deviations for the cluster-expanded values as distributed over (even very large) isomer classes are computable by way of generating functions without dealing separately with each individual isomer.

4. Chemical property trends in the formula periodic table

The formulas and accompanying parameters of the preceding section may be used to fill average values or standard deviations for the considered cluster-expansion-approximated properties into our formula periodic table. But rather than doing this in a numerical fashion we propose to do this in a pictorial fashion. Contour plots in our formula periodic table seem to be meaningful, particularly for averages. The average values of a property are, of course, defined only at the vertex points (n, m) of our table and then they likely do not match precisely the value of a chosen contour. That is, we imagine mean values changing linearly along the vertical or horizontal grid lines from any vertex (n, m) to its adjacent neighbors at ($n \pm 1$, m) or (n, $m \pm 1$), and then draw contours passing through the requisite intermediate values appearing on these grid lines between adjacent vertices. As a matter of convenience these contour lines are drawn as straight line segments joining points on the (undrawn) square grid lines corresponding to the considered value.

The contours for the heat of formation and for the magnetic susceptibility are shown in Fig. 3. Evidently the contours approach a smooth asymptotic pattern, with the two properties giving rise to contours that are roughly orthogonal. The primary variation of the ΔH_f contour values is seen to be with respect to the degree of unsaturation $u = n - m + 1$, and this can be roughly rationalized in that u simply counts (for our acyclics) the number of higher energy π bonds. For the molar susceptibility χ_m the contours seem to follow a count of atoms per isomer, weighting C atoms roughly 3 times more than an H atom. Qualitatively, such a result can be rationalized on the grounds that molar susceptibility is an "additive" property (which indeed in Pascal's classical work[32] is considered to be additive atom by atom). A contour plot could be made for the index of refraction, but we can imagine qualitatively what it must be like. Since we anticipate all indices of refraction to be between 1 and 2, the averages $\langle n_D \rangle_{n,\,m}$ should approach constants with increasing distance from the origin ($n = m = 0$) of our table, and the contours (for n_D) should therefore fan out from the origin. Indeed this is qualitatively correct, and will be quantitatively displayed in a later figure. If one considered a magnetic susceptibility per unit of mass, a similar argument

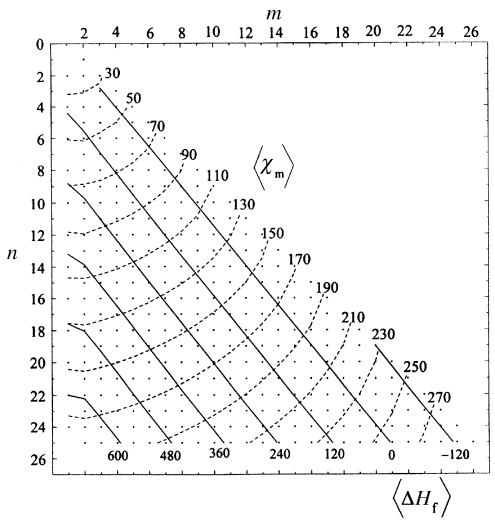

Fig. 3 Contour plots for the heat of formation and magnetic susceptibility. ΔH_f are in $-$kcal mol^{-1}; χ_m are in "magnetic units" -10^{-6} erg G^{-2} mol^{-1}.

leads one to anticipate that a similar fan-shaped set of contours should result. Contour plots for standard deviations could also be made, but it seems to us that something more informative can be done.

The standard deviations of these different properties reveal something about the extent to which properties for structures of one isomer class may overlap those of other isomer classes. Standard deviation contours would not directly reveal the overlap of the $\Delta H_f(G)$ distributions for structures G of adjacent isomer classes, because such a comparison really involves referencing standard deviations against differences of means between these different distributions (for neighboring isomer classes). Thus, we propose a type of property *overlap plot* to be made on our periodic table. This is done with a suitable depiction at various points along an average-property contour so as to show the extent of the standard deviation (at that point) referenced against the inter-contour spacing. This depiction is naturally just a line drawn normal to the contour and of a length proportional to the value of the standard deviation at that point and inversely proportional to the inter-contour property difference. That is, the length of the normal line is

$$L_P(X) \equiv \sigma_P(X) \cdot \Delta l_P/\Delta X_P \qquad (14)$$

where $\sigma_P(X)$ is the standard deviation at the considered point P, Δl_P is the normal distance along the normal from P to the next contour, and ΔX_P is the property difference between the two adjacent contours. These normal lines then generally indicate the distribution of property values from the local average on a contour toward nearby averages (*e.g.*, at an adjacent contour).

The resultant overlap plot for the case of heats of formation is given in Fig. 4. It is seen that there is rather little overlap on a scale corresponding to differences between neighboring isomer classes separated from one another in the normal direction, though there is a general increase in the extent of overlap as the degree of unsaturation u increases and also a seemingly slower increase as n increases (at constant u). The overlap plot for susceptibilities in Fig. 5 shows somewhat greater extents of overlap, but still there is (for the portion of the table considered) only notable overlap with neighbor classes separated along the (diagonal) normal direction. For the indices of refraction the overlap plot is shown in Fig. 6; it is seen that there is a notable variation in the extents of overlap, which are especially small near the alkane diagonal and increase with degree of unsaturation u. Indeed, for small u the overlaps are so small that they are not perceptible on the plot. Notably the angles are large (*i.e.*, close to 90°) between contours for $\langle\chi_m\rangle$ and the contours for either $\langle\Delta H_f\rangle$ or $\langle n_D^{(20)}\rangle$, while at the same time overlap distances are small compared to neighbor separations. Thence specification of

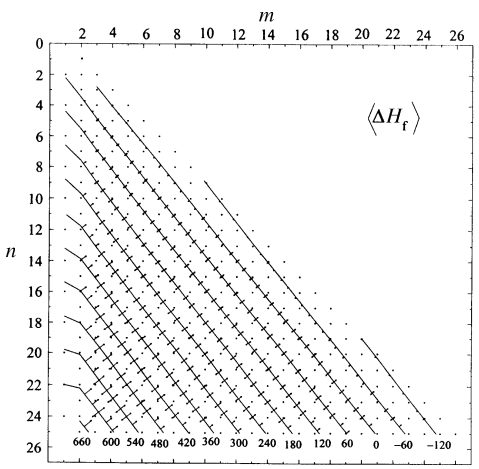

Fig. 4 Heat-of-formation overlap plot, with units of kcal mol^{-1}. The scaled values of Table 2 are used.

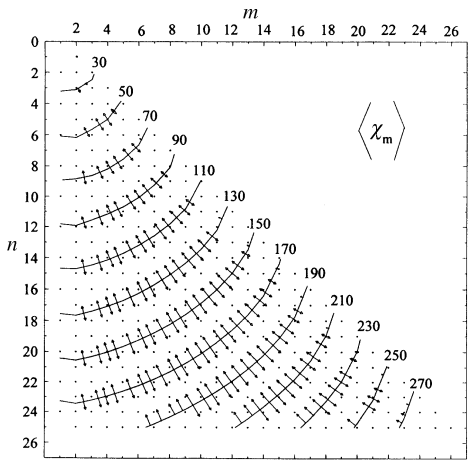

Fig. 5 Magnetic-susceptibility overlap plot, with magnetic units of -10^{-6} erg G^{-2} mol^{-1}.

either of the pairs of values $[\chi_m(G), \Delta H_f(G)]$ or $[\chi_m(G), n_D(G)]$ for a structure G tends to localize G in a single isomer class (at least as long as the structure is not too large). For ever larger structures we anticipate (in correspondence with findings for the alkanes[12]) that averages increase as $\approx n$, while standard deviations should increase as $\approx n^{1/2}$. Thus, as the number n of carbons increases, contour spacings should approach constancy, the extent of overlap should increase, and the localization to a particular isomer class as dictated by a set of property values should become more ambiguous. Generally, specification of $[\chi_m(G), \Delta H_f(G)]$ or $[\chi_m(G), n_D(G)]$ yields more information than is specified by (n, m), though for small u in our considered portion of the periodic table, one sees from the small sizes of the overlap that (n, m) provides about the same amount of information on the structure.

It is to be emphasized that the averages and standard deviations for the various property plots here are predicated on the (truncated) cluster expansion approximation of section 3. These approximations are not always adequate, or else may have contributions at higher orders, as illustrated for certain properties in ref. 33. In particular, if there is much branching, then "steric hindrance" becomes relevant, and in the context of cluster expansions may be taken into account through the introduction of contributions for (selected) larger substructures. Indeed, for ΔH_f this is often recommended and the associated contribution values are described by Cohen and Benson,[26] but the combinatoric generating-function manipulations to implement our present averages become even more challenging; thus we have not tried to include such additional terms in this work. Notably steric hindrance of overly

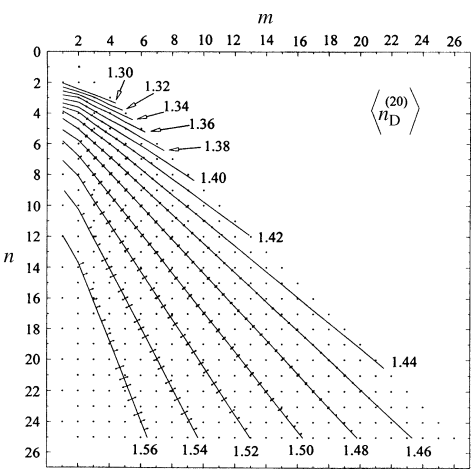

Fig. 6 Index-of-refraction overlap plot.

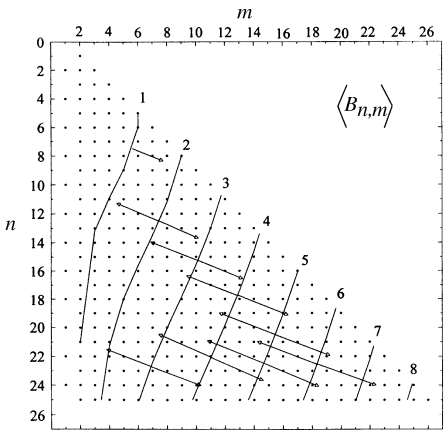

Fig. 7 Degree-of-branching overlap plot.

branched structures eventually precludes[34] their experimental isolatability, whereupon even the classical Cayley–Polya enumerations may be viewed as eventually in error.

Overlap plots may, of course, be given of other properties of acyclics. Here we consider two that should correlate to some degree with the extent of steric hindrance, and which have been previously dealt with as isomer-class averages.[14] The first is the *branching degree* defined as

$$B \equiv n_{=C\langle} + n_{-C\langle} + 2\, n_{>C\langle} \qquad (15)$$

(with the common argument G suppressed). The second such graph invariant is the *graph diameter D* defined as the longest chain of CC bonds (independent of multiplicity) occurring in an acyclic structure G—this invariant being of key relevance in the initial step of standard (IUPAC) nomenclature. If B(G) is small, then the structure G should be rather spatially extended (*i.e.*, non-compact), whereas if D(G) is small (compared to n), then the structure must be spatially compact. And of course the more compact a structure the greater the steric hindrance. The resultant overlap plots for B and D are shown in Figs. 7 and 8. The extents of overlap for these two invariants are much longer than for the molecular properties shown in Figs. 4–6, thus in Figs. 7 and 8 we display only a few standard-deviation lines. Though the contours for these two invariants are nearly orthogonal, the pair of values B(G) and D(G) for a structure G would tend to localize G to within a dozen or so isomer classes only (for the portion of the periodic table considered). Again the extents of overlap should generally increase with n. Clearly, not all properties are as localized as ΔH_f, χ_m and $n_D^{(20)}$. Because fluctuations associated with steric hindrance are less localized, one may surmise that the overlaps for molecular properties with steric hindrance

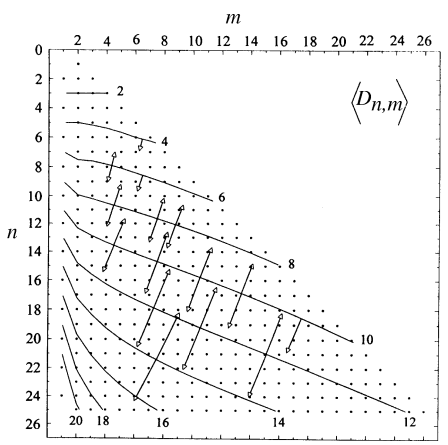

Fig. 8 Graph-diameter overlap plot.

incorporated may be larger than the values we have obtained (without inclusion of steric hindrance).

5. Conclusions

A formula periodic table for the isomer classes of all acyclic hydrocarbons has been proposed, with rows and columns respectively specifying numbers of C and H atoms. Different average characteristics of the different isomer classes are pictorially represented in this periodic table to reveal trends for heats of formation, magnetic susceptibilities, and indices of refraction. The generating-function techniques used to obtain these results are seemingly quite powerful, here dealing with classes of over 10^{12} structural isomers, though considerably greater numbers could be similarly dealt with, and other even more diverse classes of structures could be similarly treated. The property trends revealed indicate not only trends of averages but also mean extents of property distributions for each isomer class. The contours of Figs. 3–8 seem to us to be qualitatively rationalizable, but are of course quite quantitative. The extent to which distributions of various properties overlap between different locations in our periodic table is revealed by the strandard-deviation normals, appearing in Figs. 4–8. Beyond some quantitative information, it is seen that pairs of properties (χ_m, ΔH_f) or (χ_m, n_D) tend to localize a structure within a single isomer class. This may be contrasted with the distributions for two common graph invariants (branching B and diameter D), which are seen to be so broad as to exhibit a great degree of overlap between different isomer classes in all directions in the formula periodic table. That is, for the molecular properties the distributions are seen to be tighter, and the more so the lower the degree of unsaturation. Evidence is noted that there should be further attention given to effects of steric hindrance. Though we have argued that these effects should correlate with B and D, a parameterization to attempt to quantitatively augment the simple cluster expansion considered here is a matter for possible future work.

Overall, the present work provides an illustration of the beginnings of a powerful combinatoric chemical approach for comparing and screening different diverse isomer classes, with reference to a variety of properties, even when there are astronomically large numbers of isomers in individual classes. The approach goes notably beyond the classical enumerative considerations. Indeed, the fundamentals illustrated here may be viewed as a theoretical analog to the recent[35] widely acclaimed and experimentally based combinatorial chemistry. Our characterizations of even huge isomer classes might be utilized to supplement computer-based[36] combinatorial libraries, which typically treat each considered substance individually. Notably, our combinatoric analysis is capable not only of identifying averages and standard deviations for properties, but also extreme property values (within the approximation of the cluster expansion) and associated structures (as illustrated for the simpler case of alkanes in ref. 12). The possibilities for future interesting work and insight seem promising.

Acknowledgements

The authors acknowledge support from the Welch Foundation of Houston, Texas.

References

1 Z. Slanina, *Contemporary Theory of Chemical Isomerism*, D. Reidel Pub., Dordrecht, 1986, ch. 2.

2 A. von Humboldt, *Versuche uber die gereizte Muskel- und Nervenfaser, nebst Vermutungen uber den chemischem Prozess in der Tier- und Pflanzenwelt*, Berlin, 1797. A nice discussion of this work, including reproductions of the relevant text, appear in A. Kerber, *MatCh*, 1999, **39**, 127. Most histories as in ref. 1 overlook Humboldt and instead begin with J. J. Berzelius (*Ann. Phys. Chem.*, 1830, **19**, 305), who introduced the word *isomer*.

3 C. Brown, *Trans. R. Soc. Edinburgh*, 1864, **23**, 707.
4 (*a*) A. Cayley, *Philos. Mag.*, 1874, **47**, 444; (*b*) A. Cayley, *Rep. Br. Assn. Adv. Sci.*, 1887, 257.
5 H. R. Henze and C. M. Blair, *J. Am. Chem. Soc.*, 1931, **53**, 3042 and 3077; C. M. Blair and H. R. Henze, *J. Am. Chem. Soc.*, 1932, **54**, 1098 and 1538.
6 (*a*) G. Polya, *Acta Math.*, 1937, **68**, 145; (*b*) G. Polya, *Z. Kristallogr., Sect. A*, 1936, **93**, 414; (*c*) G. Polya, *Vierteljschr. Naturforsch. Ges.* (*Zurich*), 1936, **81**, 243.
7 G. Polya and R. C. Read, *Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds*, Springer–Verlag, Berlin, 1987.
8 S. Fujita, *Symmetry and Combinatorial Enumeration in Chemistry*, Springer–Verlag, Berlin, 1991.
9 N. Trinajstiċ, S. Nikoliċ, J. V. Knop, W. R. Muller and K. Szymanski, *Computational Chemical Graph Theory: Characterization, Enumeration, and Generation of Chemical Structures by Computer Methods*, Horwood/Simon & Schuster, New York, 1991.
10 F. Harary and E. Palmer, *Graphical Enumeration*, Academic Press, New York, 1973.
11 K. Balasubramanian, *Chem. Rev.*, 1985, **85**, 599.
12 (*a*) L. Bytautas and D. J. Klein, *J. Chem. Inf. Comput. Sci.*, 1998, **38**, 1063; (*b*) L. Bytautas and D. J. Klein, *J. Chem. Inf. Comput. Sci.*, 1999, **39**, 803.
13 L. Bytautas and D. J. Klein, *Croat. Chem. Acta*, in press.
14 L. Bytautas and D. Klein, *Phys. Chem. Chem. Phys.*, 1999, **1**, 5565.
15 M. Randic and N. Trinajstiċ, *Croat. Chem. Acta*, 1987, **67**, 1.
16 R. C. Read, in *Graph Theory and Applications*, ed. Y. Alavi, D. R. Lick and A. T. White, Springer–Verlag, Berlin, 1972, pp. 243–259; R. C. Read, in *Chemical Applications of Graph Theory*, ed. A. T. Balaban, Academic Press, New York, 1976, pp. 25–61.
17 (*a*) J.-R. Dias, *Acc. Chem. Res.*, 1985, **18**, 241; (*b*) J.-R. Dias, *J. Math. Chem.*, 1990, **4**, 17; (*c*) J.-R. Dias, *Polycycl. Arom. Compd.*, 1994, **4**, 87.
18 (*a*) E. A. Smolenskii, *Russ. J. Phys. Chem.* (*Engl. Transl.*), 1964, **35**, 700; (*b*) M. Gordon and J. W. Kennedy, *J. Chem. Soc., Faraday Trans. 2*, 1973, **69**, 484.
19 (*a*) J. W. Essam, M. Gordon, J. W. Kennedy and P. Whittle, *J. Chem. Soc., Faraday Trans. 2*, 1977, **73**, 1289; (*b*) J. W. Kennedy and M. Gordon, *Ann. N. Y. Acad. Sci.*, 1979, **319**, 331.
20 D. J. Klein, *Int. J. Quantum Chem., Symp.*, 1986, **20**, 153.
21 (*a*) M. I. Skvortsova, I. I. Baskin, O. L. Slovokhotova and N. S. Zefirov, *Dokl. Chem.* (*Engl. Transl.*), 1994, **336**, 496; (*b*) I. I. Baskin, M. I. Skvortsova, I. V. Stankevich and N. S. Zefirov, *Dokl. Chem.*(*Engl. Transl.*), 1994, **359**, 346; (*c*) W. Hässelbarth, *MatCh*, 1994, **31**, 7.
22 H. Kopp, *Poggendorff's Ann. Phys.*, 1855, **96**, 153 and 303.
23 S. Smiles, *The Relations between Chemical Constitution and some Physical Properties*, Longman, Green & Co., London, 1910.
24 *Handbook of Chemical Property Estimation Methods*, ed. W. J. Lyman, W. F. Reehl and D. H. Rosenblatt, McGraw–Hill, New York, 1982.
25 V. M. Tatevskii, V. A. Benderskii and S. S. Yarovoi, *Rules and Methods for Calculating the Physico-Chemical Properties of Paraffinic Hydrocarbons* (*Engl. Transl.*), Pergamon Press, Oxford, 1961; J. D. Cox and G. Pilcher, *Thermochemistry of Organic and Organometallic Compounds*, Academic Press, New York, 1970; S. W. Benson, *Thermochemical Kinetics*, John Wiley & Sons, New York, 1976.
26 N. Cohen and S. W. Benson, *Chem. Rev.*, 1993, **93**, 2419.
27 (*a*) A. I. Vogel, *J. Chem. Soc.*, 1948, 1833; (*b*) A. I. Vogel, W. T. Cresswell, G. H. Jeffery and J. Leicester, *J. Chem. Soc.*, 1952, 514.
28 T. G. Schmalz, D. J. Klein and B. L. Sandleback, *J. Chem. Inf. Comput. Sci.*, 1992, **32**, 54.
29 K. Ruud, H. Skaane, T. Helgaker, K. L. Bak and P. Jorgensen, *J. Am. Chem. Soc.*, 1994, **116**, 10135.
30 C. Barter, R. G. Meisenheimer and D. P. Stevenson, *J. Phys. Chem.*, 1960, **64**, 1312.
31 J. G. Oldenziel and N. J. Trappeniers, *Physica A*, 1976, **82**, 565 and 581.
32 P. Pascal, *Ann. Chim. Phys.*, 1910, **19**, 5.
33 M. C. McHughes and R. D. Poshusta, *J. Math. Chem.*, 1990, **4**, 227.
34 (*a*) D. J. Klein, *J. Chem. Phys.*, 1981, **75**, 5186; (*b*) P. G. de Gennes, *J. Phys. Lett.*, 1983, **44**, 351; (*c*) G. R. Newkome, V. K. Gupta, G. R. Baker and Z.-Q. Yao, *J. Org. Chem.*, 1985, **50**, 2003.
35 See, for example, any issue of *J. Comb. Chem.*, or M. A. Gallop, R. W. Barrett, W. J. Dower, S. P. A. Fodor and E. M. Gordon, *J. Med. Chem.*, 1994, **37**, 1233; R. N. Zuckermann, E. J. Martin, D. C. Spellmeyer, G. B. Stauber, K. R. Shoemaker, J. M. Kerr, B. M. Figliozzi, D. A. Goff, M. A. Siani, R. J. Simon, S. C. Banville, E. G. Brown, L. Wang, L. S. Richter and W. H. Moos, *J. Med. Chem.*, 1994, **37**, 2678; *Combinatorial Chemistry*, ed. A. W. Czarnik and S. H. DeWitt, American Chemical Society, Washington, DC, 1997.
36 See, for example, G. Grassy, B. Calas, A. Yasri, R. Lahana, J. Woo, S. Iyer, M. Kaczorek, R. Floc'h and R. Buelow, *Nature Biotech.*, 1998, **16**, 748; L. Xue, J. W. Godden and J. Bajorath, *J. Chem. Inf. Comput. Sci.*, 1999, **39**, 881.