

Alkane Isomer Combinatorics: Stereostructure Enumeration and Graph-Invariant and Molecular–Property Distributions

L. Bytautas^{†,‡} and D. J. Klein^{*,†}

Texas A&M University at Galveston, Galveston, Texas 77553-1675, and Institute of Theoretical Physics & Astronomy, Gostauto 12, 2600 Vilnius, Lithuania

Received March 11, 1999

Combinatoric manipulative methodology to characterize different isomer-class features is extended and illustrated in application to alkane isomers. Stereoisomeric enumerations and various molecular–property characteristics for structural-isomer classes are made. For a range of carbon-atom counts N we enumerate the following: the mean number of diastereomers per N -carbon structural isomer; the mean number of stereoisomers per N -carbon diastereomer; and the mean number of “conformers” per N -carbon stereoisomer. Also we consider distributions of several graph-theoretic properties, characterizing the distributions in terms of averages, variances, covariances, and extremal values. The results for these graph-theoretic invariants are used to predict (via substructural cluster expansion) some molecular properties (heats of formation and magnetic susceptibilities) averaged over these classes. The associated proper standard deviations as well as the extremal values for these cluster-expansion approximants are also obtained, and correspondent extremal structures are identified. Thence combinatoric techniques to deal neatly with astronomical numbers (even if greater than a mole) of different molecular structures and some associated physical and chemical properties are developed.

1. INTRODUCTION

The classic area of chemical isomerism has a rich history briefly reviewed in ref 1, and in addition there is a long-studied prototypical subarea of alkane isomerism. The early formal treatments focused on alkane structural isomer enumeration,^{2–4} including some consideration⁴ of related derivatives, with one or two substitutions someplace in the network. Pólya developed^{5,6} a general formalism for substitutional isomerism on a fixed skeleton but also applied⁶ the ideas to alkane isomers. Later work considered too the case of stereoisomerism^{7–9} of alkanes, and explicit enumerations up to $N = 25$ carbon atoms have been recorded by Read.⁹ Even more recently, Davies and Freyd¹⁰ concentrated on “astronomical” counts of alkane structural isomers and stereoisomers, tabulating results (in a truncated form) for selected N values up to $N = 400$. Several general reviews of isomer enumeration techniques are available,¹¹ and many combinatorics texts include a chapter or two on the general mathematical theory (usually without much reference to the original chemical motivations^{5,6} of the fundamental theory). The mathematics book of Harary and Palmer¹² is devoted to the graph-theoretic applications (again without much explicit reference to chemistry). In general, as it was pointed out by Slanina,¹ there exist different algebraic methods that can be used for enumeration and characterization of isomers, involving set theory, group theory, graph theory, and information theory. Furthermore, some of these approaches, for instance, like the one based on the group theory,¹³ can be extended to the enumeration of degenerate reaction

processes that link “equivalent” minima on potential energy surfaces.¹⁴ Formalism to deal with isomer-class properties seems perhaps to have started with consideration of their graph-theoretic diameters,^{15,16} though much of this focused more on the use of diameter as a means of subcategorization of the isomer class (such as is also more or less implicit in standard nomenclatural considerations, where the base name corresponds to a backbone diameter). Kennedy, Quintas, and co-workers¹⁷ pursued a similar idea with respect to vertex degrees. In recent work¹⁸ rather than further subcategorizing isomer classes, we have considered structural-isomer-averaged numbers of atoms of different types (primary, secondary, tertiary, and quaternary), as well as similarly averaged values for the graph diameter. Further these graph-theoretic averages were used along with appropriate substructural cluster expansions to give computed alkane isomer-class averaged molecular properties such as heats of formation, magnetic susceptibilities, and indices of refraction.

Extension of these earlier combinatoric manipulations would be desirable to characterize yet further different isomer classes. What about going beyond structural isomers to stereogeometric considerations? What about a more complete characterization of the distributions of graph invariants and molecular properties?

Here we make such extensions, thereby providing further isomer-class characterization. For a range of carbon-atom counts N an elaboration is made of some stereostructural enumerations. We enumerate the following: the mean number of diastereomers per structural isomer; the mean number of stereoisomers per diastereomer; and the mean number of “conformers” per stereoisomer. And by related techniques we compute graph-theoretic properties averaged

[†] Texas A&M University of Galveston.

[‡] Institute of Theoretical Physics & Astronomy.

over the members of structural-isomer classes. Here use is made of one standard nomenclature, where *structural isomerism* attends solely to the (graph-theoretic) connectivity pattern, *stereomerism* attends to the stereochemistry of the tetrahedral carbon atom, and *diastereomerism* attends to the stereostructure excepting that it does not distinguish overall mirror-image enantiomers. That is, diastereomerism would not distinguish a stereoisomer with the same two chiral left-handed ligands attached to a carbon from the stereoisomer with two corresponding right-handed ligands attached—but it would distinguish these from the stereoisomer with one right- and one left-handed ligand attached. The diastereomer enumeration is implicit in the work of Robinson et al.,⁸ but here it is differently approached (in a “diameter-attentive” manner), it is made to notably higher atom counts N , and it is reported in a different manner (in terms of diastereomer counts per structural isomer). The diastereomer count is quite natural in that it gives the number of conceivable chemically isolatable species having differing “scalar” properties (as various heats, condensation points, phase diagrams, and optical properties for unpolarized light). That is, the diastereomer count is the number of different sets of scalar properties for pure substances—and it identifies the number of different substances obtainable via purely asymmetric syntheses (and separations). Of course, different diastereomers for the same structural isomer should have a number of such scalar properties very similar, in that many approximate (but presumably fairly accurate) computational methods do not distinguish such structures—e.g., heat of formation approximations in terms of bond energies or of atom types or of bond types. The property approximations we use indeed are in terms of substructural features which do not differentiate diastereomers of the same structural isomer, so that for these properties we currently restrict attention to averages over structural-isomer classes. Averages of a sufficient number of graph invariants are computed to yield not only property averages but also standard deviations as well as “discorrelated” (and “absoluted”) standard deviations for these properties. Discorrelated standard deviations are basically defined like standard deviations but under the assumption that there is no correlation between the different substructure counts involved in estimating the various properties. The overlap of predicted property values (for ΔH_f and χ_{mag}) between successive structural-isomer classes turns out to be quite slight when judged from the proper standard deviations. This result contrasts with the more severe overlap estimate¹⁸ using the more easily computable “absoluted” standard deviations providing us with the upper and lower bounds which in the cases considered are comparable in magnitude to the “discorrelated” standard deviations. That is, the intra-isomer-class fluctuations in the properties here computed are rather strongly quenched because of (largely cancellative) cross-correlations between the different graph invariants (atom-type counts) in which they are expanded. As a further type of novel result we identify extreme (maximum and minimum) property values achievable via our cluster expansions for these properties, and associated extremal structures are identified too. The averages, the standard deviations, and extreme values thence yield a nice characterization of the distribution of properties within an isomer class and reveal rather little overlap between classes. Thus, the isomer classes are characterized not only in terms

of their chemical formula but also to a large extent in terms of their properties.

The enumerations and property averages are developed via fairly standard “generating-function” techniques, as illustrated in the next section, for the case of stereoisomers. In such a framework it is imagined that certain counts (as of alkane isomers) are collected together with each count appearing as a coefficient of corresponding monomial powers of a “dummy” variable in the generating function. The use of generating functions is much more general than Pólya enumeration theory and is often used both in combinatorics and in statistical mechanics. The manipulations are presented with a minimum of mathematical overhead. Modifications to deal with diastereomers and conformation counts are presented in section 3. Section 4 concentrates on the numerical enumerative results based on the two previous sections. In section 5 we develop the averages for second moments of atom-type counts, such as are utilized in section 6 to obtain standard deviations for the isomer-averaged properties. The formal definition of the “discorrelated” standard deviations along with an “absoluted” standard deviation too is made, and numerical relations to the proper standard deviations are considered. Section 7 addresses another aspect of the isomer-class property distributions without the use of the generating functions (but rather with methods more like “linear programming”). There we identify the extreme cluster-expansion values for the considered properties as well as associated “extremal structures”. Numerical results for property values are reported in section 8 to reveal “tight” correlations of properties to the structural-isomer class. Section 9 concentrates on asymptotic behavior for the quantities considered previously. Overall a promising theoretical combinatoric chemistry is developed and illustrated.

2. STEREOISOMER ENUMERATION

As already mentioned, alkane enumeration has long been a standard theoretical problem, with several tabulations of isomer counts available.^{2–4,8–11} The generating-function formalism too has been repeatedly developed, though less often^{15–20} in the manner considered here relating to generation numbers and (through-bond) diameters. Indeed though stereoisomers have previously been treated, this diameter-attentive approach seems not to have been so considered previously. In all these treatments the enumeration is first made of alkyl radicals, which are possibly “decorated”. Graph-theoretically speaking, alkyl radicals are most simply identified as *rooted trees* (each defined as a tree with a distinguished site called a *root*, which represents the radical site with a dangling bond). The graphs are “hydrogen-deleted” in the sense that only the carbon atoms and the CC bonds are explicitly represented. Now for alkanes with the degree of a site restricted to no more than 4, an n -site rooted tree may be obtained from three smaller rooted a -, b -, and c -site rooted trees (where $a + b + c = n - 1$ and $a, b, c \geq 0$) by a process as indicated in Figure 1. Once the various rooted trees are obtained, they are then joined together to form alkanes, as shown in Figure 2. For such rooted trees one may attend to *generation number* g defined as the number of carbon sites in the longest path from the root.

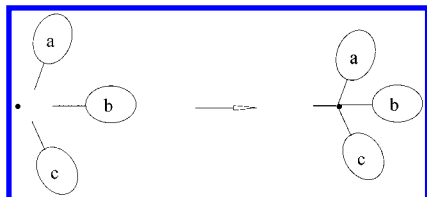


Figure 1. Illustration of the constructive manner considered for the iterative generation of alkyl radicals. Here three alkyl radicals (denoted *a*, *b*, and *c*, with the possibility that some of them might be the zero-order alkyl radical consisting of a single H-atom with a dangling bond) are joined at a new radical site.

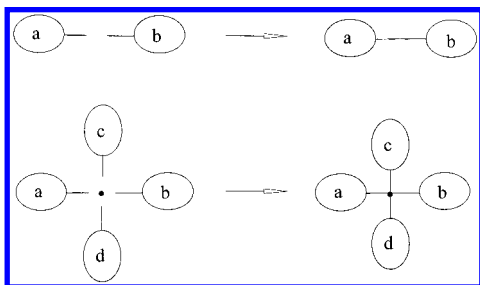


Figure 2. Two manners of joining generation-*g* radicals so as to obtain an alkane: in (a, top) two generation-*g* radicals are joined at a common bond, and in (b, bottom) four radicals are joined at a central site, with the constraint that two are of generation-*g* while the two remaining ones have a generation number $\leq g$.

Exactly what is done with this general idea depends on just what type of isomer one wishes to consider. Our earlier article¹⁸ describes the application for dealing with structural isomers.

To deal with stereoisomers or diastereoisomers, one decorates the graphs so as to attend to the “chirality” or “parity” of the different C-atom centers. That is, one can imagine that a (+) or (−) parity sign is attached to a C-atom site whenever it has four stereomerically different ligands attached thereto (such consideration including an H-atom as a possible ligand, which in our present H-deleted framework is just an absence of a C-based ligand). The (+/−) sign does not attach to a site if it has two or more ligands identical. The constructions for radicals are much like those in Figure 1, and a (+/−) sign is attached to the root (radical) site if none of the three ligands so far attached are the same, it being understood that further (later) attachments to the dangling bond at the radical site end up attaching there a final ligand which exceeds in generation number the three already attached (so that it will be distinct). The resulting sign-decorated graphs might be termed *chirographs*, which (for the nonradical case) have been formally considered before.²¹ Acyclic *chirographs* of concern here then are termed *chiotrees*. The generating function for the then relevant degree-4-limited rooted *chiotrees* of a given generation number *g* is defined as

$$\varphi_g(t) \equiv \sum_{n \geq 0} \#_{\pm}(g, n) t^n \quad (2.1)$$

where $\#_{\pm}(g, n)$ is the number of rooted *chiotrees* with sites of degree ≤ 4 and *t* is a “dummy variable”. The construction of Figure 1 leads to a recursion for these generating functions

$$\varphi_{g+1}(t) = t\{\varphi_g(t) A_2[\varphi_{<g};t] + A_2[\varphi_g;t] \varphi_{<g}(t) + A_3[\varphi_g;t]\} \quad (2.2)$$

where

$$\varphi_{<g+1}(t) = \varphi_{<g}(t) + \varphi_g(t)$$

$$A_2[f;t] = [f(t)]^2$$

$$A_3[f;t] = \{[f(t)]^3 + 2f(t^3)\}/3$$

$$A_4[f;t] = \{[f(t)]^4 + 8f(t)f(t^3) + 3[f(t^2)]^2\}/12 \quad (2.3)$$

with $\varphi_{<1}(t) \equiv 1$, *f* is some function, and $A_4[f;t]$ will be used later. The factor *t* in recursion 2.2 arises because a new site (which is to be the radical site) is added and does not already occur in the *a*-, *b*-, or *c*-groups of Figure 1. The three different terms of (2.2) may be independently rationalized. The first term concerns the case of Figure 1 where the *a*-group is of generation number *g* while the *b*- and *c*-groups have generation numbers $<g$. Such *b*- and *c*-groups are accounted for in the $[\varphi_{<g}(t)]^2$ term of $A_2[\varphi_{<g};t]$, which gives the correct stereoisomer counts, since it counts the structures where *b*- and *c*-groups are distinct twice and the ones where both *b*- and *c*-groups are identical just once. The second term of (2.2) concerns the circumstance in which the *a*- and *b*-groups have generation number *g* while the *c*-group has generation number $<g$, and the rationale for $A_2[\varphi_g;t]$ is much the same as that for $A_2[\varphi_{<g};t]$ in the first term. The third term of (2.2) concerns the case where the *a*-, *b*-, and *c*-groups are all of generation number *g*, and now up to three of these groups may be identical and the different terms of $A_3[\varphi_g;t]$ may be somewhat similarly rationalized, it being understood that for stereoisomers the parity of the site can be recognized so that odd permutations of the groups being connected are distinguished. The functional $A_4[f;t]$ has a similar utility in later manipulations. In any event all these functionals $A_n[f;t]$ are the so-called “cycle-index functions” of Pólya^{5–11} for the “alternating” group A_n of even permutations. We note that the related cycle-index functions for the full permutation group S_n

$$Z_2[f;t] \equiv \{[f(t)]^2 + f(t^2)\}/2$$

$$Z_3[f;t] \equiv \{[f(t)]^3 + 3f(t)f(t^2) + 2f(t^3)\}/6$$

$$Z_4[f;t] \equiv \{[f(t)]^4 + 6[f(t)]^2 f(t^2) + 8f(t)f(t^3) + 3[f(t^2)]^2 + 6f(t^4)\}/24 \quad (2.4)$$

arise in dealing with structural isomers, as in ref 18 and also section 5 here.

The alkyl-radical generating functions at each different generation number may be used to obtain generating functions for the nonradical alkanes. This is imagined from joining different alkyl radicals together at the “center” of an alkane. Here the *center* of a tree or *chiotree* graph may be precisely defined (in a standard fashion) as the last nonzero graphical remnant remaining from successive *prunings*, each pruning entailing the deletion of all end (degree-1) vertices. The joining together of two generation-*g* radicals at a central bond yields an alkane of diameter $2g - 1$, with a tree’s *diameter* being precisely specifiable as the largest number of bond steps between two sites of the tree. If instead the center is a single site, then there being four places of

attachment at a site, four radicals are joined to it with two of the radicals of g generations while the other two have generation number $\leq g$, thereby yielding alkanes of diameter $2g$. These odd- and even-diameter constructions are respectively indicated in Figure 2a,b. With the introduction of a generating function $P_D(t)$ for alkane counts, these center-based constructions then lead to the relations

$$P_{2g-1}(t) = Z_2[\varphi_g; t]$$

$$P_{2g}(t) = t\{(1/2)[\varphi_g(t)]^2 [\varphi_{<g}(t)]^2 + \varphi_g(t^2) \varphi_{<g}(t^2)\} + A_3[\varphi_g; t] \varphi_{<g}(t) + A_4[\varphi_g; t] \quad (2.5)$$

Here for $P_{2g}(t)$ the first, second, and third summands respectively concern center-site constructions (of Figure 2b) with two, three, and four generation- g radicals while there are respectively two, one, and zero radicals of generation number $<g$. The overall stereoisomer generating function for all alkanes then is

$$P_{\text{ster}}(t) \equiv \sum_{D \geq 0} P_D(t) \quad (2.6)$$

The auxiliary functions $\varphi_g(t)$ and $P_D(t)$ are also useful in developing various properties (in section 3). The final generating function when expanded out in powers of t is

$$P_{\text{ster}}(t) = \sum_{N \geq 0} \#_{\text{ster}, N} t^N \quad (2.7)$$

with coefficients $\#_{\text{ster}, N}$ counting the number of N -carbon alkane stereoisomers.

3. DIASTEREOMERS AND CONFORMERS

Enumeration of *diastereomers* implicates the partition of stereoisomer structures into *achiral* and *chiral* subsets. The total stereoisomer count $\#_{\text{ster}, N}$ has a chiral portion $\#_{\text{chir}, N}$, counting structures differing from their mirror images, while the achiral part $\#_{\text{achir}, N}$ counts structures that are the same as their mirror images. Since *diastereomers* by definition (see, e.g., ref 8) are stereoisomers where mirror images are not distinguished, one obtains the proper count of diastereomers by counting a pair of mirror images as a single structure in the count of stereoisomers. Thus, the following discussion is based on the generating functions $F^*(t)$ and $F^\circ(t)$ representing chiral and achiral structures, respectively. Also, we use notation for diastereomer and stereomer generating functions

$$P_{\text{diast}}(t) = F^\circ(t) + (1/2)F^*(t)$$

$$P_{\text{ster}}(t) = F^\circ(t) + F^*(t) \quad (3.1)$$

As before, we deal with the rooted trees first as they serve as the building blocks for the unrooted (complete) structures. Then the recursion relations for rooted tree generating functions that represent achiral structures are

$$\varphi_{g+1}^\circ(t) = t\{\delta_1^\circ(t) + \delta_2^\circ(t) + \delta_3^\circ(t)\} \quad (3.2)$$

where

$$\delta_1^\circ(t) = \{\varphi_g^\circ(t)[\varphi_{<g}^\circ(t^2) + \varphi_{<g}^*(t^2)]\}$$

$$\delta_2^\circ(t) = \{\varphi_{<g}^\circ(t)[\varphi_g^\circ(t^2) + \varphi_g^*(t^2)]\}$$

$$\delta_3^\circ(t) = \{\varphi_g^\circ(t)[\varphi_g^\circ(t^2) + \varphi_g^*(t^2)]\} \quad (3.3)$$

Here a term $t\delta_n^\circ(t)$ accounts for the combining of n radicals of generation number g with $3 - n$ of lesser generation number (as indicated in Figure 1, without reference to generation numbers). Note that all three terms in (3.3) share the same basic form with the only difference among them being in the generation number specification. In the term $\delta_n^\circ(t)$ the first summand corresponds to the rooted tree structure of the form $(-abb \text{ or } -aaa)$ and the second to the structures $(-aBB)$, where lower case letters here represent achiral structures and the upper case chiral ones. For the first case, where the general form is $(-abc)$ the resulting structure will be *achiral* if at least two of the substructures (where the generation number is $<g$ or $=g$) are the same, i.e., if one has $(-abb \text{ or } -aaa)$; otherwise it is *chiral*. The second case is slightly more complex, since here one deals with the chirality of substructures. The general form here is $(-aBC)$, where B and C substructures ($<g$ or $=g$) can have two signs ($+$ or $-$), one at the time. It is clear that the only possibility that the resulting structure ($g + 1$) turns out to be achiral is if two substructures are the same, i.e. if one has $(-aBB)$. Now, there can be four possibilities: $(-aB^+B^+)$, $(-aB^-B^-)$, $(-aB^+B^-)$, and $(-aB^-B^+)$. Reflecting each of them in the plane, one observes that $(-aB^+B^+)$ and $(-aB^-B^-)$ are interchanged, $(-aB^+B^-)$ remains the same, as does $(-aB^-B^+)$ also, but neither of them is superimposable with the others; i.e., all these structures are different. This implies that structures $(-aB^+B^+)$ and $(-aB^-B^-)$ are *chiral*, while structures $(-aB^+B^-)$ and $(-aB^-B^+)$ are *achiral*. The recursion relation for chiral structures is

$$\varphi_{g+1}^*(t) = t\{[\varphi_g(t) A_2[\varphi_{<g}; t] - \delta_1^\circ(t)] + [A_2[\varphi_g; t] \varphi_{<g}(t) - \delta_2^\circ(t)] + [A_3[\varphi_g; t] - \delta_3^\circ(t)]\} \quad (3.4)$$

Now for the unrooted trees with a center bond one has generating functions

$$P_{2g-1}^*(t) = (1/2)\{[\varphi_g^*(t)]^2 + 2\varphi_g^*(t) \varphi_g^\circ(t)\}$$

$$P_{2g-1}^\circ(t) = (1/2)\{[\varphi_g^\circ(t)]^2 + \varphi_g^\circ(t^2) + \varphi_g^*(t^2)\} \quad (3.5)$$

Again, as it was done for rooted trees one can easily understand the terms present in eq 3.5 if one thinks of each chiral substructure (with generation number g) as having either a $+$ or $-$ sign, which algebraically implies that there exists a factor of 2 for the polynomial expansion coefficient for any chiral structure as compared to any achiral structure. For instance the first summand in $P_{2g-1}^*(t)$ of the (3.5) term reflects the circumstance that there are four chiral structures of the form (AB) and only two chiral structures of the form (AA) (i.e., (A^+A^+) and (A^-A^-)), while the achiral term (i.e., (A^-A^+)) is represented by $\varphi_g^*(t^2)$ of $P_{2g-1}^\circ(t)$. Though all the analytical formulas given here rest on such an approach, we shall not give further explanations but rather provide the

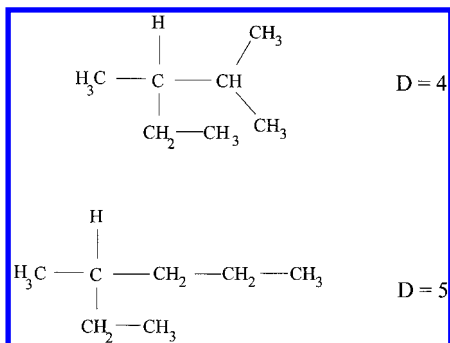


Figure 3. Two alkane structures of diameters $D = 4$ and $D = 5$ corresponding to the smallest chiral trees.

final equations only. Thus, for unrooted trees with a center site one has

$$P_{2g}^{\circ}(t) = t\{\Delta_{(2,2)}^{\circ} + \Delta_{(3,1)}^{\circ} + \Delta_{(4)}^{\circ}\}$$

$$P_{2g}^{*}(t) = t\{(1/2)[\{\varphi_g(t)\}^2 [\varphi_{<g}(t)]^2 + \varphi_g(t^2) \varphi_{<g}(t^2)] - \Delta_{(2,2)}^{\circ}\} + [A_3[\varphi_g;t] \varphi_{<g}(t) - \Delta_{(3,1)}^{\circ}] + [A_4[\varphi_g;t] - \Delta_{(4)}^{\circ}] \quad (3.6)$$

where

$$\Delta_{(2,2)}^{\circ} = \{\varphi_g^{*}(t^2) (\varphi_{<g}^{\circ}(t))^2 + \varphi_{<g}^{*}(t^2) (\varphi_g^{\circ}(t))^2 + \varphi_g^{\circ}(t^2) (\varphi_{<g}^{\circ}(t))^2 + \varphi_{<g}^{\circ}(t^2) (\varphi_g^{\circ}(t))^2\}$$

$$\Delta_{(3,1)}^{\circ} = \{\varphi_g^{*}(t^2) \varphi_g^{\circ}(t) \varphi_{<g}^{\circ}(t) + \varphi_g^{\circ}(t^2) \varphi_g^{\circ}(t) \varphi_{<g}^{\circ}(t)\}$$

$$\Delta_{(4)}^{\circ} = \{\varphi_g^{*}(t^2) (\varphi_g^{\circ}(t))^2 + \varphi_g^{*}(t^4) + \varphi_g^{\circ}(t^2) (\varphi_g^{\circ}(t))^2 + \varphi_g^{\circ}(t^4)\} \quad (3.7)$$

From eqs 3.5, on the one hand, and 3.6, on the other, one sees that their sum properly yields the generating function for stereoisomers given by (2.5). Figure 3 shows the two smallest chiral trees of diameters $D = 4$ and $D = 5$.

Conformers may be viewed as arising from an extreme isomer distinction, where rotation about single bonds is imagined to be “frozen out”. Enumeration of conformers proceeds somewhat analogously to that for structural or stereoisomers, but now with symmetry equivalence being suppressed, so that in the generating-function recurrences the occurrences of $Z_n[f;t]$ or $A_n[f;t]$ collapse simply to $[f(t)]^n$. The recursions are defined elsewhere (ref 18 and also ref 22) and may be viewed as a simple polymer statistics model for branched polymers. Such conformations for a linear chain often are identified to a specification for each nonterminal carbon as being either gauche+, gauche−, or trans—just what is involved in the standard Flory “rotational isomeric model”²³ of polymer statistics. Technically each such conformer for an alkane graph G may be defined as in one-to-one correspondence with a mapping ϕ from G into the tetrahedral lattice L such that first, neighbor pairs of sites of G are mapped to neighbor pairs of sites of L ; second, the center of G is mapped to a fixed position (site or bond) of L ; and third, distinct neighbors of a site of G are mapped to distinct neighbors of L . Viewed in this way such conformations ϕ may be seen to violate “volume exclusion” when two different sites end up mapped to the same site of L . But let us not worry about volume exclusion here.

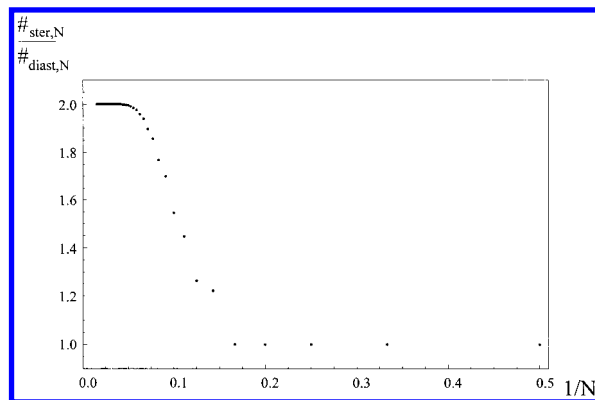


Figure 4. Plot of $\langle \#_{\text{ster}} \rangle_{\text{diast},N}$ versus $1/N$ to examine the asymptotic behavior.

4. NUMERICAL ENUMERATIVE RESULTS

The results of enumeration via our various generating functions are presented in Table 1. There in place of the diastereomer counts $\#_{\text{diast},N}$ we present the component counts $\#_{\text{achir},N}$ and $\#_{\text{chir},N}$ for achiral and chiral isomers (of N carbons), respectively. These achiral and chiral isomer counts are given as the coefficients in $F^{\circ}(t)$ and $F^{*}(t)$, respectively, and yield

$$\#_{\text{diast},N} = \#_{\text{achir},N} + \#_{\text{chir},N}/2 \quad (4.1)$$

Also in Table 1 we give mean diameters $\langle D \rangle_{\text{diast},N}$ and $\langle D \rangle_{\text{ster},N}$ of diastereomer and stereomer isomer classes as can be quite straightforwardly computed using the diameter and generation number based generation-function approach presented here (as elaborated in ref 18, for the case of structural isomers). Also we give the following: the mean number $\langle \#_{\text{diast}} \rangle_{\text{struc},N}$ of diastereomers per N -carbon structural isomer; the mean number $\langle \#_{\text{ster}} \rangle_{\text{diast},N}$ of stereoisomers per N -carbon diastereomer; and the mean number $\langle \#_{\text{conf}} \rangle_{\text{ster},N}$ of “conformers” per N -carbon stereoisomer. These three columns are simply obtained from our enumerations as ratios

$$\text{I} \equiv \langle \#_{\text{diast}} \rangle_{\text{struc},N} = \#_{\text{diast},N} / \#_{\text{struc},N}$$

$$\text{II} \equiv \langle \#_{\text{ster}} \rangle_{\text{diast},N} = \#_{\text{ster},N} / \#_{\text{diast},N}$$

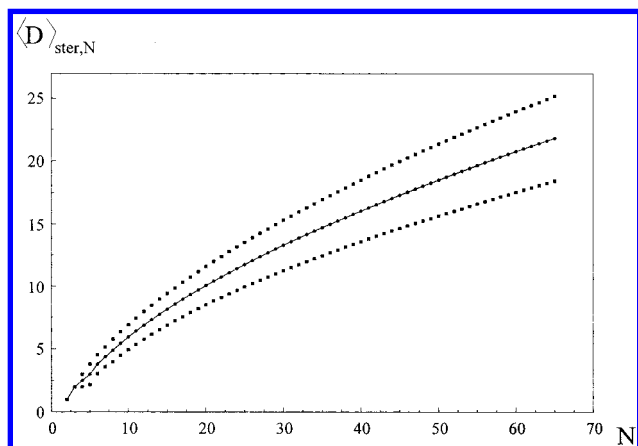
$$\text{III} \equiv \langle \#_{\text{conf}} \rangle_{\text{ster},N} = \#_{\text{conf},N} / \#_{\text{ster},N} \quad (4.2)$$

These ratios represent relative increases in distinction power as one proceeds from one level of isomerism to the successively more refined level.

Examination of Table 1 reveals several points. It seems that the mean counts $\langle \#_{\text{diast}} \rangle_{\text{struc},N}$ and $\langle \#_{\text{conf}} \rangle_{\text{ster},N}$ increase exponentially with N , whereas $\langle \#_{\text{ster}} \rangle_{\text{diast},N}$ remains bounded. Indeed one must have $\langle \#_{\text{ster}} \rangle_{\text{diast},N}$ between 1 and 2 as these are the only possible values for the number of stereomers per individual diastereoisomer, and as $N \rightarrow \infty$, we may expect that most stereoisomers are distinct from their mirror image so that then $\langle \#_{\text{ster}} \rangle_{\text{diast},N} \rightarrow 2$. This trend seems to be supported by the plot in Figure 4 where this ratio is plotted versus $1/N$. The two columns of Table 1 giving mean diameters for the diastereomers and for the stereoisomers supplement the earlier given ref 18 mean diameters for structural isomers. All three diameters turn out to be within 1% of one another. Figure 5 shows the mean diameter of stereoisomers as a function of the number N of carbons, also with the vertically displaced points one standard deviation above and below the

Table 1. Alkane Isomer Counts and Diameters^a

<i>N</i>	# _{ster,N}	# _{achir,N}	# _{chir,N}	$\langle D \rangle_{\text{diast},N}$	$\langle D \rangle_{\text{ster},N}$	# _{struc,N}	I	II	III
2	1	1	0	1.000	1.000	1	1.000	1.000 00	1.0
3	1	1	0	2.000	2.000	1	1.000	1.000 00	6.0
4	2	2	0	2.500	2.500	2	1.000	1.000 00	6.5
5	3	3	0	3.000	3.000	3	1.000	1.000 00	2.43×10
6	5	5	0	3.800	3.800	5	1.000	1.000 00	6.24×10
7	11	7	4	4.333	4.364	9	1.000	1.222 22	1.31×10^2
8	24	14	10	4.842	4.875	18	1.055	1.263 15	2.97×10^2
9	55	21	34	5.368	5.418	35	1.085	1.447 36	6.77×10^2
10	136	40	96	5.875	5.912	75	1.173	1.545 45	1.44×10^3
11	345	61	284	6.355	6.388	159	1.276	1.699 50	3.05×10^3
12	900	118	782	6.825	6.853	355	1.433	1.768 17	6.41×10^3
13	2 412	186	2 226	7.269	7.292	802	1.619	1.856 81	1.33×10^4
14	6 563	355	6 208	7.703	7.719	1 858	1.861	1.897 36	2.78×10^4
15	18 127	567	17 560	8.121	8.133	4 347	2.150	1.939 33	5.76×10^4
16	50 699	1 081	49 618	8.526	8.534	10 359	2.499	1.958 24	1.19×10^5
17	143 255	1 755	141 500	8.919	8.924	24 894	2.912	1.975 79	2.45×10^5
18	408 429	3 325	405 104	9.301	9.305	60 523	3.401	1.983 84	5.06×10^5
19	1 173 770	5 454	1 168 316	9.673	9.676	148 284	3.976	1.990 74	1.04×10^6
20	3 396 844	10 306	3 386 538	10.036	10.038	366 319	4.650	1.993 95	2.15×10^6
21	9 892 302	17 070	9 875 232	10.391	10.391	910 726	5.440	1.996 55	4.43×10^6
22	28 972 080	32 136	28 939 944	10.737	10.738	2 278 658	6.364	1.997 78	9.12×10^6
23	85 289 390	53 628	85 235 762	11.076	11.077	5 731 580	7.444	1.998 74	1.87×10^7
24	252 260 276	100 704	252 159 572	11.409	11.409	14 490 245	8.707	1.999 20	3.86×10^7
25	749 329 719	169 175	749 160 544	11.735	11.735	36 797 588	10.184	1.999 54	7.94×10^7
26	2 234 695 030	316 874	2 234 378 156	12.056	12.056	93 839 412	11.908	1.999 71	1.63×10^8
27	6 688 893 605	535 267	6 688 358 338	12.370	12.370	240 215 803	13.923	1.999 83	3.36×10^8
28	20 089 296 554	1 000 524	20 088 296 030	12.680	12.680	617 105 614	16.277	1.999 90	6.91×10^8
29	60 526 543 480	1 698 322	60 524 845 158	12.984	12.984	1 590 507 121	19.027	1.999 94	1.42×10^9
30	182 896 187 256	3 168 500	182 893 018 756	13.283	13.283	4 111 846 763	22.240	1.999 96	2.92×10^9
31	554 188 210 352	5 400 908	554 182 809 444	13.578	13.578	10 660 307 791	25.993	1.999 98	6.00×10^9
32	1 683 557 607 211	10 059 823	1 683 547 547 388	13.869	13.869	27 711 253 769	30.376	1.999 98	1.23×10^{10}
33	5 126 819 371 356	17 211 368	5 126 802 159 988	14.155	14.155	72 214 088 660	35.497	1.999 99	2.53×10^{10}
34	15 647 855 317 080	32 010 736	15 647 823 306 344	14.437	14.437	188 626 236 139	41.478	1.999 99	5.21×10^{10}
35	47 862 049 187 447	54 947 147	47 861 994 240 300	14.715	14.715	493 782 952 902	48.464	1.999 99	1.07×10^{11}
36	146 691 564 302 648	102 059 572	146 691 462 243 076	14.990	14.990	1 295 297 588 128	56.624	1.999 99	2.20×10^{11}
37	450 451 875 783 866	175 702 378	450 451 700 081 488	15.261	15.261	3 404 490 780 161	66.155	1.999 99	4.52×10^{11}
38	1 385 724 615 285 949	325 962 785	1 385 724 289 323 164	15.529	15.529	8 964 747 474 595	77.287	1.999 99	9.30×10^{11}
39	4 270 217 915 878 409	562 645 937	4 270 217 353 232 472	15.794	15.794	23 647 478 933 969	90.289	1.999 99	1.91×10^{12}
40	$1.318\,044\,618\,93 \times 10^{16}$	1 042 695 173	$1.318\,044\,514\,663 \times 10^{16}$	16.055	16.055	62 481 801 147 341	105.474	1.999 99	3.92×10^{12}

^a For decimal values the results are given in a truncated form.**Figure 5.** plot of $\langle D \rangle_{\text{ster},N}$ versus N . Averages are indicated by solid dots which are connected by a curve, while standard deviations $\sigma_N(D)$ are indicated by square dots displaced from $\langle D \rangle_N$ by $-\sigma_N(D)$ and $+\sigma_N(D)$.

averages. Notably these standard deviations turn out somewhat larger than the ones for molecular properties of the next two sections.

Further, quantitative asymptotic data are given in section 9. The present data are not only given differently (in terms of our various averages) but also are much extended beyond the earlier ref 8 $N = 14$ limit (for stereoisomers and

diastereomers), and these earlier results for $N = 13$ and 14 counts of achiral and chiral structures are corrected, though the use of the rather different analytical formula (i.e., eq 23) along with the earlier data in Table 2 of that reference (i.e., ref 8) yields results in agreement with our Table 1. Thus, the results given here are expected to be correct. The total stereoisomer counts here are in agreement with ref 8 as well as with the values of Read⁹ up through $N = 25$. We also agree with the stereoisomer counts of Davies and Freyd,¹⁰ wherever they are available.

5. ATOM-TYPE MOMENT GENERATING FUNCTIONS

The generating function techniques may be extended to compute mean values for a variety of structural invariants, which may in turn (as in the next section) be used to estimate mean values of various physicochemical properties. First, we highlight some of the formulas of ref 18 to aid in the understanding of combinatorial manipulations that follow later in this section. We start by considering the structural-isomer averages of counts for primary, secondary, tertiary, and quaternary carbons. For these four *atom types* one imagines (much as in ref 18) four different t -variables t_1, t_2, t_3, t_4 and an alkyl-radical generating function $\varphi_g(t_1, t_2, t_3, t_4) \equiv \varphi_g(\mathbf{t})$ where the powers of t_1, t_2, t_3, t_4 respectively identify the numbers of primary, secondary, tertiary, or quaternary

carbons in the species being counted (with the radical site being assigned a type as though the dangling bond is attached to another carbon, as it will be once center joinings are made). The recursion for these $\varphi_g(t, t_1, t_2, t_3, t_4)$ is analogous to that of (2.2) for the earlier $\varphi_g(t)$ —except that, first, in place of the initial t factor in (2.2) there is a t_d in correspondence with the (H-deleted graphical) degree d of the root site; second, we now deal with structural isomers rather than stereoisomers; and third, a separation of $\varphi_0(t, t_1, t_2, t_3, t_4) = 1$ from $\varphi_{<g}(t, t_1, t_2, t_3, t_4)$ is needed to properly account for the root-site type. Then in terms of $\varphi_{<g}^-(\mathbf{t}) \equiv \varphi_{<g}(\mathbf{t}) - 1$, one has

$$\varphi_{g+1}(\mathbf{t}) = t_2 F_2[\varphi_g; \mathbf{t}] + t_3 F_3[\varphi_g; \mathbf{t}] + t_4 F_4[\varphi_g; \mathbf{t}] \quad (5.1)$$

where

$$\begin{aligned} F_2[\varphi_g; \mathbf{t}] &= \varphi_g(\mathbf{t}) \\ F_3[\varphi_g; \mathbf{t}] &= \varphi_g(\mathbf{t})[\varphi_{<g}^-(\mathbf{t})] + Z_2[\varphi_g; \mathbf{t}] \\ F_4[\varphi_g; \mathbf{t}] &= \varphi_g(\mathbf{t}) Z_2[\varphi_{<g}^-(\mathbf{t})] + \\ &\quad Z_2[\varphi_g(\mathbf{t})] \varphi_{<g}^-(\mathbf{t}) + Z_3[\varphi_g; \mathbf{t}] \end{aligned} \quad (5.2)$$

These alkyl-radical generating functions then may be combined to give overall alkane structural isomer counts, with the powers of the different t_d recording the numbers of different types of atoms as shown in ref 18. One notes that this step is analogous to obtaining the stereoisomer counts of eq 2.5 from the results of eq 2.2 of the present article. Recursion 5.1 could be used to generate all the individual $\#_{g,N}$, but this leads to an intimidating multidimensional array of different indexes. Instead we make use of derivative generating functions

$$t_d \partial_d \varphi_g(\mathbf{t}) = \sum_{\mathbf{N}} N_d \#_{g,N} \mathbf{t}^{\mathbf{N}} \quad (5.3)$$

where ∂_d indicates a derivative with respect to t_d , $\mathbf{N} \equiv (N, N_1, N_2, N_3, N_4)$ with N the total number of atoms and N_d the number of degree- d atoms, $\#_{g,N}$ the count of such alkyl radicals, and $\mathbf{t}^{\mathbf{N}}$ an abbreviation for t^N times the product of the four t_d each to a power N_d . Of course these derivative generating functions may be computed recursively via a relation obtained by taking derivatives of (5.1) which are conveniently expressed in terms of the derivatives of the forms of (5.2),

$$\begin{aligned} \partial_d F_2[\varphi_g; \mathbf{t}] &\equiv \partial_d \varphi_g(\mathbf{t}) \\ \partial_d F_3[\varphi_g; \mathbf{t}] &\equiv [\partial_d \varphi_g(\mathbf{t})] [\varphi_{<g}^-(\mathbf{t})] + \\ &\quad \varphi_g(\mathbf{t}) [\partial_d \varphi_{<g}^-(\mathbf{t})] + \partial_d Z_2[\varphi_g; \mathbf{t}] \\ \partial_d F_4[\varphi_g; \mathbf{t}] &\equiv [\partial_d \varphi_g(\mathbf{t})] Z_2[\varphi_{<g}^-(\mathbf{t})] + \varphi_g(\mathbf{t}) (\partial_d Z_2 \\ &\quad [\varphi_{<g}^-(\mathbf{t})] + (\partial_d Z_2[\varphi_g; \mathbf{t}]) \varphi_{<g}^-(\mathbf{t}) + \\ &\quad Z_2[\varphi_g; \mathbf{t}] [\partial_d \varphi_{<g}^-(\mathbf{t})] + (\partial_d Z_3[\varphi_g; \mathbf{t}]) \end{aligned} \quad (5.4)$$

The $\partial_d Z_n[f; \mathbf{t}]$ are similar sorts of derivatives (based on the formulas of (2.3)) as presented in ref 18. Then the recursion

relations for the desired derivative generating functions are

$$\partial_d \varphi_{g+1}(\mathbf{t}) = t F_d[\varphi_g; \mathbf{t}] + \sum_{e=2}^4 t_e (\partial_d F_e[\varphi_g; \mathbf{t}]) \quad (5.5)$$

which we imagine is to be evaluated at $t_1 = t_2 = t_3 = t_4 = 1$ (with the variables t_d having been introduced solely so that the formal derivatives could be taken). That is, numerically one works with recursions for single-variable polynomials

$$\partial_d \varphi_{g+1}(t) = t F_d[\varphi_g; t] + \sum_{e=2}^4 t (\partial_d F_e[\varphi_g; t]) \quad (5.6)$$

and these derivative functions for alkyl radicals may be used to obtain corresponding derivative functions $\partial_d P_D(t)$ and $\partial_d P(t)$ for alkanes, by taking derivatives of polynomial functions $P_D(t)$ and $P(t)$. Defining the coefficient of t^N in $\partial_d P(t)$ to be $\partial_d \#_{\text{struc}, N}$, one may obtain the mean number of type- d atoms in isomeric N -atom alkanes as

$$\langle N_d \rangle_N = (\partial_d \#_{\text{struc}, N}) / \#_{\text{struc}, N} \quad (5.7)$$

where $\#_{\text{struc}, N}$ represents structural isomer counts for class- N .

These ideas can be extended to obtain standard deviations for the numbers of atoms of each different type. This involves taking second derivatives, so that in addition to (5.5) one has

$$\begin{aligned} \partial_c \{t_d \partial_d \varphi_{g+1}(\mathbf{t})\} &= t F_d[\varphi_g; \mathbf{t}] + t (\partial_c F_e[\varphi_g; \mathbf{t}]) + \\ &\quad 2t (\partial_d F_d[\varphi_g; \mathbf{t}]) + \sum_{e=2}^4 t t_e (\partial_c \partial_d F_e[\varphi_g; \mathbf{t}]) \end{aligned} \quad (5.8)$$

which again are used as evaluated at $t_1 = t_2 = t_3 = t_4 = 1$. The consequent second-derivative generating functions then lead to $\partial_c \partial_d P_D(t)$ and $\partial_c \partial_d P(t)$, and the coefficients of the latter might be denoted as $\partial_c \partial_d \#_{\text{struc}, N}$. Then, remembering that if the first derivative brings down a power N_d the second derivative brings down an overall factor of $N_c N_d$, the (mixed) second moments for numbers of different atom types are

$$\begin{aligned} \text{cov}_N(c, d) &\equiv \langle (N_c - \langle N_c \rangle_N) (N_d - \langle N_d \rangle_N) \rangle_N \\ &= \langle N_c N_d \rangle_N - \langle N_c \rangle_N \langle N_d \rangle_N \\ &= \{(\partial_c \partial_d \#_{\text{struc}, N}) / \#_{\text{struc}, N}\} - \\ &\quad \{(\partial_c \#_{\text{struc}, N}) (\partial_d \#_{\text{struc}, N}) / (\#_{\text{struc}, N})^2\} \end{aligned} \quad (5.9)$$

which then generalize our earlier results¹⁸ for the variances (i.e., the diagonal $c = d$ covariances).

It may be noted that the different means are not “independent”, e.g., since for any structure a sum over d of these N_d must evidently sum to N . But also there is another relation, since every time a branch is introduced another end is simultaneously introduced. That is,

$$\begin{aligned} N_1 + N_2 + N_3 + N_4 &= N \\ N_3 + 2N_4 + 2 &= N_1 \end{aligned} \quad (5.10)$$

and, of course, these same relations apply with each N_d replaced by the corresponding average $\langle N_d \rangle_N$, and this can be either used to avoid constructions or else to check the

Table 2. Covariances for Alkane C-types^a

<i>N</i>	cov(1,2)	cov(1,3)	cov(1,4)	cov(2,3)	cov(2,4)	cov(3,4)
2	0.000 000 00	0.000 000 00	0.000 000 00	0.000 000 00	0.000 000 00	0.000 000 00
3	0.000 000 00	0.000 000 00	0.000 000 00	0.000 000 00	0.000 000 00	0.000 000 00
4	-0.500 000 00	0.250 000 00	0.000 000 00	-0.500 000 00	0.000 000 00	0.000 000 00
5	-1.000 000 00	0.000 000 00	0.333 333 33	-0.111 111 11	-0.444 444 44	-0.111 111 11
6	-0.960 000 00	0.240 000 00	0.160 000 00	-0.640 000 00	-0.160 000 00	-0.160 000 00
7	-1.123 456 79	0.172 839 50	0.259 259 25	-0.530 864 19	-0.296 296 29	-0.185 185 18
8	-1.388 888 88	0.111 111 11	0.388 888 88	-0.549 382 71	-0.419 753 08	-0.327 160 49
9	-1.436 734 69	0.140 408 16	0.385 306 12	-0.620 408 16	-0.408 163 26	-0.339 591 83
10	-1.621 866 66	0.158 933 33	0.434 666 66	-0.736 533 33	-0.442 666 66	-0.418 666 66
11	-1.736 798 38	0.145 484 75	0.481 942 96	-0.748 546 33	-0.494 126 02	-0.457 576 83
12	-1.868 700 65	0.149 684 58	0.523 110 49	-0.819 385 04	-0.524 657 80	-0.520 015 86
13	-2.009 264 55	0.150 955 21	0.569 118 04	-0.870 487 43	-0.569 388 56	-0.568 576 99
14	-2.144 975 44	0.149 868 60	0.615 079 41	-0.930 483 31	-0.607 246 06	-0.630 746 10
15	-2.274 292 59	0.151 472 58	0.657 115 80	-0.984 732 23	-0.644 780 18	-0.681 787 05
16	-2.417 553 87	0.153 967 90	0.703 206 01	-1.048 319 69	-0.684 617 09	-0.740 383 87
17	-2.555 564 41	0.155 688 39	0.748 062 54	-1.106 446 79	-0.724 558 81	-0.795 070 01
18	-2.697 617 14	0.158 200 42	0.793 738 76	-1.169 033 30	-0.764 291 92	-0.852 632 46
19	-2.838 951 48	0.160 567 63	0.839 272 07	-1.229 955 15	-0.804 498 16	-0.908 819 89
20	-2.982 289 80	0.163 074 39	0.885 380 33	-1.292 540 09	-0.844 874 85	-0.966 391 29
21	-3.124 987 26	0.165 602 39	0.931 260 82	-1.354 534 02	-0.885 226 61	-1.023 329 24
22	-3.268 863 40	0.168 207 00	0.977 483 13	-1.417 350 07	-0.925 756 66	-1.080 936 06
23	-3.412 451 85	0.170 809 51	1.023 610 94	-1.479 920 27	-0.966 265 78	-1.138 301 24
24	-3.556 545 85	0.173 454 69	1.069 878 81	-1.542 871 98	-1.006 836 93	-1.195 962 58
25	-3.700 580 84	0.176 106 84	1.116 122 38	-1.605 767 63	-1.047 406 60	-1.253 553 94
26	-3.844 830 80	0.178 775 57	1.162 426 55	-1.668 832 20	-1.087 999 29	-1.311 281 06
27	-3.989 060 09	0.181 451 57	1.208 718 97	-1.731 893 93	-1.128 583 07	-1.368 990 77
28	-4.133 394 85	0.184 137 41	1.255 040 00	-1.795 043 50	-1.169 175 67	-1.426 768 66
29	-4.277 724 18	0.186 828 06	1.301 356 01	-1.858 204 67	-1.209 759 75	-1.484 548 53
30	-4.422 099 17	0.189 524 75	1.347 683 22	-1.921 412 91	-1.250 343 12	-1.542 363 40
31	-4.566 475 16	0.192 225 29	1.394 008 19	-1.984 636 67	-1.290 919 24	-1.600 186 08
32	-4.710 870 65	0.194 929 78	1.440 337 03	-2.047 888 08	-1.331 491 28	-1.658 028 52
33	-4.855 266 73	0.197 637 26	1.486 664 06	-2.111 153 61	-1.372 056 56	-1.715 879 07
34	-4.999 670 68	0.200 347 62	1.532 991 81	-2.174 437 16	-1.412 616 76	-1.773 741 91
35	-5.144 074 23	0.203 060 34	1.579 317 85	-2.237 732 49	-1.453 170 87	-1.831 611 81
36	-5.288 480 03	0.205 775 25	1.625 643 17	-2.301 040 47	-1.493 719 77	-1.889 489 96
37	-5.432 884 57	0.208 492 07	1.671 966 80	-2.364 358 06	-1.534 263 25	-1.947 373 90
38	-5.577 288 74	0.211 210 65	1.718 289 14	-2.427 685 08	-1.574 801 83	-2.005 263 77
39	-5.721 691 03	0.213 930 79	1.764 609 81	-2.491 019 89	-1.615 335 56	-2.063 158 31
40	-5.866 091 72	0.216 652 36	1.810 928 99	-2.554 362 04	-1.655 864 84	-2.121 057 30

^a The results are given in a truncated form.

numerical results from the constructions. Also, there are “sum rules” for the covariances, because of the relations of (5.10). Indeed there are three such sum rules: two from squaring each of the equations of (5.10) and one from the product of these two equations.

The numerical results for the off-diagonal ($c \neq d$) covariances are given in Table 2 for alkanes of up to $N = 40$ atoms. The averages and standard deviations (i.e., square roots of diagonal ($c = d$) variances) up through $N = 40$ were given in our earlier article,¹⁸ along with structural isomer counts which have been given several times previously.^{10,11} Evidently the off-diagonal covariances increase in magnitude with N in a manner that plausibly appears to approach linearity, just as we have already found for the (diagonal) variances.

6. PROPERTY-VALUE EXPECTATIONS AND STANDARD DEVIATIONS

Mean values for these graph invariants may be utilized to predict mean property values for a number of properties, representative ones of which are heat of formation, index of refraction, and magnetic susceptibility. That is, such properties for a particular molecule may be estimated in terms of an expansion in terms of local substructure counts, say of different types of atoms and of different types of bonds. For a structure represented by a graph G , such a property X is

represented in terms of an additive *cluster expansion* as

$$X(G) = \sum_{\xi}^G \nu(\xi; G) X(\nu, \xi) \quad (6.1)$$

where the symbol ξ labels different types of substructures, $\nu(\xi; G)$ is (at least in the current article just) the number of such substructures in G , and $X(\nu, \xi)$ is a parameter dependent on the property X but not on the structure G . For example, if ξ labels a type of atom or bond, then $X(\nu, \xi)$ is an atomic or bond contribution to the total property for the structure G . In principle the expansion if fully extended is^{24,25} exact, but it is virtually always used in a truncated (approximate) form where the substructures ξ are restricted to those of smaller size (e.g., perhaps just single sites and bonds). The substructures can be labeled to different extents (and there are different, yet complete, classes of substructures one may consider), so that different cluster expansions are possible.

The particular cluster expansions used are those in terms of atom-types (primary, secondary, tertiary, and quaternary) for the properties heat of formation, magnetic susceptibility, and index of refraction. Further details and the values for the expansion coefficients $X(\nu, \xi)$ for each property are given in our earlier paper¹⁸ (also see refs 25–28).

The computation of mean values and standard deviations is now straightforward in terms of the atom-type results and

the expansion coefficients (given in section 5). That is, one generally takes

$$\langle X(G) \rangle_N = \sum_{\xi}^G \langle \nu(\xi; G) \rangle_N X(\nu, \xi) \quad (6.2)$$

For second-moment expectations one has

$$\langle X(G)^2 \rangle_N = \sum_{\xi, \zeta}^G \langle \nu(\xi; G) \nu(\zeta; G) \rangle_N X(\nu, \xi) X(\nu, \zeta) \quad (6.3)$$

and as a consequence for the standard deviation squared one has

$$\sigma_N(X)^2 = \langle \{X(G) - \langle X(G) \rangle_N\}^2 \rangle_N = \sum_{\xi, \zeta}^G \text{cov}_N(\xi, \zeta) X(\nu, \xi) X(\nu, \zeta) \quad (6.4)$$

If the quantities $\nu(\xi, G)$ and $\nu(\zeta, G)$ were independent (or uncorrelated), which is frequently considered as the “standard” case in statistics books, then the off-diagonal covariances (with $\xi \neq \zeta$) would be 0. Regardless, these off-diagonal terms could be deleted to obtain a quantity here called the *disrelated* standard deviation

$$\sigma_{\text{dis},N}(X)^2 \equiv \sum_{\xi}^G (\sigma_N(\xi))^2 X(\nu, \xi)^2 \quad (6.5)$$

For correlated (dependent) quantities $\nu(\xi, G)$ and $\nu(\zeta, G)$, the disrelated quantity $\sigma_{\text{dis},N}(X)$ generally is not be the same as $\sigma_N(X)$, though equality still might occur if nonzero off-diagonal covariances take suitable signs and manage to mutually cancel. Of course the off-diagonal covariances might conceivably cancel part of the contribution of the diagonal terms, so that $\sigma_{\text{dis},N}(X)$ is not always a known bound to $\sigma_N(X)$. Also of some interest is the *absoluted* standard deviation

$$\sigma_{\text{abs},N}(X) \equiv \sum_{\xi}^G |\sigma_N(\xi)| |X(\nu, \xi)| \quad (6.6)$$

which involves again only the diagonal part of the covariance matrix and provides an *upper bound* to the proper standard deviations $\sigma_N(X)$, as shown in appendix A.

For our present two properties ΔH_f and χ_{mag} , the mean properties at a sequence of N values, are given in Table 3, where also we display ratios $\sigma_N(X)$ to $\sigma_{\text{abs},N}(X)$ and of $\sigma_{\text{dis},N}(X)$ to $\sigma_{\text{abs},N}(X)$. That these ratios turn out to be somewhat less than 1 indicates (for these properties) the cancellative intercorrelation among the atom-type counts. From Table 3 it is seen that these kinds of “modified” standard deviations do not differ very much from one another, the ratios of $\sigma_{\text{dis},N}(X)/\sigma_{\text{abs},N}(X)$ being close to 0.5. However, both of them differ considerably from the proper standard deviations, the ratios $\sigma_N(X)/\sigma_{\text{abs},N}(X)$ being on the order of 0.1 and 0.01 for ΔH_f and χ_{mag} , respectively. The behavior with increasing N of (proper) standard deviations of the two properties considered here can be seen in Figure 6 and 7, from which the standard deviations evidently seem to scale like $N^{1/2}$, as we shall discuss in more detail a little later (in section 9).

Table 3. Alkane Average Properties (ΔH_f and χ_{mag}) and Deviations^a

N	$-\Delta H_f$	$\sigma_N/\sigma_{\text{abs},N}$	$\sigma_{\text{dis},N}/\sigma_{\text{abs},N}$	χ_{mag}	$\sigma_N/\sigma_{\text{abs},N}$	$\sigma_{\text{dis},N}/\sigma_{\text{abs},N}$
2	20.000 00			28.580 00		
3	25.000 00			40.060 00		
4	31.200 00	0.107 14	0.640 37	51.800 00	0.0111 97	0.615 04
5	37.499 99	0.133 71	0.663 44	63.713 33	0.0191 78	0.570 47
6	42.940 00	0.116 77	0.636 89	75.228 00	0.0148 25	0.560 82
7	48.833 33	0.122 22	0.642 83	86.962 22	0.0162 95	0.559 43
8	54.933 33	0.124 05	0.640 47	98.731 11	0.0169 77	0.550 99
9	60.785 71	0.123 06	0.639 29	110.440 57	0.0167 25	0.551 09
10	66.708 00	0.122 39	0.637 30	122.146 40	0.0166 11	0.549 12
11	72.654 71	0.122 99	0.637 67	133.881 88	0.0167 80	0.548 48
12	78.579 15	0.122 75	0.636 60	145.597 69	0.0167 57	0.547 19
13	84.515 08	0.122 88	0.636 48	157.323 84	0.0168 01	0.546 74
14	90.450 43	0.122 84	0.635 85	169.045 79	0.0168 19	0.545 80
15	96.384 54	0.122 85	0.635 56	180.770 01	0.0168 35	0.545 32
16	102.321 31	0.122 81	0.635 19	192.493 13	0.0168 43	0.544 79
17	108.259 72	0.122 82	0.634 95	204.218 15	0.0168 57	0.544 40
18	114.198 18	0.122 80	0.634 68	215.942 31	0.0168 64	0.544 01
19	120.137 89	0.122 80	0.634 47	227.667 40	0.0168 72	0.543 69
20	126.078 34	0.122 78	0.634 26	239.392 43	0.0168 79	0.543 39
21	132.019 39	0.122 78	0.634 09	251.117 86	0.0168 85	0.543 12
22	137.961 02	0.122 77	0.633 92	262.843 36	0.0168 91	0.542 88
23	143.903 18	0.122 76	0.633 77	274.569 13	0.0168 96	0.542 67
24	149.845 77	0.122 75	0.633 63	286.294 98	0.0169 00	0.542 47
25	155.788 76	0.122 74	0.633 50	298.021 01	0.0169 04	0.542 28
26	161.732 08	0.122 74	0.633 39	309.747 13	0.0169 08	0.542 12
27	167.675 72	0.122 73	0.633 28	321.473 37	0.0169 11	0.541 96
28	173.619 62	0.122 72	0.633 17	333.199 69	0.0169 14	0.541 82
29	179.563 76	0.122 72	0.633 08	344.926 11	0.0169 17	0.541 68
30	185.508 12	0.122 71	0.632 99	356.652 59	0.0169 20	0.541 56
31	191.452 67	0.122 71	0.632 91	368.379 15	0.0169 23	0.541 44
32	197.397 40	0.122 70	0.632 83	380.105 76	0.0169 25	0.541 33
33	203.342 29	0.122 70	0.632 76	391.832 43	0.0169 27	0.541 23
34	209.287 31	0.122 69	0.632 69	403.559 15	0.0169 29	0.541 14
35	215.232 47	0.122 69	0.632 62	415.285 91	0.0169 31	0.541 05
36	221.177 75	0.122 69	0.632 56	427.012 72	0.0169 33	0.540 96
37	227.123 13	0.122 68	0.632 50	438.739 56	0.0169 35	0.540 88
38	233.068 61	0.122 68	0.632 45	450.466 44	0.0169 36	0.540 81
39	239.014 19	0.122 67	0.632 40	462.193 35	0.0169 38	0.540 73
40	244.959 85	0.122 67	0.632 35	473.920 29	0.0169 39	0.540 67

^a The results are given in kcal/mol and m.u.

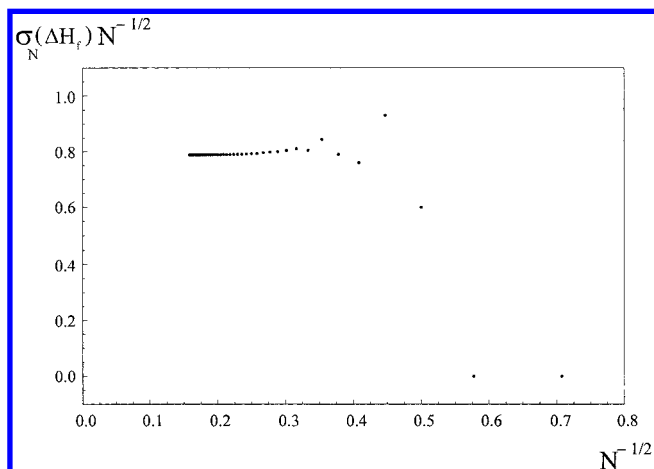


Figure 6. Plot of $\sigma(\Delta H_f)/N^{1/2}$ (in units of kilocalories per mole) versus $1/N^{1/2}$ to examine the asymptotic behavior of the standard deviations for ΔH_f .

Of course the disrelated or absolved standard deviations are simpler to compute in only requiring just the diagonal ($\xi = \zeta$) covariances, and such is what was reported in our previous article,¹⁸ where $\sigma_{\text{abs},N}(X)$ was called the “standard deviation”, though it is clear that the different atom-type counts are not independent, because of the two

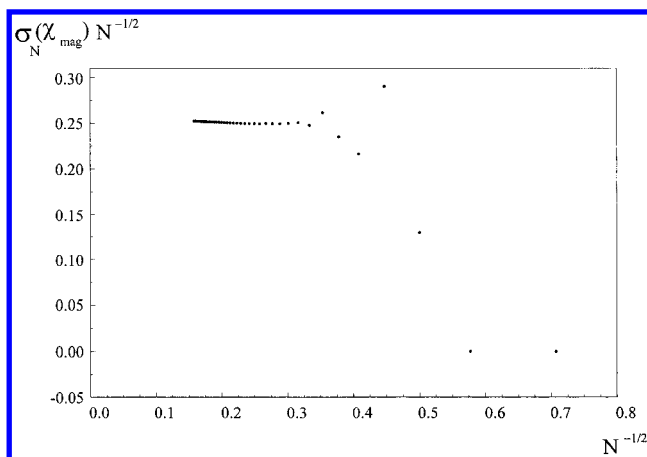


Figure 7. Plot of $\sigma_N(\chi_{\text{mag}})/N^{1/2}$ (in units magnetic units (mu) = -10^{-6} erg/(G² mol)) versus $1/N^{1/2}$ to examine the asymptotic behavior of standard deviations for χ_{mag} .

relations in (5.10). The additional formalism of section 5 to compute the proper standard deviations is quite modest, though it substantially increases the computational effort. Again the numerical results of Table 3 show that at least for ΔH_f and χ_{mag} the proper standard deviations are much different than the others (without contributions from the off-diagonal covariances).

7. STRUCTURES WITH EXTREMAL PREDICTED PROPERTY VALUES

Beyond averages and standard deviations there may be desired some additional information about property-value distributions within an isomer class. And of natural interest in this regard are the maximum and minimum values predicted for the properties. Notably for properties given in terms of truncated cluster expansions (such as for ΔH_f and χ_{mag}) this problem turns out to be relatively straightforward. In particular, in this case one just varies the graph invariants to extremize the truncated cluster expansion expression, which for our present considerations are in terms of atom types, so that it becomes

$$X(G) = \sum_{d=1}^4 N_d X_d \quad (7.1)$$

Here of course these N_d are nonnegative integers, and they also satisfy the two linear relations of (5.10). The extremization (or optimization) problem subject to such constraints presumably may be approached via powerful techniques of *linear programming*, using the *simplex algorithm*, e.g., as discussed in ref 29. But really the current circumstance is sufficiently simple that one can neatly solve the problem without general elaboration of such general formalities. First we note that because of the two conditions one may eliminate N_1 and N_2 in favor of N_3 and N_4 (as well as N which we view as fixed, for a given isomer class); thusly,

$$N_1 = 2 + N_3 + 2N_4 \quad N_2 = N - 2 - 2N_3 - 3N_4 \quad (7.2)$$

Then the expansion of (7.1) becomes

$$X(G) = x_0 + x_3 N_3 + x_4 N_4 \quad (7.3)$$

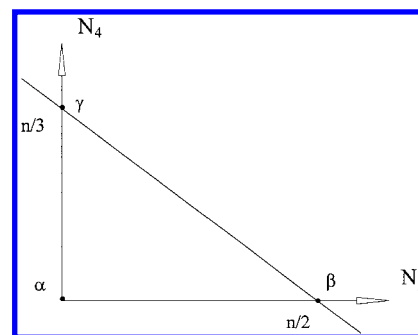


Figure 8. Portion of the (N_3, N_4) plane allowed by the constraints of (7.5).

where

$$x_0 \equiv 2X_1 + (N - 2)X_2, \quad x_3 \equiv X_3 - 2X_2 + X_1, \\ x_4 \equiv X_4 - 3X_2 + 2X_1 \quad (7.4)$$

The constraints on the remaining two variables N_3 and N_4 are (aside from their integrality)

$$N_3 \geq 0, \quad N_4 \geq 0, \quad n - 2N_3 - 3N_4 \geq 0 \quad (7.5)$$

where we abbreviate $n \equiv N - 2$. These conditions specify a triangular region of the (N_3, N_4) plane, as indicated in Figure 8. And $X(G)$ being a function of N_3 and N_4 , we may imagine the value of $X(G)$ (at a fixed N) plotted along an axis normal to this (N_3, N_4) plane. Since $X(G)$ is linear in N_3 and N_4 , these values of $X(G)$ simply identify a new plane, which in general will be slanted (nonparallel) with respect to the base (N_3, N_4) plane. Then the extremum values of $X(G)$ associated with the allowable N_3 and N_4 (as designated in (7.5)) must be realizable at the corners of this triangular region. Thus with no attention to integrality the maximum and minimum values of $X(G)$ satisfying (7.5) are among the values

$$X(\alpha) = x_0, \quad X(\beta) = x_0 + x_3 n/2, \quad X(\gamma) = x_0 + x_4 n/3 \quad (7.6)$$

where α , β , and γ simply label the three triangle corners as in Figure 8.

The remnant conditions of integrality of the N_d are resolved through a consideration of the integral- N_d points nearer these triangle corners. And these nearer integral- N_d points are readily identified through a consideration of the remainder r after dividing $n \equiv N - 2$ by 2 (in dealing with corner β) or by 3 (in dealing with corner γ). When a corner β or γ has noninteger coordinates, really one needs only to consider the integer points near β or γ such that all other integer points are within the convex hull of the considered points—that is, the (normal) distance to the plane of values of $X(G)$ is extremized when at a boundary point of the region of integral values of the (N_3, N_4) plane. These nearby points (applicable when β or γ has nonintegral coordinates) may be readily identified from appropriate plots, as indicated in Figure 9, where we use the notation $[N/q]$ to be the largest integer not exceeding N/q . For β when $n = 2q + 1$ (with $q = [N/2]$ an integer) the remainder is $r_\beta = 1$, and there are two such integral points

$$\beta_1' \text{ at } (N_3, N_4) = (q, 0) \quad \beta_1'' \text{ at } (N_3, N_4) = (q-1, 1) \quad (7.7)$$

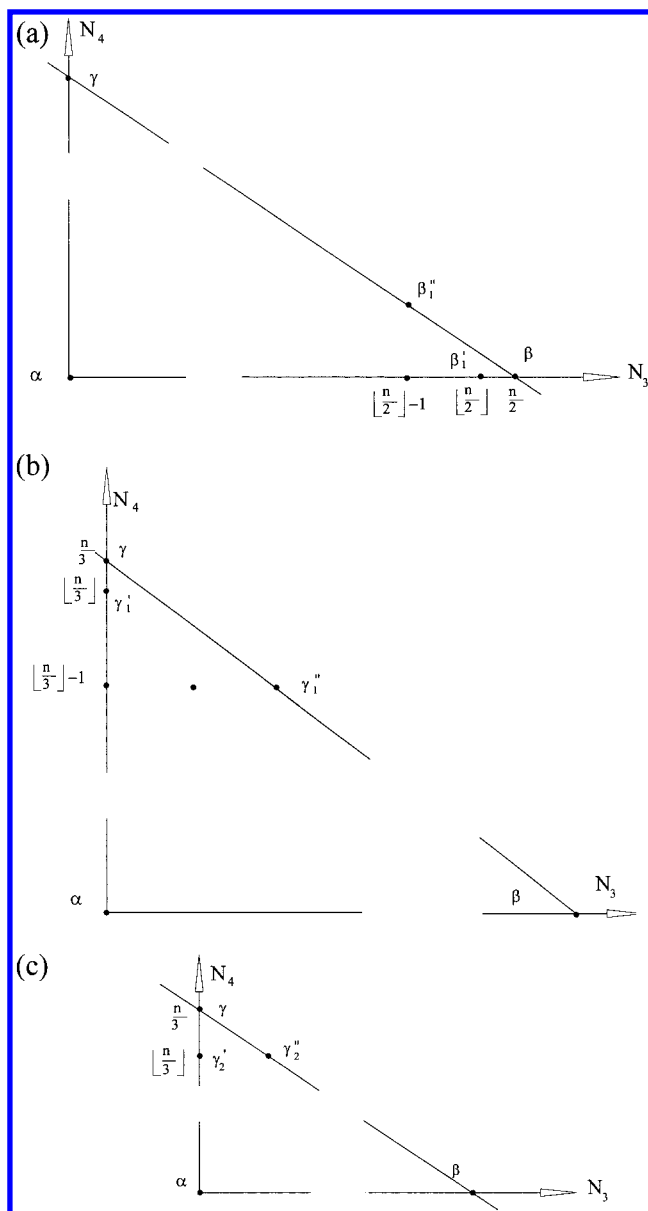


Figure 9. Detail of the (N_3, N_4) plane showing integer points, for different cases: (a) near β when $n = 2q + 1$ and $r_\beta = 1$, (b) near γ when $n = 3q + 1$ and $r_\gamma = 1$, and (c) near γ when $n = 3q + 2$ and $r_\gamma = 2$.

For γ when $n = 3q + 1$ (with $q = \lfloor N/3 \rfloor$), the remainder is $r_\gamma = 1$, and there are again two such integral points

$$\gamma_1' \text{ at } (N_3, N_4) = (q, 0) \quad \gamma_1'' \text{ at } (N_3, N_4) = (q-1, 2) \quad (7.8)$$

For γ when $n = 3q + 2$, there are yet again two such integral points

$$\gamma_2' \text{ at } (N_3, N_4) = (q, 0) \quad \gamma_2'' \text{ at } (N_3, N_4) = (q-1, 1) \quad (7.9)$$

In each of these cases then there is an additional consideration to determine which of the two points applies. One simply needs to compare the values of $X(G)$ at each of the points of the appropriate pair.

The final results for the maximum values of $X(G)$ are summarized in Table 4, where conditioning inequalities are

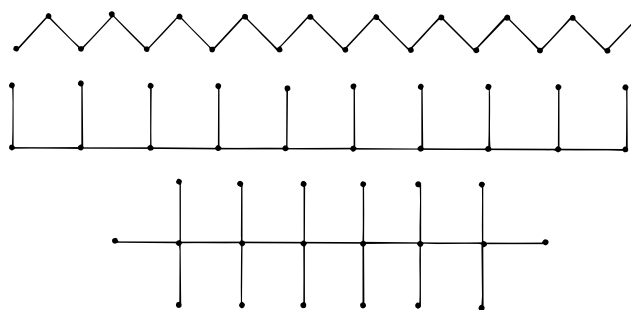


Figure 10. Type α , β , and γ extremal graphs for $N = 20$ carbon atoms (where $n = 18$ and all $r = 0$).

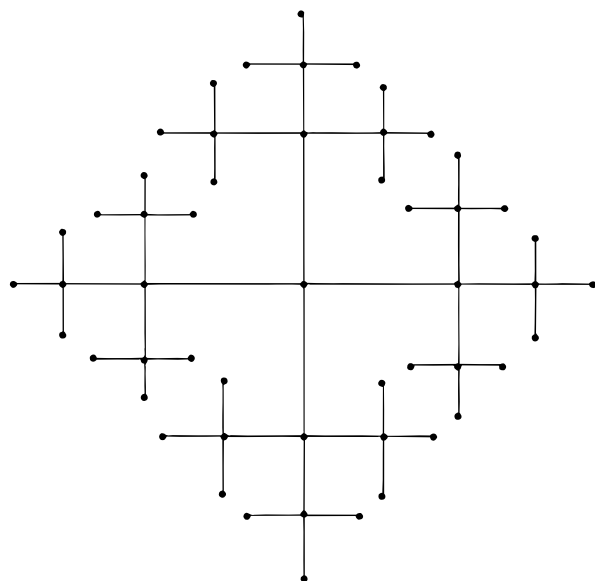
reexpressed in terms of the original N_d and X_d . It is understood that for each corner (α , β , or γ) the initial condition in this table applies to the ensuing subcases. The final results for the minimum values are obtained from Table 4 simply by interchanging \geq and \leq signs (and replacing max by min).

Further there is the question of the *extremal* molecular structures corresponding to these extremal values. Point α corresponds to the linear chain (normal alkane) structure, point β corresponds (among other possibilities) to a linear comb structure where the bulk of the sites are of degree 3, and point γ corresponds (also among other possibilities) to a linear “two-sided” comb with the bulk of the sites of degree 4. These three types of indicated structures for no-remainder cases ($r = 0$) are shown in Figure 10. If there is a nonzero remainder, there can be one or two of the internal sites of the comb with a degree diminished by 1. For points β and γ there are generally other possible structures associated with the extrema; they may be imagined: first, by deleting all the degree-1 sites from the parent linear comb structure Λ to obtain a simple linear chain; second, by rearranging these remnant sites in a pattern allowed for any alkane structure with a limitation that the number of vertices of a given degree d does not exceed that occurring in the parent structure Λ ; and third, reattaching the initially deleted degree-1 sites still as degree-1 sites. Thus, if we let the number of alkane structures at M -carbons be denoted by $\#_{\text{struc}, M}$ and the number without quaternary sites be $\#_{\text{struc}, M}'$, then the degeneracy of the structures at point α is 1; at the point β the degeneracy is $\sim N^{\#_{\text{struc}, \lfloor N/2 \rfloor}}$ with $r = r_\beta$ ($=0$ or 1); and at γ the degeneracy is $\sim N^{\#_{\text{struc}, \lfloor N/3 \rfloor}}$, where $r = r_\gamma$ ($=0, 1$, or 2). In any case, for the β and γ cases there generally are numerous extremal structures. For maximal values, with our X_d parameters for ΔH_f the case γ , with subcases γ_1' and γ_2' of Table 4, applies, and for the χ_{mag} case γ , with subcases γ_1' and γ_2' . For minimal values, with our X_d parameters for ΔH_f and, also, for χ_{mag} the case α of Table 4 applies.

To solve the extremal problem for the graph diameter (or also the mean shortest-path distances between sites), the reasoning is even more straightforward. The maximum diameter clearly corresponds to the linear chain (as in Figure 10, top) with a diameter $D = N - 1$. On the other hand, the minimum diameter case is related to the dendritic (or Bethe) tree of degree 4, an example of such a Bethe tree being indicated in Figure 11. In general such Bethe trees are obtained by joining four maximally branched g -generation radicals at a central site to give a total of $N = 2 \times 3^{g+1} - 1$ sites and a diameter of $D = 2g$. Hence for these structures $D \sim \log N$. Of course only rather a few (a fraction $\sim (\log$

Table 4. Maximal Property Values

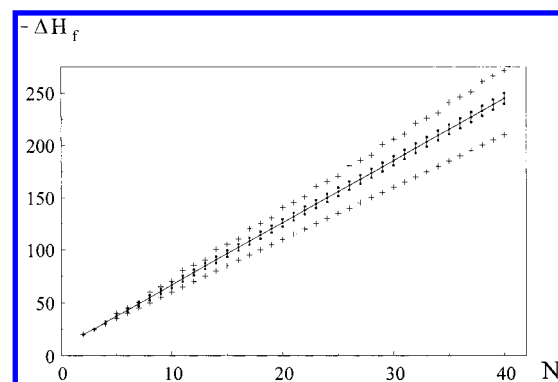
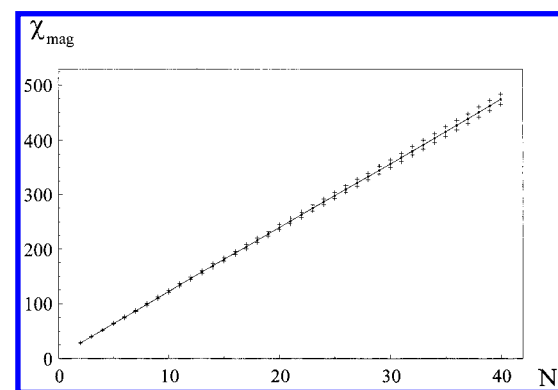
corner	conditions	N value	max $X(G)$
α	$(3X_1 + 3X_3) \leq (6X_2) \geq (4X_1 + 2X_2)$	any N	x_0
β	$(6X_2) \leq (3X_1 + 3X_3) \geq (4X_1 + 2X_2)$		
β_0		$N = 2q$	$x_0 + qx_3$
β'_1	also $(X_1 + X_4) \geq (X_2 + X_3)$	$N = 2q + 1$	$x_0 + qx_3$
β''_1	also $(X_1 + X_4) \leq (X_2 + X_3)$	$N = 2q + 1$	$x_0 + (q - 1)x_3 + x_4$
γ	$(3X_1 + 3X_3) \leq (4X_1 + 2X_2) \geq (6X_2)$		
γ_0		$N = 3q$	$x_0 + qx_4$
γ'_1	also $(X_1 + X_4) \geq (2X_3)$	$N = 3q + 1$	$x_0 + qx_4$
γ''_1	also $(X_1 + X_4) \leq (2X_3)$	$N = 3q + 1$	$x_0 + 2x_3 + (q - 1)x_4$
γ'_2	also $(X_2 + X_3) \geq (2X_2)$	$N = 3q + 2$	$x_0 + qx_4$
γ''_2	also $(X_2 + X_3) \leq (2X_2)$	$N = 3q + 2$	$x_0 + x_3 + qx_4$

**Figure 11.** Diameter $D = 6$ (dendritic) Bethe tree of $N = 53$ carbons based on degree = 4 (alkane-like) sites.

$N)/N$ values of N correspond to a (perfect) Bethe tree, but if terminal sites are deleted from a given Bethe tree, the diameter remains the same for consequent values of remnant numbers N of sites until the next smaller Bethe tree is reached. Hence these “near-Bethe” trees are the extremal structures at these intermediate N values, and still have $D \sim \log N$. At the special site counts of $N = 2 \times 3^{g+1} - 1$, there is only a single extremal structure, whereas at intermediate values there typically are many, corresponding to different choices for deleting a given number of terminal sites. Finally it may be mentioned that as D increases for these Bethe (or near-Bethe) trees, one soon reaches^{16,30} a circumstance of severe steric hindrance.

8. NUMERICAL RESULTS FOR PROPERTY VALUES

Numerical results for heat of formation and magnetic susceptibility are given in Figures 12 and 13, respectively. The central point (at each N) in these figures is the average value $\langle X \rangle_N$, and they are connected by a curve. Displaced one standard deviation $\sigma_N(X)$ above and below the average are two points connected by a vertical line, though in Figure 13 the standard deviation points (especially for smaller N) are so close to the average that they are not clearly distinguished. Finally the two outermost points (marked by “+” at each N) identify the extremal values of X . These

**Figure 12.** Plot for heat of formation values $\langle \Delta H_f \rangle_N$ in units of kilocalories per mole. Averages are indicated by solid dots, while standard deviations $\sigma_N(\Delta H_f)$ are indicated by vertical lines extending from $\langle \Delta H_f \rangle_N - \sigma_N(\Delta H_f)$ to $\langle \Delta H_f \rangle_N + \sigma_N(\Delta H_f)$, and extremal ΔH_f values are indicated by crosses.**Figure 13.** Plots for magnetic susceptibility values $\langle \chi_{\text{mag}} \rangle_N$ in magnetic units ($1 \text{ mu} = -10^{-6} \text{ erg/(G}^2 \text{ mol)}$). Averages are indicated by solid dots, while standard deviations $\sigma_N(\chi_{\text{mag}})$ are indicated by a vertical line extending from $\langle \chi_{\text{mag}} \rangle_N - \sigma_N(\chi_{\text{mag}})$ to $\langle \chi_{\text{mag}} \rangle_N + \sigma_N(\chi_{\text{mag}})$, and extremal χ_{mag} values are indicated by crosses.

results reveal something about the degree to which an identification of an isomer class correlates to a particular value of a molecular property. It is seen that the heat of formation has at smaller N (say near $N = 10$) a standard deviation range so that there is no overlap to neighboring isomer-class ranges on either side, but as N increases the extent of overlapping isomer-class ranges also increases. That is, for ΔH_f there are structural variations which become ever larger as N increases, exceeding by ever larger margins the difference between neighboring isomer classes. However, the range of overlap as measured in terms of the percentage of $\langle X \rangle_N$ seems to go down as N increases—so that, the fraction of isomer classes up to a given N overlapped by the N -atom

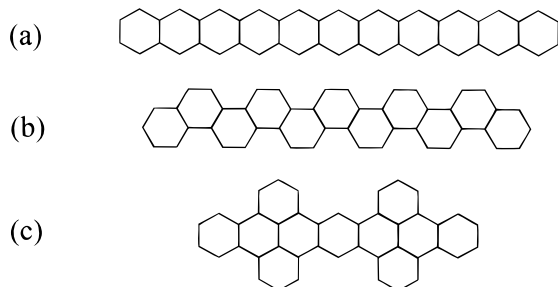


Figure 14. Three isomeric benzenoids ($C_{46}H_{26}$) of vastly different stabilities.

isomers seems ultimately to go down. That is, since the asymptotic forms reveal that the averages for ΔH_f and χ_{mag} scale as N while the corresponding standard deviations scale as $N^{1/2}$, the overlap ultimately increases as $\sim N^{1/2}$ while the fraction of N isomer classes (up to N) overlapped should be $\sim N^{-1/2}$. The value of the proportionality constant involved is neatly described in terms of a *separation number*

$$n(X) \equiv (1/2) \lim_{N \rightarrow \infty} \{(\langle X \rangle_{N+1} - \langle X \rangle_N) / N^{-1/2} \sigma_N(X)\}^2 \quad (8.1)$$

which is a sort of measure of an asymptotic accelerative rate at which N must increase in order to increase the extent of overlap: if at N the standard deviation overlaps the standard-deviation ranges of m isomer classes above (and also below), then in order to similarly overlap $m + 1$ isomer-class ranges, N would need to be increased by $\approx n(X)m$ additional carbon atoms. That is, though the overlap between isomer classes increases with N , the frequency at which the interclass overlap increases ever decreases. The rationale for the formula is presented in appendix B. For our two properties we find

$$n(\Delta H_f) \cong 28 \quad n(\chi_{\text{mag}}) \cong 1035 \quad (8.2)$$

Also there are higher-order terms in $\sigma_N(X)$ which at finite N modify the manner of approach to this asymptotic behavior (see Figures 6 and 7). Of course the property distributions extend out beyond the standard-deviation ranges. But even in considering the extremes of these distributions shown in Figures 12 and 13 they seem to give only modest overlap of these extreme tails. We note that for our present form of cluster expansion the deviations of the extremes from the average values should in general scale as $\sim N$ (whereas the standard deviations scale as $\sim N^{1/2}$).

The small interclass overlap for the properties considered here indicates that these alkane isomer classes are characterized not only in terms of their chemical formula but also to a large extent in terms of their properties (ΔH_f and χ_{mag}). Such a conclusion does not always seem to apply so well for other examples of isomer classes. For instance, if one examines the benzenoid isomers, one finds that for isomer classes of catacondensed species $C_N H_{2N+4}$ there are highly unstable polyacenes (as in Figure 14a), while at the same time there are fairly stable polyphenenes, as in Figure 14b, and highly stable (especially aromatic) so-called³¹ “fully benzenoid” species, as in Figure 14c. For our alkanes, if instead of focusing on ΔH_f and χ_{mag} , one attends to the diameters, then (as is seen in Figure 5) there is a substantial overlap between different isomer classes, and even more so

Table 5. Asymptotic Numerics for Alkanes^{a,b}

Ξ	α	β	A	B	C	κ
#struc, N	$-5/2$	1	0.6570	-0.068	6.2	2.815 460 0
#ster, N	$-5/2$	1	0.2879	-0.202	0.93	3.287 112 0
#conf, N	$-5/2$	1	0.3664	-0.574	0.6	27/4
$\langle \#_{\text{diast}} \rangle_{\text{struc},N}$	0	1	0.220	-0.181		1.167 522 2
$\langle \#_{\text{conf}} \rangle_{\text{ster},N}$	0	1	1.275	-1.282		2.053 474 2
$\langle \#_{\text{achir}} \rangle_{\text{ster},N}$	2		1.093			0.551 1
var(1,1)	1	1	0.0952	0.0305		1
cov(1,2)	1	1	-0.1441	-0.1007		1
cov(1,3)	1	1	0.0027	0.1098		1
cov(1,4)	1	1	0.0463	-0.0396		1
var(2,2)	1	1	0.2477	0.1668		1
cov(2,3)	1	1	-0.0630	-0.0316		1
cov(2,4)	1	1	-0.0405	-0.0345		1
var(3,3)	1	1	0.1181	-0.2662		1
cov(3,4)	1	1	-0.0577	0.1879		1
var(4,4)	1	1	0.0520	-0.1138		1
$-\langle \Delta H_f \rangle_N$	1	1	5.9433	7.218		1
$\langle \chi_{\text{mag}} \rangle_N$	1	1	11.726	4.872		1
$\langle \sigma_N(\Delta H_f) \rangle$	$1/2$	$1/2$	0.7836	0.0320		1
$\langle \sigma_N(\chi_{\text{mag}}) \rangle$	$1/2$	$1/2$	0.2576	-0.0330		1
$\langle D \rangle_N$	$1/2$	$1/2$	3.3117	-4.897		1
$\langle \sigma_N(D) \rangle$	$1/2$	$1/2$	0.5142	-0.7917		1

^a Asymptotic form: $\Xi = N^\alpha(A + BN^{-\beta} + CN^{-2\beta})\kappa^N$. ^b Results for entrees #struc, N and #conf, N were reported in ref 18.

if attention is paid to the extreme diameters (at $N - 1$ and $\sim \log N$).

9. ASYMPTOTICS

A further point of interest concerns the asymptotic behavior of the various quantities considered in Tables 1–3. In general one may imagine an averaged quantity Ξ_N at a carbon-atom count of N to exhibit an asymptotic form

$$\Xi_N \approx N^\alpha(A_\Xi + B_\Xi N^{-\beta} + C_\Xi N^{-2\beta})\kappa_\Xi^N \quad (9.1)$$

with the exponents α and β also dependent on Ξ but typically only taking simple rational values which depend only “weakly” on Ξ . Of course these exponents α and β are distinct from the labels of section 7 and Figure 9 for the extremal regions. The techniques of our earlier article¹⁸ enable one to obtain such asymptotic results as in Table 5. The first step in obtaining asymptotic parameters of eq 9.1 is the determination of the potentially dominant N -dependence associated with κ_Ξ (if $\neq 1$) or equivalently $t_c = 1/\kappa$. An efficient procedure which allows one to determine t_c quite readily is discussed in detail in our previous article,¹⁸ where we reported t_c values for structural isomers and conformers of alkanes. Here we only highlight some of the details as it relates to the determination of t_c for stereoisomers. As was pointed out in refs 18 and 32, our method for the determination of t_c does not always allow for an analytical solution as appears to be a case for stereoisomers of alkanes. This case involves the following generating function for arbitrary-diameter rooted stereomer structures

$$\phi(t) = 1 + (t/3)\{\phi(t)^3 + 2\phi(t^3)\} \quad (9.2)$$

Due to the presence of the term $\phi(t^3)$ one cannot so straightforwardly use this equation to solve for t_c as it was done for conformer counts in ref 18. Thus, one starts from high powers $z > 2^M$ of t , where $t^z \approx 0$ and $\phi(t^z) \approx \phi(0) = 1$, and then recur downward, obtaining at the k th step $\phi(t^{z(k)})$, where $z(k) \equiv z_k = 3^{(M-k)}$, until at the M th step we

reach $\phi(t^1) = \phi(t)$. The choice of powers z for t ($0 < t < 1$) is made with reference to a suitable small tolerance ϵ such that $t^z \ll \epsilon$, so that $\phi(t^z) = 1 + O(\epsilon)$ (while $\phi(t) = 1 + O(t)$). The procedure we just outlined implies that now one has to deal with the following equation

$$(\phi(t^z))^3 - (3/t^z)\phi(t^z) + (2\phi(t^{3z}) + (3/t^z)) = 0 \quad (9.3)$$

which follows directly from eq 9.2 with t replaced by t^z . The recursion at each step involves the solution of a polynomial (which here is a cubic) equation for $\phi(t^{z(k)})$ in terms of t and higher order $\phi(t^{z(k-1)})$ (which has already been obtained), where $z(k) = 2^{M-k}$ and $z(k-1) = 2^{(M-k+1)}$. Furthermore, if our guessed value for t was $t < t_c$, then the recursion procedure leads to a real polynomial root at the last step; otherwise if $t > t_c$, it does not. Thus within a chosen tolerance ϵ one can rather easily decide whether a given t is $> t_c$ or $< t_c$. Therefore, one can use the following procedure for deducing t_c (see refs 18 and 32):

- (1) Choose bounds $0 < t < 1$; set the initial $t = 0$ and a step size $\Delta t = 0.1$.
- (2) Increase t by Δt and calculate $\phi(t)$ as discussed near (9.2).
- (3) If the resulting root for $\phi(t)$ is *real*, then return to (2),- whereas if the root is *imaginary*, then back up one step (diminishing t by Δt) and go to (4).
- (4) If the desired accuracy (Δt) is not yet achieved, then diminish the stepsize by a factor of 10 and return to (2).

Following this procedure, after at the most 10×13 computations of $\phi(t)$ one arrives at a t_c estimate within our chosen tolerance, of $\epsilon = 10^{-13}$. The so obtained t_c critical values for alkane stereoisomer (i.e., $t_c^{(\text{ster})} \cong 0.304\,218\,409\,0746$) and structural-isomer (i.e., $t_c^{(\text{struc})} \cong 0.355\,181\,742\,3143$) counts lead to the growth factors

$$\begin{aligned} \kappa_{\text{ster}} &\cong 3.287\,112\,055\,58 \\ \kappa_{\text{struc}} &\cong 2.815\,460\,033\,17 \end{aligned} \quad (9.4)$$

where the accuracy for the determination of t_c for structural-isomer counts has been extended a few more digits than previously.¹⁸ Note also that the ranges of $0.30 < t_c^{(\text{ster})} < 0.31$ and $0.35 < t_c^{(\text{struc})} < 0.36$ were already given by Pólya.⁶ Now, with accurate κ values in hand, subsequent fittings with the κ^N dependence divided out are via least squares procedures. The results of asymptotic studies regarding data reported in Tables 1–3 are presented in Table 5, where values given as integers or as rational fractions are believed to be exact, while the decimal values are believed to be approximate. Blanks should presumably be nonzero but were not included in the fitting. E.g., for the case of $\langle \#_{\text{achir}} \rangle_{\text{ster},N}$ the straightness of the plot in Figure 15 of $\ln(\langle \#_{\text{achir}} \rangle_{\text{ster},N}/N^2)$ vs N indicates that whatever B -coefficient might apply it should be quite small. Indeed this plot is so near linear (for large N) we might speculate that B is 0 (exactly). In this plot the slope ($= \ln \kappa$) and the intercept ($= \ln A$) allow us to obtain the asymptotic fit constants reported in Table 5. The ranges of data points used in the weighted least squares fitting procedure were chosen differently for different properties. In particular, the range $\{15 \rightarrow 65\}$ was selected for the stereoisomer count fit; the range $\{20 \rightarrow 65\}$ was for $\langle \#_{\text{dias}} \rangle_{\text{struc},N}$; the range $\{15 \rightarrow 40\}$ was for $\langle \#_{\text{conf}} \rangle_{\text{ster},N}$ and for covariance data as well as property standard deviations; the range $\{10 \rightarrow 40\}$ was used for the fits of average property

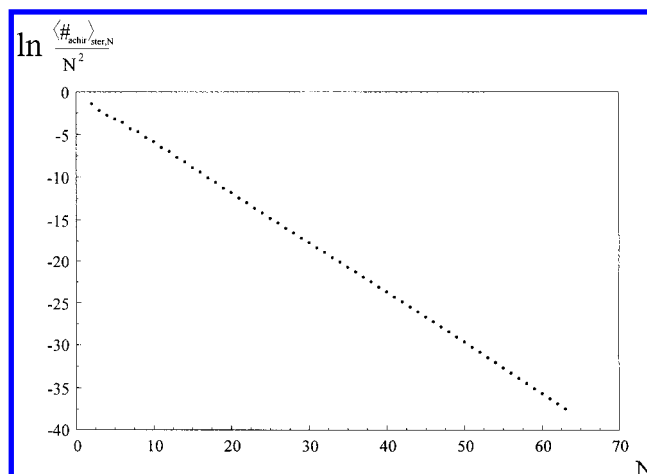


Figure 15. Plot of $\ln(\langle \#_{\text{achir}} \rangle_{\text{ster},N}/N^2)$ vs N .

data; and finally the range $\{40 \rightarrow 65\}$ was used for the fits of the stereoisomer diameter data. For more details regarding the fitting procedure, see ref 18. The numerical results for the standard deviations of ΔH_f and χ_{mag} differ from those given¹⁸ earlier because these earlier results were for the “absoluted” standard deviations. We also note that some of the data presented in Table 5 have been reported earlier, i.e., in our earlier paper¹⁸ for structural isomers and conformers and also by Davies and Freyd in ref 10 for structural isomers only. The results for structural isomers given by refs 18 and 10 essentially agree if the typo-kind error in ref 10 is corrected (namely, the numerical values for the constants A and B for the asymptotic form presented in ref 10 need to be interchanged).

10. CONCLUSIONS

Hence, an interesting and hopefully useful formal “combinatoric” chemistry is indicated, to yield not only enumerations but also (via cluster expansion) some alkane-isomer-class averaged properties, associated standard deviations, and their extremal property values. These results along with our identification of extremal structures add to our earlier illustration¹⁸ of the derivation of asymptotics (and a brief indication of a relation to systematic nomenclature). Such combinatoric techniques may be especially desirable with the combinatoric collections of molecules other than the alkanes, Pólya-theoretic enumeration techniques in fact being well adapted to treating general uncorrelated chemical substitutional isomerism on a fixed “skeleton”. That is, much of what has been done should extend to a number of other collections of molecules. As a further illustration we have considered elsewhere³² fully conjugated acyclic polyenes (and related radicals), and work on all acyclic hydrocarbons (regardless of degree of unsaturation) is under way.³³ Of course properties beyond those which are often described as “locally additive” could be especially interesting, in that optimization of a property that is “locally additive” evidently is especially straightforward, involving the inclusion of an extreme number of the local groups with an extreme local value for this property. Thus the earlier illustrated¹⁸ ability to compute the mean “Wiener” number is perhaps of interest in that it is often viewed as not “locally additive” but rather a “global” characteristic. It may be that one can ultimately mathematically simulate a number of aspects of current experimental dendrimer chemistry³⁴ or of what is currently

called³⁵ “combinatorial chemistry”. Standard isomer enumeration problems are still actively pursued, e.g., concerning different substitutional isomers³⁶ of buckminsterfullerene. But as illustrated here (and in ref 18) there seems to be much more chemically interesting combinatorics that can be developed for various areas of chemistry.

ACKNOWLEDGMENT

The authors acknowledge the support of the Welch Foundation of Houston, TX.

APPENDIX A: ABSOLUTED STANDARD DEVIATION BOUND

To establish the bound of (6.6) for the standard deviation, we start with the evident bound

$$\sigma_N(X)^2 \leq \sum_{\xi, \zeta}^G |\text{cov}_N(\xi, \zeta)| |X(\nu, \xi) X(\nu, \zeta)| \quad (\text{A1})$$

But through the use of the Schwartz inequality one sees that

$$\begin{aligned} |\text{cov}_N(\xi, \zeta)|^2 &= |(\langle \nu_\xi - \langle \nu_\xi \rangle_N | \nu_\xi - \langle \nu_\xi \rangle_N \rangle_N)|^2 \\ &\leq \langle (\nu_\xi - \langle \nu_\xi \rangle_N)(\nu_\xi - \langle \nu_\xi \rangle_N) \rangle_N \langle (\nu_\zeta - \langle \nu_\zeta \rangle_N)(\nu_\zeta - \langle \nu_\zeta \rangle_N) \rangle_N \end{aligned} \quad (\text{A2})$$

where we have here abbreviated $\nu(\xi, G) = \nu_\xi$ (and similarly for ζ). Then noting that the result on the right-hand side of this inequality is the square of $\sigma_N(\nu_\xi)$ $\sigma_N(\nu_\zeta)$, one may substitute into (A1), to obtain

$$\begin{aligned} \sigma_N(X)^2 &\leq \sum_{\xi, \zeta}^G \sigma_N(\nu_\xi) \sigma_N(\nu_\zeta) |X(\nu, \xi) X(\nu, \zeta)| = \\ &\quad \left\{ \sum_{\xi}^G \sigma_N(\nu_\xi) |X(\nu, \xi)| \right\}^2 \end{aligned} \quad (\text{A3})$$

which then establishes the desired bound $\sigma_N(X) \leq \sigma_{\text{abs}, N}(X)$.

APPENDIX B: SEPARATION NUMBER RATIONALE

Formula 8.1 for the separation number measuring the quality of the isomer-class distinguishability provided by a property X is to be rationalized. The asymptotic difference between successive isomer-class property averages is denoted

$$\Delta_X \equiv \lim_{N \rightarrow \infty} \{ \langle X \rangle_{N+1} - \langle X \rangle_N \} \quad (\text{B1})$$

Then $\langle X \rangle_N \rightarrow N\Delta_X + (\text{constant})$, as is expected for bulk (i.e., extensive) properties. Corresponding decorrelated standard deviations $\sigma_{\text{dis}, N}(X)$ (and the $\sigma_N(N_d)$ from which they are comprised) have already been argued¹⁸ to scale as $N^{1/2}$. And the proper standard deviations then must also similarly scale, because of the similarity in formulations for the decorrelated standard deviation, absolved standard deviation, and the proper standard deviation. That is, we should have a nonzero limit

$$s_X \equiv \lim_{N \rightarrow \infty} N^{-1/2} \sigma_N(X) \quad (\text{B2})$$

with $\sigma_N(X) \rightarrow s_X N^{1/2} + (\text{constant})$. Then if we consider the value for $N = N_m$ such that the standard-deviation range for

$X(G)$ of N -site G overlaps the standard-deviation range for m adjacent isomer classes (above and below), then (at least for the large- N asymptotic realm)

$$2N^{1/2} s_X \approx 2\sigma_N(X) = \langle X \rangle_{N+m} - \langle X \rangle_N \approx \Delta_X m \quad (\text{B3})$$

and evidently the value of N at which this m -fold overlap occurs is

$$N \approx (\Delta_X / 2s_X)^2 m^2 \quad (\text{B4})$$

Then the rate at which N changes per unit of overlap (i.e., the increase in N needed to achieve overlap to one more isomer class) would be

$$\frac{\partial}{\partial m} (\Delta_X / 2s_X)^2 m^2 \approx m (\Delta_X^2 / 2s_X^2) \quad (\text{B5})$$

which is essentially just $n(X)m$, with $n(X)$ having the form of eq 8.1. Hence, this motivates $n(X)$ as a separation measure.

REFERENCES AND NOTES

- (1) Slanina, Z. *Contemporary Theory of Chemical Isomerism*; Reidel Pub. Co.: Dordrecht, The Netherlands, 1986.
- (2) Cayley, A. On the Mathematical Theory of Isomers. *Philos. Mag.* **1874**, 47, 444. Cayley, A. On the Analytical Forms Called Trees with Applications to the Theory of Chemical Compounds. *Rep. Br. Assoc. Adv. Sci.* **1887**, 257–305.
- (3) Schiff, H. Zur Statistik Chemischer Verbindungen. *Ber. Deutsch. Chem. Ges.* **1876**, 8, 1542–1547. Flavitzky, F. *Ber. Deutsch. Chem. Ges.* **1876**, 9, 267. Delannoy, H. A. Sur les arbres geometriques et leur emploi dans la theorie des combinaisons chimiques. *C. R. Assoc. Fr. Avance. Sci.* **1894**, 23 (2), 102–116. Herrmann, F. Über das Problem die Anzahl der Isomeren Paraffine der Formel $C_n H_{2n+2}$ zu Bestimmen. *Ber. Deutsch. Chem. Ges.* **1897**, 9, 792, 2423–2426. Losanitsch, S. M. Die Isomerie-Arten bei den Homologender Paraffin-Riehe. *Ber. Deutsch. Chem. Ges.* **1897**, 30, 1917–1926.
- (4) Henze, H. R.; Blair, C. M. The number of structurally isomeric alcohols of the methanol series. *J. Am. Chem. Soc.* **1931**, 53, 3042–3046; The Number of Isomeric Hydrocarbons of the Methane Series. *J. Am. Chem. Soc.* **1931**, 53, 3077–3085. Blair, C. M.; Henze, H. R. The Number of Stereoisomeric & Non-Stereoisomeric Mono-Substitution Products of the Paraffins. *J. Am. Chem. Soc.* **1932**, 54, 1098–1106; The Number of Stereoisomeric & Non-Stereoisomeric Paraffin Hydrocarbons. *J. Am. Chem. Soc.* **1932**, 54, 1538–1545. Perry, A. The Number of Structural Isomers of Certain Homologs of Methane and Methanol. *J. Am. Chem. Soc.* **1932**, 54, 2918–2920. Coffman, D. D.; Blair, C. M.; Henze, H. R. *J. Am. Chem. Soc.* **1933**, 55, 252–253. Henze, H. R.; Blair, C. M. The number of structurally isomeric hydrocarbons of the ethylene series. *J. Am. Chem. Soc.* **1933**, 55, 680–686. Henze, H. R.; Blair, C. M. The Number of Structural Isomers of the More Important Types of Aliphatic Compounds. *J. Am. Chem. Soc.* **1934**, 56, 593.
- (5) Pólya, G. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta Math.* **1937**, 68, 145–254. Pólya, G. Algebraische Berechnung der Isomeren einiger Organischer Verbindungen. *Zeit. Kryst. A* **1936**, 93, 414–443. Pólya, G. Über das Anwachsen der Isomerenzahlen in den Homologen Reihe der Organischen Chemie. *Vierteljahrsschr. Naturforsch. Ges. Zürich* **1936**, 81, 243–258.
- (6) Pólya, G.; Read, R. C. *Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds*; Springer-Verlag: Berlin, 1987.
- (7) Ford, G. W.; Uhlenbeck, G. E. Combinatorial Problems in the Theory of Graphs. I. *Proc. Natl. Acad. Sci. U.S.A.* **1956**, 42, 122–128; II. *Proc. Natl. Acad. Sci. U.S.A.* **1956**, 42, 203–208. Ford, G. W.; Uhlenbeck, G. E. Combinatorial Problems in the Theory of Graphs. III. *Proc. Natl. Acad. Sci. U.S.A.* **1956**, 42, 529–535.
- (8) Robinson, R. W.; Harary, F.; Balaban, A. T. The Numbers of Chiral and Achiral Alkanes and Monosubstituted Alkanes. *Tetrahedron* **1976**, 32, 355–361.
- (9) Read, R. C. The Enumeration of Acyclic Chemical Compounds. In *Chemical Applications of Graph Theory*; Balaban, A. T., Ed.; Academic Press: New York, 1976; pp 25–61.
- (10) Davies, R. E.; Freyd, P. J. $C_{167}H_{336}$ Is the Smallest Alkane with More Realizable Isomers than the Observed Universe Has “Particles”. *J. Chem. Educ.* **1989**, 66, 278–281.

- (11) Knop, J. V.; Müller, W. R.; Szymanski, K.; Trinajstić, N. *Computer Generation of Certain Classes of Molecules*; SKTH/Kemija u industriji: Zagreb, Croatia, 1985. Trinajstić, N.; Nikolić, S.; Knop, J. V.; Müller, W. R.; Szymanski, K. *Computational Chemical Graph Theory*; Horwood: New York, 1991. Balasubramanian, K. Applications of Combinatorics and Graph Theory to Spectroscopy and Quantum Chemistry. *Chem. Rev.* **1985**, 85, 599–618. Fujita, S. *Symmetry and Combinatorial Enumeration in Chemistry*; Springer-Verlag: Berlin, 1991.
- (12) Harary, F.; Palmer, E. *Graphical Enumeration*; Academic Press: New York, 1973.
- (13) Ruch, E.; Hässelbarth, W.; Richter, B. Doppelnebenklassen als Klassenbegriff und Nomenklaturprinzip für Isomere und ihre Abzählung. *Theor. Chim. Acta* **1970**, 19, 288–300. Hässelbarth, W.; Ruch, E. Classifications of Rearrangement Mechanisms by Means of Double Cosets and Counting Formulas for the Numbers of Classes. *Theor. Chim. Acta* **1973**, 29, 259–268.
- (14) Nourse, J. G.; Mislow, K. Dynamic Stereochemistry of Tetraarylmethanes and Cognate Systems. The Role of the Permutation Subgroup Lattice. *J. Am. Chem. Soc.* **1975**, 97, 4571–4578. Klein, D. J.; Cowley, A. H. Permutational Isomerism. *J. Am. Chem. Soc.* **1975**, 97, 1633–1640. Klein, D. J.; Cowley, A. H. Permutational Isomerism with Bidentate Ligands and Other Constraints. *J. Am. Chem. Soc.* **1978**, 100, 2593–2599. Nourse, J. G. Self-Inverse and Non-Self-Inverse Degenerate Isomerizations. *J. Am. Chem. Soc.* **1980**, 102, 4883–4889. Brocas, J.; Gielen, M.; Willem, R. *The Permutational Approach to Dynamic Stereochemistry*; McGraw-Hill: New York, 1983.
- (15) Gordon, M.; Kennedy, J. W. The Counting and Coding of Trees of Fixed Diameter. *SIAM J. Appl. Math.* **1975**, 28, 376–398. Balaban, A. T.; Kennedy, J. W.; Quintas, L. The Number of Alkanes Having n Carbons and a Longest Chain of Length d . *J. Chem. Educ.* **1988**, 65, 304–313.
- (16) Klein, D. J.; Seitz, W. A. Graphs, Polymer Models, Excluded Volume, and Chemical Reality. In *Chemical Applications of Topology and Graph Theory*; King, R. B., Ed.; Elsevier: Amsterdam, 1983; pp 430–445.
- (17) Bailey, C.; Palmer, E.; Kennedy, J. W. Points by Degree and Orbit Size in Chemical Trees, I. In *The Theory and Applications of Graphs*; Wiley: New York, 1981; pp 27–43. Quintas, L. V.; Yarmish, J. Degree Distributions for Random 4-Trees and Skeletons of Symmetry Weighted (1,4)-Trees. *Congr. Num.* **1982**, 36, 115–125.
- (18) Bytautas, L.; Klein, D. J. Chemical Combinatorics for Alkane–Isomer Enumeration and More. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 1063–1078.
- (19) S. J. Cyvin's group and C.-Y. Yeh have taken such an approach, with representative applications to compounds other than alkanes including the following: Cyvin, S. J.; Cyvin, B. N.; Brunvoll, J.; Fujii, Z.; Xiaofeng, G.; Tosić, R. Graph Theoretical Studies on Fluoranthenoids and Fluorenoids: Enumeration of Some Catacondensed Systems. *J. Mol. Struct.* **1993**, 285, 179–185. Cyvin, S. J.; Cyvin, B. N.; Brunvoll, J.; Brendsdal, E. Enumeration and Classification of Certain Polygonal Systems Representing Polycyclic Conjugated Hydrocarbons: Annulated Catafusenes. *J. Chem. Inf. Comput. Sci.* **1994**, 34, 1174–1180. Cyvin, S. J.; Brunvoll, J.; Cyvin, B. N. Enumeration of Constitutional Isomers of Polyenes. *J. Mol. Struct. (THEOCHEM)* **1995**, 357, 255–261. Cyvin, S. J.; Brunvoll, J.; Cyvin, B. N.; Brendsdal, E. Enumeration of Isomers and Conformers: A Complete Mathematical Solution for Conjugated Polyene Hydrocarbons. *Adv. Mol. Struct. Res.* **1996**, 2, 213–245. Cyvin, S. J.; Cyvin, B. N.; Brunvoll, J.; Jianji, W. Enumeration of Staggered Conformers of Alkanes and Monocyclic Cycloalkanes. *J. Mol. Struct.* **1998**, 445, 127–137. Yeh, C.-Y. Theory of Acyclic Chemical Networks and Enumeration of Polyenes via Two-Dimensional Chirality. *J. Phys. Chem.* **1996**, 100, 15800–15804. Yeh, C.-Y. Counting Linear Polyenes by Excluding Structures with Steric Strain. *J. Chem. Phys.* **1996**, 105, 9706–9707. Yeh, C.-Y. Isomer Enumeration of Alkenes and Aliphatic Cyclopropane Derivatives. *J. Chem. Inf. Comput. Sci.* **1996**, 36, 854–856.
- (20) Yeh, C.-Y. Isomer enumeration of alkanes, labeled alkanes, and monosubstituted alkanes. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 912–913. Brunvoll, J.; Cyvin, B. N.; Brendsdal, E.; Cyvin, S. J. Computerized Enumeration of Staggered Alkane Conformers. *Comput. Chem.* **1995**, 19, 379–388. Cyvin, S. J.; Jianji, W.; Brunvoll, J.; Cao, S.; Li, Y.; Cyvin, B. N.; Wang, Y. Staggered conformers of alkanes: Complete solution of the enumeration problem. *J. Mol. Struct.* **1997**, 413–414, 227–239. Yeh, C.-Y. Isomerism of asymmetric dendrimers and stereoisomerism of alkanes. *J. Mol. Struct. (THEOCHEM)* **1998**, 432, 153–156.
- (21) Dreiding, A. S.; Wirth, K. The Multiplex—A Classification of Finite Ordered Point Sets in Oriented d -Dimensional Spaces. *MatCh* **1980**, 8, 341–352. Dress, A. Chirotopes and oriented matroids: Discrete structures, algebraic methods and applications. *Bayreuth. Math. Schrift.* **1986**, 21, 14–68. Tratch, S. S.; Zefirov, N. S. Combinatorial Models and Algorithms in Chemistry. Ladder of Combinatorial Objects and Its Application to Formalization of Structural Problems in Organic Chemistry. In *Principles of Symmetry and Systemology in Chemistry*; Stepanov, F. N., Ed.; Moscow State University Press: Moscow, 1987; pp 54–86. Zlatina, L. A.; Elyashberg, M. E. Generation of stereoisomers and their spatial models corresponding to the given molecular structure. *MatCh* **1992**, 27, 191–207. Tratch, S. S.; Zefirov, N. S. Algebraic Chirality Criteria and Their Applications to Chirality Classification in Rigid Molecular Systems. *J. Chem. Inf. Comput. Sci.* **1996**, 36, 448–464.
- (22) Klein, D. J.; Seitz, W. A.; Kilpatrick, J. E. Branched Polymer Models. *J. Appl. Phys.* **1982**, 53, 6599–6603. Cyvin, S. J.; Brunvoll, J.; Cyvin, B. N.; Lüttke, W. Enumeration of the Staggered Conformers of Alkanes. *Z. Naturforsch.* **1995**, 50a, 857–863.
- (23) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Wiley Interscience: New York, 1969.
- (24) Smolenskii, E. A. (English translator). Application of the Theory of Graphs to Calculations of the Additive Structural Properties of Hydrocarbons. *Russ. J. Phys. Chem.* **1964**, 35, 700–702. Gordon, M.; Kennedy, J. W. The graph-like state of matter. Part 2. LCGI Schemes for the thermodynamics of alkanes and the theory of inductive inference. *J. Chem. Soc., Faraday Trans. 2* **1973**, 69, 484–504. Essam, J. W.; Gordon, M.; Kennedy, J. W.; Whittle, P. The graph-like state of matter. Part 8. LCGI schemes and the statistical analysis of experimental data. *J. Chem. Soc., Faraday Trans. 2* **1977**, 73, 1289–1307. Klein, D. J. Chemical Graph-Theoretic Cluster Expansions. *Intl. J. Quantum Chem., Quantum Chem. Symp.* **1986**, 20, 153–173.
- (25) Schmalz, T. G.; Klein, D. J.; Sandleback, B. L. Chemical Graph-Theoretic Cluster Expansion and Diamagnetic Susceptibility. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 54–57. Bytautas, L.; Schmalz, T. G.; Klein, D. J. To be submitted for publication.
- (26) Tatevskii, V. M.; Benderskii, V. A.; Yarovoi, S. S. *Rules & Methods for Calculating the Physico-Chemical Properties of Paraffinic Hydrocarbons* (English Translation); Pergamon Press: Oxford, U.K., 1961.
- (27) Benson, S. W. *Thermochemical Kinetics*; Wiley: New York, 1976. Cohen, N.; Benson, S. W. Estimation of Heats of Formation of Organic Compounds by Additivity Methods. *Chem. Rev.* **1993**, 93, 2419.
- (28) Vogel, A. I. Physical Properties and Chemical Constitution. Part XXIII. Miscellaneous Compounds. *J. Chem. Soc.* **1948**, 1833–1854. Vogel, A. I.; Cresswell, W. T.; Jeffery, G. H.; Leicester, J. Physical Properties and Chemical Constitution. Part XXIII. Aliphatic Alloximes, Ketoximes, and Ketoxime O-Alkyl Ethers, N,N-Dialkylhydrazines, Aliphatic Ketazines, Mono- & Dialkylaminopropionitriles, Dialkyl Azodiformates, and Dialkyl Carbonates. *J. Chem. Soc.* **1952**, 514–549.
- (29) See, e.g.: Borgwardt, K. H. *The Simplex Method*; Springer-Verlag: Berlin, 1987. Karloff, H. *Linear Programming*; Progress in Theoretical Computer Science; Birkhauser: Boston 1991.
- (30) Klein, D. J. Rigorous results for branched polymer models with excluded volume. *J. Chem. Phys.* **1981**, 75, 5186–5189. deGennes, P. G.; Hervet, H. Statistics of 'Starburst' Polymers. *J. Phys. Lett.* **1983**, 44, 351–360. Newkome, G. R.; Gupta, V. K.; Baker, G. R.; Yao, Z.-Q. Cascade Molecules: A New Approach to Micelle A [27]-Arborol. *J. Org. Chem.* **1985**, 50, 2003–2008.
- (31) Clar, E. *The Aromatic Sextet*; Wiley: New York, 1972.
- (32) Bytautas, L.; Klein, D. J. Isomer Combinatorics for Acyclic Conjugated Polyenes: Enumeration and Beyond. *Theor. Chem. Acc.* **1999**, 101, 371–387.
- (33) Bytautas, L.; Klein, D. J. *Croat. Chem. Acta*, accepted for publication.
- (34) Tomalia, D. A.; Naylor, A. M.; Goddard, W. A., III. Starburst Dendrimers: Control of Size, Shape, Surface Chemistry, Topology, and Flexibility in the Conversion of Atoms to Macroscopic Materials. *Angew. Chem., Intl. Ed. Engl.* **1990**, 102, 138–175. Frechet, J. M. J. Functional Polymers and Dendrimers-Reactivity, Molecular Architecture, and Interfacial Energy. *Science* **1994**, 263, 1710–1715. Tzialis, D.; Tor, Y. Toward Self-Assembling Dendrimers—Metal Complexation Induces the Assembly of Hyperbranched Structures. *Tetrahedron Lett.* **1996**, 37, 8293–8296. Zeng, F.; Zimmerman, S. C. Dendrimers in Supramolecular Chemistry: From Molecular Recognition to Self-Assembly. *Chem. Rev.* **1997**, 97, 1681–1712.
- (35) Czarnik, A. W.; DeWitt, S. H., Eds. *Combinatorial Chemistry*; American Chemical Society: Washington, DC, 1997.
- (36) Fujita, S. Soccerane Derivatives of Given Symmetries. *Bull. Chem. Soc. Jpn.* **1991**, 64, 3215–3223. Balasubramanian, K. Combinatorics of Cluster Enumeration. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 47–54. Balasubramanian, K. Enumeration of Chiral and Positional Isomers of Substituted Fullerene Cages (C_{20} – C_{70}). *J. Phys. Chem.* **1993**, 97, 6990–6998. Fowler, P. W. Isomer Counting using Point Group Symmetry. *J. Chem. Soc., Faraday Trans.* **1995**, 91, 2241–2247. Shao, Y.; Wu, J.; Jiang, Y. Isomer counting and isomer permutation representation. *Chem. Phys. Lett.* **1996**, 248, 366–372.