

Posetic cluster expansion for substitution-reaction networks and application to methylated cyclobutanes

Teodora Ivanciuc and Douglas J. Klein*

Texas A&M University at Galveston, Galveston, TX 77553, USA
E-mail: kleind@tamug.edu

Ovidiu Ivanciuc

Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, TX 77555-0857, USA

Received 29 June 2005; revised 20 August 2005

Substitution-reaction networks frequently arise in chemistry and are naturally identifiable as “partially ordered sets” (or posets). The direction of the reaction identifies the partial ordering of the set of the member molecular species, for which various properties may be developed in light of the partial ordering. Here some general characteristics of such posets are noted, and a chemically appealing posetic cluster-expansion analysis of molecular properties is described and illustrated for methyl substitutions on cyclobutane. The results for computed energies and a few further selected properties are compared to alternative posetic approaches.

KEY WORDS: reaction networks, partial ordering, cluster expansion, property correlations

1. Introduction

Progressive reaction networks are ubiquitous throughout chemistry, as has been earlier emphasized [1,2], with a listing of several examples. A typical example considers a molecular skeleton upon, which substituents are progressively introduced one after another, with the earlier pattern of substituents not rearranging. A classical example concerns the addition of substituents of a given type to benzene, and this example in particular played [3] an important historical role in characterizing the structure of benzene. But there are many further examples [1]. Thus one may desire methodology to utilize such a reaction network, which mathematically may be viewed as a “partial ordering”. That is, if structure ζ can so result (i.e., can be so obtained by substitutive reaction) from

*Corresponding author.

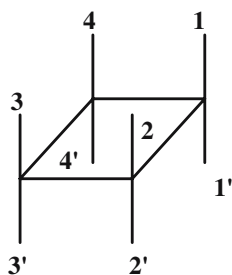


Figure 1. Standard numbering for the cyclobutane skeleton.

structure ξ , then we write $\xi \triangleright \zeta$, and identify \triangleright as a *partial ordering* (with $\xi \triangleright \zeta$ precluding $\zeta \triangleright \xi$; and also $\xi \triangleright \zeta$ and $\zeta \triangleright \eta$ implying $\xi \triangleright \eta$). To utilize these posets with elements representing chemical compounds it is naturally useful to have various means by which to interpolate on the poset so that two methods have been previously developed [2,4]. And further one may inquire about, and hopefully use, special characteristics of progressive reaction networks.

Here these goals are pursued, being illustrated in terms of an example of some modestly greater complexity than the earlier considered 13-member network of substitution in benzene. To this end consider the 8-position cyclobutane skeleton as indicated in figure 1. The choice $G=D_{4h}$ for the symmetry group distinguishes geometric isomers short of attention to full stereoisomer distinction. The consequent substitution poset for cyclobutane has 34 elements with a reaction diagram as in figure 2. In the iconic representation of the different isomeric substitution patterns, the H atoms are suppressed, and the presence of a substituent (such as a chloro or methyl group – but the same throughout the diagram) is indicated by a simple stick bond appended to the underlying square cyclobutane skeleton. Such an overall directed diagram (without the designations of the cyclobutane structures) is called a *Hasse diagram* in the mathematical context of posets.

Here different characteristics and an interpolative method for such posetic reaction networks are considered, and illustrated in the framework of this simple cyclobutane example. In as much as substitution-reaction posets offer a mathematically seemingly newly formulated special type of poset, some of their general characteristics are noted briefly in the next section, with special reference to the underlying substitution skeleton and its symmetry. Thereafter we address a primary chemical concern, of a scheme by which to utilize the posetic structure manifested by such reaction networks. This use involves a poset-attentive representation of a physico-chemical property *via* posetic substructural “cluster expansion”. Then an example property, namely molecular-mechanics energies of the polymethyl-cyclobutanes as here computed, is used to illustrate an application of the cluster expansion, and a comparison is made to two other earlier

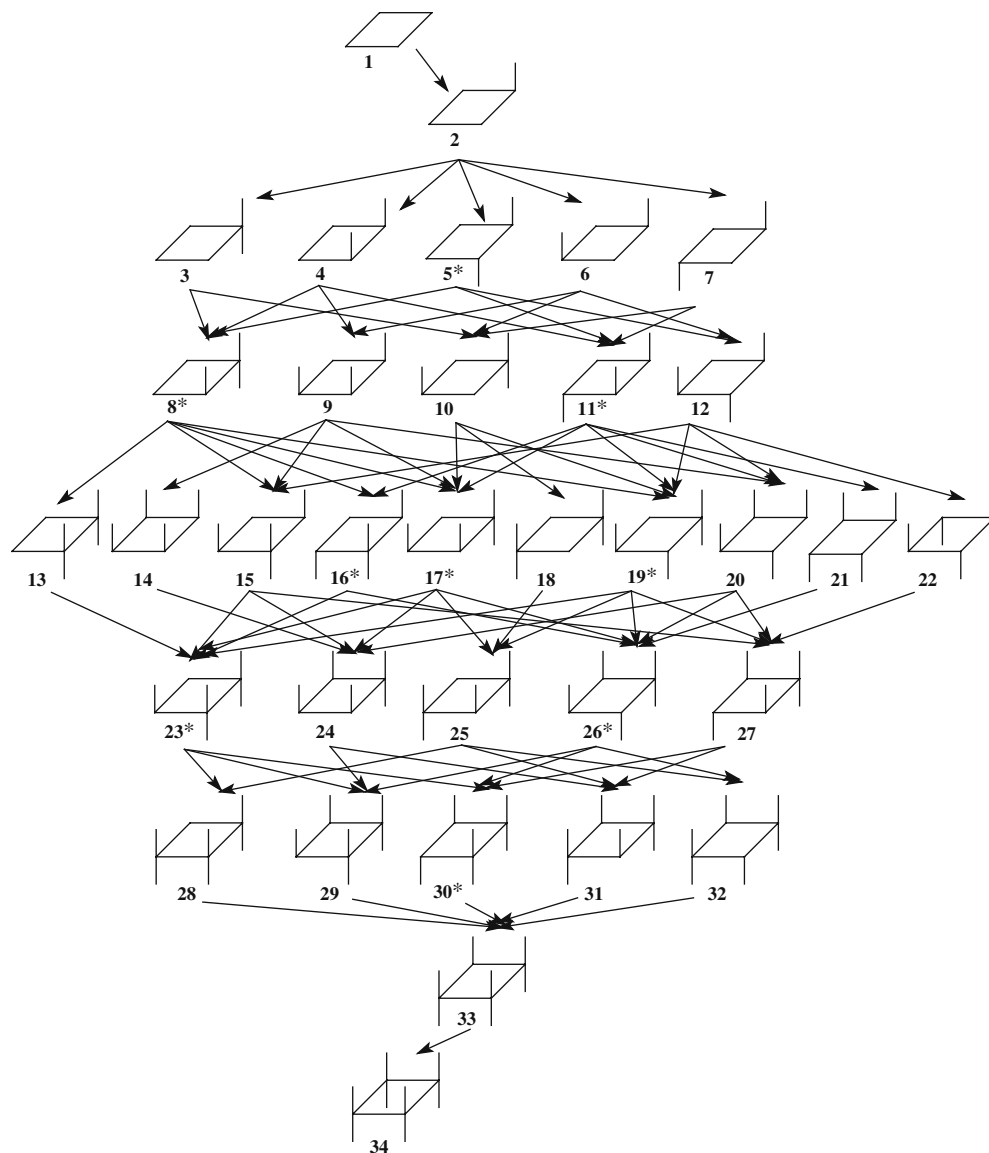


Figure 2. The cyclo-butane substitution reaction poset, showing diastereomers, based on skeletal symmetry group $G=D_{4h}$. The appended lines on cyclobutane squares indicate the sites on which an H atom of C_4H_8 has been replaced by the substituting ligand. Asterisks indicate diastereomers consisting of pairs of chiral enantiomers.

developed poset-attentive methods by which to deal with properties of the compounds identified to the members of the reaction poset. Yet further the “exact” cluster expansion is considered, again with illustration on the current molecular mechanics energies. Finally still for polymethyl-cyclobutanes, applications are made to experimentally measured boiling points, standard enthalpy of vaporiza-

tion, octanol/water partition coefficients, and logarithms of vapor pressures, for which we have found data for but a fraction ($\approx 1/2$) of the 34 possible species. Again comparison is made to an earlier splinoid scheme [4] by which to fit data.

2. Substitution-reaction poset characteristics

As a sort of preliminary inquiry several different general poset characteristics of such “substitution posets” may be noted, illustrated in the context of the cyclobutane example. With the designation of the present skeleton by \square , an n -substituted species might be indicated $\square X_n$ all of which occur in the Hasse diagrams’ n th row corresponding to the n th rank, for $n = 0-8$. The different species at the n th rank might be summarily indicated just by $i_1, i_2, \dots, i_n - \square$ with i_1, i_2, \dots, i_n the substituent-occupied positions, which may be conveniently chosen to be the lexicographically as small as possible. For example, the five disubstituted cyclobutanes would be denoted 1, 1' - \square , 1, 2 - \square , 1, 2' - \square , 1, 3 - \square , and 1, 3' - \square .

The overall appearance of the Hasse diagram may be characterized. The *height* of a poset is the length of the longest top-to-bottom chain in the network. With the substitution reaction posets being ranked, this height is evidently just 1 less than the number of ranks, so that the height is eight for the cyclobutane example. The poset *width* may be taken as the number of members of the Hasse diagram’s most extensive row (here the row at rank 4, which then gives width 10). Substitution-reaction posets are *self-dual* in the sense that the global posetic connectivity pattern of the Hasse diagram remains unchanged if the diagram is inverted with the reversal of the arrows (and if one wishes to attend to the cyclobutane structures for each isomer, then one interchanges X and H). That is, such substitution-reaction posets generally manifest a Hasse diagram with a sort of reflection symmetry about the mid-rank of the diagram.

Substitution-reaction posets \mathbb{P} are neatly definable in terms of a previous formal description [5,6] for the specification of individual isomers in terms of a permutation group G with elements acting on the skeletal positions whereat the substituents may be appended. That is, there is a set S of potential substitution sites on the molecular skeleton, and given a *configuration* with n substituents at positions $\{i_1, i_2, \dots, i_n\} \equiv C \subseteq S$, other configurations, $\{i_{p_1}, i_{p_2}, \dots, i_{p_n}\} \equiv PC$ for $P \in G$ are deemed *equivalent*. It is the equivalence classes of configurations which comprise the elements of the substitution-reaction poset \mathbb{P} – and for $\xi, \zeta \in \mathbb{P}$ one has $\xi \triangleright \zeta$ iff there are configurations $C \in \xi$ and $C' \in \zeta$ such that $C \subset C'$. Thence the poset depends on the set S of available skeletal substitution sites and the group G drawing the equivalences amongst subsets C of S .

Different choices of group G yield different posets, even with S fixed. In figure 2 we have assumed that G is specified as the permutation group

$$D_{4h} \equiv \{I, (11')(22')(33')(44')\}D_4, \quad (1)$$

where I is the identity permutation, (i_1, i_2, \dots, i_m) indicates a cyclic permutation taking i_k to i_{k+1} with $i_{m+1} \equiv i_1$, and D_4 is the group specified by

$$\begin{aligned} D_4 &\equiv \{I, (11')(24')(33')(42')\}C_4, \\ C_4 &\equiv \{I, (1234)(1'2'3'4'), (13)(24)(1'3')(2'4'), (1432)(1'4'3'2')\}. \end{aligned} \quad (2)$$

Here each of these permutation groups is isomorphic to a correspondent point group, from which the names are taken. But instead of D_{4h} , one might choose the symmetry group for cyclobutane to be $G = D_4$, thereby imagining that chiral mirror-image structures can be distinguished. In this case this (smaller) group entails greater distinguishing abilities and the isomers marked (in figure 2) with an asterisk (*) break up into enantiomeric pairs in the enlarged poset. The connections of the original poset then also break up, into pairs of connections, when one end of the connection is asterisked and the other is not. When both ends of a connection bear an asterisk (i.e., both reactant and product are chiral), then the connection is still broken in two though not into four connections (despite there being four conceivable pairings of chiral reactant and chiral product) – essentially with the two connection lines being “mirror images” of one another. The reaction diagram for $G = D_{4h}$ is evidently a condensation of that for D_4 , though we do not seek to show the diagram for D_4 in detail, as it is a little more cumbersome (with 43 members). A yet further alternative choice for the symmetry group could be as

$$G = \{I, (12)\}\{I, (34)\}, \{I, (56)\}\{I, (78)\}D_4 \quad (3)$$

whence only the graphical connectivity of the molecular structure is relevant, and the partial ordering becomes that of being a proper subgraph. In this case the isomers might be termed “constitutional” (or “structural”). Evidently there is less distinguishing amongst the (orientational) configurations, and the resultant poset is a condensation of that of figure 2. The 21-member result is as in figure 3, where now we mark with asterisks those isomers which break up if one reduces the symmetry group back to D_{4h} . For a general substitution-reaction poset it is evident that no isomer under a group G may break into more than “ $|G'|/|G|$ ”, isomers under the refinement to a subgroup $G \subseteq G'$, and the poset for G' is a (rank-preserving) condensation of that for G . (Here $|A|$ denotes the order of a set A .)

In general one sees that each member of the poset is represented by a set of configurations, obtainable from any member configuration C comprised from sites i_1, i_2, \dots, i_n by the action of $P \in G$ on each of these sites to obtain a configuration PC . Each of these configurations itself has a symmetry group which is the intersection of G with the subgroup leaving the configuration invariant – here the group which is a product of the group S_C of all permutations on the set $\{i_1, i_2, \dots, i_n\} = C$ of substituted sites, and the group S_{C^*} where C^* is the *dual*

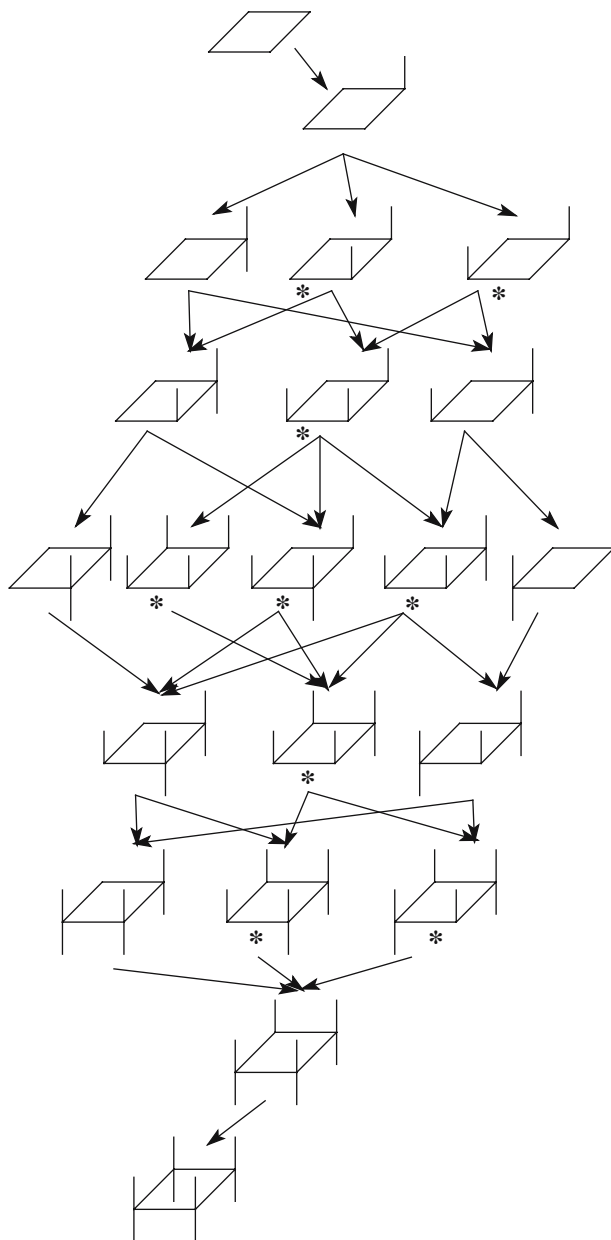


Figure 3. The cyclo-butane substitution reaction poset, showing structural isomers, based on G as in equation (1). The notation is as in figure 2, but with asterisks indicating current isomers breaking up into 2 or more isomers of figure 2.

set of skeletal sites without substituents. That is, the symmetry group of the configuration C is

$$G \cap (S_C \times S_{C^*}) \equiv G_C \quad (4)$$

and the symmetry groups for all other configurations equivalent to C are isomorphic. Each isomer $\xi \in \mathcal{P}$ then is labelled (up to isomorphism) by its own permutation group, which corresponds to a molecular point group for ξ when G itself does. Examples of these isomer symmetry groups $G_\xi (\approx G_C$ for $C \in \xi$) are indicated in figure 4. There also are shown the numbers of “types” of substituted and of unsubstituted sites, where sites i and j in a configuration C are of the same *type* iff there is a $P \in G_C$ such that $Pi=j$. One may say also that such i and j are “symmetry equivalent”, or that i and j are in the same “orbit” of sites with regard to the skeletal symmetry group G . Notably the number of orbits of substituted sites for $\xi \in \mathcal{P}$ identifies the number $d_{\xi \rightarrow}$ of successors in the Hasse diagram, and the number of orbits of substituted sites identifies the number $d_{\rightarrow \xi}$ of predecessors in the Hasse diagram. Further, at a given rank, these numbers evidently decrease as the symmetry group “increases”. That is, if G_ξ is a subgroup of G_ζ for ξ and ζ at the same rank, then $d_{\zeta \rightarrow} \geq d_{\xi \rightarrow}$ and $d_{\rightarrow \zeta} \geq d_{\rightarrow \xi}$.

Evidently there is an intimate relation between the permutation-group-theoretic representation of isomers and the current poset. It seems that the poset expressing the isomers in terms of sets of configurations has embedded within it different sorts of information, possibly of some chemical utility. One is tempted to seek to utilize the so-represented formal poset for further chemical applications.

3. Posetic cluster-expansion

It is rather natural to consider that different chemical properties might be ordered in concert with the partial ordering of a progressive reaction poset. Thence different ways of fitting data assigned to the different members of a poset might naturally be contemplated. Already we have investigated [2] a systematic simple-averaging scheme for interpolation of property values in the special case that one desires a property value for which all the immediately neighboring property values in the associated Hasse diagram are known. But here we describe yet another scheme by way of “cluster expansion”, which also has a long chemical history, dating back before molecular-structure ideas were very recognizable, ~ 150 years ago. See, e.g., Hermann Kopp’s review [7] of 1855, often dealing with homologous series. These ideas for characterization of molecular properties are manifested in more modern chemical literature as substructural cluster expansions, e.g., for heats of formation [8], or for a great diversity of other properties [9]. In a systematic formal mathematical framework, related expansions were generalized in the context of arbitrary partial orderings in Gian-Carlo Rota’s

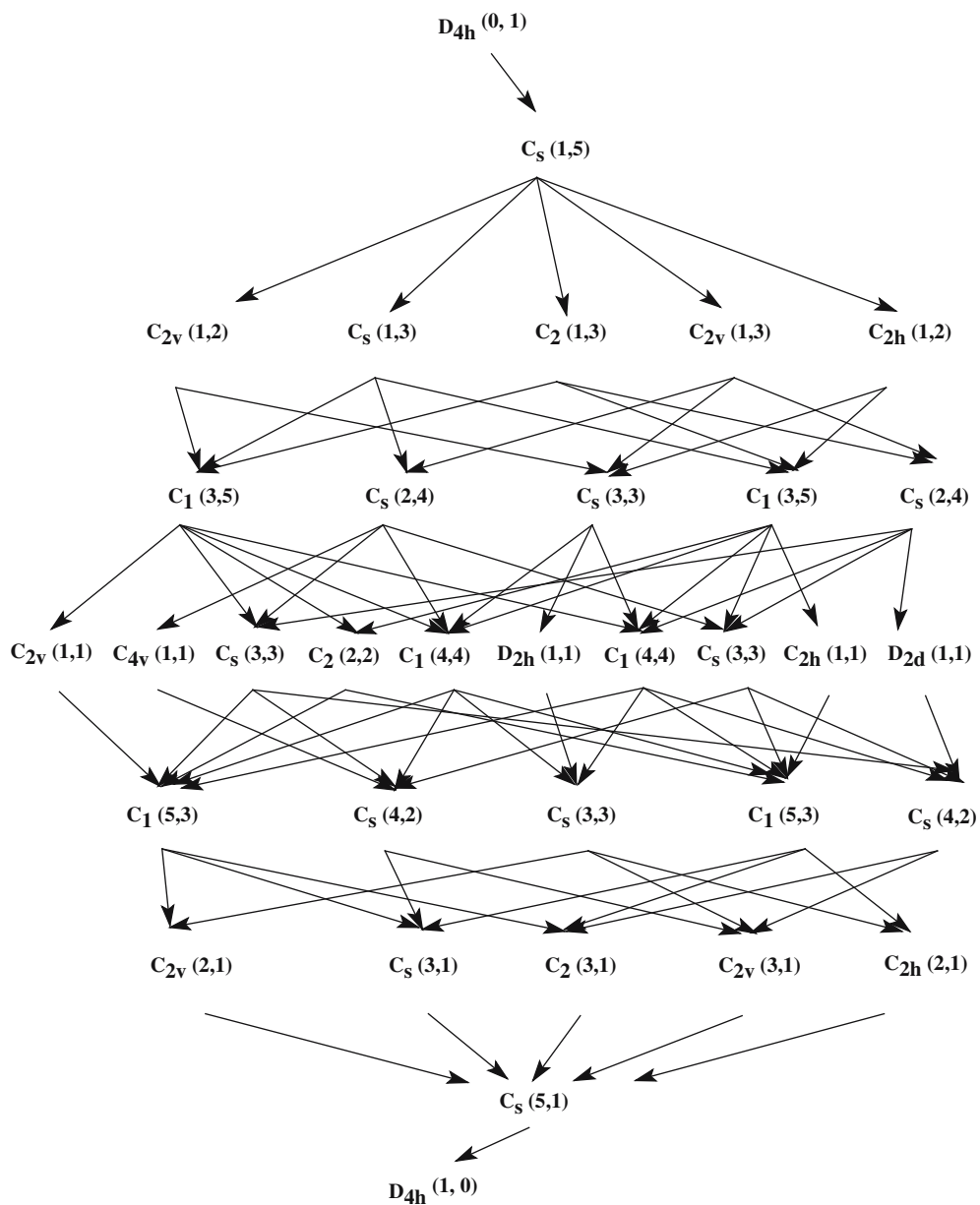


Figure 4. The symmetries for the diastereomer-based cyclo-butane substitution poset of figure 2. Further given in the present figure are the numbers of different equivalent sets of substituted and of unsubstituted sites in the isomer.

seminal paper [10] of 1964. There one imagines that a property X, ξ for a poset member ξ is expanded in terms of another related property $x(\zeta)$ for the preceding members ζ of the poset \mathbb{P}

$$X(\xi) = x(\xi) + \sum_{\zeta \triangleright \xi}^{\mathbb{P}} x(\zeta) \quad (5)$$

One may view the $x(\zeta)$ as fitting parameters, which in principle match in number the data to be fit, so that in practice one seeks circumstances where some further simplification occurs – such as the x -values diminishing as the argument increases, so that $x(\zeta)$ for suitably large ζ may be estimated as 0. A simple sort of example of such a cluster expansion is that for the substructure partial ordering where a molecular property X is expanded to leading order as a sum of contributions $x(\zeta)$ from smaller substructures ζ , say just of atoms and bonds (or bonded atom pairs), though one might sometimes also consider higher order contributions from pairs of adjacent bonds (or triples of atoms one of which is bonded to the other two). Sometimes contributions from even larger functional groups [such as a carboxylic acid group (COOH)] might be entertained. Indeed this substructural type of expansion is very common in chemistry, e.g., arguably underlying the importance of the central organizational idea of “functional groups” in organic chemistry, but also it is involved as a low-order quantitative approximation in various bond-energy schemes for computing heats of formation. The formalized idea in terms of all possible substructures has been enunciated by Smolenski [11], Gordon and Kennedy [12], and Essam *et al.* [13], and by others [14–17]. These formal chemical approaches have typically been made in terms of the (well-known) subgraph partial ordering, though at least formally similar cluster expansions may be made for rather arbitrary posets.

Presently such a general cluster-expansion approach is considered, making reference to an expanded poset \mathbb{P}_0 underlying the actual reaction poset \mathbb{P} as of figures 2–4. This expanded poset \mathbb{P}_0 simply entails the distinguishing of all substituent positions, e.g., so that for our present example there would be eight single-substituent configurations with the substituent appearing at any one of the eight positions, 1, 1', 2, 2', 3, 3', 4, 4'. Generally this expanded poset \mathbb{P}_0 is that formed with the trivial group $G = \{\mathbb{I}\}$, and the elements of \mathbb{P}_0 are simply single configurations (each simply corresponding to one of the $2^{|S|}$ subsets of S). Then the cluster expansion appears as

$$\begin{aligned} X(\emptyset) &= x(\emptyset), \\ X(i) &= x(\emptyset) + x(i), \quad i \in S, \\ X(i, j) &= x(\emptyset) + x(i) + x(j) + x(i, j), \quad i, j \in S \\ &\dots \end{aligned} \quad (6)$$

and generally for a configuration subset C of the full set S of sites

$$X(C) = \sum_{C' \subseteq C} x(C'). \quad (7)$$

But now this can be related more directly to the substitution-reaction poset \mathbb{P} , with collections of equivalent configurations mediated by the general group \mathbb{G} , by taking

$$X(PC) = X(C) \text{ and } x(PC) = x(C), \quad \forall P \in \mathbb{G} \quad (8)$$

For example, in the present cyclo-butane case, all site contributions are equal, so that $X(i) = X(1)$ and $x(i) = x(1), \forall i \in \{1, 2, 3, 4\}$, and there are just five distinct types of pair contributions, namely $x(1, 1'), x(1, 2), x(1, 2'), x(1, 3), x(1, 3')$. Then elimination of duplicate names for different x -contributions, gives our current general *cluster expansion*

$$X(\xi) = \sum_{\zeta \triangleright \xi} n(\zeta, \xi) \cdot x(\zeta) + x(\xi), \quad \forall \xi \in P \quad (9)$$

where $n(\zeta, \xi)$ is the number of ways in which the configurational arrangements of ζ occur as substructures in a configuration $C \in \xi$. That is, if ξ and ζ are identified in terms of subsets C and C' of S using the group \mathbb{G} , e.g., as $\xi = \mathbb{G}C \equiv \{\mathbb{G}C : P \in \mathbb{G}\}$, then

$$n(\mathbb{G}C', \mathbb{G}C) = |\{C'' \subseteq \mathbb{G}C' : C'' \in \mathbb{G}C\}|. \quad (10)$$

And for the present particular case the expansion appears as

$$\begin{aligned} X(\emptyset) &= x(\emptyset), \\ X(1) &= x(\emptyset) + x(1), \\ X(1, 1') &= x(\emptyset) + 2x(1) + x(1, 1'), \\ X(1, 2) &= x(\emptyset) + 2x(1) + x(1, 2), \\ X(1, 2') &= x(\emptyset) + 2x(1) + x(1, 2'), \\ X(1, 3) &= x(\emptyset) + 2x(1) + x(1, 3), \\ X(1, 3') &= x(\emptyset) + 2x(1) + x(1, 3'), \\ X(1, 1', 2) &= x(\emptyset) + 3x(1) + x(1, 1') + x(1, 2) + x(1, 2') + x(1, 1', 2), \\ X(1, 1', 3) &= x(\emptyset) + 3x(1) + x(1, 1') + x(1, 3) + x(1, 3') + x(1, 1', 3), \\ X(1, 2, 3) &= x(\emptyset) + 3x(1) + 2x(1, 2) + x(1, 3) + x(1, 2, 3), \\ X(1, 2, 3') &= x(\emptyset) + 3x(1) + x(1, 2) + x(1, 2') + x(1, 3') + x(1, 2, 3'), \\ X(1, 2', 3) &= x(\emptyset) + 3x(1) + 2x(1, 2') + x(1, 3) + x(1, 2', 3), \\ X(1, 1', 2, 2') &= x(\emptyset) + 4x(1) + 2x(1, 1') + 2x(1, 2) + 2x(1, 2') \\ &\quad + 4x(1, 1', 2) + x(1, 1', 2, 2') \end{aligned} \quad (11)$$

...

In fact, this is reminiscent of the traditional substructural cluster expansions, say of some sort of energy: $x(\emptyset)$ would be the energy of the basic (unsubstituted) reference structure; the $x(i)$ would be an energy for a single site; the $x(i, j)$ would be the additional energy of interaction between a pair of sites i and j ; the $x(i, j, k)$ would be higher-order 3-site interactions; *etc.* Frequently one might imagine that the (irreducible) interaction energies should diminish in size as the number of sites increases – perhaps to such an extent that higher-order interactions might be neglected.

This cluster expansion of (9) is not of the form of (5). But in fact so long as $n(\xi, \zeta)$ is some function for which $n(\xi, \xi) \neq 0, \forall \xi \in \mathbb{P}$, and $n(\xi, \zeta) = 0, \forall \xi \triangleright \zeta$, many of the general results arising [10] with (5) carry over [14,17] to (9), as well as yet further choices, and we still refer to these as cluster expansions.

Notably the cluster-expansion equations of (11) may be systematically inverted, step by step. For the present case, giving

$$\begin{aligned}
 x(\emptyset) &= X(\emptyset), \\
 x(1) &= X(1) - X(\emptyset), \\
 x(1, 1') &= X(1, 1') - 2X(1) + X(\emptyset), \\
 x(1, 2) &= X(1, 2) - 2X(1) + X(\emptyset), \\
 x(1, 2') &= X(1, 2') - 2X(1) + X(\emptyset), \\
 x(1, 3) &= X(1, 3) - 2X(1) + X(\emptyset), \\
 x(1, 3') &= X(1, 3') - 2X(1) + X(\emptyset), \\
 x(1, 1', 2) &= X(1, 1', 2) - X(1, 1') - X(1, 2) - X(1, 2') + 3X(1) - X(\emptyset), \\
 x(1, 1', 3) &= X(1, 1', 3) - X(1, 1') - X(1, 3) - X(1, 3') + 3X(1) - X(\emptyset), \\
 x(1, 2, 3) &= X(1, 2, 3) - 2X(1, 2) - X(1, 3) + 3X(1) - X(\emptyset), \\
 &\dots
 \end{aligned} \tag{12}$$

as completed (in general and in detail) in the section after the next. Such relations may then be used to determine the x -values from the earlier less-substituted X -values, so that with the neglect of higher-order x there result predictions for the more-substituted X values. Another approach is of course to least-squares fit a set of lower-order x to a wide range of X values (and in chemical applications involving bond energies, possibly even with the next corrections thereafter, this is what usually is done).

Again it is noted that the presently proposed cluster expansion differs modestly from that previously proto-typically described [11–17] in terms of (smaller) subgraphs of the molecular graph. In the present scheme, in place of attention to graphical molecular structures, attention is instead directed to the group-mediated posetic structure, and the underlying poset \mathbb{P}_0 of “configurations”. Moreover, the current approach has something in common with so-called “linear free-energy” methods, e.g., as promulgated by Hammett [18] and Taft [19], though there is much more recent work (e.g. [20]), usually in application to properties (logarithms of different rate or equilibrium constants) other than considered here. And though their scheme may be viewed to concern species occurring

in reaction networks, the focus of their work has tended to be on a given pattern of substitution under the variation of one of the substituents, rather on a fixed substituent under variation of the substitution pattern. A natural extension for future work would be to combine these two approaches – if the present scheme should prove successful in its currently considered applications.

4. Application to polymethyl-cyclobutane energies

The cluster expansion of the preceding section may be applied to ground-state methyl-substituted cyclo-butane molecular-mechanics energies $E_{MM}(\xi)$. These energies (via the MM+ force field from HyperChem 3 version) are reported in table 1, where we have referenced all the energies against that of unsubstituted cyclo-butane, so that $E_{MM}(\emptyset) = 0$. That is, the initial value of $E_{MM}(\emptyset)$ has been subtracted from all the initially computed energies. In addition to the computed molecular mechanics energies, table 1 gives cluster-expansion results as explained below, excepting the last column which is explained in the next section.

To apply the cluster expansion in a chemically plausible manner, we retain just the 1- and 2-site terms of the preceding section, now abbreviating them

$$\begin{aligned} \alpha &\equiv x(1), & \beta_0 &\equiv x(1, 1'), & \beta_1 &\equiv x(1, 2), & \beta'_1 &\equiv x(1, 2'), \\ \beta_2 &\equiv x(1, 3), & \beta'_2 &\equiv x(1, 3'). \end{aligned} \quad (13)$$

There are different ways in which these parameters may be fit. One is by inversion, using X -values (or here E_{MM} -values) up through two-site isomers, much as suggested in the preceding section. Then with the presumption that the higher-order x are 0, one obtains predicted X values for n -site isomers with $n > 2$. This might be imagined to be a delicate matter, as this is purely extrapolative. Alternatively one may simply take the expressions for all the different X in terms of the x , assuming that the higher-order x are 0, but then least-squares fit the parameters $\alpha, \beta_0, \beta_1, \beta'_1, \beta_2,$ and β'_2 for all the available X values.

We then make these two types of fits for the cyclo-butane energies of table 1. We find by inversion

$$\begin{aligned} \alpha &= -3592.98, & \beta_0 &= 1.84, & \beta_1 &= 1.23, & \beta'_1 &= 0.06, & \beta_2 &= 0.03, \\ \beta'_2 &= 0.02 \end{aligned} \quad (14)$$

in units of kcal/mole. The associated statistics (standard deviation s and correlation coefficient r) are indicated in Table 2 (under the identification of “mildly truncated inversion”). *Via* least-squares regression, we find

$$\begin{aligned} \alpha &= -3593.322, & \beta_0 &= 2.388, & \beta_1 &= 1.545, & \beta'_1 &= 0.360, & \beta_2 &= 0.078, \\ \beta'_2 &= 0.059 \end{aligned} \quad (15)$$

Table 1
 Computed (MM+ HyperChem molecular mechanics) and fitted energies (in kcal/mole) for methylated-cyclobutane derivatives (numbered as in Figure 2).

No.	E_{MM}	E_{fit}	$E_{\text{MM}} - n\alpha$	$E_{\text{fit}} - n\alpha$	e_{exact}
2	-3592.98	-3593.322	0.342	0	-3592.98
3	-7184.12	-7184.257	2.524	2.387	1.84
4	-7184.73	-7185.099	1.914	1.545	1.23
5	-7185.9	-7186.284	0.744	0.36	0.06
6	-7185.93	-7186.566	0.714	0.078	0.03
7	-7185.94	-7186.585	0.704	0.059	0.02
8	-10775.66	-10775.673	4.31	4.297	0.15
9	-10776.41	-10776.798	3.556	3.168	0.04
10	-10777.08	-10777.442	2.886	2.524	-0.03
11	-10777.62	-10778.002	2.346	1.964	0.01
12	-10778.8	-10779.169	1.166	0.797	-0.01
13	-14364.66	-14364.703	8.628	8.585	0.40
14	-14366.63	-14366.387	6.658	6.901	0.15
15	-14367.1	-14367.012	6.188	6.276	0.04
16	-14367.14	-14367.031	6.148	6.257	0.02
17	-14367.33	-14367.313	5.958	5.975	0.01
18	-14368.24	-14368.24	5.048	5.048	0.02
19	-14368.55	-14368.499	4.738	4.789	0.01
20	-14369.22	-14369.341	4.068	3.947	0.02
21	-14369.26	-14369.36	4.028	3.928	-0.00
22	-14371.63	-14371.731	1.658	1.557	0.03
23	-17956.03	-17955.983	10.58	10.627	0.02
24	-17957.27	-17957.107	9.34	9.503	-0.00
25	-17958.26	-17957.751	8.35	8.859	-0.01
26	-17958.72	-17958.311	7.89	8.299	-0.01
27	-17959.91	-17959.478	6.7	7.132	-0.03
28	-21544.79	-21544.875	15.142	15.057	0.05
29	-21545.83	-21545.717	14.102	14.215	0.05
30	-21547.31	-21546.903	12.622	13.029	0.02
31	-21547.94	-21547.185	11.992	12.747	0.01
32	-21548.16	-21547.204	11.772	12.728	0.02
33	-25134.21	-25134.25	19.044	19.004	0.00
34	-28719.77	-28721.237	26.806	25.339	-0.05

again in kcal/mole, with the associated statistics indicated in table 2 (under the identification of “mildly truncated least-squares). The results *via* either method are quite good, though the least-squares fitting procedure is seen from the standard deviation s to be preferred – for a more explicit comparison than the statistical data of table 2, we report the individual least-squares fitted energies in the third column of table 1. One sees that the 1-site α term is by far the largest, so that the interactions between different substituents (methyls) are relatively minor, though they are needed for a really good fit for s . Further, it is noteworthy that

Table 2
Statistics for energy fits

Method	r	s (kcal/mole)
cluster expansion		
(a) mildly truncated inversion	0.999 ⁺	1.023
(b) mildly truncated least squares	0.999 ⁺	0.495
(c) strongly truncated inversion	0.999 ⁺	2.278
(d) strongly truncated least squares	0.999 ⁺	2.227
poset average	0.999 ⁺	0.788
splinoid	0.999 ⁺	0.906

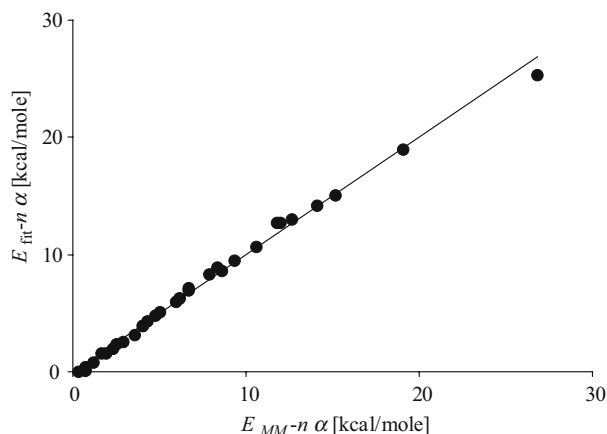


Figure 5. The fitted α -term shifter energies plotted versus the MM+ HyperChem molecular mechanics obtained energies, all in kcal/mole.

the pair terms (i.e., the β terms) decrease in size with the distance between the pair of sites. That the inversion and least-squares methods agree so closely while also the parameters have a chemically appealing behavior is an indication of the adequacy and chemical relevance of the cluster-expansion method for the present property. And to facilitate an even more detailed comparison the contributions from the (rather large) α -term are subtracted out in the fourth and fifth columns (with the value of α taken from the least-squares fit). It is seen at this level of detail, especially as displayed in figure 5, that the rank-2 β -terms are important, for s .

Further in table 2, statistics for two other poset-attentive schemes are reported for comparison. A “poset-average” method [2] just takes any X value to be the average of a mean preceding X -value and a mean succeeding X -value. That is, for an isomer ξ in the interior of the Hasse diagram one estimates $X(\xi)$ as

$$\tilde{X}(\xi) \equiv \{X(\text{pre-}\xi) + X(\text{post-}\xi)\}/2 \quad (16)$$

where $X(\text{pre-}\xi)$ is the mean of the X -values for isomers immediately preceding ξ in the Hasse diagram, and $X(\text{post-}\xi)$ is the mean for isomers immediately following ξ . Evidently this straight forward procedure is “parameter-free”, and is of utility when data is missing from an isolated interior point of the poset. Although not missing data in the present circumstance, one still can compare the predicted $\tilde{X}(\xi)$ values to the actual $X(\xi)$ values, to obtain standard deviation (s) and correlation coefficient (r), thereby providing an indicator of poset relevance. This requires immediately preceding and succeeding X -data, so that the statistics here are just for the (32) interior isomers.

Fourthly we also indicate in table 2 statistics for a posetic splinoid fitting method [4], which may be viewed as a generalized version of the poset-average procedure, fitting a smooth curve through the whole of the poset such as to intersect the subset of known values exactly. Again currently all energy values are known, so that an estimator of poset relevance is obtained when one leaves out X -values one at a time to make a prediction $\tilde{X}(\xi)$ for each of the 34 $X(\xi)$, then yielding s and r values. Details of this splinoid-fitting procedure are discussed elsewhere [4].

Yet further one may imagine schemes which attend less closely to the reaction-network poset. One such scheme would simply focus on the number of substituents – thereby attending only to rank in the poset, with further details of the poset “washed out”. This amounts to taking all β -terms=0 and retaining only α . This fitting is reported in table 2 as the “strongly truncated” cluster expansion, which again can be done either by inversion or by least-squares fitting α . From table 2 it is seen that though these two less poset-attentive schemes yield very good r -values, their s -values are notably worse, so that overall these relatively poset-inattentive schemes fare relatively poorly. This notable improvement for the posetic schemes validates attention to the poset and the utility of our associated schemes.

From amongst the poset-attentive schemes one sees that the present (6-parameter) poset-attentive cluster-expansion scheme is quite accurate, both *via* inversion and *via* least-squares fitting. In particular, both correlation coefficients r exceed 0.999, and further the predicted versus quantum-chemical data are compared in figure 5. The simple “poset-average” and splinoid methods have very good r -values (>0.999), and the standard deviations are fairly decent, comparing with that for the (mildly truncated) inversion-fitted cluster expansion. The best standard deviation and thence here preferred scheme is the here developed (mildly truncated) least-squares fitted cluster-expansion approach.

5. Complete cluster expansion

A further test of the cluster expansion may be made *via* a “complete” inversion where all the higher order terms are computed. That is the $x(\xi)$ for all ξ

are considered, and obtained *via* inversion as from equation (9), which can be rewritten in a matrix notation as

$$\vec{X} = \mathbf{n}\vec{x} \quad (17)$$

where \vec{X} and \vec{x} are vectors with the respective ξ^{th} components $X(\xi)$ and $x(\xi)$, while \mathbf{n} is the matrix with $(\xi, \zeta)^{\text{th}}$ element $n(\xi, \zeta)$. Thence inversion is represented by

$$\vec{x} = \mathbf{n}^{-1}\vec{X} \quad (18)$$

with the inverse matrix here being guaranteed to exist since \mathbf{n} has non-zero elements (=1) along the diagonal and otherwise is lower triangular (when one orders the rows and columns so as not to violate the partial order). For the present case of the cyclobutane reaction network this inverse appears as in table 3. The inversion process can be done by a numerical matrix-inversion routine or analytically as suggested in the earlier discussion leading to equation (12). In any case it may be observed that the elements of \mathbf{n}^{-1} are all integers and many elements are 0.

Indeed this inversion may be made fairly explicitly for the general case. One may start with the inversion of the underlying poset \mathbb{P}_0 which individually distinguishes all sites. This case appears of the form as in (5), and the consequent inverse $\mu = \mathbf{n}^{-1}$ has elements called the *Möbius* function $\mu(C', C)$ on this (underlying) poset \mathbb{P}_0 . Moreover, this special poset, being the family of subsets of the parent set S with the subset partial ordering, is well known, indeed long before Rota [10] introduced the general nomenclature and theory. E.g., see the textbook of Riordan [21] under the topic of “inclusion and exclusion”. This Möbius function for \mathbb{P}_0 may be given as

$$\mu(C, C') = \begin{cases} (-1)^{|C|-|C'|}, & C' \subseteq C, \\ 0, & \text{otherwise} \end{cases} \quad (19)$$

as is not too hard to prove, say inductively as C proceeds down the Hasse diagram from \emptyset . See e.g., [22,23]. Then realizing that the reaction-network poset \mathbb{P} with elements $\xi = GC$ and \mathbf{n} defined as in (10), one sees that the matrix elements of the inverse are

$$[\mathbf{n}^{-1}]_{\xi, \zeta} = \sum_{C \subseteq \xi} \mu(C, C'), \quad \zeta = GC' \quad (20)$$

from which it follows that

$$[\mathbf{n}^{-1}]_{\xi, \zeta} = \begin{cases} 1, & \xi = \zeta \\ (-1)^{\text{rank}(\xi) - \text{rank}(\zeta)} n(\xi, \zeta), & \zeta \triangleright \xi \\ 0, & \text{otherwise} \end{cases} \quad (21)$$

Table 3
Inverse matrix $\mu = n^{-1}$.

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34			
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	-1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	1	-2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	1	-2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5	1	-2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
6	1	-2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7	1	-2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
8	-1	3	-1	-1	-1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
9	-1	3	-2	0	-1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
10	-1	3	-1	0	0	-1	-1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
11	-1	3	0	-1	-1	0	-1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
12	-1	3	0	0	-2	-1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
13	1	-4	2	2	2	0	0	-4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
14	1	-4	0	4	0	2	0	0	-4	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
15	1	-4	1	2	2	1	0	-2	-1	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
16	1	-4	1	2	2	0	1	-2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
17	1	-4	1	2	1	1	1	-1	-1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
18	1	-4	2	0	0	2	2	0	0	-4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
19	1	-4	1	1	2	1	1	-1	0	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
20	1	-4	0	2	2	1	1	0	-1	0	-2	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
21	1	-4	0	2	2	0	2	0	0	0	-4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
22	1	-4	0	0	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
23	-1	5	-2	-3	-3	-1	-1	5	1	1	2	1	-1	0	-1	-1	0	-1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
24	-1	5	-1	-4	-2	-2	-1	2	4	1	2	1	0	-1	-1	0	-2	0	0	-1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
25	-1	5	-2	-2	-2	-2	-2	2	1	4	2	1	0	0	0	-2	-1	-2	0	-2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	
26	-1	5	-1	-3	-3	-1	-2	2	1	1	5	1	0	0	0	-1	-1	0	-1	-1	-1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
27	-1	5	-1	-2	-4	-2	-1	2	1	1	2	4	0	0	-1	0	0	-2	-1	0	-1	0	-1	0	0	0	0	1	0	0	0	0	0	0	0	0	
28	1	-6	3	4	4	2	2	-8	-2	-4	-4	-2	2	0	2	2	4	1	4	0	0	0	0	-4	0	-2	0	0	1	0	0	0	0	0	0	0	
29	1	-6	2	5	4	2	2	-6	-4	-2	-6	-2	1	1	2	2	4	0	2	2	1	0	-2	-2	0	-2	0	0	1	0	0	0	0	0	0	0	
30	1	-6	2	4	5	2	2	-6	-2	-2	-6	-4	1	0	2	2	2	4	2	1	1	0	-2	0	0	-2	0	0	1	0	0	0	0	0	0	0	
31	1	-6	2	4	4	3	2	-4	-4	-4	-4	0	1	2	0	4	1	4	2	0	1	0	0	1	0	-2	0	0	0	1	0	0	0	0	0	0	
32	1	-6	2	4	4	2	3	-4	-2	-4	-8	-2	0	0	2	4	1	4	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	-1	7	-3	-6	-6	-3	-3	10	5	5	10	5	-2	-1	-4	-4	-8	-1	-8	-4	-2	-1	6	3	3	6	3	-1	-2	-1	-1	-1	-1	-1	1	0	
34	1	-8	4	8	8	4	4	-16	-8	-8	-16	-8	4	2	8	8	16	2	16	8	4	2	-16	-8	-8	-16	-8	4	8	4	8	4	4	4	-8	1	

The integerness and frequent occurrence of 0-values is now quite clearly manifested. This then joins the list [13,23,24] of similarly explicit inversions, usually for Möbius functions (as associated with the inversion of equations as (5)).

Granted this invertability the $x(\xi)$ are seen to be just alternative transformed forms in which to present the data $X(\xi)$ - but a utility of the inversion is to check to see whether in complete detail our presumption of the diminishing higher order contributions is fully justified. In the last column of table 1 the resultant exact cluster energies e_ξ (i.e., the x_ξ when $X = E_{MM}$) are given. It is seen that there is a degree of agreement with the fitted values of equation (15), so that the validity of the whole approach is further supported. Moreover, this analysis allows the identification of structure 13 (of figure 2 and also denoted 1, 1', 2, 2' - □) as entailing an anomalous correction - as well as structures 8 and 14 to a somewhat lesser degree. Notably this is considerably different than the 7- and 8-substituted structures (34, 31, and 32), which the least-squares fitting singles out as the most problematic. Many chemists might imagine that the fully substituted species 34 as well as other highly substituted structures might be exceptional in some sense. But this expectation might be mixed some with the expectation (and the actuality, in the present case) that these highly substituted species are at an extreme (i.e., a maximum) for properties amongst the considered structures. The exact cluster expansion factors out this aspect to make a refined assessment, to say that whatever behavior these highly substituted structures have is in fact largely accounted for in contributions from earlier less substituted structures.

Further as a side note it may be seen that for the present type of reaction posets, the cluster expansion may be viewed to have a form like a so-called "lattice-gas" (or 'Ising-type) model of statistical mechanics. That is, for a configuration $C \subseteq S$ one may identify C through the use of 0,1-valued "indicator" variables n_i such that $n_i = 1$ if $i \in C$ and $n_i = 0$ if $i \notin C$, whence $X(C)$ may also be written as

$$X(C) \equiv X(n_1, n_2, \dots, n_{|C|}) = \sum_{C'}^{\subseteq S} x(C') \prod_i^{\in C'} n_i. \quad (22)$$

For the energy example then $X(n_1, n_2, \dots, n_{|C|})$ would be termed the energy function, and a corresponding Hamiltonian operator can be presented in terms of operators $\hat{C} \equiv \prod_{k \in C} \hat{n}_k$ where \hat{n}_k is a "number" operator acting on a basis of vectors $|n_1, \dots, n_{|S|}\rangle$ as

$$\hat{n}_k |n_1, \dots, n_{|S|}\rangle \equiv \delta(n_k, 1) |n_1, \dots, n_{|S|}\rangle. \quad (23)$$

Then the Hamiltonian operator is

$$\hat{X} \equiv \sum_{\xi} x(\xi) \sum_{C \in \xi} \hat{C}. \quad (24)$$

The $x(\xi)$ appear as irreducible interactions, and the identity amongst the different $x(C)$ corresponding to the same isomer $\xi = GC$ is manifested in \widehat{X} commuting with the elements of the group G . Sometimes such models are recast in terms of “spin” operators $\hat{\sigma}_k \equiv 2\hat{n}_k - 1$. Further then in statistical mechanics, generating functions for such X would for the overall “ensemble” provide the partition function, which in turn leads to various thermodynamic properties. But here a more detailed analysis isomer by isomer is desired, framed in terms of a general property X (further examples of which are next addressed).

6. Application to further properties

Application may be sought to a more “typical” circumstance involving other experimentally measured properties, of which here we consider: boiling point (at 1.0 atm), standard heat of vaporization, octanol/water partition coefficient (under standard conditions), and logarithm of the vapor pressure (at 25°C). There are two aspects of greater challenge now: first, the available data sets for these properties are incomplete, in having property values known for only $k < |S|$ structures of the poset; and second, it might be imagined that for some of these properties there is less *a priori* rationale to expect cluster-expansion additivity to so rapidly converge (as found for the energies). Both of these distinctions indicate that to implement the cluster expansion the least-squares fitting approach is recommended (over the inversion approach). Thence using data [25] for $k=17$ of the cyclobutane species, the (6-parameter) cluster expansion was applied. In table 4 we give these cluster-expansion fits for these four properties, whence it is seen that the statistical fits all appear quite reasonable, each with $r > 0.99$. The octanol/water partition coefficient is fit very well, with $r = 0.999$, and further with the β -parameters all essentially =0, indicating an especially simple dependence of this parameter. For the boiling point, heat of vaporization, and logarithm of vapor pressure the β parameters do not so significantly dominate (in comparison with the energies of the preceding section) nor do they so rapidly decrease with distance between substituent methyls. In these cases the interaction parameters (i.e., the β parameters) have a somewhat different interpretation than for the energy. Particularly for the boiling point and vapor pressure, interactions between distinct (but isomorphic) molecules should be important in effecting condensation, so that a pair of methyls on separate molecules can be in close contact, even though the pair correspond to a distant pair on the same molecule. In such a case the β parameters might simulate the effects of this intermolecular interaction (whence there is no necessary special great proclivity for them to be larger when the methyls are adjacent). That is, it is quite reasonable that the sizes of the β parameters indicate in part which methyls are more likely to come in contact in intermolecular interactions between pairs of molecules. For the partition coefficients, pairs of the considered

Table 4
Cluster expansion fits to various properties.

No.	Property	Statistics parameters		Parameters for cluster expansion method						
		r	s	$\alpha \equiv x(1)$	$\beta_0 \equiv x(1, 1')$	$\beta_1 \equiv x(1, 2)$	$\beta'_1 \equiv x(1, 2')$	$\beta_2 \equiv x(1, 3')$	$\beta'_2 \equiv x(1, 3')$	
1	bp (°C)	0.991	1.724	25.020	-5.593	-0.843	-1.330	-2.136	-0.369	
2	ΔH_{vap} (kJ/mol)	0.999	0.166	2.451	-0.559	-0.132	-0.154	-0.182	-0.101	
3	log P	0.999	0.0001	0.4920	0.049	0.000	0.000	0.000	0.000	
4	$-\log v_p$ (torr)	0.989	0.034	0.396	-0.086	-0.002	-0.011	-0.027	0.006	

molecule do not come in contact, but the β -terms of the cluster expansion have to do with which methyl/solvent interactions are modified more by the change of solvent (from water to octanol) – and here each polymethyl-cyclobutane molecule interacts with several water molecules so that again there is no expectation that the interaction be significantly more important for two methyl groups in close proximity. Still in using the cluster expansion we are at least hoping that there is a degree of “additivity” manifested in the earlier cluster expansion terms being more dominant, though the second rank (i.e., β terms) need not be ordered as to separation between the pairs of methyls. In any event for all the presently considered properties quite reasonable cluster-expansion fits are seen from table 4 to be obtained.

Similarly with such incomplete data, the splinoid fitting is recommended over the simple poset-averaging method. The splinoid-fitting scheme fits the given data exactly and predicts the unknown data. An estimate of the accuracy of the method may be obtained using just the k (here =17) known values, by successively leaving out each one of these values to make $k = 17$ further fits each of $k=16$ known points, whence the 17th known but deleted value is predicted and may be compared to the known experimental value. The standard deviation for these k predictions then is surmised to be an estimate of the standard deviation for the remaining experimentally unknown but predicted values. Thence the statistics for this approach for each of the four considered properties is given in table 5, where also the known data and predictions are given. The splinoid fit entails a collection of cubic polynomials one on each link of the Hasse diagram, with the cubic polynomials joined at the nodes of the Hasse diagram so as to satisfy appropriate equality and smoothness conditions, so that perhaps the spline is most concisely specified by simply giving the known and fitted property values. The predicted values appear in boldface in table 5 while the known points are in ordinary (non-boldface) type. Again for each of the four properties the correlation coefficients are $r > 0.99$. This is rather similar to the cluster-expansion fit, for which however the standard deviations appear slightly smaller. The situation with these properties seems to parallel the comparisons obtained with the molecular-mechanics energies.

Further comparisons of the cluster-expansion and splinoid fits for the boiling point and water/octanol partition coefficient are made in figures 6 and 7. The corresponding plots for the standard heat of vaporization and the room temperature vapor pressures have a similar general appearance (though we have not displayed them here). Overall both these schemes seem to work quite well.

7. General Discussion and Prognosis

In conclusion, quantitative predictions have been entertained to predict properties for species falling within a progressive reaction poset. As a prelim-

Table 5

Compound name, experimental and predicted values of boiling point (bp, (°C)), standard enthalpy of vaporization (ΔH_{vap} , (kJ/mole)), octanol/water partition coefficient (log P) and vapor pressure at 25°C ($-\log v_p$, (torr)) for methyl-cyclobutane derivatives with the splinoid method.

No.	Compound name	CAS register number	Property				
			bp(°C)	ΔH_{vap} (kJ/mol)	log P	$-\log v_p$ (torr)	
1	□	287–23–0	12.5	24.19	2.257	3.072	
2	1 □	598–61–8	36.3	26.96	2.749	2.704	
3	1,1'□	18931–71–0	56.0	28.65	3.290	2.387	
4	1,2 □		59.6	28.99	3.249	2.326	
5	1,2'□	345202–05–3	60.0	29.00	3.241	2.320	
6	1,3 □	2398–09–6	57.5	28.78	3.241	2.362	
7	1,3'□	2398–10–9	60.5	29.04	3.241	2.314	
8	1,1',2 □		81.6	30.90	3.765	1.953	
9	1,2,3 □		83.5	31.06	3.742	1.921	
10	1,1',3 □	75017–20–8	81.7	30.90	3.782	1.954	
11	1,2,3'□		83.7	31.08	3.744	1.918	
12	1,2',3 □		83.5	31.06	3.740	1.920	
13	1,1',2,2'□	52897–98–0	97.1	32.27	4.323	1.685	
14	1,2,3,4 □	73493–50–2	107.1	33.16	4.225	1.507	
15	1,2,2',3 □	2893–95–0	102.1	32.71	4.274	1.597	
16	1,2,2',3'□		102.2	32.73	4.270	1.593	
17	1,1',2,3 □		102.3	32.73	4.272	1.592	
18	1,1',3,3'□	24642–79–3	97.1	32.27	4.323	1.685	
19	1,1',2',3 □		102.3	32.73	4.272	1.592	
20	1,2',3,4 □	73493–48–8	107.1	33.16	4.225	1.507	
21	1,2',3',4 □	73493–47–7	107.1	33.16	4.225	1.507	
22	1,2',3,4'□	73493–49–9	107.1	33.16	4.225	1.507	
23	1,1',2,2',3 □	75017–21–9	116.7	34.03	4.815	1.330	
24	1,1',2,3,4 □		119.9	34.31	4.789	1.275	
25	1,1',2,3,3'□	57905–86–9	116.7	34.03	4.815	1.330	
26	1,1',2',3,4 □		119.5	34.28	4.792	1.282	
27	1,1',2,3',4 □		119.93	34.31	4.789	1.275	
28	1,1',2,2',3,3'□		132.5	35.47	5.342	1.036	
29	1,1',2,2',3,4 □		133.1	35.52	5.337	1.028	
30	1,1',2,2',3',4 □		133.1	35.52	5.337	1.028	
31	1,1',2,3,3',4 □		133.2	35.53	5.336	1.025	
32	1,1',2',3,3',4 □		133.0	35.51	5.338	1.029	
33	1,1',2,2',3,3',4 □		147.7	36.86	5.869	0.749	
34	1,1',2,2',3,3',4,4'□	18998–25–9	162.3	38.25	6.389	0.453	
			0.997	0.995	0.999	0.997	<i>r</i>
			3.005	0.334	0.029	0.053	<i>s</i>

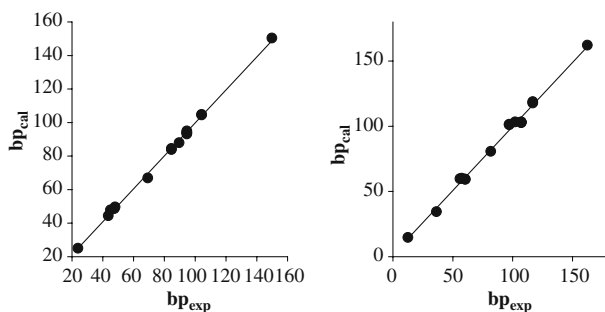


Figure 6. Fitted versus experimental boiling points, first with the fitting *via* the cluster expansion method, and second *via* the splinoid method.

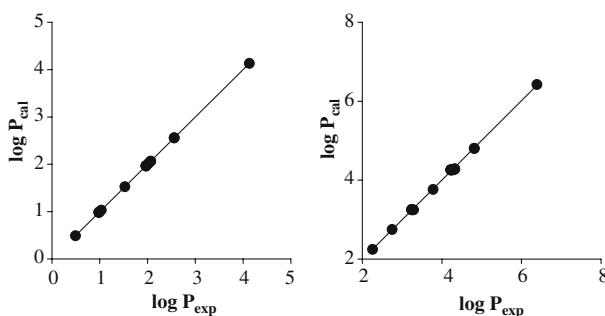


Figure 7. Fitted versus experimental water/octanol partition coefficients, first with the fitting *via* the cluster-expansion method, and second *via* the splinoid method.

inary consideration toward characterizing the general circumstance, some general “combinatorial” aspects of these posets have been noted. Further a chemically appealing posetic cluster-expansion method for various properties has been described, in general in equation (5), in terms of lower-order contributions. This method has been illustrated in different modes of application, and further it is compared with other poset-based methods. The work here indicates that quite favorably accurate property predictions (at least for several suitable properties) may be made by different methods attentive to the posetic structure, which though long known in chemistry seems to have been relatively little used previously. Evidently now one has a choice of two fairly general poset-attentive methods which yet are not overly demanding of known data. These are as follows:

- the currently developed cluster expansion; and
- an elsewhere-developed [4] posetic splinoid method.

So far each method appears comparably accurate – indeed at least for the studied properties highly accurate, with $r > 0.99$ – and though comparable the

present cluster-expansion scheme seems to be somewhat better. In more special circumstances (with a greater abundance of known data from which to interpolate), a simpler “poset-average” method is also applicable, as earlier studied [2]. The 34-member cyclo-butane-substitution poset here is of modestly greater size than the 13-member benzene-substitution poset previously studied with regard to the more special “poset averaging” method. Presumably the larger the poset, the greater likelihood of missing data being so frequent as to cause problems for the simple poset averaging – whence the cluster-expansion and splinoid methods are more widely recommended. The present cluster expansion differs from the other entertained posetic methods (average-poset and splinoid fit) in providing a “model” in the more common chemical-statistical sense – the model being summarily specified as in equations (14) or (15). Further this model often has a traditional chemical interpretation in terms contributions first from the basic substituents and second from the interactions amongst the substituents. The splinoid (and sometimes the poset average method) still can be used to make predictions, when so desired. The expectation that the cluster-expansion and splinoid fits are applicable even with many missing data points in the poset is supported with the results here, for the boiling point, heat of vaporization, water/octanol partition coefficient, and vapor pressure.

We see no reason why considerably larger posets might not be reasonably handled *via* these posetic methods. We propose that future work be undertaken to consider such larger posets, and a variety of further properties, including biological activities of various sorts. As such in going beyond molecular structure to fit activities in concert with the overall reaction network poset this might be termed a *quantitative-superstructure/activity relationship* or *QSSAR* scheme. Overall notable success is evidenced in the current initial investigation thereby indicating a potentially wide-range of possible chemical applications.

Acknowledgment

The authors acknowledge the support (through grant BD-0894) from the Welch Foundation of Houston, Texas.

References

- [1] D.J. Klein and L. Bytautas, *Commun. Math. Comput. Chem. (MATCH)* 42 (2000) 261.
- [2] T. Ivanciuc and D.J. Klein, *J. Chem. Inf. Comp. Sci.* 44 (2004) 610.
- [3] W. Körner, *Gazz. Chim. Ital.* 4 (1874) 305, as summarized by H. E. Armstrong, *J. Chem. Soc.* 29 (1876) 204.
- [4] T. Doslić and D. J. Klein, *J. Comp. Appl. Math.* 177 (2005) 175.
- [5] E. Ruch and I. Ugi, *Topics Stereochem.* 4 (1969) 99.
- [6] I. Ugi, D. Marquarding, H. Kusacek, G. Gokel and P. Gillespie, *Angew. Chem. Intl. Edn.* 9 (1970) 703.

- [7] H. Kopp, Poggendorff's Ann. Phys. 96 (1855) 153 and 303.
- [8] J.D. Cox and G. Pilcher, *Thermochemistry of Organic and Organometallic Compounds* (Academic Press, NY, 1970).
- [9] W.J. Lyman, W.F. Reehl, and D.H. Rosenblatt eds, *Handbook of Chemical Property Estimation Methods*, (McGraw-Hill, New York, 1982).
- [10] G.C. Rota, Zeit. Wahrscheinlichkeitstheorie 2 (1964) 340.
- [11] E.A. Smolenski, Russ. J. Phys. Chem. [Engl. trans.] 38 (1964) 700.
- [12] M. Gordon and J.W. Kennedy, J. Chem. Soc., Faraday Trans. II 69 (1973) 484.
- [13] J.W. Essam, J.W. Kennedy, M. Gordon and P. Whittle, J. Chem. Soc. Faraday II 73 (1977) 1289.
- [14] D.J. Klein, Intl. J. Quantum Chem. 20 S (1986) 153.
- [15] M.I. Skvortsova, I.I. Baskin, O.L. Slovokhotova and N. S. Zefirov, Dok Chem. Engl. trans. 336 (1994) 496.
- [16] I.I. Baskin, M.I. Skvortsova, I.S. Stankevich and N.S. Zefirov, Dok Chem.(Engl. trans.) 359 (1994) 346.
- [17] D.J. Klein, T.G. Schmalz and L. Bytautas, SAR and QSAR Environ. Res. 10 (1999) 131.
- [18] L.P. Hammett, J. Am. Chem. Soc. 59 (1937) 96.
- [19] R.W. Taft and I.C. Lewis, J. Am. Chem. Soc. 80 (1958) 2436.
- [20] C. Hansch, A. Leo and R. W. Taft, Chem. Rev. 91 (1991) 165.
- [21] J. Riordan, *An Introduction to Combinatorial Analysis* (Wiley, NY, 1958).
- [22] C. Berge, *Principles of Combinatorics* (Academic Press, NY, 1971).
- [23] C. Domb, Graph Theory and Embeddings, in: of Phase Transitions and Critical Phenomena eds. C. Domb and M.S. Green vol.3 (Academic Press, NY, 1974) pp. 1–90.
- [24] J. W. Essam, Disc. Math. 1 (1971) 83.
- [25] From a CAS web-site (<http://cas.org./Scifinder/Scholar>) with the CAS registry numbers given in table 5.