# The Combined Fragmentation and Systematic Molecular Fragmentation Methods

Michael A. Collins,*,† Milan W. Cvitkovic,‡ and Ryan P. A. Bettens*,‡

†Research School of Chemistry, Australian National University, Canberra ACT 0200, Australia
‡Department of Chemistry, National University of Singapore, 3 Science Drive 3, Singapore 117543

S Supporting Information

**CONSPECTUS:** Chemistry, particularly organic chemistry, is mostly concerned with functional groups: amines, amides, alcohols, ketones, and so forth. This is because the reactivity of molecules can be categorized in terms of the reactions of these functional groups, and by the influence of other adjacent groups in the molecule. These simple truths ought to be reflected in the electronic structure and electronic energy of molecules, as reactivity is determined by electronic structure. However, sophisticated ab initio quantum calculations of the molecular electronic energy usually do not make these truths apparent. In recent years, several computational chemistry groups have discovered methods for estimating the electronic energy as a sum of the energies of small molecular fragments, or small sets of groups. By decomposing molecules into such fragments of adjacent functional groups, researchers can estimate the electronic energy to chemical accuracy; not just qualitative trends, but accurate enough to understand reactivity. In addition, this has the benefit of cutting down on both computational time and cost, as the necessary calculation time increases rapidly with an increasing number of electrons. Even with steady advances in computer technology, progress in the study of large molecules is slow.

In this Account, we describe two related "fragmentation" methods for treating molecules, the combined fragmentation method (CFM) and systematic molecular fragmentation (SMF). In addition, we show how we can use the SMF approach to estimate the energy and properties of nonconducting crystals, by fragmenting the periodic crystal structure into relatively small pieces. A large part of this Account is devoted to simple overviews of how the methods work.

We also discuss the application of these approaches to calculating reactivity and other useful properties, such as the NMR and vibrational spectra of molecules and crystals. These applications rely on the ability of these fragmentation methods to accurately estimate derivatives of the molecular and crystal energies. Finally, to provide some common applications of CFM and SMF, we present some specific examples of energy calculations for moderately large molecules. For computational chemists, this fragmentation approach represents an important practical advance. It reduces the computer time required to estimate the energies of molecules so dramatically, that accurate calculations of the energies and reactivity of very large organic and biological molecules become feasible.

## 1. INTRODUCTION

In principle, predicting the physical properties of a molecule is as simple as solving the quantum equations of motion of its electrons and nuclei. In practice, solving these equations for large molecular systems is a major challenge. The central difficulty is that the calculation time required increases rapidly as the number of electrons increases. The computation time of a relatively simple Hartree−Fock calculation scales at a staggering $O(N_{basis}^4)$, where $N_{basis}$ is the number of basis functions used to describe the electronic wave function. Advanced, correlated, methods scale faster still, for example, $O(N_{basis}^5)$ and $O(N_{basis}^7)$. Even with steady advances in computer technology, progress in the study of large molecules is slow.

An emerging and effective strategy to make calculations tractable for large systems is the fragmentation approach. Fragment-based methods work by dividing a chemical system into smaller subsets ("fragments"), performing quantum chemistry calculations on each fragment, and combining the fragment properties into a property for the whole. This distinguishes fragmentation, which treats the entire system quantum-mechanically, from the multiscale approach, which treats part of the system quantum-mechanically and part with semiempirical methods or molecular mechanics. Provided it generates fragments of a limited size, a fragmentation method has a computational complexity which is only *linearly* proportional to the number of atoms in the system. Fragmentation methods are also inherently parallelizable, since computations can be performed on different fragments independently.

Several fragmentation methods have been developed over the last two decades.[1−10] These methods differ in how they generate fragments, perform calculations, and recombine the results. We will not attempt to review all these approaches in this brief Account, as many are described in the accompanying Accounts.

This Account focuses on the systematic molecular fragmentation (SMF) method of Collins and co-workers and the combined fragmentation method (CFM) of Bettens and co-workers. Both methods, while distinct, employ the same underlying approach: gather bonded atoms into disjoint groups, form overlapping fragments from these groups, perform calculations on the fragments, and combine the fragment results using an inclusion/exclusion-principle-based expression. In the following sections, we present the SMF and CFM procedures in sufficient detail to understand and compare their implementations. We then indicate the utility of these methods for the ab initio calculation of energies, reactions, and properties of large molecules and crystals, and present examples of their performance.

## 2. BONDS AND GROUPS

Fragmentation works best when atoms in the same functional group are kept together in all computations. To ensure this, both SMF and CFM divide the atoms of the target molecule into groups, which are later used to define fragments. Groups are based on the bonded connectivity of the molecule.

### 2.1. Grouping Atoms in SMF

**Defining Bonds.** The following rules are used to define bonds that correspond to a chemist's normal view. (i) Two atoms share a *single* bond if the sum of their covalent radii + $b_{tol}$ exceeds the distance between them (we set $b_{tol} = 0.4$ Å). (ii) A single bond is replaced by a *multiple* bond if the bond length is much shorter ($b_{tol} = 0.07$ Å), unless either atom has its normal valence (e.g., an oxygen has two single bonds). (iii) Hydrogen bonds can be defined as a separate type of single bond if the distance between a hydrogen single bonded to O or N is within 2.4 Å of another O or N atom.

Note that the CC bonds in an *aromatic* polyene are sufficiently short to be defined as multiple bonds.

In addition, to follow the energy along a reaction path on which bonds are formed or broken, one would define these bonds to exist, irrespective of the atom−atom distance. This ensures continuity of the fragmentation along the reaction path.

**Forming Groups.** Each atom is considered to be its own group, except that: (i) Two atoms connected by a multiple bond share the same group. (ii) Hydrogen atoms share the same group with the "heavy" atom to which they are bonded. (iii) Where formal charges are present, if atom A is connected by a single bond to a formally charged atom B (or to an atom multiple-bonded to B), then atom A is in the same group as B. For example, $-(CH_2)-COO^-$ is a single group (including the $CH_2$ ensures the charge is localized within one group).

### 2.2. Grouping Atoms in CFM

**Defining Bonds.** Bonding in CFM is based on the Lewis structure of the molecule, which can be generated by any means (e.g., user-specified, using criteria (i) and (ii) of SMF, etc.)

**Forming Groups.** Groups of bonded atoms are formed by the rules given in Table 1. These rules do not necessarily define a unique set of groups, so to ensure maximum computational efficiency CFM uses an algorithm to select the grouping that minimizes the number of atoms in each group.[11]

**Table 1. Hard (H) and Soft (S) Rules for Forming Sets of Atoms to Define Groups in CFM**

| no. | rule |
|---|---|
| H1 | Singly valent (e.g., hydrogen and halogen) atoms are placed in the same group as the atom to which they are bonded. |
| H2 | Atoms bonded to each other by two or more bonds are placed in the same group. |
| H3 | Any atom in group, $i$, may not be bonded to more than one other atom outside of group $i$. |
| S1 | Atoms contained in an aromatic system are placed in the same group. |
| S2 | −OH and −SH atoms are placed in the same group as the atom to which they are bonded. |

Hard rules must be obeyed when fragmenting valence bonded systems. Soft rules need not be strictly followed, but are recommended to ensure consistency in the number of atoms in each group and to avoid "capping hydrogens" being too close to one another (vide infra). Monomers in a cluster obviously reside in different groups. Generally, CFM groups are larger than groups in SMF. For example, in an alkyl chain, each group contains two $CH_2$ units.

## 3. FRAGMENTATION

So, a molecule is a collection of groups connected by single bonds. Fragments are formed as combinations of these groups. The following sections explain how fragments are formed, calculations performed, and the results combined, in SMF and CFM.

### 3.1. Systematic Molecular Fragmentation (SMF)

The original SMF algorithm[10,12] has been replaced by a closely related approach called systematic molecular fragmentation by annihilation (SMFA), which is more easily implemented for very large molecules.[13] In particular, SMFA correctly treats the energy of cyclic structures without recourse to the post-fragmentation "ring repair" procedure of SMF.

**The SMFA Algorithm.** We consider a simple example, but the algorithm applies generally. Consider a chainlike molecule, $M$, that has $N$ groups, connected by single bonds:

$$M = G_1 G_2 ... G_N \tag{3.1.1}$$

The total electronic energy is written as $E(M)$:

$$E(M) = E(G_1 G_2 ... G_N) \tag{3.1.2}$$

We fragment the molecule (*keeping the geometry within each fragment fixed*) as follows.

(i) Remove some arbitrary group $G_k$:

$$M - G_k \rightarrow G_1 G_2 ... G_{k-1} + G_{k+1} ... G_N \tag{3.1.3}$$

The energy of the molecule is related to the energies of the fragments:

$$E(M) = E(G_1 G_2 ... G_{k-1}) + E(G_{k+1} ... G_N) + dE_1 \tag{3.1.4}$$

where $dE_1$ denotes an energy correction caused by the annihilation of $G_k$.

(ii) Remove from $M$ all groups that are separated from $G_k$ by *at least* Level groups in the bonded sequence (this defines "Level"):

$$M - G_1 G_2 ... G_{k-Level-1} - G_{k+Level+1} .... G_N$$

$$\rightarrow G_{k-Level} ... G_{k-1} G_k G_{k+1} ... G_{k+Level} \tag{3.1.5}$$

and

$$E(M) = E(G_{k-\text{Level}}...G_{k-1}G_kG_{k+1}...G_{k+\text{Level}}) + dE_2 \tag{3.1.6}$$

(iii) Remove from $M$ both $G_k$ and all the groups removed in (ii):

$$M - G_k - G_1G_2...G_{k-\text{Level}-1} - G_{k+\text{Level}+1}....G_N$$
$$\rightarrow G_{k-\text{Level}}...G_{k-1} + G_{k+1}...G_{k+\text{Level}} \tag{3.1.7}$$

and

$$E(M) = E(G_{k-\text{Level}}...G_{k-1}) + E(G_{k+1}...G_{k+\text{Level}}) + dE_3 \tag{3.1.8}$$

We now assume that if Level is sufficiently large, the energy correction due to removal of all groups separated from $G_k$ by at least Level groups is the same whether $G_k$ is removed at the same time. That is, $dE_3 \approx dE_1 + dE_2$.

Hence, if we add eqs 3.1.4 and 3.1.6 and subtract eq 3.1.8:

$$E(M) \approx E(G_1G_2....G_{k-1}) + E(G_{k+1}...G_N)$$
$$+ E(G_{k-\text{Level}}...G_{k-1}G_kG_{k+1}...G_{k+\text{Level}})$$
$$- E(G_{k-\text{Level}}...G_{k-1}) - E(G_{k+1}...G_{k+\text{Level}}) \tag{3.1.9}$$

corresponding to

$$M \rightarrow G_1G_2....G_{k-1} + G_{k+1}....G_N + G_{k-\text{Level}}...G_{k-1}G_kG_{k+1}...$$
$$G_{k+\text{Level}} - G_{k-\text{Level}}...G_{k-1} - G_{k+1}...G_{k+\text{Level}} \tag{3.1.10}$$

The fragments with negative coefficients correspond to overlaps between the fragments with positive coefficients. If there are no groups in $M$ that are separated from any other group by Level groups, then fragmentation at this "Level" is not possible.

The algorithm now takes each of the fragments on the rhs (right-hand side) of eq 3.1.10 and applies the SMFA procedure, (i) to (iii), again and again, until all fragments cannot be further fragmented. For a chain of five groups, $G_1G_2G_3G_4G_5$, the molecule is decomposed as follows:

$$G_1G_2G_3G_4G_5 \rightarrow G_1G_2 + G_2G_3 + G_3G_4 + G_4G_5 - G_2$$
$$- G_3 - G_4 \text{ (Level 1)}$$
$$\rightarrow G_1G_2G_3 + G_2G_3G_4 + G_3G_4G_5 - G_2G_3$$
$$- G_3G_4 \text{ (Level 2)}$$
$$\rightarrow G_1G_2G_3G_4 + G_2G_3G_4G_5 - G_2G_3G_4$$
$$\text{(Level 3)} \tag{3.1.11}$$

and so on for higher Levels (for larger molecules). At Level 1, the interaction of each group with its $\alpha$ substituents is included in the fragments. At Level 2, $\beta$ substituents are included, and so on.

Hydrogen atoms are appended to restore the valence of all functional groups, where single bonds have been broken in forming the final fragments. Explicitly, the Cartesian coordinates of a hydrogen "cap" in fragment $n$ is given by $\mathbf{X}(H)$:[10,14]

$$\mathbf{X}(H) = \mathbf{X}(j) + \frac{r(j) + r(H)}{r(j) + r(m)}[\mathbf{X}(m) - \mathbf{X}(j)] \tag{3.1.12}$$

where $r$ denotes a standard covalent radius for the element, $j$ is in fragment $n$, and the $j\cdots m$ bond was broken in forming the fragment. Equation 3.1.12 produces sensible bond lengths for the capping $j\cdots H$ bond.

Finally, the molecule is represented by a sum of fragments, $F_n$:

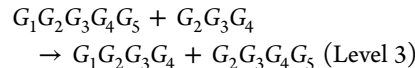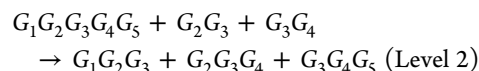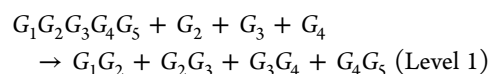$$M \rightarrow \sum_{n=1}^{N_{\text{frag}}} f_n F_n \tag{3.1.13}$$

and the associated "bonded" energy is given by

$$E_b = \sum_{n=1}^{N_{\text{frag}}} f_n E(F_n) \tag{3.1.14}$$

The coefficients, $f_n$, are integers. The SMFA procedure is fully automated for any molecular structure.[10,12,13] The number of fragments, $N_{\text{frag}}$, is proportional to $N$, so the ab initio CPU time for the rhs of eq 3.1.14 is proportional to $N$ (not $N^4$, $N^5$, or $N^7$).

As the value of Level increases, more account of the bonding environment is included in the fragments, and eq 3.1.14 becomes more accurate. SMFA is related to some older concepts in thermochemistry.[10] Rearranging eq 3.1.11 gives some chemical reactions:

$$G_1G_2G_3G_4G_5 + G_2 + G_3 + G_4$$
$$\rightarrow G_1G_2 + G_2G_3 + G_3G_4 + G_4G_5 \text{ (Level 1)}$$

$$G_1G_2G_3G_4G_5 + G_2G_3 + G_3G_4$$
$$\rightarrow G_1G_2G_3 + G_2G_3G_4 + G_3G_4G_5 \text{ (Level 2)}$$

$$G_1G_2G_3G_4G_5 + G_2G_3G_4$$
$$\rightarrow G_1G_2G_3G_4 + G_2G_3G_4G_5 \text{ (Level 3)}$$

Level 1 reactions are known as *isodesmic* reactions,[15] Level 2 as *homodesmotic* reactions,[16] and Level 3 are (less well) known as *isoperiochic*, *superhomodesmic*, or *hyperhomodesmotic* reactions.[10,17,18] As the Level increases, we expect these reactions to become closer to thermoneutral, as the bonding environment in reactants and products becomes more similar.

**Treatment of Nonbonded Interactions.** Equation 3.1.14 approximates the energy associated with chemical bonding. No fragment contains groups that are separated from other groups by more than Level bonds. However, the total energy should contain contributions from the interaction of more distantly connected groups. This "nonbonded" energy, $E_{\text{nb}}$, is also evaluated using fragmentation.

If we fragment the molecule at Level 1:

$$M \rightarrow \sum_{n_1=1}^{N_{\text{frag}}^{(1)}} f_{n_1}^{(1)} F_{n_1}^{(1)} \tag{3.1.15}$$

then the nonbonded energy is given by[10,12]

$$E_{nb} = \frac{1}{2} \sum_{n_1=1}^{N_{\text{frag}}^{(1)}} \sum_{n_2=1}^{N_{\text{frag}}^{(1)}} f_{n_1}^{(1)} f_{n_2}^{(1)} E[F_{n_1}^{(1)} \leftrightarrow F_{n_2}^{(1)}]_{\text{allowed}} \tag{3.1.16}$$

where $E[F_{n_1}^{(1)} \leftrightarrow F_{n_2}^{(1)}]_{\text{allowed}}$ denotes the energy of interaction between two fragments, if that interaction is "allowed"; meaning that the interaction was not included in $E_b$.[19] These fragment–fragment interactions can be evaluated ab initio in the usual way:

$$E[F_{n_1}^{(1)} \leftrightarrow F_{n_2}^{(1)}] = E[F_{n_1}^{(1)} + F_{n_2}^{(1)}] - E[F_{n_1}^{(1)}] - E[F_{n_2}^{(1)}]$$

(3.1.17)

Alternatively, if the two Level 1 fragments are well-separated in space, then the interaction energy can be accurately approximated using perturbation theory, including contributions from electrostatics, induction, and dispersion.[19,20]

**SMFA for Very Large Molecules.** During the SMFA fragmentation process, a large number of fragments are generated, most of which cancel with *identical* fragments which have opposite coefficients ($f_n$), so that the final value of $N_{frag}$ is of order $N$. The computation time and memory required to handle this large number of fragments is made negligible by the following algorithm.[13]

The molecule is subjected to a sequence of *compressions*. The functional groups are arbitrarily combined into exclusive pairs of directly bonded groups. These pairs of groups are bonded to other pairs, so we can form pairs of pairs. Repeating the compression $K$ times means that the number of "groups" (actually pairs of pairs⋯of groups) in the final "molecule" is reduced by a factor approaching $2^K$ (we stop when the number of "groups" is reduced to about 10).

Now, the compressed molecule is fragmented at the chosen Level to produce a set of fragments, $F_n$, and associated coefficients, $f_n$, as above. Then, each fragment is *expanded* by replacing each "group" by the pair of "groups" which it comprises. We simply repeat this "reverse iteration" process of fragmentation followed by expansion until no more expansions are possible. Fragmenting small molecules is computationally inexpensive and cancellation of fragments with opposite coefficients occurs at each reverse step. The total computation time and memory were found to be negligible for biological molecules containing some thousands of atoms.[13] Random reordering of all 2048 atoms in a protein did not change the final fragments at Levels 1 to 4, indicating that the procedure produces a unique set of fragments.[13]

**Fragmentation of Crystals.** Crystals are *periodic* structures that are determined by the position of the atoms in one *unit cell* and the *lattice vectors* that define the shape and dimension of the unit cell. Fragmentation is a novel and potentially very useful approach to the electronic structure of *nonconducting* crystals. As a simple example, consider a one-dimensional crystal with three groups, A−C, per unit cell. The infinite crystal structure would look like this:

$$\cdots A_{n-1}B_{n-1}C_{n-1}A_nB_nC_nA_{n+1}B_{n+1}C_{n+1}\cdots$$
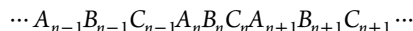
From eq 3.1.11, it should be clear that the Level 3 fragmentation of this lattice is

$$\cdots + C_{n-1}A_nB_nC_n + A_nB_nC_nA_{n+1} + B_nC_nA_{n+1}B_{n+1}$$
$$+ C_nA_{n+1}B_{n+1}C_{n+1} + \cdots$$
$$\cdots - C_{n-1}A_nB_n - A_nB_nC_n - B_nC_nA_{n+1} - C_nA_{n+1}B_{n+1} - \cdots$$
$$= \sum_{n=-\infty}^{n=\infty} \left[ \begin{array}{l} A_nB_nC_nA_{n+1} + B_nC_nA_{n+1}B_{n+1} + C_nA_{n+1}B_{n+1}C_{n+1} \\ - B_nC_nA_{n+1} - C_nA_{n+1}B_{n+1} - A_{n+1}B_{n+1}C_{n+1} \end{array} \right]$$

(3.1.18)

The infinite lattice is defined by these 6 fragments which define one unit cell, and the lattice energy *per unit cell.*

The SMFA scheme above is applied to a general 3D crystal structure with some relatively simple modifications.[14,21] Fragmentation of a large segment of crystal, surrounding some central unit cell, proceeds as for a molecule, except that

any fragment that does not contain a central-unit-cell group is immediately discarded. Any fragment which is a translation (by one or more unit cells) of another fragment is discarded. The remaining fragments comprise a single unit cell, as in eq 3.1.18. In general, the crystal structure C is given by

$$C \rightarrow \sum_{l_1=-\infty}^{l_1=\infty} \sum_{l_2=-\infty}^{l_2=\infty} \sum_{l_3=-\infty}^{l_3=\infty} \sum_{n=1}^{N_{frag}} f_n F_n(l_1, l_2, l_3)$$

(3.1.19)

where $l_1$, $l_2$, and $l_3$ are integers which label unit cells in the crystal. Since every unit cell has the same structure, the energy of the crystal *per unit cell* (due to bonded interactions) is

$$E_b^{UC} = \sum_{n=1}^{N_{frag}} f_n E[F_n(0, 0, 0)]$$

(3.1.20)

So, the computation for an infinite crystal is merely that for $N_{frag}$ fragments. *Any* level of ab initio quantum chemistry can be employed to estimate the crystal energy. In addition, if the crystal structure has high symmetry, many of the $F_n(0,0,0)$ fragments only differ by orientation. For example, the diamond lattice at Level 2 involves 24 molecular fragments, but only 2 of these are unique, so only 2 ab initio energy calculations are necessary.

As for molecules, there are "nonbonded" interactions, which are accounted for using *allowed* interactions between Level 1 fragments.[21]

The crystal structure is determined by minimizing eq 3.1.20 wrt (with respect to) both the atomic positions and lattice vectors.[14] The second derivatives of the crystal energy wrt the atomic positions can be readily calculated from the Hessians of the individual fragments, so that all the crystal vibration (phonon) frequencies can be evaluated.[21]

It is important to note that this fragmentation approach fails for metals or any conducting crystal, for the same reason that benzene cannot be fragmented. Electron delocalization ensures that benzene is a functional group by itself, and that a metal cannot be decomposed into small fragments.

**Fragmentation of Crystal Surfaces.** A crystal can be cleaved along a plane of symmetry. The bonding on the surfaces of these cleaved crystals is radically disturbed, so such surfaces may act as catalysts; hence the interest in the structure, properties and reactivity of cleaved crystals. Fragmentation takes account of the fact that the bonding near the surface is different from the bulk.[22]

Two lattice vectors can be chosen to lie in the cleavage plane (with indices $l_2$ and $l_3$). Then fragmentation of the cleaved crystal can be written as[22]

$$C_{surf} \rightarrow \sum_{l_1=-\infty}^{-1} \sum_{l_2=-\infty}^{\infty} \sum_{l_3=-\infty}^{\infty} \sum_{n=1}^{N_{frag}} f_n F_n(l_1, l_2, l_3)$$
$$+ \sum_{l_2=-\infty}^{\infty} \sum_{l_3=-\infty}^{\infty} \sum_{n=1}^{N_{frag}^{surf}} g_n G_n(l_2, l_3)$$

(3.1.21)

Here fragmentation of the bulk crystal describes the structure only in the range $-\infty < l_1 \leq -1$. The structure is different at the surface ($l_1 = 0$) and is therefore described by a different set of fragments,[22] which are summed across the surface plane.

**Embedded Charges.** Induction is difficult to describe using fragmentation.[23,24] Each group in a molecule or crystal is polarizable, so an electric dipole moment is induced in each

group by the *net* electric field at each group. The induced dipole moments have an energy in the presence of the local field, the *induction energy*. The important point is that the *net* field is a collective property of the whole molecule or crystal. An approximate solution to this problem was first implemented in the context of fragmentation by Li and co-workers,[8] and widely adopted thereafter. The electric field at a molecular fragment is modeled by the field produced by charges (and possibly higher electric moments) at the positions of atoms in the rest of the molecule. The atomic charges have been calculated in a number of ways,[8,25−28] including by a natural population analysis,[29,30] or using Stone's distributed multipole approach.[31] The charges may be "refined" in an iterative process. Once the charges have been determined, all the fragment energies in eqs 3.1.14 and 3.1.17 are obtained from ab initio calculations in which these charges are included. Care has to be taken so that "double counting" of charge−charge interactions is removed.[8] For formally charged or highly polar molecules, this approach appears to accurately account for induction.[27]

### 3.2. Combined Fragmentation Method (CFM)

Ultimately both SMF and CFM approximate the total energy of a molecule or cluster by the expression

$$E_{tot} = \sum_{i=1}^{N_{frag}} f_i E(\hat{h}F_i) \tag{3.2.1}$$

where, as before, $f_i$ is an integer coefficient and $E(\hat{h}F_i)$ represents the energy of fragment $F_i$ (possibly hydrogen-capped, signified by $\hat{h}$). The CFM algorithm for obtaining the $f_i$ and $F_i$ for all $N_{frag}$ is given in Figure 1. We discuss each of the steps shown in the figure below.

**The CFM Algorithm.** As indicated in Figure 1, groups must first be defined, as in section 2.2. A "precursory" fragmentation is performed. The precursory fragmentation in CFM is identical to Level 1 of SMF, except that the hydrogen capping procedure uses standard bond lengths rather than the covalent radii used in SMF. However, we remind the reader that the groups in CFM are larger than those in SMF, so the precursory fragments in CFM are significantly larger than SMF at Level 1. It is, in principle, possible to perform a precursory fragmentation in CFM according to the SMF Level 2 (or higher level) procedure. In practice, however, Level 1 precursory fragmentation works best, but the method is scalable in this regard.

After performing the precursory fragmentation, all fragments are interacted with one another. Here we define an interaction energy as

$$\mathcal{E}(F_i, F_j) = E(\hat{h}\{F_i \cup F_j\}) - E(\hat{h}F_i) - E(\hat{h}F_j) + E(\hat{h}\{F_i \cap F_j\}) \tag{3.2.2}$$

which is a generalized version of eq 3.1.17 that allows for fragments $F_i$ and $F_j$ to overlap with one another (i.e., to contain some of the same groups). It is readily seen from eq 3.2.2 that if $F_i \subseteq F_j$ or $F_j \subseteq F_i$ then $\mathcal{E}(F_i, F_j) = 0$.

The total CFM interaction energy [analogous to eq 3.1.16] is then

$$\mathcal{E} = \sum_{j>i}^{N_{frag}} f_i^P f_j^P \mathcal{E}(F_i, F_j) \tag{3.2.3}$$

Note that in eq 3.2.3 we do not use the qualifier "allowed" as in SMF eq 3.1.16. Equation 3.2.3 is then added to the initial
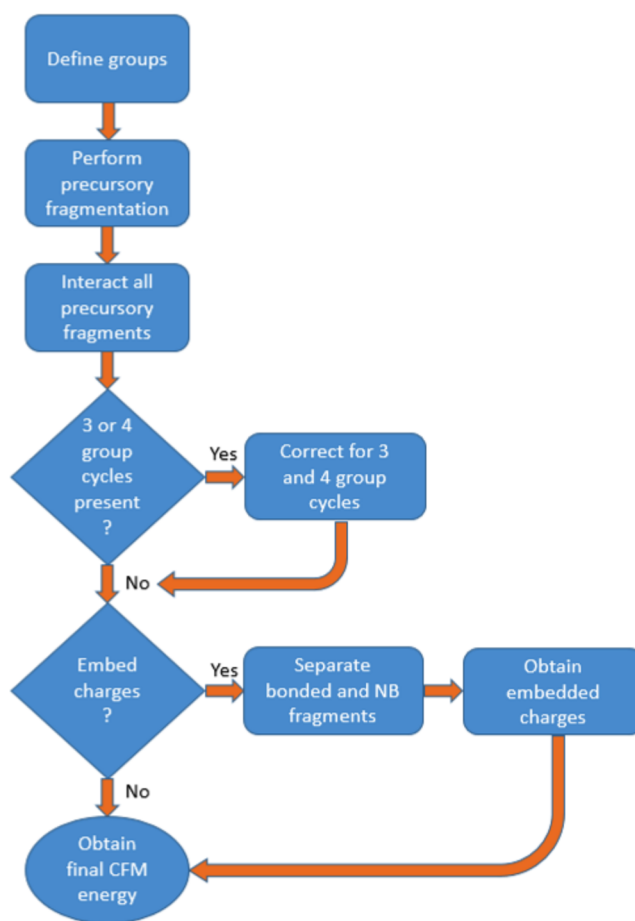


**Figure 1.** CFM algorithm. See the text for details of each step.

precursory fragmentation expression to obtain the total (combined) fragmentation energy expression for CFM:

$$E_{frag} = \sum_{i=1}^{N_{frag}} f_i E(\hat{h}F_i) + \sum_{j>i}^{N_{frag}} f_i^P f_j^P \mathcal{E}(F_i, F_j) \tag{3.2.4}$$

The above procedure can produce a very large number of fragment calculations for large molecules. However, it is noted that the vast majority of the nonbonded interaction energies may be computed virtually exactly through use of perturbation theory, with the accuracy being dependent upon the distance between interacting monomers.[32]

If the target molecule to be fragmented contains no three or four group cycles, then eq 3.2.4 represents the final fragmentation expression. However, if such cycles exist, then the modifications detailed below are applied to eq 3.2.4 to improve accuracy, without significant additional computational expense.

Application of eq 3.2.3 to a three-group cycle, $G_i G_j G_k$, where $G_k$ is also valence bonded to $G_i$, yields $\mathcal{E} = 0$. While this is correct based on eq 3.2.3, it misses the three-group interaction energy, which CFM readily computes in noncycle situations. It is therefore prudent to amend the CFM energy expression by adding the three-body interaction energy defined by

$$\varepsilon(G_i, G_j, G_k) = E(\hat{h}\{G_i \cup G_j \cup G_k\}) - E(\hat{h}\{G_i \cup G_j\})$$
$$- E(\hat{h}\{G_i \cup G_k\}) - E(\hat{h}\{G_j \cup G_k\})$$
$$+ E(\hat{h}G_i) + E(\hat{h}G_j) + E(\hat{h}G_k) \qquad (3.2.5)$$

to the total energy expression. This correction is applied for each three-cycle present in the system that has not already been accounted for.

Application of eq 3.2.3 to a four-group cycle, $G_iG_jG_kG_l$, where $G_l$ is also valence bonded to $G_i$, yields a CFM energy expression that double-counts the four-group interaction energy $\varepsilon$-$(G_iG_jG_kG_l)$. This occurs because the precursory fragments $G_iG_j$ and $G_kG_l$ produce, when interacted, the exact same fragment $G_iG_jG_kG_l$ as does the fragments $G_lG_i$ and $G_jG_k$. Thus, the energy of a cyclic four-group fragment will be double-counted in the final CFM energy expression. To correct for this, the extra four-group interaction energy is removed for each four-cycle present in the system that has not already been removed.

**Embedded Charges.** If there is a significant coulomb field pervading the entire molecular system, fragments may be embedded in a set of point charges as described in SMF. As with others, we have found that greater accuracy is obtained by including embedded charges, especially when monopoles are present in the system under consideration (e.g., protonated or deprotonated functional groups). In each fragment energy calculation in CFM, charges are placed at the location of all heavy atoms not present in the fragment being computed. Point charges are determined using Stone's distributed multipole analysis (DMA).[31] These charges must be determined iteratively. They can be approximately obtained by dividing CFM energy expression into bonded and nonbonded contributions. Only the bonded contributions are then utilized to obtain the iterated charges that are then placed around all fragments to obtain the final CFM energy.

## 4. APPLICATIONS AND EXAMPLES

### Reactions of Molecules and Crystal Surfaces Using SMF

It is important to realize that the fragment energies represent not only the energy of a fragment molecule at some single configuration, but the potential energy surface (PES) as a function of its configuration. The motion of the atomic nuclei is governed by the PES which is given by eq 3.1.14 as a sum over many fragment PESs.[35] So, for example, the PES for the dynamics of $n$-pentane is given at Level 1 fragmentation by a sum of PES for ethane and methane [see eq 3.1.11]:

$$E[CH_3(CH_2)_3CH_3]$$
$$= E[CH_3CH_3] + E[CH_3CH_3] + E[CH_3CH_3]$$
$$+ E[CH_3CH_3] - E[CH_4] - E[CH_4] - E[CH_4]$$

More interestingly, the reaction of another molecule with $n$-pentane is governed by a PES that is given by a sum of fragment PESs. At Level 1, the PES for a hydrogen atom reacting with $n$-pentane would be given by

$$E[H + CH_3(CH_2)_3CH_3]$$
$$= E[H + CH_3CH_3] + E[H + CH_3CH_3]$$
$$+ E[H + CH_3CH_3] + E[H + CH_3CH_3]$$
$$- E[H + CH_4] - E[H + CH_4] - E[H + CH_4]$$

The PES for H + $n$-pentane is a function in 48 dimensions. The PES for H + ethane is a function in only 21 dimensions. The *volume* of a molecular configuration space increases *exponentially* with the dimension. So, fragmentation *exponentially* reduces the task of constructing PES for molecular reactions. The same approach, via eq 3.1.21, yields a tractable PES for reactions on crystal surfaces.[36]

For general reactions, where multiple bonds might be formed and multiple bonds broken, all such bonds must be included in the description of bonding, as indicated in section 2.1, to ensure continuity of the fragmented PES. As yet, no global PES for such multiple bond forming reactions have been produced via SMFA.

### Potential Energy Correction Surfaces Using SMF

We could construct a more accurate surface for H + $n$-pentane using Level 2 fragmentation. However, a more efficient approach is to write

$$E[M] = E[\text{Level 1}] + \{E[\text{Level 2}] - E[\text{Level 1}]\}$$

This form is more efficient because the term $\{E[\text{Level 2}] - E[\text{Level 1}]\}$ is not a PES, but a potential energy correction surface (PECS).[35] A PECS is a flat, relatively featureless function because it only describes the difference of Level 2 fragment energies minus their Level 1 approximations. Relatively flat, featureless, functions are easier to approximate (requiring fewer ab initio calculations) than normal PES, which usually vary over tens or hundreds of kJ mol$^{-1}$ for chemical reactions.

So, the general point is that fragmentation provides a feasible approach to the chemical reactions of large molecules and cleaved crystal surfaces.[36]

### Energy Derivatives and Properties

The derivatives of the molecular energy wrt the atomic coordinates are given by the sum over derivatives of the fragment energies in eqs 3.1.14 and 3.1.16.[10] The fragment energy derivatives wrt H caps are assigned to the atoms via eq 3.1.12.[14] Ab initio methods that provide analytic gradients for the fragments thus provide analytic gradients for the whole molecule.[10,14] These derivatives are sufficiently accurate that accurate geometry optimization can be carried out using the energy gradients (and Hessians), and vibrational frequencies can be evaluated using the combined fragment Hessians.[10,18,37] Most molecular properties can be calculated as derivatives of the molecular energy wrt some quantity (eg the dipole moment is the derivative wrt an external electric field). So, molecular properties can be calculated from the derivatives of the fragment energies.

The NMR spectrum of a molecule provides a good illustration. The chemical shifts of various nuclei within a molecule are obtained from the trace of the nuclear magnetic shielding tensor, $\sigma_n$, for nucleus $n$. This tensor is the second derivative of the electronic energy with respect to an applied external magnetic field, **B**, and magnetic moment of the nucleus, **m**$_n$:

$$\sigma_n = \frac{\partial^2 E}{\partial \mathbf{B} \partial \mathbf{m}_n} \tag{4.0.1}$$

Within the fragmentation approximation, by applying the derivatives to expression 3.2.1, we readily obtain

$$\sigma_n = \sum_{i=1}^{N_{\text{frag}}} f_i \frac{\partial^2 E_i}{\partial \mathbf{B} \partial \mathbf{m}_n} \tag{4.0.2}$$

Although perhaps not obvious, the power of molecular fragmentation is well illustrated in eq 4.0.2. For example, if one was interested in the chemical shift of a particular nucleus in a large protein, then $\sigma_n$ need only be calculated for those few fragments that contain the nucleus of interest. Furthermore, nonbonded interactions involving fragments far from the nucleus of interest have little impact on the shielding tensor, so $\sigma_n$ can be readily obtained from a few small molecular fragment computations. CFM, and its precursor,[9] yield very accurate values of $\sigma_n$ compared to values computed from the full molecular wave function:[33,34] For a test set of proteins, the rms error in $\sigma$ for $^1$H, $^{13}$C, $^{15}$N, $^{17}$O, and $^{33}$S was 0.033, 0.218, 0.630, 2.919, and 0.040 ppm, respectively.[34]

CFM has been applied to other molecular properties, for example, central and distributed multipoles,[11] and molecular electrostatic potentials[22,38] (ESP), which are important for modeling interaction energies as part of rational drug design. The ESP is defined as

$$\phi_{\text{ESP}}(\mathbf{r}) = \sum_A^{\text{nuclei}} \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} - \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \, d\mathbf{r}' \tag{4.0.3}$$

where $Z_A$ is the nuclear charge on nucleus $A$ and $\rho(\mathbf{r}')$ is the electron density of the molecule. Because the ESP is exactly additive we can readily apply fragmentation to estimate $\phi_{\text{ESP}}(\mathbf{r})$,

$$\phi_{\text{ESP}}(\mathbf{r}) = \sum_{i=1}^{N_{\text{frag}}} f_i \left( \sum_{A_i}^{\text{nuclei}_i} \frac{Z_{A_i}}{|\mathbf{r} - \mathbf{R}_{A_i}|} - \int \frac{\rho_i(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}_i'|} \, d\mathbf{r}_i' \right) \tag{4.0.4}$$

where $\rho_i(\mathbf{r}')$ is the electron density of fragment $F_i$. The fragment ESP can be obtained either directly from the fragment electronic wave function, or from a distributed multipole analysis. The former approach is more accurate, while the latter is computationally efficient and just as accurate as the former provided (a) at least up to distributed quadrupoles are included and (b) $\mathbf{r}$ lies well outside the van der Waals radii of the atoms in the target molecule.

Other applications of CFM include[11] obtaining the density matrix of the complete molecule, or the self-consistent reaction field (SCRF) energy of a target molecule by embedding the fragments in a continuum dielectric material, as one would do for the target molecule.

## Examples

Over the last several years, hundreds of molecules, and a few crystals and crystal surfaces, have been studied using CFM and SMF. The accuracy and efficiency of these methods for calculating energies, energy gradients and Hessians, and other properties has been well established. Hence, we will not attempt an exhaustive review of these results, but merely present a few illustrative examples.

Figure 2 presents the CPU time for MP2/6-31G energy calculations, from ref 19, for a set of 82 moderate-sized molecules (which could be fragmented at Level 3) versus the
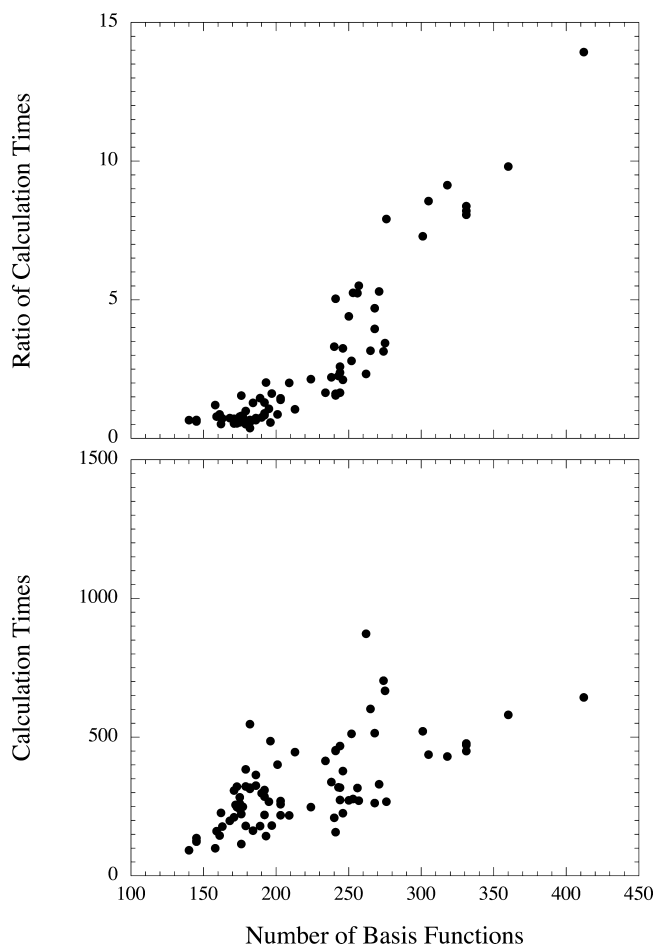


**Figure 2.** Upper: The ratio of the CPU time for a MP2/6-31G calculation of the whole molecule to the CPU time for the corresponding fragmentation calculation, at Level 3 plus nonbonded contributions, is shown versus the number of basis functions. Lower: The CPU time (in s) for SMF Level 3 including nonbonded contributions.

number of basis functions required for the whole molecule. The ratio of the CPU time required for the whole molecule to the corresponding SMF time is also shown. The significant computational efficiency of SMF is evident, even for relatively low numbers of basis functions. The CPU time for a given number of basis functions fluctuates substantially due to the varied structures of the molecules and the associated variation of the Level 3 fragments, but linear scaling is still indicated.

In ref 21, the structure of the $\alpha$-quartz crystal was optimized (using energy gradients and Hessians) with SMFA Level 3 at MP2/aug-cc-pVDZ. Figure 3 presents the energy of this optimized structure for MP2, CCSD, and CCSD(T) at various levels of fragmentation. This figure indicates the convergence of the energy versus Level of fragmentation and the applicability of very high level ab initio methods to crystals using SMFA.

In addition, Figure 4 depicts four molecules (of differing types) considered herein. The Cartesian coordinates for these molecules are included in the Supporting Information. Table 2 presents the errors in SMFA and CFM estimates of the total electronic energy of these four molecules. Some calculations have included the use of embedded charges to indicate the utility of this approach. Since the cost of an electronic structure calculation rises rapidly with the size of the basis set, it is clear that both CFM and SMFA provide significant computational
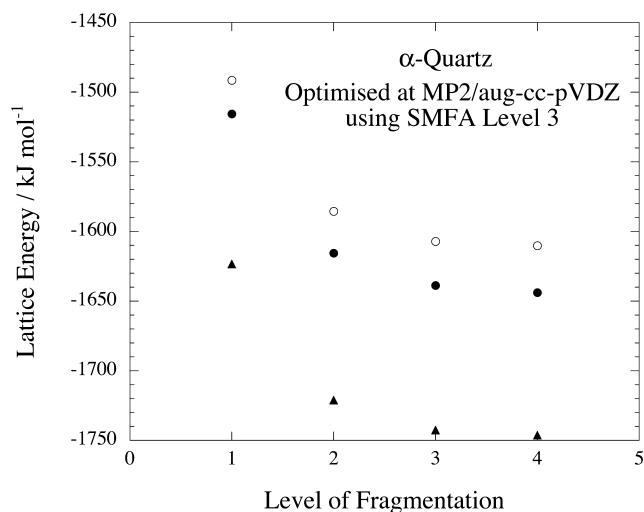
**Figure 3.** Lattice energy (per mole of SiO₂) of the SMFA Level 3 MP2/aug-cc-pVDZ optimized geometry for α-quartz, for MP2 (▲), CCSD (○), and CCSD(T) (●) at different Levels of fragmentation, including long-range electrostatics and dispersion.
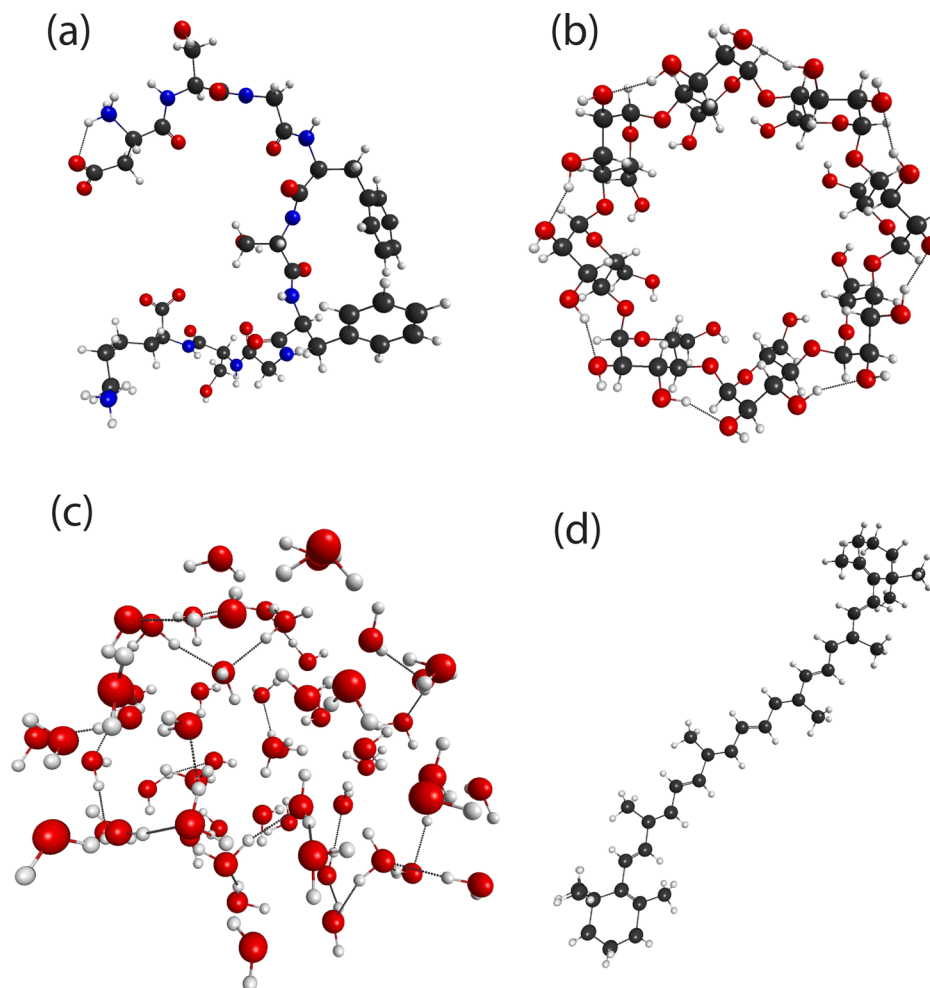
savings, since "Max basis" is very much smaller than "Basis (whole)". In all cases, one or both methods provide accurate estimates of the energy. The energy of a molecule such as β-

carotene is estimated very easily and accurately, presumably due to the absence of long-range interactions between charge distributions. Note that the bonding rules for both CFM and SMFA ensure that the β-carotene backbone is composed of alternating single and double bonds (the single bonds can be broken). In contrast, γ-cyclodextrin and water clusters contain a dense collection of highly polar groups, so that larger fragments (and thus larger values of "Max basis") and embedded charges are required to provide a reasonable estimate of the energy.

Table 3 presents an example (for the protein in Figure 2) of the accuracy of the energy estimate by SMFA as a function of the Level of fragmentation. Note that the number of separate ab initio calculations and the average size of the basis set also varies with the Level of fragmentation. This table indicates the utility of a method that provides systematic convergence of energy estimates.

## 5. CONCLUDING REMARKS

CFM and SMF provide the means to apply our chemical understanding of bonding and functional groups to the accurate calculation of molecular electronic energies, properties, and reactions. The dramatic reduction in the computer time that these fragmentation methods provide means that accurate quantum chemistry methods can be applied to large organic and biological molecules, and even to some crystalline materials.



**Figure 4.** Ball and stick figures for (a) the peptide 1OAI; (b) γ-cyclodextrin; (c) 57 water molecules; (d) β-carotene.

2783

dx.doi.org/10.1021/ar500088d | Acc. Chem. Res. 2014, 47, 2776–2785

**Table 2. Error in Total Electronic Energies for the Molecules of Figure 4[a]**

| system | method | stoichiometry | basis (whole) | max basis CFM | error CFM (mE_h) | max basis SMFA | error SMFA (mE_h) |
|---|---|---|---|---|---|---|---|
| $\beta$-carotene | HF[b] | $C_{40}H_{56}$ | 888 | 420 | $-0.026^d$ | $252^{df}$ | $0.245^{df}$ |
|  | B3LYP[b] |  |  | 420 | $0.592^d$ | $252^{df}$ | $0.326^{df}$ |
| $\gamma$-cyclodextrin | HF[b] | $C_{48}H_{80}O_{40}$ | 1824 | 612 | $3.042^d$ | $306^g$ | $5.343^g$ |
|  | B3LYP[b] |  |  | 612 | $5.638^d$ | $306^g$ | $4.350^g$ |
| 1OAI (peptide) | HF[b] | $C_{41}H_{58}N_{10}O_1$ | 1362 | 360 | $1.380^d$ | $324^h$ | $1.940^h$ |
|  | HF[b] |  |  | 360 | $-0.901^e$ |  |  |
| $(H_2O)_{57}$ | HF[c] | $O_{57}H_{114}$ | 2337 | 164 | $6.648^d$ | $451^i$ | $1.107^i$ |

[a]"Basis (whole)" means the number of basis functions for the whole molecule; "max basis" is the maximum number of basis functions for a single fragment calculation. [b]6-311G(d) basis set. [c]aug-cc-pVDZ basis set. [d]Without embedded charges. [e]With embedded charges. [f]Level 3. [g]Level 4 with embedded charges representing all OH groups. [h]Level 3 with embedded charges on formally charged groups only. [i]Level 3 with embedded charges on all atoms.

**Table 3. For the Peptide 1OAI, the SMFA Error in the Estimated Energy [HF/6-311G(d), as in Table 2] is Shown versus Level[a]**

| level | max basis | mean basis | number of fragments | error (mE_h) |
|---|---|---|---|---|
| 2 | 324 | 93 | 105 | 16.831 |
| 3 | 324 | 119 | 104 | 1.940 |
| 4 | 336 | 129 | 99 | 0.701 |
| 5 | 417 | 149 | 97 | 0.140 |

[a] "Number of fragments" is the total number of *ab initio* calculations, and "Mean basis" is the average size of the basis set over all fragments.

## ASSOCIATED CONTENT

### Ⓢ Supporting Information

The Cartesian coordinates of the molecules in Figure 4. This material is available free of charge via the Internet at http://pubs.acs.org/

## AUTHOR INFORMATION

### Corresponding Authors

*E-mail: collins@rsc.anu.edu.au,.
*E-mail: ryan.pa.bettens@nus.edu.sg.

### Notes

The authors declare no competing financial interest.

### Biographies

**Michael A. Collins** [BSc (1974) and PhD (1977), University of Sydney] was a postdoctoral fellow at MIT, a Fulbright Scholar at the University of Chicago, and a Nuffield Fellow at Cambridge University. He has been a Professor in the area of Theoretical Chemical Physics at the Australian National University since 2002.

**Milan W. Cvitkovic** [BA (2013) Carleton College] is a Fulbright Scholar in Theoretical and Computational Chemistry at the National University of Singapore.

**Ryan P. A. Bettens** [BSc (1986) University of Queensland, and PhD (1992), Monash University] was a postdoctoral fellow at ETH, Zurich; Ohio State University; and Australian National University. He has been an Associate Professor in the area of Theoretical and Computational Chemistry at the National University of Singapore since 2008.

## ACKNOWLEDGMENTS

## REFERENCES

(1) Fedorov, D. G.; Kitaura, K. *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*; CRC Press: Boca Raton, FL, 2009.

(2) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2012**, *112*, 632−672.

(3) Zhang, D. W.; Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein-molecule interaction energy. *J. Chem. Phys.* **2003**, *119*, 3599−3605.

(4) He, X.; Zhang, J. Z. H. A new method for direct calculation of total energy of protein. *J. Chem. Phys.* **2005**, *122*, 031103.

(5) Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. Molecular Tailoring approach for simulation of electrostatic properties. *J. Phys. Chem.* **1994**, *98*, 9165−9169.

(6) Babu, K.; Gadre, S. R. Ab initio quality one-electron properties of large molecules: Development and testing of molecular tailoring approach. *J. Comput. Chem.* **2003**, *24*, 484−495.

(7) Li, S. H.; Li, W.; Fang, T. An efficient fragment-based approach for predicting the ground-state energies and structures of large molecules. *J. Am. Chem. Soc.* **2005**, *127*, 7215−7226.

(8) Li, W.; Li, S. H.; Jiang, Y. S. Generalized energy-based fragmentation approach for computing the ground-state energies and properties of large molecules. *J. Phys. Chem. A* **2007**, *111*, 2193−2199.

(9) Bettens, R. P. A.; Lee, A. M. A new algorithm for molecular fragmentation in quantum chemical calculations. *J. Phys. Chem. A* **2006**, *110*, 8777−8785.

(10) Deev, V.; Collins, M. A. Approximate ab initio energies by systematic molecular fragmentation. *J. Chem. Phys.* **2005**, *122*, 154102.

(11) Le, H. A.; Tan, H. J.; Ouyang, J. F.; Bettens, R. P. A. Combined Fragmentation Method: A Simple Method for Fragmentation of Large Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 469−478.

(12) Collins, M. A.; Deev, V. A. Accuracy and efficiency of electronic energies from systematic molecular fragmentation. *J. Chem. Phys.* **2006**, *125*, 104104.

(13) Collins, M. A. Systematic Fragmentation of Large Molecules by Annihilation. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7744−7751.

(14) Netzloff, H. M.; Collins, M. A. Ab initio energies of non-conducting crystals by systematic fragmentation. *J. Chem. Phys.* **2007**, *127*, 134113.

(15) Hehre, W. J.; Ditchfield, R.; Radom, L.; Pople, J. A. Molecular orbital theory of the electronic structure of organic compounds. V. Molecular theory of bond separation. *J. Am. Chem. Soc.* **1970**, *92*, 4796−4801.

(16) George, P.; Trachtman, M.; Bock, C. W.; Brett, A. M. An alternative approach to the problem of assessing stabilization energies in cyclic conjugated hydrocarbons. *Theor. Chim. Acta* **1975**, *38*, 121−129.

(17) Baldridge, K. K.; Gordon, M. S. Potentially Aromatic Metallocenes. *J. Am. Chem. Soc.* **1988**, *110*, 4204−4208.

(18) Hess, B. A., Jr.; Schaad, L. J. Ab Initio Calculation of Resonance Energies. Benzene and Cyclobutadiene. *J. Am. Chem. Soc.* **1983**, *105*, 7500.

(19) Addicoat, M. A.; Collins, M. A. Accurate treatment of nonbonded interactions within systematic molecular fragmentation. *J. Chem. Phys.* **2009**, *131*, 104103.

(20) Mullin, J. M.; Roskop, L. B.; Collins, M. A.; Gordon, M. S. Systematic Fragmentation Method and the Effective Fragment Potential: An Efficient Method for Capturing Molecular Energies. *J. Phys. Chem. A* **2009**, *113*, 10040−10049.

(21) Collins, M. A. Ab initio lattice dynamics of nonconducting crystals by systematic fragmentation. *J. Chem. Phys.* **2011**, *134*, 164110.

(22) Frankcombe, T. J.; Collins, M. A. Potential energy surfaces for gas-surface reactions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 8379−8391.

(23) Kitaura, K.; Sawai, T.; Asada, T.; Nakano, T.; Uebayasi, M. Pair interaction molecular orbital method: an approximate computational method for molecular interactions. *Chem. Phys. Lett.* **1999**, *312*, 319−324.

(24) Song, L.; Han, J.; Lin, Y-l.; Xie, W.; Gao, J. Explicit Polarization (X-Pol) Potential Using ab Initio Molecular Orbital Theory and Density Functional Theory. *J. Phys. Chem. A* **2009**, *113*, 11656−11664.

(25) Duan, L. L.; Tong, Y.; Mei, Q. G.; Zhang, J. Z. H. Quantum study of HIV-1 protease-bridge water interaction. *J. Chem. Phys.* **2007**, *127*, 145101.

(26) Le, H. A.; Lee, A. M.; Bettens, R. P. A. Accurately Reproducing Ab Initio Electrostatic Potentials with Multipoles and Fragmentation. *J. Phys. Chem. A* **2009**, *113*, 10527−10533.

(27) Pruitt, S. R.; Addicoat, M. A.; Collins, M. A.; Gordon, M. S. The fragment molecular orbital and systematic molecular fragmentation methods applied to water clusters. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7752−7764.

(28) Reid, D. M.; Collins, M. A. Molecular electrostatic potentials by systematic molecular fragmentation. *J. Chem. Phys.* **2013**, *139*, 184117.

(29) Reed, A. E.; Weinhold, F. Natural bond orbital analysis of near-Hartree−Fock water dimer. *J. Chem. Phys.* **1983**, *78*, 4066−4073.

(30) Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural population analysis. *J. Chem. Phys.* **1985**, *83*, 735−746.

(31) Stone, A. J. Distributed Multipole Analysis: Stability for Large Basis Sets. *J. Chem. Theory Comput.* **2005**, *1*, 1128.

(32) Bettens, R. P. A.; Lee, A. M. On the accurate reproduction of ab initio interaction energies between an enzyme and substrate. *Chem. Phys. Lett.* **2007**, *449*, 341−346.

(33) Lee, A. M.; Bettens, R. P. A. First Principles NMR Calculations by Fragmentation. *J. Phys. Chem. A* **2007**, *111*, 5111−5115.

(34) Tan, H. J.; Bettens, R. P. A. Ab initio NMR chemical-shift calculations based on the combined fragmentation method. *Phys. Chem. Chem. Phys.* **2013**, *15*, 7541−7547.

(35) Collins, M. A. Molecular potential energy surfaces constructed from interpolation of systematic fragment surfaces. *J. Chem. Phys.* **2007**, *127*, 024104.

(36) Frankcombe, T. J.; Collins, M. A. Growing Fragmented Potentials for Gas-Surface Reactions: The Reaction between Hydrogen Atoms and Hydrogen-Terminated Silicon (111). *J. Phys. Chem. C* **2012**, *116*, 7793−7802.

(37) Collins, M. A. Molecular forces, geometries and frequencies by systematic molecular fragmentation including embedded charges. Submitted for publication.

(38) Sahu, N.; Singh, G.; Gadre, S. R. Exploring Structures and Energetics of Large OCS Clusters by Correlated Methods. *J. Phys. Chem. A* **2013**, *117*, 10964−10972.