A New Algorithm for Molecular Fragmentation in Quantum Chemical Calculations

Ryan P. A. Bettens* and Adrian M. Lee†

Department of Chemistry, National University of Singapore, 3 Science Drive 3, Singapore 117543

Received: April 4, 2006; In Final Form: May 19, 2006

In this study, we present a "black-box" method for fragmenting a molecule with a well-defined Kekulé or valence-bond structure into a significant number of smaller fragment molecules that are more amenable to high level quantum chemical calculations. By taking an appropriate linear combination of the fragment energies, we show that it is possible in many cases to obtain highly accurate total energies when compared to the total energy of the full molecule. Our method is derived from the approach reported by Deev and Collins,[43] but it contains significant unique elements, including an isodesmic approach to the fragmentation process. Using a method such as that described in this work it is in principle possible to obtain very accurate total energies of systems containing hundreds, if not thousands, of atoms as the approach is subject to massive parallelization.

1. Introduction

The primary goal of physical chemistry is to produce a theoretical framework within which to predict the physical properties as well as the chemical reactivity of all matter. Implicitly, application of the Schrödinger and Dirac equations does this, but the ab initio methods available to solve these equations suffer from a steep scaling problem that generally prohibits the accurate estimation of properties for large systems. Achieving linear scaling with respect to the size of the system being considered is greatly desired and is a necessary criterion if the dream of ab initio calculations of systems of many thousands of atoms is to be realized.

In the recent past, much has been done to reduce the nonlinear scaling inherent in many of the key steps in traditional ab initio methods to near linearity. Linear-scaling methods are primarily based on the principle of quantum locality[1] or "near-sightedness",[2] that the properties of a certain observation region of only one or a few atoms are only weakly influenced by factors that are spatially far away from this observation region. This can be achieved by limiting to a local region of space the physical span of the electronic degrees of freedom.[3] Careful consideration of such underlying physics and improved mathematical methods have led to linear scaling in, inter alia, the calculations of the Coulomb[4−10] and exchange[11,12] integrals, and in alternative approaches to the direct diagonalization of the Fock matrix.[13−23] A comprehensive discussion of linear-scaling methods in electronic structure calculations can be found in the review given by Goedecker.[1]

An alternative approach to large system problems is the hybrid quantum-mechanical/molecular-mechanical (QM/MM) approach[24−27] such as the ONIOM method,[28] in which a part of the system deemed important for properties of interest is treated at a high level of accuracy, while the rest of the system is treated approximately. Such an approach provides a powerful means to study biological systems. Recently, applications to potential energy surfaces and transition state structures for enzyme reactions using QM/MM methods have been reported.[29−32] However, quantum mechanics and molecular mechanics are inherently incompatible with one another, and this leads to an interface problem between the QM and MM regions. Resolving this problem remains the main difficulty in the QM/MM approach.

The "divide-and-conquer" method developed by Yang[16,17] to circumvent the cubic scaling inherent in the diagonalization of the Fock matrix has generated a new category of approaches known as fragment-based methods. In fragment-based methods, rather than treat the whole system at once, the system is divided into a set of subsystems (or fragments). Conventional quantum-mechanical calculations are then performed on all the fragments, the results of which are then combined in some way to determine various properties of the whole system. Many fragment-based approaches have been developed: the adjustable density matrix assembler approach[33−35] of Exner and Mezey; the fragment molecular orbital approach[36−39] of Kitaura and co-workers; the elongation method[40,41] of Imamura and co-workers; the molecular tailoring approach[42−44] of Gadre and co-workers; the ab initio fragment orbital-based theory[45] of Das et al.; the kernel energy method[46] of Huang et al.; the molecular fractionation with conjugate caps scheme[47−51] developed by Zhang and co-workers. These methods are inherently linearly scaling and have been demonstrated to yield quite accurate properties for various types of large molecules. It has been noted[38] that perhaps the best reason for employing a fragment-based approach is the ease of achieving massive parallelization in comparison to whole-system solutions.

Two recent studies,[52,53] based on the molecular fractionation with conjugate caps approach,[47] have shown that the total ground-state energy of a molecule can be directly computed by fragmenting the molecule and taking a linear combination of the resultant fragment total energies. These results were achieved at both the Hartree−Fock and post-Hartree−Fock levels of theory. In addition, Li et al.[52] implemented geometry optimization in their work, while Deev and Collins[53] showed more generally that accurate gradients and second derivatives of the energy can also be obtained through the use of their fragment-based method. Both sets of authors have pointed out that for certain systems fragmentation may not give accurate results but

* Corresponding author. Fax: +65 6779 1691. Telephone: +65 6516 2846. E-mail: chmbrpa@nus.edu.sg.
† E-mail: chmaml@nus.edu.sg.

10.1021/jp062104n CCC: $33.50
Published on Web 06/17/2006

also that there exists a wide variety of systems for which their methods work well.

In the work of Li et al., no "black-box" algorithm was presented for determining the fragments, so each individual system under consideration requires a specialist to decide just which fragments should be considered. On the other hand, Deev and Collins have put forward a scheme for generating a hierarchy of molecular fragments whereby higher levels in the hierarchy produce molecular fragments of larger size but approximate the total energy more reliably.

The work presented here is derived from the approach taken by Deev and Collins, but with unique elements as described below. Often, but not always, the fragment set generated by our method is identical to that generated in the study by Deev and Collins. However, our approach has several advantages. First, our method is able to automatically deal with certain situations where fragmentation, as described by Deev and Collins, cannot be performed. Examples include fragmentation of cyclic systems at higher levels in the hierarchy; specifically our method allows the fragmentation of a six-group ring at level 3 and higher. Second, our method is simpler and much faster to implement. Third, on those occasions where our fragment set differs from Deev and Collins, we find that our fragments are smaller. Finally, the fragment set obtained using our method is unique for a given Kekulé or valence-bond structure, a criterion that has not yet been verified by Deev and Collins for all molecular morphologies. Our approach to computing the nonbonding interactions is unsophisticated, significantly simpler and less encompassing than previous treatments, as the main thrust of the work presented here is the fragmentation procedure.

The paper is organized in the following manner. In section 2, we describe our methodology and present the fragmentation algorithm. In section 3, we validate our method by applying it to a variety of chemical systems. Further, we give an example of a situation (in section 3.4) where our method produces different fragments compared with Deev and Collins. Finally, a brief summary is given in section 4.

2. Methodology

In their seminal paper, Hehre et al. introduced the term "isodesmic" to describe a chemical reaction in which the number of bonds of each formal type is conserved, but the relationship among the bonds is altered.[54] The energy of an isodesmic reaction is normally interpreted as a measure of deviation from the additivity of bond energies, and therein lies the popularity of employing isodesmic reactions in thermochemical calculations.[55] The method of fragmentation developed in this study employs an approach that is isodesmic in nature.

Any molecule can be represented as a collection of functional groups that are interconnected in some manner, with each group containing at least one polyvalent atom. If we consider a linear molecule that is composed of just three groups (G_1, G_2, and G_3), then a decomposition of this parent molecule into fragment molecules can be described by the following isodesmic reaction:

$$G_1G_2G_3 + H^{(1)}G_2H^{(2)} \rightarrow G_1G_2H^{(2)} + H^{(1)}G_2G_3 \qquad (1)$$

where $H^{(1)}$ and $H^{(2)}$ are "capping" hydrogen atoms. It can be seen that the capping hydrogen atoms in the fragment molecules replace bonding connections in the parent molecule. For instance, $H^{(1)}$ is bonded to G_2 whenever G_1 is not present in a particular fragment molecule, and $H^{(2)}$ is bonded to G_2 whenever G_3 is not present. Each capping hydrogen atom is placed 1 Å away from the relevant polyvalent atom in G_2 in the direction of the severed bond vector, i.e., $H^{(1)}$ lies 1 Å away from G_2 along the G_2G_1 bond vector. The use of hydrogen atoms assumes that the connections between groups are single bonds and so ensures that the number of single bonds in the reaction is conserved.

In their study, Hehre et al. further noted that a satisfactory method of predicting energies of bond separation in isodesmic reactions would be sufficient to predict the total energy of the parent molecule, provided that the energies of the bond separation fragments are known. Our ansatz is that if the bond separation fragments both maintain the geometry found in the parent molecule and are large enough to fully describe the local electronic environment, then these energies of bond separation are approximately zero.

Applying this ansatz to the above isodesmic reaction, i.e., assuming $\Delta_r E \approx 0$, we arrive at the following energy for the parent molecule:

$$E(\mathrm{P}) \approx E(\mathrm{F}_1) + E(\mathrm{F}_2) - E(\mathrm{F}_3) \qquad (2)$$

where $P = G_1G_2G_3$, $F_1 = G_1G_2H^{(2)}$, $F_2 = H^{(1)}G_2G_3$, and $F_3 = H^{(1)}G_2H^{(2)}$. Note that G_2 can be identified as the molecular overlap between fragments F_1 and F_2, and together with the relevant capping hydrogen atoms forms fragment F_3. Thus, the energy of the parent molecule can be approximated by a linear combination of energies of fragment molecules. In this paper, because of the way they are employed in eq 2, fragments F_1 and F_2 are referred to as "positive" fragments and fragment F_3 is referred to as a "negative" fragment. To reiterate, a critical point to note is that the geometry of each fragment is *exactly* the same as the geometry in the parent molecule, except for the capping hydrogen atoms.

By examining a limiting case of reaction 1 we arrive at a situation in which $\Delta_r E$ is *exactly* zero. This case is illustrated below:

$$G_1G_2\cdots G_3 + H^{(1)}G_2\cdots H^{(2)} \rightarrow G_1G_2\cdots H^{(2)} + H^{(1)}G_2\cdots G_3 \qquad (3)$$

In this example, G_3 and $H^{(2)}$ are moved to an infinite distance away from G_2. In this instance $\Delta_r E$ is exactly zero as the fragments on the left-hand side of reaction 3 are now identical to those on the right-hand side. Of course, a similar reaction is obtained by moving G_1 and $H^{(1)}$ to an infinite distance away from G_2 whereby $\Delta_r E = 0$. These simple examples offer some insight into when isodesmic fragmentation will lead to accurate total energies of parent molecules. For instance, if we allow G_3 to approach G_2 on both sides of reaction 3, then it is clear that in the parent molecule (on the left-hand side) G_3 "sees" G_1 whereas on the right-hand side in the second fragment G_3 "sees" $H^{(1)}$. Similarly as $H^{(2)}$ approaches G_2, $H^{(2)}$ "sees" $H^{(1)}$ on the left-hand side, but "sees" G_1 on the right-hand side. If G_2 is large such that the electronic environment in G_1 is only weakly influenced by the presence of G_3, then a similar conclusion can be reached for the interactions between $H^{(1)}$ and $H^{(2)}$ in the negative fragment and between G_1 and $H^{(2)}$ and G_3 and $H^{(1)}$ in the positive fragments. Using these arguments, it is clear that the most accurate total energies of parent molecules using isodesmic fragmentation will be achieved when G_2 is large spatially. This conclusion has two important consequences which are intrinsic to the fragmentation method described below. First, fragmentation becomes more accurate as the positive fragments become larger. In a hierarchical sense such as utilized in this study, this is achieved by having more groups in the positive

fragments. Second, fragmentation is most accurate when the overlap between positive fragments to form a negative fragment is maximal. These conclusions are valid for any fragmentation scheme that employs negative fragments to cancel overlapping contributions from positive fragments.

A helpful way to visualize our method for fragmenting a molecule is as follows. Consider the electronic environment of any particular group or region within a molecule. If one is to accurately describe the electronic structure about this group or region in a fragment then one should include as much of the parent molecule as possible that surrounds this group or region. So, for a level-1 fragmentation, we consider only the connections between groups to be important, and the collection of positive fragments involves forming all possible pairs of connected groups within the parent molecule. At level 2 our focus is a group itself, that is not a terminal group (a terminal group is any group that has a connectivity of one), and the collection of positive fragments involves connecting all groups that are connected to our central focus group. Higher levels are built up from these two levels. At level 3, we form all the positive fragments by connecting all groups directly connected to the positive fragments of level 1. At level 4, we form all the positive fragments by connecting all groups directly connected to the positive fragments of level 2. At level 5, we form all the positive fragments by connecting all groups directly connected to the positive fragments of level 3, etc. If at any point a positive fragment is generated that is a subfragment of another positive fragment, it is deleted since the larger fragment is a better description of the two. To obtain the final fragmentation energy of the molecule we essentially add together all the energies of the "positive" fragments and subtract from this energy all the things that have been counted two or more times in the positive fragments, i.e. the overlaps.

2.1. The Algorithm. The general algorithm for all levels is described in this paragraph with the details provided in the following paragraphs. A summary of the algorithm is given in Figure 1. (a) Form the initial set of positive fragments. (b) Add the caps to the initial fragments. These caps are hydrogen atoms or other groups from the original parent molecule depending on the proximity of capping hydrogen atoms from different groups in the initial fragments. (c) Eliminate from the new fragment list any fragments that are completely contained within other fragments. This new, and final, list of fragments defines the positive fragments. The coefficients multiplying the energies of the positive fragments to be used in obtaining the total energy of the molecule are, as yet, unknown—we term these coefficients the positive-fragment coefficients. (d) Solve for the negative-fragment set. (e) The negative-fragment coefficients are dependent on the positive-fragment coefficients; thus, in this step we derive these coefficients in terms of the currently unknown positive-fragment coefficients. (f) Use singular value decomposition to obtain the minimal norm solution for the positive-fragment coefficients. Very often, but not always, this solution produces a set of positive-fragment coefficients that are all unity and, as a consequence, negative-fragment coefficients that are integers.

(a) Form the initial set of positive fragments. At level 1 this involves forming all pairs of groups that are connected within the original molecule. At level 2, we consider each group in the molecule and form a fragment around it by connecting all groups directly connected to it. Fragments formed from terminal groups are always subfragments of other positive fragments. All subfragments are deleted. Level 3 is an extension of level 1. We start with the level-1 fragment set, and for each level-1

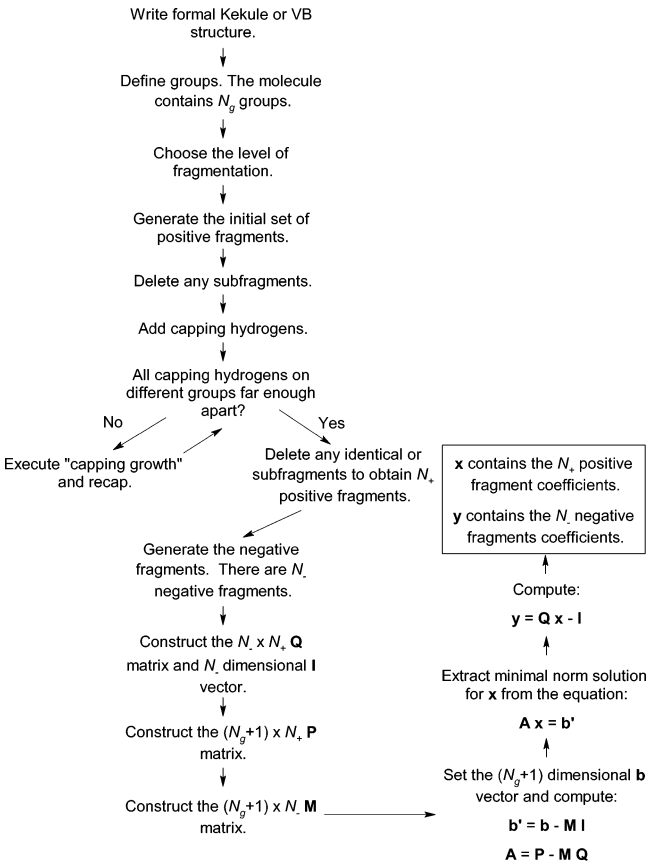


Figure 1. Summary of the algorithm used to determine fragments with the details described in the text.

fragment we directly connect all groups directly connected to it, thus forming a level-3 fragment. All subfragments are deleted. Level 4 is an extension of level 2. We start with the initial level-2 fragment set, and for each level-2 fragment we directly connect all groups directly connected to it, thus forming a level-4 fragment. All subfragments are deleted. Thus, for level N, where N is an integer greater than 2, we begin with the initial set of positive fragments from level $N - 2$, and connect all groups directly connected to it, deleting all subfragments that might be generated. The above procedure generates the initial positive-fragment set.

(b) Add the caps to the initial positive fragments. For each initial fragment, hydrogen caps are added along the bond that is broken and placed at a distance of 1.0 Å from the heavy atom. Each capping hydrogen atom is looped over in turn and the distance between it and other capping hydrogen atoms located on different groups is calculated. If any distance between capping hydrogen atoms is less than 2.2 Å then the looping is stopped and the group that should be where the capping hydrogen is located is added into the fragment and the hydrogen capping process is restarted. This process continues until the fragment has no hydrogen caps that are located on different groups closer than 2.2 Å. It is this procedure that can cause fragments to increase in size —we term this occurrence as "capping growth".

(c) Eliminate redundant fragments. As the procedure followed in step b can yield larger fragments, it is possible that the modified fragment list contains either (i) identical fragments or (ii) fragments that are subfragments of larger fragments. Thus, the list of positive fragments is searched and those that are the same as, or subfragments of, other fragments are deleted. This procedure produces the final list of positive fragments.

(d) Solve for the negative fragments. Level 1 is the simplest, and is most easily described if we also include step e here; step f is unnecessary for level 1 as we can always use unity as the positive-fragment coefficients. At level 1, all negative fragments are actually made up of single (capped) groups. Each group must appear at least once in the positive fragment list. If a group appears more than once then the energy of this capped group must be subtracted from the total energy of all the positive fragments. The number of times the energy of a capped group must be subtracted is one less than the total number of times the group appears in the positive fragment list.

For levels 2 and above, a more sophisticated procedure is required. Each positive fragment is looped through in turn and overlapped with all remaining positive fragments. "Overlap" here means, "that part of the two fragments that are in common". From the set of overlap fragments generated from a single positive fragment, we eliminate (i) any overlap fragments that are subfragments of this set of overlap fragments, and (ii) any identical copies of overlap fragments in this set. Each overlap fragment from this reduced set is then added to the list of negative fragments unless the overlap fragment is already in the negative-fragment list. After looping through all the positive fragments, the set of negative fragments is obtained.

(e) Obtain the negative-fragment coefficients in terms of the positive-fragment coefficients. The relationship between the negative- and positive-fragment coefficients can be conveniently expressed in matrix form. Here **Q** represents such a matrix with dimensions $N_- \times N_+$, where N_- and N_+ are the number of negative and positive fragments, respectively. **Q** is a matrix of ones and zeros where an entry of one for element $Q_{i,j}$ indicates that negative fragment i is a subfragment of positive fragment j. It is important that overcounting does not occur for the negative fragments. This is possible if a negative fragment is a subfragment (or sub-subfragment, etc.) of another negative fragment. Therefore, the rows of **Q** are sorted in descending order so that the largest negative fragment, in terms of number of groups in the fragment, appears in row one. Starting at negative fragment one and ending with fragment $N_- - 1$, each negative fragment is then looped over in turn and compared with the remaining negative fragments. If fragment j, e.g., is found to be a subfragment, then row j in **Q** must have the row corresponding to the parent negative fragment subtracted from it. Care must also be taken to keep track of the number of times this occurs. A vector that will be used in the solving the positive fragment coefficients needs to be updated. This vector initially has the value of one for every element and is N_- dimensional. We label it here as **l**. Each time row j in **Q** is altered because it is a subfragment of some other negative fragment, element j in **l** must have the parent element subtracted from it. The reason for the existence of **l** will become clear in the next paragraph.

(f) Solve for the positive-fragment coefficients and the dependent negative-fragment coefficients. A fragment, positive or negative, can be represented as an $N_g + 1$ dimension vector, where N_g is the number of groups in the parent molecule. The $(N_g + 1)$th element of this vector represents the number of capping hydrogen atoms in the fragment. This vector contains a one in element j if group j is present in the fragment, but is zero otherwise. When all the fragments are collected together they form the $(N_g + 1) \times N_+$ matrix **P** for the positive fragments or the $(N_g + 1) \times N_-$ matrix **M** for the negative fragments. If the negative fragment coefficients were independent of the positive fragment coefficients, then we could write down the matrix expression,

$$\mathbf{Bz} = \mathbf{c} \tag{4}$$

where **z** would be a $N_+ + N_-$ dimensional vector and $\mathbf{B} = \mathbf{P}|\mathbf{M}$, i.e., the $(N_g + 1) \times (N_+ + N_-)$ fragment matrix. The vector, **c**, is $(N_g + 1)$ dimensional and has the values of one in every element except the $(N_g + 1)$th which must be zero. Equation 4 simply ensures that no matter how we add together the energies of the fragments (given by the coefficients in **z**), we must ensure that the net number of times each group appears is 1, and that the total number of capping hydrogen atoms is 0 as is the case in the molecule.

Equation 4 is not very restrictive, and cannot be expected to produce a satisfactory total energy for the parent molecule in general. We would like to ensure that the energy of each negative fragment is subtracted from the linear combination of energies of the positive fragments based on the number of times, less one, the negative fragment is a subfragment of the positive fragments. We can ensure this occurs through the use of the **Q** matrix and the **l** vector. That is, we will solve for the positive-fragment coefficients, **x**, and having obtained these we can obtain the negative-fragment coefficients, **y**, via

$$\mathbf{y} = \mathbf{Qx} - \mathbf{l} \tag{5}$$

The number of times each group in the parent molecule appears in the positive fragments is given by **Px**. The $(N_g + 1)$th element of this vector is also the number of times that capping hydrogen atoms appear in the positive fragments. The number of times each group appears in the negative fragments is **My**. The $(N_g + 1)$th element of this vector is also the number of times that capping hydrogen atoms appear in the negative fragments. The difference between these two vectors must be equal to the number of times each group appears in the parent molecule, i.e., once, and the $(N_g + 1)$th element must be zero as the molecule contains no capping hydrogen atoms. This latter vector we write as **b**. Thus, we wish to solve,

$$\begin{aligned} \mathbf{Px} - \mathbf{My} &= \mathbf{b} \\ \mathbf{Px} - \mathbf{M}(\mathbf{Qx} - \mathbf{l}) &= \mathbf{b} \\ (\mathbf{P} - \mathbf{MQ})\mathbf{x} &= \mathbf{b} - \mathbf{M}\,\mathbf{l} \\ \mathbf{Ax} &= \mathbf{b}' \end{aligned} \tag{6}$$

where $\mathbf{A} = \mathbf{P} - \mathbf{MQ}$, and $\mathbf{b}' = \mathbf{b} - \mathbf{Ml}$ and **x** is the positive-fragment coefficients.

The $(N_g + 1) \times N_+$ **A** matrix can be decomposed via singular value decomposition (SVD) and the solution space obtained. We say "space" here, because very often the solution is not unique. Even if we apply the conditions that $x_i > 0$ for all elements of **x** and $y_i > 0$ for all elements of **y** there is still very often no unique solution. However, the "best" solution should be that solution which gives all positive fragments equal weight (if such a solution exists). The solution we choose is the minimal norm, which very often turns out to be a vector where all the elements of **x** are indeed 1. All other possible solutions can be obtained from the null space which is conveniently obtained from the SVD as those column vectors in the **V** matrix (the SVD of $\mathbf{A} = \mathbf{UWV}^t$ as the **W** is diagonal) that correspond to $w_{i,i} = 0$. Thus, the minimal norm solution plus *any* linear combination of the null space vectors also satisfies eq 6, although not necessarily always with the positive conditions described above.

In explicitly identifying the negative fragments as the overlap between positive fragments, the method described above is necessarily isodesmic in nature in that isodesmic reactions such as reaction 1 can be created from the parent molecule and the negative and positive fragments. Our method also ensures that

the maximum overlap between positive fragments is always preserved in the negative-fragment list.

2.2. Nonbonding Interactions. We deal with nonbonding interactions in a very simple manner, different from the way other workers deal with them. The nonbonding energy is significantly smaller than the bonding energy in the molecule. However, for generality, it is possible to consider nonbonding interactions in a hierarchy of levels as was done with the bonding interactions. We define a level 1 nonbonding treatment as one that involves forming the positive nonbonding fragments by pairing each group in the molecule with every other group in the molecule. If both of these groups appear in the same positive "bonding" fragment list then this pair is deleted from the positive nonbonding fragment list. Furthermore, if capping hydrogen atoms on different groups come within 2.2 Å of each other then this pair is also deleted from the positive nonbonding fragment list. Note that this never occurs if the level of fragmentation of the molecule is greater than 2. The negative nonbonding fragments are simply made up of capped groups from the parent molecule. The number of times each of these fragments (i.e., groups at level 1 nonbonding treatment) appears in the positive nonbonding fragment list, less 1, is the multiplier by which the energy of the negative nonbonding fragment must be multiplied and subtracted from the total energy of the positive nonbonding fragments.

Our studies thus far have shown that no higher level nonbonding treatment is warranted for obtaining the nonbonding energy. We also note that the procedure described above can generate a huge number of small fragments, something on the order of, but less than $N_+(N_+ - 1)/2$. Clearly cutoffs can be introduced to exclude the computation of negligible nonbonding interactions for groups spatially far from each other. However, in the present work, where nonbonding calculations were performed, we chose to include all possible interactions.

3. Examples

Gaussian 03[56] was used for all calculations involved in this work.

3.1. *n*-Hexane. We choose this molecule as it is the same as that chosen by Deev and Collins so that we can illustrate the ease of generating the fragments and that the final result in this simple case is identical. Each carbon atom and the valence bonded hydrogen atoms constitute a single group. We therefore represent hexane as 123456. At level 1 all the bonded pairs results in the positive-fragment list: 12, 23, 34, 45, 56. The negative fragments are immediately obvious as 2, 3, 4, and 5 as these groups have been counted twice. Thus, the energy of *n*-hexane is simply the sum of the positive fragment energies less the sum of the negative-fragment energies.

At level 2, we begin at nonterminal group 2 (starting at group 1 leads to a subfragment that is deleted) and connect all groups directly connected to it to form our first fragment, i.e., 123. We continue with groups 3, 4, and 5 to yield the positive fragments: 123, 234, 345, 456. The negative fragments are the overlap fragments, which are 23, 34, and 45. Fragments 3 and 4, generated from the overlap of 123 with 345 and 234 with 456 respectively are deleted because they are subfragments of 23 and 34 respectively which were generated by overlapping 123 with 234 and 234 with 345, respectively. The **P**, **M**, and **Q** matrices are then

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 2 & 2 & 1 \end{bmatrix}, \quad \mathbf{M} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 2 & 2 & 2 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

The **1** vector is simply a three dimension vector of ones, and the **b** vector is a seven dimension vector of ones except for the seventh element which has the value of 0. The final **A** matrix, **b′** vector, and solution vector are

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & -2 & -2 & -1 \end{bmatrix}, \quad \mathbf{b}' = \begin{bmatrix} 1 \\ 0 \\ -1 \\ -1 \\ 0 \\ 1 \\ -6 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

By inspection, the solution can be seen to be unique (as it is in the case of all parent molecules described by a linear chain of groups) and to be given by the **x** vector above. The **y** vector is then computed via eq 5 and found to be a three-dimensional vector of ones. Thus, the energy of *n*-hexane is simply the sum of the positive fragment energies less the sum of the negative fragment energies.

Level 3 is similar to the positive fragment set being 1234, 2345, and 3456 each generated by considering the all the nonterminal pairs 23, 34 and 45, respectively. The negative fragments generated by overlapping 1234 with 2345 and 3456 are 234, and 34, but 34 is deleted as it is a subfragment of 234. Likewise positive fragment 2345 generates the negative fragment 345.

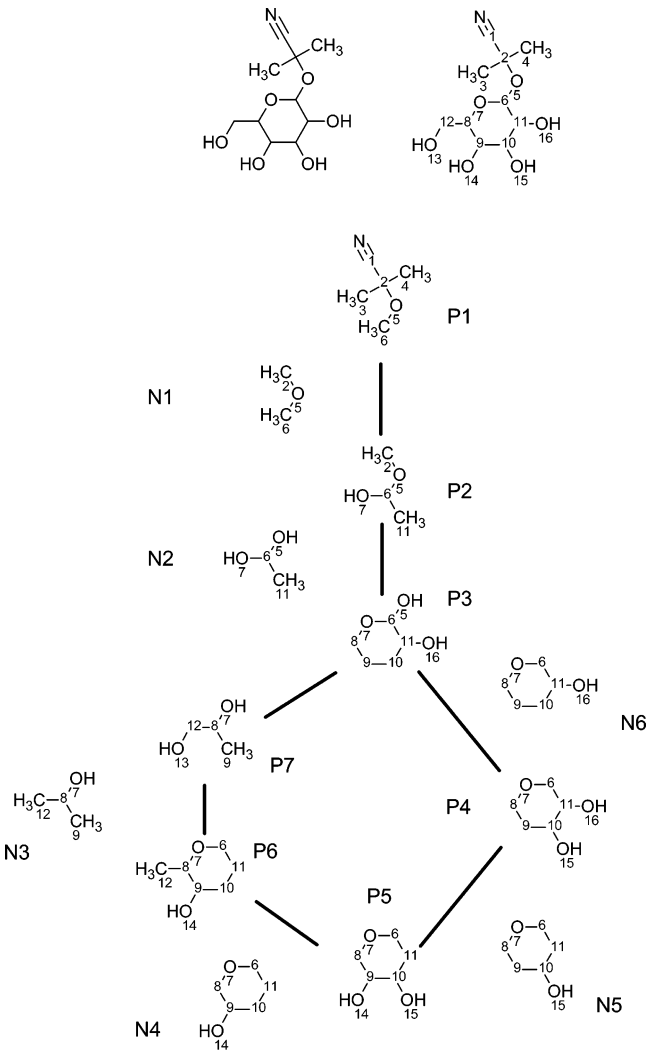
Setting up the **P**, **M**, and **Q** matrices and the **l** and **b** vectors are similar to before. Computing the **A** matrix and **b′** vector again reveal that the solution is a three-dimensional vector of ones and the resulting **y** vector a two-dimensional vector of ones. Thus, the energy of *n*-hexane is simply the sum of the positive fragment energies less the sum of the negative fragment energies.

Finally, at level 4, we have only two positive fragments 12345 and 23456, and the negative fragment 2345. Without any calculation we can immediately see that $E(n\text{-hexane}) = E(12345) + E(23456) - E(2345)$. Following the above procedure produces the same unique result. *n*-Hexane is too small to decompose at level 5.

3.2. Linamarin. Our next example is much more complex and illustrates capping growth during the fragmentation of a six-membered ring at level 3. The *O*-glycoside is illustrated in Scheme 1. This molecule contains 34 atoms of which 17 are heavies and contains 16 groups (the nitrile forms a single group). The arbitrary group numbering is also indicated in Scheme 1. For brevity we shall consider only the level-3 decomposition. Included in Scheme 1 are all the bonding positive and negative fragments that can be formed from linamarin, with the positive fragments connected by solid lines. The negative fragments from which the positive fragments are derived are indicated next to the solid lines connecting the two parent fragments.

The positive fragments are derived as follows. As discussed previously, the initial set of level-3 positive fragments are generated from the initial set of level-1 positive fragments. However, any positive fragments that are generated from level-1

SCHEME 1: Linamarin and Its Level-3 Decomposition (See Text)[a]



[a] Fragments labeled P*n* are positive fragments, while those labeled N*n* are negative fragments.

fragments that contain a terminal group will eventually be deleted as these fragments will be subfragments of other level-3 positive fragments. Thus, we will only consider those level-1 positive fragments that do not contain terminal groups.

The first such fragment is the fragment formed from groups 2 and 5 (designated [2,5]). Connecting all groups directly connected to these results in fragment P1 in Scheme 1. Fragment P2 is generated similarly from the level-1 fragment [5,6]. We now move around the glucose ring in a clockwise manner, and find the next level-1 fragment as [6,11]. Using our rule to generate a level-3 fragment we obtain [5,6,7,11,16]. However, the capping hydrogen atoms on groups 7 and 10 are too close to one another, resulting in capping growth, which eventually leads to the fragment P3. Similarly, consideration of the level-1 fragments [11,10], [10,9], and [9,8] lead to the level-3 fragments P4, P5, and P6, respectively. There are two more level-1 fragments in the glucose ring, and those are [8,7] and [7,6]. The fragments generated from these are [6,7,8,9,12] and [5,6,7,8,11] respectively. However, both of these fragments are subfragments of other positive fragments, namely P6 and P3 respectively, and so are deleted. There remains only one other non terminal level-1 fragment, and that is [8,12]. Directly

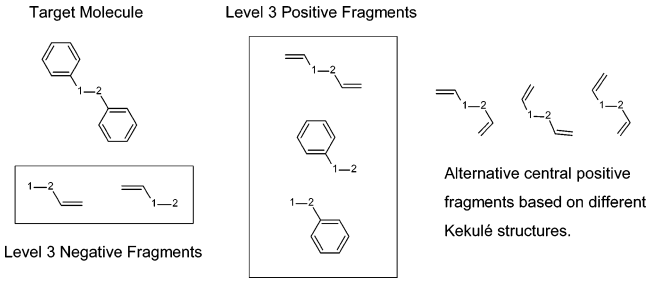


Figure 2. Level three fragmentation of 1,1′-ethane-1,2-diyldibenzene.

connecting all groups to these groups yields fragment P7, thus completing the set of positive bonding fragments generated at level 3.

The negative fragments are formed by overlapping each of the seven positive fragments with each other. Performing this operation results in the fragments shown next to the lines drawn connecting each of the positive fragments. Note that no negative fragment is drawn that overlaps P3 with P7, i.e., [7,8,9] as this overlap fragment is a subfragment of the overlap between P7 and P6, i.e. N3, and between P3 and P4, i.e., N6. Solving for the positive-fragment coefficients yields unity for P1−P7, which results in a value of −1 for all negative-fragment coefficients. This solution, however, is not unique. A single vector spans the null space and has two nonzero elements $1/\sqrt{2}$ P6 and $-1/\sqrt{2}$ P7.

3.3. 1,1′-Ethane-1,2-diyldibenzene. We consider the fragmentation of the title compound (Figure 2) as an illustration of how benzene-type systems are fragmented. As with all structures, formal bonding indicates the connectivity as well as which bonds can and cannot be broken. Thus, a Kekulé structure[57] is required for benzene. Each double bond is considered as a single group, so benzene is made up of three groups. Even at level 1, when pairs of groups are considered that both exist within benzene, capping growth ensures that the whole ring is part of the fragment. However, once the focus leaves groups within a benzene ring, defining benzene as three groups rather than one results in smaller fragments without sacrificing accuracy.

We illustrate the above point by examining the level three decomposition of 1,1′-ethane-1,2-diyldibenzene, given in Figure 2. The level-1 ethane fragment, given by the labeled groups "1" and "2", produces the first positive level-3 fragment given in Figure 2. The remaining two positive fragments are produced by considering all the remaining level-1 fragments of the title compound. Note, however, that any of the positive fragments on the right-hand side of Figure 2 would have been obtained if the alternate Kekulé structures had been drawn for both benzenes. The difference between the total energy at the B3LYP/6-31G(d) (using pure d basis functions) level of theory and the level-3 fragmentation energy for all four possible structures is less than 0.25 millihartree.

3.4 Alternative Fragmentation. As a simple example where our method of fragmentation differs from that of Deev and Collins we shall consider the level-3 fragmentation of a molecule with the morphology shown in Figure 3.

Table 1 provides the fragmentation using the Deev and Collins method and the method presented in this work. Clearly the two methods, for this morphology, give different fragments. At level-1 and -2 identical fragments and coefficients are obtained, but at level-3, the fragments given by the method described in this work are smaller. Here both methods give unique fragments.

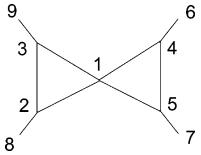


Figure 3. Example of the morphology to be fragmented under the Deev and Collins scheme and the scheme presented in this work. Groups are labeled with numbers.

TABLE 1: Deev and Collins (D&C) Fragmentation vs the Fragmentation Given in This Work for the Morphology Given in Figure 3

	D&C Level-3		level-3	
	coef	frag	coef	frag
1	1	1234567	2/3	12367
2	1	1234589	2/3	12389
3	−1	12345	2/3	123456
4			2/3	123457
5			2/3	123458
6			2/3	123459
7			-1/3	1238
8			-1/3	1239
9			-1/3	1456
10			-1/3	1457
11			−12/3	12345

4. Results and Discussion

The molecules chosen as test cases for our method are given in Figures 4 and 5. The selected species are representative of a wide variety of organic compounds containing C, N, O, H, P, S, and Cl. We also chose a range of structures and methods for testing our method with results essentially independent of level of theory and specific geometry as indicated in Table 2. Individual groups were defined as either single heavy atoms, or double or triple bonds. Exceptions to this choice are as follows. The four membered ring of taxol was chosen as a single group as was the $-NO_2$ of ranitidine and SPO_2 of VX gas. We always consider amide groups as single groups such as in folic acid, similarly with nitrogen atoms directly attached to benzene rings. In the latter case, a double bond in benzene together with the adjacent nitrogen atom is treated as a single group. Note that benzene rings are *not* considered as single groups, rather they are considered as being made up of three double bonds, i.e., three groups. Thus, some arbitrariness is associated with which possible Kekulé structure is chosen for benzene. However, once a specific formal structure is assigned to a molecule and the groups chosen, then the set of positive fragments generated at a given level are absolutely unique.

***cis*-3-Hexenal.** The smallest molecule studied in this work, and is made up of only five groups. Thus, it is not possible to fragment at level 4. It is clear from Table 1 that this system is already well described at level 2, which is likely due to the fact that it is a relatively small molecule.

VX Gas. It contains the third period elements, S and P, and is made up of 13 groups. Our relatively simple nonbonding treatment tends to overestimate the effects of nonbonding for this relatively large basis set, with agreement within 1 millihartree occurring only at level 4.

Octadecanonene, *β,β*-Carotene, and Vitamin A. The former is more properly named 1,3,5,7,9,11,13,15,17-octadecanonene is closely related to the central conjugated system of *β,β*-carotene. As is clearly shown in Table 2, these highly conjugated systems are well described (to within 1 millihartree) at level 4, and to within a few millihartrees at level 3. Note, however, that at level 4, up to five double bonds can be included in a positive fragment.

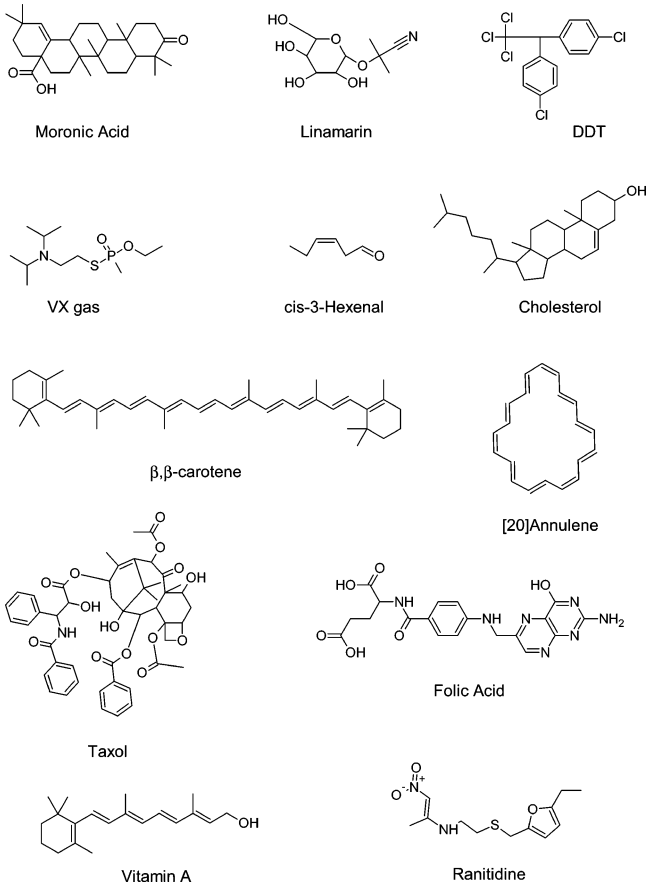


Figure 4. Some of the molecules studied in this work.

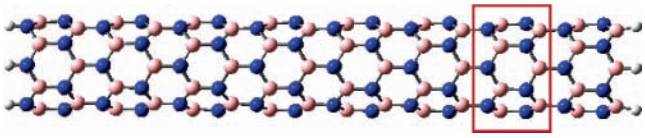


Figure 5. BN nanotube taken from the work of Li et al. and used in this work. The definition of a group is illustrated by the set of atoms contained within the red box.

[20]Annulene. We included this formally antiaromatic system here to illustrate that it may be possible to successfully fragment such systems. The level 3 results tend to suggest that fragmentation is possible, but the poorer level 4 results seem to suggest otherwise. We also tried fragmenting the aromatic [18]annulene, and found, not surprisingly, that fragmentation fails in this case. At present, at least for aromatic systems, accurate energies can only be obtained if at least one fragment contains the entire aromatic structure. Note, however, that when the focus of the fragmentation lies fully *outside* the aromatic system, fragmentation of the aromatic leads to accurate total energies as illustrated in Table 2 by DDT, taxol, and folic acid.

Ranitidine. The energy of this compound is already well reproduced at level 3, with nonbonding interactions playing very little role in the total energy of the system.

Moronic Acid. After including nonbonding interactions at level 3, which apparently are captured at level 4, agreement is well below the millihartree level.

Folic Acid. Agreement to about 1 millihartree is achieved at level 3 and beyond. For the configuration chosen, nonbonding plays little role.

Cholesterol. Nonbonding plays a slightly more significant role for this system, and its inclusion is necessary to improve

TABLE 2: Single Point Energies Less the Level *n* (L*n*) Fragmentation Energy or the Level *n* Fragmentation Energy Including a Nonbonding Correction (L*n* + NB)[a]

molecule (structure derived from) and method(s)	L1	L2	L3	L3 + NB	L4	L4 + NB
cis-3-hexenal (B3LYP/6-31G(d) 5d)						
HF/6-311G(d,p) 5d	−2.51	−0.44	−0.01	−0.05		
MP2/6-311G(d,p) 5d	−8.77	−0.67	−0.01	0.03		
MP3/6-311G(d,p) 5d	−6.61	−0.62	−0.01	0.03		
MP4(SDQ)/6-311G(d,p) 5d	−6.80	−0.61	−0.01	0.03		
QCISD(T)/6-311G(d,p) 5d	−7.67	−0.70	−0.02	0.03		
VX gas (B3LYP/6-31G(d) 5d)						
B3LYP/6-311+G(2d,p)	−35.51	8.53	2.40	−3.65	0.13	−0.15
octadecanonene (B3LYP/6-31G(d) 5d)						
B3LYP/6-311+G(2d,p)	−16.46	−5.40	−1.88	−1.84	−0.66	−0.61
linamarin (AM1)						
B3LYP/6-31G(d,p) 5d	−128.06	5.82	−2.61	1.25	0.36	0.77
vitamin A (AM1)						
B3LYP/6-311+G(2d,p)	−7.92	4.73	0.32	−0.18	0.28	0.18
DDT (B3LYP/6-31G(d) 5d)						
B3LYP/6-311+G(2d,p)	9.87	10.94	1.13	−3.82	1.18	0.90
ranitidine (B3LYP/6-31G(d) 5d)						
HF/6-31G(d,p) 5d	−33.06	5.71	1.02	0.51	1.12	0.79
MP2/6-31G(d,p) 5d	−56.67	−0.09	−0.18	0.55	0.44	0.42
moronic Acid (AM1)						
B3LYP/6-31G(d) 5d	−32.65	49.78	8.50	−0.64	−0.41	−0.30
folic acid (B3LYP/6-31G(d) 5d)						
B3LYP/6-31G(d) 5d	10.63	6.69	0.04	−1.09	−1.25	0.63
cholesterol (AM1)						
HF/6-31G(d)	−23.46	32.22	6.31	−2.37	−0.27	−0.13
MP2/6-31G(d)	−158.30	−0.61	1.61	0.88	−2.65	0.06
[20]annulene (B3LYP/6-31G(d) 5d)						
B3LYP/6-31G(d) 5d	−12.43	−4.67	0.46	0.38	2.23	2.21
β,β-carotene (B3LYP/6-31G(d) 5d)						
B3LYP/6-31G(d) 5d	−29.14	−0.05	−2.27	−2.12	−0.91	−0.73
Taxol (A) (B3LYP/6-31G(d) 5d)						
B3LYP/6-31G(d) 5d	−93.64	22.52	−0.43	19.81	−25.07	0.37
Taxol (B) (B3LYP/6-31G(d) 5d)						
B3LYP/6-31G(d) 5d	−93.19	24.45	5.61	14.92	−19.58	0.65
BN nanotube—zigzag (from ref 19)						
HF/3-21G	−13.31	−1.28	0.04	—	0.61	—

[a] Units are millihartree.

agreement at levels 3 and 4. Including nonbonding interactions at level 4 reduces the error to significantly less than 1 millihartree.

Taxol. Two possible structures were reported by Mastropaolo et al.[58] labeled "A" and "B". On the basis of the information provided in their publication, we optimized both structures at the B3LYP/6-31G(d) 5d level under the constraints given in Table 2 of Mastropaolo et al. It was noted that significant nonbonding interactions existed in the final optimized structures which included hydrogen bonding, and our results reflect this. The agreement at level 3 is fortuitous as the inclusion of nonbonding interactions leads to a significant error. At level 4, once nonbonding interactions are included excellent agreement is finally obtained for the total energy.

BN Nanotube—Zigzag Conformation. The structure used for this computation was from Li et al. However, the groups chosen in this system were six-membered ring collars circumventing the tube as illustrated in Figure 5. Thus, the tube is essentially an eight-group straight-chain system. With such a choice for the groups, the agreement with the total energy is excellent at and beyond level 3.

5. Conclusion

Utilizing isodesmic reactions, we have presented an automated method for the systematic hierarchical fragmentation of large molecules. We have shown for a wide variety of molecules that by computing the energies of the fragments and combining these energies linearly the total energy of the molecule can be computed with increasing accuracy as one moves up the hierarchy of fragmentation. The algorithm described here is similar in various aspects to a competing method described by Deev and Collins, yet differs fundamentally. The method described in our work provides, for a given classical chemical structure, fragments that are absolutely unique. We are also able to fragment a much wider variety of molecules than earlier work, and our algorithm is simple, efficient and easily implemented.

The benefit obtained in utilizing fragmentation in computing total energies (and dependent properties) of large molecules is substantial. Not only does CPU time now scale approximately linearly with molecular size for a wide variety of molecular systems, the method described here is directly amenable to parallelization, e.g., one fragment per CPU—for certain very large molecular systems technology no longer restricts the ability to compute accurate energies, but only computational expense.

There are some unsatisfactory aspects of the current method. Our 2.2 Å rule for capping growth requires a more sophisticated approach. Essentially, if capping growth leads to the cyclization of a fragment when the focus group (or groups) is within the ring that grows, then whenever *any* group (or groups) within the ring is the focus capping growth should occur—our 2.2 Å rule achieves this at present, but implementation of the algorithm just described is required. The treatment of nonbonding interactions is clearly overly simplistic, and it requires further improvement for accurate results in systems where nonbonding

plays a major role. No solution exists for the linear combination of energies of fragments of some highly connected molecules. There exists work-arounds when this situation arises, e.g., by introducing dummy atoms, or redefining what constitutes a group, but we have no "black-box" solution for this at present. We hope to address this in future work.

There are many other applications for which the fragmentation approach can be applied that have not been explicitly investigated in this work. Some examples include the fragmentation of transition metal complexes. Our preliminary studies have indicated that for certain metals, and with slight modification to the fragmentation scheme, transition metal complexes can be fragmented successfully. The use of symmetry coupled with fragmentation allows the calculation of essentially infinitely large systems. The BN nanotube given here is a specific example as the high symmetry requires the computation of only a few fragments from which a very large system can be readily constructed. Future work also needs to address the proper treatment of charges within molecules so that proteins and nucleic acids, and other highly charged species, can be automatically fragmented and have their energies (and dependent properties) accurately computed.

In summary this work and others reveals the potential of fragmentation to completely revolutionize computational chemistry for large systems, but much work is still needed to increase the number of systems directly amenable to the method.

Acknowledgment. We thank Professor M. A. Collins (RSC, Canberra, Australia) for his helpful discussions on this topic. We also thank the National University of Singapore, Faculty Research Grant for supporting this work.

References and Notes

(1) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085.
(2) Kohn, W. *Phys. Rev. Lett.* **1996**, *76*, 3168.
(3) Parr, R. G.; Yang, W. *Annu. Rev. Phys. Chem.* **1995**, *46*, 701.
(4) Greengard, L.; Rokhlin, V. *J. Comput. Phys.* **1987**, *73*, 325.
(5) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. *Chem. Phys. Lett.* **1994**, *230*, 8.
(6) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. *Chem. Phys. Lett.* **1996**, *253*, 268.
(7) Kutteh, R.; Aprà, E.; Nichols, J. *Chem. Phys. Lett.* **1995**, *238*, 173.
(8) Strain, M. C.; Scuseria, G. E.; Frisch, M. J. *Science* **1996**, *271*, 51.
(9) Challacombe, M.; Schwegler, E.; Almlöf, J. *J. Chem. Phys.* **1996**, *104*, 4685.
(10) Pérez-Jordá, J. M.; Yang, W. *J. Chem. Phys.* **1997**, *107*, 1218.
(11) Schwegler, E.; Challacombe, M. *J. Chem. Phys.* **1996**, *105*, 2726.
(12) Burant, J. C.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **1996**, *105*, 8969.
(13) Goedecker, S.; Colombo, L. *Phys. Rev. Lett.* **1994**, *73*, 122.
(14) Goedecker, S. *J. Comput. Phys.* **1995**, *118*, 261.
(15) Goedecker, S.; Teter, M. *Phys. Rev. B* **1995**, *51*, 9455.
(16) Yang, W. *Phys. Rev. Lett.* **1991**, *66*, 1438.
(17) Yang, W. *J. Chem. Phys.* **1991**, *94*, 1208.
(18) Lee, T.-S.; York, D. M.; Yang, W. *J. Chem. Phys.* **1996**, *105*, 2744.
(19) Zhao, Q. S.; Yang, W. *J. Chem. Phys.* **1995**, *102*, 9598.
(20) Li, X. P.; Nunes, W.; Vanderbilt, D. *Phys. Rev. B* **1993**, *47*, 10891.
(21) Daw, M. *Phys. Rev. B* **1993**, *47*, 10895.
(22) Millam, J. M.; Scuseria, G. E. *J. Chem. Phys.* **1997**, *106*, 5569.
(23) Li, X.; Millam, J. M.; Scuseria, G. E.; Frisch, M. J.; Schlegel, H. B. *J. Chem. Phys.* **2003**, *119*, 7651.
(24) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(25) Singh, U. C.; Kollman, P. *J. Comput. Chem.* **1986**, *7*, 718.
(26) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.
(27) Gao, J.; Xia, X. *Science* **1992**, *258*, 631.
(28) Stevensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357.
(29) Castillo, R.; Andres, J.; Moliner, V. *J. Am. Chem. Soc.* **1999**, *121*, 12140.
(30) Varnai, P.; Richards, W. G.; Lyne, P. D. *Proteins* **1999**, *37*, 218.
(31) Mullholland, A. J.; Lyne, P. D.; Karplus, M. *J. Am. Chem. Soc.* **2000**, *122*, 534.
(32) Ridder, L.; Mulholland, A. J.; Rietjens, I. M. C.; Vervoort, J. *J. Am. Chem. Soc.* **2000**, *122*, 8728.
(33) Exner, T. E.; Mezey, P. G. *J. Phys. Chem.* **2002**, *106*, 11791.
(34) Exner, T. E.; Mezey, P. G. *J. Comput. Chem.* **2003**, *24*, 1980.
(35) Exner, T. E.; Mezey, P. G. *J. Phys. Chem.* **2004**, *108*, 3599.
(36) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayashi, M. *Chem. Phys. Lett.* **1999**, *313*, 701.
(37) Kitaura, K.; Sugiki, S.-I.; Nakano, T.; Komeiji, Y.; Uebayashi, M. *Chem. Phys. Lett.* **2001**, *336*, 163.
(38) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *120*, 6832.
(39) Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett.* **2004**, *389*, 129.
(40) Imamura, A.; Aoki, Y.; Maekawa, K. *J. Chem. Phys.* **1991**, *95*, 5419.
(41) Korchowiec, J.; Gu, F. L.; Imamura, A.; Kirtman, B.; Aoki, Y. *Int. J. Quantum Chem.* **2005**, *102*, 785.
(42) Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. *J. Phys. Chem.* **1994**, *98*, 9165.
(43) Babu, K.; Gadre, S. R. *J. Comput. Chem.* **2003**, 484.
(44) Babu, K.; Ganesh, V.; Gadre, S. R.; Ghermani, N. E. *Theor. Chem. Acc.* **2004**, *111*, 255.
(45) Das, G. P.; Yeates, A. T.; Dudis, D. S. *Int. J. Quantum Chem.* **2004**, *120*, 6832.
(46) Huang, L.; Massa, L.; Karle, J. *Int. J. Quantum Chem.* **2005**, *103*, 808.
(47) Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2003**, *119*, 3599.
(48) Zhang, D. W.; Xiang, Y.; Zhang, J. Z. H. *J. Phys. Chem. B* **2003**, *107*, 12039.
(49) Gao, A. M.; Zhang, D. W.; Zhang, J. Z. H.; Zhang, Y. *Chem. Phys. Lett.* **2004**, *394*, 293.
(50) Mei, Y.; Zhang, D. W.; Zhang, J. Z. H. *J. Phys. Chem. A* **2005**, *109*, 2.
(51) He, X.; Zhang, J. Z. H. *J. Chem. Phys.* **2005**, *122*, 12039.
(52) Li, S.; Li, W.; Fang, T. *J. Am. Chem. Soc.* **2005**, *127*, 7215.
(53) Deev, V.; Collins, M. A. *J. Chem. Phys.* **2005**, *122*, 154102.
(54) Hehre, W. J.; Ditchfield, R.; Radom, L.; Pople, J. A. *J. Am. Chem. Soc.* **1970**, *92*, 4796.
(55) Benson, S. W.; Cohen, N. In *Computational Thermochemistry: Prediction and Estimation of Molecular Thermodynamics*; Irikura, K. K., Frurip, D. J., Eds.; American Chemical Society: Washington, DC, 1998; p 20.
(56) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision B.01. Gaussian, Inc.: Wallingford, CT, 2004.
(57) Kekulé, A. *Bull. Soc. Chim.* **1865**, *3*, 98.
(58) Mastropaolo, D.; Camerman, A.; Luo, Y.; Brayer, G. D.; Camerman, N. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6920.