Structural determination of large molecules through the reassembly of optimized fragments

Jung-Goo Lee [a,*], Yoon Sup Lee [b], Christopher Roland [a]

[a] Center for High Performance Simulations (CHiPS) and Department of Physics, North Carolina State University, Raleigh, NC 27695, USA
[b] Department of Chemistry and School of Molecular Science (BK21), KAIST, Deajeon 205-701, South Korea

ARTICLE INFO

ABSTRACT

The accurate determination of the optimized structures of large molecules is, computationally quite expensive. As an alternate to the conventional approaches to structural optimization, we have explored the accuracy and speed-up obtained when variants of the fragmentation optimization and recombination method (FORM) are used. Specifically, the method was applied to eight prototypical molecules – n-decane, hexa-alanine, a long conjugate hydrocarbon molecule, a large polar conjugated molecule, large (5,5) armchair single-walled carbon nanotubes, a salen–aluminum complex and a multiply H-bonded system (two conformers of vancomycin aglycon with Di-N-acetyl-L-Lys-D-Ala-D-Ala – without optimizing the structure of the whole molecules. We find that FORM can predict the structure of these molecules with an acceptable accuracy, all at a computational cost that is 2–11 times less than conventional quantum mechanical methods at the Hartree–Fock (HF), density functional theory (DFT) and MP2 level of accuracy. FORM may therefore represent a viable approach for the fast structural predictions of large molecules.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Ab initio quantum mechanics is perhaps the most significant computational tool available to today's theoretical chemists, being able to provide information as the structural and electronic properties of different molecular systems at a useful and desired accuracy. Unfortunately, accurate quantum chemistry simulations are computationally quite expensive, both in terms of CPU and memory requirements. Until recently, the application of *ab initio* methods has therefore been limited to relatively "small" molecular systems. Methods aimed at reducing the computational costs of such calculations are therefore of great interest to the theoretical community [1,2]. Typically, the computational cost for a conventional quantum calculation is $O(N^x)$, where N is a measure of the system size, such as the number of electrons or the number of basis functions, and the exponent x a measure of the computational cost associated with the method, e.g., $x \approx 2$–3 for both Hartree–Fock (HF) and density functional theory (DFT), and $x \approx 4$–5 for MP2 energy calculations. One appealing and intuitively obvious approach is to break up the molecule into smaller fragments. The fragments are then optimized separately, and are then subsequently reassembled for analysis. Ever since the 1990s, there has been considerable interest in exploring the ramifications of such "divide and conquer" (DC) methods for electronic structure calculations [3–21]. Indeed, it may be argued that this approach is the basis for many kinds of linear scaling methods—i.e., $O(N)$ methods. While the DC approach was initially applied to computing DFT-based energies in 1991 [3,4], it has since been used with HF MP2 [5,8,13–16,18,19] and semiempirical systems [9–13]. The DC method, where density matrix elements of a molecule are constructed from those of fragments [3–12,14], has previously been used to calculate the atomic charges (Coulson, Mulliken and Natural) [20], the electrostatic potentials [5,7,8], molecular electron densities [6,8,17] and NMR chemical shifts [21] of a variety of molecules, with results that are competitive with conventional *ab initio* methods. The DC approach has also been applied to solvated macromolecular systems at the semiempirical level in tandem with dielectric continuum model [22–25] or with discrete water model [26]. In a discrete approach, individual solvent molecules are treated explicitly.

In terms of the DC gradient methods, the final structure of a molecule is constructed from its fragments. The molecular structure obtained in such a way should be close to the real optimum structure, with an acceptable level of accuracy. Generally, this is difficult to achieve for the following obvious reasons. First, the potentials in the so-called "buffer" regions – which connect the

* Corresponding author.
    E-mail address: jglee@chips.ncsu.edu (J.-G. Lee).

different molecular fragments – are typically not accurate in DC approaches, especially when more complex molecules are involved. Second, small differences in bond lengths, bond angles and torsions may cause relatively large structural errors, particularly when linear polar molecules are involved. Finally, it is difficult to accurately include the non-bonded interactions between distant atoms without violating the spirit of the DC approach. This means, that there exist non-zero coupling terms between different internal coordinates on the force constant matrix of a molecule which are left out in some of the DC methods. Incomplete force constants or approximate forces can lead to an inaccurate molecular structure. Because of

these problems, the determination of large molecular structures is difficult, especially when compact systems are involved. As will be discussed, through a modification of existing DC methods, we have made progress, and partially overcome these issues. For instance, many DC methods use a very small number of atoms in a given fragment in order to maximize CPU savings. We find, however, that higher accuracy and flexibility may be achieved with the use of different kinds of fragmentation processes, while still obtaining reasonable CPU savings.

Focusing specifically on geometry optimization studies, we note that there have been several such investigations with DC
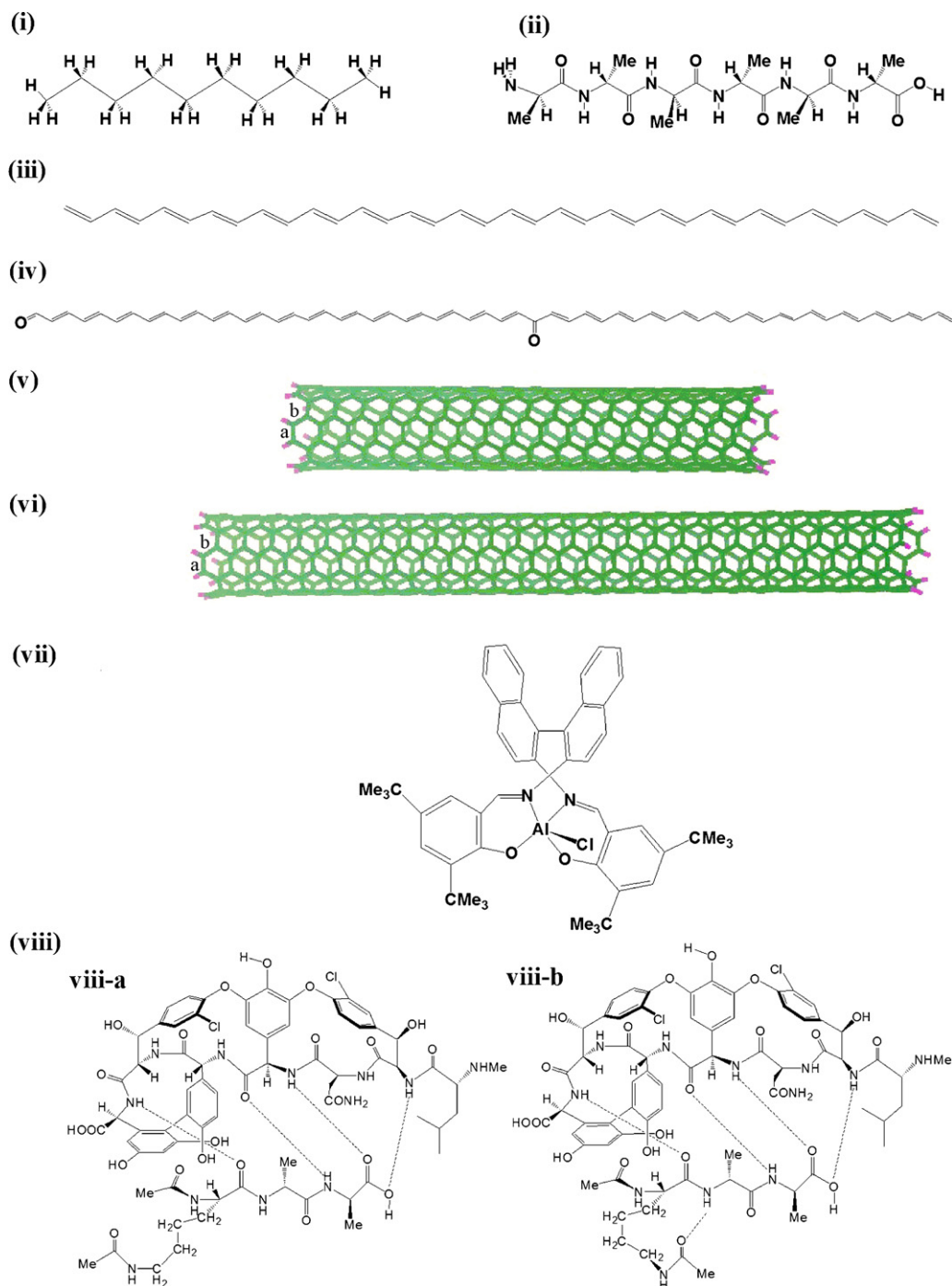

Fig. 1. The prototypical molecules investigated with the FORM method. (i) n-decane ($C_{10}H_{22}$); (ii) hexa-alanine ($C_{18}H_{32}N_6O_7$); (iii) conjugate hydrocarbon ($C_{36}H_{38}$); (iv) conjugated system with carbonyls ($C_{58}H_{58}O_2$); (v) SWCNT ($C_{340}H_{20}$); (vi) SWCNT ($C_{520}H_{20}$); (vii) aluminum–salen ($C_{50}H_{54}AlClN_2O_2$); (viii) VAT ($C_{69}H_{80}Cl_2N_{12}O_{23}$).

methods within a quantum mechanical framework. For instance, Zhao and Yang reported on a DFT-based optimal structure of the single-chain β-tetrapeptide glycine, with a total of 31 atoms using the DC analytical gradient method [27], with approximate Pulay forces [28,29]. In order to avoid using the approximated forces, the DC electronic structure computation is followed by the numerical gradient method [12,29]. There are also two recent DC studies on the structural optimization of both nonpolar and polar linear-chained systems (20 atoms or less) [30], and large one-dimensional systems (large oligomers of acetylene and BN nanotubes) [31]. Both studies have extended the work on molecular fractionation with conjugated caps [18] and total energies followed by analytical first derivatives [30,31] and the second derivatives [30] expressed in terms of its fractions. In each DC optimization cycle, the following steps are generally taken; (i) fragmentation of a molecule; (ii) energy and gradient calculations of fragments; (iii) recombination of the fragment forces according to the analytical gradient equation [28–31]; (iv) estimate of a new geometry. These steps (i–iv) are repeated many times, *i.e.*, until the optimum geometry is obtained. On the other hand, the optimization procedure we use is based on the following; (1) fragmentation of a molecule; (2) optimization of fragments; (3) recombination of the optimized fragments with the aid of computer graphics. We name this approach the FOR method or FORM. Therefore, FORM, is a modification of the DC gradient method. These steps (1–3) may be repeated a few more times for large molecules with different choice of fragmentation as described in detail in the next section.

There are other kinds of algorithms for the DC geometry optimization [32–34] where fragmented hybrid-delocalized [32] or localized-primitive [33,34] internal coordinates are used in the transformations between internal and Cartesian coordinates (and forces). There are some reports on linear scaling methods for geometry optimization without the DC approach, through the use of the fast transformations between redundant internal coordinates and Cartesian coordinates [35,36]. Also, there are several investigations of the geometry optimization algorithms using the geometry optimization using direct inversion in the iterative subspace (GDIIS) method [37–40] with screened Cholesky decomposition for the coordinate transformations [41]. These coordinate and force transformations also become the bottleneck in gradient calculations, especially, for large molecules while FORM does not require these transformations because of the direct construction of the Cartesian coordinates for a large molecule from its fragments with the aid of computer graphics. Transition state searches have been carried out by dividing a whole molecular system into reaction core part and its environment [32,42,43] as an extension of the DC geometry optimization for ground states.

As we show, the structural determination of large one-dimensional non-ionic molecules including large single-walled carbon nanotubes (SWCNT) is relatively straightforward. However, the structural optimization of large three-dimensional or compact molecules can be quite involved specifically because non-bonded interactions – which reach across the entire molecule – may be involved. Therefore, FORM for optimization of compact molecules with many cations and/or anions in gas phase may not work well [44]. In the following section, we illustrate how to determine the structures of a large compact system as well as a one-dimensional system by using FORM, and the structures are compared with those obtained by conventional HF, B3LYP and MP2 calculations [45].

2. Computational methodology

We illustrate the structural optimization with FORM with eight prototypical molecules: (i) *n*-decane ($C_{10}H_{22}$), (ii) hexa-alanine ($C_{18}H_{32}N_6O_7$), (iii) a long conjugate hydrocarbon ($C_{36}H_{38}$), (iv) a

long conjugate hydrocarbon with two carbonyl groups ($C_{58}H_{58}O_2$), two large (5,5) armchair single-walled carbon nanotubes [46,47], (v) $C_{340}H_{20}$, (vi) $C_{520}H_{20}$, (vii) 2,2'-bis(3,5-di-tert-butyl-2-hydroxybenzylideneamino)-1,1'binaphthyl aluminum chloride ($C_{50}H_{54}AlClN_2O_2$, a salen–aluminum complex) and (viii) two conformers ((viii-a) and (viii-b)) of vancomycin aglycon [48,49] with tripeptide, namely, Di-*N*-acetyl-L-Lys-D-Ala-D-Ala ($C_{69}H_{80}Cl_2N_{12}O_{23}$, in short, VAT). These are shown in Fig. 1. In order to effectively benchmark FORM, the optimized structures for these molecules were obtained with conventional HF/6-31G* calculations (molecules (i), (ii), (iv) and (vii)), B3LYP/4-31G (molecules (v) and (vi)), MP2/6-31G* (molecules (i), (iii) and (vii)) and HF/4-31G (conformers (viii-a) and (viii-b)). The initial geometries for these molecules and their fragments were obtained from a molecular mechanics force minimization using the Cerius2 package [50]. No symmetries for employed for molecules (i), (ii), (vii), (viii), while C_s symmetries were used for molecules (iii) and (iv). For molecules (v) and (vi), D_{5d} symmetries were used. All *ab initio* calculations were carried out with the Gaussian03 package [45]. In terms of the geometry optimization, there are a number of standard procedures used in *ab initio* gradient calculations such as quasi-Newton methods using approximate Hessians obtained from gradient information [51–53], conjugated gradient method [54] and geometry optimization using direct inversion in the iterative subspace methods. The GDIIS method uses information of forces and coordinates from previous optimization cycles, and was
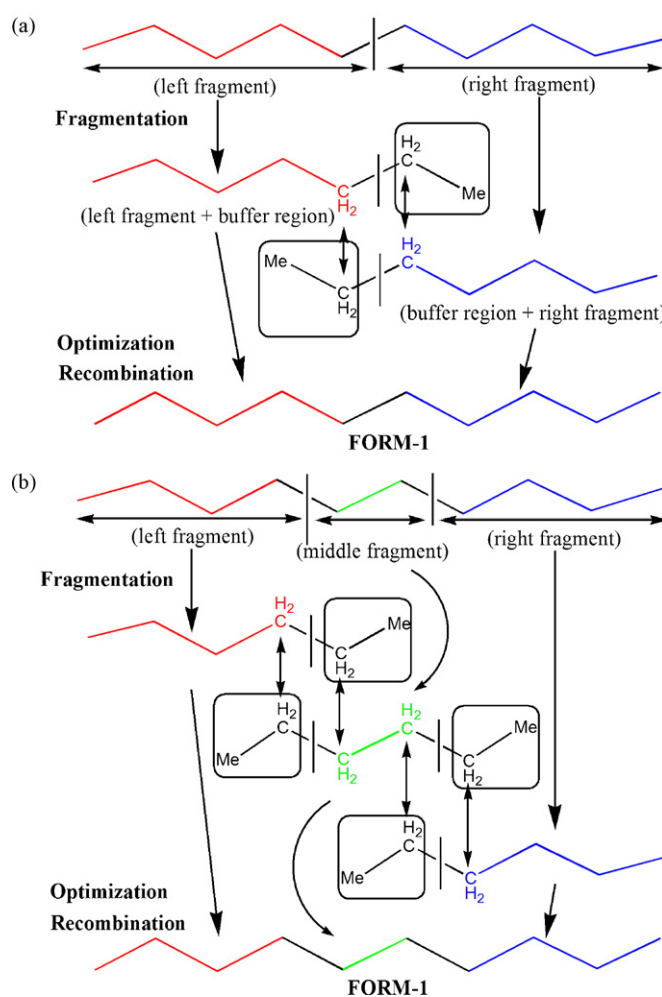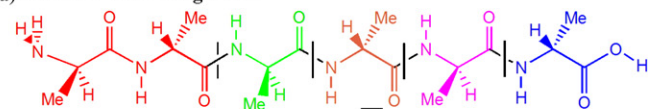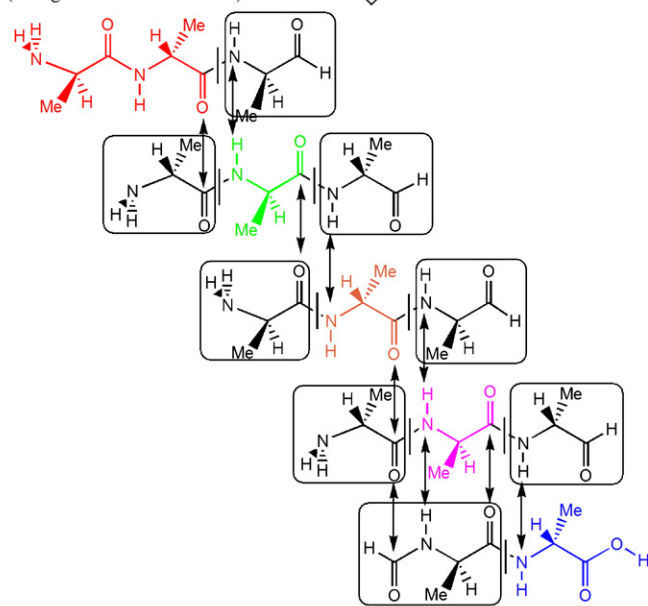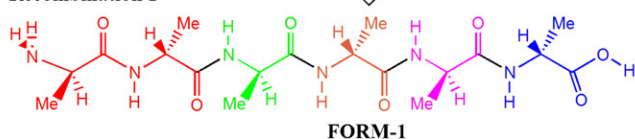

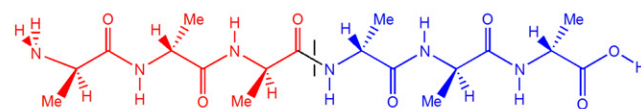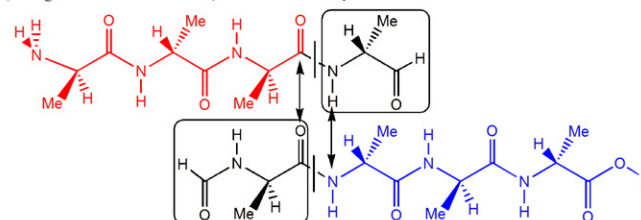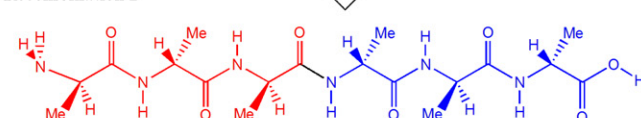**Fig. 2.** Details of the structural determination of *n*-decane via FORM using (a) two and (b) three fragments, respectively.
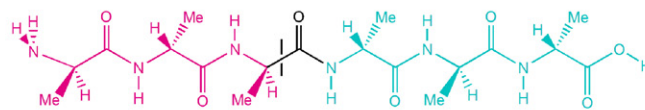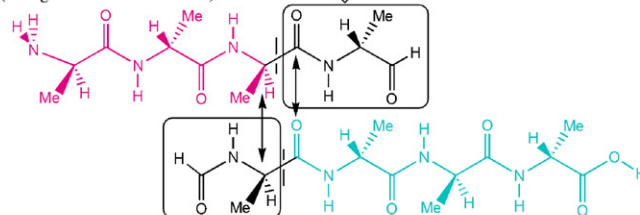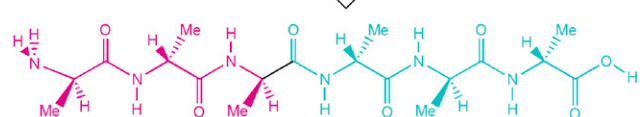
Fig. 3. Structural determination of hexa-alanine via FORM using (a) five, (b) two, and (c) two fragments, respectively. Note that the "cut-lines" in (b) and (c) are different.
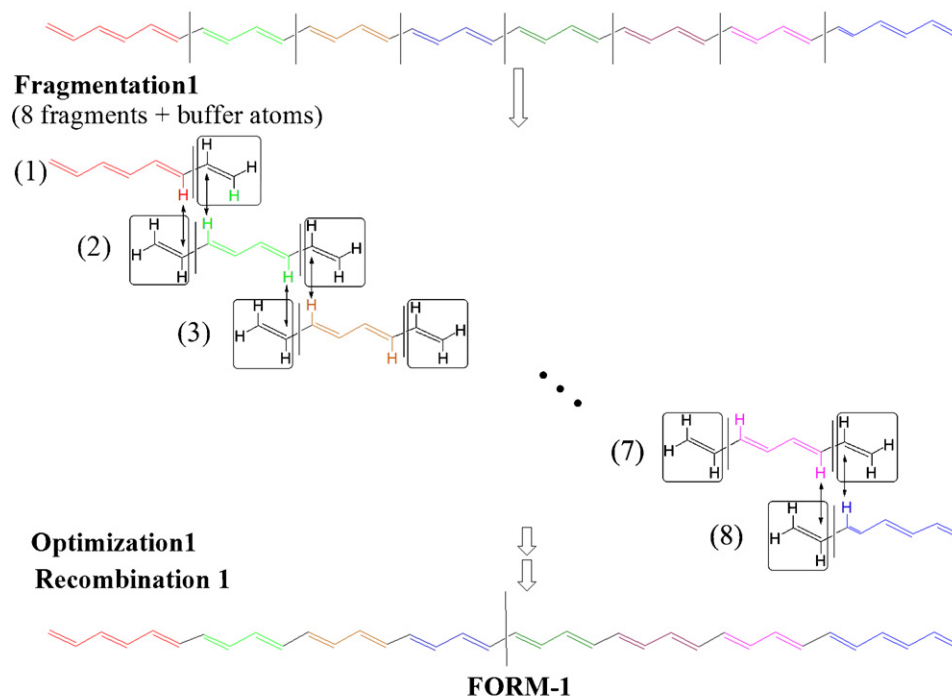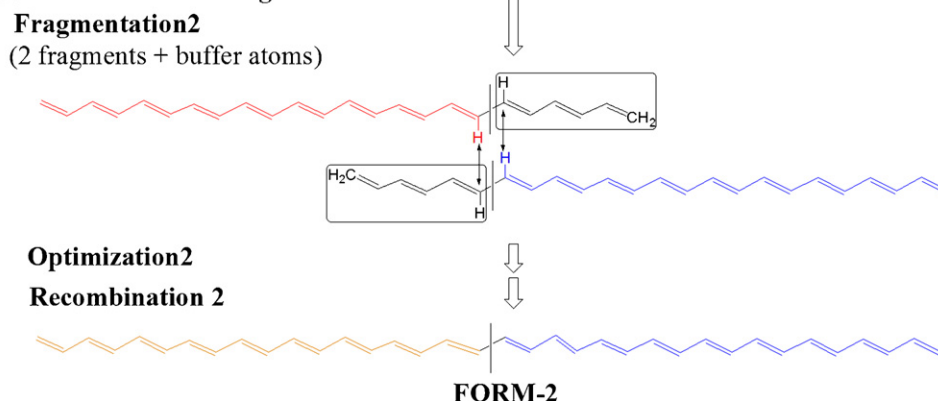
Fig. 4. Structural determination of $C_{36}H_{38}$ using FORM with (a) eight and (b) two fragments, respectively.

employed for all the optimizations of both molecules and fragments in this study.

The FORM approach consists of interaction of a basic fragmentation (F), optimization (O) and recombination (R) cycle. The details of effective fragmentation require some chemical intuition, and are best explained through the chosen examples. In terms of the number of iterations, we discuss results for one FOR-cycle for molecule (i), two FOR-cycles (FORFOR) for molecules (ii), (iii) and (iv), and three (for (v) and (vii)) to four (for (vi)) FOR-cycles, respectively. We first discuss in general terms each of the FORM steps.

Step I (fragmentation 1): The molecules are first divided into fragments, whose number ranges from 2 to 15, as shown in Figs. 2a, 3a and 4a for molecules (i), (ii) and (iii), respectively [55]. This is followed by the addition of buffer atoms, which are the atoms shown in the added rectangular (or oval) regions. For instance, Fig. 2 illustrates the fragmentation process for decane, which was divided into two (three) fragments. Buffer atoms added correspond the CH_2CH_3 units. Fig. 3a illustrates the case of hexa-alanine,

divided into five fragments. Note that the buffer atoms chosen now include several "neighboring" atoms, so that the immediate environment of the fragments mimics that of the complete molecule as much as possible. The conjugated molecules ((iii) and (iv)) were divided into eight (four) fragments, respectively, with up to two (four) neighboring heavy atoms included in the buffer region. The carbon nanotubes (molecules (v) and (vi)) were divided into 15 and 20 equal fragments. Finally, we explored the fragmentation of the salen–aluminum complex (or VAT conformers) into four different fragments [55].

Step 2 (optimization 1): All the fragments with the buffer atoms were optimized using the GDIIS optimizer.

Step 3 (recombination 1): The structures were then reconstituted from the optimized fragments. This was accomplished through the superposition of several atoms surrounding the cut lines of each of the fragment molecules, followed by the elimination of the overlapping atoms in the buffer regions.

We now examine in detail how FORM is carried out for the example molecules, starting with the case of *n*-decane. The vertical

Table 1
Comparison of the optimized structures for decane (molecule (i)) as obtained with FORM and the conventional HF/6-31G* and MP2/6-31G* calculations

	FORM-1 with two fragments	FORM-1 with three fragments	HF (MP2)
Total energy (a.u.)	−391.506562 (−392.820290)	−391.506560 (−392.820284)	−391.506565 (−392.820292)
Energy difference (kcal/mol)	0.0019 (0.0017)	0.0031 (0.0051)	0 (0)
Maximum force (a.u.)	0.000423 (0.000953)	0.000496 (0.000790)	<0.000450 (<0.000450)
RMS force (a.u.)	0.000086 (0.000091)	0.000111 (0.000165)	<0.000300 (<0.000300)
RMSD (Å)	0.0012 (0.0011)	0.0015 (0.0009)	0 (0)

MP2 numbers are quoted in parentheses.

Table 2
Comparison of the optimized structures for hexa-alanine $C_{18}H_{32}N_6O_7$ (molecule (ii)) as obtained with FORM and the conventional HF/6-31G* calculations

	FORM-1	FORM-2s	FORM-2d	HF
Total energy (a.u.)	−1551.137120	−1551.137316	−1551.137351	−1551.137374
Energy difference (kcal/mol)	0.159	0.036	0.014	0
Maximum force (a.u.)	0.001949	0.0011360	0.0011360	<0.000450
RMS force (a.u.)	0.000425	0.000219	0.000190	<0.000300
RMSD (Å)	0.0324	0.0136	0.0087	0

cut line in Fig. 2a divides the molecule into two equal left- and right-fragments, each consisting of C_5H_{11}. Buffer atoms, shown in the rectangular squares were then added. The natural choice here proved to be two CH_2 groups (i.e., CH_2CH_2). In addition, another H-atom was added to the endgroup in order to ensure charge neutrality. Ultimately, this ended up producing two identical (C_5H_{11})-CH_2CH_3 fragments, which were then optimized by standard means. In order to recombine the molecule, the two fragments are positioned such that the four carbons and eight hydrogens can overlap, as shown in Fig. 2a. After the superposition, the excess atoms from the left- and right-buffer regions of the fragments were simply eliminated. This procedure involves a single FORM cycle, and is hence simply termed FORM-1. Note that the reconstruction of the molecule is accomplished graphically, so that no coordinate transformations are needed. This significantly saves CPU time, especially, for large molecules. Breaking the molecule up into two fragments maximizes the size of the fragment molecules produced. We also tested the case of three fragments, which is illustrated in Fig. 2b.

While the FORM-1 procedure turns out to be adequate for a relatively simple molecule such as n-decane, more elaborate procedures are needed to control the errors associated with more complicated molecules. The key feature here is to repeat the FOR steps, using different fragments for each cycle. Specifically, choosing larger fragments enables one to at least partially include in the effects of the long-range interactions as illustrated. For molecules (ii), (iii) and (iv), the molecular geometry was constructed out of a two-FOR cycle (termed FORM-2). The procedure begins with a FORM-1 cycle in which the molecule is divided into a relatively large number of fragments, as illustrated in Figs. 3b and c, and 4b. Under FORM-2, the cycle is then repeated, except that a different fragmentation procedure – one that results in a relatively small number of larger fragments, i.e., two larger fragments – is used. Clearly, depending on the kinds of molecules under consideration, there are some ambiguity and latitude in constructing these fragments. For example, Fig. 3b and c illustrates two possible choices for hexa-alanine. For Fig. 3b, the cut line and buffer regions for left and right fragments are similar to the choice made for FORM-1. We name the FORM-2 recombined in this way as FORM-2s.

On the other hand, for Fig. 3c, the cut line and the buffer region of left fragment differ from the choice made from FORM-1 and FORM-2s, which is termed as FORM-2d. Fig. 4b illustrates a similar fragmentation for the two conjugate molecules. After this second

fragmentation of the molecules, the molecular fragments were optimized separately subject to the condition that the procedure was terminated once the maximum force became less than some predefined small number (0.001 a.u.). This has the effect of reducing the errors around the cut lines incurred during the first FOR cycle, and at least partially accounting for any long-range interaction. After this optimization step, the molecular fragments were recombined thereby completing the FORM-2 procedure.

For the single-walled carbon nanotube ($C_{340}H_{20}$) and the salen–aluminum complex, three FOR cycles (FORM-3) were needed to produce the fairly accurate structures. For the SWCNT (salen–aluminum complex), the method involves successive FOR steps involving 15 (4), 4 (3) and finally 2 (2) fragments. We note, however, that only a single additional gradient operation was needed during the last optimization step, since most of the internal forces already obtained were sufficiently small. Clearly, depending on the complexity of the molecule under consideration, the number of FOR cycles may be increased or modified in such a way, as to achieve better accuracy without significant increases in CPU time or memory requirements. A ground state geometry of SWCNT ($C_{520}H_{20}$) was obtained simply by recombining the two ($C_{340}H_{20}$)

Table 3
Comparison of the optimized structures for the conjugated system $C_{36}H_{38}$ (molecule (iii)) as obtained with FORM and the conventional MP2/6-31G* calculations

	FORM-1	FORM-2	MP2
Total energy (a.u.)	−1389.679474	−1389.681332	−1389.681374
Energy difference (kcal/mol)	1.192	0.026	0
Maximum force (a.u.)	0.007078	0.001667	<0.000450
RMS force (a.u.)	0.001856	0.000267	<0.000300
RMSD (Å)	0.01678	0.00378	0

Table 4
Comparison of the optimized structures for the conjugated system with carbonyls (molecule (iv)) as obtained with FORM and the conventional HF/6-31G* calculations

	FORM-1	FORM-2	MP2
Total energy (a.u.)	−2379.498443	−2379.498536	−2379.498539
Energy difference (kcal/mol)	0.0602	0.0019	0
Maximum force (a.u.)	0.003373	0.001038	<0.000450
RMS force (a.u.)	0.000370	0.000068	<0.000300
RMSD (Å)	0.0562	0.0373	0

Table 5
Comparison of the optimized bond structures for SWNT $C_{340}H_{20}$ (molecule (v)) as obtained with FORM and the conventional B3LYP/4-31G calculations

	FORM-1	FORM-2	FORM-3	B3LYP
Total energy (a.u.)	−12952.405158	−12952.462093	−12952.462321	−12952.462665
Energy difference (kcal/mol)	36.086	0.359	0.216	0
Maximum force (a.u.)	0.011690	0.001549	0.001373	<0.000450
RMS force (a.u.)	0.003009	0.000266	0.000193	<0.000300
RMSD (Å)	0.0646	0.0027	0.0015	0

fragments, where the fragment was computed from a single gradient calculation of FORM-3.

In the case of 2 VAT conformers ((viii-a) and (viii-b)), both thresholds of the maximum forces during the FORM-1, FORM-2 and FORM-3 steps were reduced to 0.00045 a.u. because the conformers have the H-bonds between vancomycin and ligand, and the floppy torsional angles on the L-Lys part of the tripeptide. Each conformer was divided to five fragments in the FORM-1 step. Two fragments were then taken from FORM-1 to FORM-2. In FORM-3, two large fragments were generated from FORM-2. Finally, the left fragment of FORM-3 and the FORM-3 structure (as right fragment) were taken to regenerate the last structure. Here, these last two fragments were not optimized but directly recombined at the different cut lines. The last structure is termed as FORM-3-FR because of the absence of the optimization step.

3. Results and discussion

Results are summarized in Tables 1–9 and Figs. 5–8, which give a detailed comparison between the structures obtained with the FORM methods, and the ones with the conventional quantum mechanics calculations. Generally speaking, the agreement between the FORM results and the regular calculations is quite good, showing that FORM may indeed prove to be a reasonably reliable method for the structural optimization of molecules. We note that all the energies quoted in these tables were obtained by conventional quantum mechanics means, in order to ensure their accuracy. Also, with standard benchmarking, we find that the conventional HF and B3LYP energy calculations scale with the number of basis functions N as $O(N^x)$, where $x \sim 2.1$–2.9, while the gradients scale with the number of atoms n as $O(n^y)$, where $y \sim 2$–3. For the MP2, this is somewhat higher with energies scaling as near $O(N^x)$ where $x \sim 4$ for molecule (i) and $x \sim 5$ for molecule (v).

Turning the results for each of the eight prototypical molecules, Table 1 and Fig. 5 summarize the results for the n-decane molecule as obtained with FORM-1 using either two or three fragments. Agreement between the FORM structures and the optimized HF/6-31G* structure is excellent, with only small deviations. For instance, the RMSD for the molecule is 0.0012 Å (0.0015) when two (three) fragments are used. As may be expected, the structural errors are small as well. For example, the largest error for a C–C bond length occurs around the "cut" lines. Deviations are 0.0007 Å (0.0011) when two (three) fragments are used. For the bond angles, the largest error (0.1°) also occurs around the cut-lines. The energy differences are also quite small, in the 10^{-3} kcal/mol range. The small RMSD implies that the maximum forces are also small, being either *less than* or *very close to* the default threshold value (0.000450 a.u.) of the Gaussian03 program. The MP2/6-31G* results showed similar trends in terms of the energies and structural characteristics. Note that both FORM and the conventional methods predict a final geometry with an C_{2h} symmetry, even though the initial geometry did not contain such symmetries. Overall, the FORM results with two fragments gave slightly better results that FORM with three fragments, indicating that the use of higher number of fragments potentially leads to a proliferation of

errors. In terms of the savings in CPU time, FORM with two (three) fragments was found to be 2.7 (3.6) times faster than the HF calculations, and 2.3 (4.2) times faster than the MP2 runs. These are reasonable savings, given the size of the molecule.

As expected, the structural errors incurred with FORM increase as the complexity of molecule increases. Table 2 presents results for the hexa-alanine complex at the HF level, with the molecule being first constructed out of five (for FORM-1) and then two (FORM-2) fragments. We note that there are two different structures (FORM-2s and FORM-2d) associated with FORM-2, as described previously. The FORM-1 results are close in energy, the error being 0.16 kcal/mol, while the maximum force (0.001949 a.u.) and the RMS force (0.000425 a.u.) are still relatively large. This results in a relatively large RMSD of 0.0324 Å. A closer examination of the bond lengths and bond angles shows that these structural differences are relatively small with differences of at most 0.003 Å for the bond lengths and 0.4° for the bond angles, respectively. Most of the errors occur around the bonds bisecting the "cut" lines of the fragments. In contrast, the torsional angles show relatively large errors (up to 1.4°). These errors occur at the 4 torsional angles bisecting the cut lines of the back bone atoms, and the four torsional angles $\tau(O=C-N-H$, with errors) ranging from 0.8° to 2.2° degrees. In FORM-2s and FORM-2d, the errors of the

Table 6
Comparison of the optimized bond lengths (Å) for SWNT $C_{340}H_{20}$ (molecule (v)) as obtained with FORM and the conventional B3LYP/4-31G calculations

No.	FORM-1: a (b)	FORM-2: a (b)	FORM3: a (b)	B3LYP: a (b)
1	1.383 (1.417)	1.374 (1.428)	1.374 (1.428)	1.374 (1.428)
2	1.428 (1.440)	1.438 (1.434)	1.438 (1.434)	1.438 (1.435)
3	1.428 (1.440)	1.437 (1.417)	1.436 (1.417)	1.436 (1.417)
4	1.428 (1.440)	1.423 (1.428)	1.423 (1.428)	1.424 (1.428)
5	1.428 (1.440)	1.436 (1.428)	1.436 (1.429)	1.435 (1.429)
6	1.428 (1.440)	1.435 (1.418)	1.436 (1.419)	1.436 (1.421)
7	1.428 (1.440)	1.423 (1.428)	1.424 (1.428)	1.426 (1.428)
8	1.428 (1.440)	1.436 (1.428)	1.436 (1.428)	1.436 (1.428)
9	1.428 (1.440)	1.435 (1.418)	1.435 (1.418)	1.435 (1.421)
10	1.428 (1.440)	1.423 (1.428)	1.423 (1.428)	1.426 (1.428)
11	1.428 (1.440)	1.436 (1.428)	1.436 (1.428)	1.435 (1.428)
12	1.428 (1.440)	1.435 (1.418)	1.436 (1.419)	1.435 (1.420)
13	1.428 (1.440)	1.423 (1.428)	1.424 (1.428)	1.426 (1.429)
14	1.428 (1.440)	1.436 (1.428)	1.436 (1.428)	1.436 (1.428)
15	1.428 (1.440)	1.435 (1.417)	1.435 (1.418)	1.436 (1.418)
16	1.428 (1.440)	1.416 (1.434)	1.417 (1.434)	1.418 (1.434)
17	1.469 (1.417)	1.454 (1.428)	1.453 (1.429)	1.453 (1.429)

Table 7
Comparison of the optimized structures for SWNT $C_{520}H_{20}$ (molecule (vi)) as obtained with FORM and the conventional B3LYP/4-31G calculations

	FORM-4	B3LYP
Total energy (a.u.)	−19803.407711	−19803.408021
Energy difference (kcal/mol)	0.019	0
Maximum force (a.u.)	0.001163	<0.000450
RMS force (a.u.)	0.000162	<0.000300
RMSD (Å)	0.0013	0

Table 8
Comparison of the optimized structures for the salen–aluminum complex $C_{50}H_{54}AlClN_2O_2$ (molecule (vii)) as obtained with FORM and the conventional HF/6-31G* and MP2/6-31G* calculations

	FORM-2	FORM-3	HF (MP2)
Total energy (a.u.)	−2885.158723	−2885.159516 (−2892.503909)	−2885.159719 (−2892.504139)
Energy difference (kcal/mol)	0.628	0.127 (0.144)	0 (0)
Maximum force (a.u.)	0.006330	0.001795 (0.001989)	<0.000450 (<0.000450)
RMS force (a.u.)	0.000806	0.000275 (0.000226)	<0.000300 (<0.000300)
RMSD (Å)	0.0791	0.0222 (0.0669)	0 (0)

MP2 numbers are quoted in parentheses.

Table 9
Comparison of the optimized structures for VAT $C_{69}H_{80}Cl_2N_{12}O_{23}$ (conformers (viii-a) and (viii-b)) as obtained with FORM and the conventional HF/4-31G calculations

	FORM-3-FR		HF	
	viii-a	viii-b	viii-a	viii-b
Total energy (a.u.)	−5945.544779	−5945.550954	−5945.546989	−5945.553251
Energy difference (kcal/mol)	1.39	1.44	0	0
Relative energy (kcal/mol)	3.87	0	3.93	0
Maximum force (a.u.)	0.011815	0.010571	<0.000450	<0.000450
RMS force (a.u.)	0.000712	0.000718	<0.000300	<0.000300
RMSD (Å)	0.0810	0.0836	0	0

bond lengths and angles are further reduced to the 0.002 Å and 0.2° level, respectively. The errors of the 2 torsional angles associated with the cut line (τ(CCNC) or τ(OCNH)) still persist, with an error of 1.6° or 1.4° for FORM-2s, while those of FORM-2d are significantly reduced (0.0° or 0.1°). These differences ensure that the maximum forces, RMS forces, RMSDs and energies of both FORM-2s and

Fig. 5. A comparison of the bond lengths (Å) and bond angles (°) for *n*-decane as obtained using FORM with: three fragments (no parenthesis); two fragments (in the round parenthesis); via conventional *ab initio* calculations (in the square parenthesis).
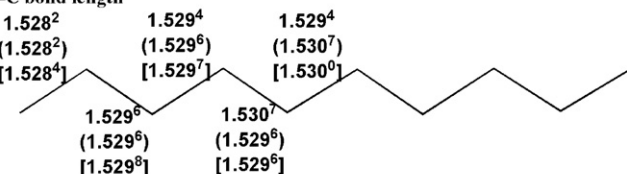
FORM-2d are significantly reduced over those of FORM-1. Moreover, the accuracy of FORM-2d is higher than that of FORM-2s. This is because FORM-2d can reduce the geometry errors than FORM-2s by choosing the cut different from in FORM-1. In terms of CPU timing, structural optimization with the FORM-2 approach is a factor of 3.9 times faster than with the conventional HF method.

Table 3 and Fig. 6 show results for conjugated molecule (iii) as obtained via FORMs. The molecule is initially constructed out of eight fragments (FORM-1), and then further refined using the FORM-2 procedure. We note that the FORM-1 results – when compared to the MP2/6-31G* structure – are characterized by relatively large errors in both the energy and the geometry. For FORM-1, the large errors in bond lengths occur not only at the cut lines (0.012–0.015 Å) but also at all bonds on the backbone atoms between second left C–C single bond and second right one (0.007–0.012 Å). As seen from the MP2 optimum geometry, there are distinct differences in the bond lengths between the C=C (and C–C) bonds in the middle of the chain and those at, or near, the terminal ends. The two terminal C=C bonds (1.3475 Å) are the shortest, and the C=C bond length gradually increases until both C=C bonds in the middle are 1.3670 Å. The opposite trend is observed for the C–C bonds. Specifically, the C–C bond decreases from 1.4476 Å (outside bonds) to 1.4304 Å (middle bond). When a large number of fragments are used (FORM-1), and the fragment molecule is small, this trend in bond lengths is replicated resulting in relatively large structural errors. For instance, a maximum force of 0.0071 a.u. is found at the middle C–C bond, a RMS force of 0.0019 a.u, an RMSD of 0.0167 Å and an energy error of 1.2 kcal/mol. On the other hand, the errors in the bond angles are relatively small (less than or equal to 0.2°). Therefore, the FORM-1 approach with relatively small fragments for this long conjugate system produces misleading results. When FORM-2 with a larger fragment molecule ($C_{24}H_{26}$) is carried out, the errors in geometry and energy are drastically reduced. This is because the larger fragment takes into account of the inside CC bond lengths to a certain extent. As a result, the errors of FORM-2 are 0.00378 Å for the RMSD and 0.026 kcal/mol for the total energy, which may be deemed acceptable. The CPU saving for molecules (iii) is a factor of 5.3 over conventional MP2 method.
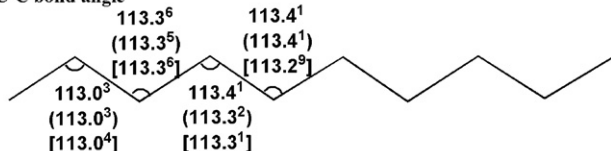
Similar trends are observed in the structural optimization of molecule (iv), with results given in Table 4. Since FORM-1 here involves four (as opposed to eight) fragments, the FORM-1 results
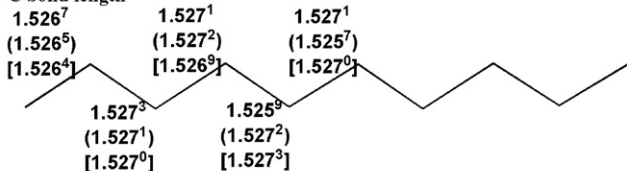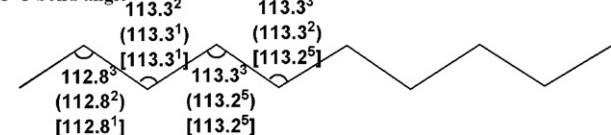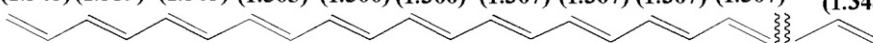
CC bond length
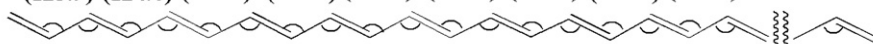
1.347	1.356	1.356	1.356	1.356	1.356	1.356	1.356	1.356	1.356		1.347
1.348	1.359	1.363	1.365	1.365	1.365	1.365	1.365	1.365	1.365		1.348
(1.348)	(1.359)	(1.363)	(1.365)	(1.366)	(1.366)	(1.367)	(1.367)	(1.367)	(1.367)		(1.348)

1.449	1.444	1.446	1.444	1.446	1.444	1.446	1.444	1.446	1.449
1.448	1.438	1.434	1.433	1.432	1.432	1.433	1.433	1.433	1.448
(1.448)	(1.438)	(1.434)	(1.432)	(1.431)	(1.431)	(1.431)	(1.430)	(1.430)	(1.448)

C-C=C bond angle

123.9	124.0	123.8	124.0	123.8	124.0	123.8	124.0	123.8
123.8	124.0	124.0	124.0	124.0	124.0	124.0	123.9	124.0
(123.9)	(124.0)	(124.0)	(124.0)	(124.0)	(124.0)	(123.9)	(123.9)	(124.0)

123.8	124.0	123.8	124.0	123.8	124.0	123.8	124.0	123.8		123.9
123.9	123.9	124.0	124.0	123.9	124.0	124.0	124.0	124.0		123.8
(123.8)	(123.9)	(124.0)	(124.0)	(124.0)	(123.9)	(123.9)	(124.0)	(124.0)		(123.9)

Fig. 6. Bond lengths (Å) and bond angles (°) of $C_{36}H_{38}$ obtained by FORM-1 (1st row), FORM-2 (2nd row) and conventional MP2 calculations (in parentheses).

are somewhat better than those obtained for molecule (iii). As for molecule (iii), FORM-1 results for molecule (iv) displays errors at the left and right cut lines. The errors in bond lengths bisecting these two cut lines are 0.0055 Å at the left cut and 0.0054 Å at the right cut, respectively. These errors are not as substantial as in the case of (iii). The FORM-1 bond angle errors are also small (less than or equal to 0.3°). As a result, the error of FORM-1 in energy is only 0.06 kcal/mol. The FORM-2 results reduce these geometrical errors significantly. In fact, in terms of the internal coordinates, the molecule is characterized by 117 bond lengths and 174 bond angles. Between these, only two forces (one 0.00104 and one 0.00046 a.u.) exceeded the default threshold force. In terms of the bond lengths, only three bonds localized in the buffer region differ by 0.001 Å from the optimum bond lengths, while the rest differ by less than 0.0005 Å. In terms of the angles, eight angles differ by less than 0.11° from their optimum angles, while the rest differ by less than 0.05°. FORM-2 reduced the structural RMSD to 0.0373 Å (from 0.0562 Å), and the energy difference to 0.0019 kcal/mol (from 0.0602 kcal/mol). Therefore, the polar aspect of this molecular system does not affect the accurate structural prediction of FORM, provided a proper fragmentation is chosen. On the other hand, Deev and Collins [30] reported rather large DC errors of energies (and presumably geometries as well) in cases where terminal groups have significant partial charges. The CPU saving for molecules (iv) is a factor of 3.9 over conventional HF method.

The results for SWCNT($C_{340}H_{20}$) are shown in Tables 5 and 6. There are two types of bonds for SWCNT; perpendicular to the tube axis (a) and partially along the axis (b) as shown in Fig. 1 (molecule (v)). Apparently, SWCNT constructed from 15 $C_{60}H_{20}$ fragments did not reproduce the optimum structures, very well. As listed in Table 6, not only FORM-1 predicted longer a (or shorter b) for the far left (No.1) and far right (No.17) CC bonds than the conventional B3LYP geometry, but also it showed 15 identical bonds of a (or b) inside the tube (Nos. 2–16). Therefore, the small fragment like $C_{60}H_{20}$ does not represent any part of the large SWCNT, which was also observed in the case of the long conjugated system ($C_{36}H_{38}$). On the other hand, larger fragments (FORM-2 and FORM-3) improved the errors much better than FORM-1, and both reproduced the optimum bond lengths of a and b within 0.003 Å. Also, both FORM-2 and FORM-3 predicted a periodicity of three sections inside the tube as the conventional B3LYP did [45]. The RMSDs and the errors of total energies are 0.065 Å

and 36.09 kcal/mol for FROM-1, 0.0027 Å and 0.36 kcal/mol for FORM-2, and 0.0015 Å and 0.22 kcal/mol for FORM-3, respectively. The cpu saving by FORM-3 is a factor of 4.7 times over conventional B3LYP method. As described earlier, a ground state geometry of the larger SWCNT ($C_{520}H_{20}$) was obtained from two ($C_{340}H_{20}$) fragments where the FORM-3 structure of ($C_{340}H_{20}$) was optimized by just one gradient cycle to compute the new geometry. Therefore, the geometry of ($C_{520}H_{20}$) obtained in this way is actually FORM-4. The trend of the bond lengths (a and b) for $C_{520}H_{20}$ turned out to be very similar to that for $C_{340}H_{20}$. The RMSD and the energy error of FORM-4 (Table 7) are 0.0013 Å and 0.19 kcal/mol, respectively, which is again in very good agreement with the conventional B3LYP method [56].

The results for salen–aluminum complex are summarized in Table 8 and Fig. 7. Generally speaking, the results obtained with three fragments and FORM-2 are less good than the results obtained with FORM-3 and two subsequent fragments. Energy differences between the FORM-2 and FORM-3 structures are 0.63 kcal/mol and 0.13 kcal/mol at the HF level, and 0.14 kcal/mol for FORM-3 at the MP2 level. Similarly, all the forces are lower for FORM-3 results, and all residual forces residing around the three cut lines for FORM-3. The bond lengths and bond angles obtained by both FORM-3 and conventional MP2 methods are shown in Fig. 7 where only the bond lengths and bond angles up to 3rd neighbors from the Al atom are shown. The largest error of FORM-3 in bond length (0.019 Å) is one of the Al–N bonds. This error along with other large errors would be reduced if more steps or better choices of buffer regions are considered. However, we have not pursued further elaboration of the method because our main purpose of the paper is to show the applicability of FORM to highly compact systems like this molecule (vii). Since the experimental X-ray structure for this salen–aluminum complex is available [57], a detailed comparison with our theoretical structures was feasible. The experimental structure appears to be slighter closer to the HF results, with RMSDs of 0.2804 Å for the HF and 0.2964 Å for the MP2 results, respectively. However, the experimental structure has another chiral form of (vii) and two $Me_2C(OH)_2$ molecules in its unit cell, in addition to the current conformation. We therefore carried out a more detailed comparison between of the partial complex near the Al center, where the effect of the other molecules should be minimized. The RMSD between the experimental and theoretical structures of the first nearest neighbors of the Al atoms
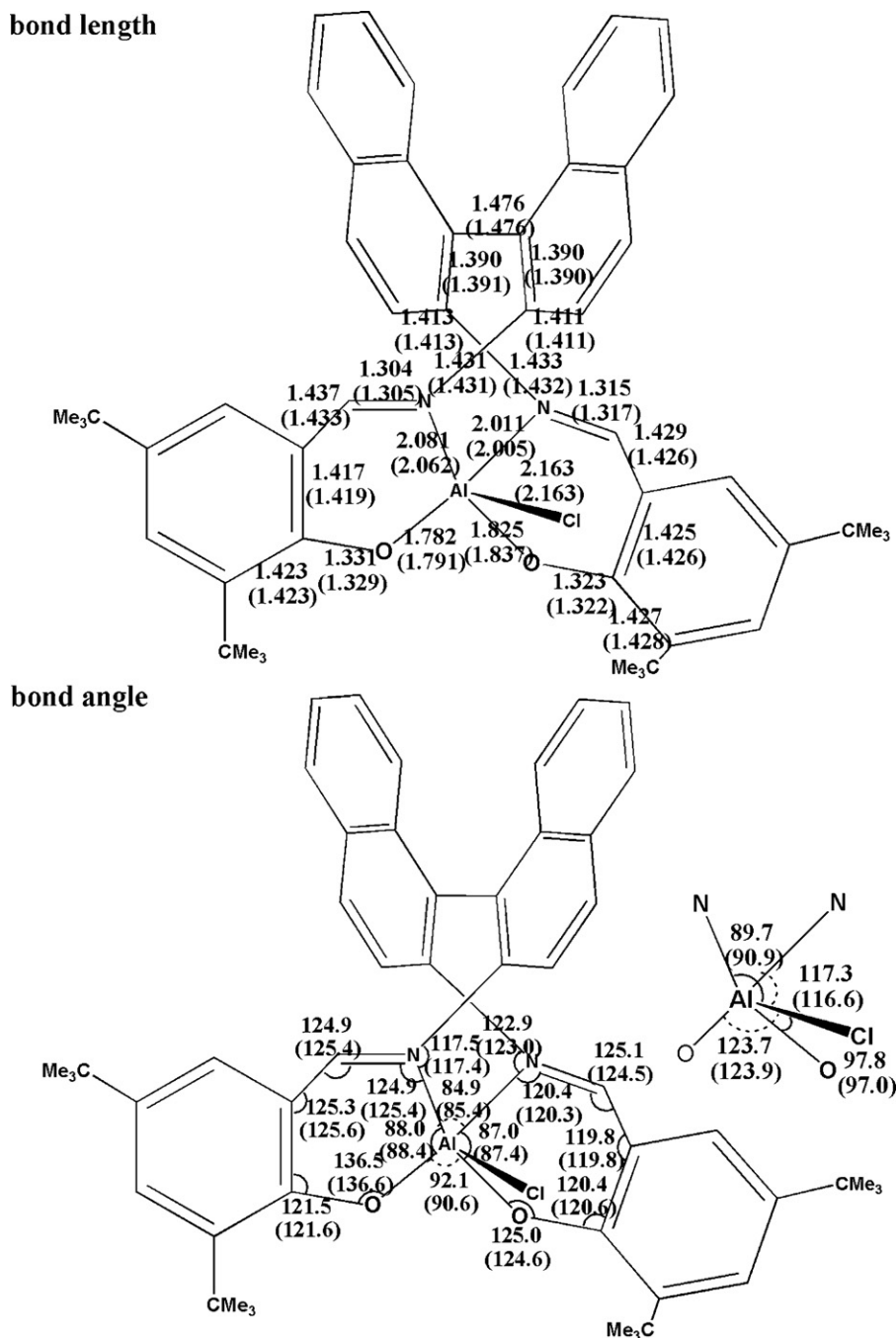
bond length

bond angle

Fig. 7. Bond lengths (Å) and bond angles (°) of $C_{50}H_{54}AlClN_2O_2$ obtained by FORM-3 (conventional MP2 calculations in parentheses).

was 0.0553 Å and 0.0473 Å for the respective HF and MP2 calculations. As far as the core of the structure is concerned, the MP2 results are closer to the experimental results. Finally, in terms of CPU time, FORM-3 was a factor of 2.0 times faster than the HF calculations, and 11 times faster than the MP2 runs. The time savings of FORM-3 over the MP2 runs is particularly significant. This is because the conventional MP2 calculations scale as near $O(N^5)$ for the compact system such as (vii).

Finally, the results for two conformers of VAT are shown in Table 9 and Fig. 8. The energy differences between FORM-3-FR and conventional methods turned out to be 1.39 kcal/mol for (viii-a) and 1.44 kcal/mol for (viii-b), respectively. For both conformers, the total energies by FORM-3-FR are lower by about 0.4 kcal/mol than those by FORM-3. Again, alternation of the cut lines improves

the errors in energies and geometries. Furthermore, the relative energy (3.87 kcal/mol) between these two conformers by FORM-3-FR is in an excellent agreement with that (3.93 kcal/mol) by the conventional method. All the H-bond lengths by FORM-3-FR and the conventional method are shown in Fig. 8. Note that the largest error among all valence and H-bonds comes from one of the H-bonds for each conformer. The RMS forces by FORM-3-FR are relatively small although the maximum forces are still large. The maximum force of each (viii-a) or (viii-b) occurred around the one of the bonds bisecting the cut lines. These maximum forces and RMD forces would be reduced by taking further optimization cycle with different cut lines. Overall, we obtained very accurate relative energy between two conformers of VAT and fairly accurate geometry for the H-bonded system. The conventional geometry
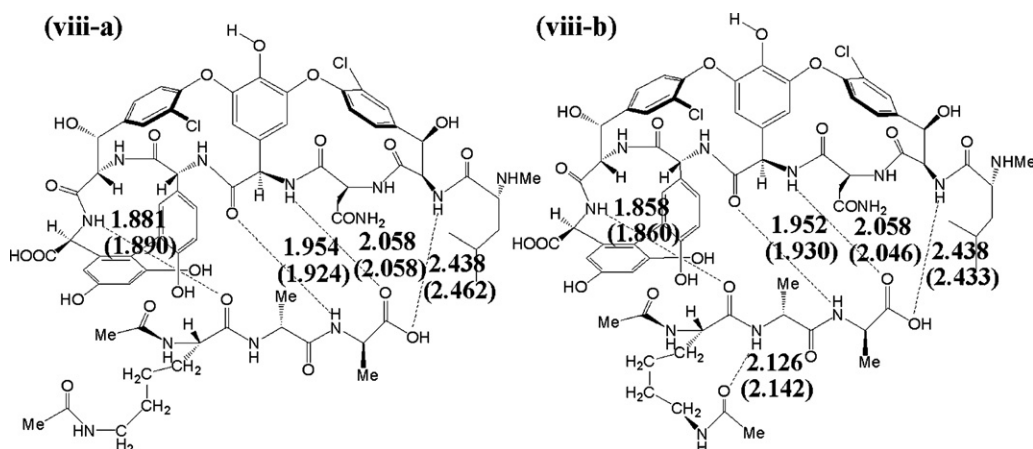
Fig. 8. Bond lengths (Å) of VAT ((viii-a) and (viii-b)) obtained by FORM-3-FR (conventional HF calculations in parentheses).

optimization method starting from the MM structure of (viii-a) did not lead to the structure of (viii-a) but to that of (viii-b) while the FORM approach starting from the same MM structure lead to the targeted local structure of (viii-a). This is another merit of the FORM approach. The cpu saving for (viii-b) at the HF level is a factor of 1.5 over the conventional method, and that at the MP2 level would be much more substantial as seen in the case of the salen–aluminum complex.

4. Conclusions

In summary, we have explored the structural optimization of eight prototypical molecules ranging from *n*-decane (32 atoms) to SWCNT (540 atoms) with variants of the fragmentation optimization and recombination method. Detailed comparisons between the FORM and conventional optimum structures (as obtained by HF, DFT and MP2) show good agreement between the two. Most importantly, the FORM structures were obtained with CPU savings of 2–11 times over the regular quantum chemistry methods. There are a number of possible extensions to the FORM approach, which we leave to the future. These include the following calculations:

(1) One electron properties such as ESP charges can be easily computed by DC density matrix elements, as several groups previously have investigated [58].
(2) Hessians can be constructed from fragments, which leads to computations of the vibrational frequencies [58]. These vibrational terms together with rotational and translational terms can be used to compute free energies.
(3) Transition state calculations can be done by FORM.
(4) Finally, structural determination of large molecules with discrete solvent systems can be performed by FORM [59].

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2008.06.004.

References

[1] S. Goedecker, Linear scaling electronic structure methods, Rev. Mod. Phys. 71 (1999) 1085–1123.
[2] S.Y. Wu, C.S. Jayanthi, Order-N methodologies and their applications, Phys. Rep. 358 (2002) 1–74.
[3] W. Yang, Direct calculation of electron density in density-functional theory, Phys. Rev. Lett. 66 (1991) 1438–1441.
[4] W. Yang, Direct calculation of electron density in density-functional theory: implementation for benzene and a tetrapeptide, Phys. Rev. A 44 (1991) 7823–7826.
[5] J.-G. Lee, R.A. Friesner, Atomic charges for large molecules derived from electrostatic potentials: fragment density matrix approach, J. Phys. Chem. 97 (1993) 3515–3519.
[6] P.D. Walker, P.G. Mezey, Molecular electron density lego approach to molecule building, J. Am. Chem. Soc. 115 (1993) 12423–12430.
[7] S.R. Gadre, R.N. Shirsat, A.C. Limaye, Molecular tailoring approach for simulation of electrostatic properties, J. Phys. Chem. 98 (1994) 9165–9169.
[8] K. Babu, S.R. Gadre, Ab initio quality one-electron properties of large molecules: development and testing of molecular tailoring approach, J. Comput. Chem. 24 (2003) 484–495.
[9] S.L. Dixon, K.M. Merz Jr., Semiempirical molecular orbital calculations with linear system size scaling, J. Chem. Phys. 104 (1996) 6643–6649.
[10] S.L. Dixon, K.M. Merz Jr., Fast accurate semiempirical molecular orbital calculations for macromolecules, J. Chem. Phys. 107 (1997) 879–893.
[11] A. van der Vaart, V. Gogonea, S.L. Dixon, K.M. Merz Jr., Linear scaling molecular orbital calculations of biological systems using the semiempirical divide and conquer method, J. Comput. Chem. 21 (2000) 1494–1504.
[12] A. van der Vaart, D. Suarez, K.M. Merz Jr., Critical assessment of the performance of the semiempirical divide and conquer method for single point calculations and geometry optimizations of large chemical systems, J. Chem. Phys. 113 (2000) 10512–10523.
[13] T.-S. Lee, D.M. York, W. Yang, Linear-scaling semiempirical quantum calculations for macromolecules, J. Chem. Phys. 105 (1996) 2744–2750.
[14] T. Akama, M. Kobayashi, H. Nakai, Implementation of divide-and-conquer method including Hartree–Fock exchange interaction, J. Comput. Chem. 28 (2007) 2003–2012.
[15] K. Kitaura, T. Sawai, T. Asada, T. Nakano, M. Uebayasi, Pair interaction molecular orbital method: an approximate computational method for molecular interactions, Chem. Phys. Lett. 312 (1999) 319–324.
[16] K. Fukuzawa, K. Kitaura, M. Uebayasi, K. Nakata, T. Kaminuma, T. Nakano, Ab initio quantum mechanical study of the binding energies of human estrogen receptor a with its ligands: an application of fragment molecular orbital method, J. Comput. Chem. 26 (2004) 1–10 (and references therein).
[17] R. Santamaria, J.A. Mondragon-Sanchez, M.A. Cunningham, Building wave functions for large molecules from their fragments, Phys. Rev. A 64 (2001) 1–8, 042501.
[18] D.W. Zhang, J.Z.H. Zhang, Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein–molecule interaction energy, J. Chem. Phys. 119 (2003) 3599–3605.
[19] R.P.A. Bettens, A.M. Lee, A new algorithm for molecular fragmentation in quantum chemical calculations, J. Phys. Chem. A 110 (2006) 8777–8785.
[20] J.-G. Lee, H.Y. Jeong, H. Lee, An efficient method to compute partial atomic charges of large molecules using reassociation of fragments, Bull. Korean Chem. Soc. 24 (2003) 369–376.
[21] B. Wang, E.N. Brothers, A. van der Vaart, K.M. Merz Jr., Fast semiempirical calculations for nuclear magnetic resonance chemical shifts: a divide-and-conquer approach, J. Chem. Phys. 120 (2004) 11392–11400.

[22] D.M. York, T.-S. Lee, W. Yang, Parameterization and efficient implementation of a solvent model for linear-scaling semiempirical quantum mechanical calculations of biological macromolecules, Chem. Phys. Lett. 263 (1996) 297–304.

[23] H. Liu, M. Elstner, E. Kaxiras, T. Frauenheim, J. Hermans, W. Yang, Quantum mechanics simulation of protein dynamics on long timescale, Proteins 44 (2001) 484–489.

[24] J. Khandogin, K. Musier-Forsyth, D.M. York, Insights into the regioselectivity and RNA-binding of nity of HIV-1 nucleocapsid protein from linear-scaling quantum methods, J. Mol. Biol. 330 (2003) 993–1004.

[25] J. Khandogin, D.M. York, Quantum descriptors for biological macromolecules from linear-scaling electronic structure methods, Proteins 56 (2004) 724–737.

[26] L.M. Westerhoff, K.M. Merz Jr., Quantum mechanical description of the interactions between DNA and water, J. Mol. Graphics Modell. 24 (2006) 440–455.

[27] Q. Zhao, W. Yang, Analytical energy gradients and geometry optimization in the divide-and-conquer method for large molecules, J. Chem. Phys. 102 (1995) 9598–9603.

[28] W. Yang, T.-S. Lee, A density-matrix divide-and-conquer approach for electronic structure calculations of large molecules, J. Chem. Phys. 103 (1995) 5674–5678.

[29] J.P. Lewis, S. Liu, T.-S. Lee, W. Yang, A linear-scaling quantum mechanical investigation of cytidine deaminase, J. Comput. Phys. 151 (1999) 242–263.

[30] V. Deev, M.A. Collins, Approximate ab initio energies by systematic molecular fragmentation, J. Chem. Phys. 122 (2005) 1–12, 154102.

[31] S. Li, W. Li, T.J. Fang, Am. Chem. Soc. 127 (2005) 7215–7226.

[32] S.R. Billeter, A.J. Turner, W. Thiel, Linear scaling geometry optimisation and transition state search hybrid delocalised internal coordinates, Phys. Chem. Chem. Phys. 2 (2000) 2177–2186.

[33] K. Németh, O. Coulaud, G. Monard, J.G. Ángyán, An efficient method for the coordinate transformation problem of massively three-dimensional networks, J. Chem. Phys. 114 (2001) 9747–9753.

[34] K. Németh, M. Challacombe, The quasi-independent curvilinear coordinate approximation for geometry optimization, J. Chem. Phys. 121 (2004) 2877–2885.

[35] B. Paizs, G. Fogarasi, P. Pulay, An efficient direct method for geometry optimization of large molecules in internal coordinates, J. Chem. Phys. 109 (1998) 6571–6576.

[36] B. Paizs, J. Baker, S. Sihai, P. Pulay, Geometry optimization of large biomolecules in redundant internal coordinates, J. Chem. Phys. 113 (2000) 6566–6572.

[37] P. Pulay, Convergence acceleration of iterative sequences. The case of scf iteration, Chem. Phys. Lett. 73 (1980) 393–398.

[38] P. Pulay, Improved SCF convergence acceleration, J. Comput. Chem. 3 (1982) 556–560.

[39] P. Csaszar, P. Pulay, Geometry optimization by direct inversion in the iterative subspace, J. Mol. Struct. 114 (1984) 31–34.

[40] O. Farkas, H.B. Schlegel, Methods for optimizing large molecules. Part III. An improved algorithm for geometry optimization using direct inversion in the iterative subspace (GDIIS), Phys. Chem. Chem. Phys. 4 (2002) 11–15.

[41] O. Farkas, H.B. Schlegel, Geometry optimization methods for modeling large molecules, Theor. Chem. 666–667 (2003) 31–39.

[42] Q. Zhao, J.B. Nicholas, Transition-state optimization using the divide-and-conquer method: reaction of trans-2-butene with HF, J. Chem. Phys. 104 (1996) 767–770.

[43] C. Lennartz, A. Schäfer, F. Terstegen, W. Thiel, Enzymatic reactions of triosephosphate isomerase: a theoretical calibration study, J. Phys. Chem. B 106 (2002) 1758–1767.

[44] Also, it may be difficult to obtain accurate molecular structures with soft (shallow) potential surfaces by FORM.

[45] M.J. Frisch, et al., Gaussian 03 Revision C. 02, Gaussian Inc., Wallingford, C.T, 2004

[46] Z. Zhou, M. Steigerwald, M. Hybertsen, L. Brus, R.A. Friesner, Electronic structure of tubular aromatic molecules derived from the metallic (5,5) armchair single wall carbon nanotube, J. Am. Chem. Soc. 126 (2004) 3597–3607.

[47] Z. Zhou, M.Y. Sfeir, L. Zhang, M.S. Hyberstsen, M. Steigerwald, L. Brus, Graphite, tubular pahs, and the diffuse interstellar bands, Astrophys. J. 638 (2005) L105–L108.

[48] J.-G. Lee, C. Sagui, C. Roland, First principles investigation of vancomycin and teicoplanin binding to bacterial cell wall termini, J. Am. Chem. Soc. 126 (2004) 8384–8385.

[49] J.-G. Lee, C. Sagui, C. Roland, Quantum simulations of the structure and binding of glycopeptide antibiotic aglycons to cell wall analogues, J. Phys. Chem. B 109 (2005) 20588–20596.

[50] Cerius2 4.8, User's Manual, Accelrys Inc., San Diego, CA, USA, 2003.

[51] H.B. Schlegel, Optimization of equilibrium geometries and transition structures, J. Comput. Chem. 3 (1982) 214–218.

[52] H.B. Schlegel, in: D.R. Yarkony (Ed.), Modern Electronic Structure Theory, World Scientific, Singapore, 1999.

[53] P. Pulay, G. Fogarasi, Geometry optimization in redundant internal coordinates, J. Chem. Phys. 96 (1992) 2856–2860.

[54] R. Fletcher, Practical Methods of Geometry Optimization, second edition, Wiley Interscience, Chichester, 1981.

[55] Fragmentation details for other molecules are shown in supplementary material.

[56] The geometry of C520H20 by the conventional B3LYP calculations started from that by FORM-4. Therefore, the cpu information is unavailable.

[57] CCDC 157312, Cambridge Crystallographic Data Centre.

[58] Preliminary calculation results of the ESP charges, the diagonal force constants of bond stretching and the vibrational frequencies for hexa-alanine by FORM/HF/6-31G* showed good agreement with those obtained by the conventional method (the numbers are listed in the supplementary material).

[59] However, application of FORM to certain implicit solvation models may not work well.