Ab Initio Quality One-Electron Properties of Large Molecules: Development and Testing of Molecular Tailoring Approach

K. BABU, SHRIDHAR R. GADRE
Department of Chemistry, University of Pune, Pune-411 007, India

Received 25 May 2002; Accepted 21 August 2002
Published online 12 February 2003 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/jcc.10206

Abstract: The development of a linear-scaling method, viz. "molecular tailoring approach" with an emphasis on accurate computation of one-electron properties of large molecules is reported. This method is based on fragmenting the reference macromolecule into a number of small, overlapping molecules of similar size. The density matrix (DM) of the parent molecule is synthesized from the individual fragment DMs, computed separately at the Hartree–Fock (HF) level, and is used for property evaluation. In effect, this method reduces the $O(N^3)$ scaling order within HF theory to an $n \cdot O(N'^3)$ one, where n is the number of fragments and N', the average number of basis functions in the fragment molecules. An algorithm and a program in FORTRAN 90 have been developed for an automated fragmentation of large molecular systems. One-electron properties such as the molecular electrostatic potential, molecular electron density along with their topography, as well as the dipole moment are computed using this approach for medium and large test chemical systems of varying nature (tocopherol, a model polypeptide and a silicious zeolite). The results are compared qualitatively and quantitatively with the corresponding *actual* ones for some cases. This method is also extended to obtain MP2 level DMs and electronic properties of large systems and found to be equally successful.

 J Comput Chem 24: 484–495, 2003

Key words: linear scaling methods; molecular tailoring approach; molecular orbital calculations on large molecules; electrostatic potential; electron density; topography

Introduction

The last 2 decades have witnessed an explosive growth in the development and applications of computational methods, in particular the *ab initio* ones, to problems of chemistry, physics, and biology. This is a result of the continuous enhancement in the computational power as well as development of faster and more reliable algorithms for performing Hartree–Fock (HF), post-HF, and density functional theory (DFT) calculations. However, all these *ab initio* methods suffer from a formidable scaling order,[1,2] which is practically $O(N^3)$ for the HF calculations on medium-sized molecules ($N^2 \log N$ for very large molecules) going up to $O(N^7)$ for correlated methods. Hence, practical applications of these *ab initio* methods are limited to systems with less than 100 atoms, at the HF level of theory. On the other hand, coupled-cluster and configuration interaction methods are able to treat systems containing utmost a few tens of atoms. This constraint has prompted theoreticians to look for alternatives that can circumvent such a high scaling order. One such effective alternative is offered by the linear scaling implementation of the *ab initio* methods.

Several linear scaling-type algorithms, wherein the computational cost scales linearly with the size of the system, are reported in the literature.[3–21,24,25,27] The most popular of these methods are the divide-and-conquer (DC) method due to Yang,[4] and the density matrix minimization (DMM) method proposed by Li et al.[5] The DC method was originally formulated for constructing the electron densities of macromolecules from the densities of smaller molecular mimics. This was later extended to density matrices by Yang and Lee.[6] The DC method involves partitioning the system into disjoint subsystems and defining a partition matrix for each of the subsystem. The electron density or density matrix of the subsystems are computed employing the corresponding local Hamiltonians. A common Fermi energy for the entire system is used to ensure normalized representation of the electrons. The error due to truncation at the boundary of the subsystems is systematically eliminated by introducing a buffer region of predefined radius that surrounds each subsystem. A method similar to this, termed the "molecular tailoring approach" was also developed independently by Gadre et al.[7] This approach involves construction of the density

Correspondence to: S. R. Gadre; e-mail: gadre@chem.unipune.ernet.in

Contract/grant sponsor: Council of Scientific and Industrial Research (CSIR), New Delhi

matrix (DM) of the supermolecule from block matrices of smaller molecular mimics, each representing a part of the supermolecule.

The DC methodology has been widely applied to large systems by several workers in recent years. Lewis et al.[8] have successfully employed the DC method for studying the geometry of the active site of an enzyme, cytidine deaminase. Merz and coworkers[9–11] have extensively tested their DC algorithm on a variety of applications. They have developed a semiempirical DC method[9] that is implemented in their program[10] DIVCON 99, based upon the frozen density matrix approximation put forth by Ermolaeva et al.[11] A solvation model for large molecules has also been developed by Gogonea et al.[12] by extending the use of DC for the large solute molecule while representing the solvent potential by the Poisson–Boltzmann technique. They have proposed an interaction-energy decomposition scheme for large molecules, and illustrated[13] its applicability for biomolecules and water clusters. Bliznyuk et al.[14] have reported a comparison of MESP inside the potassium ion channel of a protein, computed using a linear scaling semiempirical method and molecular mechanics methods.

Matta[15] has recently employed a DC type of algorithm to evaluate the electron densities and perform Bader's AIM[22] analysis. A complex oripavine molecule, PEO has been employed for testing the method at the HF/6-31G(d) level. Molecular properties such as the electronic charge, total energy, and the volume enclosed by density isosurface, reconstructed from the fragment molecules, are compared with those computed from the complete parent molecule. The results indicate a good agreement of the reconstructed properties with the *actual* ones (here and elsewhere in this article *actual* refers to the results obtained by performing an *ab initio* level calculation on the complete molecule). Khandogin et al.,[16] while reporting empirical linear scaling investigation on solvation of biomolecules and computing their electrostatics based descriptors, have emphasized the need for *ab initio* linear scaling methods that serve as a reliable tool for such studies.

Li et al.[5] proposed the DMM method, for bypassing the diagonalization step in the SCF procedure. The real space localization of DM is central to this approach. A typical DM element $P_{ij} \rightarrow 0$ as $R_{ij} \rightarrow \infty$, leading to a largely sparse DM. The electronic energy in terms of DM can be written as

$$E = Tr(PH) = \sum_{i,j} P_{ij}H_{ji} \tag{1}$$

This DM of the chemical system can be arrived at by minimizing the energy, E with respect to nonzero P_{ij}s subject to the normalization and idempotency constraints. The normalization constraint is achieved by introducing the chemical potential, μ as the Lagrange multiplier, while the McWeeney purification transformation[23] is employed to impose the idempotency constraint. The grand potential is then minimized using the equation

$$\Omega = Tr[(3\mathrm{P}^2 - 2\mathrm{P}^3)(H - \mu)] \tag{2}$$

The most time-consuming step in this process is the evaluation of P^2 that can be efficiently carried out by sparse matrix multiplication methods. Challacombe[17] has pointed out that the purification scheme in the case of a nonorthogonal basis can be represented as

$$\mathrm{P}' = 3\mathrm{PSP} - 2\mathrm{PSPSP} \tag{3}$$

The DMM method works best for insulators wherein the matrix elements P_{ij} die out rapidly with the increasing R_{ij}. Such a scheme was also independently proposed by Daw.[18] Challacombe[17] has reported a simplified DMM and benchmarked it on a series of water clusters and estane polymer observing a near-linear scaling behavior and reasonably accurate total energies. However, the DMM and other similar methods [cf. ref. 3(a)] have been developed and applied largely in the context of solid-state physics, except for those reported by Challacombe.[17]

Scuseria et al.[19,20] have applied the fast multipole methods (FMM) for linear scaling computation of the Coulomb integrals, a principal constituent of the bottleneck step in the HF and DFT calculations. This approach categorizes the Coulomb interactions into the far-field (FF) and near-field (NF) types. The NF Coulomb integrals are evaluated analytically, while the FF integrals are computed using multipole expansion methods. They have reported a near-linear scaling Coulomb integral evaluation using their Gaussian very fast multipole method (GvFMM) wherein the multipole expansion is maximally truncated to the desired accuracy. With this approach, they have achieved a scaling order of $N^{1.35}$ for a series of graphene sheets, while a conventional integral evaluation, for this system, is seen to scale as $N^{2.11}$. Scuseria et al.[20] have recently discussed linear scaling computation of the three bottleneck steps involved in the DFT calculations, the evaluation of Coulomb integrals, computing the exchange-correlation part and diagonalization of the Hamiltonian. They have advocated the use of GvFMM for the Coulomb problem and a conjugate gradient density matrix search method based on the Lee–Nunes–Vanderbilt[5] DMM method to obtain the accurate DM, instead of diagonalizing the Hamiltonian. The exchange-correlation quadrature is obtained using the LinXC[21] scheme proposed by them earlier. They have utilized this for computing energies of a series of two- and three-dimensional water clusters (72 to 1152 water molecules) as well as a 1026-atom RNA segment.

Mezey et al.[24,25] have developed and widely applied a method to compute *ab initio* quality MEDs of proteins based on the MEDLA (molecular electron density lego assembler) database. The MEDLA database consists of electron densities of a collection of very small, commonly occurring molecular fragments. Twenty-one of such small fragments, rather atom types or groups, have been defined for the case of proteins and polypeptides. In this approach the protein/macromolecule is divided into a number of small, mutually exclusive fragments. The closest match of each of the fragment, to the unit in the MEDLA database is identified. The density of the macromolecule is then constructed as a sum of fragment densities from the MEDLA database using the following procedure

$$\rho(\mathbf{r}) = \sum_{k=1}^{n} \rho^k(\mathbf{r}) \tag{4}$$

where

$$\rho^k(\mathbf{r}) = \sum_{k=1}^{n} P_{ij}^k \varphi_i(\mathbf{r}) \varphi_j(\mathbf{r}) \quad (5)$$

and $P_{ij}^k = P_{ij}$ if centers i and j belong to same fragment, $= 0.5*P_{ij}$ if only one of i or j belongs to the fragment, and $= 0$ otherwise.

It has been shown qualitatively and quantitatively that the MEDs constructed from MEDLA database, for small-sized test cases, match reasonably well with the corresponding HF/6-31G(d,p) values obtained by a calculation on the whole system. MEDs of various proteins computed by this approach have been reported, with the largest molecule contaning ~1400 atoms. The numerical similarity index [cf. eq. (6)] proposed by Richards et al.[26] has also been reported for these test cases. This index serves as a quantitative comparison of the scalar field computed over a three-dimensional grid encompassing the molecule from two data sets X and Y.

$$S_{XY}(f) = B_{XY}(f)/[T_X(f) * T_Y(f)]^{1/2} \quad (6)$$

where $T_X(f)$ and $T_Y(f)$ are the number of values in data X and Y, respectively, that fall above "f," and B_{XY} is the number of points having a value above "f" in both X and Y. They have reported the $S_{XY}(0.001)$ to lie between 0.98 to 0.99 for three test cases. In addition, a point by point comparison of the values using a newly defined index $L(a,a',X,Y)$ has also been provided.[24]

The MEDLA database may have diverse geometries of a given fragment. It should be noted however, that the best match in the database is chosen for every fragment still ignoring minor geometric differences between them. If the fragment is drastically different from the corresponding match in the MEDLA database the density of the fragment is recalculated. This serves as a major advantage of this approach, that is, it requires minimal computation for obtaining MED of any macromolecule. In contrast, the fragments used by Mezey et al.[24] are not large enough and do not take cognizance of the complete local environment, particularly the long range (through bond and through space) interactions, specific to the system under consideration. Further, error in obtaining other electronic properties such as the MESP, which are weighted by some factors, may be expected to be larger. Mezey et al.[25] have also reported an extension of this approach incorporating database of fragment DMs. This scheme has been referred to as adjustable density matrix assembler (ADMA). Recently, Mezey et al.[27] have reported the computation of MESP for polypeptide and other macromolecules using this approach.

The molecular tailoring approach, which is in principle similar to the Yang's DC method, has been independently developed earlier by Gadre et al.[7] Within this scheme, the test molecule is scissored into small overlapping fragment molecules, SCF calculations performed on all these fragments, and the DM of the parent molecule stitched together from the fragment molecule DMs. This synthesized DM was seen to be almost idempotent and the number of electrons and total electronic energies produced from it were in good agreement with the corresponding *actual* values. The order of error in the "synthesized" molecular electrostatic potential (MESP) was seen to be about 2%. However, it may be noted that this method, as implemented earlier, required manual creation of the fragments, which is possible only for small and simple molecules. Hence, a need was felt for developing an algorithm for appropriately fragmenting a large molecule in an automated manner and enabling a solution of problems incorporating larger basis sets.

As seen above, the emphasis of the work on linear scaling methods, reported so far, has been the geometries and energies of large systems, especially solids. With a few exceptions, such as those reported by Matta,[15] Challacombe,[17] and Mezey,[24,25,27] not much work has been reported on the molecular properties. Further, a substantial work, reported in the literature, has been carried out with semiempirical procedures. A survey of the recent literature would reveal the need for a practical method to perform *ab initio* investigations on large molecular systems. This is felt necessary, because molecular electron density (MED) and MESP-based descriptors are being widely used in QSAR and other biological applications. Keeping this in view, it is felt worthwhile to appraise the accuracy of one-electron properties generated by the molecular tailoring approach at a reliable *ab initio* level.

In this work, the development of an efficient algorithm for fragmenting the given supermolecule with a potential applicability to systems containing ~1000 atoms is reported. We present test calculations at the HF and MP2 levels using STO-3G and 6-31G(d,p) basis functions. One-electron properties such as the MESP, MED, topography of these two scalar fields, and the dipole moment have been evaluated by this method for a given molecular geometry. Further, a critical comparison of these results with *actual ab initio* counterparts is carried out wherever possible.

Methodology

The basic outline of the molecular tailoring approach remains similar to that proposed in the earlier[7] article. However, in its newer form, applicable to large molecular systems, the program has been further refined, generalized, and systematized. It involves three steps: (1) generating the fragments from the parent molecule, (2) obtaining density matrices of the fragments by performing *ab initio* calculations at an appropriate level of theory, and (3) construction of the super-DM from the constituent fragment DMs.

A program in FORTRAN 90 has been developed to cut the large molecule into a number of small overlapping fragment molecules. The algorithm of the program is given in Figure 1. The program consists of three parts: the preprocessing, fragmentation, and postprocessing. The preprocessing involves generation of the molecular graph and identification of sensitive regions viz. π-bonds, aromatic rings, and functional groups (which need to be kept intact). Such regions are replaced by pseudoatoms, and the molecular graph is regenerated with these pseudoatoms incorporated. This process is illustrated in the schematic of the complete fragmentation procedure on a test molecule depicted in Figure 2. The labels "A" and "A′" shown in structure II are the pseudoatoms that replace the encircled regions in structure I. The numbers in braces along with pseudo atoms indicate the number of atoms they represent.

In the fragmentation process, the molecular graph is subjected to repeated shrinking and expansion processes. The molecular graph is shrunk starting from every pendant (terminal, hanging) atom and traveling inward simultaneously. In the process, the bulk

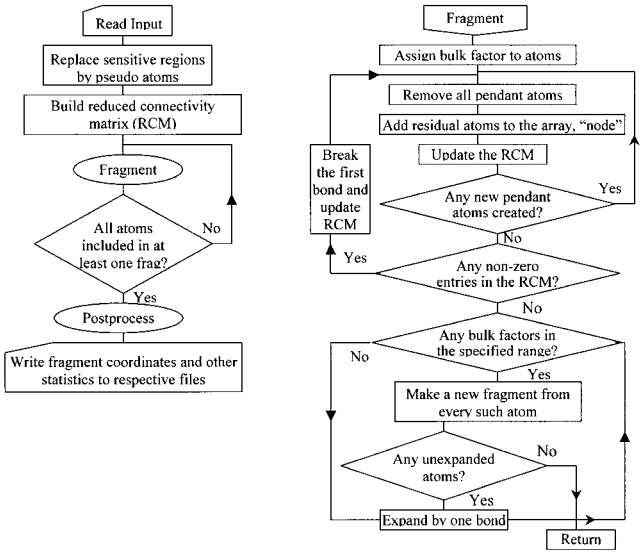

Figure 1. Flow chart depicting an overview of the program (left) and a detailed description of the fragmentation procedure (right). The RCM stores, for every atom, the number of atoms bonded to it and a list of those bonded atoms.

factor of each atom, that is, the number of atoms that are shrunk into that atom, is tracked. If there are no pendant atoms left at an intermediate stage (implying the presence of a closed ring-like structure), the first surviving bond in the array is broken and the shrinking continued. The shrinking process finally leads to one or more isolated nutshell atoms. This process is shown in Figure 2 going from structure II to structure III with the numbers at the atom positions indicating the respective bulk factors. A progressive expansion process is then effected from each of the nutshell atoms by a simultaneous expansion along all the branches to a length of one bond at a time. At every stage, the terminal atoms having a bulk factor within the (fragment) size range, specified in the input, are identified and are severed from their anchoring atoms. Each such atom, along with all those atoms shrunk into it, forms a separate fragment molecule. In Figure 2, the labels F001, F002, and F003 indicate three mutually exclusive fragment molecules generated from such atoms.

Completion of one cycle of the shrinking and expansion results in one or more fragment molecules and a reduced parent molecule. This process is repeated over and again recursively until all the atoms of the parent molecule are transferred into mutually exclusive fragments. Such recursive process has been undertaken to obtain uniform-sized fragments. However, the fragments, being mutually exclusive, suffer from truncation errors at the scissored bonds. Consequently, the atoms near the severed ends are poor mimics for the corresponding atoms in the parent molecule. These atoms need better representation in some other fragments. Hence, for each bond severed in the fragmentation process, an overlapping fragment is added. The overlapping fragment is obtained by traveling from both the atoms of the severed bond to equal width along all branches until the number of atoms included in the overlapping fragment reaches a threshold fragment size. This results in a complete set of fragments with every near-neighbor atom pairs (those separated by less than four bonds) of the parent molecule being mimicked reasonably well in at least one of the fragment.

The fragments created in the above manner need some processing (postprocessing) before being used to mimic the parent molecule DM. Adding an overlapping fragment for every severed bond results in a large number of fragments with many of them differing in just a few atoms. In postprocessing, such similar fragments are identified and merged into a single fragment, subject to the size constraint. In addition, all the unsatisfied valences are saturated with (dummy) hydrogen atoms, using appropriate bond lengths and orientation. Because weak bonds are treated equivalent to the covalent bonds during the fragmentation, the fragments may have small groups of atoms bound to the fragment only by a weak bond. Such loose groups, with less than four atoms are removed from the fragment molecule. Any close contacts due to the dummy hydrogen atoms are corrected by removing a few atoms from that site or by adding few more atoms from the parent molecule to make that region complete.

The next and very crucial task is to identify an appropriate fragment that would have the best mimic for a given DM element of the parent molecule. For this purpose, all the fragments containing the given near neighbour atom pair (a,b) of the parent molecule are examined. If R_a is the shortest distance (number of bonds) of the atom "a" to the nearest severed bond in the fragment and R_b, that of atom "b," then the fragment that has the largest value for $\mathrm{Min}(R_a,R_b)$, denoted as D_{ab}, is considered the best fragment for all the DM elements constituting the atom pair (a,b). In this manner, a validation matrix whose dimension is that of the number of atoms in the fragment is created for each fragment and stored. The elements of the matrix have a value of 0 or 1, depending on whether the given atom pair in the corresponding fragment is the best mimic or not. All the entries involving the dummy hydrogen atoms are set to zero. This matrix is crucial for synthesizing the parent DM. The first step of the procedure completes with this process. Figure 3 displays all the fragments generated from the test case shown in Figure 1. The fragments labeled F001, F002, and F003 are the mutually exclusive fragments, while F004 and F005 are the overlapping fragments. The program takes input in the form of fragment size range (minimum size and maximum size) apart from the Cartesian coordinates. The fragmentation process does not need any manual intervention/inspection until the creation of final fragments and the validation matrices, which help in generation of super-DM from the fragment DMs.

In the next step, calculations are performed on all the fragments created, at a desired level of theory (HF, MP2, etc.) using *ab initio* programs such as GAUSSIAN,[28] GAMESS,[29] or INDMOL.[30] Each DM element is read in by a program and referred to the corresponding validation matrix to decide whether the element has to be retained or to be ignored. The values collected from all the fragments are sorted as per their global indices (parent molecule) and stored in the form of three direct access files (DAF). The first file (DM file) stores only the numerically significant ($|P_{ij}| \geq 10^{-9}$) elements of the synthesized DM, as eight-byte records. The second file is an index file that stores the one-dimensional (global) index of the corresponding DM elements of the first file. It is stored as eight-byte integers. The third file is an atom pair file (lower triangular matrix). For every atom A it contains two four-byte

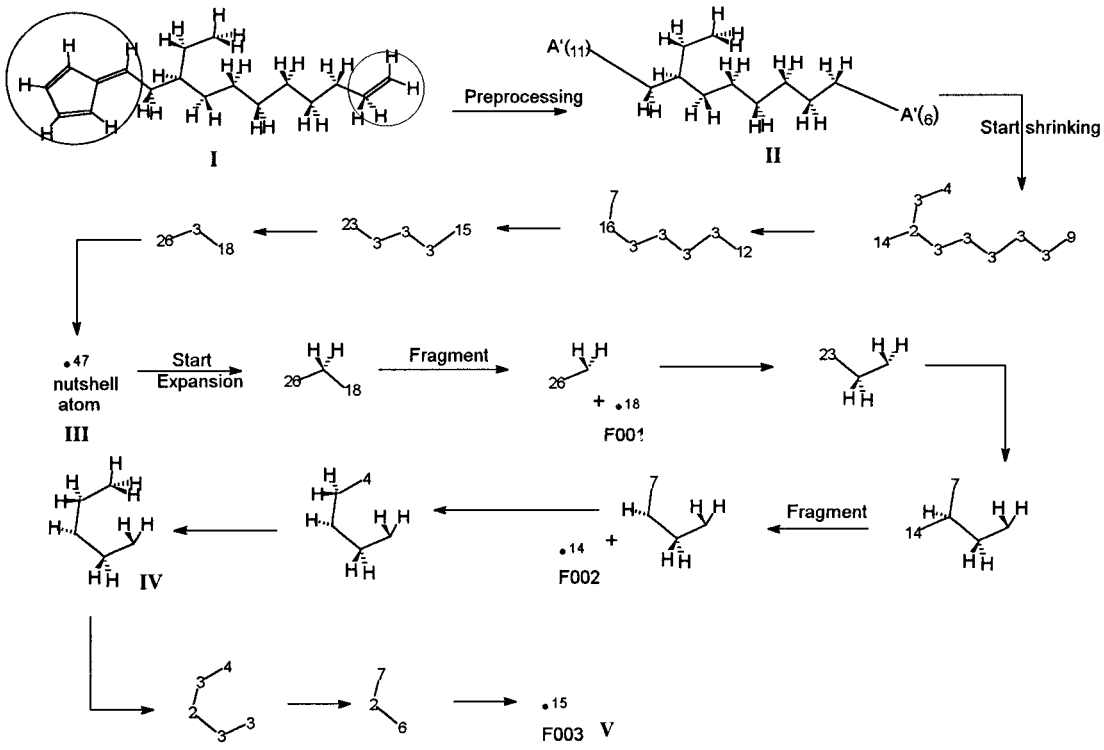

Figure 2. Schematic representation of fragmentation on a test molecule. This includes preprocessing, shrinking, and expansion procedures. The labels "A" and "A′" shown in structure II are the pseudoatoms that replace the encircled regions in structure I. All the numbers in the structures indicate bulk factors of the atoms (see text for details).

integers. These are the record numbers of the DM file where the first and the last DM elements pertaining to atom A in the first index ($P_{ij,}\ i \in A; j \leq i$) are stored.

The above method is economic, and provides quick access of the DMs for large molecular systems having thousands of atoms. The largest system studied by us so far (results to be published later) consists of 1482 atoms and 15210 basis functions at the HF/6-31G(d,p) level of theory. The storage requirements for this system are 56 MB, 28 MB, and 12 KB for the DM file, index file, and the atom pair file respectively, totaling to just 84 MB. This small storage requirement for this large system with a three-dimensional spatial extent shows that the disk storage requirements in this method are modest. In addition, the memory requirements are governed by the size of the fragment molecules (typically 40 atoms at present), and are met by even the present-day desktop computers. The synthesized DM thus obtained is used for computing the Trace(PS) to check its correctness. The elements of DM are uniformly scaled with an appropriate factor to correct the deviation of Trace(PS) from the number of electrons in the system, resulting in a normalized DM. Any property evaluation program, with minor modifications to take advantage of the storage structure of the DM, can be used to compute one-electron properties of the macromolecule.

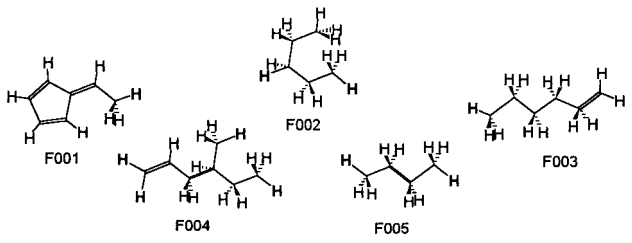

Figure 3. Fragments created from the fragmentation scheme shown in Figure 2. See text for details.

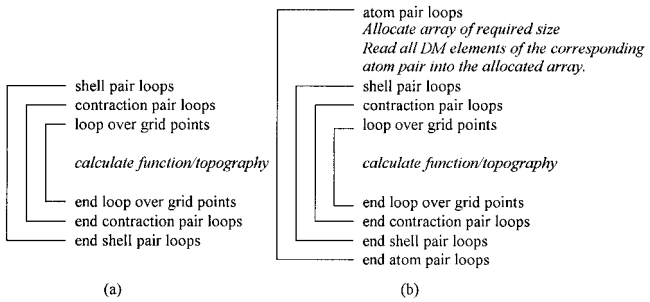

Figure 4. The old (a) and new (b) loop structurees in UNIPROP[31] for computing electronic properties. The new algorithm employs the synthesized DM generated by the molecular tailoring approach.

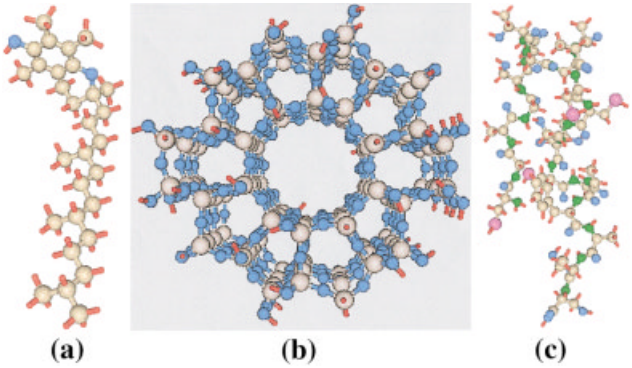

Figure 5. Structures of (a) tocopherol, (b) ZSM-5 silicalite zeolite, and (c) a model polypeptide.

It has been pointed out that property calculations employing linear scaling methods are rather scarce in the literature. For instance, it has been noticed by Matta[15] that MESP of the supermolecule can be constructed by using multipole moments of the fragments. The necessity of the development of a computer code for this purpose based on multipole moments has been mentioned.[15] However, a direct evaluation of MESP can be effected without employing the approximation of multipole moments. The program UNIPROP developed by Gadre et al.[31] has been adopted for this purpose of property calculation. Modifications to the UNIPROP program have been made enabling computation of properties using the synthesized DM in the DAF format. A new loop over atoms has been introduced, wherein the P_{ij}s due to atom A pairing with other atoms B ($i \in A$; $j \in B$ and $B \leq A$) are read into the RAM and their contribution to the property are computed before going to the next set of P_{ij}s. This modification, which has minimal memory requirements, is shown in the Figure 4. The largest memory required would be for the P_{ij} set with the last atom in the first index. All the graphical depictions of molecules and their electronic properties shown in this article have been generated using the visualization package UNIVIS-2000.[33]

Results and Discussion

Three distinct chemical systems (cf. Fig. 5) of varying nature have been chosen for testing the present algorithm. They are: (1) 5,7,8,-trimethyltocol ($C_{29}H_{50}O_2$) (TMT), a naturally occurring tocopherol with the strongest vitamin-E activity (81 atoms). This molecule has a head region comprising of an aromatic ring fused with a heterocyclic ring and a hydrocarbon tail region. This rather small molecule has been chosen for extensive benchmarking of the properties computed by molecular tailoring approach, in comparison with the *actual* ones, at different levels of theory, including the correlated one. (2) A model polypeptide ($C_{70}H_{112}O_{29}N_{22}S_5$) built of 22 amino acid residues and 238 atoms. It consists of two coils of α-helix and a small β-sheet with intervening hydrogen bonds between them. This molecule can be considered as a prototype system for proteins and other biological macromolecules. (3) ZSM-5 silicious-zeolite cluster ($Si_{128}O_{284}H_{92}$) comprising of 504 atoms with a three-dimensional structure. It is hoped that catalyzed reactions could be studied by modeling the MED and MESP of such substrates.

TMT ($C_{29}H_{50}O_2$)

The molecule has been fragmented in three different ways by providing distinct size ranges in the input. The statistical data of the fragments produced for each fragmentation scheme were analyzed to grade the schemes. The analysis includes the number of fragments generated, the total number of atoms produced in all the fragments put together, the smallest and largest fragment sizes, and the depth factor for each near neighbour atom pairs.

The statistical data are tabulated in Table 1 for three fragmentation schemes. A good scheme would be the one with minimum number of optimum sized (between 20 and 40 atoms) fragments. The depth factor, D_{ab}, is a measure of the amount of original environment around (a,b) that is retained in the fragment, and hence, reflects the quality of the constituent DM elements. D_{ab} is useful for *a priori* quantification of the "goodness" of the fragments produced by a scheme. Table 2 provides a summary of the depth factors for all near-neighbor atom pairs that are separated by four bonds are less. The best scheme is then chosen for synthesizing the DM and computing the properties. However, in the first test case, properties have been computed using all the schemes for the sake of comparison. Table 1 also gives the Trace(PS), a measure of the correctness of DM, obtained from the synthesized DMs for all the three Schemes. Scheme 1 produces accurately, the number of electrons in the system, while Schemes 2 and 3 have larger errors in Trace(PS). Table 3 provides a quantitative comparison of the synthesized DM with the *actual* one. It is seen that

Table 1. Fragment Size Statistics of Three Different Schemes for TMT Molecule along with the Time Required to Obtain the DMs at the HF/6-31G(d, p) Level.

Scheme	No. of fragments	Total atoms	Smallest frag size	Largest frag size	Av. size	Total CPU time[a] (min)	Trace (PS)
Actual	1	81				320	240.0000
1	6	225	33	41	37.5	250	239.9996
2	11	279	19	28	25.4	204	239.9583
3	11	306	19	31	27.8	252	239.9259

See text for details.

[a]Calculations carried out on a PC with Pentium4@1.5 GHz processor and 512 MB SDRAM.

Table 2. Quality of Atom Pairs as Represented in Fragments for Three Fragmentation Schemes of TMT.

Scheme	B_{ab}	Total no. of pairs	No. of pairs with D_{ab} = 1	2	3	4
1	0	81	0	0	8	22
	1	82	0	2	22	21
	2	158	0	8	40	42
	3	220	2	22	59	60
2	0	81	0	16	27	20
	1	82	7	37	19	14
	2	158	22	64	38	20
	3	220	43	87	66	9
3	0	81	0	10	25	24
	1	82	2	28	27	13
	2	158	10	51	50	25
	3	220	23	83	66	30

B_{ab} is the number of bonds between atom pairs under consideration. Depth value (D_{ab}) is the shortest number of bonds, from atoms a and b, to the nearest truncated end in the fragment, chosen to be best mimic for atom pair (a, b). See text for details.

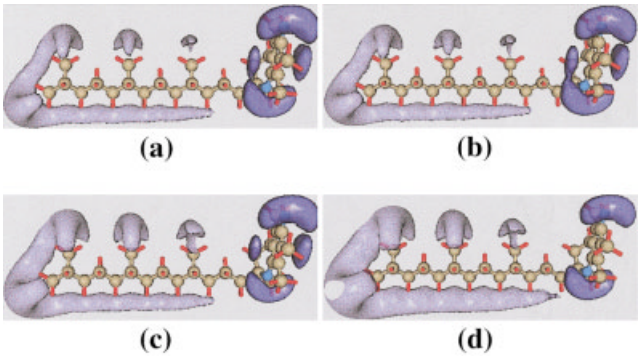

Figure 6. Comparison of MESP isosurfaces of TMT computed using (a) *actual* DM, and the synthesized DMs obtained by (b) Scheme 1, (c) Scheme 2, and (d) Scheme 3 at the HF/6-31G level of theory. The dark blue isosurfaces are of MESP value -57.75 kJ mol^{-1}, while the light blue ones are of MESP -5.25 kJ mol^{-1}.

most of the numerically significant DM elements have an error of less than 1%. These DMs are then uniformly scaled using the factor N_{el}/Trace(PS), where N_{el} is the number of electrons in the system. This normalized DM is employed for calculating the one-electron properties.

A pictorial comparison of the MESP obtained with these schemes is depicted in Figure 6. In this figure, (a) is the MESP computed by performing SCF on the entire molecule, while (b), (c), and (d) are those computed using the synthesized DMs obtained by Schemes 1, 2, and 3, respectively. Two isosurfaces of different MESP values (dark blue: -57.75 kJ mol^{-1}, light blue: -5.25 kJ mol^{-1}) are shown in these figures to compare the results in different function ranges. The features observed in the dark blue surface are that due to the lone pair of electrons of the two oxygen atoms and the π-electrons of the aromatic ring that is seen to be missing in (d) and partly diminished in (c). The hydrocarbon tail gives rise to weakly negative MESP isosurfaces (light blue), of which only that in (b) has identical size with corresponding surface of (a). Thus, the plot from Scheme 1 [cf. Fig. 6b], is seen to have a striking similarity with the *actual* result, (a) in both the shape and size of the surfaces. This observation agrees with the predictions from the precomputation analysis using the data given in Tables 1 and 2. The MESP and MED topography have been mapped using the *actual* and synthesized DMs. Table 4 provides MESP minima corresponding to the lone pair of electrons of the two oxygen atoms (CPs 1–4) and minima due to the hydrocarbon tail region (CPs 5–8). The max-error of the MESP value at the CP is about 1% for Scheme 1, while the max-deviation of the CP position is 0.026 Å. The errors in the results from Schemes 2 and 3 are higher. Similar analysis of errors in the MED topography shown in Table 5 indicates a negligible error for Scheme 1, while a small error is observed for Schemes 2 and 3. This warrants that the error associated with any density-related property would also be negligible for a good fragmentation scheme.

The above analyses reveal that the Scheme 1, in addition to being the best mimic, is capable of accurately reproducing the

Table 3. Analysis of Error in the Synthesized DM with Respect to the *Actual* Ones for Different Schemes of TMT and the Model Polypeptide.

Molecule/Level	Scheme	$\lvert P_{ij}\rvert \geq 1.0$ N_{val}	N_{cor}	$1.0 > \lvert P_{ij}\rvert \geq 0.1$ N_{val}	N_{cor}	$0.1 > \lvert P_{ij}\rvert \geq 0.01$ N_{val}	N_{cor}
TMT	1	31	31	1035	1015	9568	8289
[HF/6-31G(d, p)]	2	31	31	1035	901	9568	4890
	3	31	31	1035	943	9568	5818
TMT	1	37	37	772	771	3593	3409
[MP2/STO-3G]	2	37	37	772	740	3593	2545
Model polypeptide	1	324	317	3121	2337	11582	5659
[HF/STO-3G]	2	324	322	3121	2639	11695	6772

The notations N_{val} and N_{cor} indicate the number of values of the *actual* DM that fall within the given range and the number of the corresponding values in the synthesized DM that have less than 1% error.

Table 4. MESP Topography of TMT Computed from the *Actual* and Synthesized DMs.

			Scheme 1		Scheme 2		Scheme 3	
CP no.	CP type	$V(\mathbf{r})$	$V'(\mathbf{r})$	Δr	$V'(\mathbf{r})$	Δr	$V'(\mathbf{r})$	Δr
1	(3,+3)	−190.88	−193.00	0.015	−159.75	0.155	−189.94	0.058
2	(3,+3)	−195.42	−197.59	0.011	−154.21	0.069	−191.08	0.017
3	(3,+3)	−191.87	−191.65	0.001	−190.95	0.001	−182.51	0.024
4	(3,+3)	−187.32	−186.26	0.001	−185.64	0.001	−177.76	0.018
5	(3,+3)	−21.66	−21.75	0.013	−23.45	0.040	−26.24	0.043
6	(3,+3)	−15.54	−15.48	0.007	−15.60	0.059	−20.78	0.054
7	(3,+3)	−16.27	−15.97	0.026	−21.23	0.108	−22.15	0.070
8	(3,+3)	−15.83	−15.63	0.008	−15.26	0.025	−19.30	0.008

$V(\mathbf{r})$ and $V(\mathbf{r}')$ are the MESP values, in kJ mol^{-1}, at the CPs obtained from the *actual* and synthesized DMs, respectively. The distance between the CPs, obtained from the synthesized DM and the *actual* one is denoted as Δr.

actual results. A comparison of the dipole moments computed in a similar manner is shown in Table 6. Scheme 1 is seen to produce an error less than 1%, while results from the other schemes show larger errors. In addition to comparisons of electronic properties, the total electronic energy of the molecule has also been computed using the synthesized DM from Scheme 1 at the HF/STO-3G level of theory. The total energy of the system without iterating the DM turned out to be 1261.63143 a.u. while the *actual* energy at the HF/STO-3G level of theory is 1261.61205 a.u, that is, an error of the order 10^{-2}. However, performing one iteration of the SCF procedure starting from the synthesized DM lead to an improved energy that is correct up to five decimal places.

Model Polypeptide ($C_{70}H_{112}O_{29}N_{22}S_5$)

This system has 238 atoms and 998 electrons. The amino acid residue sequence is CYS-GLY-ALA-GLY-SER-CYS-THR-GLY-GLY-ALA-ALA-CYS-CYS-SER-THR-CYS-GLY-ALA-PHE-ALA-THR-SER. The motivation for the choice of this system is to check whether the fragmentation scheme is able to successfully capture features of such systems with weak interactions. The biological significance of such a case need not be overemphasized.

Molecular tailoring of the model polypeptide is done employing three different schemes whose results are presented in Table 7. It may be noticed that the first two schemes are of reasonable size (total number of atoms less than 1200). The third scheme has large fragments and could consume large CPU time. Hence, it is not considered for further work. The total CPU times taken for HF/STO-3G calculations on the fragments from Schemes 1 and 2 are 32 and 63 min, while the *actual* calculation takes 497 min. Thus, the CPU time advantage, viz. a factor of 8 to 16, at this simple level of calculation is evident from this example. The "goodness" of the synthesized DM can be assessed from analysis reported in Table 3. The synthesized DM from Scheme 2 has reasonably small error with most of the values unaccounted for in the table having an error between 1 to 5%. Thus, the synthesized DM is expected to capture the salient features of the corresponding *actual* one.

The MESP isosurfaces of value −57.8 kJ mol^{-1} are depicted for the *actual* HF/STO-3G (a) and the synthesized DMs [(b) and (c)] in Figure 7. This figure vividly brings out the overall agreement in these MESP plots, although some subtle differences are also noticeable. This may be attributed to somewhat unbalanced representations of the DM elements in the chosen fragmentation schemes. A good agreement among the corresponding topography of MESP is brought out by Table 8. It is seen from this table that both the schemes are able to mimic the *actual* topography (in terms of the nature of CP and the corresponding MESP) quite well.

Table 5. Comparison of MED Bond CPs for TMT Molecule.

			Scheme 1		Scheme 2		Scheme 3	
CP no.	CP type	$\rho(\mathbf{r})$	$\rho'(\mathbf{r})$	Δr	$\rho'(\mathbf{r})$	Δr	$\rho'(\mathbf{r})$	Δr
1	(3,−1)	0.35322	0.35321	0.000	0.35331	0.004	0.35277	0.006
2	(3,−1)	0.36115	0.36115	0.000	0.36092	0.006	0.36111	0.001
3	(3,−1)	0.36748	0.36750	0.000	0.36789	0.002	0.36740	0.000
4	(3,−1)	0.36125	0.36126	0.000	0.36106	0.005	0.36122	0.002
5	(3,−1)	0.30428	0.30432	0.000	0.30419	0.000	0.30443	0.000
6	(3,−1)	0.29994	0.29994	0.000	0.30025	0.001	0.30026	0.000

The densities at the CPs obtained from the *actual,* $\rho(\mathbf{r})$ and synthesized, $\rho'(\mathbf{r})$ DMs and the relative positions of CPs from the synthesized DMs with respect to the *actual* ones, denoted as Δr are provided.

Table 6. Dipole Moment of Test Molecules, in Debye Units, Obtained from the *Actual* and Synthesized DMs.

Molecule/Level	*Actual*	Scheme 1	Scheme 2	Scheme 3
TMT [HF/6-31G(d, p)]	2.894	2.962	2.486	2.470
TMT [MP2/STO-3G]	2.400	2.413	2.234	—
Polypeptide [HF/STO-3G]	22.460	22.054	20.406	—
Polypeptide [HF/6-31G(d, p)]	—	32.278	26.626	—

However, Scheme 2 represents the positions of the CPs somewhat better than the Scheme 1.

We next report the results of MESP calculation carried out at HF/6-31G(d,p) level. The Trace (PS) for the synthesized DMs turn out to be 998.197 and 998.307 for Schemes1 and 2, respectively (cf. Table 7). The corresponding synthesized MESP isosurfaces of values -52.5 kJ mol^{-1} (blue) and 52.5 kJ mol^{-1} (red) along with the embedded contours of same values are displayed in Figure 8. The shapes of the isosurfaces in Figure 8a and b are strikingly similar, although quantitatively different in some regions. At the HF/6-31G(d,p) the number of basis functions for the supermolecule is 2470 and would be a formidable task to carry out SCF calculation on the supermolecule. Hence, the MESP features from the synthesized DM cannot be compared with the *actual* ones in the present case. However, it is gratifying to note that the MESP features exhibited by the two schemes resemble each other quite well, and hence could be taken as a good representation of 6-31G(d,p) level MESP for this model polypeptide.

ZSM5 Silicious Zeolite ($Si_{128}O_{284}H_{92}$)

The modeled system has large 10-membered rings surrounded by five-memebered rings and four such layers one over another forming a mimicked cage of 10-membered rings. This zeolite cluster has been scissored into 121 fragment molecules of sizes ranging from 29 to 51 atoms, the average size being 43.5. The computational time taken to obtain the synthesized DMs at the HF/STO-3G level is 7.5 h on a Pentium-IV processor at 1.5 GHz clock speed, while the time taken at the HF/6-31G(d,p) level is ~50 days on a SUN ultrasparc 450 machine (with our benchmarking, this time would correspond to about 12.5 days on a Pentium4 at 1.5 GHz). The number of electrons (actual: 4156) produced by the synthesized DMs at these two levels are 4155.62 and 4158.44, respectively with the corresponding errors being 0.01 and 0.04%. Figure 9a depicts the isosurface of MESP -105 kJ mol^{-1} computed at the HF/6-31G(d,p) level. Only the MESP features inside the cage are of interest in this system. The cage is rich with negative MESP features, while the five-membered rings surrounding this region are void of any negative features. The MESP features inside the cage are brought out by Figure 9b displaying the MESP textured on the van der Waals surface of the molecule. Patches of blue color indicating charge localization within the cage can be observed. In addition to the steric reasons, these regions of electron localization would guide the adsorption of an organic molecule at the cage site.

With the increase in computational power, it is becoming mandatory to report *ab initio* results that incorporate the correlation effects. Apart from the DFT based methods, the Møller–Plesset method is the most widely used one. Among the three test cases discussed above, the feasibility of performing an MP2 calculation on the parent molecule exists only for the TMT molecule. Hence, the DM of TMT has been synthesized at the MP2/STO-3G level using the fragment sets reported in Table 1. The time taken for obtaining the *actual* and the synthesized DMs from Schemes 1 and 2 are 168, 40.5, and 20 min, respectively on a SGI ORIGIN 200 workstation (with R10000 processor at 180 MHz and 256 MB RAM). Despite the small size of the parent molecule, the time taken to synthesize the DM (using Scheme 1, which is shown to produce excellent results at the HF level) is only one-fourth of that required for obtaining the *actual* DM. The corresponding number of electrons produced by the synthesized DMs from Schemes 1 and 2 are 239.995 and 239.957, that is, an error of 0.002% for the DM from Scheme 1. A quantitative analysis of the DMs is given in Table 3 along with the HF results. Figure 10 features the MESP textured on a 0.002 a.u. MED surface for the results from *actual* (a) and the synthesized DM (b). The color mapping ranges from blue to red (through white) as the MESP goes from negative to positive (through zero). This figure shows that the error in (b) is not visually perceivable. Thus, it appears that the fragmentation scheme implemented at the correlated level works quite well: the error being comparable with that at the HF level.

A quantitative comparison has also been carried out for the MEDs of some cases computed over a parallelepiped encompassing the parent molecule . The numerical comparison index,[26] $S_{XY}(0.001)$ obtained for MED of TMT (Scheme 1) turns out to be

Table 7. Fragment Size Statistics for Three Different Schemes on the Model Polypeptide.

Scheme	No. of fragments	Total atoms	Smallest frag. size	Largest frag. size	Av. size	CPU time[a] (min)	Trace (PS)
Actual						497.0	998.00
1	35	1020	22	34	29.1	32.3 (1017)	998.14 (998.20)
2	31	1199	30	46	38.7	63.2 (1901)	998.18 (998.31)
3	35	1546	38	55	44.2		

The time required to obtain the synthesized DMs at the HF/STO-3G (6-31G(d, p)) levels and the number of electrons produced by the synthesized DMs are given.

[a]All the calculations were carried out on a PC with Pentium@1.5 GHz processor and 512 MB SDRAM.

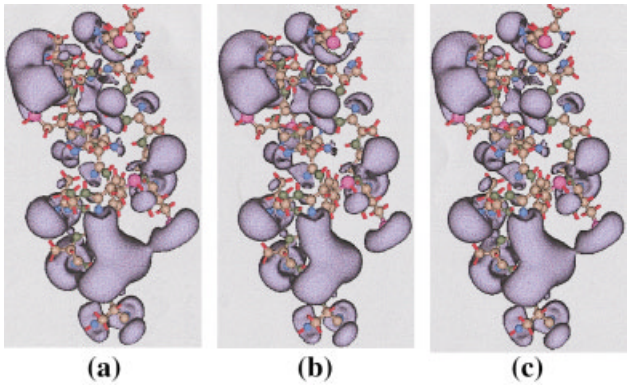

Figure 7. Isosurfaces of MESP (HF/STO-3G) value -57.8 kJ mol^{-1}, for the model polypeptide, obtained by employing Scheme 1 (b) and Scheme 2 (c), along with the actual results (a). See text for details.

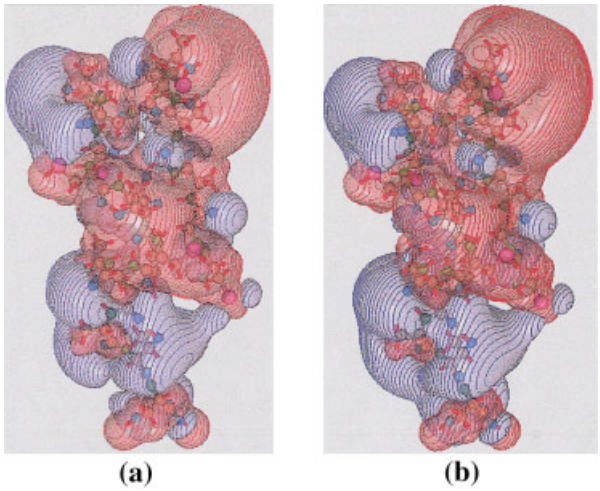

Figure 8. Isosurfaces and contours of MESP values -52.5 kJ mol^{-1} (blue) and 52.5 kJ mol^{-1} (red) computed at the HF/6-31G(d,p) level using the DMs obtained through fragmentation Schemes 1 (a) and 2 (b), for the model polypeptide.

0.9998 and 0.99997 at the HF/6-31G(d,p) and MP2/STO-3G levels, respectively. In the case of peptide (Scheme 1) the index for MED is 0.9997. A point-vise comparison of the values has also been carried out. In the case of TMT (Scheme 1), 18,950 values of 19,325 values that are above 0.001 a.u. density have an error of less than 0.5%. Similarly, for the peptide molecule there are 59,444 points with MED greater than 0.001 a.u. and 54,960 of them have an error of less than 0.5%. It should be noted that our approach produces results that are more accurate representation of the *ab initio* properties than those reported by Mezey et al.[24,25]

Concluding Remarks

Linear scaling methods have gained popularity for effective treatment of large chemical systems at relatively less computational effort. However, a survey of the literature reveals that they are applied largely to solids, mostly within DFT or semiempirical framework and the stress has been on electronic energies. There are scanty references to the calculation of molecular properties. The electronic properties govern many weak interactions, for example, those between proteins with other biomolecules, drug molecules with substrates, zeolite cages with the adsorbed reactants, etc. Therefore, it is important to study the one-electron properties. The efficacy and the accuracy of these properties obtained from linear-scaling type procedures need to be tested. The present study reports the findings of such an appraisal. There have been earlier works[15,17,24] dealing with one-electron properties of large molecules. In particular, the approach of Mezey et al.[24] utilizes a prestored database of fragment densities/DMs. This enables a rapid computation of the electronic densities. However, in the present approach the fragments are generated on a case-by-case basis enabling a better simulation of the local atomic environment, albeit at a higher computational cost.

The development of a practical method for linear scaling computation of *ab initio* one-electron properties for large molecules in a given, fixed geometry is presented. A critical analysis of the DMs synthesized by our method and the one-electron properties such as the MESP, MED, and dipole moment, computed from the synthesized DM, has been carried out. The error in the synthesized DM has been found to lie within 1% for most of the values that are numerically greater than 0.1. These values are the major contributors to the electronic properties. Although the DM elements that are numerically less than 0.1 are larger in number, their contributions are in general scaled down further by the exponent terms in the equation for property evaluation. A fragment size statistics also provides *a priori* quantification of the "goodness" of the fragments. This enables the fragmentation of the molecule by several schemes and choosing an appropriate one before performing the actual calculation. The method has been applied to three test cases.

The results for the first case are very encouraging, with excellent reproduction of all the three electronic properties discussed, using the synthesized DM from Scheme 1. However, there is only a nominal reduction in the time required to compute the synthesized DM compared to the *actual* one. This is due to the small size (81 atoms) of the supermolecule. However, at the MP2 level of

Table 8. Comparison of the MESP Values and Positions of Some Representative CPs Computed at the HF/STO-3G Level.

			Scheme 1		Scheme 2	
CP no.	CP type	$V(\mathbf{r})$	$V'(\mathbf{r})$	Δr	$V'(\mathbf{r})$	Δr
1	(3,+3)	−338.75	−331.82	0.004	−327.22	0.003
2	(3,+3)	−350.45	−346.49	0.003	−344.65	0.003
3	(3,+3)	−313.31	−299.03	0.007	−305.55	0.006
4	(3,+3)	−280.45	−259.49	0.007	−271.36	0.009
5	(3,+3)	−277.88	−277.18	0.007	−273.61	0.007
6	(3,+3)	−271.47	−274.46	0.005	−269.12	0.005

$V(\mathbf{r})$ and $V'(\mathbf{r})$ are the MESP at the CPs obtained using the actual DM and synthesized DMs, respectively. The error in the position of the CPs obtained using synthesized DMs is denoted as Δr.

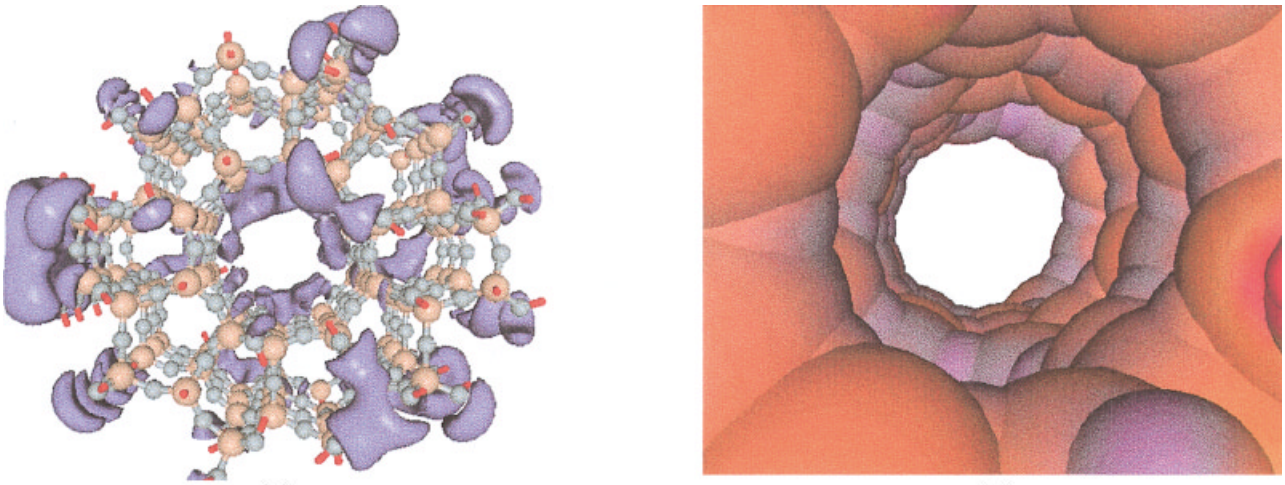

Figure 9. MESP plots of the ZSM-5 zeolite computed at the HF/6-31G(d,p) level of theory. (a) Isosurfaces of MESP -105 kJ mol^{-1} (b) MESP textured on the molecular van der Waals surface.

theory, a time advantage by a factor of ~4 is seen even for this small test molecule. An increase in the size of the parent system would lead to a linear increase in the time required to obtain the synthesized DM and for large molecules, one can expect a distinct advantage in the computation time. This has been observed in the second test case. Results from two fragmentation schemes are compared with the *actual* ones for this case. Although the Scheme 2 produces better results for this case the MESP from Scheme 1 is seen to be better in some regions indicating a slight nonuniform representation of the fragments. The CPU times reported for Scheme 2 show a factor of 8 (a*ctual* vs. synthesized) at the HF/STO-3G level. An extrapolation of this result to HF/6-31G(d,p) level leads to an estimated CPU time of about 39 days (on P4 at 1.5 GHz) as against the timing of 1.3 days (leading to a factor of 30) required for obtaining the synthesized DM at this level. Similar consideration extended to the case 3, the zeolite molecule would reveal a substantial advantage in computer time. In this case, even an HF/STO-3G level calculation on the parent molecule is a daunting task.

The MED obtained by the fragmentation scheme shows negligible error in all the benchmarked cases. However, the weight factors employed for extracting MESP and dipole moments would imply somewhat larger errors for these two entities. The former shows a typical error of 2 to 5%, whereas the latter is seen to be in somewhat larger error (typically 2 to 10%) due to the weight factor **r** representing expectation value of the dipole operator. In summary, a practical method for constructing the molecular electron density, electrostatic potential, and dipole moment at the HF or correlated levels has been reported. Although the benchmarking of total electronic energy has not been carried out in an exhaustive manner, the present work indicates that a fairly accurate value of the former can be obtained by employing the synthesized DM for the supermolecule. Further, it is expected that the convergence to the HF value would be fairly fast.

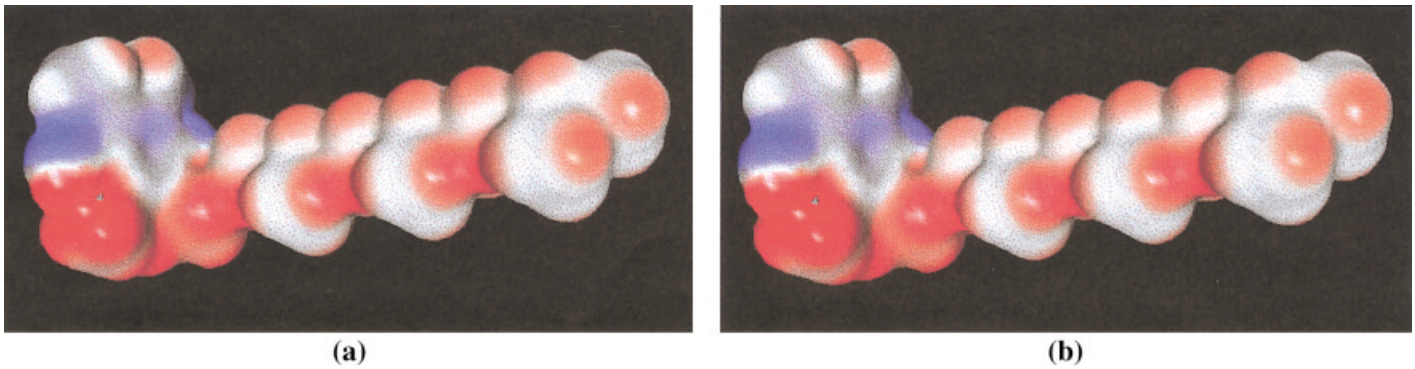

Figure 10. A comparison of MESP textured on MED isosurface (0.002 a.u.) for the TMT molecule, computed at the MP2/STO-3G level of theory, obtained by (a) performing *ab initio* caluclation on the complete molecule, and (b) molecular tailoring approach with fragmentation Scheme 1. See text for details.

Our method has a potential for effectively handling molecules comprising of a few thousand atoms and can be fruitfully employed for studying a variety of problems such as biomolecular interactions, catalytic reactions, as well as properties of molecules in crystals. Further, this method, in principle, does not have limitations on the size of the systems studied, although the computational cost would scale almost linearly with the system size. With the continuing increase in the computing power of the desktop computers one can foresee that it would not be long before the tool of *ab initio* quantum chemistry is taken into the realm of macroscopic systems such as proteins, DNA, molecular assemblies, etc.

Acknowledgments

The authors thank the Centre for Development of Advanced Computing (C-DAC) for extending the National PARAM Supercomputing Facility (NPSF), where a part of the calculations reported in this work has been carried out.

References

1. (a) Szabo, A.; Ostlund, N. S. Modern Quantum Chemistry; McGraw-Hill: New York, 1989; (b) Hehre, W. J.; Random, L.; Schleyer, P. v. R. Ab Initio Molecular Orbital Theory; John Wiley: New York, 1986; (c) Pople, J. A.; Beveridge, D. L. Approximate Molecular Orbital Theory; McGraw-Hill: New York, 1970.
2. (a) Almlöf, J.; Faegri, K.; Korsell, K. J Comp Chem 1982, 3, 385; (b) Strout, D. L.; Scuseria, G. E. J Chem Phys 1993, 102, 8448.
3. (a) See, for a review, Goedecker, S. Rev Mod Phys 1999, 71, 1085; (b) Yang, W.; Pérez–Jordá, J. M. Encyclopedia of Computational Chemistry; Schleyer, P. v. R., Ed.; John Wiley: New York, 1998; p 1496.
4. Yang, W. Phys Rev Lett 1991, 66, 1438.
5. Li, X. P.; Nunes, R. W.; Vanderbilt, D. Phys Rev B 1993, 47, 10891.
6. Yang, W.; Lee, T.-S. J Chem Phys 1995, 163, 5674.
7. Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. J Phys Chem 1994, 98, 9165.
8. Lewis, J. P.; Carter, C. W., Jr.; Hermans, J.; Pan, W.; Lee, T.-S.; Yang, W. J Am Chem Soc 1998, 120, 5407.
9. (a) Dixon, S. L.; Merz, K. M., Jr. J Chem Phys 1996, 104, 6643; (b) Dixon, S. L.; Merz, K. M., Jr. J Chem Phys 1997, 107, 879.
10. Dixon, S. L.; van der Vaart, A.; Gogonea, V.; Vincent, J. J.; Brothers, E. N.; Westerhoff, L. M.; Merz, K. M., Jr. DivCon 99, The Pennsylvania State University, University Park, PA 16802, 1999.
11. Ermolaeva, M. D.; van der Vaart, A; Merz, K. M., Jr. J Phys Chem A 1999, 103, 1868.
12. Gogonea, V.; Merz, K. M. Jr. J Phys Chem A 1999, 103, 5171.
13. Van der Vaart, A.; Merz, K. M., Jr. J Phys Chem A 1999, 103, 3321.
14. Bliznyuk, A. A.; Rendell, A. P.; Allen, T. W.; Chung, S.-H. J Phys Chem 2001, 105, 12674.
15. Matta, C. F. J Phys Chem A 2001, 105, 11088.
16. Khandogin, J.; Hu, A.; York, D. M. J Comp Chem 2000, 21, 1562.
17. Challacombe, M. J Chem Phys 1999, 110, 2332.
18. Daw, M. S. Phys Rev B 1993, 47, 10895.
19. Strain, M. C.; Scuseria, G. E.; Frisch, M. J. Science 1996, 271, 51.
20. Scuseria, G. E. J Phys Chem A 1999, 103, 4782.
21. Stratmann, R. E.; Scuseria, G. E.; Frisch, M. J. Chem Phys Lett 1996, 257, 213.
22. Bader, R. F. W. Atoms in Molecules—A Quantum Theory; Oxford University Press: Oxford, 1990.
23. McWeeny, R. Rev Mod Phys 1962, 126, 1028.
24. Walker, P. D.; Mezey, P. G. J Am Chem Soc 1994, 116, 12022, and references therein.
25. Mezey, P. G. J Math Chem 1995, 18, 141, and references therein.
26. Good, A.; Richards, W. G. J Chem Inf Sci 1992, 33, 112.
27. (a) *Ab Initio* Quality Electrostatic Potentials of Proteins: An application of the ADMA Approach; Exner, T. E.; Mezey, P. G. (Unpublished); (b) *Ab Initio* Quality Molecular Properties for Macromolecules Using the ADMA Approach; Exner, T. E.; Mezey, P. G. (Unpublished).
28. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head–Gordon, M.; Gonzalez, C.; Pople, J. A. Gaussian 94; Gaussian Inc.: Pittsburgh, PA, 1995.
29. (a) Granovsky, A. A. (http://classic.chem.msu.su/gran/gamess/index.html); (b) The package GAMESS; Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. J Comput Chem 1993, 14, 1347.
30. The package INDMOL, for sequential and parallel computation of *ab initio* molecular orbitals; Shirsat, R. N.; Limaye, A. C.; Gadre, S. R. J Comp Chem 1993, 14, 445.
31. Shirsat, R. N.; Bapat, S. V.; Gadre, S. R. Chem Phys Lett 1992, 200, 373; Package UNIPROP can evaluate *ab initio* electronic properties such as MESP, MED, the Laplacian of MED, electron momentum density, topography of all the above properties, internal electric field, electronic moments up to the third order.
32. Limaye, A. C.; Gadre, S. R. Curr Sci (India) 2001, 80, 1296.