The Field-Adapted ADMA Approach: Introducing Point Charges

Thomas E. Exner*,† and Paul G. Mezey‡,§

Fachbereich Chemie, Universität Konstanz, D-78457 Konstanz, Germany, Department of Chemistry and Department of Physics and Physical Oceanography, Memorial University of Newfoundland, St. John's, NF, Canada A1B 3X7, and Department of Organic Chemistry, Eötvös Loránd University, Pázmány Péter Sétány 1/A, H-1117 Budapest, Hungary

Received: November 12, 2003; In Final Form: February 23, 2004

New developments of the adjustable density matrix assembler (ADMA) approach to macromolecular quantum chemistry are described, based on the original fuzzy density matrix fragmentation scheme combined with an approach of using point charges to approximate the effects of additional, distant parts of a given macromolecule in the quantum chemical calculation of each fragment. The ADMA approach divides a macromolecule (the target molecule) into fuzzy fragments, for which conventional quantum chemical calculations are performed using moderate-sized "parent molecules" which contain both the fragment and all the local interactions of the fuzzy fragment with its surroundings within a preselected distance. For any such distance criterion, that is, for any size limit for the parent molecules, the computational time scales linearly with the size of the macromolecule. As demonstrated in earlier papers, in the original, linear-scaling ADMA approach, the accuracy is fully controlled by this distance, and with a large enough distance criterion nearly exact results are obtained when compared with the conventional Hartree−Fock method. In the new field-adapted ADMA method the same accuracy can be achieved using a smaller distance criterion for the parent molecules if in each parent molecule calculation point charges are also used to represent distant parts of the macromolecule. This allows one to use smaller parent molecules and faster overall calculations resulting in the same overall accuracy that can be achieved only with larger parent molecules in the original ADMA method. Specifically, in the quantum chemical calculations determining the fragment density matrices, each parent molecule is placed within a point-charge field representing the rest of the macromolecule. Consequently, not only the short-range interactions within the actual parent molecule, but also the approximate effects of longer-range electrostatic interactions present in the rest of the macromolecule, are included in the new fragment density matrices. With a number of test calculations of small oligopeptides and proteins, it is shown that the inclusion of partial charges is an efficient tool to obtain results of a uniform accuracy for all these test cases, and that this approach can be used to reduce the need to include longer-range interactions by explicit quantum chemical calculation for much larger parent molecules for the fragments. With a large increase in accuracy and the decrease in computational demand, the field-adapted ADMA approach is now able to describe efficiently very large biomolecular systems at the ab initio quality level.

I. Introduction

Standard ab initio quantum chemical calculations, like the traditional Hartree−Fock method, are only feasible for small to medium-sized molecules due to the large amount of computer power needed. Even with the large increase in computer speed of every new hardware generation, this fact will not change soon because of the high power scaling behavior (O(N^3) or worse) of these methods with system size. To circumvent this problem, two of the linear-scaling, fuzzy electron density construction methods, the numerical MEDLA (molecular electron density loge approach) technique[1−4] and the more advanced adjustable density matrix assembler (ADMA) approach, have been proposed,[5−8] leading to the first ab initio quality protein calculation over the 1000 atom limit,[1−4] where the term "ab initio quality" refers to the demonstrated fact that the macromolecular result obtained with a given basis set is at least as accurate as the conventional Hartree−Fock result with a slightly smaller basis set.[1−4] The term linear scaling is meant literally, and not in the restricted sense of referring to a family of methods involving specific selections for integrals and linearized approximations used for the diagonalization of matrices in the Hartree−Fock method or alternative approximations based on density functional approaches, leading to linear-scaling or nearly linear-scaling methods.[9−18] The MEDLA and ADMA methods have been rigorously tested, both for actual linear scaling and for accuracy, using a wide selection of examples of actual macromolecules, including several proteins. [1−8,19,20] The comparisons with alternative linear-scaling methods[9−18] are very encouraging, whenever actual results with such alternative approaches are available. Note, specifically, that besides the quoted MEDLA and ADMA results,[1−8,19,20] very few actual error analyses for linear-scaling quantum chemical protein calculations have been published so far.

In the ADMA approach, a macromolecule is divided into fuzzy fragments, for which conventional quantum chemical calculations are performed using moderate-sized "parent mol-

* Corresponding author. Phone: +49-7531-88-2024. Fax: +49-7531-88-3898. E-mail: exner@chemie.uni-konstanz.de.
† Universität Konstanz.
‡ Canada Research Chair in Scientific Modeling and Simulation, Memorial University of Newfoundland. E-mail: pmezey@mun.ca.
§ Albert Szent-Györgyi Professor, Eötvös Loránd University.

10.1021/jp037447p CCC: $27.50
Published on Web 04/17/2004

ecules" which contain the fuzzy fragment with some additional local surroundings, including all the local interactions of the fragment within a selected distance. The obtained electron density matrices of the fragments are combined to approximate the electron density matrices of the whole molecules. By choosing a large enough distance criterion, the results of an ideal direct macromolecular ab initio computation (still unattainable for most macromolecules of more than 1000 atoms) can be approximated to any desired accuracy.

It has been shown[19] that an ADMA accuracy of better than 1 kcal/mol compared to the conventional Hartree−Fock results can be obtained with a distance criterion of 12 Å. This is the distance up to which local interactions of the fragment are reproduced in the parent molecules. Although, relative to a conventional Hartree−Fock result, a 1 kcal/mol accuracy for a protein of total energy of tens or hundreds of millions of kcal/mol is remarkable, nevertheless, if needed, these results can be further improved by using a larger distance criterion.[19] However, due to the increased amount of computer time needed for the quantum chemical calculations of larger parent molecules, such an accurate ADMA approach in its current implementation is efficient only for single-point calculations. Note, however, that a much improved integral management scheme, fully exploiting the unique sparse matrix features of the ADMA method and avoiding the computation of a large number of unnecessary integrals (still present due to our current use of a standard integral package), will result in considerable time savings, within the same linear-scaling framework.

As an alternative, we present here a new approach to reduce the distance criterion required to obtain accurate ab initio quality results within the ADMA method. In this approach, the distant atoms of the target molecule, not included in the parent molecules, are represented in the quantum chemical calculations of the parent molecules as point charges. In this way, the interactions between a given fragment and distant atoms, not included in the corresponding parent molecule, are not completely ignored but are included in an approximate fashion. As will be shown, this drastically increases the accuracy of the ADMA approach especially for smaller fragment surroundings (small distance criterion for parent molecules) and, therefore, decreases the size of the surroundings needed to be included in the calculations for a desired accuracy of ab initio quality.

In this paper the arguments and detailed error analyses justifying the original linear-scaling ADMA method will not be repeated; the readers are directed to references.[1−8,19,20]

The paper is organized as follows: Section II recalls the relevant aspects of the original ADMA method and outlines the new additions. This includes the definitions of the various types of point charges used. In section III, a detailed evaluation of the various parameters is given using a set of actual molecules with a sufficient range of structural variations to provide confidence for actual protein calculations, and also having sizes suitable for the large number of conventional computations, as well as having sufficient complexity to serve as tests for the ADMA and FA-ADMA methods. These small test molecules have also been already used as test cases in earlier publications.[19,20] In section IV, the results for larger protein systems are summarized. In particular, among the results both the relative and absolute energies and their errors are significant, not only in placing the power of the method in proper perspective but also to serve as the basis for numerical comparisons in future methodological developments. In the final section conclusions are drawn and perspectives are given.

II. Methodology

The additive fuzzy density fragmentation (AFDF) principle[1−8] and the adjustable density matrix assembler (ADMA) method[5−8,19,20] provide effective computational tools for the calculation and analysis of macromolecular electron densities and other molecular properties. Using the conventional Hartree−Fock−Roothaan−Hall formalism,[21] these MO-based, additive fuzzy density fragmentation methods avoid artificial fragment boundaries and provide local molecular fragments fully analogous to complete molecules. Whereas the fragments are artificial, they do not have boundaries; hence, these fragments follow closely the way electron density diminishes with distance in real molecules. This is an advantage in contrast to some other possible fragmentation schemes of subdivision where artificial boundary surfaces occur, since such fragments with boundaries, if they originate from different molecules, can never match perfectly, as follows directly from the holographic electron density theorem.[22] Consequently, either density gaps or local density doubling (that is, locally −100% or +100% error) must necessarily occur, precisely in the critical bonding region between fragments, if one attempts to use such fragments with boundaries from smaller molecules to build models for larger molecules. By contrast, due to the fuzzy fragmentation in the AFDF approaches, neither gaps nor doubling can occur in the density, and the overall errors of the method can be made as small as desired, simply by increasing the distance criterion. The mathematical background as well as a large number of tests and applications especially for the primary molecular property of electron density are well documented in the literature,[1−4,19,20,23,24] and thus only a short description of the method is given here.

The electron density $\rho(\vec{r})$ of a molecule can be expressed in terms of a basis set of n atomic orbitals $\varphi_i(\vec{r})$ ($i = 1, 2, ..., n$) used for the expansion of the molecular wave function and the density matrix $\mathbf{P}$ of elements P_{ij} determined for the given nuclear configuration using the specified basis set:

$$\rho(\vec{r}) = \sum_{i=1}^{n}\sum_{j=1}^{n} P_{ij}\,\varphi_i(\vec{r})\,\varphi_j(\vec{r}) \tag{1}$$

Following the additive fuzzy density fragmentation (AFDF) principle,[1−8] the first step in the generation of local fuzzy electron density fragments of the macromolecule under study, referred to as the 'target' molecule, is to subdivide the set of nuclei of the molecule into a set of mutually exclusive families of nuclei denoted by f^k, $k = 1, ..., m$.

The next step is the generation of m "parent" molecules. At its central region, each of these parent molecules contains one of the nuclear families f^k with the same local nuclear geometry as in the target macromolecule, and furthermore, some additional nuclear families within a selected distance from the surroundings of family f^k, with the same local arrangement as in the target macromolecule. In addition, at the periphery of the parent molecule some additional atoms, usually hydrogens, are added, in order to avoid "dangling bonds".

The fragment density matrix of the nuclear family f^k is then defined according to the Mulliken−Mezey scheme[5−8] as

$$P_{ij}^{k} = \begin{cases} P_{ij} \text{ if both of } \varphi_i(\vec{r}) \text{ and } \varphi_j(\vec{r}) \text{ are centered on a nucleus of } f^k \\ 0.5P_{ij} \text{ if precisely one of } \varphi_i(\vec{r}) \text{ or } \varphi_j(\vec{r}) \text{ is centered on a nucleus of } f^k \\ 0 \text{ otherwise} \end{cases} \tag{2}$$

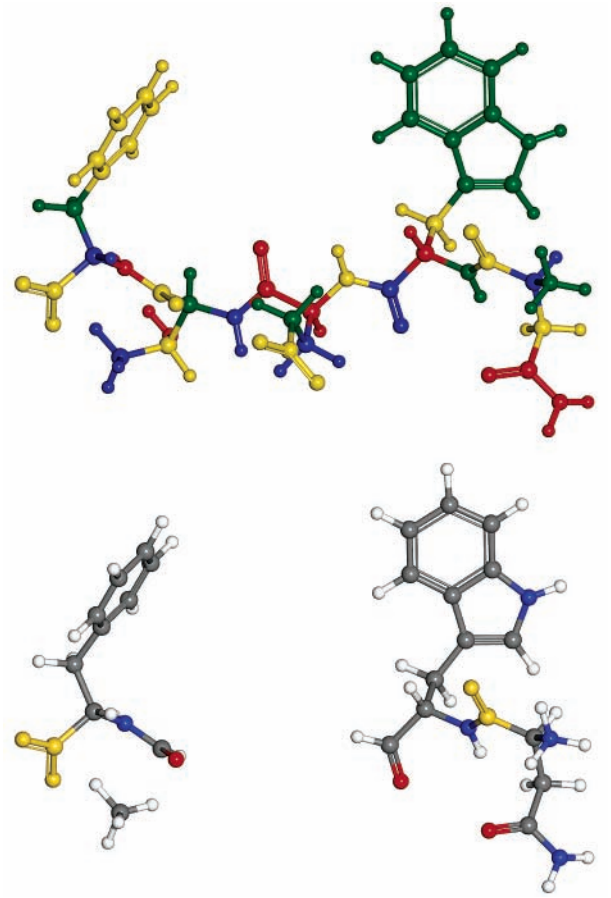

Figure 1. Two representative fragments with their 4 Å surroundings. In the upper part of the figure, the molecule is shown with the atoms color-coded according to the assignment to the fragments. In this case, the molecule was divided into 27 fragments. In the lower part, two of these fragments and their 4 Å surroundings are shown. The fragment on the left-hand side is a CO_2 group, and the fragment on the right-hand side is a CH group, both shown in yellow. The atoms of the surroundings are color-coded by atom type. As can be seen in the parent molecule for the CO_2 group, atoms less than 4 Å away from the central nuclear family but not bonded covalently to the central part within the range of the distance criterion are also included in the surroundings and hence in the parent molecule, so a given "parent molecule" may actually become a pair of molecules. Also, the hydrogen atoms filling remaining missing valences can be seen.

Within each parent molecule, this scheme gives an exactly additive decomposition of the complete density matrix of the parent molecule.

For the generation of the parent molecules and the calculation of the fragment density matrices, an automated procedure is used that was outlined in earlier publications.[19,20] In Figure 1, two representative fragments plus the surrounding are shown to illustrate the fragmentation scheme used. The information on the fragments plus surroundings are saved in pdb format and converted to a z-matrix file, which is suitable as input for a Hartree−Fock self-consistent field calculation using the Gaussian 98[25] or some similar program. The quantum chemical calculations for the parent molecules of all fragments are then initiated automatically.

A good approximation of the density matrix of the target molecule can be obtained to any desired accuracy if fragment density matrices are taken from the set of small "parent" molecules where the accuracy can be controlled by the distance criterion, that is, by the size of these parent molecules.

With the i and j indices redefined and expressed for the macromolecular list of orbitals, built from the same AO basis as those of the parent molecules, the total density matrix of the target macromolecule can be expressed as

$$P_{ij} = \sum_{k=1}^{m} P_{ij}^{k} \tag{3}$$

Note that from each parent molecule only the fragment density matrix corresponding to the central fragment is used in further calculations; hence, it is accurate within the limits set by the distance criterion chosen for the surroundings within the parent molecule. To use this approximation of the density matrix in further calculations an additional criterion should be fulfilled: the integration of the electron density over the total space must result in the total number of electrons. Due to the combination of fragment density matrices, this condition is not always exactly fulfilled. In order to remove this error, each density matrix element is multiplied by the quotient of the real number of electrons and the number obtained by the integration of the fragment electron density:

$$P_{ij}^{\text{norm}} = P_{ij} \frac{n_{\text{el}}}{\int \rho(\vec{r})\, \mathrm{d}\vec{r}} \tag{4}$$

Within the ADMA approach, the total energy of the target molecule can be calculated following the standard Hartree−Fock formalism using the ADMA approximation of the total density matrix instead of the ideal, directly calculated (and for most large molecules still unattainable) exact macromolecular density matrix:

$$E_{\text{HF}} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}((F_{ij} + H_{ij}^{\text{core}}) \sum_{k=1}^{m} P_{ij}^{k}) + V_{\text{NN}} \tag{5}$$

with

$$F_{ij} = H_{ij}^{\text{core}} + \sum_{r=1}^{b}\sum_{s=1}^{b}\left(\left[(ij|rs) - \frac{1}{2}(is|rj)\right]\sum_{k=1}^{m} P_{ij}^{k}\right)$$

The accuracy of this approach depends only on the reproducibility of the local surroundings of each macromolecular fragment density within the parent molecules, which can be improved to any desired accuracy by increasing the distance criterion, that is, the size of the parent molecules. It has been shown that large surroundings of more than 10 Å radius are needed to achieve truly high-quality results, which differ from the results of the standard Hartree−Fock method by 1 kcal/mol (0.001 56 hartree) or less.[19] Calculations with these large surroundings are time consuming and are not very efficient for large molecules especially in energy minimization algorithms. Therefore, other methods, besides simply increasing the size of the surroundings, must be developed to include, at least approximately, additional parts of the target molecule in the calculations of the parent molecules.

One possibility is to use only smaller surroundings in a full quantum chemical manner and include further surroundings as point charges. This can be done, on the one hand, using partial charges parametrized for use in a molecular mechanics force field analogous to mixed quantum mechanics/molecular mechanics (QM/MM) methods.[26−31] On the other hand, the ADMA approach is suitable to calculate ab initio quality partial charges according to the definition of Mulliken[32−35] or Löwdin.[36,37] With the use of these charges, a scheme, which we will call the field-adapted ADMA approach or, in short, FA-ADMA, can be

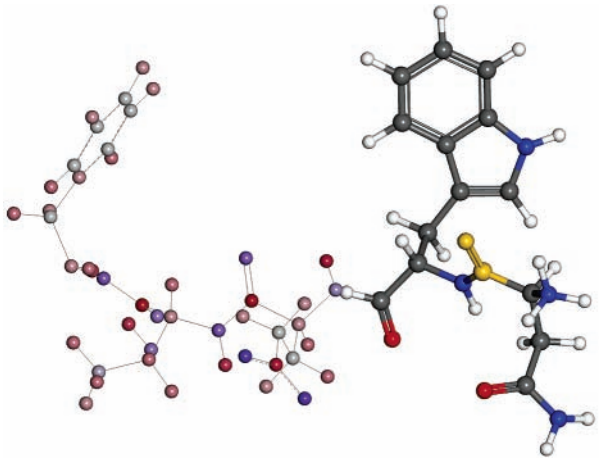

Figure 2. Partial charge surroundings used in the calculation of the fragment shown on the right-hand side of Figure 1. The fragment and the surroundings included in the quantum chemical calculation are shown in the same representation as in Figure 1 with the fragment in yellow and the surroundings color-coded by atom types. The partial charges are depicted as spheres color-coded by the sign of the charge (blue, negative; gray, neutral; and red, positive). The connections between the partial charges represent the bonding skeleton of the target molecule.

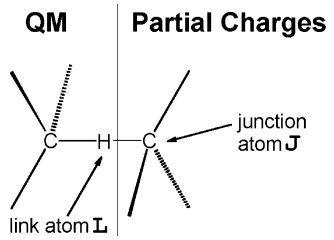

Figure 3. The link atom for the case of a C−C bond, in which the carbon on the left-hand side is treated quantum chemically. The carbon atom **J** (junction atom) on the right-hand side is substituted in the parent molecule by a hydrogen atom **L** (link atom) and is included only as a partial charge.

developed, consistent with the original ADMA approach and not dependent on any empirical parametrization. However, these charges are not known a priori, and they must be calculated in a self-consistent fashion. First, a standard ADMA calculation is performed. This is then used for the calculation of the partial charges of all atoms of the target molecule. These partial charges are then used to represent distant parts of the macromolecule. The partial charges of all atoms of the target molecule, which are not included in one specific parent molecule, are incorporated in the quantum chemical calculation of this parent molecule (see Figure 2). From the assembled new results for the macromolecule, the partial charges of all atoms of the target molecule are recalculated. These steps are repeated, and the atomic charges and fragment density matrices are optimized iteratively by additional ADMA calculations. The iterative cycle is ended when the difference between the partial charges in two successive iterations is less than a given threshold. The converged macromolecular density matrix is used for the calculation of the macromolecular Hartree−Fock energy and possibly other properties.

One problem of the fragmentation scheme (and many others used in mixed QM/MM methods[26−31]) in concern of including partial charges is how to deal with the atoms (junction atoms, **J**) of the target molecule substituted by hydrogen atoms (link atoms, **L**) in the parent molecules (see the illustration in Figure 3). These atoms are already partly included in the quantum chemical calculations due to the substitution but must be also included as partial charges because of the differences in electronegativity. To circumvent this problem, a number of different approaches were explored. These included the use of the partial charge of atom **J**, of the partial charge of **J** scaled by a certain factor, and of no charge on atom **J** at all. In all of these approaches the additional point charges were placed on the center of the substituted atoms.

In the next section, the various approaches outlined above are compared using small test cases, and the optimal treatment of the junction atom is identified. This optimal model is then used in section IV for the actual calculations of ab initio quality energies for proteins.

III. Test Calculations on Small Oligopeptides

The various approaches to introduce point charges in ADMA calculations were first tested on one hexapeptide (Asp-Tyr-Gln-Arg-Leu-Asn) and one pentapeptide (Asn-Trp-Glu-Thr-Phe) already used in earlier publications.[19,20] These test cases provide a sufficient variety of structural features for amino acids; they are large enough so that the accuracy of the approximation can be examined but also small enough to compare the results easily to direct quantum chemical calculations.

With these calculations, our aim was to find out, which model, that is, which scaling factor of the partial charges of the junction atoms **J**, gives the best overall accuracy. Therefore, the results for the FA-ADMA method were first compared to the direct calculations using Gaussian 98[25] and also to the original ADMA method using the STO-3G basis set.[38,39]

As described in the Methodology, several FA-ADMA approaches have been tested, using either the partial charge of **J**, a certain percentage of the partial charge of **J** (10% to 90% in 10% increments), or no charge at all. The results for three different sizes for surroundings, that is, for three distance criteria (3, 4, and 5 Å) are illustrated in Figures 4 and 5 using a graphical representation for the first and second peptides, respectively. For comparison, the energies obtained with the direct Hartree−Fock method and the original ADMA approach are also shown in these figures. The comparison of the numerical values in a tabular form can be obtained as Supporting Information.

In both figures, it can be seen that the energies using the scaled partial charges for the border atoms are lower than the energies of the original ADMA approach and run through a minimum at a value between 60% and 80%. Therefore, neither the total partial charges nor the case of no partial charges on the border atoms is the ideal choice but these partial charges should be scaled by an empirical factor. It can also be seen that the FA-ADMA version using Löwdin charges is much less dependent on the actual value of this scaling factor than the one using Mulliken charges. This can be explained by the larger values of the Mulliken charges compared to Löwdin charges. While comparing the results for the various surroundings, it was noticed that the absolute improvements due to the inclusion of partial charges decrease with increasing size of the surroundings (as expected for the exact convergence of the original ADMA method with the distance criterion), but this lesser significance is more than compensated by the large improvements resulting from the increase of the surroundings. In this sense, the ADMA calculations with surroundings of 3 and 4 Å can only be used as rough approximations even if partial charges are included. For these small surroundings the inclusion of partial charges cannot compensate for the loss of flexibility in the basis set due to the small number of atoms and, therefore, to the fewer basis functions in the parent molecules. On the other hand, the FA-ADMA calculations using surroundings of 5 Å reproduce

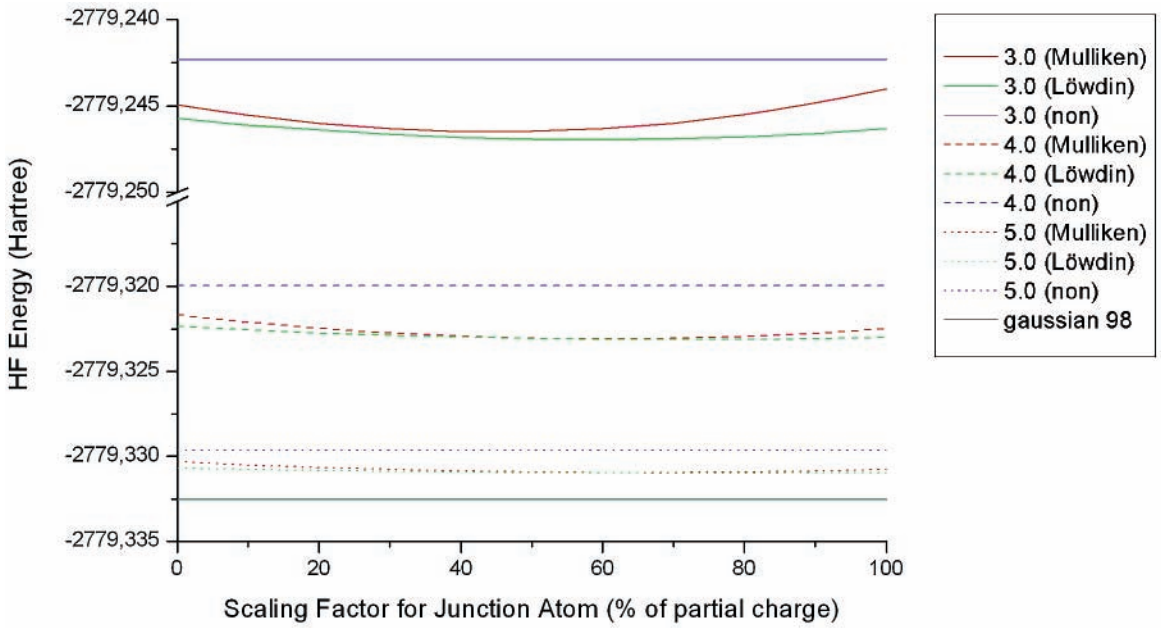

Figure 4. Graphical comparison of results of two versions of the new FA-ADMA approach using Mulliken charges and Löwdin charges, the original ADMA approach (no point charges, indicated by "non"), and conventional Hartree−Fock calculations for peptide 1 (Asp-Tyr-Gln-Arg-Leu-Asn). The STO-3G basis set and three different distance parameters (3.0, 4.0, and 5.0 Å) were used. The Hartree−Fock energy is given in atomic units, and the *x*-axis corresponds to the percentage of the partial charge of the junction atom.

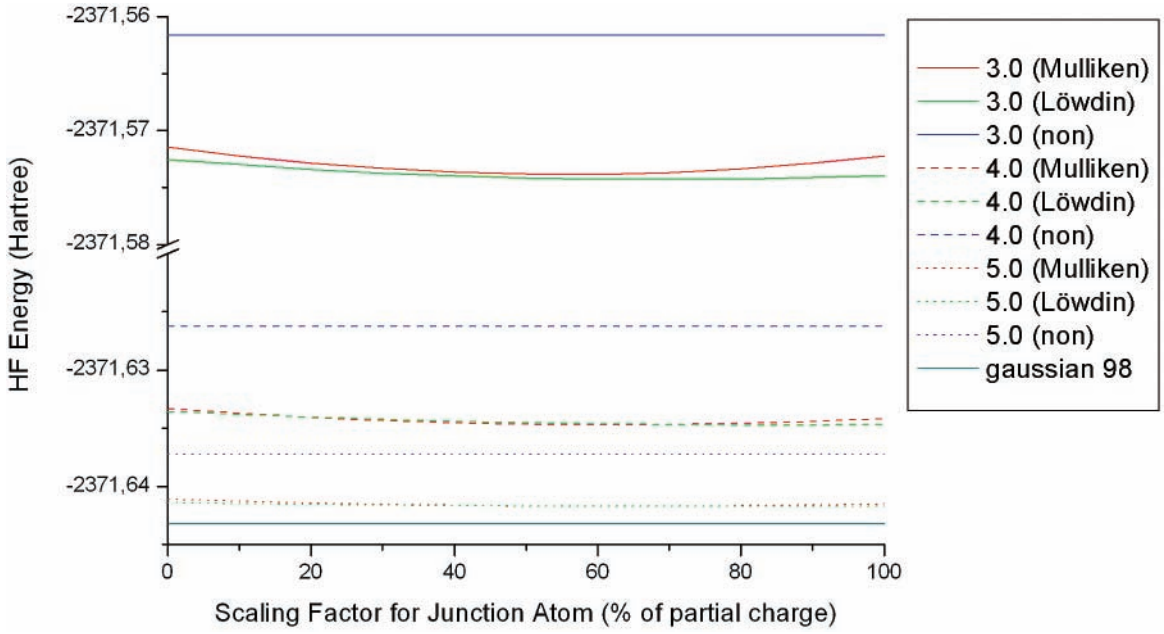

Figure 5. Graphical comparison of results of two versions of the new FA-ADMA approach using Mulliken charges and Löwdin charges, the original ADMA approach (no point charges, indicated by "non"), and conventional Hartree−Fock calculations for peptide 2 (Asn-Trp-Glu-Thr-Phe). The STO-3G basis set and three different distance parameters (3.0, 4.0, and 5.0 Å) were used. The Hartree−Fock energy is given in atomic units, and the *x*-axis corresponds to the percentage of the partial charge of the junction atom.

the energies of the conventional Hartree−Fock calculations to within a few millihartree as well as the results of the original ADMA approach obtained with much larger surroundings, as will be shown in the following section. This also demonstrates that inclusion of partial charges can indeed improve the results of ADMA calculations and reduce the size of surroundings needed to reach a required accuracy.

In order to show that the essential conclusions from the results obtained until now are valid for each basis set, we performed additional calculations for the test peptides using the 6-31G** basis set.[40−44] The results are included in the Supporting Information. All these additional calculations confirm the findings stated so far. The only difference for this larger basis set is that in the case of the Mulliken charges, the results using small charge scaling percentages (<40%) of the border atoms are worse than the results of the original ADMA approach, but the new approach still gives an improvement in the region between 60% and 90% for charge scaling.

IV. Protein Studies

As was shown in the previous section using the small peptide examples, reliable improvements can be obtained in comparison to the original ADMA approach if partial charges based on the

TABLE 1: Results of the FA-ADMA Approach for 16 Proteins, Selected as Test Cases for the Evaluation of the Methodology[a]

molecule	energy using the direct Hartree−Fock method (hartree)	energy using the original ADMA approach (hartree)	energy using Mulliken charges (hartree)	energy using Löwdin charges (hartree)
α-conotoxin pnib from *Conus pennaceus* (1akg)	−6747.33250543	−6747.32169419 (0.01081124)	−6747.32932632 (0.00317911)	−6747.32944709 (0.00305834)
charybdotoxin (1cmr)	−13868.5318985	−13868.5133924 (0.0185061)	−13868.5257477 (0.0061508)	−13868.5261723 (0.0057262)
α-conotoxin imi (1cnl)	−5764.91797185	−5764.91376105 (0.00421080)	−5764.91636099 (0.00161086)	−5764.91636150 (0.00161035)
crambin (1cnr)	−17775.2218029	−17775.2040419 (0.0177610)	−17775.2139234 (0.0078795)	−17775.2145941 (0.0072088)
human endothelin-1 (1edn)	−9839.46446225	−9839.45481533 (0.00964692)	−9839.46105919 (0.00340306)	−9839.46125501 (0.00320724)
epidermal growth factor subdomain of human thrombomodulin (1fgd)	−7512.31620690	−7512.29206004 (0.02414686)	−7512.31323944 (0.00296746)	−7512.31339153 (0.00281537)
μ-conotoxin giiib (1gib)	−9935.89211088	−9935.87370430 (0.01840658)	−9935.88804847 (0.00406241)	−9935.88817200 (0.00393888)
gramicidin A ion channel (1grm)	−12227.0526450	−12227.0476126 (0.0050324)	−12227.0470165 (0.0056285)	−12227.0474716 (0.0051734)
bovine lactoferricin (1lfc)	−11198.8742803	−11198.8479174 (0.0263629)	−11198.8704567 (0.0038236)	−11198.8706841 (0.0035962)
α-conotoxin mii (1m2c)	−7019.17869883	−7019.16974461 (0.00895422)	−7019.17587875 (0.00282008)	−7019.17608963 (0.00260920)
ω-conotoxin mviia (1omg)	−10985.5137301	−10985.4935835 (0.0201466)	−10985.5090142 (0.0047159)	−10985.5092023 (0.0045278)
α-conotoxin pni1 (1pen)	−6723.33774734	−6723.32773135 (0.01001599)	−6723.33532953 (0.00241781)	−6723.33543146 (0.00231588)
rp 71955 (tricyclic peptide active against HIV-1) (1rpb)	−8432.14950919	−8432.14416581 (0.00534338)	−8432.14511484 (0.00439435)	−8432.14544276 (0.00406643)
tertiapin (1ter)	−9578.80398501	−9578.77796648 (0.02601853)	−9578.80017971 (0.00380530)	−9578.80034551 (0.00363950)
vacuolar targeting peptide from Na-propi (1vtp)	−9885.60187070	−9885.57878673 (0.02308397)	−9885.59849984 (0.00337086)	−9885.59862187 (0.00324883)
trypsin inhibitor II (EETI II) (2eti)	−11845.1703595	−11845.1574208 (0.0129387)	−11845.1666788 (0.0036807)	−11845.1669950 (0.0033645)

[a] The STO-3G basis set and surroundings of 6 Å were used. For the ADMA calculations, the errors compared to the direct calculations are given in brackets. The scaling factor for the Mulliken charges and for the Löwdin charges of the border atoms **J** were set to 60% and 80%, respectively.

definition of Mulliken or Löwdin are used for distant atoms in the calculations for each parent molecule. Such charges are placed at the location of the nucleus of each atom of the target molecule not explicitly included in the quantum chemical calculation of a specific parent molecule. The partial charges of the border atoms are scaled by a certain factor. Therefore, we will use this method for calculations of ab initio quality energies of proteins. Because there are no standardized test sets for very large molecular systems as there are for small molecules such as the G2[45,46] and G3[47,48] test sets for the validation of new quantum chemical methods, we decided to use structures taken from the protein data bank (PDB).[49] In this database thousands of experimentally determined protein structures are stored representing a wide variety of structural and functional motives found in living organisms. We therefore think that taking a random subset of these structures should produce a diverse test set not biased by the artificial generation of test structures. In all, we took 16 protein structures from the PDB,[49] which were selected for these test calculations due to their relatively small size, on the one hand, and also for their value in numerical comparisons of both absolute and relative energies in past and future studies of methodological studies (we have used one of them already in earlier publications[2,9,20]). These preliminary tests are used to get a feeling of the absolute errors introduced by the fragmentation scheme and the size dependence of these errors. Additional investigations to study the accuracy of the new method in describing energy differences between different conformations, protonation states, and complex structures are on their way.

For three of these proteins, α-conotoxin pnib from *Conus pennaceus* (1akg), α-conotoxin imi (1cnl), and the gramicidin A ion channel (1grm), studies were performed to determine the optimum value of the scaling factor. With the use of the STO-3G basis set,[38,39] the results for two different surroundings (5 and 6 Å) were calculated and are also provided as Supporting Information. The results for these proteins show the same trends as observed for the peptide examples. In all but one case, the energies using the scaled partial charges for the border atoms run through a minimum for some intermediate scaling factor leading to energies lower than those of the original ADMA approach. Only for gramicidin A, the original ADMA approach gives slightly better energies than the new FA-ADMA approach using partial charges. However, for the gramicidin A protein, the original ADMA approach already gives a much better result than for the other proteins chosen as test cases and examples, and what actually happens is that this good result can be reproduced also with the FA-ADMA approach. In all examples, small peptides as well as proteins, if one takes Mulliken charges, the minimum of the energy lies around a scaling factor of 60%. Therefore, this value was chosen in the studies of the other proteins. In the case of Löwdin charges, the optimum percent value varies more from example to example and it lies in the range between 70% and 90%. But, as mentioned before, the energies calculated using Löwdin charges are not as sensitive

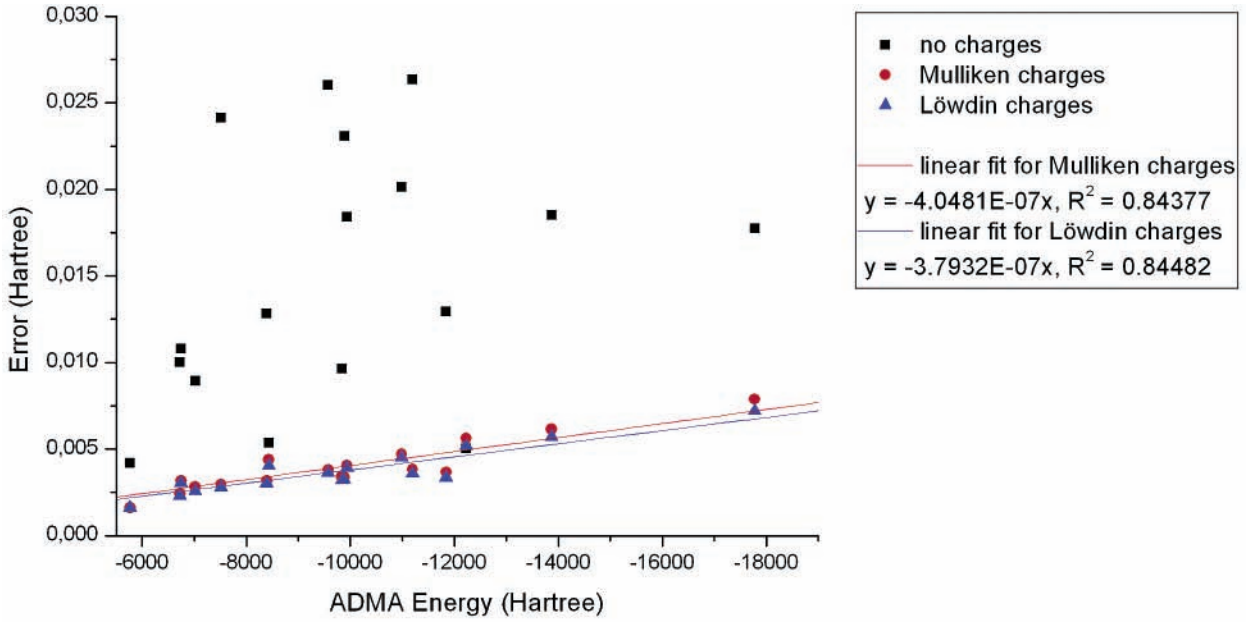

Figure 6. Correlation of the energy error of the FA-ADMA and the original ADMA approaches with the total energy (in atomic units).

to the actual value of the scaling factor as the ones using Mulliken charges so that almost identical results can be obtained with scaling factors lying in this range. Therefore, the value was set to 80% for the remaining calculations.

In Table 1, the results for all 16 proteins using the optimal scaling factors are summarized. These proteins have been selected as test cases, partly due to their relatively small size among proteins and also due to the fact that they represent a wide range of structural features and provide a variety of challenges for new computational methodologies. When the actual results are compared, it is evident that the inclusion of partial charges is more important for some proteins than for others. For example, there is no improvement at all in the case of the gramicidin A (1bdw) dimer, while the energy for the epidermal growth factor subdomain of human thrombomodulin (1fgd) decreases by 21 mhartree, and the error is reduced by almost 90%. This can be explained by the different charge distributions of the two proteins. The inclusion of partial charges seems most significant if highly polarized functional groups or formal charges appear in the molecule. If the molecule is mainly built from apolar groups, which is the case for the transmembrane protein gramicidin A, the partial charges will adopt small values and have only a minor influence on the electron density of the parent molecules. This small polarization is also responsible for the fact that very good approximations can be obtained for apolar molecules already with the original ADMA approach using smaller parent molecules. For polar and highly charged molecules the errors in the energies are much higher, and the inclusion of partial charges is needed to obtain results of the same accuracy. This is well demonstrated by the large error of the original ADMA approach in the case of the epidermal growth factor, which has a total charge of -6 and 8 formally charged groups.

To demonstrate this point from two different perspectives, in Figures 6 and 7 the energy errors of the new FA-ADMA and the original ADMA approaches (given in atomic units) are correlated with the total energy and the number of electrons in the molecule, respectively. These two figures show a high degree of similarity, but also important differences. Whereas the correlations are very poor for the original ADMA method, there are excellent correlations for the FA-ADMA approaches, for both total energy and electron count. Whereas for the FA-ADMA results patterns of points for the total energy and electron count are nearly identical, this is not the case for the ADMA method, especially for the range of larger errors, where the two figures differ considerably. Although the energy errors are remarkably small even for the original ADMA method, for this method neither the total energy nor the electron count can provide a reliable estimate for the energy error; on the basis of the two figures, these two approaches give different estimates, especially in the case of relatively large energy errors. By contrast, for FA-ADMA, both the total energy and the electron count give excellent error estimates, and they do this consistently, showing little difference between the two approaches.

Both of these diagrams can be exploited in further improving the accuracy of the energy calculations. As expected, in the FA-ADMA approach, the errors increase linearly with the number of electrons and also with the total energy, due to their extensive nature. In this way, the good results of the FA-ADMA approach can be further improved by the use of an empirical scaling to compensate for the errors introduced by the fragmentation scheme.

Overall, by including partial charges, the mean absolute error for the 16 proteins could be reduced from 15.08 mhartree for the original ADMA approach to 3.99 mhartree for Mulliken charges and 3.76 mhartree for Löwdin charges. With the use of the correlations from the Figures 6 and 7 for an empirical scaling, the errors can be further reduced. In this way, the mean absolute error with Mulliken charges is only 0.49 mhartree and 0.53 mhartree using the correlation with the total energy and the number of electrons, respectively. The errors using Löwdin charges are 0.44 mhartree and 0.50 mhartree, respectively. One should note that these two approaches lead to similar, but not identical, results. If one intends to use these empirical scaling factors reliably, they must be tested further in future investigations with additional examples and especially with various conformations of the same protein, establishing which correlation gives the better results.

Finally, we will discuss the results of crambin (1cnr) in more detail since this is the protein that has been the subject of the most quantum chemical studies, and the performance of the new method can be assessed by comparisons with earlier ADMA

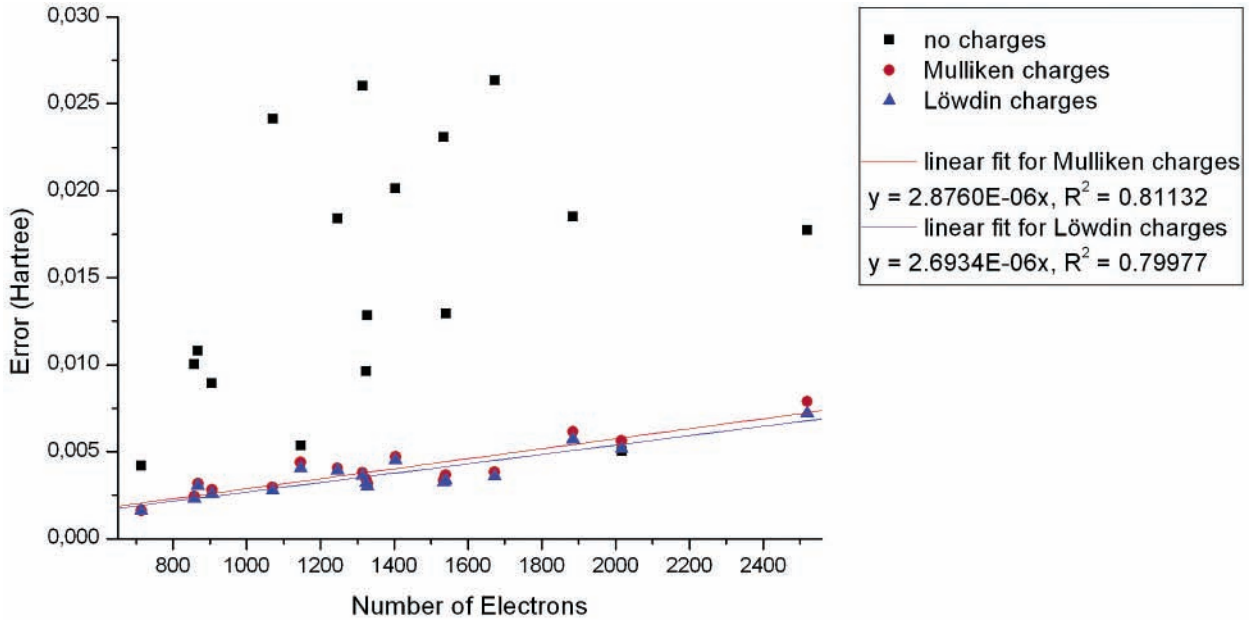

Figure 7. Correlation of the energy error of the FA-ADMA and the original ADMA approaches (in atomic units) with the number of electrons in the molecule.

results. We used this molecule in an earlier publication[19] to study in detail the influence of the size of the surroundings on the accuracy of the original ADMA approach. By comparing these earlier results with present ones, we can conclude that the inclusion of the partial charges using the FA-ADMA approach with surroundings of 6 Å leads to the same accuracy that can be obtained only with surroundings of 7−8 Å in the original ADMA approach. Furthermore, using the empirical scaling of the FA-ADMA method, the results surpass the accuracy of the original ADMA approach with the largest surroundings of 12 Å used in the earlier study.[19] The additional computation time required for the point charges in the FA-ADMA method is negligible when compared to the quantum chemical calculations of the parent molecules; however, due to the need for self-consistent calculations of the partial charges, that is achieved by iterations, a number of additional calculations are required in the FA-ADMA calculations, each approximately equivalent to one full ADMA computation with small parent molecules. In most cases of the present study on an extensive set of peptides and proteins, four cycles have been found to be sufficient to reach self-consistency. By taking the computer time requirement of four ordinary ADMA computations with a surroundings smaller by 1.5 Å, the overall performance of the FA-ADMA approach is better than that of the original ADMA approach for a given accuracy. This is even more significant if larger basis sets are used, due to the O(N^4) scaling behavior of the standard Hartree−Fock method.

V. Conclusion

In this paper, we have described the introduction of point charges into the ADMA approach resulting in a new field-adapted ADMA approach. In this method, a macromolecule, the target molecule, is first divided into fuzzy fragments, for which conventional quantum chemical calculations are performed using moderate-sized "parent molecules" which contain all the local interactions of the fragment within a preselected distance. The other parts of the target molecule not included in a specific parent molecule are approximated as point charges in the quantum chemical calculation of this parent molecule. It was shown that atom-centered partial charges based on the definitions of Mulliken[32−35] or Löwdin[36,37] are well suited to improve the results of the ADMA approach for a given value of the distance criterion up to which the local surroundings of each fragment are explicitly included in the quantum chemical calculation. In this way, a smaller size of the surroundings, that is, smaller parent molecules, can be used in FA-ADMA calculations to obtain a certain degree of accuracy, which results in a much improved performance.

As these charges are not known a priori, a self-consistent method was developed. First, a standard ADMA calculation is performed, which is used to calculate the partial charges of the target molecule. These partial charges are then optimized by a cycle of additional ADMA calculations, in which the partial charges are incorporated in the quantum chemical calculation of the parent molecules. One problem in using partial charges, also seen in mixed QM/MM approaches,[26−31] is how to deal with junction atoms, which are at the border between the quantum system and the system included as point charges. These atoms are substituted with hydrogen atoms in the quantum system and are, therefore, partly included in the quantum chemical calculations. The test calculations presented in this paper show that the best results can be obtained if the partial charges of these atoms are scaled by a certain factor. The factor was empirically determined to 60% and 80% of the original partial charge for Mulliken and Löwdin charges, respectively. To further improve these results, and to build on experiences obtained considering smaller molecules,[50,51] investigations are on their way to use different scaling factors as well as alternative link atoms (like pseudohalogen atoms used in QM/MM methods) for different kinds of junction atoms.

A number of test calculations were performed first using small oligopeptides to evaluate the method and then using 16 protein structures taken from the protein data bank.[49] These tests show that the inclusion of the partial charges results in a large improvement of the calculated energies especially for highly polarized and formally charged molecules. This can be explained by the long-range behavior of the electrostatic interactions, which are partially neglected in the original ADMA approach but are better approximated in the new one. Therefore, we think that on the one hand the treatment of the parts of the target

molecule not included in a specific parent molecule as partial charges is a very important step in the development of a fast and reliable ADMA program, which can be used as a standard quantum chemical application. Additionally, partial charges could also be of benefit for other approaches, which divide a large molecule into fragments such as the divide-and-conquer[10,11] or the "molecular tailoring" approach.[12,13] On the other hand, we think that it is very important to include highly polarized and charged molecules in a test set to evaluate a given fragmentation scheme. For these systems, the original ADMA approach as well as most alternative methods give considerable errors in the calculated energies, even if the computed electron densities are reasonable. For mainly apolar molecules, these errors are much smaller, and the inclusion of partial charges gives no or almost no actual improvement. But for an off-the-shelf method, which can be applied even by nonexperts to a large number of different systems, the results for all these systems should be within a comparable accuracy. This can only be guaranteed if the test set is diverse enough to cover examples from all the different regions of the application space. Although we concentrate on proteins in this paper, we expect that our test set of molecules provides a wide range of variety and it complies with the above condition. Based on the peptide and protein results, we expect that the FA-ADMA macromolecular quantum chemistry approach can be used efficiently for a wide variety of other large molecules.

Acknowledgment. We sincerely thank Professor J. Brickmann for the supply of the MOLCAD II module and the Natural Sciences and Engineering Research Council of Canada, the Canada Foundation for Innovation, as well as the Albert Szent-Györgyi Project of Hungary for the financial support of this work. Thomas Exner gratefully acknowledges the Alexander von Humboldt Foundation for a Feodor Lynen research fellowship.

Supporting Information Available: Comparisons of the results of the FA-ADMA approach with results of direct Hartree−Fock calculations and the original ADMA approach for the two test peptides and three proteins, in both tabular and graphical forms. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Walker, P. D.; Mezey, P. G. *J. Am. Chem. Soc.* **1993**, *115*, 12423.
(2) Walker, P. D.; Mezey, P. G. *J. Am. Chem. Soc.* **1994**, *116*, 12022.
(3) Walker, P. D.; Mezey, P. G. *Can. J. Chem.* **1994**, *72*, 2531.
(4) Walker, P. D.; Mezey, P. G. *J. Math. Chem.* **1995**, *17*, 203.
(5) Mezey, P. G. *J. Math. Chem.* **1995**, *18*, 141.
(6) Mezey, P. G. *Adv. Quantum Chem.* **1996**, *27*, 163.
(7) Mezey, P. G. *Int. J. Quantum Chem.* **1997**, *63*, 39.
(8) Mezey, P. G. *Int. Rev. Phys. Chem.* **1997**, *16*, 361.
(9) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085.
(10) Yang, W. *J. Mol. Struct. (THEOCHEM)* **1992**, *255*, 461.
(11) St-Amant, A. A Gaussian Implementation of Yang's Divide-and-Conquer Density-Functional Theory Approach. In *Chemical Applications of Density-Functional Theory*; Laird, B. B., Ross, R. B., Ziegler, T., Eds.; ACS Symposium Series 629; American Chemical Society: Washington, DC, 2002; pp 70−84.
(12) Babu, K.; Gadre, S. R. *J. Comput. Chem.* **2003**, *24*, 484.
(13) Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. *J. Phys. Chem.* **1994**, *98*, 9165.
(14) Bowler, D. R.; Bush, I. J.; Gillan, M. J. *Int. J. Quantum Chem.* **2000**, *77*, 831.
(15) Gogonea, V.; Merz, K. M., Jr. *J. Phys. Chem. A* **1999**, *103*, 5171.
(16) Li, X.-P.; Nunes, R. W.; Vanderbilt, D. *Phys. Rev. B* **1993**, *47*, 10891.
(17) Scuseria, G. E.; Ayala, P. Y. *J. Chem. Phys.* **1999**, *111*, 8330.
(18) Friesner, R. A.; Murphy, R. B.; Beachy, M. D.; Ringnalda, M. N.; Pollard, W. T.; Dunietz, B. D.; Cao, Y. *J. Phys. Chem. A* **1999**, *103*, 1913.
(19) Exner, T. E.; Mezey, P. G. *J. Comput. Chem.* **2003**, *24*, 1980.
(20) Exner, T. E.; Mezey, P. G. *J. Phys. Chem. A* **2002**, *106*, 11791.
(21) Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69.
(22) Mezey, P. G. *Mol. Phys.* **1999**, *96*, 169.
(23) Mezey, P. G.; Walker, P. D. *Drug Discovery Today* **1997**, *2*, 6.
(24) Mezey, P. G. *J. Comput. Chem.* **1998**, *19*, 1344.
(25) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.9; Gaussian, Inc.: Pittsburgh, PA, 1998.
(26) Gao, J. Methods and Applications of Combined Quantum Mechanical and Molecular Mechanical Potentials. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1995; Vol. 7, pp 119−185.
(27) Amara, P.; Field, M. J. Combined Quantum Mechanical and Molecular Mechanical Potentials. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. C., Clark, T., Gasteiger, J., Kollmann, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, UK, 1998; pp 431−437.
(28) Ruiz-López, M. F.; Rivail, J.-L. Combined Quantum Mechanical and Molecular Mechanical Approaches to Chemical and Biochemical Reactivity. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. C., Clark, T., Gasteiger, J., Kollmann, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, UK, 1998; pp 437−448.
(29) Gao, J. Hybrid Quantum Mechanical/Molecular mechanical (QM/MM) Methods. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. C., Clark, T., Gasteiger, J., Kollmann, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, UK, 1998; pp 1257−1263.
(30) Metz, K. M., Jr.; Stanton, R. V. Quantum Mechanical/Molecular Mechanical (QM/MM) Coupled Potentials. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. C., Clark, T., Gasteiger, J., Kollmann, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, UK, 1998; pp 2330−2343.
(31) Tomasi, J.; Pomelli, C. S. Quantum Mechanics/Molecular Mechanics (QM/MM). In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. C., Clark, T., Gasteiger, J., Kollmann, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, UK, 1998; pp 2343−2350.
(32) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.
(33) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1841.
(34) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 2338.
(35) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 2343.
(36) Löwdin, P.-O. *Adv. Quantum Chem.* **1970**, *5*, 185.
(37) Cusachs, L. C.; Politzer, P. *Chem. Phys. Lett.* **1968**, *1*, 529.
(38) Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969**, *51*, 2657.
(39) Collins, J. B.; Schleyer, P. v. R.; Binkley, J. S.; Pople, J. A. *J. Chem. Phys.* **1976**, *64*, 5142.
(40) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724.
(41) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.
(42) Hariharan, P. C.; Pople, J. A. *Mol. Phys.* **1974**, *27*, 209.
(43) Gordon, M. S. *Phys. Lett.* **1980**, *76*, 163.
(44) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.
(45) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221.
(46) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063.
(47) Curtiss, L. A.; Raghavachari, K.; Redfern, P C.; Pople, J A. *J. Chem. Phys.* **2000**, *112*, 7374.
(48) Curtiss, L. A.; Redfern, P. C.; Rassolov, V.; Kedziora, G.; Pople, J. A. *J. Chem. Phys.* **2001**, *114*, 9287.
(49) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.
(50) Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290.
(51) Zhang, Y.; Lee, T.-S.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46.