JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

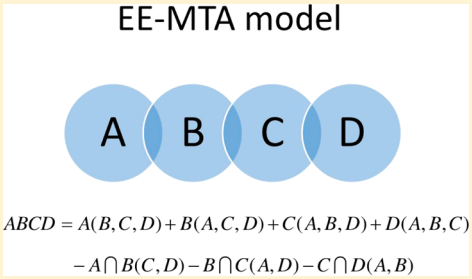
Electrostatically Embedded Molecular Tailoring Approach and Validation for Peptides

Miho Isegawa, Bo Wang, and Donald G. Truhlar*

Department of Chemistry, Chemical Theory Center, and Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455-0431, United States

Supporting Information

ABSTRACT: We add higher-order electronic polarization effects to the molecular tailoring approach (MTA) by embedding each fragment in background charges as in combined quantum mechanical and molecular mechanical (QM/MM) methods; the resulting method considered here is called electrostatically embedded MTA (EE-MTA). We compare EE-MTA to MTA for a test peptide, Ace-$(Ala)_{20}$-NMe, and we find that including background charges (embedding charges) greatly improves the performance. The fragmentation is performed on the basis of amino acids as monomers, and several sizes of fragment are tested. The fragments are capped with either hydrogen cap atoms or tuned fluorine cap atoms. The effective core potential of the tuned fluorine cap atom is optimized so as to reproduce the proton affinity for seven types of tetrapeptide, and the proton affinity calculated with the tuned cap atom shows a lower mean unsigned error than that obtained by using a hydrogen cap atom. In the application to the test peptide, these generically tuned cap atoms show better performance compared with the hydrogen cap atom for both the electronic energy and the energy difference between an α helix and a β sheet (in the latter case, 1.0% vs 2.7% when averaged over three sizes of fragments and two locations for cut bonds). Also, we compared the accuracy of several charge redistribution schemes, and we find that the results are not particularly sensitive to these choices for the Ace−$(Ala)_{20}$−NMe peptide. We also illustrate the dependence on the choice of charge model for the embedding charges, including both fixed embedding charges and embedding charges that depend on conformation.



1. INTRODUCTION

Fragment-based approaches constitute one of the most powerful classes of quantum chemistry tools for obtaining properties of a large system such as a protein. Although local density functionals (i.e., those with no Hartree−Fock exchange integrals or screened Hartree−Fock exchange and no nonlocal correlation contributions) have made it possible to compute the electronic energy of an undivided system including electron correlation with a computational cost having asymptotic scaling of N_e^3 (where N_e is the number of electrons), such calculations are still expensive to apply to large systems, and they are usually not as accurate as hybrid density functional calculations or post-Hartree−Fock correlated wave functional methods, both of which have higher scaling (N_e^4 or higher). In order to remove the N_e^3 dependency in the diagonalization procedure of the Kohn−Sham operator, Yang et al. have suggested a divide-and-conquer approach that makes it possible to avoid the N_e^3 procedure by partitioning the density matrix.[1−7] Other methods to reduce the scaling are based on recognizing that the most time-consuming step in the formation of the Kohn−Sham matrix is the calculation of two-electron integrals and approximating these with classical monopole−multipole or multipole−multipole interactions.[8−10] In the present article, we consider fragmentation approaches that are equally applicable to density functional theory and all kinds of wave function theory, from Hartree−Fock calculations to (for example) multireference coupled cluster theory at an arbitrary excitation level.

The molecular tailoring approach (MTA)[11−19] is the fragment-based approach under consideration here, and we simply draw the reader's attention to other fragment methods such as the fragment molecular orbital method (FMO),[20−28] the molecular fractionation with conjugate caps (MFCC) method,[29,30] the effective fragment potential (EFP),[31] the systematic molecular fragmentation (SMF) method,[32,33,34] the generalized energy-based fragmentation (GEBF)[35−37] method, the explicit polarization (X-Pol) method,[38,39] the kernel energy method (KEM),[40−42] the electrostatically embedded many-body (EE-MB) method,[43−46] the ternary interaction model,[47] the electrostatically embedded many-body expansion of the correlation energy (EE-MB-CE),[48,49] the molecules in molecules (MIM) method,[50] the multilevel fragment-based approach (MFBA),[51] the hybrid many-body interaction (HMBI) method,[52] and the many-overlapping-body expansion.[53] Some of these methods, e.g., EE-MB, MIM, and GEBF, show distinctive similarities to the method proposed here and deserve special emphasis for that reason.

The MTA is applicable to a wide range of systems from intermolecular interactions of small clusters to conformational

Received: September 29, 2012
Published: February 12, 2013

stability of proteins using any electronic structure method, for example, density functional theory, post-HF methods such as CCSD(T), and semiempirical methods including the density-functional-based tight binding (DFTB) method. In the MTA, the computational time is reduced proportionally to the number of CPUs one uses, so massive parallelization may be used to greatly reduce the computational time, and this is one of the chief advantages of the MTA.

Mahadevi et al.[18] have applied the MTA to benzene clusters with the MP2/6-31G++G** level of calculation for fragments, and they showed that the deviation of the electronic energy from that of full quantum mechanical (QM) calculations is within the range of chemical accuracy. Rahalkar el al.[16] applied the MTA to the alanine polypeptide, inorganic clusters including pure water, and a small protein, and they compared the performance with the FMO method. Their comparison showed that MTA provides stable accuracy with smaller dependences on the cluster or peptide size as compared with FMO method.

In order to increase the accuracy of MTA, one needs to consider several effects. The first is the combined effect of electronic polarization and long-range electrostatic interactions, which are not explicitly included in most of the previous MTA calculations, although the short-range and medium-range interactions within the fragment selected are included. For neutral species, this can lead to relatively accurate results; however, the errors may be larger in ionic or highly polar systems,[51] and it is important to consider the interfragment electrostatic interactions by including the background charges due to the other fragments. We will do that in the present paper by introducing the electrostatically embedded molecular tailoring approach (EE-MTA) in which the long-range electrostatic interactions are implicitly included on the basis of a procedure similar to that used in electrostatically embedded[54−56] combined QM/MM schemes.

The next consideration for possibly improving the accuracy is the type of cap atom (also called link atom) that is introduced at the fragment boundaries to satisfy the valence requirements at the bonds that are cut to make the fragments and to mimic the original electronic structure of the fragment. The most widely used cap-atom approach is the hydrogen cap atom. One of the advantages of the hydrogen cap atom is the small interaction with the rest of atoms in the QM region and other fragments, and this advantage is also seen in the localized orbital approach.[57] A more advanced method than the hydrogen cap atom is the tuned atom in which a pseudo- or effective core potential[58−63] is used, and parameters are adjusted to reproduce specific properties. We also mention the pseudobond method[58,59] and the quantum capping potentials method,[60] which have similar objectives. Recently, a general procedure was introduced[63] for tuning a fluorine cap atom for combined quantum mechanical and molecular mechanical (QM/MM) calculations. In this procedure, a pseudopotential is added to the fluorine atom, and the coefficient of the extra exponential function in the pseudopotential is optimized for each system. It was found that the tuned fluorine atom performs better than the general hydrogen cap atom, especially for the case that the partitioning into fragments occurs close to the region of the reaction center. This method could also be used in other contexts, such as the EE-MB method; here we apply it in the EE-MTA, but with a significant change. In particular, rather than tune the fluorine for each case, as in the original method, we attempt to derive general parameters suitable for use in all peptides.

The third consideration is the size of the fragments; the accuracy of the MTA or the EE-MTA should increase as one takes the sizes of fragments larger, although the quantitative effect of increasing fragment size depends on the structure of the molecule. However the computational cost becomes expensive for larger fragments, and one needs to carefully choose the size of the fragments as a compromise with the calculation cost.

The fourth consideration is the place where a molecule is partitioned; usually it is desirable to cut a nonpolar bond in which the electronegativity is similar between the two atoms involved in the bond. In the MTA and the EE-MTA, the energies of the boundary parts of the fragments are canceled against each other, so it is expected that significant differences due to the partitioning place will not be observed. This has been shown to be true for the MTA, and here we examine it for the EE-MTA.

The improvement in going from the MTA to the EE-MTA should be important for polar and charged systems. In addition, we examine four additional considerations: (i) the size of the fragments, (ii) the place of partitioning, (iii) the effect of the type of background charge, and (iv) the type of cap atom. With regard to the cap atoms, we compare the use of widely used hydrogen cap atoms to tuned fluorine cap atoms, where the parameter of the fluorine cap atom is determined so as to minimize the error of the proton affinity as compared to full QM calculations for seven tetrapeptides. The test systems employed here for EE-MTA are the α helix and β sheet conformations of alanine polypeptide, Ace−$(Ala)_{20}$−NMe, where Ace is an *N*-terminal acetyl group and NMe is a C-terminal *N*-methyl group, and the electronic energy and conformational energy are compared with those calculated by full QM calculations.

2. EE-MTA

2.1. EE-MTA with Coulomb Overcounting Term. The electronic energy of the whole peptide based on the EE-MTA is given by

$$E^{\text{EE-MTA}} = \sum E_I - \sum E_{I\cap J} + \sum E_{I\cap J\cap K} + \ldots + (-1)^{N-1} \sum E_{I\cap J\ldots\cap N} - E_{\text{OC}} \quad (1)$$

where E_I is the electronic energy of fragment I which is coupled with the electrostatic potential due to the rest of fragments; $E_{I\cap J}$ is the energy of the overlapping part of two fragments, I and J; the remaining terms prior to the last term involve overlap of three or more fragments; and the last term, E_{OC}, is the sum of the Coulomb interactions that were overcounted. The overcounting issue has been discussed elsewhere[33,35] and is further discussed below. In the present work, we truncate eq 1 to second order:

$$E^{\text{EE-MTA}} \approx \sum E_I - \sum E_{I\cap J} - E_{\text{OC}} \quad (2)$$

This formula is best understood by an example. Consider a chain polymer with 20 monomers, which will be fragmented into 16 overlapping pentamers: 1−2−3−4−5, 2−3−4−5−6, ..., 16−17−18−19−20. There are 15 overlaps, namely 2−3−4−5, 3−4−5−6, ..., 16−17−18−19. Since the overlaps are tetramers, such a calculation is labeled as a "pentamer−tetramer" fragmentation scheme, which denotes that pentamers are

stitched together (by the "tailor") using tetramers. The first sum in eq 2 is over the pentamers, and the second sum is over the tetramers.

The overcounted electrostatics term in eq 2 may be calculated in more than one way. The simplest way is given by

$$E_{\mathrm{OC}} = \sum_{k}^{N_f} \sum_{l=k+M}^{N_f} \sum_{i(\in k)} \sum_{j(\in l)} \frac{Q(i,k)Q(j,l)}{R(i,j)} \tag{3}$$

where i and j are labels for atomic sites, k is a fragment label, $Q(i, k)$ is the partial atomic charge on atom i in monomer k, and M is 5 when we use the pentamer−tetramer scheme. This term is rationalized as follows: Straightforward examination of the number of occurrences of each Coulomb interaction shows that Coulomb interactions between fragments that are never contained in a common fragment are counted twice. For example, label the pentamers as 1 through 16, and label the tetramers as 2 through 16 (the label is the number on the first fragment). Consider, as an illustration of the issue, Coulomb interactions of atoms in fragment 8 with those in fragment 11. These are counted eight times (pentamers 4−11) and subtracted seven times (tetramers 5−11), so their net count is one, which is correct. However, the Coulomb interactions between atoms in fragments 8 and 13 are counted 10 times (pentamers 4−13) and subtracted only eight times (tetramers 5−8 and 10−13), so they are overcounted. Therefore, we subtract all Coulomb interactions between point charges in fragment 8 and those in fragment 13. Similarly, we subtract all other overcounted Coulomb interactions, as indicated in eq 3.

The overcounting scheme just explained will be called the straight Coulomb overcounting scheme. For our final calcuations, we used a better method, but it will be easier to explain that after we give more details of the electrostatic embedding, so we postpone discussion of the final overcounting procedure to section 2.2.

In the example just discussed, 14 of the pentamers and all of the tetramers need to be capped at both ends to satisfy dangling bonds; the first and last pentamer each need only one cap each. Capping will be accomplished with cap atoms, as explained below. Each capped oligomer (i.e., pentamer or tetramer) is called a capped primary system (CPS). If, instead of pentamers, the first-order fragments of the chain polymer are trimers, we get a "trimer−dimer" calculation. If they are heptamers, we get a "heptamer−hexamer" calculation, and so forth.

In the general case, one might consider a trade-off between using smaller fragments with higher orders or larger fragments with lower orders. In the present study, we did not examine higher orders of many-body interactions (higher than second order) because the trade-off of making the fragment size smaller when one calculates higher orders of many-body interactions would make the cap−cap atom interactions larger in the doubly capped fragments, and getting the caps too close to one another could be a significant source of error or uncertainty. So it seems most reasonable to study how accurate the method can be at second order.

In the present study, all fragment and dimer energies are calculated by hybrid Kohn−Sham density functional theory, in which the energy E_I of fragment I is given by

$$E_I = \frac{1}{2} \sum_{\mu} P_{\mu\nu}(H_{\mu\nu}^{\mathrm{core}} + F_{\mu\nu}) + \sum_{\alpha} \sum_{\beta>\alpha} \frac{Z_\alpha Z_\beta}{R_{\alpha\beta}} + \sum_{\alpha,A} \frac{q_A Z_\alpha}{R_{\alpha,A}} \tag{4}$$

where μ and ν represent atomic basis functions, Z_α is the charge of the nucleus of atom α of the capped primary system, CPS*, where the asterisk means the CPS is electrostatically embedded, q_A is the background charge of atom A as obtained by electrostatic potential fitting or density population analysis, $R_{\alpha\beta}$ and $R_{\alpha,A}$ are the nucleus−nucleus distance between atoms, between α and β and between α and A, respectively, $P_{\mu\nu}$ is an element of the density matrix, and the core Hamiltonian is defined as

$$H_{\mu\nu}^{\mathrm{core}} = \int \phi_\mu \left[-\frac{1}{2}\nabla^2 \right] \phi_\nu \, \mathrm{d}\mathbf{r} + \int \phi_\mu \left[-\left(\sum_{\alpha} \frac{Z_\alpha}{|\mathbf{R}_\alpha - \mathbf{r}|} + \sum_{A} \frac{q_A}{|\mathbf{R}_A - \mathbf{r}|} \right) \right] \phi_\nu \, \mathrm{d}\mathbf{r} \tag{5}$$

where the first term is the kinetic energy of electrons, and the second term represents the attractive interactions of the nucleus and the electron and of the electron and the background charge. Finally $F_{\mu\nu}$ is an element of the Fock matrix, and it is given by

$$F_{\mu\nu} = H_{\mu\nu}^{\mathrm{core}} + \sum_{\lambda=1} \sum_{\sigma=1} P_{\mu\nu} \left[(\mu\nu|\lambda\sigma) - \frac{1}{2}\left(\frac{X}{100}\right)(\mu\lambda|\nu\sigma) \right] + \left(\frac{1-X}{100}\right) V_{\mu\nu}^{\mathrm{x}} + V_{\mu\nu}^{\mathrm{c}} \tag{6}$$

where X is the percentage of Hartree−Fock exchange, $(\mu\nu|\lambda\sigma)$ is a two-electron integral in Mulliken's notation, and $V_{\mu\nu}^{\mathrm{x}}$ and $V_{\mu\nu}^{\mathrm{c}}$ are matrix elements of the local part of the exchange potential and the density-functional correlation potential, respectively.

All of the CPSs have one or two cap atoms, which may be either hydrogen atoms or tuned fluorine atoms. The latter have pseudopotentials modified from CRENBL effective core potentials.[64] These are called tuning potentials, and they have the form

$$U_{\mathrm{tuning}}(r) = U(r) + U_0(r) + \sum_{lm} [U_l(r) - U_0(r)] |lm\rangle\langle lm| \tag{7}$$

where $|lm\rangle\langle lm|$ are spherical harmonic projectors, U_0 and U_l are the original CRENBL potentials (representing 1s electrons) given by

$$U_l(r) = \frac{1}{r^2} \sum_{j} C_{lj} r^{n_{lj}} \, \mathrm{e}^{-\alpha_{lj} r^2} \tag{8}$$

where the parameters, C_{lj}, n_{lj}, and α_{lj} are taken from ref 64, and $U(r)$ is an extra exponential function

$$U(r) = C \exp\left[-\left(\frac{r}{r_0} \right)^2 \right] \tag{9}$$

where C is a coefficient to be optimized, and r_0 is taken to be 1.0 bohr. The basis set that corresponds to the core orbital of fluorine was not excluded in the previous capped atom study,[63] but here we use the effective core potential in the conventional way; that is, the basis function corresponding to the 1s orbital is replaced by the pseudopotential. As a consequence, the optimized parameter cannot be taken from the previous study.

dx.doi.org/10.1021/ct300845q | J. Chem. Theory Comput. 2013, 9, 1381−1393

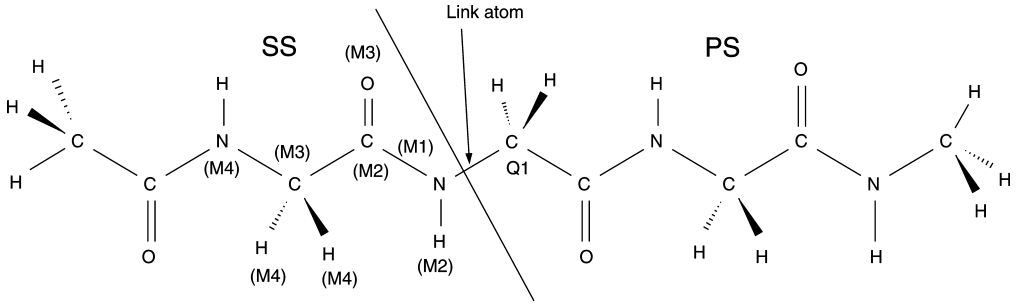


Figure 1. Definitions of Q1, M1, M2, M3, and M4 for the case where the interfragment bond cutting is of the cap(N)-C_O type, and the peptide is Ace−$(Gly)_3$−NMe. The cap atom is not shown, but the "link atom" arrow shows where it would be placed after the cut is made.

The tuned cap F atom is placed on the bond vector of the M1 and Q1 atoms (see Figure 1), as defined previously; thus it is at

$$\mathbf{r}_L = \mathbf{r}_{Q1} + d^0_{Q1-L}\frac{\mathbf{r}_{M1} - \mathbf{r}_{Q1}}{|\mathbf{r}_{M1} - \mathbf{r}_{Q1}|} \tag{10}$$

where $\mathbf{r}_L$, $\mathbf{r}_{Q1}$, and $\mathbf{r}_{M1}$ are the positions of the cap atom, the Q1 atom, and the M1 atom, respectively, and d^0_{Q1-L} is the standard bond length as used in the previous tuned cap atom study.[64] The definitions of Q1 and M1 (and also M2, which will be used below) are illustrated in Figure 1. The standard bond lengths that we employed are 1.09 and 1.01 Å for C−H and N−H bonds for using a hydrogen cap atom and 1.33 and 1.41 Å for C−F and N−F bonds for using a tuned fluorine cap atom.

When we cap a fragment, the charge near the boundary is redistributed to avoid unrealistically large electrostatic interactions. All the tested redistribution schemes are balanced; that is, they preserve the sum of the charges of the system by requiring the sum of the embedding charges to be an integer;[73,74] this is called a balanced redistributed charge. A balanced redistributed charge scheme satisfies[73,74]

$$q^{ES} = q^{CPS} + q^{M1} + \sum_{i \neq M1} q_i^{SS} \tag{11}$$

where q^{ES} is the sum of charges of the entire system, q^{CPS} is the charge of the capped primary system, q^{M1} is the charge on the M1 atom, and q_i^{SS} is charge of atomic site i in the secondary system. In a balanced calculation (all calculations reported here are balanced), q^{CPS} is zero or an integer, and the sum of the other two terms on the right-hand is required to be zero or an integer. In the charge redistribution schemes used in the present study, q^{M1} is equal to zero, and all embedding charges are centered at other non-CPS nuclei. Therefore, when one determines the q_i^{SS}, one needs to redistribute charges from the CPS and M1 to atomic sites of SS, excluding M1.[70,74] In the RC2 scheme,[74] the redistributed charge is evenly redistributed over all M2 atoms. In the AMBER-2 scheme,[73] the redistributed charge is evenly redistributed over the centers of all of the secondary atoms (a secondary atom is any atom that is not a CPS atom), and in the RC3 scheme,[74] the charge is evenly redistributed on M2 and M3 atoms.

In section 2.2, we will need to consider combined QM−MM calculations, and as a preliminary discussion prior to that section, we need to explain the relation of electrostatically embedded fragment calculations to standard combined QM/MM calculations. In QM/MM calculations, one writes the energy as

$$E^{QM/MM} = E_{QM} + E_{QM-MM} + E_{MM-MM} \tag{12}$$

where the first, second, and third terms on the right-hand side are the QM energy, the QM−MM interactions terms, and the MM−MM interaction terms. The electrostatically embedded fragment calculations discussed above are equivalent to the sum of the first two terms where only the electrostatics are included in the QM−MM term (i.e., no van der Waals or valence terms), and with the straight Coulomb overcounting scheme, we do not need the third term. In section 2.2, we will, however, need the third term. In our original scheme employing redistributed charges for the QM−MM terms,[70] we used unredistributed charges in the MM−MM terms; we will call this option 1 for the MM−MM terms. In our more recent work employing balanced and redistributed charges in the QM−MM terms,[63,74] we employed balanced but unredistributed charges in the MM−MM terms (even though the charges were balanced and redistributed in the QM−MM terms), we call this option 2 for the MM−MM terms. In section 2.2, we will consider a more consistent option, namely to use balanced and redistributed charges for the MM−MM term; we call this option 3 for the MM−MM terms.

One more complication must be discussed. In a typical force field, the MM−MM electrostatic interactions are not the same as the QM−MM electrostatic interactions; for example, the MM−MM electrostatic interactions do not include the 1,2- and 1,3-interactions (although they are not omitted in the QM−MM term), and (depending on the force field used) sometimes the 1,4 electrostatic interactions are excluded or reduced as well. We may call this the force field MM−MM electrostatics. However, in the discussion below, we will need to refer to calculations in which the MM−MM electrostatics include all interactions; we will denote this by referring to MM−MM electrostatics without exclusions.

2.2. EE-MTA with the GEBF Overcounting Correction Scheme. The overcounting scheme of eq 3 is called the straight Coulomb overcounting scheme. For our final calculations, we used a better method, which will be called the GEBF overcounting scheme, and which is explained next.

As mentioned above, the GEBF method[35] is very similar to MTA. One difference is that GEBF includes electrostatic embedding but employing an iterative method, in contrast to the noniterative method used here. GEBF also differs from EE-MTA in the way that fragments are assigned (it uses a distance cutoff) and in the way that one corrects from overcounting of electrostatic interactions. The overcounting difference is worth examining in more detail. In the GEBF method, eq 1 is replaced by

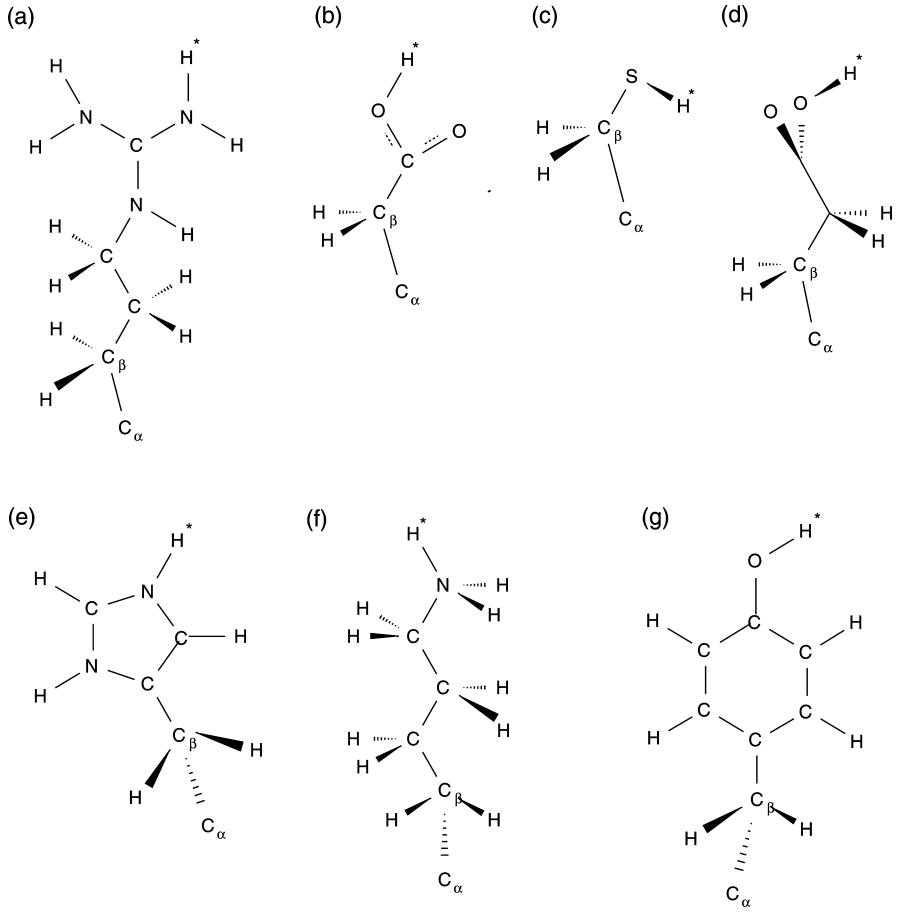


Figure 2. The positions of deprotonation on the side chains of X in Ace−Gly−X−Gly−NMe are given by asterisks in the illustrated side chains. (a) Arginine, (b) protonated aspartic acid, (c) cysteine, (d) protonated glutamic acid, (e) protonated histidine, (d) lysine, (f) tyrosine.

$$E^{\mathrm{GEBF}} = \sum \tilde{E}_I - \sum \tilde{E}_{I\cap J} + \sum \tilde{E}_{I\cap J\ldots\cap K} + \ldots + (-1)^{N-1} \sum \tilde{E}_{I\cap J\ldots\cap N} - \left(\sum_{k=1}^{M} C_m - 1\right) \sum_{A} \sum_{B>A} \frac{Q_A Q_B}{R_{AB}} \quad (13)$$

where M is the total number of constructed systems (for example, in our pentamer−tetramer calculations, it is the number of pentamers plus the number of tetramers), and C_m is the value (plus or minus one) of the coefficient in front of the $\tilde{E}_I$, $\tilde{E}_{I\cap J}$, ..., $\tilde{E}_{I\cap J\ldots\cap N}$ values in eq 13. $\tilde{E}_I$ differs from E_I in that $\tilde{E}_I$ includes the MM−MM electrostatic interaction terms, whereas E_I does not, and where the final term of eq 13 is zero for the present application (it would not be zero for all possible fragmentation schemes, but it is zero for the present ones because $\sum_{k=1}^{M} C_m$ equals unity in the present case).

In the remainder of this section, we only consider the case where $\sum_{k=1}^{M} C_m$ equals unity. In this case, the difference between eqs 1 and 13 is that the overcounting term of eq 3 is replaced by including the electrostatic MM−MM interaction terms (without exclusions) in the fragment energies. It is easy to see that these two schemes would be identical (i) if no covalent bonds were cut in making the fragments (thus avoiding the need for link atoms or capping atoms and meaning that the MM charges used in all steps of the calculations need neither balancing nor redistribution) or (ii) if covalent bonds are cut but there is no balancing of MM charges and no redistribution at the boundaries. Since the overcounting term has to correct overcounting of Coulomb interactions in the QM/MM terms by means of an MM term, the correction should be as consistent as possible. Equation 3, however, by employing straight Coulomb interaction of unbalanced and unredistributed MM charges, is not as consistent as possible, and it would be hard to modify it to take account of adjusted charges because the adjustment on a given site depends on which fragment is being calculated. To remedy this, we employ the GEBF overcounting scheme with the consistent option 3 mentioned at the end of the section 2.1. In this way, the MM charges used to evaluate the MM−MM electrostatics are the same charges at the same locations as those used in the embedded QM calculations. We tested both the straight electrostatics overcounting term and GEBF overcounting correction scheme with option 3 for the MM−MM interaction terms, and we found that the GEBF overcounting correction scheme is much more accurate. Therefore, we employ this scheme in our EE-MTA method. All numerical results in this paper were obtained with the GEBF overcounting correction scheme.

3. PARAMETERIZATION OF THE PSEUDOPOTENTIAL OF THE LINK F ATOM

A tuned F atom will be denoted as F*. In the previous studies,[63] the parameter in the effective core potential of the cap F* atom was determined so as to reproduce the Mulliken charge separation estimated by a full QM calculation (the details of the procedure are given in the previous paper[63]).

dx.doi.org/10.1021/ct300845q | J. Chem. Theory Comput. 2013, 9, 1381−1393

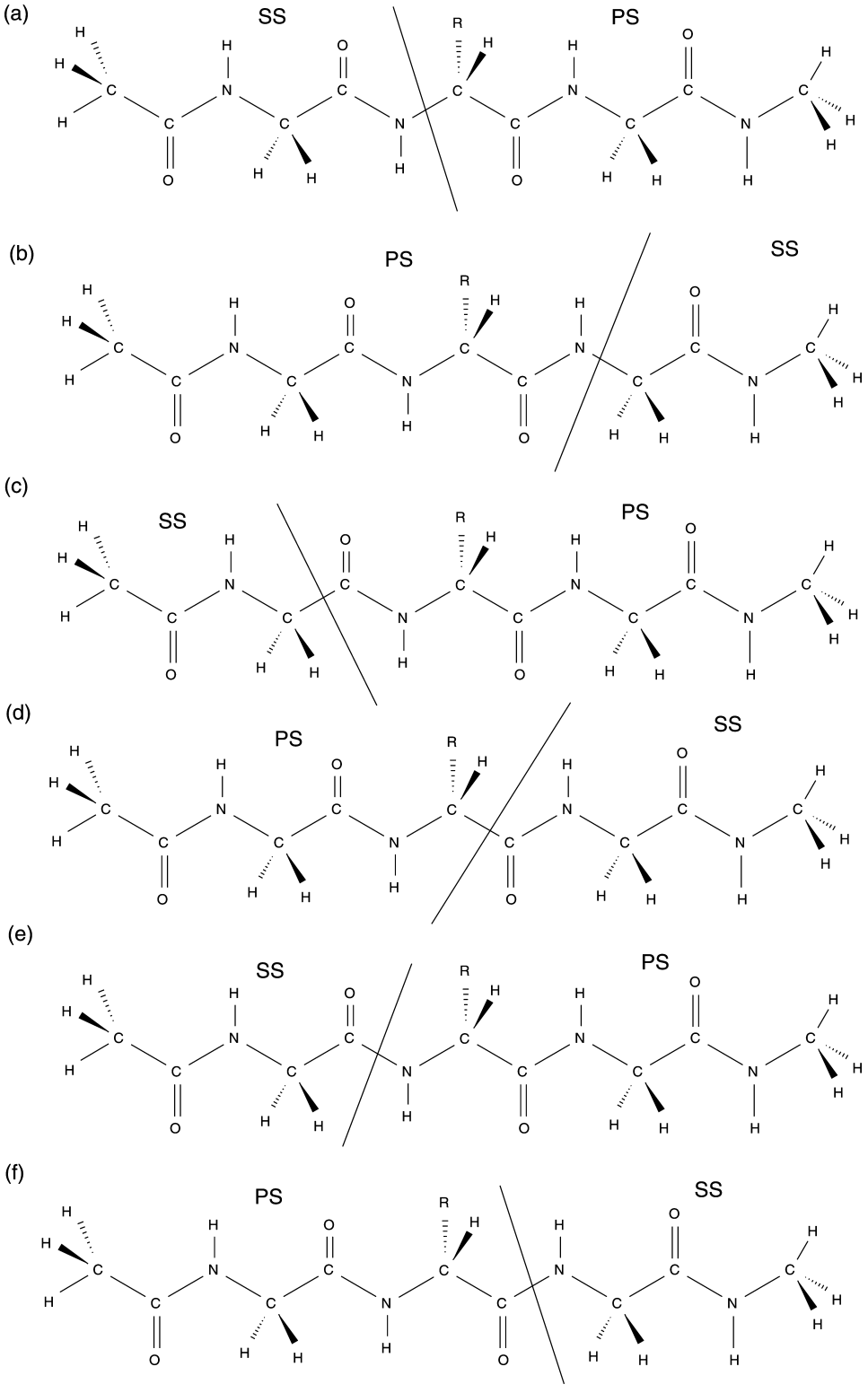


Figure 3. Location of the QM/MM partition in the protonation calculations. The side marked PS is the primary subsystem, which is treated by QM, and the side marked SS is the secondary subsystem, which is treated by MM.

However, the dependence of the Mulliken charge on the choice of basis set is fairly large; in addition, the partial charges depend greatly on the choice of other methods, e.g., methods of Mulliken population analysis,[65] Hirshfeld population analysis,[66] natural bond population analysis,[67] electrostatic potential fitting,[68] or CM5,[69] that can be used to derive partial charges, and the resultant parameters depend on which method one selects. To avoid this dependence, the present parameters are determined to minimize the mean unsigned error of the proton affinity.

The advantages of the use of proton affinity for the parametrizations of the fluorine cap atom are that (i) the proton affinity, unlike the partial charge, is an observable property—therefore it is possible to use experimental data in

the parametrization; (ii) one can use diffuse basis sets with a high-level of ab initio calculation which yield reasonably accurate values for the experimentally observable proton affinity, whereas this would usually result in unphysical electron populations if one used Mulliken population analysis; and (iii) the reactant (protonated) and product (deprotonated) are well-defined, and the parameter of the effective core potential depends on both the initial and final states, which makes the parametrization straightforward.

The proton affinity is calculated in the framework of the QM/MM scheme by

$$E^{\mathrm{PA,QM/MM}} = -\lfloor E^{\mathrm{QM/MM}}(\mathrm{AH}^+) - E^{\mathrm{QM/MM}}(\mathrm{A})\rfloor \qquad (14)$$

where $E^{\mathrm{QM/MM}}(\mathrm{AH}^+)$ and $E^{\mathrm{QM/MM}}(\mathrm{A})$ represent the QM/MM energy of the protonated and deprotonated species, respectively. The QM/MM energy is defined by[70]

$$\begin{aligned} E^{\mathrm{QM/MM}} = {} & E(\mathrm{QM;\ CPS^{**}}) + [E(\mathrm{val;\ ES}) \\ & - E(\mathrm{val;\ CPS})] + E(\mathrm{Coul;\ SS}) \\ & + [E(\mathrm{vdW;\ ES}) - E(\mathrm{vdW;\ CPS})] \end{aligned} \qquad (15)$$

where E(QM;CPS**) is quantum mechanical energy of the QM system that is coupled by the electrostatic potential to the secondary subsystem; E(val;ES) and E(val;CPS) are the MM energies for the valence interactions (for the entire system and for the CPS) corresponding to bond stretching, bond angle stretching, and torsional terms (which are described as a harmonic or trigonometric potentials in most MM force fields); E(vdW;ES) and E(vdW;CPS) are MM energies for the van der Waals interactions; and E(Coul;SS) is the Coulomb interaction energy of the secondary subsystem. Since the optimized geometry of the protonated form is used for the energy calculation of the deprotonated form, all of the MM terms are canceled in eq 15, and only the energy of the first term for the protonated and deprotonated form is required in the present case.

For parametrization, we used seven tetrapeptides, each of which contains two glycine residues and one other amino acid, called X; the structure is written as Ace−Gly−X−Gly−NMe, to indicate that the N terminal and C terminal are capped by Ace and NMe, respectively. The residue X can be Arg (arginine), Ash (protonated aspartic acid), Cys (cysteine), Glh (protonated glutamic acid), Hip (protonated histidine), Lys (lysine), or Tyr (tyrosine). The hydrogen to be removed in the deprotonated form is shown in Figure 2 with an asterisk. Figure 3 shows the partitioning scheme used for the protonation calculations.

For the parametrization, the CM5 method[69] is used to obtain background charges. The charges determined for the protonated form of the Ace−Gly−X−Gly−NMe are also used for the deprotonated form.

In total, six parameters are determined for the fluorine cap atom, depending on which kind of bond is broken and which side of the broken bond is capped. The partitioning locations are shown in Figure 3. For example, part a shows a broken N−C_α bond capped as F*−C_α; we denote this as F*(N)−C_α. Using the same shorthand notation, the other types of capping considered are (b) F*(C_α)−N, (c) F*(C_α)−C_O, where C_O denotes a carbonyl carbon, (d) F*(C_O)−C_α, (e) F*(C_O)−N, and (f) F*(N)−C_O. Where no confusion will result, we shorten these to F*(N)−C_α, F*−N, F*−C_O, F*−C_α, (e) F*−N, and F*−C_O. It is noted that each of these cases should have different optimized parameters. The parameters so determined are to be used for peptides with any amino acid sequence and any charge state.

All geometries are optimized for the protonated form of Ace−Gly−X−Gly−NMe by using the M06-2X density functional with the 6-31G* basis set. These geometries were started from a β sheet conformation, and hence they are not intended to represent global minima. The Cartesian coordinates of the optimized geometry for each X are given in the Supporting Information (Tables S1−S7).

The CM5,[69] electrostatic potential (ESP) fitting by Merz−Kollman scheme,[71] and Mulliken[72] charges were derived using the M06-2X density functional with the 6-31G* basis set. For the calculations on the polypeptide, the CM5, ESP, and Mulliken charges are determined by full QM calculations for the whole system for each of the two conformations.

4. COMPUTATIONAL DETAILS FOR THE MTA AND EE-MTA CALCULATIONS

Two secondary structural motifs, namely the α helix and the β sheet, of Ace−$(Ala)_{20}$−NMe were optimized without constraints on the dihedral angles starting from the Ramachandran angles (φ, ψ) = (−47°, −57°) for the α helix and (−119°, 113°) for the β sheet conformation using the *AMBER99* force field,[75] where the threshold of the root mean squared (RMS) gradient was set to 0.04 kcal mol^{-1} Å^{-1}, and the determined geometries are given in the Supporting Information (Tables S8 and S9) and shown in Figure 4. These geometries are used for all of the MTA and EE-MTA calculations. All of the MTA and EE-MTA calculations were performed using the M06-2X density functional with the 6-31G* basis set.

In the present work, as mentioned in section 2, we truncate the expansion at second order, as in eq 2. Although the inclusion of the higher-order many-body interactions should in principle provide more accurate results, the error due to the interactions of cap atoms becomes larger as the CPSs become smaller. In addition, the computational cost becomes slightly more expensive at higher order. Therefore, the second-order expansion is reasonable. The background charge used for the EE-MTA is a restrained electrostatic potential (RESP) charge[76,77] from *AMBER02*,[78] which is determined by the standard AMBER procedure, in particular using B3LYP/cc-pVTZ//HF/6-31G* for a dipeptide (Ace−X−NMe, where X is each amino acid) in a continuum solvent for two important conformations, and the charge in one amino acid unit of alanine is customized by fixing the charges of the blocking groups, as explained in ref 78.

5. RESULTS AND DISCUSSION

5.1. Parametrization of F* against Proton Affinities. As the coefficient of the extra exponential in the effective core potential becomes large and positive, the electron affinity is reduced. When $C = 0$, the electron affinity of the cap atom is equal to that of the fluorine atom, and when $C < 0$, the electron affinity is larger than that of the fluorine atom.

Table 1 gives the parameters optimized for seven peptides plus the parameters in the original CRENBL effective core potential. The optimized parameter is 0.78 for F*(M1) bound to C_α(Q1) (case a in Figure 3), and it is 2.34 for F*(M1) bound to N(Q1) (see Figure 3b), which shows that a fluorine cap atom binding to the α-carbon requires a larger electron affinity as compared to that binding to the amide nitrogen to

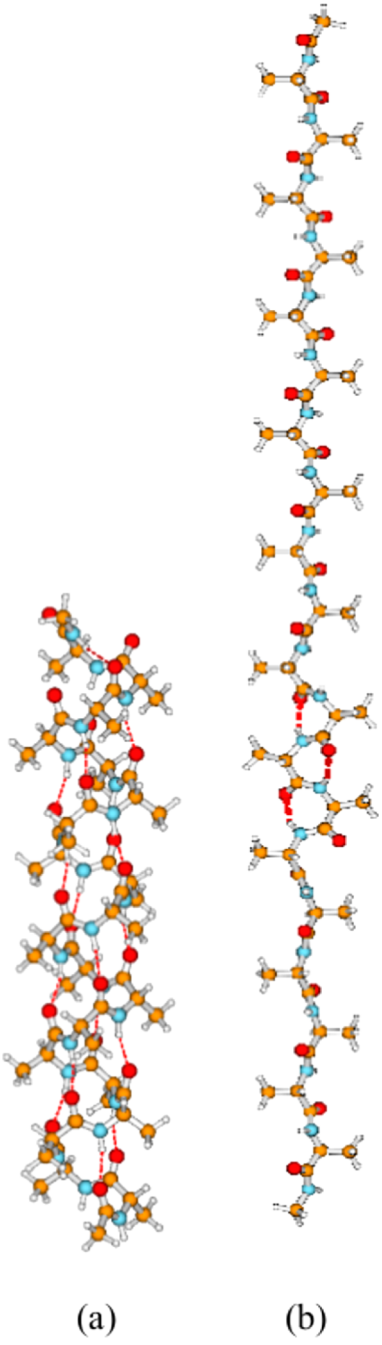


Figure 4. Optimized geometries of Ace−$(Ala)_{20}$−NMe for two conformations, (a) α helix and (b) parallel-β sheet, with the AMBER99 force field.

reproduce the electronic structure obtained by full QM calculations, and these C values both correspond to a larger electron affinity than that of hydrogen. We obtained quite different parameters, 0.78 and −3.87, for F*(M1) bound to C_α(Q1), where one of them is substituting for the amide nitrogen and the other is substituting for the C_O. The latter is electron withdrawing due to the electronegative oxygen; thus the optimized parameter is reasonable. Such a large difference is also seen for the F*(M1) bound to N(Q1).

Table 2 gives the resultant proton affinities using the optimized parameters, and these are compared with the proton affinities calculated by using a conventional hydrogen cap atom, which is denoted as H_L. The deviations from full QM

Table 1. Optimized Parameters of the Effective Core Potential of a Cap F* Atom and Parameters Used in the Original CRENBL Effective Potential

	original CRENBL			
n_{ij}[a]	α_{ij}[a]	C_{ij}[a]	C[b,c]	
	U_0		(a) F*−C_α	(b) N−F*
2	2.8835	12.685306	0.78	2.34
2	3.1077	19.302589		
`	5.6122	1.002179	(c) F*−C_O	(d) C_α−F*
0	2.8146	2.245349	0.65	−3.87
	U_1		(e) F−N	(f) C_O−F*
2	44.5166	6.72324	−1.64	1.50
2	12.9487	0.929649		
1	132.4967	1.526734		

[a]Taken from ref 63. [b]In eq 9. [c]Location of the partition for a−f is shown in Figure 3. C_α is an α-carbon, and C_O is carbonyl carbon.

Table 2. Signed Errors of Proton Affinity (kcal/mol) of Ace−Gly−X−Gly−NMe Calculated by EE-QM[a]

	QM/MM[b,c,d]				full QM[e]
	(a) F*−C_α		(b) N−F*		
	F*	H_L	F*	H_L	
Arg	−0.33	0.68	−5.40	−5.33	276.70
Ash	0.38	2.16	0.30	−1.03	363.40
Cys	0.53	2.84	0.20	−1.48	357.60
Glh	−0.15	1.38	1.15	−0.29	368.10
Hip	−5.04	−3.67	−1.27	−2.65	264.20
Lys	0.02	1.30	−11.69	−11.89	260.50
Tyr	−0.52	0.87	−0.05	−0.87	368.60
MUE[f]	0.99	1.84	2.87	3.36	
	(c) F*−C_O		(d) C_α−F*		
	F*	H_L	F*	H_L	
Arg	−0.19	0.27	−4.7	−4.7	
Ash	0.31	1.24	2.4	10.9	
Cys	0.38	1.48	4.0	10.0	
Glh	0.3	1.06	−1.3	5.9	
Hip	−1.64	−0.57	−5.4	4.0	
Lys	−0.13	0.43	−9.0	−12.4	
Tyr	0.22	0.96	−0.1	5.1	
MUE[f]	0.45	0.86	3.85	7.58	
	(e) F*−N		(f) C_O−F*		
	F*	H_L	F*	H_L	
Arg	−0.99	3.36	−5.82	−5.64	
Ash	0.32	7.97	5.20	4.96	
Cys	−0.04	9.54	2.91	2.92	
Glh	0.48	5.24	0.06	−0.57	
Hip	−4.92	−2.77	−0.25	−0.19	
Lys	−1.37	4.65	−14.85	−14.79	
Tyr	−0.68	3.35	1.72	1.94	
MUE[f]	1.26	5.27	4.40	4.43	

[a]The optimized parameters of the tuned fluorine cap atom are used. The signed errors are compared with those of EE-QM calculations with a conventional hydrogen cap atom. [b]The M06-2X/6-31G* basis set is used for the QM part, and the local minimum geometry is also determined by M06-2X/6-31G*. [c]CM5 charges from full QM calculations on the protonated form (these charges are used as embedding charges for both the protonated and unprotonated molecules). [d]The balanced RC2 charge redistribution scheme in ref 73 is used. [e]M06-2X/6-31G//M06-2X/6-31G*. [f]Mean unsigned error.

calculation are very small, with the values being under 1.0 kcal/mol for C_α−N. On the other hand, the MUE error is fairly large for N(Q1)−C_α for both tuned fluorine and the hydrogen cap atom, and the large error mainly comes from positively charged Ace−Gly−Lys−Gly−NMe in the protonated form. Overall, the tuned fluorine atom shows better performance than a hydrogen atom.

5.2. Ace−$(Ala)_{20}$−NMe. *5.2.1. Effect of Background Charge and Size of Fragment.* Table 3 compares the electronic energy of the α helix to that of the β sheet, including the energy difference of these two conformations as calculated by MTA (without background charges) and by EE-MTA (with background charges). For this table, the conventional hydrogen cap atom is used. In particular, Table 3 gives the electronic energy by MTA calculations for five sizes of fragments, and one can see that the relative energy becomes close to that of full QM when one uses larger fragments and that the energy of the α helix depends more on the size of the fragments than does the energy of the β sheet. The background charges in EE-MTA remedy the high MTA relative energy of the α helix as compared with the relative energy by the full QM calculation. The effect of electronic polarization is not obvious in the β sheet conformation. Thus, the improvement of the energy difference of the two conformations comes from the improvement in energy of the α helix.

Table 3. MTA and EE-MTA Electronic Energies (in hartrees) of Ace−$(Ala)_{20}$−NMe for Two Motifs, α Helix and Parallel-β Sheet, and Relative Energy of the Two Motifs[a]

fragment unit[b]	E_α	E_β	$E_\alpha - E_\beta$	|% error|
		CPS-MTA		
trimer (19)−dimer (18)	−5192.80567	−5192.88396	0.07829	138
tetramer (18)−trimer (17)	−5192.99501	−5192.88828	−0.10673	49
pentamer (17)−tetramer (16)	−5193.00802	−5192.88714	−0.12088	42
hexamer (16)−pentamer (15)	−5193.02488	−5192.88757	−0.13731	34
heptamer (15)−hexamer (14)	−5193.05100	−5192.88721	−0.16379	22
		EE-MTA with H_L		
trimer (19)−dimer (18)	−5193.05037	−5192.88705	−0.16332	22
tetramer (18)−trimer (17)	−5193.06990	−5192.88730	−0.18260	13
pentamer (17)−tetramer (16)	−5193.09035	−5192.88730	−0.20305	2.8
hexamer (16)−pentamer (15)	−5193.09124	−5192.88732	−0.20392	2.4
heptamer (15)−hexamer (14)	−5193.09311	−5192.88731	−0.20580	1.5
		ref. (M06-2X/6-31G*)		
	−5193.09618	−5192.88729	−0.20889	

[a]The cap atom is H_L. The embedding charges are *AMBER* RESP charges. The charge redistrbution scheme is balanced RC2. The partitioning is cap-C_α...N-cap. The QM subsystems are treated by M06-2X/6-31G*. [b]The number of fragments of a given size is in parentheses.

The small errors in the relative energies, when large enough fragments are considered, confirm the good performance of the molecular tailoring approach in the previous studies.[16,18] For CPS-MTA, the largest improvement of the energy difference is seen when increasing from the trimer−dimer fragmentation (percentage error: 138%) to the tetramer−trimer fragmentation (percentage error: 49%), and this large change is due to the α helix. This largest change can be explained by the fact that in this partition (cap-C_α...N-cap), the tetramer is the smallest fragment that includes the important intramolecular hydrogen-bond interaction of an α helix between the i and i + 4 peptide units. Since the computational cost increases as one makes the fragment size larger, one wants to choose the smallest fragment size that gives the target accuracy, and Table 3 shows that pentamer−tetramer or hexamer−pentamer fragmentation is a reasonable choice for EE-MTA, although the errors with respect to the full QM calculation are only 5−6 millihartrees, respectively. The heptamer−hexamer calculation reduces this to 3 millihartrees, which is 1.9 or 0.1 kcal/mol per residue. We shall see further improvement in the relative energy of the two conformations in later tables as we improve the other algorithmic choices. However, already in Table 3, we see the encouraging result that the tetramer−trimer EE-MTA result is more accurate than the heptamer−hexamer MTA result for the difference in energy of the two structures.

5.2.2. Location of the Cut Bonds. Table 4 shows the electronic energy of Ace−$(Ala)_{20}$−NMe for three different partitioning positions: (i) cap-C_α(Q1)···N(Q1)-cap, (ii) cap-C_O(Q1)···C_α(Q1)-cap, and (iii) cap-N(Q1)...C_O(Q1)-cap. As can be seen in Table 4, the error depends on both the sizes of the fragments and the locations of the partitions. Partitions ii and iii show higher accuracy than partition i in only four of the 10 cases. For the trimer−dimer and hexamer−pentamer cases, the partitioning between the backbone C_O and the α-carbon or between C_O and nitrogen are the best choices. Taking a broader view of the results, one can say that the partitioning between the backbone C_O and the α-carbon is the most successful strategy. The relative accuracies of the partitions are related in part to the polarities of the cut bonds, but the common assumption that partitioning of a polar bond lowers the accuracy is only partly borne out. Since the boundary portions are canceled out in the MTA, it is expected that the large dependence on the location of the partition comes to a large extent from the electrostatic interactions of background charges, and the sensitivity of the results to overcounting correction is consistent with that.

It is encouraging that the error in conformational energy difference is less than or equal to 6.0% in nine of 15 cases (nine of 12 cases if we exclude the trimer−dimer results); this is 1.3 millihartrees, which is less than 8 kcal/mol (0.4 kcal/mol per residue).

For tetramers or larger, the least accurate location to cut a bond is usually choice (iii), which may be explained in terms of the secondary-structure hydrogen bonds of the α-helix. For schemes (i) and (ii), a tetramer contains one hydrogen bond, a pentamer contains two, etc., but for scheme (iii) a tetramer has no hydrogen bond, a pentamer has one, a hexamer has two, etc.

5.2.3. Charge Redistribution and Type of Background Charge. Table 5 shows the results of comparisons of three types of charge redistributions, which are explained at the end of section 2 and which are indicated in column 2 of Table 5. The tested charge types are (i) the RESP charge from *AMBER02*[78,79] (these charges differ from those of the *AMBER99* force field in which the charge is derived using the Hartree−Fock level with the 6-31G* basis set and for which the dipole moment is overestimated in the gas phase[80]), (ii) CM5 charges, which are derived so as to reproduce the gas-

Table 4. EE-MTA Electronic Energies (in hartrees) of Ace−$(Ala)_{20}$−NMe for Two Motifs, α Helix and Parallel-β Sheet, and Relative Energy of the Two Motifs[a]

fragment unit	E_α	E_β	$E_\alpha - E_\beta$	|% error|
		(i) $H_L-C_\alpha...N-H_L$		
trimer (19)−dimer (18)	−5193.05037	−5192.88705	−0.16332	22
tetramer (18)−trimer (17)	−5193.06990	−5192.88730	−0.18260	13
pentamer (17)−tetramer (16)	−5193.09035	−5192.88730	−0.20305	2.8
hexamer (16)−pentamer (15)	−5193.09124	−5192.88732	−0.20392	2.4
heptamer (15)−hexamer (14)	−5193.09311	−5192.88731	−0.20580	1.5
		(ii) $H_L-C_O...C_\alpha-H_L$		
trimer (19)−dimer (18)	−5193.07682	−5192.88800	−0.18881	9.6
tetramer (18)−trimer (17)	−5193.08399	−5192.88766	−0.19633	6.0
pentamer (17)−tetramer (16)	−5193.08984	−5192.88737	−0.20247	3.1
hexamer (16)−pentamer (15)	−5193.09142	−5192.88738	−0.20404	2.3
heptamer (15)−hexamer (14)	−5193.09245	−5192.88732	−0.20513	1.8
		(iii) $H_L-N...C_O-H_L$		
trimer (19)−dimer (18)	−5193.05835	−5192.88724	−0.17111	18
tetramer (18)−trimer (17)	−5193.02495	−5192.88540	−0.13955	33
pentamer (17)−tetramer (16)	−5193.11832	−5192.88726	−0.23106	11
hexamer (16)−pentamer (15)	−5193.09154	−5192.88703	−0.20450	2.1
heptamer (15)−hexamer (14)	−5193.08392	−5192.88734	−0.19658	5.9
		ref. (M06-2X/6-31G*)		
	−5193.09618	−5192.88729	−0.20889	

[a]The embedding charges are *AMBER* RESP charges. The charge redistribution scheme is balanced RC2. The QM subsystems are treated by M06−2X/6-31G*.

phase dipole moments (either experimental or calculated by a high level electronic-structure method[69]), (iii) ESP charges,[71] and (iv) Mulliken charges. Note that charge types ii, iii, and iv are determined by full QM calculations for the whole system for each conformation. All four sets of charges are tabulated in the Supporting Information, but before we discuss the differences among the results with various charges models, we should consider the charge redistribution schemes, which we do next.

As can be seen in Table 5, there is only a slight difference among the results for the three redistribution schemes, even for the F* cap atoms (where one might have expected a greater sensitivity due to the F* being closer to the secondary subsystem). Such a small dependency on the charge redistribution scheme has also been observed in a previous study[63] in which the various redistributed charge schemes were tested for proton affinities of a variety of organic compounds.

Now we can return to the question of charge models, i.e., the different ways to obtain the charges that represent the secondary subsystems, as summarized above and as indicated in column 1 of Table 5. These tests employ the balanced redistributed charge method, RC2. One can see in Table 5 that using charge distributions obtained from Mulliken analysis of the full QM calculations for each conformation yields more accurate conformational energies than does using the RESP fixed charge model for each conformation. However, using the CM5 method or ESP fitting for each conformation does not improve the results over RESP. Since the partial atomic charges of types ii, iii, and iv are determined by full quantum mechanical calculations for the whole system for each conformation, whereas in practical applications one would probably determine the charges for smaller subsystems, the results here are probably about the best we can do with fixed background charges. Further work would be required to completely sort out the dependence on charge model, but the range of values of the errors in Table 5 does show that the choice of charge model can be important.

5.2.4. Hydrogen Cap Atom and Tuned Fluorine Cap Atom. Table 5 also compares the performance of hydrogen cap atoms

Table 5. EE-MTA Electronic Energies (in hartrees) of Ace−$(Ala)_{20}$−NMe for Two Motifs, α Helix and Parallel-β Sheet, and Their Relative Energy[a]

charge	charge RD	E_α	E_β	$E_\alpha - E_\beta$	|% error|
			$H_L-C_\alpha...N-H_L$		
RESP[b]	RC2	−5193.09035	−5192.88730	−0.20305	2.8
RESP[b]	AMBER-2	−5193.09275	−5192.88986	−0.20289	2.9
RESP[b]	RC3	−5193.08949	−5192.88733	−0.20216	3.2
CM5[c]	RC3	−5193.08404	−5192.88732	−0.19672	5.8
ESP[c]	RC2	−5193.08827	−5192.88730	−0.20097	3.8
Mulliken[c]	RC2	−5193.09201	−5192.88728	−0.20473	2.0
			$F^*-C_\alpha...N-F^*$		
RESP[b]	RC2	−5193.08859	−5192.88710	−0.20149	3.5
RESP[b]	AMBER-2	−5193.09167	−5192.88888	−0.20279	2.9
RESP[b]	RC3	−5193.08998	−5192.88702	−0.20296	2.8
CM5[c]	RC3	−5193.08162	−5192.88738	−0.19424	7.0
ESP[c]	RC2	−5193.08812	−5192.88727	−0.20084	3.9
Mulliken[c]	RC2	−5193.09196	−5192.88730	−0.20467	2.0
			ref. (M06-2X/6-31G*)		
		−5193.09618	−5192.88729	−0.20889	

[a]The fragment size is pentamer−tetramer. The QM subsystems are treated by M06-2X/6-31G*. [b]*AMBER* RESP charges. [c]Determined by full QM calculation.

to that of the tuned fluorine cap atoms for the case of pentamers stitched together by tetramers. One can see that the two kinds of cap atoms yield fairly similar deviations from a full QM calculation for the electronic energy in both conformations. One sees that the interaction between the fluorine cap atom and the PS calculated in the first term in eq 1 is mostly canceled in the second term. If one calculates a reaction energy such as a proton affinity, which is more sensitive to the electron density distribution, it is expected that the performance of the fluorine link atom and the H link atom will differ more, as we see in Table 2, where the tuned fluorine atom performs much better than the hydrogen link atom.

Eight of the 10 calculations in Table 5 lead to errors of 3.9% or less (5 millihartrees), which corresponds to 2.5 kcal/mol, which is 0.25 kcal/mol per residue, even though the table is restricted to fragments no larger than pentamers.

Table 6 shows results for the tuned F* cap atom for two locations of the cut bonds and includes hexamers and heptamers as well as pentamers, and all six results may be compared to results in Table 4 for the hydrogen cap atom. (A difference between hydrogen link atom and fluorine link atom is not observed for cap-Cα(Q1)···N(Q1)-cap, so the table shows results only for cap-C_O (Q1)···Cα(Q1)-cap and cap-C_O (Q1)···N(Q1)-cap.) Comparing the six cases between Tables 4 and 6, we see that the error with the fluorine cap atom is smaller for all six cases. Averaging the mean unsigned percentage error over the six cases gives an average percentage error almost 3 times larger (2.7% vs 1.0%) for hydrogen cap atoms than for tuned fluorine cap atoms. This comparison shows that our attempt to parametrize a general set of tuned F* atoms for peptides was successful.

Table 6. EE-MTA Electronic Energies (in hartrees) of Ace−$(Ala)_{20}$−NMe for Two Motifs, α Helix and Parallel-β Sheet, and Their Relative Energy[a]

fragment unit	E_α	E_β	$E_\alpha - E_\beta$	|% error|
		(ii) F*−C_O...C_α−F*		
pentamer (17)−tetramer (16)	−5193.09166	−5192.88729	−0.20437	2.2
hexamer (16)−pentamer (15)	−5193.09421	−5192.88735	−0.20686	1.0
heptamer (15)−hexamer (14)	−5193.09568	−5192.88730	−0.20839	0.2
		(iii) F*−N...C_O−F*		
pentamer (17)−tetramer (16)	−5193.09468	−5192.88729	−0.20739	0.7
hexamer (16)−pentamer (15)	−5193.09283	−5192.88735	−0.20548	1.6
heptamer (15)−hexamer (14)	−5193.09544	−5192.88731	−0.20812	0.4
		ref. (M06-2X/6-31G*)		
	−5193.09618	−5192.88729	−0.20889	

[a]The notations ii and iii in this table refer to the locations of the cut bonds, as explained near the beginning of section 5.2.2; thus these results may be compared to those for the same positions of the cut bonds in Table 4, as discussed in section 5.2.4. The embedding charges are *AMBER* RESP charges. The charge redistribution scheme is balanced RC3. The QM subsystems are treated by M06-2X/6-31G*.

6. CONCLUDING REMARKS

In this study, we have suggested an electrostatically embedded MTA method in which the long-range electrostatic interactions are explicitly included in an MTA calculation in the manner of electrostatically embedded QM/MM calculations with an MM overcounting term evaluated the same way as in the GEBF scheme.

We examined five variables to learn how they affect the calculated energies of two conformations (α helix and β sheet) of Ace−$(Ala)_{20}$−NMe and the calculated energy difference between these conformations: (a) polarization by extrafragment electrostatic interactions, including dependence on the type of background charge (i.e., the charge model), (b) the sizes of the fragments, (c) the locations of the sites at which the peptide is fragmented (i.e., what kind of bond is cut), (d) the redistribution of charges near the fragment boundaries, and (e) the type of cap atom, including new generically tuned fluorine cap atoms derived in this work.

We found that electronic polarization has a large effect on the energies and relative energies; the effect is especially significant in the α helix conformation, in which the long-range interactions are very important.

We also showed the importance of taking a large enough fragment size. In the present cases, reasonably good results are obtained with tetramers, pentamers, hexamers, or heptamers as the large fragments.

The location of the fragmentation does make a difference; however, this difference is often smaller than that due to the difference associated with the sizes of the fragments.

The charge redistribution scheme has only a small effect on the results, but the choice of charge model is more significant. Partial atomic charges that depend on conformation are not systematically more reliable than conformation-independent charges.

The generically tuned F* atoms perform better than hydrogen link atoms in all six cases that are sensitive to the choice of cap atoms.

Altogether, we presented three tests of EE-MB for tetramer fragments, 16 tests for pentamer fragments, five tests for hexamer fragments, and five tests for heptamer fragments. Averaging over all these cases, the average error in the $\alpha-\beta$ energy difference is reduced from 49% by MTA to 17% by EE-MTA for tetramers, from 42% by MTA to 3.7% by EE-MTA for pentamer fragments, from 34% to 1.9% for hexamer fragments, and from 22% by MTA to 2.0% by EE-MTA for heptamer fragments. The average unsigned error in EE-MTA as compared to that in MTA is a factor of 2.8 lower for tetramers, a factor of 11 lower for pentamers, a factor of 18 lower for hexamers, and a factor of 11 lower for heptamers.

The new EE-MTA method can be used for any peptide, protein, or other biopolymer as well as for more general problems. One could also apply it to proteins interacting with substrates and coenzymes. Applications to a variety of problems in biological chemistry would be of great interest.

One may classify fragment methods in a variety of ways: for example, some methods are density based (they approximate the electron density by fragment methods and then calculate an energy from the final density), and others are energy-based (they calculate the final energy from the energies of the fragments). Some methods use a single electronic structure level, and others use more than one electronic structure level[44,51,81] (e.g., different levels for different fragments or an incremental approach where the whole system is treated at a lower level, and higher-order corrections are calculated based on a fragment approach). The EE-MTA method is a single-level, energy-based method, and one possible improvement is

to use more than one level, as in the many-body expansion of the correlation energy.[48,49] Even among single-level energy-based methods, though, there is great variety and perhaps room for improvement. As discussed by Suarez et al.[82] and Herbert and Richard,[83] one may divide fragmentation methods into two general types: nonoverlapping and overlapping. (Mayhall and Raghavachari[53] propose a somewhat different partitioning into inclusion−exclusion methods and many-body methods.) The nonoverlapping methods are those in which each atom of the entire system (except possibly for boundary atoms in some methods) appears in at most one fragment. Single-level, energy-based nonoverlapping methods include the local self-consistent-field (LSCF) method,[84] the molecular-orbital derived empirical potential for liquids (MODEL),[85] the fragment molecular orbital (FMO) method,[20] molecular fragmentation with conjugate caps (MFCC),[29] the variational explicit polarization (X-Pol) method,[38] and many other examples. The overlapping methods are methods in which some portions of the system belong to more than one fragment. Single-level, energy-based overlapping methods include the electrostatically embedded many-body (MB) expansion,[43] the molecular tailoring approach (MTA),[13,18] the electrostatically embedded molecular tailoring approach (EE-MTA) proposed here, the closely related generalized energy-based fragmentation (GEBF),[35−37,86,87] the systematic molecular fragmentation method (SMF),[88] the isodesmic fragmentation method (IFM),[89−91] the combined fragmentation method (CFM),[92] and others. The power of the overlapping methods is especially clear when one considers chain polymers like polypeptides and proteins, in which every residue can be strongly affected by its near neighbors on both sides along the chains—so any nonoverlapping method would necessarily treat some strong interactions as interfragment interactions, whereas an overlapping method can treat them as intrafragment interactions, which is usually more accurate. The present results for applying EE-MTA to a polypeptide are very encouraging. The overlapping methods are very general, and we see them as having enormous potential for a variety of uses. We look forward to further development, refinement, and improvements, further validation, and more applications.

■ ASSOCIATED CONTENT

Supporting Information

The Cartesian coordinates of the studied peptides and the partial atomic charges used for electrostatic embedding. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: truhlar@umn.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under grant no. CHE09-56776.

■ REFERENCES

(1) Yang, W. *Phys. Rev. Lett.* **1991**, *66*, 1438.
(2) Yang, W. *Phys. Rev. A* **1991**, *44*, 7823.
(3) Lee, C.; Yang, W. *J. Chem. Phys.* **1992**, *96*, 2408.
(4) Lu, J. P.; Yang, W. *Phys. Rev. B* **1994**, *49*, 11421.
(5) York, D.; Lu, J. P.; Yang, W. *Phys. Rev. B* **1994**, *49*, 8526.
(6) Zhao, Q.; Yang, W. *J. Chem. Phys.* **1995**, *102*, 9598.
(7) Yang, W.; Lee, T.-S. *J. Chem. Phys.* **1995**, *103*, 5674.
(8) White, C. A.; Head-Gordon, M. *J. Chem. Phys.* **1994**, *101*, 6593.
(9) Strain, M. C.; Scuseria, G. E.; Frisch, N. J. *Science* **1996**, *271*, 51.
(10) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. *Chem. Phys. Lett.* **1994**, *230*, 8.
(11) Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. *J. Phys. Chem.* **1994**, *98*, 9165.
(12) Babu, K.; Gadre, S. R. *J. Comput. Chem.* **2003**, *24*, 484.
(13) Ganesh, V.; Dongare, R. K.; Balanarayan, P.; Gadre, S. R. *J. Chem. Phys.* **2006**, *125*, 104109.
(14) Elango, M.; Subramanian, V.; Rahalkar, A. P.; Gadre, S. R.; Sathyamurthy, N. *J. Phys. Chem. A* **2008**, *112*, 7699.
(15) Rahalkar, A. P.; Ganesh, V.; Gadre, S. R. *J. Chem. Phys.* **2008**, *129*, 234101.
(16) Rahalkar, A. P.; Katouda, M.; Gadre, S. R.; Nagase, S. *J. Comput. Chem.* **2010**, *31*, 2405.
(17) Yeole, S. D.; Gadre, S. R. *J. Chem. Phys.* **2010**, *132*, 094102.
(18) Mahadevi, A. S.; Rahalkar, A. P.; Gadre, S. R.; Sastry, G. N. *J. Chem. Phys.* **2010**, *133*, 164308.
(19) Yeole, S. D.; Gadre, S. R. *J. Chem. Phys.* **2011**, *134*, 084111.
(20) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701.
(21) Nakano, T.; Kaminuma, T.; Sato, T.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2000**, *318*, 614.
(22) Kitaura, K.; Sugiki, S.-I.; Nakano, T.; Komeiji, Y.; Uebayasi, M. *Chem. Phys. Lett.* **2001**, *336*, 163.
(23) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *120*, 6832.
(24) Fedorov, D. G.; Ishida, T.; Kitaura, K. *J. Phys. Chem. A* **2005**, *109*, 2638.
(25) Fedorov, D. G.; Ishimura, K.; Ishida, T.; Kitaura, K.; Pulay, P.; Nagase, S. *J. Comput. Chem.* **2007**, *28*, 1476.
(26) Sawada, T.; Fedorov, D. G.; Kitaura, K. *Int. J. Quantum Chem.* **2009**, *109*, 2033.
(27) Murata, K.; Fedorov, D. G.; Nakanishi, I.; Kitaura, K. *J. Phys. Chem. B* **2009**, *113*, 809.
(28) Fedorov, D. G.; Jensen, J. H.; Deka, R. C.; Kitaura, K. *J. Phys. Chem. A* **2008**, *112*, 11808.
(29) Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2003**, *119*, 3599.
(30) Jiang, N.; Ma, J.; Jiang, Y. *J. Chem. Phys.* **2006**, *124*, 114112.
(31) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293.
(32) Deev, V.; Collins, M. A. *J. Chem. Phys.* **2005**, *122*, 154102.
(33) Collins, M. A.; Deev, V. A. *J. Chem. Phys.* **2006**, *125*, 104104. Collins, M. A.; Deev, V. A. *J. Chem. Phys.* **2009**, *131*, 104103.
(34) Pruitt, S. R.; Addicoat, M. A.; Collins, M. A.; Gordon, M. S. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7752.
(35) Li, W.; Li, S.; Jiang, Y. *J. Phys. Chem. A* **2007**, *111*, 2193.
(36) Hua, W. J.; Fang, T.; Li, W.; Yu, J. G.; Li, S. H. *J. Phys. Chem. A* **2008**, *112*, 10864.
(37) Hua, S. G.; Hua, W. J.; Li, S. H. *J. Phys. Chem. A* **2010**, *114*, 8126.
(38) Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2008**, *128*, 234108.
(39) Wang, Y.; Sosa, C.; Cembran, A.; Truhlar, D. G.; Gao, J. *J. Phys. Chem. B* **2012**, *116*, 6781.
(40) Huang, L.; Massa, L.; Karle, J. *J. Int. Quantum Chem.* **2005**, *103*, 808.
(41) Huang, L.; Massa, L.; Karle, J. *J. Int. Quantum Chem.* **2006**, *106*, 447.
(42) Huang, L.; Massa, L.; Karle, J. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 9541.
(43) Dahlke, E. E; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46.
(44) Dahlke, E. E; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1342.
(45) Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 33.

(46) Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 1573.
(47) Hirata, S.; Valiev, M.; Dupuis, M.; Xantheas, S. S.; Sugiki, S.; Sekino, H. *Mol. Phys.* **2005**, *103*, 2255.
(48) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1342.
(49) Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 33.
(50) Mayhall, N. J.; Raghavachari, K. *J. Chem. Theory Comput.* **2011**, *7*, 1336.
(51) Řezač, J.; Salahub, D. R. *J. Chem. Theory Comput.* **2010**, *8*, 2669.
(52) Wen, S.; Nanda, K.; Huang, Y.; Beran, G. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7578.
(53) Mayhall, N. J.; Raghavachari, K. *J. Chem. Theory Comput.* **2012**, *5*, 1573.
(54) Harrison, M. J.; Burton, N. A.; Hillier, I. H. *J. Am. Chem. Soc.* **1997**, *119*, 12285.
(55) Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185.
(56) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718.
(57) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714.
(58) Zhang, Y.; Lee, T.-S.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46.
(59) Zhang, Y. *J. Chem. Phys.* **2005**, *122*, 24114.
(60) DiLabio, G. A.; Hurley, M. M.; Christiansen, P. A. *J. Chem. Phys.* **2002**, *116*, 9578.
(61) Koga, N.; Morokuma, K. *Chem. Phys. Lett.* **1990**, *172*, 243.
(62) Nasuluzov, V. A.; Ivanova, E. A.; Shor, A. M.; Vayssilov, G. N.; Birkenheuer, U.; Rösch, N. *J. Phys. Chem. B* **2003**, *107*, 2228.
(63) Wang, B.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10556.
(64) Pacios, L. F.; Christiansen, P. A. *J. Chem. Phys.* **1985**, *82*, 2664.
(65) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.
(66) Löwdin, P.-O. *J. Chem. Phys.* **1950**, *18*, 365.
(67) Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211.
(68) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.
(69) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2012**, *8*, 527.
(70) Lin, H.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 3991.
(71) Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. *J. Comput. Chem.* **1990**, *1*, 431.
(72) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.
(73) Walker, R. C.; Crowley, M. F.; Case, D. A. *J. Comput. Chem.* **2008**, *29*, 1019.
(74) Wang, B.; Truhlar, D. G. *J. Chem. Theory Comput.* **2010**, *6*, 359.
(75) (a) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179. (b) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
(76) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.
(77) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620.
(78) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999.
(79) Cieplak, P.; Caldwell, J.; Kollman, P. *J. Comput. Chem.* **2001**, *22*, 1048.
(80) St.-Amant, A.; Cornell, W. D.; Kollman, P. A.; Halgren, T. A. *J. Comput. Chem.* **1995**, *16*, 1483.
(81) Beran, G. *J. Chem. Phys.* **2009**, *130*, 164115.
(82) Suarez, E.; Diaz, N.; Suarez, D. *J. Chem. Theory Comput.* **2009**, *5*, 1667.
(83) Richard, R. M.; Herbert, J. M. *J. Chem. Phys.* **2012**, *137*, 064113.
(84) Assfeld, X.; Rivail, J.-L. *Chem. Phys. Lett.* **1996**, *263*, 100.
(85) Gao, J. *J. Phys. Chem. B* **1997**, *101*, 657.
(86) Hua, S.; Xu, L.; Li, W.; Li, S. *J. Phys. Chem. B* **2011**, *115*, 11462.
(87) Li, W. *J. Chem. Phys.* **2013**, *138*, 14106.
(88) Collins, M. A.; Deev, V. A. *J. Chem. Phys.* **2006**, *125*, 104104.
(89) Bettens, R. P. A.; Lee, A. M. *J. Phys. Chem. A* **2006**, *110*, 8777.
(90) Lee, A. M.; Bettens, R. P. A *J. Phys. Chem. A* **2007**, *111*, 5111.
(91) Le, H.-A.; Lee, A. M.; Bettens, R. P. A. *J. Phys. Chem. A* **2009**, *113*, 10527.
(92) Le, H.-A.; Tan, H.-J.; Ouyang, J. F.; Bettens, R. P. A *J. Chem. Theory Comput.* **2012**, *8*, 469.