A subsystem density-functional theory approach for the quantum chemical treatment of proteins

Christoph R. Jacob[a] and Lucas Visscher[b]
Department of Theoretical Chemistry, Faculty of Sciences, Vrije Universiteit Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, Netherlands

(Received 15 February 2008; accepted 13 March 2008; published online 15 April 2008)

We present an extension of the frozen-density embedding (FDE) scheme within density-functional theory [T. A. Wesolowski and A. Warshel, J. Phys. Chem. **97**, 8050 (1993)] that can be applied to subsystems connected by covalent bonds, as well as a practical implementation of such an extended FDE scheme. We show how the proposed scheme can be employed for quantum chemical calculations of proteins by treating each constituting amino acid as a separate subsystem. To assess the accuracy of the extended FDE scheme, we present calculations for several dipeptides and for the protein ubiquitin. © *2008 American Institute of Physics*. [DOI: 10.1063/1.2906128]

I. INTRODUCTION

The treatment of systems relevant for the understanding of biological processes presents a major challenge for theoretical chemistry.[1] Applications of quantum chemical methods, in particular, density-functional theory (DFT), to systems such as proteins or enzymes are not only hampered by the increase of the required computational power with the size of the studied system[2,3] but also by the fact that the interpretation of the results of such calculations becomes increasingly difficult for large systems since in many cases a lot of superfluous information is obtained (see, e.g., Ref. 4).

Subsystem approaches (for examples, see Refs. 5–11), in which the total system is decomposed into a number of constituting fragments that are each treated individually, are a very attractive alternative to a full quantum chemical treatment of large biological systems for several reasons. First, subsystem methods are in general more efficient than a conventional treatment of the full system. Since the computational effort needed for the calculation of one subsystem is usually independent of the size of the full system, one obtains methods that naturally scale linearly with the system size.[5–7] Second, from a chemist's point of view a partitioning into subsystems provides a more natural way for the interpretation of the results since it offers a picture in terms of the chemical building blocks, such as the individual amino acids constituting a protein or interacting chromophores in natural light harvesting systems.[12,13] Finally, subsystem approaches provide the possibility to focus on interesting parts of the system, such as the active site of an enzyme. Since the subsystems are treated individually, it is easily possible to employ a more accurate treatment only for one or a few selected subsystems of interest.[14–17]

One example of a subsystem approach for the treatment of proteins is the molecular fractionation with conjugate cap (MFCC) scheme developed by Zhang and co-workers.[8,18,19] In this scheme, the protein is partitioned into the constituting amino acids, and capping groups are added to saturate bonds between different subsystems.

For the case of dialanine, this partitioning is illustrated in Fig. 1. The dialanine is cut at the peptide bond and in both fragments a capping group is added. In Fig. 1, these capping groups are chosen as the complementary part of the peptide bond (i.e., a N—H group or a C=O group, respectively) plus a terminating methyl group. This choice can be considered as minimal since it uses the smallest capping groups that preserve the electronic structure of the peptide bond. In the original MFCC scheme, larger caps also including side chains of the neighboring amino acids are used. The two introduced caps can be joined to a "capping molecule," in the case illustrated in the figure, N-methylacetamide. This partitioning can be easily generalized to larger peptides and an arbitrary number of subsystems. In this case, several capping

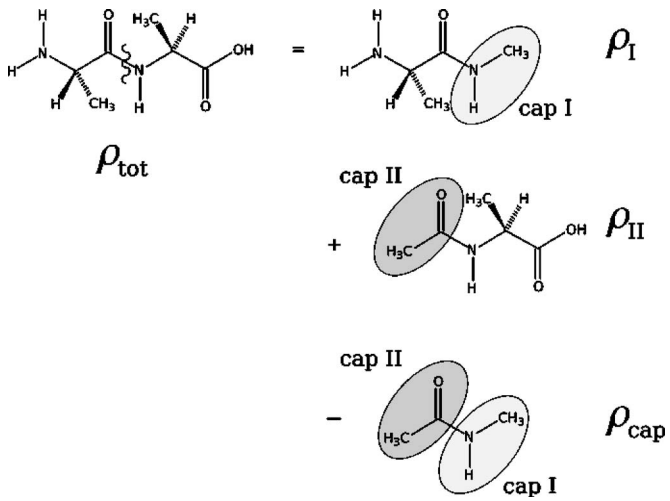

FIG. 1. Partitioning of the electron density in the MFCC scheme, illustrated for dialanine. In the original MFCC scheme, larger caps containing also the side chains of the neighboring amino acid have been used. This partitioning can be easily generalized to larger polypeptides.

[a]Present address: ETH Zurich, Laboratorium für Physikalische Chemie, Wolfgang-Pauli-Strasse 10, 8093 Zurich, Switzerland. Electronic mail: christoph.jacob@phys.chem.ethz.ch.
[b]Electronic mail: visscher@chem.vu.nl.

molecules are introduced that each contain one of the covalent bonds that were cut when defining the subsystems.

Given such a partitioning, the electron density of the full system ρ_{tot} is, in the case of two subsystems illustrated in Fig. 1, obtained as

$$\rho_{\mathrm{tot}}(\boldsymbol{r}) = \rho_{\mathrm{I}}(\boldsymbol{r}) + \rho_{\mathrm{II}}(\boldsymbol{r}) - \rho_{\mathrm{cap}}(\boldsymbol{r}), \tag{1}$$

where ρ_{I} and ρ_{II} are the electron densities of subsystems I and II (including the corresponding capping groups), respectively, and ρ_{cap} is the electron density of the capping molecule. In the general case of n_{sub} subsystems and n_{cap} capping molecules, the total electron density is given by

$$\rho_{\mathrm{tot}}(\boldsymbol{r}) = \sum_{i=1}^{n_{\mathrm{sub}}} \rho_i(\boldsymbol{r}) - \sum_{j=1}^{n_{\mathrm{cap}}} \rho_j^{\mathrm{cap}}(\boldsymbol{r}), \tag{2}$$

where ρ_i is the electron density of subsystem i (including the corresponding capping groups) and ρ_j^{cap} is the electron density of cap molecule j.

In the MFCC scheme, the electron densities ρ_i of the subsystems as well as those of the cap molecules ρ_j^{cap} are obtained from calculations performed for the isolated molecules, i.e., no effect of the other subsystems on these electron densities is included. Therefore, the polarization of the individual subsystems due to their environment, in particular sequentially adjacent amino acids or subsystems forming direct hydrogen bonds, is neglected. In variants of the MFCC scheme, this deficiency has been addressed by using larger capping groups to include the effect of sequentially adjacent amino acids,[19] by including additional caps to model the effects of hydrogen bonding,[20] or by introducing an electrostatic embedding potential of the environment to describe long-range polarization effects.[21,22]

A different path is followed by the fragment molecular orbital (FMO) method by Kitaura and co-workers.[9,23,24] In this scheme, a protein is also partitioned into the constituting amino acids, but no caps are applied. Instead, each subsystem is calculated in the presence of the full electrostatic potential, i.e., the nuclear potential of all atoms as well as the Coulomb potential of the electrons in all subsystems. However, since with such a purely electrostatic embedding potential a description of the covalent bonds between subsystems is not possible, a special treatment of these bonds is necessary. Therefore, for these bonds, a minimal basis set of localized orbitals is used in order to obtain bonds which closely resemble those in an appropriate model system.[23]

However, such a treatment introduces several problems. Since a minimal basis set is used for the bonds connecting different subsystems, a polarization of these bonds is not possible. In addition, such a treatment also limits the size of the basis sets that can be employed to rather small basis sets since larger basis sets containing diffuse functions will offer additional flexibility at these bonds. Furthermore, the purely electrostatic embedding potential will give a rather poor description of short-range interactions such as hydrogen bonding and might also lead to problems with charge leaking into the environment when larger basis sets are used.[25] To improve the accuracy of the calculated total energy, within the FMO scheme one usually applies a correction, which is obtained from calculations on selected pairs of subsystems.[26] However, such a correction will not improve the calculated electron density and molecular properties.

Another conceptually very appealing subsystem approach is offered by the frozen-density embedding (FDE) scheme within DFT (Refs. 14 and 27) and related methods,[17,28,29] which are based on a subsystem formulation of DFT proposed by Cortona.[5] In the FDE scheme, the total electron density is partitioned into possibly overlapping electron densities of subsystems, which are each individually calculated. The effect of the environment on a particular subsystem is included by the use of an effective embedding potential, which contains, in addition to the electrostatic potential of the nuclei in the environment and the Coulomb potential of the electron density of the environment, also a component of the nonadditive exchange-correlation energy and of the nonadditive kinetic energy.[14,27] Compared to other subsystem approaches, the FDE scheme has the advantages that it includes the effect of the environment in an accurate and improvable way,[16,30,31] and that it provides a treatment that is exact in the exact functional limit.[27]

However, the FDE scheme relies on the use of an approximate functional for the nonadditive kinetic energy and its applicability is, therefore, limited by the available kinetic energy functionals. While it has been shown that accurate results can be obtained for van der Waals complexes,[32–34] hydrogen-bonded complexes,[35–38] solvent systems,[15,30,39–41] or even for small parts of proteins,[42–45] FDE currently cannot be applied to subsystems connected by covalent bonds.

One possibility for extending the applicability of FDE to covalently bonded systems is the introduction of capping groups in a similar way as in the MFCC scheme.[46] This way, it would be possible to circumvent the insufficient accuracy of the available kinetic-energy functionals since it is no longer necessary to describe the covalent bonds connecting the subsystems using an approximate nonadditive kinetic-energy functional. Instead, these bonds are replaced by the bonds to the newly introduced caps, which are treated within the individual subsystems. This approach is similar to the introduction of link atoms[47] or connection atoms[48] in combined quantum mechanics/molecular mechanics (QM/MM) schemes.

Such an extension of the FDE scheme has been proposed earlier by Casida and Wesolowski.[46] However, they did not present an implementation of the proposed scheme, and their formalism does not enforce the positivity of the total electron density.

In this paper, we present an extension of the FDE scheme to a partitioning of the electron density similar to the one used in the MFCC scheme. In our extension, we explicitly include a constraint that ensures that the total electron density is positive in any point in space. Furthermore, we present an implementation of the proposed formalism and demonstrate the accuracy of our approach using calculations for dipeptides as well as for the protein ubiquitin.

This work is organized as follows. In Sec. II, it is shown how the FDE scheme can be generalized to the MFCC partitioning. This is followed by a presentation of an implementation of this generalization and of the computational details

in Sec. III. In Sec. IV, applications to different test systems are presented. First, calculations on small dipeptides have been performed in order to assess the accuracy of the generalized FDE scheme. These will be discussed in Sec. IV A. Second, in Sec. IV B, it will be shown how the generalized FDE scheme can be applied to the protein ubiquitin. Finally, concluding remarks are given in Sec. V.

II. THEORY

To obtain a formulation of the FDE scheme that can be employed for the description of covalent bonds even with the available kinetic-energy functionals, we make use of the partitioning of the electron density of the MFCC scheme (cf. Fig. 1) instead of the usual partitioning into two partitions, i.e., the densities of a nonfrozen subsystem and the density of a frozen environment. In the simplest case of only two subsystems, the total electron density is then expressed as

$$\rho_{\mathrm{tot}}(\boldsymbol{r}) = \rho_{\mathrm{I}}(\boldsymbol{r}) + \rho_{\mathrm{II}}(\boldsymbol{r}) - \rho_{\mathrm{cap}}(\boldsymbol{r}), \tag{3}$$

where ρ_{I} and ρ_{II} are the densities of the subsystems I and II, both containing a group of capping atoms to saturate the covalent bond that was cut, and ρ_{cap} is the density of these caps. Since with this partitioning, three partitions are used to represent the total electron density, this extended FDE scheme will be referred to as 3-FDE in the following, to distinguish it from the conventional two-partition FDE scheme.

Using this partitioning, the DFT total energy can be written as a functional of the densities ρ_{I}, ρ_{II}, and ρ_{cap} as[46]

$$\begin{aligned} E[\rho_{\mathrm{I}},\rho_{\mathrm{II}},\rho_{\mathrm{cap}}] = E_{\mathrm{NN}} &+ \int (\rho_{\mathrm{I}}(\boldsymbol{r}) + \rho_{\mathrm{II}}(\boldsymbol{r}) - \rho_{\mathrm{cap}}(\boldsymbol{r}))(v_{\mathrm{I}}^{\mathrm{nuc}}(\boldsymbol{r}) + v_{\mathrm{II}}^{\mathrm{nuc}}(\boldsymbol{r}) - v_{\mathrm{cap}}^{\mathrm{nuc}}(\boldsymbol{r}))d\boldsymbol{r} \\ &+ \frac{1}{2}\int \frac{(\rho_{\mathrm{I}}(\boldsymbol{r}) + \rho_{\mathrm{II}}(\boldsymbol{r}) - \rho_{\mathrm{cap}}(\boldsymbol{r}))(\rho_{\mathrm{I}}(\boldsymbol{r}') + \rho_{\mathrm{II}}(\boldsymbol{r}') - \rho_{\mathrm{cap}}(\boldsymbol{r}))}{|\boldsymbol{r}-\boldsymbol{r}'|} d\boldsymbol{r}d\boldsymbol{r}' + E_{xc}[\rho_{\mathrm{I}} + \rho_{\mathrm{II}} - \rho_{\mathrm{cap}}] \\ &+ T_s[\rho_{\mathrm{I}}] + T_s[\rho_{\mathrm{II}}] - T_s[\rho_{\mathrm{cap}}] + T_s^{\mathrm{nadd}}[\rho_{\mathrm{I}},\rho_{\mathrm{II}},\rho_{\mathrm{cap}}], \end{aligned} \tag{4}$$

where E_{NN} is the nuclear repulsion energy, $v_{\mathrm{I}}^{\mathrm{nuc}}$, $v_{\mathrm{II}}^{\mathrm{nuc}}(\boldsymbol{r})$, and $v_{\mathrm{cap}}^{\mathrm{nuc}}(\boldsymbol{r})$ are the electrostatic potentials of the nuclei in subsystem I, subsystem II, and in the capping molecule, respectively, $E_{xc}[\rho]$ is the exchange-correlation energy functional, $T_s[\rho]$ is the kinetic energy of the noninteracting reference system, and $T_s^{\mathrm{nadd}}[\rho_{\mathrm{I}},\rho_{\mathrm{II}},\rho_{\mathrm{cap}}]$ is the nonadditive kinetic energy, which is defined as

$$\begin{aligned} T_s^{\mathrm{nadd}}[\rho_{\mathrm{I}},\rho_{\mathrm{II}},\rho_{\mathrm{cap}}] = T_s[\rho_{\mathrm{I}} + \rho_{\mathrm{II}} - \rho_{\mathrm{cap}}] \\ - T_s[\rho_{\mathrm{I}}] - T_s[\rho_{\mathrm{II}}] + T_s[\rho_{\mathrm{cap}}]. \end{aligned} \tag{5}$$

To determine ρ_{I} for a given (frozen) density ρ_{II} and a given frozen cap density ρ_{cap}, one has to minimize the above total energy functional with respect to ρ_{I} under the constraint that ρ_{I} integrates to the number of electrons N_{I}. However, for the partitioning chosen here, an additional constraint has to be introduced to ensure that one obtains a total density that is positive in any point in space. An unconstrained minimization, as it was done in an earlier attempt to generalize FDE to three overlapping partitions,[46] leads to total densities that are negative in the cap region and thus unphysical. Therefore, we introduce the constraint that in the region of cap I, i.e., near the cap atoms that were introduced to subsystem I, the density in subsystem I should equal the density of the capping molecule, i.e.,

$$\rho_{\mathrm{I}}(\boldsymbol{r}) = \rho_{\mathrm{cap}}(\boldsymbol{r}) \quad \text{for } \boldsymbol{r} \in V_{\mathrm{I}}^{\mathrm{cap}}, \tag{6}$$

where $V_{\mathrm{I}}^{\mathrm{cap}}$ is some suitably defined cap region. This way it is ensured that after subtracting the density of the capping molecule ρ_{cap}, one still obtains a positive total density. However, it should be noted that no constraint applies to the density of the original subsystem I (i.e., subsystem I excluding the cap).

To take the constraint of Eq. (6) into account, the Lagrange multiplier $\lambda(\boldsymbol{r})$ is introduced. Outside this cap I region, where no constraint applies, $\lambda(\boldsymbol{r})$ is set to zero. This leads to the condition

$$\begin{aligned} 0 = \frac{\delta}{\delta\rho_{\mathrm{I}}}\Bigg[E[\rho_{\mathrm{I}},\rho_{\mathrm{II}},\rho_{\mathrm{cap}}] + \mu\left(\int \rho_{\mathrm{I}}(\boldsymbol{r})d\boldsymbol{r} - N_{\mathrm{I}}\right) \\ + \int \lambda(\boldsymbol{r})[\rho_{\mathrm{I}}(\boldsymbol{r}) - \rho_{\mathrm{cap}}(\boldsymbol{r})]d\boldsymbol{r}\Bigg], \end{aligned} \tag{7}$$

where $\lambda(\boldsymbol{r})=0$ for $\boldsymbol{r} \notin V_{\mathrm{I}}^{\mathrm{cap}}$, and after introducing the Kohn–Sham (KS) orbitals $\{\phi_i^{(\mathrm{I})}\}$ of subsystem I to calculate the noninteracting kinetic energy $T_s[\rho_{\mathrm{I}}]$, one obtains—in analogy to the conventional two-partition FDE scheme—the KS-like equations,

$$\begin{aligned} &\left[-\frac{\nabla^2}{2} + v_{\mathrm{eff}}^{\mathrm{KSCED}}[\rho_{\mathrm{I}},\rho_{\mathrm{II}}](\boldsymbol{r})\right]\phi_i^{(\mathrm{I})}(\boldsymbol{r}) = \epsilon_i\phi_i^{(\mathrm{I})}(\boldsymbol{r}), \\ &i = 1,\ldots,N_{\mathrm{I}}/2, \end{aligned} \tag{8}$$

where the effective potential in these equations is given by

$$v_{\mathrm{eff}}^{\mathrm{KSCED}}[\rho_{\mathrm{I}},\rho_{\mathrm{II}}](\boldsymbol{r}) = v_{\mathrm{eff}}^{\mathrm{KS}}[\rho_{\mathrm{I}}](\boldsymbol{r}) + v_{\mathrm{eff}}^{\mathrm{emb}}[\rho_{\mathrm{I}},\rho_{\mathrm{II}}](\boldsymbol{r}) + \lambda(\boldsymbol{r}) = \begin{cases} v_{\mathrm{eff}}^{\mathrm{KS}}[\rho_{\mathrm{I}}](\boldsymbol{r}) + v_{\mathrm{eff}}^{\mathrm{emb}}[\rho_{\mathrm{I}},\rho_{\mathrm{II}}](\boldsymbol{r}) & \text{for } \boldsymbol{r} \notin V_{\mathrm{I}}^{\mathrm{cap}} \\ v_{\mathrm{cap}}(\boldsymbol{r}) & \text{for } \boldsymbol{r} \in V_{\mathrm{I}}^{\mathrm{cap}} \end{cases} . \tag{9}$$

It consists of the KS potential of subsystem I, $v_{\mathrm{eff}}^{\mathrm{KS}}[\rho_{\mathrm{I}}]$, containing the usual nuclear potential, the Coulomb potential of the electrons, and the exchange-correlation potential, as well as an effective embedding potential,

$$\begin{aligned} v_{\mathrm{eff}}^{\mathrm{emb}}[\rho_{\mathrm{I}},\rho_{\mathrm{II}}](\boldsymbol{r}) &= v_{\mathrm{II}}^{\mathrm{nuc}}(\boldsymbol{r}) - v_{\mathrm{cap}}^{\mathrm{nuc}}(\boldsymbol{r}) \\ &+ \int \frac{\rho_{\mathrm{II}}(\boldsymbol{r}') - \rho_{\mathrm{cap}}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} d\boldsymbol{r}' \\ &+ \left.\frac{\delta E_{\mathrm{xc}}[\rho]}{\delta\rho}\right|_{\rho=\rho_{\mathrm{I}}+\rho_{\mathrm{II}}-\rho_{\mathrm{cap}}} - \left.\frac{\delta E_{\mathrm{xc}}[\rho]}{\delta\rho}\right|_{\rho=\rho_{\mathrm{I}}} \\ &+ \frac{\delta T_s^{\mathrm{nadd}}[\rho_{\mathrm{I}},\rho_{\mathrm{II}},\rho_{\mathrm{cap}}]}{\delta\rho_{\mathrm{I}}}, \end{aligned} \tag{10}$$

and the Lagrange multiplier $\lambda(\boldsymbol{r})$, which is nonzero only within the cap region. This Lagrange multiplier acts as an additional potential, that is defined by the constraint of Eq. (6), i.e., it has to be determined such that Eq. (6) is fulfilled. Therefore, one can equally well determine the full potential in the cap I region according to this constraint, and replace the full potential in the cap I region by a cap potential $v_{\mathrm{cap}}(\boldsymbol{r})$.

As a first guess for this cap potential, the potential that was used in the calculation of the cap density itself can be used, i.e.,

$$v_{\mathrm{cap}}(\boldsymbol{r}) \approx v_{\mathrm{cap}}^{\mathrm{nuc}}(\boldsymbol{r}) + \int \frac{\rho_{\mathrm{cap}}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} d\boldsymbol{r}' + \left.\frac{\delta E_{\mathrm{xc}}[\rho]}{\delta\rho}\right|_{\rho=\rho_{\mathrm{cap}}} . \tag{11}$$

To determine the cap potential more accurately, this first guess can be improved using a modified version of a scheme for the calculation of the KS potential from the electron density, restricted to the cap region. Several such schemes have been proposed in the literature, e.g., by Wang and Parr,[49] by van Leeuwen and Baerends,[50] by Zhao *et al.*,[51] by Colonna and Savin,[52] and by Kadantsev and Scott.[53] Because of its conceptual simplicity, in our implementation we have employed the procedure proposed by van Leeuwen and Baerends[50] and determine the cap potential iteratively as

$$v_{\mathrm{cap}}^{\mathrm{new}}(\boldsymbol{r}) = f \cdot v_{\mathrm{cap}}^{\mathrm{old}}(\boldsymbol{r}) \quad \text{with } f = \frac{\rho_{\mathrm{I}}^{\mathrm{old}}(\boldsymbol{r})}{\rho_{\mathrm{cap}}(\boldsymbol{r})}, \tag{12}$$

where $v_{\mathrm{cap}}^{\mathrm{old}}$ is some approximate cap potential, $\rho_{\mathrm{I}}^{\mathrm{old}}$ is the density in subsystem I calculated using this cap potential, and ρ_{cap} is the target cap density. At points where the density is larger than the target density, this procedure will increase the repulsive potential, and it will decrease the repulsive potential in the case that the density is too small. To ensure that the potential is always positive (i.e., repulsive), the nuclear potential has to be subtracted from the total potential, i.e., the iterative procedure is applied to the total electron potential only, while the nuclear potential is kept constant.

While the procedure outlined above will for given densities ρ_{II} and ρ_{cap} yield a density ρ_{I} of subsystem I, the initial density used for subsystem II can also be updated by employing freeze-and-thaw cycles,[31] i.e., by interchanging the role of the nonfrozen and frozen subsystems while keeping the cap density fixed. As in conventional two-partition FDE, this will in general be necessary to described the polarization of subsystem II by its environment correctly.

It is important to note that, even though in the cap regions the density is constraint to equal the density of the cap molecule, the density is still flexible everywhere when calculated using freeze-and-thaw cycles. Since in the calculation on subsystem I no constraint applies in the cap II region and the same holds in the calculation on subsystem II for the cap I region, the density can, in principle, be polarized everywhere. However, the constraints on the total density that are imposed by the requirement that the subsystem densities equal the cap densities in the respective cap regions introduce a restriction on the possible total densities. Since the number of electrons in the cap regions is fixed, also the number of electrons in subsystems I and II, respectively, is fixed and it is not possible to move electrons between the two subsystems. However, by choosing the cap density appropriately, it is possible to minimize the error caused by this restriction. The consequences of this restriction will be investigated in more detail in Sec. IV.

The 3-FDE scheme described above for the case of two subsystems can be easily generalized to an arbitrary number of subsystems by replacing the frozen density ρ_{II} by the sum of the frozen densities $\Sigma_{i=1}^{n_{\mathrm{fr}}}\rho_i^{\mathrm{fr}}$, and the cap density ρ_{cap} by the sum of the cap densities $\Sigma_{i=1}^{n_{\mathrm{cap}}}\rho_i^{\mathrm{cap}}$, i.e., the total density is expressed as

$$\rho_{\mathrm{tot}}(\boldsymbol{r}) = \rho_{\mathrm{I}}(\boldsymbol{r}) + \sum_{i=1}^{n_{\mathrm{fr}}} \rho_i^{\mathrm{fr}}(\boldsymbol{r}) - \sum_{j=1}^{n_{\mathrm{cap}}} \rho_j^{\mathrm{cap}}(\boldsymbol{r}). \tag{13}$$

It should be noted that one separate cap fragment has to be introduced for each covalent bond that is cut by the introduction of the partitioning into subsystems, and that a cap potential $v_{\mathrm{cap}}^{(j)}$ has to be determined for each cap that is introduced in subsystem I, i.e., there might be several distinct cap regions.

III. IMPLEMENTATION AND COMPUTATIONAL DETAILS

The generalized three-partition FDE scheme described above has been implemented in the Amsterdam density functional (ADF) package.[54,55] The implementation is based on the flexible, fragment-based implementation of FDE in this program package.[16,40]

Our implementation makes use of the fact that, in contrast to most other DFT packages, ADF employs a numerical

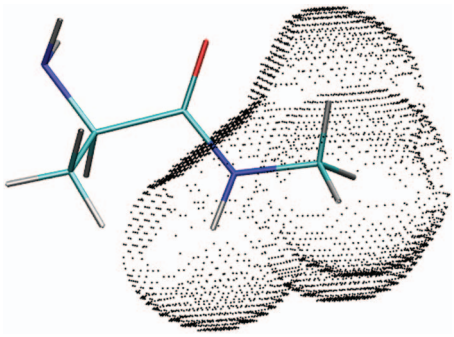

FIG. 2. (Color) Plot of the cap I region of subsystem I in the calculation of dialanine. Graphics: VMD (Ref. 66).

integration scheme for the calculation of matrix element of the KS potential, i.e., the full KS potential is actually available in all grid points used in the numerical integration. This makes it very easy to replace the potential in cap I region by a special cap potential.

This cap I region is in our implementation chosen such that all grid points are included that are (a) closer than $d_{\mathrm{cap}}=3.0$ Bohr from an atom contained in the cap and (b) not closer to an atom that is not part of the cap than to a cap atom. In Fig. 2, the shape of the cap region defined in this way is shown for the case of subsystem I of the dialanine molecule discussed earlier. In the test calculations, we found that the results are not sensitive to the choice of d_{cap}, as long as it is sufficiently large.

Within this cap I region, the full potential is initially replaced by a first guess for the cap potential as given in Eq. (11). This first guess is then improved using the following iterative procedure. During this procedure, the cap potential is only updated if the SCF convergence, measured by the norm of the commutator of the Fock matrix and the density matrix, is below 10^{-3}. If this is not the case, the cap potential of the previous iteration is kept unchanged.

First, a constant shift is applied to the potential until the absolute value of the error in the number of electrons,

$$\Delta N_{\mathrm{cap}}=\int_{V_{\mathrm{cap}}}\rho_{\mathrm{I}}(\boldsymbol{r})d\boldsymbol{r}-\int_{V_{\mathrm{cap}}}\rho_{\mathrm{cap}}(\boldsymbol{r})d\boldsymbol{r}, \tag{14}$$

is smaller than 0.05. This constant shift is chosen as $0.2\Delta N_{\mathrm{cap}}$ (in a.u.).

Second, after the number of electrons in the cap is converged, the potential within the cap region is further refined using the procedure by van Leeuwen and Baerends, as described above, until the error in the cap density, defined as

$$\frac{1}{N_{\mathrm{cap}}}\sqrt{\int_{V_{\mathrm{cap}}}(\rho_{\mathrm{I}}(\boldsymbol{r})-\rho_{\mathrm{cap}}(\boldsymbol{r}))^2d\boldsymbol{r}}, \tag{15}$$

where $N_{\mathrm{cap}}=\int_{V_{\mathrm{cap}}}\rho_{\mathrm{cap}}(\boldsymbol{r})d\boldsymbol{r}$ is the number of electrons in the cap region, is smaller than a given convergence threshold. In our calculations, we employed a threshold of 10^{-4}.

When updating the potential according to Eq. (12), a damping factor of 0.05 is applied, i.e., $f=0.05(\rho_{\mathrm{I}}^{\mathrm{old}}(\boldsymbol{r})/\rho_{\mathrm{cap}}(\boldsymbol{r}))$, and it is ensured that no too large steps are taken by requiring $0.99<f<1.01$. Furthermore, the cap potential is only adjusted at grid points where $\rho_{\mathrm{cap}}(\boldsymbol{r})>0.01$ in order to avoid convergence problems.

In the calculation of the embedding potential, care has to be taken in the regions of cap II (cf. Fig. 1) of the frozen subsystem II. In this region, the densities ρ_{II} and ρ_{cap} should be equal and the corresponding contributions to the embedding potential should cancel. However, if approximate densities are used, this condition is not fulfilled and large contributions to the embedding potential arise in the cap II region, where also the density of the active subsystem I is large. Therefore, the contributions of functions on the atoms contained in cap II to the electrostatic part of the embedding potential as well as the corresponding frozen density in this cap region are explicitly set to zero. It should be noted that this is not necessary as soon as one freeze-and-thaw cycle is employed since in this case, the constraint applied in the previous calculation ensures that $\rho_{\mathrm{II}}=\rho_{\mathrm{cap}}$.

All 3-FDE calculations were performed as follows. As a starting point, the density of all fragments, including the cap fragments, was calculated for the isolated molecules in the absence of any environment. The total density obtained from these calculations corresponds to the MFCC scheme. Using the embedding potential constructed from these densities of the isolated molecules, the densities of the individual subsystems were calculated in separate 3-FDE calculations. To introduce no dependency on the ordering of the subsystems, these updated densities were not used in the construction of the embedding potential for the other subsystems. The density obtained using this procedure will be referred to as 3-FDE(0) in the following.

In the calculations employing freeze-and-thaw cycles, these freeze-and-thaw cycles were performed by successively updating the electron densities of the subsystems, and the updated density was immediately used in the construction of the embedding potential for calculation of the next subsystem. This can be repeated several times until the total density is converged. These calculations will be referred to as 3-FDE(n), where n is the number of freeze-and-thaw cycles employed for each subsystem. In all these calculations, the cap densities were kept fixed at the densities calculated for the isolated cap fragments. It should be noted that the intermediate results of these 3-FDE(n) calculations ($n\geqslant 1$) will in general depend on the order, in which the subsystem densities are updated. However, once the update procedure has converged, the total density and other properties of interest should not depend on the update order.

For the exchange-correlation potential, we applied the generalized-gradient approximation functional BP86, consisting of the exchange functional by Becke[56] and the correlation functional by Perdew.[57] In the calculations on dipeptides presented in Sec. IV A, we employed the TZ2P basis set from the ADF basis set library which is of triple-ζ quality and contains two sets of polarization functions. In the calculations on ubiquitin presented in Sec. IV B, the DZP basis set from the ADF basis set library, which is of double-ζ quality and contains one set of polarization functions, was used.

In all 3-FDE calculations, we applied the PW91k kinetic-energy functional[58] for the nonadditive kinetic-energy component of the embedding potential. We used the

monomolecular basis set expansion,[59] i.e., only basis functions of atoms in the calculated subsystems were included.

For partitioning the investigated systems and to automate the MFCC and 3-FDE calculations, we made use of the recently developed PYADF scripting framework.[60] For the manipulation of structures and geometric coordinates, PYADF relies on the OPENBABEL package.[61,62]

IV. RESULTS AND DISCUSSION

To assess the accuracy of both the MFCC scheme and the 3-FDE scheme described above, one can compare the electron density calculated using these schemes to the electron density calculated in a conventional, supermolecular DFT calculation. Comparing the electron densities directly is preferable to comparing, e.g., total energies or molecular properties, since it will reveal possible problems that might otherwise remain hidden.[34,38]

For the FDE scheme, the calculated electron density should in the exact limit be identical to the one from the conventional, supermolecular DFT calculation.[27,59] However, because of the introduced approximations, i.e., (a) the use of an approximate kinetic-energy functional, (b) differences in the basis set expansion,[35] (c) the constraints imposed by the use of a given cap density, and possibly (d) the use of no or only a limited number of freeze-and-thaw cycles, there will be differences between the two.

To measure the size of the difference between the electron density ρ_{KS} calculated in a supermolecular DFT calculation and the electron density ρ_{FDE} calculated in the 3-FDE calculation (or the electron density ρ_{MFCC} calculated using the MFCC scheme), the following quantities can be used[63] (N_{tot} is the total number of electrons):

(1) The integrated absolute error in the electron density,

$$\Delta^{\mathrm{abs}} = \frac{1}{N_{\mathrm{tot}}} \int |\rho_{\mathrm{KS}}(\boldsymbol{r}) - \rho_{\mathrm{FDE}}(\boldsymbol{r})| d\boldsymbol{r}. \tag{16}$$

(2) The integrated root mean square error in the electron density,

$$\Delta^{\mathrm{rms}} = \frac{1}{N_{\mathrm{tot}}} \sqrt{\int (\rho_{\mathrm{KS}}(\boldsymbol{r}) - \rho_{\mathrm{FDE}}(\boldsymbol{r}))^2 d\boldsymbol{r}}. \tag{17}$$

(3) The magnitude of the error in the dipole moment,

$$|\Delta\mu| = |\mu_{\mathrm{KS}} - \mu_{\mathrm{FDE}}| = \left| \int (\rho_{\mathrm{KS}}(\boldsymbol{r}) - \rho_{\mathrm{FDE}}(\boldsymbol{r})) \boldsymbol{r}\, d\boldsymbol{r} \right|. \tag{18}$$

The first two of these measures, Δ^{abs} and Δ^{rms}, both indicate the absolute size of the difference between the two densities. The third measure $|\Delta\mu|$ depends not only on the absolute size of the difference density but also on its spatial distribution, i.e., it indicates "how far" electron density has moved between the different calculations.

Both Δ^{abs} and Δ^{rms} were calculated by numeric integration on the same integration grid that was used by ADF in the supermolecular DFT calculation. For the calculation of $|\Delta\mu|$, the dipole moments calculated analytically by ADF were used.

TABLE I. Integrated absolute error Δ^{abs} in the electron density, root mean square error Δ^{rms} in the electron density, and magnitude of the error in the dipole moment $|\Delta\mu|$ in the MFCC and 3-FDE calculations of the investigated test systems (see text for details).

		$\Delta^{\mathrm{abs}} \times 10^3$	$\Delta^{\mathrm{rms}} \times 10^3$	$\lvert\Delta\mu\rvert$ (D)
Ala–Ala	MFCC	0.45	0.034	0.09
	3-FDE(0)	0.70	0.040	0.14
His–Leu	MFCC	0.66	0.027	0.26
	3-FDE(0)	0.49	0.031	0.19
Ala–Ala+env	MFCC	4.04	0.196	3.47
	3-FDE(0)	1.36	0.070	0.87
	3-FDE(1)	0.93	0.055	0.37
	3-FDE(2)	0.90	0.052	0.27
H^+–His–Leu	MFCC	2.20	0.094	1.15
	3-FDE(0)	1.66	0.106	0.79
H^+–Ala–Ala	MFCC	2.98	0.172	0.84
	3-FDE(0)	3.17	0.243	0.69

A. Test calculations on dipeptides

As a first set of test systems we have used the dipeptides consisting of two alanines (Ala–Ala) and of histidine and leucine (His–Leu), as well as the corresponding protonated dipeptides H^+–Ala–Ala and H^+–His–Leu. Dialanine has been protonated at the terminal amino group, while in H^+–His–Leu, the imidazole ring in the side chain of histidine has been protonated. In the MFCC and 3-FDE calculations, these have been partitioned into two subsystems, as shown in Fig. 1.

As a simple model for the environment of a peptide bond in a protein, we have also performed calculations for a complex of a dialanine molecule with two *N*-methylacetamide molecules, which form hydrogen bonds to dialanine, as in an α-helix or β-sheet substructure of a protein. In the MFCC and 3-FDE calculations, this complex has been partitioned into four subsystems: the two *N*-methylacetamide molecules and the two subsystems of dialanine.

For all test systems, the structures have been optimized in a conventional supermolecular DFT calculation. As shown in Fig. 1, *N*-methylacetamide has been used as capping molecule. It should be noted that in the original MFCC scheme[8,18] and its variants,[19] larger caps which at least also include the side chains of the involved amino acids are used. With such larger caps, the MFCC scheme would be exact for the dipeptides considered here. For the atoms in the cap molecule that are also present in the original molecule, the same atomic coordinates have been used. The remaining hydrogen atoms have been added using the hydrogen addition routine of the OPENBABEL package.

The calculated measures for the error in the electron density for all investigated test systems are given in Table I. Isosurface plots of the difference densities are shown in Figs. 3–6.

For dialanine, already the simple MFCC scheme is very accurate and yields an electron density that agrees well with the supermolecular DFT calculation. The only difference visible in the isosurface plot of the difference density in Fig.

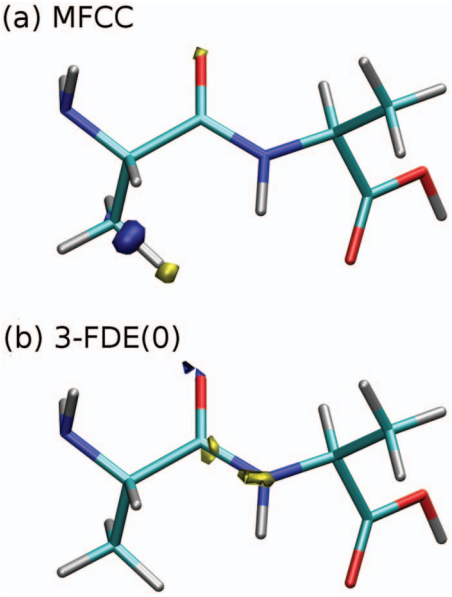

FIG. 3. (Color) Isosurface plots (contour value of 0.001 a.u.) of the difference densities between (a) the MFCC calculation and (b) the 3-FDE(0) calculation and the conventional supermolecular DFT calculation for the Ala–Ala dipeptide. Graphics: VMD (Ref. 66).

3(a) is the polarization of a methyl C–H bond in the N-terminal alanine. This polarization is caused by the polar carboxylic acid group at the C-terminus. Since both subsystems are treated as isolated in the MFCC scheme, it misses this effect.

In the 3-FDE calculation, the error with respect to the supermolecular DFT calculation is similarly small. In this case, both subsystems feel the effect of the other subsystem through the embedding potential. Therefore, the polarization effects missing in the MFCC scheme are correctly captured. However, Fig. 3(b) shows that a small error is introduced in the cap regions. This error is mainly caused by the thresholds applied on the convergence of the cap potentials.

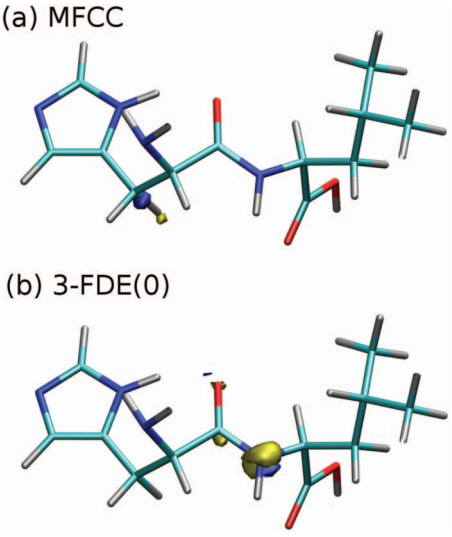

FIG. 4. (Color) Isosurface plots (contour value of 0.001 a.u.) of the difference densities between (a) the MFCC calculation and (b) the 3-FDE(0) calculation and the conventional supermolecular DFT calculation for the His–Leu dipeptide. Graphics: VMD (Ref. 66).

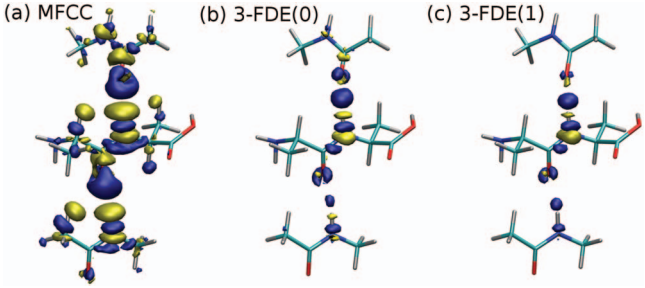

FIG. 5. (Color) Isosurface plots (contour value of 0.001 a.u.) of the difference densities between (a) the MFCC calculation, (b) the 3-FDE(0) calculation, and (c) the 3-FDE(1) calculation employing one freeze-and-thaw cycle and the conventional supermolecular DFT calculation for the complex of dialanine with two *N*-methylacetamide molecules. Graphics: VMD (Ref. 66).

The situation is similar for the histidine-leucine dipeptide. Both with the MFCC scheme and with the 3-FDE scheme, the errors in the electron density are small. As can be seen from the plots in Fig. 4, the MFCC scheme misses the polarization of a C–H bond in the histidine subsystem due to the carboxylic acid group in the leucine subsystem, while the 3-FDE scheme introduces small differences in the cap regions, especially at the peptide bond between the two subsystems.

For the complex of dialanine with two *N*-methylacetamide molecules (Ala–Ala+env), the electron density calculated with the MFCC scheme significantly deviates from the density obtained from the supermolecular DFT calculation. As the magnitude of the error in the dipole moment of 3.47 D shows, these deviations are unacceptably large. The isosurface plot of the difference density in Fig. 5(a) shows the source of this error. Since all subsystems are treated as isolated, the polarization of the dialanine by the neighboring *N*-methylacetamide molecules is not captured.

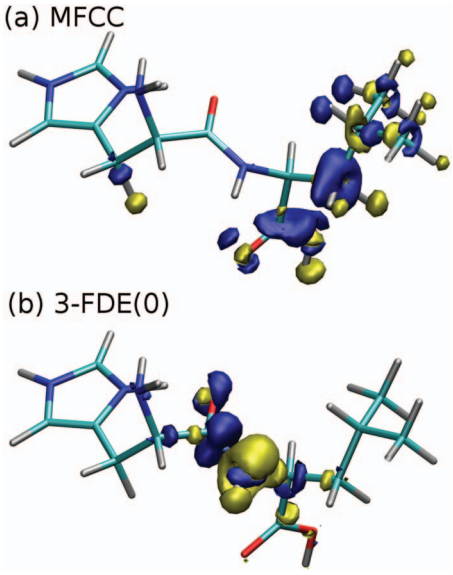

FIG. 6. (Color) Isosurface plots (contour value of 0.001 a.u.) of the difference densities between (a) the MFCC calculation and (b) the 3-FDE(0) calculation and the conventional supermolecular DFT calculation for the protonated H^+–His–Leu dipeptide. Graphics: VMD (Ref. 66).

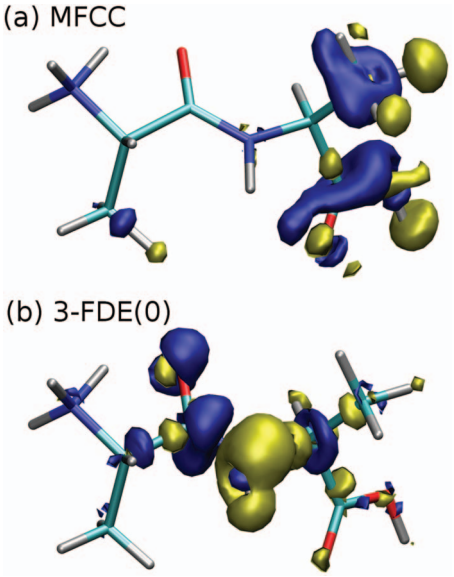

FIG. 7. (Color) Isosurface plots (contour value of 0.001 a.u.) of the difference densities between (a) the MFCC calculation and (b) the 3-FDE(0) calculation and the conventional supermolecular DFT calculation for the protonated H^+–Ala–Ala dipeptide. Graphics: VMD (Ref. 66).

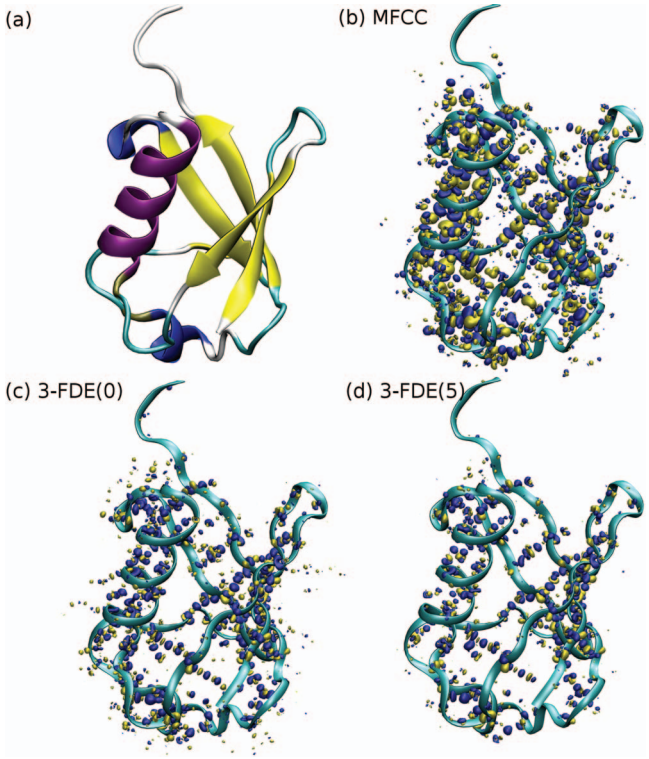

FIG. 8. (Color) (a) Plot of the secondary structure of ubiquitin. [(b)–(d)] Isosurface plots (contour value of 0.002 a.u.) of the difference densities between (b) the MFCC calculation, (c) the 3-FDE(0) calculation, and (d) the 3-FDE(5) calculation and the conventional supermolecular DFT calculation for ubiquitin. Graphics: VMD (Ref. 66).

Furthermore, the hydrogen bonds between the dialanine and the *N*-methylacetamide molecules are not correctly described.

In the 3-FDE(0) calculation, this error is reduced significantly [cf. Fig. 5(b)] since the FDE scheme is able to describe both the polarization effects as well as the effects of hydrogen bonding.[35,38,64] The integrated absolute error and the integrated rms error are both reduced by roughly a factor of 3, and the magnitude of the error in the dipole moment is reduced by a factor of approximately 4.

This error can be further reduced by performing freeze-and-thaw cycles in the 3-FDE calculation. As has been shown earlier, freeze-and-thaw cycles are crucial for an accurate description of hydrogen bonds within FDE.[38,64] Compared to the MFCC calculation, the 3-FDE calculation employing freeze-and-thaw cycles reduces the magnitude of the error in the dipole moment by more than an order of magnitude to 0.27 D. This is only slightly larger than the error for the dipeptides discussed above. As Fig. 5(c) indicates, the remaining error is caused by deficiencies in the description of the hydrogen bonds in the FDE scheme. This description could possibly still be improved by including basis functions on some atoms of the frozen subsystems.[38]

The most difficult test cases are the protonated dipeptides because in these systems the effect of the protonated amino acid on the other subsystem is very large. As can be clearly seen for the protonated histidine-leucine dipeptide in Fig. 6(a), the MFCC scheme misses these polarization effects. This leads to a large error in the calculated electron density.

In the 3-FDE calculation, these polarization effects are correctly described since the effect of the protonated histidine is included in the embedding potential felt by the leucine subsystem. However, as Fig. 6(b) shows, the 3-FDE scheme introduces an additional error at the peptide bond. This error stems from the fact that because of the requirement that for both subsystems the density in the cap region equals the density of the isolated cap molecule, the distribution of the electrons between the two subsystems is fixed by the cap density. Therefore, the polarization of the bond between the two subsystems cannot be described correctly within the 3-FDE scheme in this case.

For the protonated dialanine, both effects are even more pronounced. In the MFCC calculation [cf. Fig. 7(a)], a large error appears because the polarization of the C-terminal alanine by the protonated N-terminal alanine is not captured. In the 3-FDE calculation [cf. Fig. 7(b)], a similarly large error is introduced because the polarization of the bond between the two subsystems is not correctly described.

In summary, the test calculations presented above show a good agreement of both the electron densities calculated with the MFCC scheme and the 3-FDE scheme with the ones from the conventional supermolecular DFT calculations for the simple dipeptides, in which polarization effects are small. The effects of polarization due to neighboring N-methylacetamide molecules (as a simple model of a protein environment) as well as the effects of hydrogen bonding are well described by the 3-FDE scheme, while these effects are missing in the MFCC scheme.

If there are very strong polarization effects, as in the case of the protonated dipeptides, the 3-FDE scheme is able to model the polarization effects on the other subsystems, but it misses the polarization of the bond between the subsystems. However, it should be mentioned that also other subsystem methods, such as the FMO scheme, in which a minimal basis

set of localized orbitals is used to describe these bonds,[23] suffer from the same problem. Furthermore, it should be noted that it is possible to polarize the electron density at the atoms on either side of these bonds, as the example of the Ala–Ala+env complex shows.

Within the 3-FDE scheme, it is possible to achieve a better description of the bond between the subsystems by using a cap density in which the electron density of this bond resembles the true charge distribution more closely. This could for instance be achieved by employing larger caps, as it is done in the MFCC I and MFCC II schemes,[8,18,19] or by applying an appropriate embedding potential in the calculation of the cap density.

B. Calculation of the electron density of the protein ubiquitin

To test the applicability and accuracy of the proposed 3-FDE scheme for larger systems, we performed calculations for the protein ubiquitin. Ubiquitin is a small protein consisting of 76 amino acids, which contains the most important secondary structural elements, an α-helix, a β-sheet, a short 3_{10}-helix, and several unstructured loops. The secondary structure of ubiquitin is shown in Fig. 8(a). Because of its moderate size (1231 atoms, of which 602 are nonhydrogen atoms), it is still possible to perform a conventional, supermolecular DFT calculation on the full system. This calculation can, as for the small test systems discussed in the previous section, be used as a reference for both the MFCC and the 3-FDE scheme.

We employed the x-ray structure of human ubiquitin (pdb code 1UBQ),[65] and no water molecules were included. The hydrogen atoms were added using the hydrogen addition routine of OPENBABEL. For all side chains, we chose to use the neutral form. This might not agree with the situation in the crystal or in solution, but the purpose here is only the comparison of the different subsystem methods, and for such a comparison, it is only important that the same structure and protonation state are used in all cases.

The protein was partitioned into 76 subsystems, with each constituting amino acid as one subsystem, and 75 capping molecules were introduced. The positions of the cap atoms were determined as described for the dipeptides in the previous section. Again, it should be noted that we employ smaller caps as in the original MFCC scheme, where also the side chains of the adjacent amino acids are included in the caps.

The calculated measures for the error in the electron density calculated with both the MFCC and the 3-FDE scheme are given in Table II, and isosurface plots of the difference density are shown in Fig. 8.

With the MFCC scheme, a large error in the electron density is obtained. As already shown in the previous section, the MFCC scheme misses the polarization of the individual amino acids caused by their environment, as well as the effects of hydrogen bonds between different subsystems. Within a protein, these effects become even more important than for the simple test systems studied in the previous section. This can also be seen from the plot of the difference density in Fig. 8(b). Most striking is the very large magnitude of the error in the dipole moment of 29.6 D. This large error can be explained by the insufficient description of hydrogen bonding. Since within the α-helix and β-sheet substructures the hydrogen bonds are aligned, the error in their description adds up to a large error in the dipole moment.

TABLE II. Integrated absolute error Δ^{abs} in the electron density, root mean square error Δ^{rms} in the electron density, and magnitude of the error in the dipole moment $|\Delta\mu|$ in the MFCC and 3-FDE calculations of ubiquitin (see text for details).

	$\Delta^{abs}\times 10^3$	$\Delta^{rms}\times 10^3$	$\lvert\Delta\mu\rvert$ (D)
MFCC	5.05	0.046	29.60
3-FDE(0)	3.92	0.030	4.57
3-FDE(1)	2.72	0.026	10.28
3-FDE(2)	2.66	0.026	5.89
3-FDE(3)	2.66	0.026	5.06
3-FDE(4)	2.66	0.026	5.60
3-FDE(5)	2.66	0.026	5.21

In the 3-FDE(0) calculation, a large improvement with respect to the MFCC calculation can be seen. Both the integrated absolute error and the rms error in electron density significantly decrease, and the magnitude of the error in the dipole moment is reduced by roughly a factor of 6. The improvement is also obvious from the plot of the difference density in Fig. 8(c).

Further freeze-and-thaw cycles improve the reduction of the integrated absolute error and the RMS error in electron density slightly. However, the magnitude of the error in the dipole moment increases after one freeze-and-thaw cycle but decreases again in further cycles.

As Fig. 8(d) shows, the remaining error is mainly caused by the description of the hydrogen bonds. As already seen in the previous section, the FDE scheme still introduces a small error for hydrogen bonds, even though it offers a significant improvement over the MFCC scheme. Further improvements might be achieved by using a larger basis set and by including basis functions on selected atoms of the frozen subsystem.[38]

V. CONCLUSIONS

We have developed an extension of the FDE scheme, that is—in contrast to the conventional FDE scheme—applicable also for subsystems connected by covalent bonds. By introducing capping groups in analogy to the MFCC scheme,[8] it is possible to express the electron density of the total system as a sum of the densities of the individual subsystems (including suitable capping groups) minus the densities of capping molecules, which compensate the electron density of the introduced caps.

The densities of all subsystems are optimized individually, subject to an effective embedding potential which represents the effect of the environment. In contrast to an earlier attempt,[46] in this optimization a constraint is applied to the electron density in the regions of the caps, in order to ensure the positivity of the total electron density. To satisfy this constraint, a special cap potential has to be determined in these regions.

Test calculations on dipeptides show that this 3-FDE scheme can accurately model both the polarization of a subsystem due to its environment as well as the effects of hydrogen bonding. However, since the electron density in the cap region is constrained, the distribution of the electrons between the subsystems is also fixed, and a polarization of the bond between the subsystems cannot be accounted for. However, similar problems also appear in related subsystem methods.[23] For the 3-FDE scheme, this can possibly be rectified by employing a more suitable cap density, which will be explored in future work.

The 3-FDE scheme developed here can be applied for the quantum chemical treatment of proteins, which can be partitioned into the constituting amino acids. The electron density of each amino acid is then separately calculated in the presence of the embedding potential of all other subsystems. This is demonstrated for the protein ubiquitin, where we show that the 3-FDE scheme is able to reproduce the electron density of the conventional DFT calculation on the full protein with good accuracy. As we can show, the 3-FDE scheme is significantly more accurate than the simpler MFCC scheme, especially because it is able to describe the hydrogen bonds within α-helix and β-sheet structures adequately.

However, since the purpose of the implementation used in this work was the testing of the proposed 3-FDE scheme, the current implementation was not developed with a main focus on computational efficiency. Therefore, the computer time needed for the 3-FDE calculation of one subsystem is still approximately 25–50 times larger than for a calculation on the isolated subsystem. The main reason for this is the slow convergence of the procedure used to determine the cap potentials. Usually, in total 100–150 SCF iterations are needed in the 3-FDE calculation of one subsystem. It should, however, be straightforward to speed up the 3-FDE calculations significantly by expanding the cap potential in a suitable basis set instead of calculating it in each integration grid point. Further improvements can be achieved by employing direct inversion of the iterative subspace (DIIS) in the SCF procedure, which in the current implementation cannot be used in combination with the determination of the cap potentials.

The subsystem formalism presented in this paper is particularly suited for applications where focus can be placed on a small part of a protein, such as an active site of an enzyme, or for the calculation of rather localized molecular properties. In this case, only a few subsystems have to be accurately treated by employing 3-FDE with several freeze-and-thaw cycles, while for the electron density of subsystems further away from the region of interest, the simple MFCC approximation can be used. Such a strategy is similar to the one chosen in applications of FDE for modeling solvent effects on molecular properties.[15,30,39–41] This might open the way to a number of interesting applications of the 3-FDE scheme proposed here to systems of biological relevance.

ACKNOWLEDGMENTS

The authors would like to thank Johannes Neugebauer (ETH Zurich) for helpful comments and discussions. The authors further thank The Netherlands Organization for Scientific Research (NWO) for financial support via the TOP program and gratefully acknowledge computer time provided by the Dutch National Computing Facilities (NCF).

[1] *Atomistic Approaches in Modern Biology*, Topics in Current Chemistry Vol. 268, edited by M. Reiher (Springer, Berlin, 2007).
[2] C. J. Cramer, *Essentials of Computational Chemistry* (Wiley, New York, 2002).
[3] F. Jensen, *Introduction to Computational Chemistry*, 2nd ed. (Wiley, Chichester, 2007).
[4] C. Herrmann, J. Neugebauer, and M. Reiher, New J. Chem. **31**, 818 (2007).
[5] P. Cortona, Phys. Rev. B **44**, 8454 (1991).
[6] M. Iannuzzi, B. Kirchner, and J. Hutter, Chem. Phys. Lett. **421**, 16 (2006).
[7] F. Shimojo, R. K. Kalia, A. Nakano, and P. Vashishta, Comput. Phys. Commun. **167**, 151 (2005).
[8] D. W. Zhang and J. Z. H. Zhang, J. Chem. Phys. **119**, 3599 (2003).
[9] D. G. Fedorov and K. Kitaura, J. Phys. Chem. A **111**, 6904 (2007).
[10] R. P. A. Bettens and A. M. Lee, J. Phys. Chem. A **110**, 8777 (2006).
[11] H. M. Netzloff and M. A. Collins, J. Chem. Phys. **127**, 134113 (2007).
[12] J. Neugebauer, J. Chem. Phys. **126**, 134116 (2007).
[13] J. Neugebauer, J. Phys. Chem. B **112**, 2207 (2008).
[14] T. A. Wesolowski and A. Warshel, J. Phys. Chem. **97**, 8050 (1993).
[15] J. Neugebauer, M. J. Louwerse, E. J. Baerends, and T. A. Wesolowski, J. Chem. Phys. **122**, 094115 (2005).
[16] Ch. R. Jacob, J. Neugebauer, and L. Visscher, J. Comput. Chem. **29**, 1011 (2008).
[17] N. Govind, Y. A. Wang, and E. A. Carter, J. Chem. Phys. **110**, 7677 (1999).
[18] A. M. Gao, D. W. Zhang, J. Z. Zhang, and Y. Zhang, Chem. Phys. Lett. **394**, 293 (2004).
[19] Y. Mei, D. W. Zhang, and J. Z. H. Zhang, J. Phys. Chem. A **109**, 2 (2005).
[20] Y. Mei, E. L. Wu, K. L. Han, and J. Z. H. Zhang, Int. J. Quantum Chem. **106**, 1267 (2006).
[21] X. He and J. Z. H. Zhang, J. Chem. Phys. **124**, 184703 (2006).
[22] Y. Mei, C. Ji, and J. Z. H. Zhang, J. Chem. Phys. **125**, 094906 (2006).
[23] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Lett. **313**, 701 (1999).
[24] T. Nakano, T. Kaminuma, T. Sato, Y. Akiyama, M. Uebayasi, and K. Kitaura, Chem. Phys. Lett. **318**, 614 (2000).
[25] A. Laio, J. VandeVondele, and U. Rothlisberger, J. Chem. Phys. **116**, 6941 (2002).
[26] K. Kitaura, T. Sawai, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Lett. **312**, 319 (1999).
[27] T. A. Wesolowski, in *Computational Chemistry: Reviews of Current Trends*, edited by J. Leszczynski (World Scientific, Singapore, 2006), Vol. 10.
[28] E. V. Stefanovich and T. N. Truong, J. Chem. Phys. **104**, 2946 (1995).
[29] T. Klüner, N. Govind, Y. A. Wang, and E. A. Carter, Phys. Rev. Lett. **86**, 5954 (2001).
[30] Ch. R. Jacob, J. Neugebauer, L. Jensen, and L. Visscher, Phys. Chem. Chem. Phys. **8**, 2349 (2006).
[31] T. A. Wesolowski and J. Weber, Chem. Phys. Lett. **248**, 71 (1996).
[32] T. A. Wesolowski, Y. Ellinger, and J. Weber, J. Chem. Phys. **108**, 6078 (1998).
[33] T. A. Wesolowski and F. Tran, J. Chem. Phys. **118**, 2072 (2003).
[34] Ch. R. Jacob, T. A. Wesolowski, and L. Visscher, J. Chem. Phys. **123**, 174104 (2005).
[35] T. A. Wesolowski, J. Chem. Phys. **106**, 8516 (1997).
[36] T. A. Wesolowski, J. Am. Chem. Soc. **126**, 11444 (2004).
[37] J. Neugebauer and E. J. Baerends, J. Phys. Chem. A **110**, 8786 (2006).
[38] K. Kiewisch, G. Eickerling, M. Reiher, and J. Neugebauer, J. Chem. Phys. **128**, 044114 (2008).
[39] J. Neugebauer, M. J. Louwerse, P. Belanzoni, T. A. Wesolowski, and E. J. Baerends, J. Chem. Phys. **123**, 114101 (2005).
[40] J. Neugebauer, Ch. R. Jacob, T. A. Wesolowski, and E. J. Baerends, J. Phys. Chem. A **109**, 7805 (2005).
[41] R. E. Bulo, Ch. R. Jacob, and L. Visscher, J. Phys. Chem. A **112**, 2640 (2008).
[42] M. Štrajbl, G. Hong, and A. Warshel, J. Phys. Chem. B **106**, 13333

(2002).
[43] M. H. M. Olsson, G. Hong, and A. Warshel, J. Am. Chem. Soc. **125**, 5025 (2003).
[44] G. Hong, E. Rosta, and A. Warshel, J. Phys. Chem. B **110**, 19570 (2006).
[45] Y. Xiang and A. Warshel, J. Phys. Chem. B **112**, 1007 (2008).
[46] M. E. Casida and T. A. Wesolowski, Int. J. Quantum Chem. **96**, 577 (2004).
[47] M. J. Field, P. A. Bash, and M. Karplus, J. Comput. Chem. **11**, 700 (1990).
[48] I. Antes and W. Thiel, J. Phys. Chem. A **103**, 9290 (1999).
[49] Y. Wang and R. G. Parr, Phys. Rev. A **47**, R1591 (1993).
[50] R. van Leeuwen and E. J. Baerends, Phys. Rev. A **49**, 2421 (1994).
[51] Q. Zhao, R. C. Morrison, and R. G. Parr, Phys. Rev. A **50**, 2138 (1994).
[52] F. Colonna and A. Savin, J. Chem. Phys. **110**, 2828 (1999).
[53] E. S. Kadantsev and M. Scott, Phys. Rev. A **69**, 012502 (2004).
[54] Amsterdam density functional program (ADF), Theoretical Chemistry, Vrije Universiteit Amsterdam, 2007 (http://www.scm.com).
[55] G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders, and T. Ziegler, J. Comput. Chem. **22**, 931 (2001).
[56] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[57] J. P. Perdew, Phys. Rev. B **33**, 8822 (1986).
[58] A. Lembarki and H. Chermette, Phys. Rev. A **50**, 5328 (1994).
[59] T. A. Wesolowski, H. Chermette, and J. Weber, J. Chem. Phys. **105**, 9182 (1996).
[60] Ch. Jacob, R. E. Bulo, and L. Visscher, PYADF, A Scripting Framework for Quantum Chemistry, 2008.
[61] The OPENBABEL package, Version 2.1.1 (http://openbabel.sourceforge.net/).
[62] G. R. Hutchison, P. Murray-Rust, H. Rzepa, C. Steinbeck, J. Wegner, and E. L. Willighagen, J. Chem. Inf. Model. **46**, 991 (2006).
[63] Y. A. Bernard, M. Dulak, J. W. Kaminski, and T. A. Wesolowski, J. Phys. A **41**, 055302 (2008).
[64] R. Kevorkyants, M. Dulak, and T. A. Wesolowski, J. Chem. Phys. **124**, 024104 (2006).
[65] S. Vijay-Kumar, C. E. Bugg, and W. J. Cook, J. Mol. Biol. **194**, 531 (1987).
[66] W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graphics **14**, 33 (1996).