JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Quantum-Chemical Electron Densities of Proteins and of Selected Protein Sites from Subsystem Density Functional Theory

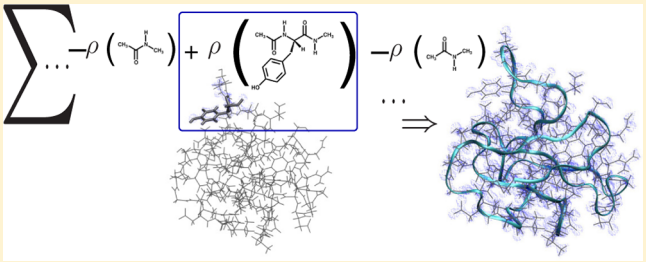
Karin Kiewisch,[†] Christoph R. Jacob,*[,‡] and Lucas Visscher*[,†]

[†]Amsterdam Center for Multiscale Modeling, VU University Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands

[‡]Center for Functional Nanostructures and Institute of Physical Chemistry, Karlsruhe Institute of Technology (KIT), Wolfgang-Gaede-Str. 1a, 76131 Karlsruhe, Germany

Supporting Information

ABSTRACT: The ability to calculate accurate electron densities of full proteins or of selected sites in proteins is a prerequisite for a fully quantum-mechanical calculation of protein−protein and protein−ligand interaction energies. Quantum-chemical subsystem methods capable of treating proteins and other biomolecular systems provide a route to calculate the electron densities of proteins efficiently and further make it possible to focus on specific parts. Here, we evaluate and extend the 3-partition frozen-density embedding (3-FDE) scheme [Jacob, C. R.; Visscher, L. *J. Chem. Phys.* **2008**, *128*, 155102] for this purpose. In particular, we have extended this scheme to allow for the treatment of disulfide bridges and charged amino acid residues and have introduced the possibility to employ more general partitioning schemes. These extensions are tested both for the prediction of full protein electron densities and for focusing on the electron densities of a selected protein site. Our results demonstrate that 3-FDE is a promising tool for the fully quantum-chemical treatment of proteins.



1. INTRODUCTION

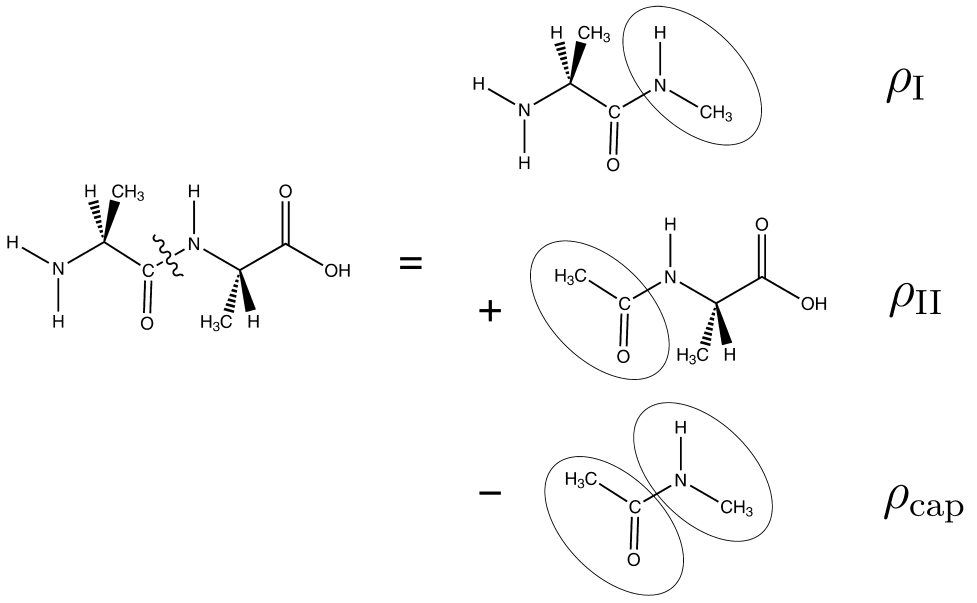
The biological function of proteins is governed by their interaction with other molecules. Protein−protein interactions are involved in, e.g., signal transduction, protein−RNA interactions are important for protein synthesis, and the action of drug molecules is mediated via their interaction with proteins.

Theoretical methods are an essential tool to study the structure of the complexes as well as strength and specificity of their interactions. For protein−protein interactions, determining binding affinity amounts to a huge task that has seen much progress in recent years (see ref 1 and references therein), whereas for protein−ligand complexes a variety of computational methods is routinely applied in the search for drug leads (see ref 2 and references therein). Here, much of the effort is focused on developing better scoring functions for the estimation of binding affinity approximates. Many available scoring functions rely heavily on semiempirical parametrizations, which gives rise to a number of shortcomings. To name just one example, polarization effects are difficult to treat accurately in the commonly applied force-field based methods, since the use of polarizable force fields is still nonstandard.[3,4]

Quantum-chemical methods provide a rigorous and parameter-free approach to the description of the binding energy. Quantum-mechanical (QM) calculations yield a wealth of data since they are not specifically designed to describe only the energy or a few properties but to determine molecular orbitals and electron densities. These contain all information on the system and could be used to formulate more detailed QM-based scoring functions. A first step in this direction is the use of the full electron density instead of point charges for the evaluation of the electrostatic interaction energy. While providing high accuracy and flexibility, ab initio quantum-chemical methods have the drawback of a restriction of the system size that can be dealt with, which is due to their steep scaling of computational effort with system size. Although developers strive to extend the applicability of ab initio quantum-chemical methods to larger molecules by taking advantage of linear-scaling approaches,[5,6] they are still not routinely applied to molecules of the size of an average protein.[7,8] For this reason, up until recently, mostly semiempirical QM methods[9,10] or quantum-mechanics/molecular-mechanics (QM/MM) approaches[11] have been applied to evaluate protein−ligand bond strengths. Multilevel methods such as QM/MM approaches[12] focus on parts of the molecules by employing a QM-type method for a local active site and describing the environment by a more approximate MM method, where link atoms[13] or boundary atoms (see refs 14) might be used to describe the boundary. However, through the coupling of two very different methods, problems arise at the boundary of the two regions.

A variety of fragment-based approaches have been proposed to overcome the size problem of the brute-force supermolecular

Received: October 10, 2012
Published: March 19, 2013



$$\rho_{\mathrm{tot}}(\mathbf{r}) = \rho_{\mathrm{I}}(\mathbf{r}) + \rho_{\mathrm{II}}(\mathbf{r}) - \rho_{\mathrm{cap}}(\mathbf{r})$$

Figure 1. Schematic illustration of the partitioning of peptide bonds in the MFCC and 3-FDE scheme, shown for dialanine. The total electron density $\rho_{\mathrm{tot}}(\mathbf{r})$ is obtained by summing up the fragment densities $\rho_{\mathrm{I}}(\mathbf{r})$ and $\rho_{\mathrm{II}}(\mathbf{r})$ and subtracting the cap density $\rho_{\mathrm{cap}}(\mathbf{r})$.

approach. They all rely on utilizing chemical knowledge to divide the system into smaller subsystems, e.g., cutting up a large molecule into fragments. For very large molecules, such schemes are more efficient since the computational effort scales with the number of subsystems and at most linear with the system size. These subsystem methods differ in the way they treat the fragment boundaries and the interactions between the subsystems. For extensive current reviews, cf. refs 15−17.

Subsystem methods are not only a practical way to overcome size restrictions, they also offer a conceptually attractive way of looking at a system in terms of its chemical constituents. In typical applications only part of a biomolecule is of interest, for instance, the interface in protein−protein interactions or a binding pocket in protein−ligand docking. A fragment-based approach allows one to employ an accurate but computationally demanding setup for only the important parts of the system. If an accurate treatment of the whole system is desired, one may still benefit from a fragment-based approach as it facilitates interpretation of the results, while in principle being able to reproduce the exact result.

An approach suitable for the treatment of large molecules, especially proteins, has been proposed by Zhang and Zhang: the molecular fractionation with conjugate cap (MFCC) scheme.[18−20] In this scheme, proteins are partitioned into their constituting amino acids, and capping groups are added to saturate the dangling bonds, two of which put together build a cap molecule, cf. Figure 1. The density of this cap molecule is subtracted when adding fragment results in order to obtain the density and other properties of the complete system. Thus, the total electron density of the molecule is obtained as the sum over n_{sub} subsystem electron densities minus the sum of the n_{cap} cap electron densities.

$$\rho_{\mathrm{tot}}(\mathbf{r}) = \sum_{i=1}^{n_{\mathrm{sub}}} \rho_i(\mathbf{r}) - \sum_{j=1}^{n_{\mathrm{cap}}} \rho_j^{\mathrm{cap}}(\mathbf{r}) \qquad (1)$$

In improved MFCC schemes, incorporating effects of the environment, such as hydrogen-bonds or electrostatic embedding, or including polarization by neighboring amino acids has been addressed,[21−24] and MFCC schemes have subsequently been used to calculate protein−ligand interaction energies; see refs 24 and 25 and references therein.

The 3-FDE method, an extension of the subsystem density functional theory (DFT) approach frozen-density embedding (FDE),[26] has been developed by two of us.[27] Its name reflects the way covalent bonds between fragments are dealt with: by introducing capping groups each severed bond is represented by three fragments, so that covalent bonds are treated in a similar way as in the MFCC scheme. Within the 3-FDE scheme, the influence of all other subsystems is included in the calculation of the individual fragments via an effective embedding potential, which is in principle exact and depends on the density and the nuclei of the surrounding fragments. The density of the 3-FDE caps is fixed to the density within the molecules built by the two capping groups and is later subtracted. This constraint for the cap densities ensures that the total density is positive, but gives rise to minor errors in the density of the cap region. The mutual polarization of the fragments within FDE is accounted for by so-called freeze-and-thaw (FT) cycles in which the frozen densities of the subsystems are updated iteratively.[28] For hydrogen-bond interactions, FDE reproduces the densities of supermolecular calculations well[29] and has been successfully applied to describe solvent effects on molecular properties[30−34] and biomolecular systems.[35,36]

Capping groups were introduced in 3-FDE to circumvent problems in approximating the kinetic-energy contribution to the embedding potential occurring for systems connected by covalent bonds.[37] Alternative methods exist which address this problem more directly using optimized effective potential techniques, see e.g., refs 38−41. Other QM-based fragmentation methods employ alternative approaches for the description

of the fragment boundaries and encounter different problems. Some of these methods rely on reproducing the molecular orbitals at the cut bonds, e.g., the fragment molecular orbital method (FMO).[42−44]

We employ the 3-FDE method since it combines the accuracy of a QM method with the intrinsic favorable linear-scaling of a fragment-based method. Compared to QM/MM methods, 3-FDE provides a more detailed description of the environment with built-in polarization and automatic system-specific capping but still allows focusing in on specific fragments. Previously, it has been shown with a pilot implementation that 3-FDE can be applied as a subsystem method for subsystems connected by covalent bonds.[27] Here, we want to extend 3-FDE toward a method that can be applied to study protein−ligand interaction quantum chemically. This requires the ability to focus on selected parts of a protein by, first, employing a more accurate description for some fragments and, second, defining fragments in the most suitable fashion. For this purpose the partitioning scheme has to be more general and flexible. Even more importantly, in order to be applicable to a wider class of proteins, a treatment of disulfide bridges and of charged residues has to be made possible.

This work is organized as follows: In section 2, the computational methodology is briefly outlined. In the following we present the necessary extensions of the 3-FDE scheme, beginning with the partitioning of the protein in section 3, demonstrated for ubiquitin, followed by the introduction of caps for disulfide bonds shown for cystine in section 4 and the treatment of charged side groups in section 5, with small charged dipeptides as test cases. We then go on to show the application of these extensions to the bubble protein in section 6, and discuss possible application to generate electrostatic potentials for docking purposes, before presenting our conclusions in section 8.

2. COMPUTATIONAL METHODOLOGY

The theory of the 3-partition FDE scheme and its implementation in the Amsterdam density functional package[45,46] are given in detail in ref 27. Here, we restrict ourselves to a brief sketch of the salient points. The partitioning we use is shown in Figure 1. The capping groups employed here preserve the electronic structure of the severed peptide bond as much as possible while still being small at the same time. The first step of the 3-FDE calculations consists of calculations on the isolated capped fragments and cap molecules. The total density obtained from summing up their densities will be denoted as MFCC as this is equivalent to the procedure proposed by Zhang and Zhang.[18] These fragment densities are then used to construct the embedding potential for the first 3-FDE calculations, in which each fragment is optimized in the presence of the unrelaxed frozen densities. The density assembled from these calculations will be denoted as 3-FDE(0).

The most outstanding feature of the 3-partitioning scheme is the fact that the density added at the fragment boundaries in the capping groups by construction equals the density in the cap molecule that is subtracted when the density of the total molecule is obtained from the various fragment calculations. This constraint has to be introduced to ensure that the total density is positive everywhere in space.[27] In practice, within a fragment calculation, in the region of the cap(s) a fixed target density has to be reproduced. In order to fulfill this condition, in the Kohn−Sham-like (KS) equations used to determine the orbitals ϕ_i of the active subsystem I and its density under the influence of a frozen density ρ_{II}, the embedding potential $v_{eff}^{FDE}[\rho_I,\rho_{II}](\mathbf{r})$ within the cap region is replaced by a cap potential which is constructed such that it yields the target density:

$$\left[-\frac{\nabla^2}{2} + v_{eff}^{FDE}[\rho_I, \rho_{II}](\mathbf{r})\right]\phi_i^{(I)}(\mathbf{r}) = \varepsilon_i\phi_i^{(I)}(\mathbf{r}) \tag{2}$$

$i = 1, ..., N_I/2$, with N_I the number of electrons in subsystem I. The effective potential is given by

$$v_{eff}^{FDE}[\rho_I, \rho_{II}](\mathbf{r}) = \begin{cases} v_{eff}^{KS}[\rho_I](\mathbf{r}) + v_{eff}^{emb}[\rho_I, \rho_{II}](\mathbf{r}) & \text{for } \mathbf{r} \notin V_I^{cap} \\ v_{cap}(\mathbf{r}) & \text{for } \mathbf{r} \in V_I^{cap} \end{cases} \tag{3}$$

which outside of the cap region V_I^{cap} consists of the KS potential of subsystem I, $v_{eff}^{KS}[\rho_I]$, and of the effective embedding potential

$$\begin{aligned} v_{eff}^{emb}[\rho_I, \rho_{II}](\mathbf{r}) = {} & v_{II}^{nuc}(\mathbf{r}) - v_{cap}^{nuc}(\mathbf{r}) \\ & + \int \frac{\rho_{II}(\mathbf{r}') - \rho_{cap}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, d\mathbf{r}' \\ & + \left.\frac{\delta E_{xc}[\rho]}{\delta\rho}\right|_{\rho=\rho_I+\rho_{II}-\rho_{cap}} - \left.\frac{\delta E_{xc}[\rho]}{\delta\rho}\right|_{\rho=\rho_I} \\ & + \frac{\delta T_s^{nadd}[\rho_I, \rho_{II}, \rho_{cap}]}{\delta\rho_I} \end{aligned} \tag{4}$$

where v_{II}^{nuc} and v_{cap}^{nuc} are the electrostatic potentials of the nuclei in subsystem II and in the cap group, E_{xc} is the exchange−correlation energy, and T_s^{nadd} is the nonadditive kinetic energy. Within the cap region, the effective potential is given by the cap potential V_{cap}. For details and derivations, see ref 27. The 3-FDE scheme is geared to keep the errors introduced on partitioning the protein small, but there are still two sources of error connected to it. First, cutting the peptide bond and introducing caps fixes the number of electrons per fragment and does not allow for polarization of the peptide bond. Second, flaws in the approximate kinetic-energy functional yield errors in the treatment of hydrogen bonds and other interactions between fragments.

The cap region is defined by the grid points which are closer than 3 bohr from an atom in the cap region and not closer to a noncap atom. The cap potential v_{cap} which yields the target cap density is obtained iteratively. Starting from the potential of the isolated cap molecule, the cap potential is updated in a self-consistent field (SCF) cycle once convergence as measured by the norm of the commutator of the Fock matrix and the density matrix falls below 10^{-3} a.u. Depending on the error in the number of electrons in the cap region, either a constant shift of $0.2\Delta N_{cap}$ is applied to all grid points to adjust the number of electrons in the cap, or the cap potential is updated at each grid point using the scheme of van Leeuwen and Baerends.[47] There, the updated cap potential is calculated as

$$v_{cap,el}^{new}(\mathbf{r}) = f v_{cap,el}^{old}(\mathbf{r}) \qquad \text{with } f = \frac{\rho_I^{old}(r)}{\rho_{cap}(r)}$$

where $v_{cap,el}$ denotes the total electronic potential in the cap region, (i.e., the nuclear potential is constant and not part of the updates). In order to prevent too large changes in the update

procedure, several measures are taken. A damping factor is employed, i.e., $f = 1 + \mathrm{LBdamp}[(\rho_I^{old}(r)/\rho_{cap}(r)) - 1]$, and f is kept within a range of $1 \pm \mathrm{LBmaxstep}$. The default values for LBdamp and LBmaxstep are 0.25 and 0.05, respectively. Updates are furthermore restricted to grid points where $\rho_{cap}(\mathbf{r}) > 0.01$ e·bohr^{-3}. The cap potential is regarded as converged as soon as the error in the cap density, defined as

$$\Delta_{cap} = \frac{1}{N_{cap}} \sqrt{\int_{V_{cap}} (\rho_I(\mathbf{r}) - \rho_{cap}(\mathbf{r}))^2 \, d\mathbf{r}}$$

drops below a threshold of 10^{-4}.

A few slight modifications have been made to the original scheme of ref 27. In the original implementation, in the case of several caps on a fragment, all cap potentials were updated until all of them met the convergence criteria. Here, we found that this lead to oscillations in some cases. Therefore every cap potential that has converged, is now kept fixed. In practice, it turns out that when the SCF convergence criteria are met, all cap potentials are still converged. An additional difficulty we encountered is that, for some fragments, the norm of the commutator of the Fock matrix and the density matrix never meets the update criterion when starting from the potential of the isolated cap molecule, so that the cap potential update scheme does not start. This behavior cannot be predicted beforehand, so that if it occurs, we now force potential update steps in the 30th, 45th, etc., SCF cycle until the normal update scheme starts. So far, this scheme has proven to work robustly. In case the default parameters do not lead to cap convergence, a particular fragment calculation can be redone with, e.g., a smaller LBdamp value. For the systems considered here, only a small number of fragment calculations requires such manual attention.

For all calculations, the exchange−correlation functional BP86[48,49] and the kinetic-energy functional PW91k[50] were employed. For the 3-FDE and the supermolecular DFT calculations of the proteins, the DZP basis set from the ADF basis set was used. In principle, a larger basis set could be employed for the 3-FDE calculations, but this is not feasible for the supermolecular reference calculations for the proteins. Where possible, i.e., for the small test calculations on dipeptides, we used a TZ2P basis set. In the FT cycles (results denoted as 3-FDE(n)), densities were updated after a full FT cycle had been performed and not immediately. For the frozen fragments the exact densities, i.e., not the density represented by using fit functions, were employed in the calculation of the nonadditive exchange−correlation and kinetic-energy potential. As standard damping factor for the potential updates, LBdamp = 0.25 was kept. For larger fragments, it turned out that a smaller damping factor was needed, i.e., for partitioning into fragments of the size of two amino acids, LBdamp = 0.15 was used, and for three amino acids and larger fragments, LBdamp = 0.1. Fragment calculations that did not converge were also repeated with LBdamp = 0.1. For the bubble protein with charged residues, calculations with the continuum solvation method COSMO have been run for both the supermolecular and the 3-FDE case with the default parameters for water in ADF. In the case of 3-FDE, the coordinates of the supermolecule were used to construct the COSMO surface to ensure a consistent treatment of solvation in the same cavity.

All aspects relating to the partitioning, as well as running the 3-FDE calculations with a number of FT cycles are taken care of with the scripting framework PyADF,[51] which uses OpenBabel[52,53] routines for the manipulation of the protein structures. The extensions to the existing scripting framework to allow for different partitionings, partial updates in the FT cycles, and disulfide bonds are described in the following sections.

For the analysis of the MFCC and 3-FDE densities, the same quantities as described in refs 27 and 54 were used, i.e., the integrated absolute error in the total density,

$$\Delta^{abs} = \frac{1}{N_{tot}} \int |\rho_{KS}(\mathbf{r}) - \rho_{FDE}(\mathbf{r})| \, d\mathbf{r} \tag{5}$$

the integrated root-mean-square error in the electron density

$$\Delta^{rms} = \frac{1}{N_{tot}} \sqrt{\int (\rho_{KS}(\mathbf{r}) - \rho_{FDE}(\mathbf{r}))^2 \, d\mathbf{r}} \tag{6}$$

and the magnitude of the error in the dipole moment

$$|\Delta\mu| = |\mu_{KS} - \mu_{FDE}| = \left| \int (\rho_{KS}(\mathbf{r}) - \rho_{FDE}(\mathbf{r}))\mathbf{r} \, d\mathbf{r} \right| \tag{7}$$

Additionally, we calculated the angle between the dipole moment vectors ϕ_μ

$$\phi_\mu = \arccos\left(\frac{\mu_{KS}\mu_{FDE}}{|\mu_{KS}||\mu_{FDE}|} \right) \tag{8}$$

The first two measures can also be applied locally if the integration is restricted to a certain volume. For the numerical integration, the same integration grid as used by ADF in the supermolecular calculation was employed. For neutral molecules, $|\Delta\mu|$ can also be given as a percentage of the absolute value of the dipole moment. For charged molecules, the dipole moment is origin-dependent, and hence, only the magnitude of the error in the absolute value is meaningful.

3. STRATEGIES FOR PARTITIONING PROTEINS IN 3-FDE

First, we explore the most suitable strategy for partitioning proteins in 3-FDE. As a test case, we have chosen ubiquitin, which is a regulatory protein playing a crucial role in the degradation of proteins and thereby in numerous biological processes.[55] Its structure is well-studied.[56] Ubiquitin contains only 76 amino acids (1231 atoms, of which 602 non-hydrogen), which makes it still possible to perform full calculations for comparison and contains different structural elements: an α-helix, a β-sheet, a short 3_{10}-helix, and several turns and loops. Its very compact and tightly hydrogen-bonded structure[56] involves overlapping densities and directional polarization effects between neighboring residues. These constitute challenging conditions for fragment-based calculations, and ubiquitin is, hence, a good model case. Here, we consider a hypothetical neutral form, where all side groups are kept neutral. Previously it has been demonstrated that for ubiquitin the 3-FDE approach can be employed to calculate the electron density for a whole protein using a partitioning into individual amino acids. For practical purposes usually a subregion of the protein is particularly relevant for the specific problem at hand. For the concrete example of ubiquitin, this could, e.g., be the face of the protein including residues Leu8, Ile44, and Val70 which is described to take part in protein−protein interactions.[57] For a study of protein−protein interactions, an accurate description of this interacting face would be most important.

In this section, we first discuss the reproduction of the density of the full protein, introducing different partitionings of the protein into subsystems. Subsequently, we focus our attention on reproducing only the density of a small region. On the technical side, this corresponds to studying the influence of the size or composition of the fragment and the number of density updates.

With the original implementation of the 3-FDE method in ADF and PyADF, proteins were partitioned such that each amino acid constituted one fragment, and the same number of FT cycles was used for all fragments. Rather than cutting the whole protein in individual amino acids, it is also possible to combine a number of amino acids in one fragment. This can initially speed up a 3-FDE calculation. Which size is most efficient depends on the computational overhead per cap. In addition, a small number of caps yields in principle more accurate results, but apart from the drawback of the increased computational requirements for larger fragments, cap convergence is also hampered in larger fragments.

While such a partitioning into larger fragments merely in terms of the size can be automatized, it is also possible to use chemical insight in the system. For instance, one main source of error in the 3-FDE densities is due to the description of hydrogen bonds between different fragments, so that in some cases it might be desirable to keep these bonds in the same fragment. This means grouping amino acids belonging to a helix or another secondary structure element together. Some protein structures might require larger fragments, e.g., if they contain bonds that cannot be easily capped, such as a metal atom coordinated by several Cys residues. An additional consideration in the choice of the fragment size is the treatment of charged amino acid residues in the fragmentation scheme.

Options were added to the PyADF routines which partition the protein. First, automatic partitioning into fragments containing m amino acids, where m can now be an integer larger than 1 has been added. Second, the user can define all or a part of the fragments by specifying lists of residue numbers.

Apart from the fragment size, it is the number of updates per fragment, i.e. the number of fragment calculations, which determines the computational cost. Limiting these updates to a small region of interest, e.g., a binding pocket, while using only one FT cycle or unmodified fragment densities for the rest of the system saves on computational effort, while giving a smooth transition between different levels of approximation.

On the scripting level, a new routine for this partial update option has been added to allow to specify the number of updates for every fragment, which corresponds to mixing of 3-FDE(n) results for different n.

We have partitioned the protein ubiquitin in five different ways (see Figure 2 for a representation of the partitioning): three simple partitions according to the fragment size (1 (A), 2 (B), or 3 (C) amino acids), chemically motivated partitioning into five large fragments (D), consisting of residues 1−20, 21−38, 39−52, 53−63, and 64−76, where the capped residues were chosen such that they are far away from the face including Leu8, Ile44, and Val70, and one partitioning (E) focusing on the "interacting face"; the two neighboring residues on each side of Leu8, Ile44, and Val70 were included in the same fragments, while the remaining part of the protein was partitioned into fragments each containing one amino acid.

The partitioning according to fragment size (A−D) serves to illustrate the influence of the partitioning on the quality of the density. Fewer caps reduce the error introduced in the density

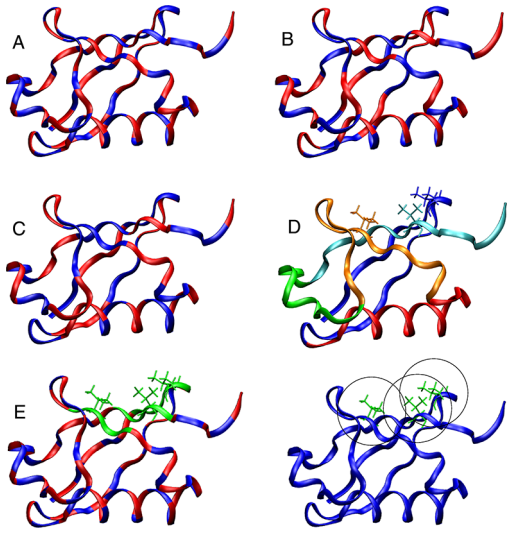


Figure 2. Partitioning applied to ubiquitin. Fragments are indicated by different colors. For partitioning D and E, Leu8, Ile44, and Val70 are represented by sticks. In the lowermost right panel, the three overlapping spheres of 5 Å radius around the geometric centers of these three amino acids are depicted. Graphics created with VMD.[58]

since a significant part of it is due to the caps. Comparing partitioning A−C to D gives us an indication how large the contribution of the error introduced in the cap region is compared to errors caused by the kinetic-energy functional for the description of hydrogen bond and other interactions between the fragments. Since partitioning D contains only four caps, the error must in this case mainly be attributed to noncap errors.

3.1. Electron Density of Ubiquitin. Error measures for the densities and dipole moment of ubiquitin obtained with partitionings A, B, C, and D are given in Table 1. The partitioning into fragments consisting of only one amino acid, A, is the same as that used in ref 27.

For 3-FDE calculations, there are two main possibilities for performing the density updates in the FT cycles. Updated fragment densities can either be used immediately as frozen density in the calculation of the next fragment, or the same set of frozen densities can be retained until all densities have been updated after a full cycle. The first, sequential option, will be called "sequential FT", while updating all fragment densities at once is called "parallel FT". The conceptional advantage of the latter scheme is the independence of the results on the order in which fragments are updated. This independence of order is also achieved in embedding schemes that simultaneously determine new fragment densities, such as in ref 59. Such a parallel FT scheme also has practical advantages. First, a fragment calculation in a FDE cycle that did not converge can be rerun with different parameters without effect on all other fragments' results in this cycle. Second, in a parallelized scheme, calculations on different fragments can be distributed to different machines easily.

For the MFCC and 3-FDE(0) results obtained for partitioning A, we observe errors that are in the same order of magnitude as in the earlier results but which are not identical. These small differences can be attributed to slight changes in the implementation in the ADF program suite and in openbabel and the use of the fitted density for the calculation of the exchange−correlation potential instead of the nonfitted density used in the earlier results. With the parallel FT density

dx.doi.org/10.1021/ct3008759 | J. Chem. Theory Comput. 2013, 9, 2425−2440

Table 1. Integrated Absolute Error Δ^{abs} in the Electron Density (e·bohr^{-3}), Root Mean Square Error Δ^{rms} in the Electron Density (e·bohr^{-3}), and Error in the Magnitude and Direction of the Dipole Moment, $\Delta\mu$ (D) and ϕ_μ, in the MFCC and 3-FDE(n) Calculations of Ubiquitin Using Different Fragment Sizes[a]

		$\Delta^{abs} \times 10^3$	$\Delta^{rms} \times 10^3$	$\lvert\Delta\mu\rvert$ (D)	ϕ_μ (deg)
A	MFCC	4.44	0.039	17.51	13.6
	3-FDE(0)	3.69	0.028	4.52	3.4
A, parallel FT	3-FDE(1)	2.42	0.023	5.03	4.0
	3-FDE(2)	3.91	0.028	8.50	3.9
	3-FDE(3)	2.71	0.025	4.26	3.5
A, sequential FT	3-FDE(1)	2.38	0.021	10.30	8.4
	3-FDE(2)	2.35	0.021	6.17	3.4
	3-FDE(3)	2.36	0.021	6.97	3.8
B, parallel FT	MFCC	4.20	0.038	17.01	12.2
	3-FDE(1)	1.86	0.018	2.65	2.2
C, parallel FT	MFCC	3.99	0.036	14.12	10.0
	3-FDE(1)	1.67	0.015	2.96	1.8
D, parallel FT	MFCC	2.52	0.025	2.15	1.7
	3-FDE(1)	0.85	0.009	0.91	0.2

[a]The dipole moment in the supermolecular reference calculation is μ = (16.61, −53.46, 42.35) D, $|\mu|$ = 70.19 D.

update scheme, in the 3-FDE(1) to 3-FDE(3) results, we observe an oscillation in the results with respect to the FT cycles. The density errors do not converge with increasing number of FT cycles, but Δ^{abs} and Δ^{rms} decrease from MFCC, 3-FDE(0), to 3-FDE(1), before they increase again in the 3-FDE(2) and 3-FDE(3) results. Repeating the calculations for partitioning A with the sequential FT cycles, for 3-FDE(1) to 3-FDE(3) errors in the density converge. Thus, the oscillation in the density errors with the simultaneous update of all densities in parallel FT does not occur with the immediate update of the sequential FT. This is because in sequential FT there is a mixture of updated and old densities, so that fragment densities already react to polarization in one part of a system resulting in a continuous, smooth change. Compared to the oscillatory behavior, this smooth change is more desirable, however, as pointed out earlier, it is accompanied with a dependency of the results on the fragment update order and other practical disadvantages.

For the errors in the dipole moments, similar observations are made: with parallel FT, $|\Delta\mu|$ oscillates for calculations using different numbers of FT cycles, while for sequential FT, the convergence behavior is smoother. Note, however, that also for sequential FT, $|\Delta\mu|$ does not decrease monotonously, but still oscillates slightly.

Apart from the practical advantages, one can still justify the use of parallel FT updates based on the resulting error values: For FDE(1) with continuously updated densities, the Δ^{rms} error is already converged, the Δ^{abs} error is already practically converged, and only the error in the dipole moment is still large. The FDE(1) results with parallel update are very similar (Δ^{abs} error 2.42 × 10^{-3} vs 2.38 × 10^{-3} e·bohr^{-3} and Δ^{rms} error 0.023 × 10^{-3} vs 0.021 × 10^{-3} e·bohr^{-3}), but the error in the dipole moment is much smaller (5.03 vs 10.30 D). The cost of any FT cycle equals the cost of m fragment calculations, with m being the number of fragments, and determines the cost of the calculation. Since a reasonable description is already achieved in the first FDE calculation, i.e., 3-FDE(0), practical application will only employ multiple FT cycles if this is easily affordable in special cases. All results presented in the following will refer to the parallel FT scheme since that will be convenient for future applications.

Now we turn to the results exploring different partitionings. The results for fragments containing two- and three residue-partitionings (B and C, respectively) show smaller errors than the results obtained with partitioning A, and this holds also for C compared to B with exception of the error in the dipole moment obtained with FDE(1). For partitioning B and C, similar convergence patterns with the number of FT cycles are found as for partitioning A (cf. Table S2 of the Supporting Information). What might be surprising at first glance is the fact that the errors decrease only slightly when the number of caps is reduced by a half or a third. This is due to the fact that errors are not only introduced by cutting the protein at the peptide bonds and introducing capping groups, but also by the imperfect description of other interactions between the fragments, foremost hydrogen bonds, using an approximate kinetic-energy functional. The results of partitioning D underline this further: only 4 caps are present in this molecule, the α-helix is treated in one fragment while the β-sheet has been cut between connected strands twice. These cuts through the β-sheet cause a major contribution to the error density (cf. Figure 3). The importance of the secondary structure and the

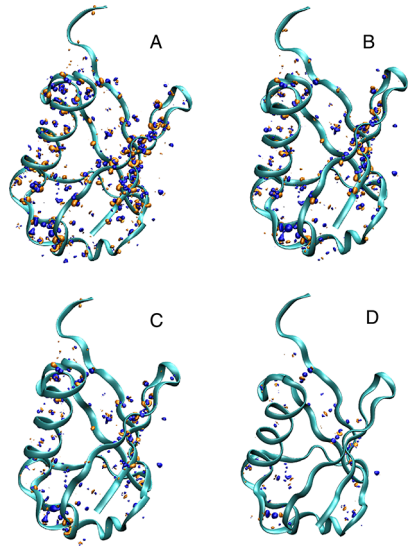


Figure 3. Isosurface plots (contour value of 0.002 e/bohr^{-3}) of the difference densities between the FDE(1) calculations and the conventional supermolecular DFT calculation for the ubiquitin molecule using partitionings A–D. Graphics created with VMD.[58]

treatment of hydrogen bonds is also reflected in the large error in the dipole moment in the MFCC calculations of partitionings A to C: in these partitionings, the α-helix of ubiquitin, which has a large contribution to the dipole moment, is cut into small fragments, which are not polarized in the MFCC calculation. For partitioning D, where the α-helix is treated within one fragment, the error in the dipole moment is considerably reduced. This is in agreement with earlier work on MFCC,[21] in which the importance of the hydrogen bonds within secondary structure elements has been observed as well.

3.2. Interacting Face of Ubiquitin. In the following, we focus on the description of the density of only a part of ubiquitin, which will be relevant for its interactions with other

dx.doi.org/10.1021/ct3008759 | J. Chem. Theory Comput. 2013, 9, 2425–2440

proteins. This small part was for this purpose defined as the volume of three overlapping spheres with a radius of 5 Å around the geometric centers of residues Leu8, Ile44, and Val70. "Focusing" on one region means trying to achieve a good description for that particular part of the protein, while choosing a fast approximate description for the rest of the protein. In the case of the 3-FDE approach this translates to starting from the sum of the isolated fragment densities and including polarization effects by adding one FDE cycle, 3-FDE(0), and possibly a second FT cycle, 3-FDE(1), for a *selected* number of fragments. This is comparable to the partial update scheme used for an FDE water environment description by Neugebauer et al. in ref 33.

At the same time, one can employ larger fragments in order to reduce errors introduced by caps. The procedure to achieve a locally improved description of the density is thus clear-cut: update the fragment densities in the region of interests and close to it, and/or combine fragments to larger fragments in that region. For ubiquitin, we tested both of these aspects. For partitioning A−F (here A denotes the partitioning, and F, a focus on a selected part), 3-FDE calculations were performed, for the 11 amino acids within the three overlapping spheres and for 4 neighboring amino acids (cf. Table S1 of the Supporting Information (SI)), the resulting protein density is a combined MFCC+3-FDE(0) density. For the 11 amino acids, further 3-FDE calculations were run, yielding a density consisting of MFCC densities for the residues that are not part of the interacting face, FDE(0) densities for the four neighboring fragments, and FDE(1) densities for the 11 selected fragments. These results will be denoted as MFCC+3-FDE(1). For MFCC +3-FDE(0), 15 FDE fragment calculations were performed, for MFCC+3-FDE(1), another 11 FDE fragment calculations had to be run, which contrasts to 76 FDE fragment calculations for each 3-FDE(0) and 3-FDE(1) results for the full protein.

In Table 2, the integrated Δ^{abs} and Δ^{rms} errors of the electron density, where the integration was restricted to grid points within the selected volume of three overlapping spheres, are shown. The results for partitioning A−F and MFCC+3-FDE(0) are very close to the full 3-FDE(0) calculation. For MFCC+3-FDE(0), the results are identical to 3-FDE(0) if all fragments contributing to the density in this region are updated. As can be seen from small differences, some more distant fragments do contribute to some extent. The MFCC+3-FDE(1) density will be different, since in this case the fragment densities of the environment are partly of MFCC and partly of FDE(0)-type. This is reflected in the Δ^{abs} and Δ^{rms} errors, but these are still sufficiently close to the full 3-FDE(1) results.

Next, we assess the impact of choosing larger fragments at the interacting face with partitioning E. Here, residues Leu8, Ile44, and Val70 are grouped together with their respective two neighboring residues on each side, yielding five-residue fragments. The remainder of the protein was partitioned into one-residue fragments (see the SI for details). 3-FDE(0) calculations were performed first (E−F1) for the three five-residue fragments and the direct neighbor residues on each side, resulting in a total of nine FDE(0) fragment calculations. In this case, FDE(1) calculations were performed only for the three five-residue fragments. Second, FDE(0) calculations were run for two neighbor residues on each side (E−F2, total of 15 FDE(0) calculations), and the FDE(1) calculations included also the direct neighbors of the three-residue fragments (nine calculations). For both cases, there are also differences in the MFCC+3-FDE(0) compared to the 3-FDE(0) densities, which

Table 2. Integrated Absolute Error Δ^{abs} in the Electron Density ($e \cdot bohr^{-3}$), Root Mean Square Error Δ^{rms} in the Electron Density ($e \cdot bohr^{-3}$) within a Radius of 5 Å around the Geometric Center of the Residues Leu8, Ile44, and Val70, in the MFCC and 3-FDE(n), and Combined MFCC+3-FDE Calculations of Ubiquitin Using Different Fragment Sizes[a]

	no. frag calc		$\Delta^{abs} \times 10^3$	$\Delta^{rms} \times 10^3$
A	76	MFCC	4.67	0.141
	76	3-FDE(0)	4.12	0.109
	76	3-FDE(1)	2.82	0.091
A−F	15	MFCC+3-FDE(0)	4.26	0.114
	11	MFCC+3-FDE(1)	3.44	0.102
E	64	MFCC	4.16	0.137
	64	3-FDE(0)	2.58	0.070
	64	3-FDE(1)	1.49	0.051
E−F1	9	MFCC+3-FDE(0)	2.77	0.080
	3	MFCC+3-FDE(1)	2.52	0.074
E−F2	15	MFCC+3-FDE(0)	2.71	0.077
	9	MFCC+3-FDE(1)	2.32	0.069
D	5	MFCC	3.04	0.113
	5	3-FDE(0)	1.10	0.045
	5	3-FDE(1)	0.95	0.040

[a]For comparison, the number of fragment calculations (no. frag calc) is given.

is again due to the fact that not all fragments in the volume we integrate over are included in the 3-FDE(0) cycle. This explains the improvement observed when including one neighbor. In partitioning E compared to A, there are two amino acids missing in the 3-FDE(0) calculation since we only include the direct covalently bound neighbors of the five-residue fragments around Leu8, Ile44, and Val70. Not taking the spatial arrangement into account causes larger differences between 3-FDE(0) and MFCC+3-FDE(0), but the difference density contributions in the outer parts of the volume will be less important. In the MFCC+3-FDE(1) calculations, the inclusion of one more neighboring fragment in the updates going from E−F1 to E−F2 yields again a notable improvement in the density compared to the fully updated 3-FDE(1). The density errors for partitioning scheme E are in general smaller than the density errors for partitioning scheme A. The results for partitioning D can be regarded as FDE reference calculations for this integrating volume, as there are no caps within or close to the volume. All density differences result thus from (hydrogen-bond) interactions between the different strands. Partitioning E comes already relatively close to these results. For all different partitionings, improvement of the density can be achieved by adding only a few 3-FDE(1) calculations on top. However, taking into account the relatively small improvement, for practical purposes, MFCC+3-FDE(0) might often already be good enough.

4. TREATMENT OF DISULFIDE BONDS IN 3-FDE

Disulfide bonds are important for folding and stability of many proteins. Treatment of these covalent bonds between residues is therefore a necessary feature of a generally applicable 3-FDE method. In the 3-FDE routines within the scripting framework PyADF, disulfide caps were introduced analogously to the MFCC scheme described in ref 60 (see Figure 4), i.e., the dangling bonds were capped with a SCH_3 group on both sides.

dx.doi.org/10.1021/ct3008759 | J. Chem. Theory Comput. 2013, 9, 2425−2440

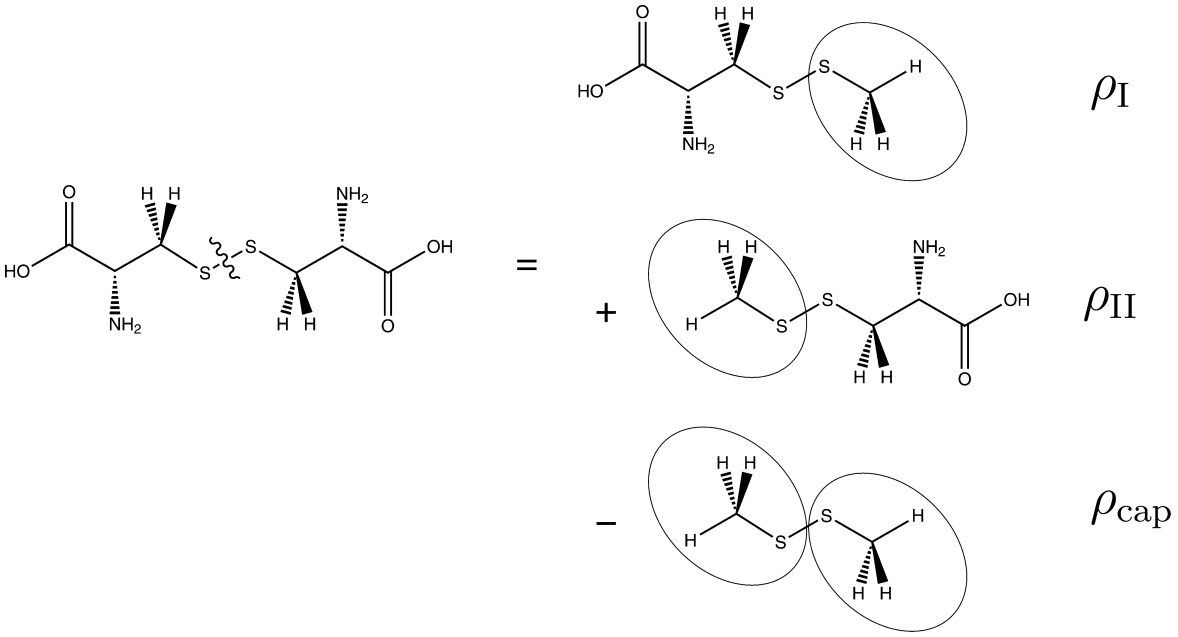


Figure 4. Schematic illustration of the partitioning of disulfide bonds in the MFCC and 3-FDE scheme, shown for cystine.

In order to test this, we consider the dipeptide cystine consisting of two cysteines connected by a disulfide bond. The structure of cystine was optimized employing the exchange–correlation functional BP86 and the TZ2P basis set. This basis set was also used for the supermolecular reference, MFCC and 3-FDE calculations.

For cystine, the Δ^{abs} error, the Δ^{rms} error, and $|\Delta\mu|$ for the MFCC and the 3-FDE(n) calculations are in the same order of magnitude or slightly smaller (see Table 3 and Table S3 of the Supporting Information) than for the description of peptide bonds in ref 27. As was the case for dialanine, for cystine the MFCC overall density errors are slightly smaller than for 3-FDE(n). Likewise, the small difference in the densities in the case of MFCC originates from missing polarization effects on the carboxylic and amino group (see Figure 5a).

Table 3. Integrated Absolute Error Δ^{abs} in the Electron Density (e·bohr^{-3}), Root Mean Square Error Δ^{rms} in the Electron Density (e·bohr^{-3}), and Error in the Magnitude and Direction of the Dipole Moment, $\Delta\mu$ (D) and ϕ_μ, in the MFCC and 3-FDE(n) Calculations of the Cystine and of Three Dipeptides with Charged Residues

		$\Delta^{\mathrm{abs}} \times 10^3$	$\Delta^{\mathrm{rms}} \times 10^3$	$\lvert\Delta\mu\rvert$ (D)	ϕ_μ (deg)
cystine	MFCC	0.44	0.020	0.17	5.0
	3-FDE(0)	0.51	0.025	0.06	1.5
	3-FDE(1)	0.49	0.025	0.06	0.6
Ala–Arg$^+$	MFCC	1.45	0.058	0.81	0.0
	3-FDE(0)	1.09	0.059	0.54	0.0
	3-FDE(1)	0.99	0.058	0.40	0.0
Ala–Asp$^-$	MFCC	1.29	0.062	0.68	0.1
	3-FDE(0)	1.04	0.070	0.22	0.0
	3-FDE(1)	0.94	0.071	0.12	0.0
Asp$^-$–Arg$^+$	MFCC	4.46	0.183	6.79	4.5
	3-FDE(0)	2.62	0.118	4.12	2.3
	3-FDE(1)	2.41	0.118	3.87	1.7
	3-FDE(2)	2.44	0.119	3.78	1.6

The fact that MFCC already performs particularly well for this test case can be attributed to the symmetric nature of the disulfide bond between two Cys residues. The errors in the density are even smaller than for dialanine, which could be a consequence of the disulfide bond being less polar than a peptide bond. Even though the error measures for the density are small for MFCC, the deviation in the angle of the dipole moment is at 5.0° fairly large. The missing polarization effects are captured in the 3-FDE(n) calculation at the cost of small differences in the cap region, since the density close to the cap gets slightly distorted because of the constraint in the cap region (see Figure 5). The direction of the dipole moment is considerably improved by FDE(1). As for the capped peptide bonds, already one FT cycle is sufficient for the disulfide bonds, and further cycles do not decrease the error in the density significantly. These observations should to a large extent also be true for disulfide bridges in a protein, since the bond between two Cys residues will be nonpolar in most cases.

5. TREATMENT OF CHARGED AMINO ACID RESIDUES IN 3-FDE

A realistic modeling of a protein needs to include correct protonation states and, therefore, has to be able to treat charged amino acid residues. Within the 3-FDE scheme, an (integer) charge is assigned to the fragment containing the charged group. This leads to charges which are localized by construction. Previously, only calculations on positively charged dipeptides were performed. Here, we extend the treatment of charged residues to negative charges and to zwitterionic dipeptides.

While for both systems tested earlier,[27] H$^+$–His–Leu and H$^+$–Ala–Ala, density errors are by a factor of 3–4 larger than for the neutral dipeptides, the dipeptide which was protonated at the side group (His) instead of the N-terminus, exhibited smaller density errors. These were attributed to the polarization of the peptide bond, which is not described properly by the caps introduced in 3-FDE. Instead, in 3-FDE the number of

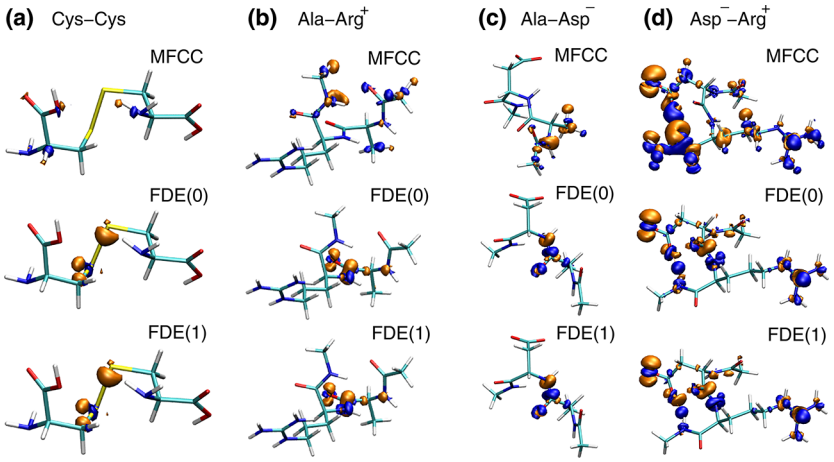


Figure 5. Isosurface plots of the difference densities between the MFCC and FDE(n) calculations and the conventional supermolecular DFT calculation for (a) cystine (contour value of 0.0005 e/bohr^{-3}) and the (b) Ala−Arg$^+$, (c) Ala−Asp$^-$, and (d) Asp$^-$−Arg$^+$ dipeptides with caps (contour value of 0.001 e/bohr^{-3}). Graphics created with VMD.[58]

electrons in each fragment is fixed and a charge-transfer between subsystems is not possible.

For the dipeptides here we chose an approach identical to the procedure followed in the calculations for proteins. We prepared the test molecules for charged residues by cutting out pairs of amino acids from protein structures on pdb files (Asp31−Arg32 from 1ubq, Asp14−Ala15 and Ala33−Arg34 from 1e3v) and using the atom coordinates of neighboring amino acids to generate NH−CH_3 and CO−CH_3 caps. Consequently, the dipeptides were (de)protonated at the side groups. The positions of the missing hydrogen atoms were determined using openbabel routines without further geometry relaxation for the resulting structures, as this is difficult in a gas phase model. The error measures for the electron densities of the three test systems are given in Table 3 and Table S4 of the Supporting Information, isosurface plots of the difference densities are depicted in Figure 5b−d.

For the dipeptides with charged side groups, larger polarization errors can be expected and are indeed found. For the capped Ala−Arg$^+$ dipeptide, the MFCC density deviates on the caps added to both the Ala and Arg residues and on the methyl group of the Ala residue, which is due to missing polarization by the other fragment. This is remedied in the 3-FDE(n) calculations, again at the expense of some errors introduced in the cap region. One can observe a negative difference density on the Arg$^+$ residue side of the cap, and a positive difference density on the Ala residue side of the cap, i.e., the 3-FDE(n) density is smaller than the reference density on the Arg$^+$ side. The Δ^{abs} error decreases from 1.45×10^{-3} to 1.09×10^{-3} and 0.99×10^{-3} e·bohr^{-3} going from MFCC to 3-FDE(0) and 3-FDE(1), respectively, which is larger than for neutral dipeptides, but still smaller than for the H$^+$−His−Ala and H$^+$−Ala−Ala test systems mentioned earlier. These larger errors are mainly due to the constraint of a fixed number of electrons in the fragmentation which prohibits charge transfer between the fragments. An indication of the amount of this polarization is given by the Voronoi partial charges[61] in the reference calculation, which show that the Ala and Arg residues should carry charges of +0.05 and +0.95, respectively.

A more significant improvement than for the densities can be achieved in the error of the dipole moment, which is a factor of 2 smaller for the 3-FDE(1) calculation compared to the MFCC calculation. The angular deviation is small for MFCC as well as the 3-FDE calculations. The Δ^{rms} error is 30−50% larger than for the neutral dipeptides and fairly constant for MFCC and 3-FDE(n) calculations.

In the MFCC density of the capped Ala−Asp$^-$ dipeptide, differences with respect to the reference density occur mainly on the noncharged capped Ala fragment. The polarization via the peptide bond in the supermolecule is reflected in Voronoi partial charges of −0.15 and −0.85 on the Ala and Asp residues, respectively. The shape of the 3-FDE(n) difference density isosurfaces bear similarity to the Ala−Arg$^-$ dipeptide in the cap region, but with reversed signs of the difference densities. Thus, Δ^{abs} and Δ^{rms} are of the same order of magnitude for the Ala−Asp$^-$ as for the Ala−Arg$^+$ dipeptide. Likewise, the most significant improvement is found for the error in the dipole moment, from 0.68 (MFCC) to 0.12 D (FDE(1)), with the MFCC+3-FDE dipole moments well aligned with the reference, while the Δ^{abs} error decreases by ca. 30% and the Δ^{rms} error is even larger for the 3-FDE(1) than for the MFCC calculation.

The capped Arg$^+$−Asp$^-$ dipeptide presents a more challenging case for the evaluation of the performance of the 3-FDE method, since the comparison to a supermolecular reference calculation is not clear-cut. The supermolecular reference calculation is a gas phase calculation for a zwitterionic structure. Usually, the neutral structure is stable in the gas phase, while the zwitterionic structure is more stable when solvation effects are included. This is reflected in the fact that a level shift of 0.03 hartree had to be applied in order to converge the gas phase calculation. Thus, the reference calculation also represents a somewhat artificial situation. Charge transfer of a larger extent can be expected because of the opposite charges on the residues. In the supermolecular calculation, the Voronoi partial charges show a transfer of 0.35 e with respect to fragments with a formal charge of ±1. Apart from the peptide bond, the Asp−Arg dipeptide features a hydrogen bond between the carboxylic and an amino group. The charge transfer at a hydrogen bond from a charged system can be as large as 0.2 e per hydrogen bond,[62] so that the differences in the density due to neglecting the polarization via the hydrogen bond and the peptide bond can be the same order of magnitude. The restriction of integer number of electrons in the fragment calculations therefore leads to considerable differences in the density with respect to the supermolecular

dx.doi.org/10.1021/ct3008759 | J. Chem. Theory Comput. 2013, 9, 2425−2440

calculation. This kind of restriction which constitutes a drawback for reproducing the supermolecular density with delocalized charges is on the other hand very well suited for other purposes, e.g., to study charge-transfer reactions[63] or electron transfer.[64−66] For example, with the constrained DFT method,[67,68] a fixed number of electrons in a certain volume can be achieved by constructing an additional potential analogously to the cap potential.

The consequences can also be observed in the MFCC difference density, where it can be seen from the negative sign (in Figure 5 blue) that the density in the supermolecule is larger at the guanidinium group and at large parts of the cap which forms a hydrogen bond with the carboxylate group of Asp. Moreover, in the MFCC difference density, there are regions where the polarization through the other fragment is not described. This is reflected in Δ^{rms} and Δ^{abs} errors about a factor of 3 larger than for the singly charged dipeptides and a large difference in the dipole moment of 6.73 D. Although the restriction to an integer number of electrons remains, the polarization through the other fragment in the FDE(n) calculations significantly decreases the Δ^{rms} and Δ^{abs} errors and improves on the dipole moment error (still 3.87 D for FDE(1), which is on the order of 10% of the total dipole moment). The angular deviation ϕ_μ of the dipole moment decreases from 4.5° to under below 2°. Also the isosurface plots of the difference densities show smaller regions for the same difference density value.

As a summary of this part, we conclude that peptide bonds dividing a charged from a neutral residue, result in larger density differences than found for neutral dipeptides due to the lack of polarization of the peptide bond and the constraint of integer number of electrons in the fragments. However, reasonable densities and dipole moments are obtained.

6. THE BUBBLE PROTEIN AS CHALLENGING TEST CASE FOR 3-FDE

Ubiquitin is a nearly ideal model system because of its very moderate size, its high resolution X-ray structure, and its structural features. However, it does not contain any disulfide bonds. In order to validate the 3-FDE(n) scheme also for proteins containing disulfide bonds, another small exemplary protein was chosen: the bubble protein (pdb 1uoy[69]). The bubble protein is described as a defensin[70] with an interesting charge distribution consisting of an acidic and a basic pole.[69] It contains four disulfide bonds and a salt bridge and has a small enough size (64 amino acids, 452 non-hydrogen atoms) to allow for supermolecular reference calculations. An X-ray structure is available with a resolution of 1.5 Å and shows some β-sheet structures, but no α-helix. The bubble protein can serve as an example for disulfide bonds as well as the treatment of charged side groups.

As a first step, the density of the bubble protein with neutral residues was calculated using a partitioning into fragments of the size of one amino acid. The calculated errors are given in Table 4, and isosurface plots of the difference densities for the MFCC and 3-FDE(n) calculations are depicted in Figure 6. The errors in the density are only slightly larger than for the analogous calculation on ubiquitin while the error in the dipole moment of the MFCC calculations is much smaller than for the ubiquitin. This can be attributed to the absence of an α-helix.

6.1. Describing Localized Charges with 3-FDE. In order to obtain densities suitable for the calculation of interactions between a protein and other molecules, it is essential to be able

Table 4. Integrated Absolute Error Δ^{abs} in the Electron Density (e·bohr^{-3}), Root Mean Square Error Δ^{rms} in the Electron Density (e·bohr^{-3}), and Error in the Magnitude and Direction of the Dipole Moment, $|\Delta\mu|$ (D) and ϕ_μ, in the MFCC and 3-FDE(n) Calculations of the Neutral Bubble Protein[a]

	$\Delta^{abs} \times 10^3$	$\Delta^{rms} \times 10^3$	$\lvert\Delta\mu\rvert$ (D)	ϕ_μ(°)
MFCC	4.54	0.045	7.30	5.0
3-FDE(0)	3.90	0.032	4.90	3.6
3-FDE(1)	2.74	0.028	4.54	5.3
3-FDE(2)	4.21	0.034	0.77	1.1
3-FDE(3)	3.04	0.030	5.47	6.4

[a]The dipole moment in the supermolecular reference calculation is μ = (−22.49, −11.03, 31.08 D), $|\mu|$ = 39.92 D.

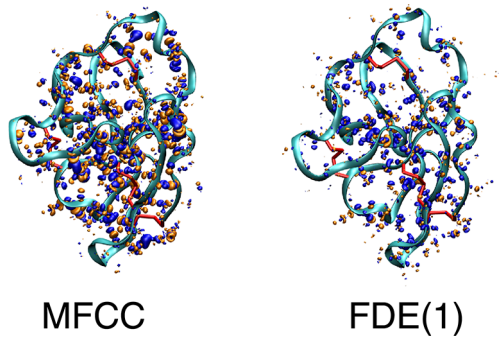


Figure 6. Isosurface plots (contour value of 0.002 e/bohr^{-3}) of the difference densities between the MFCC and FDE(1) calculations and the conventional supermolecular DFT calculation for the neutral bubble protein. Graphics created with VMD.[58]

to use the correct protonation state of the protein.[71] Handling both positively and negatively charged sites on one molecule is, however, problematic in supermolecular DFT calculations because this gives rise to a small gap between the highest occupied and lowest unoccupied molecular orbitals. This makes it difficult to converge the calculation to the desired physical charge state, requiring additional measures such as level shifts, restarts of the calculation from nonaufbau occupations, and adjustment of the direct inversion in the iterative subspace (DIIS) parameters. These problems were already observed for the zwitterionic capped dipeptide formed from Arg^+ and Asp^-. For such a small molecule one can analyze the convergence in detail and arrive at a proper procedure, but this quantum-chemical trial-and-error approach is not suitable for a full protein. Instead, one wants a robust protocol to yield a physically reasonable density with the correct protonation states.

Such a protocol is provided by the 3-FDE approach because the localization of the charge is a priori defined with an integer number of electrons assigned to each fragment, preventing the charge from delocalizing over different fragments and thereby avoiding convergence problems. A drawback is that the localization may be too strong, eliminating also physically reasonable charge transfer to neighboring fragments. In order to test for such effects we aim to assess the quality of a density by comparison with a supermolecular reference density. To this end, we need to determine a converged supermolecular density that corresponds to a minimum in the electronic structure degrees of freedom.[72] For a gas-phase model, this is difficult to achieve, because the application of convergence aids for

dx.doi.org/10.1021/ct3008759 | J. Chem. Theory Comput. 2013, 9, 2425−2440

enforcing the correct charge state may steer the calculation to a saddle point. An alternative is to employ a more physical model and allow for charge stabilization by the environment. If solvation by a polar solvent is taken into account, e.g., by a continuum solvation model such as the conductor-like screening model (COSMO),[73] the charges at the protein surface are stabilized by countercharges and supermolecular calculations can be converged to the correct charge state. Therefore, we compare the MFCC and 3-FDE calculations to the only available reference density, i.e., the one obtained from a supermolecular calculation including solvation described by the COSMO model. Note, however, that at this stage solvation is not included in the MFCC and 3-FDE calculations. The resulting difference densities are depicted in Figure 7, and the integrated difference density measures are given in Table 5.

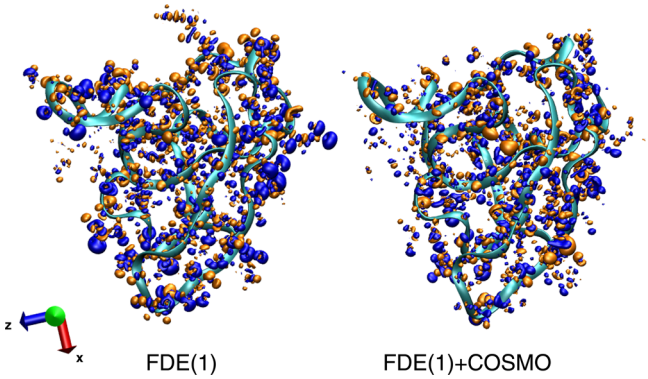


Figure 7. Isosurface plots (contour value of 0.002 e/bohr^{-3}) of the difference densities between the FDE(1) calculations and the conventional supermolecular DFT calculation (with COSMO) for the bubble protein with charged residues. Graphics created with VMD.[58]

Table 5. Integrated Absolute Error Δ^{abs} in the Electron Density (e·bohr^{-3}), Root Mean Square Error Δ^{rms} in the Electron Density (e·bohr^{-3}), and Error in the Magnitude and Direction of the Dipole Moment, $|\Delta\mu|$ (D) and ϕ_μ, in the MFCC and 3-FDE(n) Calculations of the Bubble Protein with Charged Residues

		$\Delta^{abs} \times 10^3$	$\Delta^{rms} \times 10^3$	$\lvert\Delta\mu\rvert$ (D)	ϕ_μ (deg)
	MFCC	7.05	0.063	22.67	2.2
w/o solvation	3-FDE(0)	7.58	0.060	80.16	1.8
	3-FDE(1)	6.39	0.054	64.48	2.6
	3-FDE(2)	7.22	0.057	65.03	1.6
	3-FDE(3)	6.70	0.056	71.28	2.7
COSMO	3-FDE(0)	6.70	0.052	72.66	2.1
	3-FDE(1)	5.69	0.047	57.95	2.7

While for the MFCC results, the Δ^{abs} and Δ^{rms} differences are only about one-third larger than for the neutral protein, for the 3-FDE(n) results, they are about a factor of 2 larger. More striking are the huge differences in the dipole moment, e.g., more than 80 D for the FDE(0) results. The comparison of the fragment calculations to the supermolecular reference obviously deserves closer attention. In the fragment calculations, each residue constitutes one fragment, where the Asp and Glu residues carry a negative and Arg and Lys residues a positive charge, summing up to a total charge of the bubble protein in our calculation of +1. When superposing isolated charged fragments in the MFCC results, the individual dipole moments add up to a relatively large dipole moment. If one allows these densities to relax in the FDE(0) calculation, the electron densities shift within the flexibility given by the basis functions of this fragment and charge separation decreases, resulting in lower dipole moments. In the supermolecular calculation without solvation, electrons can be delocalized over several fragments, so that an even lower dipole moment would be expected if the calculation could be converged. However, in the supermolecular calculation with solvation, charges at the surface of the protein are stabilized, resulting in a larger dipole moment than for MFCC. This large dipole moment with COSMO solvation is in line with observations by Pichierri[74] and explains the larger dipole moment differences for FDE(n) results relative to the MFCC results. Thus, these larger deviations in the FDE(n) calculations compared to MFCC are not a shortcoming of the 3-FDE scheme but appear because we are comparing a 3-FDE not including solvation to a supermolecular calculation that does include solvation.

Locally the density of the 3-FDE calculation resembles the supermolecular reference density, but these small differences add up to a large difference in dipole moment, particularly along the x-axis. This suggests that electron density is shifted along the basic funnel and acidic poles,[69] which extend in direction of the x-axis. In order to obtain a $\Delta\mu$ of 80 D along one axis, 0.5 electrons need to be displaced by about 33.4 Å (for comparison: the extension along the x-direction of this molecule is 27.5 Å). Ordering the residues along the x-axis and summing the difference densities of the FDE(0) results for both the first and second 32 residues yields about 0.5 electrons difference, confirming this shift of electrons in the x direction. Through the stabilization by the COSMO countercharges at the surface of the cavity, the density of negatively charged groups is drawn toward the surface of the molecule, with the largest difference close to the surface. Inside the protein the density is just shifted without causing large local changes. This is observable in the big difference density lobes located at the surface of the molecule displayed in Figure 7.

A more straightforward comparison would be to apply the continuum solvation model also in the 3-FDE(n) calculation. In the 3-FDE fragment calculation, this would mean that the fragment "feels" the embedding potential of all other fragments and additionally the field by the solvation method. For efficient applications of 3-FDE, this requires a smart implementation to prevent an "iterative overload" in the determination of the capping potential when both the solvation surface charges and the SCF density change during the same iterative process. As a simple first approximation to such a combination of 3-FDE with COSMO, we ran 3-FDE calculations where for the construction of the surface charges only the active density was taken into account within the cavity for the full protein, but the effect of the frozen densities was neglected. With this treatment, the difference density close to the surface becomes smaller and also the overall difference measures Δ^{abs} and Δ^{rms} as well as the dipole moment error are reduced. One should keep in mind, however that even with a consistent solvation treatment with the same surface charges, a charge shift through the molecule is by construction not possible with 3-FDE. Still, even for such a challenging example, the 3-FDE method reproduces the densities quite well locally. Besides integrating the COSMO treatment in 3-FDE, further improvement is possible by microsolvation, e.g., adding a number of explicit water

molecules. The FDE method has earlier been demonstrated to work well for this kind of solvation treatment.[31,33]

6.2. Focusing on the Funnel and Salt-Bridge in the Bubble Protein. Though the precise function of the bubble protein is yet unknown, within the structure of the bubble protein, a potential recognition site was described as including the residues Arg12, Phe27, and Trp43 lying in a funnel;[69] see Figure 8a. We define this site by spheres of 5 Å around the geometrical centers of these residues and performed MFCC+3-FDE(n) calculations as an example for a focused density calculation.

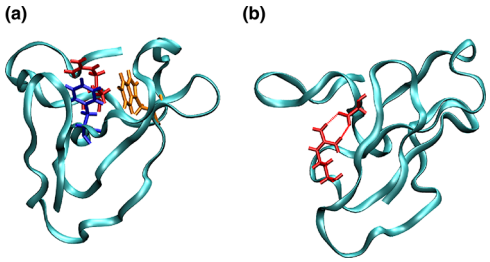


Figure 8. (a) Possible recognition site within the bubble protein molecule: Arg12, Phe27, and Trp43 and (b) salt-bridge between Asp10 and Arg32. Graphics created with VMD.[58]

For a one-amino-acid-per-fragment partitioning (A−F), the 15 residues within the overlapping spheres (cf. Table S5 of the Supporting Information) were updated in the FDE(0) and FDE(1) calculations, while the 6 neighboring residues to Arg12, Phe27, and Trp43, which do not lie within the spheres, are included only in the FDE(0) updates.

To avoid the above-mentioned convergence problems related to zwitterions or fragments with high charges, automatic partitioning is restricted to result in only one charged residue per fragment. The simplest choice for such a partitioning is to use single residue fragments, but because results should improve for larger fragments we decided to perform additional test calculations to explore whether other partitionings could also be worthwhile. We thus performed calculations for a partitioning, where Arg12, Phe27, and Trp43 were treated together in two 3- and one 2-residue fragment with their neighboring fragments (E−F, cf. Table S5 of the Supporting Information). In the last case, the 3-residue fragment including Lys42 lead to SCF convergence problems, which is why Lys42 was left out. The two next neighbors on each side of the three larger fragments were updated in the 3-FDE(0) calculation, while the larger fragments were updated in FDE(0) as well as FDE(1) calculations.

The 3-FDE calculations were compared to the supermolecular calculation including solvation, the resulting error measures are given in Table 6, and the difference density for the FDE(1) calculation with full updates is shown in Figure 8a. Compared to the density errors for the full protein, the Δ^{abs} errors are comparable while Δ^{rms} errors are larger for the funnel region. This was also the case for the ubiquitin face, although slightly less pronounced. For the funnel region, an improvement of the Δ^{abs} error over the MFCC results can only be achieved with the second FDE cycle, FDE(1). For partitioning A−F, with MFCC+3-FDE(0) practically the same results are achieved within the spheres with 21 instead of 64 fragment calculations (A). Adding another 15 fragment calculations for MFCC+3-FDE(1) further improves the results, does not reach the level of 3-FDE(1) with full updates (64 fragment calculations), and does not even perform better than MFCC. In order to improve with respect to the MFCC results, more neighboring fragments therefore have to be updated. This is probably due to the—compared to the neutral fragments—higher charges close by, which require a larger range in which fragments are updated.

Table 6. Integrated Absolute Error Δ^{abs} in the Electron Density (e·bohr^{-3}) and Root Mean Square Error Δ^{rms} in the Electron Density (e·bohr^{-3}) in the MFCC and 3-FDE(n) Calculations of the Bubble Protein with Charged Residues for Arg12, Phe27, Trp43 in the Funnel and the Salt-Bridge between Asp10 and Arg32[a]

	no. frag calc		$\Delta^{abs} \times 10^3$	$\Delta^{rms} \times 10^3$
funnel				
A	64	MFCC	6.01	0.141
	64	3-FDE(0)	7.94	0.160
	64	3-FDE(1)	5.46	0.115
A−F	21	MFCC+3-FDE(0)	7.93	0.161
	15	MFCC+3-FDE(1)	6.59	0.134
E	59	MFCC	6.00	0.144
	59	3-FDE(0)	7.39	0.148
	59	3-FDE(1)	5.06	0.103
E−F	15	MFCC+3-FDE(0)	7.37	0.152
	3	MFCC+3-FDE(1)	6.71	0.138
salt-bridge				
A	64	MFCC	7.08	0.194
	64	3-FDE(0)	5.73	0.136
	64	3-FDE(1)	4.85	0.121

[a]For comparison, the number of fragment calculations (no. frag calc) is also given. All MFCC and 3-FDE calculations are performed without solvation, while the supermolecular reference calculation includes COSMO solvation.

Employing larger fragments, as done in partitioning E−F indeed decreases the density errors. Performing the 15 fragment calculations for MFCC+FDE(0) reaches nearly the full 59 update level. With another additional three fragment calculations for MFCC+FDE(1), the errors decrease further but do not reach the agreement of the MFCC+3-FDE(1) calculation of partitioning A.

In Table 6, results for the integration over another region of the bubble protein, i.e., overlapping spheres around the salt bridge between Asp1 and Arg32 (see Figure 8b), are given. These two oppositely charged residues are connected via two hydrogen bonds and represent a potentially difficult case for FDE. The Δ^{abs} and Δ^{rms} values indicate that such a situation does not lead to more difficulties than e.g., the funnel region, but to the contrary, already the first FDE cycle yields an improvement over the MFCC results. Note again that the MFCC and 3-FDE calculations do not include solvation. Therefore, the larger errors can in part be attributed to the stabilization of charges at the protein surface due to the inclusion of a COSMO solvent environment in the supermolecular calculation.

7. TOWARD QUANTUM-CHEMICAL DOCKING SIMULATIONS

So far, we have focused on the quality of the electron density as this is a fundamental quantity from which many molecular properties can be derived. In this final section, we will provide an assessment of the expected accuracy of the resulting electrostatic potential and the efficiency of the procedure

relative to conventional calculations. Both aspects are important for the applicability of the 3-FDE scheme in future applications in quantum-chemical docking simulations, which we plan to explore in more detail in future work.

One property that is of direct relevance to chemical reactivity and molecular recognition is the electrostatic potential $V(\mathbf{r}) = V_{nuc}(\mathbf{r}) + \int[\rho(\mathbf{r})/|\mathbf{r} - \mathbf{r}'|]\, d^3r'$ at possible interaction sites. This is the most important ingredient for estimating the interaction energy of a protein with some probe molecule in docking simulations. In order to assess errors in this quantity, we will look at the electrostatic potential (calculated from the fitted electron density) on a grid surrounding the protein for selected MFCC, 3-FDE, and supermolecular reference calculations. As a convenient grid we employed the coordinates of the COSMO surface segments generated with the default parameters for water. For each point r_i in this grid we can calculate the absolute error $\Delta V(\mathbf{r}_i) = |V^{KS}(\mathbf{r}_i) - V^{FDE}(\mathbf{r}_i)|$ of the electrostatic potential obtained from the MFCC or 3-FDE calculation relative to one from the supermolecular KS reference calculation. For a local analysis, only the grid points (COSMO segments) belonging to the amino acids of the interaction site are employed.

For a probe charge q, the error $\Delta V(\mathbf{r}_i)$ in the electrostatic potential causes an error $\Delta E(\mathbf{r}_i) = q\Delta V(\mathbf{r}_i)$ in the interaction energy. Thus, an average error in the interaction energy with a probe charge can be calculated as

$$\Delta E^{probe} = \frac{\sum_i S_i|\Delta E(\mathbf{r}_i)|}{\sum_i S_i} = \frac{q\sum_i S_i|\Delta V_i(\mathbf{r}_i)|}{\sum_i S_i} \quad (9)$$

where we weight each grid point with the area S_i of the corresponding COSMO surface element. For the probe charge, we chose an (arbitrary) value of $q = 0.1$ e, which is the order of magnitude of the atomic partial charges in typical probe molecules. In addition, we can calculate a relative error in the interaction energy as

$$\Delta V^{rel} = \frac{\sum_i S_i|\Delta V(\mathbf{r}_i)|}{\sum_i S_i|V^{KS}(\mathbf{r}_i)|} \quad (10)$$

which corresponds to a relative error in the electrostatic potential because the value of the probe charge q drops out.

The measures have to be considered as an upper bound for the errors in the electrostatic interaction energy with a probe molecule. Typically, we can expect that the errors are similar for neighboring grid points. Therefore, for a neutral probe molecule, the errors in the electrostatic potential will cause errors of similar magnitude and opposite sign for positive and negative partial charges that largely cancel. In addition, in docking simulations the position of the probe molecule will adjust to the changed electrostatic potential, possibly decreasing the error in the interaction energy further. Finally, one can expect further error cancellation if only the difference between the interaction energies of different probe molecules are of interest.

In Table 7, the results for ΔE^{probe} and for ΔV^{rel} are given, with the Δ^{abs} error in the electron density added for comparison. Results are shown for the interacting face of ubiquitin (see section 3.2) and for the funnel of the bubble protein (see section 6.2), using the partitioning into single amino acid fragments. In all cases, we can observe a clear correlation between the error in the density and the error in the electrostatic potential, which is in turn related to the error in the interaction energy with a probe charge. If the error in the electron density becomes smaller, also the error in the electrostatic potential around the molecule decreases. For the interacting face of ubiquitin, ΔE^{probe} decreases from 13.0 kJ/mol with MFCC to 10.4 kJ/mol with FDE(1), which corresponds to relative errors of 36.6% and 29.2%, respectively. For the funnel of the bubble protein, the decrease is even more pronounced with 22.8 kJ/mol for MFCC and 10.9 for FDE(0), i.e., the error in the electrostatic potential and in the interaction energy is halved. For both proteins, a similar error as for the full FDE(0) or FDE(1) calculations is already achieved in the MFCC+FDE(0) or MFCC+FDE(1) calculations, in which only the amino acids at the interacting site are updated.

Table 7. Integrated Absolute Error Δ^{abs} in the Electron Density (e·bohr^{-3}), Average Error ΔE^{probe} in the Interaction Energy with a Probe Charge of $q = 0.1$ e (kJ/mol), and Average Relative Error in the Electrostatic Potential, ΔV^{rel} for MFCC and 3-FDE(n) calculations[a]

		$\Delta^{abs} \times 10^3$	ΔE^{probe}	ΔV^{rel}
ubiquitin (interaction site)				
A	MFCC	4.67	13.0	36.6%
	3-FDE(0)	4.12	12.9	36.3%
	3-FDE(1)	2.82	10.4	29.2%
A−F	MFCC+3-FDE(0)	4.26	11.2	31.5%
	MFCC+3-FDE(1)	3.44	10.9	30.8%
bubble protein (funnel)				
A	MFCC	6.01	22.8	26.2%
	3-FDE(0)	7.94	10.9	11.5%
	3-FDE(1)	5.46	12.6	14.1%
A-F	MFCC+3-FDE(0)	7.93	16.7	14.5%
	MFCC+3-FDE(1)	6.59	19.9	17.3%

[a]All errors for ubiquitin refer to the interacting face, while all errors for the bubble protein refer to the funnel site.

Comparing the density and potential errors for FDE(1) and MFCC+FDE(1), one notices that while the error in the density is larger for MFCC+FDE(1), the errors in the electrostatic potential are very similar. This is understandable, as the density errors in MFCC+FDE(1) compared to FDE(1) will be larger at the interface between the MFCC and the 3-FDE regions. This is further away from the grid points at which the electrostatic potential is probed, so it affects the error in the electrostatic potential only slightly. The order of magnitude of the errors in the interaction energy with a small probe charge of more than 10−20 kJ/mol and of relative errors of approximately 15−35% seems rather large at first glance. However, as discussed above, these have to be considered as an upper bound for errors in the interaction energy with real probe molecules. Here, the definition of ΔE^{probe} and for ΔV^{rel} in terms of absolute values does not allow for any kind of error cancellation. Moreover, the errors in the interaction energies will further decrease if a partitioning into larger fragments is considered. We also note that errors of ca. 20 kJ/mol were also observed in the MFCC study in ref 24 for a partitioning into single amino acid fragments. A better assessment of the error by docking probe molecules using the obtained electrostatic potentials is planned for a follow-up study but falls outside the scope of the current work. We expect that the decrease in the error measures considered here will correspond to a significant decrease in error in relative energies when comparing different probe molecules.

dx.doi.org/10.1021/ct3008759 | J. Chem. Theory Comput. 2013, 9, 2425−2440

Finally, the efficiency of the quantum-chemical calculation is the main bottleneck for their routine application in docking simulations. Even though efficiency of the implementation has not been our main focus so far, we will provide some timings comparing fragment and supermolecular calculations (cf. Table 8) to indicate the performance. All calculations mentioned were

Table 8. Exemplary Timings for Selected Supermolecular, MFCC, and 3-FDE Calculations Run on 16 Cores[a]

ubiquitin		KS	6:24
	A	MFCC	0:46
		3-FDE(0)	14:56
	A, parallel FT	3-FDE(1)	11:00
		total	26:42
	A	MFCC	0:46
	A−F	MFCC+3-FDE(0)	3:14
		MFCC+3-FDE(1)	2:15
		total	6:15
bubble protein		KS	4:37
w/o solvation	A	MFCC	0:40
		3-FDE(0)	12:12
		3-FDE(1)	11:43
		total	24:35
	A	MFCC	0:40
w/o solvation	A−F	MFCC+3-FDE(0)	5:10
		MFCC+3-FDE(1)	3:15
		total	9:05

[a]Times are given separately for each step as well as the total time (h:min).

performed on a compute cluster with 2.26 GHz Intel 8-core nodes, each equipped with 24 GB memory and connected via Infiniband. For the supermolecular calculations, 16 cores distributed over 4 nodes were used because of the memory requirements per process, while for the 3-FDE calculations, 2 nodes with 8 cores each could be employed. As we only use a small number of nodes, the scaling of ADF with the number of cores is almost linear. For our examples, the supermolecular reference calculation is still significantly faster (6 times for the bubble protein and 4 times for the larger ubiquitin) than FDE(1) with full updates. This is due to the slow convergence of the cap generation procedure, which requires many iterations to converge. The cost per iteration for the 3-FDE calculations is, however, significantly smaller than a supermolecular calculation as is visible in the individual timings for the MFCC calculation for which the cap potential is not required. We also note that the 3-FDE(1) update is slightly faster than the 3-FDE(0) calculations as fewer iterations are needed to converge.

Already for these two examples we can see that the ratio between the supermolecular and 3-FDE timings gets more favorable when the size of the protein is increased, as expected. Before the point of equal computational times would be reached, however, the supermolecular calculation would probably not be feasible anymore, as the required memory grows quadratically with system size, while the memory requirements for 3-FDE remain constant. This was already evident for our relatively small test systems. However, even with the current implementation 3-FDE becomes advantageous if only a smaller part of the protein is of interest. For ubiquitin, the required computer time is similar for MFCC+FDE(1) as for the supermolecular calculation, while the MFCC+FDE(0) calculation is significantly faster while providing results of similar accuracy. Note that, as the iterations for the caps are the bottleneck in the current implementation, the computational effort decreases for the partitionings with larger fragments. Thus, the timings presented here can be considered as a worst-case scenario. For future applications, further improvements in the determination of the cap potential are possible, for instance by expanding the cap potential in a basis set.[41] We will address such performance enhancements in future work.

8. CONCLUSIONS

We have extended the 3-FDE scheme such that it is applicable to a wide range of proteins, including proteins containing disulfide bonds and proteins with charged side groups. Further we have introduced more flexibility into the setup to allow the user to define partitioning schemes using chemical insight. With partial updates in the iterative 3-FDE scheme, focusing on selected parts of the protein becomes possible. This will be essential for the fully quantum-chemical description of protein−protein and protein−ligand interactions.

For proteins with neutral side groups, the electron density and the dipole moment can be reproduced with good accuracy with two 3-FDE cycles, one 3-FDE cycle already yields decent densities and dipole moments and an improvement over the MFCC densities. Small artifacts can be observed in the density of the cap regions, so that the MFCC results yield a better description for the backbone, but for the side groups 3-FDE takes the influence of the environment into account. Larger fragment sizes decrease the errors in the density, which is especially important for secondary structure elements. For a hotspot in a protein, choosing larger fragment sizes in combination with including a few neighboring fragments leads to accurate results at reduced costs.

Proteins with charged side groups pose a challenge for DFT calculations in general because of the charges localized on different parts of the molecule. In the standard KS-DFT calculation solvation is a way to stabilize these charges. With 3-FDE, localization of charges is achieved by assigning them to fragments, so that a solvation treatment is not necessary to achieve convergence. It precludes a flow of electrons via peptide and hydrogen bonds, but this might turn out to be an advantage for cases where the charge needs to be localized.

The current implementation of the cap potential construction is the major computer-time-consuming part which leaves room for large improvements. For small proteins, there is therefore currently no reduction of computer time with respect to the conventional supermolecular DFT calculation. Large systems on the other hand, which are not feasible for conventional DFT calculations due to memory requirements and increased difficulties in the treatment of localized charges, are still possible with 3-FDE. However, one is usually not interested in electron densities of full proteins, but rather in an accurate description of some interacting site. For describing protein−protein and protein−ligand interactions, this will be an interacting face or a binding pocket, respectively. Here, we have demonstrated that in such scenarios, 3-FDE can provide accurate electron densities of selected parts of a protein at a significantly reduced computational cost. This can pave the way to accurate quantum-chemical docking simulations.

■ ASSOCIATED CONTENT

Supporting Information

Tables with details on how ubiquitin and the bubble protein were partitioned and tables with errors in electron density and dipole moment for all studied systems containing results for more FT iterations. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: +49 721 608 48032. Fax: +49 721 608 48496. E-mail: christoph.jacob@kit.edu (C.R.J.). Tel.: +31 20 59 87624. Fax: +31 20 59 87629. E-mail: l.visscher@vu.nl (L.V.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge financial support by the German Academic Exchange Service, DAAD, (fellowship within the Postdoc-Programme for K.K.), the DFG-Center for Functional Nanostructures at KIT (C.R.J.), and the Netherlands Organization for Scientific Research, NWO, (L.V. and K.K.), as well as computer time provided by the Dutch National Computing Facilities (NCF) at the LISA facility at SURFsara.

■ REFERENCES

(1) Moal, I. H.; Agius, R.; Bates, P. A. Protein−protein binding affinity prediction on a diverse set of structures. *Bioinformatics* **2011**, *27*, 3002−3009.

(2) Sotriffer, C. In *Protein-Ligand Interactions*, 1st ed.; Gohlke, H., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2012; pp 237−263.

(3) Mackerell, A. D., Jr. Empirical Force Fields for Biological Macromolecules: Overview and Issues. *J. Comput. Chem.* **2004**, *25*, 1584−1604.

(4) Lopes, P. E. M.; Roux, B.; MacKerell, A. D., Jr Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: theory and applications. *Theor. Chem. Acc.* **2009**, *124*, 11−28.

(5) Goedecker, S. Linear scaling electronic structure methods. *Rev. Mod. Phys.* **1999**, *71*, 1085−1123.

(6) Ochsenfeld, C.; Kussmann, J.; Lambrecht, D. S. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Cundari, T. R., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, 2007; pp 1−82.

(7) Fox, S. J.; Pittock, C.; Fox, T.; Tautermann, C. S.; Malcolm, N.; Skylaris, C.-K. Electrostatic embedding in large-scale first principles quantum-mechanical calculations on biomolecules. *J. Chem. Phys.* **2011**, *135*, 224107.

(8) Fox, S. J.; Wallnoefer, H. G.; Fox, T.; Tautermann, C. S.; Skylaris, C.-K. First Principles-Based Calculations of Free-Energy of Binding: Application to Ligand Binding in a Self-Assembling Superstructure. *J. Chem. Theory Comput.* **2011**, *7*, 1102−1108.

(9) Vasilyev, V.; Bliznyuk, A. Application of semiempirical quantum chemical methods as a scoring function in docking. *Theor. Chem. Acc.* **2004**, *112*, 313−317.

(10) Dobeš, P.; Fanfrlík, J.; Řezáč, J.; Otyepka, M.; Hobza, P. Transferable scoring function based on semiempirical quantum mechanical PM6-DH2 method: CDK2 with 15 structurally diverse inhibitors. *J. Comput. Aided Mol. Des.* **2011**, *25*, 223−235.

(11) Hayik, S. A.; Dunbrack, R., Jr.; Merz, K. M., Jr. Mixed Quantum Mechanics/Molecular Mechanics Scoring Function To Predict Protein−Ligand Binding Affinity. *J. Chem. Theory Comput.* **2010**, *6*, 3079−3091.

(12) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227−249.

(13) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.* **1990**, *11*, 700−733.

(14) Senn, H.; Thiel, W. In *Atomistic Approaches in Modern Biology*; Reiher, M., Ed.; Topics in Current Chemistry; Springer: Berlin/Heidelberg, 2007; Vol. *268*; pp 173−290.

(15) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2012**, *112*, 632−672.

(16) Gomes, A. S. P.; Jacob, C. R. Quantum-chemical embedding methods for treating local electronic excitations in complex chemical systems. *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.* **2012**, *108*, 222−277.

(17) Neugebauer, J. Chromophore-specific theoretical spectroscopy: From subsystem density functional theory to mode-specific vibrational spectroscopy. *Phys. Rep.* **2010**, *489*, 1−87.

(18) Zhang, D. W.; Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein−molecule interaction energy. *J. Chem. Phys.* **2003**, *119*, 3599−3605.

(19) Gao, A. M.; Zhang, D. W.; Zhang, J. Z. H.; Zhang, Y. An efficient linear scaling method for ab initio calculation of electron density of proteins. *Chem. Phys. Lett.* **2004**, *394*, 293−297.

(20) Mei, Y.; Zhang, D. W.; Zhang, J. Z. H. New Method for Direct Linear-Scaling Calculation of Electron Density of Proteins. *J. Phys. Chem. A* **2005**, *109*, 2−5.

(21) Mei, Y.; Wu, E. L.; Hand, K. L.; Zhang, J. Z. H. Treating Hydrogen Bonding in Ab Initio Calculation of Biopolymers. *Int. J. Quantum Chem.* **2006**, *106*, 1267−1276.

(22) Jiang, N.; Ma, J.; Jiang, Y. Electrostatic field-adapted molecular fractionation with conjugated caps for energy calculations of charged biomolecules. *J. Chem. Phys.* **2006**, *124*, 114112.

(23) Li, S.; Li, W.; Fang, T. An Efficient Fragment-Based Approach for Predicting the Ground-State Energies and Structures of Large Molecules. *J. Am. Chem. Soc.* **2005**, *127*, 7215−7226.

(24) Antony, J.; Grimme, S. Fully *ab initio* Protein−Ligand Interaction Energies with Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2012**, *33*, 1730−1739.

(25) Zhang, D. W.; Xiang, Y.; Zhang, J. Z. H. New Advance in Computational Chemistry: Full Quantum Mechanical ab Initio Computation of Streptavidin−Biotin Interaction Energy. *J. Phys. Chem. B* **2003**, *107*, 12039−12041.

(26) Wesolowski, T. A.; Warshel, A. Frozen Density Functional Approach for ab Initio Calculations of Solvated Molecules. *J. Phys. Chem.* **1993**, *97*, 8050−8053.

(27) Jacob, C. R.; Visscher, L. A subsystem density-functional theory approach for the quantum chemical treatment of proteins. *J. Chem. Phys.* **2008**, *128*, 155102.

(28) Wesolowski, T. A.; Weber, J. Kohn-Sham equations with constrained electron density: an iterative evaluation of the ground-state electron density of interacting molecules. *Chem. Phys. Lett.* **1996**, *248*, 71−76.

(29) Kiewisch, K.; Eickerling, G.; Reiher, M.; Neugebauer, J. Topological analysis of electron densities from Kohn−Sham and subsystem density functional theory. *J. Chem. Phys.* **2008**, *128*, 044114.

(30) Jacob, C. R.; Neugebauer, J.; Jensen, L.; Visscher, L. Comparison of frozen-density embedding and discrete reaction field solvent models for molecular properties. *Phys. Chem. Chem. Phys.* **2006**, *8*, 2349−2359.

(31) Bulo, R. E.; Jacob, C. R.; Visscher, L. NMR Solvent Shifts of Acetonitrile from Frozen Density Embedding Calculations. *J. Phys. Chem. A* **2008**, *112*, 2640−2647.

(32) Neugebauer, J.; Louwerse, M. J.; Baerends, E. J.; Wesolowski, T. A. The merits of the frozen-density embedding scheme to model solvatochromic shifts. *J. Chem. Phys.* **2005**, *122*, 094115.

(33) Neugebauer, J.; Jacob, C. R.; Wesolowski, T. A.; Baerends, E. J. An Explicit Quantum Chemical Method for Modeling Large Solvation Shells Applied to Aminocoumarin C151. *J. Phys. Chem. A* **2005**, *109*, 7805−7814.

(34) Neugebauer, J.; Louwerse, M. J.; Belanzoni, P.; Wesolowski, T. A.; Baerends, E. J. Modeling solvent effects on electron spin resonance hyperfine couplings by frozen-density embedding. *J. Chem. Phys.* **2005**, *123*, 114101.
(35) Neugebauer, J. Photophysical Properties of Natural Light-Harvesting Complexes Studied by Subsystem Density Functional Theory. *J. Phys. Chem. B* **2008**, *112*, 2207−2217.
(36) Neugebauer, J. Subsystem-Based Theoretical Spectroscopy of Biomolecules and Biomolecular Assemblies. *Chem. Phys. Chem.* **2009**, *10*, 3148−3173.
(37) Fux, S.; Kiewisch, K.; Jacob, C.; Neugebauer, J.; Reiher, M. Analysis of Electron Density Distributions from Subsystem Density Functional Theory Applied to Coordination Bonds. *Chem. Phys. Lett.* **2008**, *461*, 353−359.
(38) Goodpaster, J. D.; Ananth, N.; Manby, F.; Miller, T. F., III Exact nonadditive kinetic potentials for embedded density functional theory. *J. Chem. Phys.* **2010**, *133*, 084103.
(39) Goodpaster, J. D.; Barnes, T. A.; Miller, T. F., III Embedded density functional theory for covalently bonded and strongly interacting subsystems. *J. Chem. Phys.* **2011**, *134*, 164108.
(40) Fux, S.; Jacob, C.; Neugebauer, J.; Visscher, L.; Reiher, M. Accurate frozen-density embedding potentials as a first step towards a subsystem description of covalent bonds. *J. Chem. Phys.* **2010**, *132*, 164101.
(41) Jacob, C. R. Unambiguous optimization of effective potentials in finite basis sets. *J. Chem. Phys.* **2011**, *135*, 244102.
(42) Fedorov, D. G.; Kitaura, K. Extending the Power of Quantum Chemistry to Large Systems with the Fragment Molecular Orbital Method. *J. Phys. Chem. A* **2007**, *111*, 6904−6914.
(43) Fedorov, D. G.; Nagata, T.; Kitaura, K. Exploring chemistry with the fragment molecular orbital method. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7562−7577.
(44) Fischer, B.; Fukuzawa, K.; Wenzel, W. Receptor-specific scoring functions derived from quantum chemical models improve affinity estimates for *in silico* drug discovery. *Proteins Struct. Funct. Bioinf.* **2008**, *70*, 1264−1273.
(45) *Amsterdam Density Functional program. Theoretical Chemistry*; Vrije Universiteit: Amsterdam, 2012; http://www.scm.com.
(46) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. Chemistry with ADF. *J. Comput. Chem.* **2001**, *22*, 931−967.
(47) van Leeuwen, R.; Baerends, E. J. Exchange−correlation potential with correct asymptotic behaviour. *Phys. Rev. A* **1994**, *49*, 2421−2431.
(48) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098−3100.
(49) Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **1986**, *33*, 8822−8824.
(50) Lembarki, A.; Chermette, H. Obtaining a gradient-corrected kinetic-energy functional from the Perdew−Wang exchange functional. *Phys. Rev. A* **1994**, *50*, 5328−5331.
(51) Jacob, C. R.; Beyhan, S. M.; Bulo, R. E.; Gomes, A. S. P.; Götz, A. W.; Kiewisch, K.; Sikkema, J.; Visscher, L. PyADF − A scripting framework for multiscale quantum chemistry. *J. Comput. Chem.* **2011**, *32*, 2328−2338.
(52) O'Boyle, N.; Banck, M.; James, C.; Morley, C.; Vandermeersch, T.; Hutchison, G. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33.
(53) *The Open Babel Package*, version 2.2.3; http://openbabel.org (accessed Feb 2012).
(54) Bernard, Y. A.; Dulak, M.; Kaminski, J. W.; Wesolowski, T. A. The energy-differencees based exact criterion for testing approximations to the functional for the kinetic energy of non-interacting electrons. *J. Phys. A* **2008**, *41*, 055302.
(55) Hershko, A.; Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **1998**, *67*, 425−479.
(56) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of Ubiquitin Refined at 1.8 Å Resolution. *J. Mol. Biol.* **1987**, *194*, 531−544.
(57) Dikic, I.; Wakatsuki, S.; Walters, K. J. Ubiquitin-binding domains −- from structures to functions. *Nature Rev. Mol. Cell Biol.* **2009**, *10*, 659−671.
(58) Humphrey, W.; Dalke, A.; Schulten, K. VMD — Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(59) Iannuzzi, M.; Kirchner, B.; Hutter, J. Density functional embedding for molecular systems. *Chem. Phys. Lett.* **2006**, *421*, 16−20.
(60) Chen, X. H.; Zhang, D. W.; Zhang, J. Z. H. Fractionation of peptide with disulfide bond for quantum mechanical calculation of interaction energy with molecules. *J. Chem. Phys.* **2004**, *120*, 839−844.
(61) Fonseca Guerra, C.; Handgraaf, J.-W.; Baerends, E. J.; Bickelhaupt, F. M. Voronoi Deformation Density (VDD) Charges: Assessment of the Mulliken, Bader, Hirshfeld, Weinhold, and VDD Methods for Charge Analysis. *J. Comput. Chem.* **2004**, *25*, 189−210.
(62) Fonseca Guerra, C.; Szekeres, Z.; Bickelhaupt, F. M. Remote Communication in a DNA-Based Nanoswitch. *Chem.—Eur. J.* **2011**, *17*, 8816−8818.
(63) Pavanello, M.; Neugebauer, J. Modelling charge transfer reactions with the frozen density embedding formalism. *J. Chem. Phys.* **2011**, *135*, 234103.
(64) Hong, G.; Rosta, E.; Warshel, A. Using the Constrained DFT Approach in Generating Diabatic Surfaces and Off Diagonal Empirical Valence Bond Terms for Modeling Reactions in Condensed Phases. *J. Phys. Chem. B* **2006**, *110*, 19570−19574.
(65) Xiang, Y.; Warshel, A. Quantifying Free Energy Profiles of Proton Transfer Reactions in Solution and Proteins by Using a Diabatic FDFT Mapping. *J. Phys. Chem. B* **2008**, *112*, 1007−1015.
(66) Rosta, E.; Warshel, A. Origin of Linear Free Energy Relationships: Exploring the Nature of the Off-Diagonal Coupling Elements in SN2 Reactions. *J. Chem. Theory Comput.* **2012**, *8*, 3574−3585.
(67) Wu, Q.; Van Voorhis, T. A direct optimization method to study constrained systems within density functional theory. *Phys. Rev. A* **2005**, *72*, 024502.
(68) Kaduk, B.; Kowalczyk, T.; Van Voorhis, T. Constrained Density Functional Theory. *Chem. Rev.* **2012**, *112*, 321−370.
(69) Olsen, J. G.; Flensburg, C.; Olsen, O.; Bricogne, G.; Henriksen, A. Solving the structure of the bubble protein using the anomalous sulfur signal from single-crystal in-house Cu Kα diffraction data only. *Acta Cryst. D* **2004**, *60*, 250−255.
(70) Seibold, M.; Wolschann, P.; Olsen, O. Joint occurrence of the bubble protein and mycophenolic acid in *Penicillium brevicompactum Dierckx*. *Monatsh. Chem.* **2011**, *142*, 1309−1315.
(71) Gohlke, H.; Klebe, G. Approaches to the Description and Prediction of the Binding Affinity of Small-Molecule Ligands to Macromolecular Receptors. *Angew. Chem., Int. Ed.* **2002**, *41*, 2644−2676.
(72) Høst, S.; Jansík, B.; Olsen, J.; Jørgensen, P.; Reine, S.; Helgaker, T. A ground-state-directed optimization scheme for the Kohn−Sham energy. *Phys. Chem. Chem. Phys.* **2008**, *10*, 5344−5348.
(73) Klamt, A.; Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perk. Trans. 2* **1993**, 799−805.
(74) Pichierri, F. Computation of the permanent dipole moment of α-chymotrypsin from linear-scaling semiempirical quantum mechanical methods. *J. Mol. Struct.* **2003**, *664−665*, 197−205.