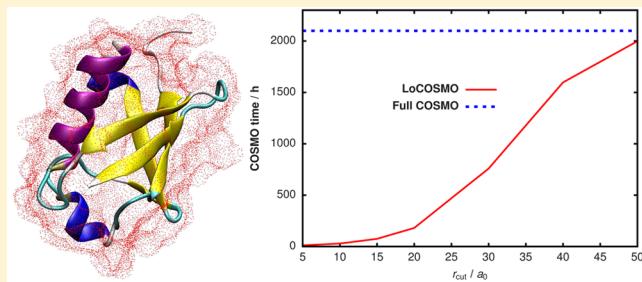


A Local Variant of the Conductor-Like Screening Model for Fragment-Based Electronic-Structure Methods

Albrecht Goez and Johannes Neugebauer*

Theoretische Organische Chemie, Organisch-Chemisches Institut and Center for Multiscale Theory and Computation, Westfälische Wilhelms-Universität Münster, Corrensstraße 40, 48149 Münster, Germany

ABSTRACT: Due to steadily rising computational power and sophisticated modeling approaches, increasingly larger molecular systems can be modeled with ab initio methods. An especially promising approach is subsystem methods, where the total system is broken down into smaller subunits that can be treated individually. If an implicit solvent environment such as the conductor-like screening model (COSMO) is included in the description, then additional environmental effects can be incorporated at relatively low cost. For very large systems described with subsystem methods, however, the solution of the COSMO equations can actually become the bottleneck of the calculation. A prominent example in this area is biomolecular systems such as proteins, which can, for instance, be described with frozen density embedding (FDE), especially the related 3-FDE approach. In this article, we present an alternative COSMO variant, which exploits the subsystem nature of the underlying electronic description and has been implemented in a development version of the Amsterdam Density Functional program suite (ADF). We show that the computational cost for the solvent model can be reduced dramatically while retaining the accuracy of the regular description. We compare several schemes for density and surface charge updates and assess the effect of the single tuning parameter.



1. INTRODUCTION

The theoretical modeling of biomolecular systems is an essential part of modern chemistry and physics. Driven by numerous methodological developments as well as growing computational power at hand, ever larger systems can be modeled with increasing accuracy.¹ Especially with regard to molecular dynamics (MD) simulations, parametrized force field descriptions have been very successful at describing certain properties of large composite systems such as proteins and protein aggregates.² However, due to the empirical nature of these methods, not all quantities of interest can be obtained. For instance, the description of energy or charge transfer processes in photosynthetic antenna complexes and reaction centers inherently requires an electronic-structure method to describe the excited states of the system.^{3,4} Such research questions create the need for methods that can, on one hand, describe the electronic structure of extended systems while, on the other hand, being computationally efficient enough to remain applicable.

One such method is subsystem DFT^{5,6} in its frozen density embedding (FDE) formulation⁷ (for recent reviews, see refs 8 and 9), which has been successfully employed to obtain information about excitation energies and excitonic couplings in photosynthetic light-harvesting complexes.^{10–12} However, a persistent problem with all available nonadditive kinetic energy functionals for FDE is their inability to describe covalently bound fragments.¹³ A pragmatic solution to this problem is presented by the three-partition frozen density embedding (3-FDE) scheme,^{14–16} which combines ideas from molecular

fractionation with conjugate caps (MFCC)¹⁷ with FDE. As has been shown previously, 3-FDE makes it possible to generate fully quantum-chemical densities for entire proteins. Several supermolecular properties such as electron densities and dipole moments^{15,16} as well as properties of embedded building blocks (e.g., excitation energies of cofactors¹²) could be obtained in this way.

Most naturally occurring light-harvesting complexes are transmembrane proteins, whereas one prominent example of a water-soluble pigment–protein complex is the well-known Fenna–Matthews–Olson (FMO) complex.¹⁸ Although, in principle, it would be possible to describe the environment around such a protein with FDE, the required number of solvent or lipid molecules is so large that calculations become unfeasible, especially if an iterative density update scheme is used to model mutual polarization.

An alternative to an explicit treatment of the surroundings is presented by continuum models. One possibility is to use a statistically averaged solvent density in the spirit of the three-dimensional reference interaction site model (3D-RISM)^{19,20} as the frozen density in a FDE calculation.^{21,22} Efficient and easy-to-use alternatives, which are very popular in computational chemistry, are apparent surface charge (ASC) models. In this context, the polarizable continuum model (PCM)^{23,24} and the conductor-like screening model (COSMO),²⁵ especially in its advanced formulation, COSMO-RS,^{26,27} are among the most

Received: August 31, 2015

Published: October 22, 2015

prominent examples. In both models, the environment is described by a number of point charges on a cavity surface, which is constructed around the solute.

When dealing with very large systems such as light-harvesting proteins, the size of the cavity can become enormous, leading to a vast number of ASCs (see [Section 2.3](#)). If very efficient fragment-based electronic-structure methods are used, then the computational effort in terms of CPU time as well as memory requirements for the ASC model can become so large that it dominates the overall calculation.¹² A number of schemes to reduce the computational complexity of COSMO have been put forward, with the most notable being the domain decomposition COSMO (ddCOSMO) scheme.^{28–30} However, its advantage in terms of efficiency vanishes if the embedded system contains a large region described by a quantum-mechanical Hamiltonian.³¹ Since the goal of most FDE calculations is to describe the entire system quantum-mechanically, the ddCOSMO scheme does not present an advantage over a regular COSMO calculation in this case. Another possibility to reduce the numerical effort is the fast multipole method (FMM),³² which offers an efficient treatment of the numerous electrostatic interactions between the ASCs. This method has been successfully combined with the PCM scheme³³ by Scalmani and co-workers.

In this work, we propose a straightforward variant of the original COSMO scheme, which is especially adapted to the use of subsystem methods such as FDE. Our method makes use of the fact that at every point in a FDE calculation only a small part of the density (around the active system) is expected to change. Therefore, only the ASCs in spatial proximity to that region have to be subjected to the optimization procedure. We show that for large systems the new algorithm leads to a tremendous speedup while maintaining an accurate description. If only one update cycle over the fragment densities is applied, then this scheme is an approximation to regular COSMO. However, if several so-called freeze-and-thaw (FT) cycles³⁴ (see [Section 2.1](#)) are carried out (as is often the case), then the variant proposed here converges to the results of a regular COSMO calculation.

Although we use the new method only in conjunction with 3-FDE and apply it to proteins, we would like to emphasize that the underlying idea is not only transferable to regular FDE but also can be easily generalized to other fragment-based methods as well, such as the fragment molecular orbital method by Kitaura et al.^{35–37} Several other examples of such methods can be found in a recent review.³⁸ Furthermore, since the surface charges determined in a regular COSMO run are the basis for a COSMO-RS calculation, the method should be applicable in the context of the latter formulation as well. If the resulting active spaces of ASCs are large enough, then it might also be beneficial to combine our new algorithm with the FMM; however, as the underlying quantum-mechanical treatment usually limits the active part of the system to relatively small fragments, we rarely expect this to be the case.

This article is organized as follows: In [Sections 2.1](#) and [2.2](#), we briefly review the foundations of 3-FDE and regular COSMO. We then go on to describe the newly developed local COSMO algorithm (LoCOSMO) in [Section 2.3](#). In [Section 3](#), the investigated test systems are introduced, and the obtained results are presented and discussed in [Section 4](#). We summarize and conclude in [Section 5](#).

2. THEORETICAL BACKGROUND

2.1. Three-Partition Frozen Density Embedding (3-FDE). In this section, the most important points of the 3-FDE method will be briefly reviewed. For a detailed discussion, we refer the reader to refs 15 and 16.

The basis of the 3-FDE method is the partitioning of the supersystem into subunits according to the MFCC scheme. When dealing with proteins, the total molecule is usually split into chunks of several amino acid residues, with each fragment being cut at the peptide bond connecting it to the next fragment. The dangling bonds of the resulting oligopeptides are then saturated with small capping groups, where the most common choice is uncharged amino acid termini. The fragmentation and capping are graphically represented in [Figure 1](#).

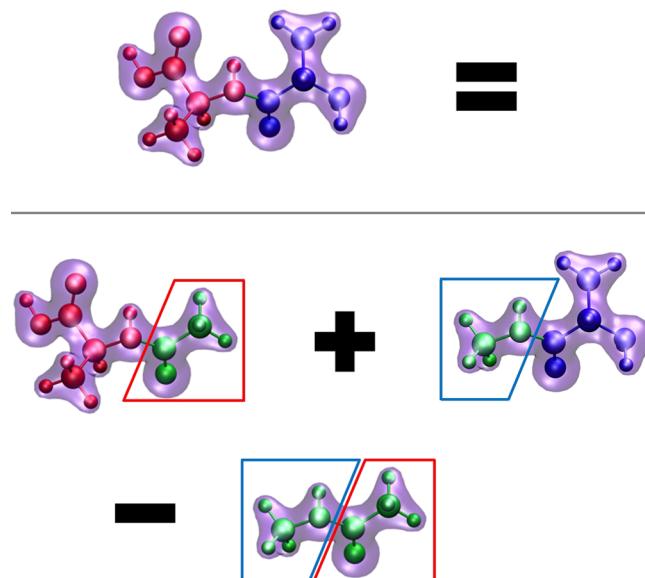


Figure 1. Schematic representation of the partitioning scheme used for MFCC and 3-FDE calculations on peptides or proteins.

Next, each capped fragment is subjected to an isolated calculation to determine an initial density. In addition, pairs of adjacent capping groups are linked together to form so-called cap molecules and subjected to a calculation as well. A first rough approximation to the supermolecular density can be obtained by summing the densities of the fragments and subtracting the ones of the cap molecules

$$\rho_{\text{tot}} = \sum_A^K \rho_A^{\text{capped}} - \sum_B^L \rho_B^{\text{cap}} \quad (1)$$

Here, K is the total number of fragments and L is the corresponding number of cap molecules. Naturally, the resulting supermolecular density contains large errors, particularly in the regions where bonds have been cut to create the subsystems. Furthermore, unphysical negative densities can arise due to the subtraction procedure.

The central idea of the 3-FDE scheme is to improve the results of this straightforward approach by updating each fragment density under the influence of an embedding potential constructed from the densities of all other fragments (analogous to regular FDE). This process can be repeated iteratively until self-consistency is reached. However, an

additional constraint is introduced, which forces the density in the cap regions of the capped fragments to be equal to the one in the corresponding cap molecules. In this way, the possibility of obtaining negative densities is eliminated. The Kohn–Sham-like equations of the 3-FDE approach can be written as

$$\left[-\frac{\nabla^2}{2} + v_{\text{eff}}^{\text{3-FDE}}[\rho_I, \rho_{II}](\vec{r}) \right] \phi_i^{(I)}(\vec{r}) = e_i \phi_i^{(I)}(\vec{r}) \quad (2)$$

where ρ_I denotes the density of the currently active subsystem and ρ_{II} represents the frozen density containing contributions from all of the other capped fragments (i.e., ρ_{II} is calculated from eq 1 but with the sum over A excluding the active fragment). The effective potential $v_{\text{eff}}^{\text{3-FDE}}$ contains not only the regular Kohn–Sham potential of the active fragment $v_{\text{eff}}^{\text{KS}}$ and the embedding potential due to the inactive fragments $v_{\text{eff}}^{\text{emb}}$ but also enforces the mentioned equality constraint for the cap regions V_I^{cap} by applying a special cap potential $v_{\text{cap}}(\vec{r})$

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

The explicit form of the potentials can be found in refs 15 and 16. If the density of the whole system is of interest, then each fragment density can, in turn, be relaxed in the embedding potential generated by all the other “frozen” fragments in individual calculations. This approach is often called freeze-and-thaw (FT)³⁴ and can, in principle, be applied until a completely self-consistent density is obtained.

2.2. The Conductor-Like Screening Model (COSMO).

The brief introduction of the COSMO algorithm presented in this section has been adapted from ref 39. For an in-depth discussion of the method, we refer the reader to that paper as well as the original publication by Klamt and Schüürmann.²⁵

In COSMO, the total electrostatic part of the solvation energy is expressed as

$$E_{\text{solv}} = \frac{1}{2} \sum_{\mu} \sum_{\nu} q_{\mu} A_{\mu\nu} q_{\nu} + \sum_A \sum_{\mu} q_{\mu} B_{A\mu} Z_A + \sum_{\mu} q_{\mu} C_{\mu} \quad (4)$$

where q_{μ} are the apparent surface charge magnitudes and Z_A are the atomic numbers of the nuclei in the system. The interaction matrix elements $A_{\mu\nu}$, $B_{A\mu}$, and C_{μ} are defined as follows

$$A_{\mu\nu} = \begin{cases} |\vec{r}_{\mu} - \vec{r}_{\nu}|^{-1} & \text{for } \mu \neq \nu \\ 1.07 \sqrt{\frac{4\pi}{S_{\mu}}} & \text{for } \mu = \nu \end{cases} \quad (5)$$

$$B_{A\mu} = |\vec{r}_{\mu} - \vec{r}_A|^{-1} \quad (6)$$

$$C_{\mu} = \int_V V_{\mu}(\vec{r}) \rho(\vec{r}) d\vec{r} \quad (7)$$

Here, \vec{r}_{μ} , \vec{r}_{ν} , and \vec{r}_A are the positions of the two apparent surface charges μ and ν and the nucleus A , respectively. S_{μ} is the surface area associated with charge μ , and $V_{\mu}(\vec{r})$ is the potential due to this charge, which acts on the electron density $\rho(\vec{r})$ inside the cavity.

In the ADF program,^{40,41} there are two ways to calculate the interaction vector C_{μ} . Each element can be evaluated either as a sum over all points of the numerical integration grid

$$C_{\mu} = \sum_k V_{\mu}(\vec{r}_k) \rho(\vec{r}_k) W(\vec{r}_k) \quad (8)$$

where the sum runs over all grid points k at positions \vec{r}_k with weights $W(\vec{r}_k)$, or as the electrostatic potential due to the electron density at the ASC positions r_{μ} in the form

$$C_{\mu} = V_{\text{Coul}}(\vec{r}_{\mu}) \quad (9)$$

Due to the much smaller number of point charges compared to integration points, the second choice is preferred and has been made the default.³⁹ For FDE calculations, it is indeed the only possible scheme, as a supermolecular grid is usually not available. The Coulomb potential can be easily evaluated from the atom-centered density fit functions. As preparation for this work, we have implemented and tested the combination of FDE and COSMO in the ADF program. This generalization is straightforward, as the only difference to supermolecular COSMO is that the Coulomb potential due to the frozen fragments has to be included in eq 9.

Since the positions of the ASCs are determined before the calculation, the only free parameters to minimize the solvation energy in eq 4 are their magnitudes q_{μ} . The corresponding set of equations reads

$$\frac{\partial E_{\text{solv}}}{\partial q_{\mu}} = \sum_{\nu} A_{\mu\nu} q_{\nu} + \sum_A B_{A\mu} Z_A + C_{\mu} ! = 0 \quad (10)$$

which, for all charges, can be put in matrix form as

$$A\vec{q} = -(B\vec{Z} + \vec{C}) \quad (11)$$

Apart from a direct inversion of the matrix A , iterative schemes like the Gauss–Seidel or Jacobi method can be used. In the implementation of COSMO in ADF, the minimum-residual biconjugate gradient method was deemed to be the most stable and still remains the default today.³⁹

2.3. Local COSMO (LoCOSMO). **2.3.1. Mathematical Formulation.** Usually, the solution of the COSMO equations is responsible only for a small fraction of the total calculation time. However, subsystem methods like FDE can reduce the cost of the quantum-chemical part of a calculation so much that assembling the matrix A and iteratively solving the linear equation system in eq 11 actually become the bottleneck for large systems such as proteins. This is particularly unfortunate for subsystem DFT calculations, where the active system might be very small but still has to be embedded in the cavity for the supersystem. In Table 1, the size of A , B , and \vec{C} for different example systems is listed. It can be easily seen that matrix A becomes huge, thus dominating the computational effort. However, in FDE calculations using the freeze-and-thaw

Table 1. Number of COSMO Matrix Elements for Different Example Systems^a

matrix	dimension	H ₂ O	Ala ₄	ubiquitin	FMO
A	$O([n^{\text{tot}}]^2)$	12 769	2 399 401	4.2×10^8	7.0×10^9
B	$O(n^{\text{tot}} \cdot n^{\text{nuc}})$	339	66 607	2.5×10^7	5.8×10^8
\vec{C}	$O(n^{\text{tot}})$	113	1549	2.1×10^4	8.4×10^4

^aThe dimensions of the matrices are given in terms of the total number of ASCs (n^{tot}) and the number of nuclei in the system (n^{nuc}).

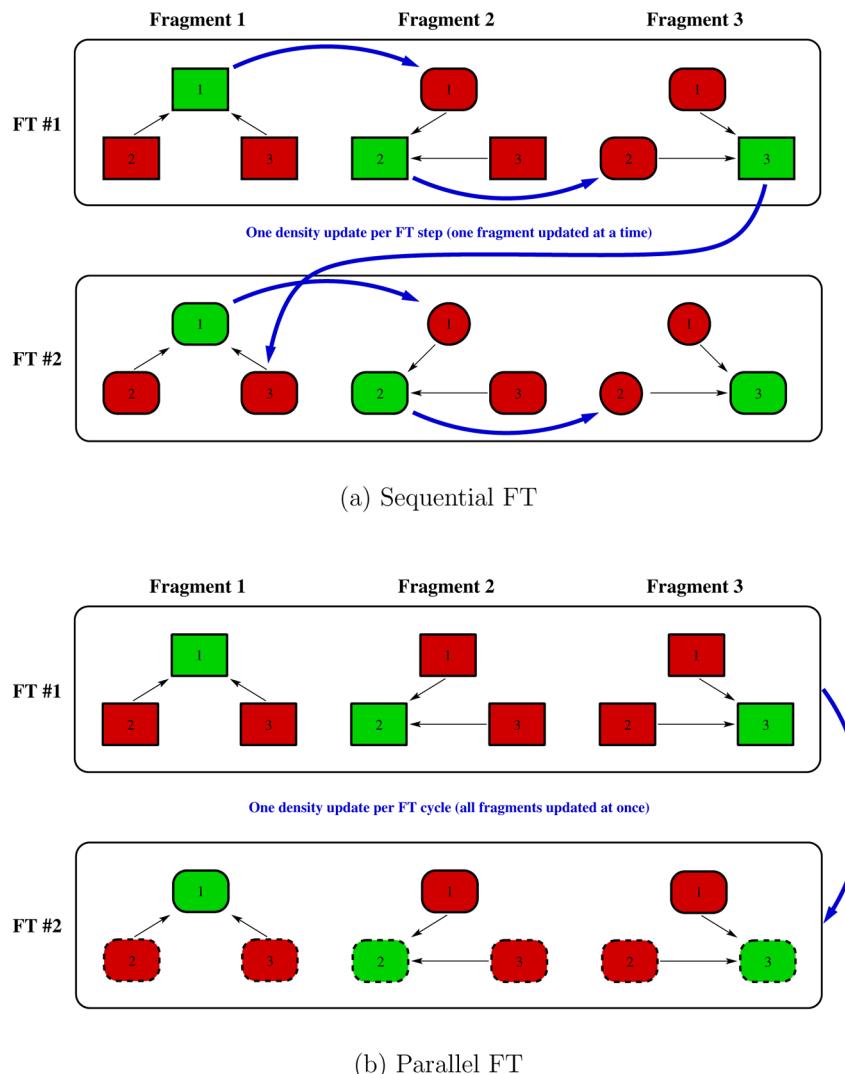


Figure 2. Graphical representation of the sequential and parallel FT schemes for a three-fragment example. Thin black arrows denote the influence of an embedding potential generated by the frozen fragments (red) on the active fragment (green). Rectangular shapes signify fragments that have not been updated yet, whereas rounded corners and spherical shapes denote one and two density updates, respectively. Bold blue arrows show the passing of updated information between calculations. The dashed outlines in parallel FT calculations 4–6 signify that the updated density is not equivalent to the one from the sequential FT calculations.

approach, only one comparatively small subsystem is active in every SCF cycle. Consequently, only the ASCs close to this active system are expected to change with respect to the updated electron density.

The essence of the LoCOSMO algorithm is to include only a subset of active charges in eq 10. This leads to a reduced set of equations, since only the partial derivatives with respect to the magnitudes of the active charges have to be taken into account. In the following example with four ASCs, the charges q_1 and q_2 belong to the active subset, whereas q_3 and q_4 are inactive.

The total problem

$$\begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{pmatrix} = \begin{pmatrix} L_1 \\ L_2 \\ L_3 \\ L_4 \end{pmatrix} \quad (12)$$

is thus reduced to

$$\begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} L_1 \\ L_2 \end{pmatrix} \quad (13)$$

where $\vec{L} = -(B\vec{Z} + \vec{C})$; see eq 11. Naturally, if only two charges are subjected to the optimization procedure, then the solution vector \vec{q} has only two variable components. Consequently, all terms $A_{\mu\nu}q_{\nu}$, where q_{ν} is an inactive charge, are constant and can be transferred to the right-hand side

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} L_1 \\ L_2 \end{pmatrix} - \begin{pmatrix} A_{13} & A_{14} \\ A_{23} & A_{24} \end{pmatrix} \begin{pmatrix} q_3 \\ q_4 \end{pmatrix} \quad (14)$$

which also guarantees that the submatrix to be inverted is quadratic.

The LoCOSMO procedure thus reduces the size of the effective A matrix from ($n^{\text{tot}} \times n^{\text{tot}}$) to ($n^{\text{act}} \times n^{\text{act}}$), where n^{tot} and n^{act} refer to the total number of charges and the number of

active charges, respectively. In our implementation, the decision regarding whether a charge is included in the active set is based solely on its distance to the nuclei of the active subsystem. Each ASC is included only if its position is closer than a user-defined threshold r_{cut} to any nucleus belonging to the active system. Note that we do not neglect any interactions in our scheme, which would be difficult due to the long-range character of the Coulomb interactions. We merely do not allow the magnitudes of the inactive charges to change from some initial value over the course of the iterative COSMO procedure within a given fragment calculation.

2.3.2. Density Choice and ASC Initialization. In every fragment calculation, two crucial choices determine the outcome of the procedure. On one hand, the frozen density generating the embedding potential has to be chosen. On the other hand, the ASC distribution has to be initialized in some way at the beginning of the first FT cycle. The passing of updated information about the ASCs between individual calculations can also be handled in different ways, but, as described below, this is determined by the density choice.

2.3.2.1. Density Choice. For the very first fragment calculation, the initial density will almost always be a sum of the isolated fragment densities (for a regular FDE calculation) or the MFCC density (in the case of a 3-FDE calculation). However, as soon as an updated density for the first fragment has been determined, the embedding potential acting on the next fragment can be constructed either from the updated or the initial density. From a pragmatic point of view, we prefer the first method because it always employs the best available approximation for the fragment densities in a given step. Hence, we expect it to converge in a smaller number of update steps. Even though a dependence on the ordering of the subsystems is introduced, the final result for the converged total density is usually independent of this ordering (some examples are given in refs 42 and 43). In the literature, this process has been called sequential FT.¹⁶ In contrast, it is also possible to use only the initial fragment densities for the embedding potential in every fragment calculation throughout the whole FT cycle. This has the advantage that the whole procedure becomes trivially parallelizable, as the individual fragment calculations are then independent of each other. Therefore, this setup is sometimes referred to as parallel FT.¹⁶ It should be noted that also sequential FT calculations can, of course, be distributed over several CPU cores by employing the regular parallelization scheme in ADF, even though not more than one fragment calculation can be run simultaneously. The distinction between sequential and parallel FT is graphically represented in Figure 2.

2.3.2.2. ASC Initialization. The second important choice in setting up a LoCOSMO calculation is how to initialize the magnitudes of the inactive charges in the first step. The two main possibilities are either to carry out one full COSMO cycle (where all charges are active) or to set all values to a uniform number (e.g., the negative of the total molecular charge, $-q_{\text{mol}}$, divided by the total number of ASCs, n^{tot} , which is the default in ADF) and gradually build up the adapted distribution over the FT cycles. In the following, we will refer to these two strategies as pre-calc and build-up, respectively.

Naturally, the second method is the preferred one because for very large systems even a single full COSMO cycle might be prohibitively expensive due to CPU or memory restrictions. Still, an important advantage of the former method is the fact that meaningful results for all fragments can be obtained already

within the first FT cycle, whereas the build-up scheme requires at least one additional production pass after the first cycle.

2.3.2.3. ASC Passing. In order to be consistent with the construction of the embedding potential, the fashion in which the initial ASC distribution is passed on from one fragment calculation to the next must be determined by the FT type. When using the sequential FT scheme, the updated ASC distribution should always represent the starting point for the next fragment. In contrast, if a parallel FT cycle is carried out, then all fragment calculations start from the same initial density and thus should also use identical initial ASC distributions. This means that no passing of ASC information is necessary, which ensures trivial parallelizability.

In the case of a sequential FT calculation, the resulting charge distribution from the last fragment in one FT cycle can simply be used as a starting point for the next. In contrast, with the parallel FT scheme, a new distribution has to be assembled from the individual fragment results. This creates difficulties due to the fact that we use overlapping regions of active charges around the fragments. As a result, there are many charges for which multiple values are available from different fragment calculations. In this work, we explore different strategies for constructing an average distribution from the individual calculations, which turns out not to be trivial (see Section 4.1.2).

In summary, two choices have to be made when applying the LoCOSMO algorithm. Just like in every FDE calculation, a density update scheme (sequential or parallel FT) has to be chosen, which, in turn, determines whether ASC information has to be passed between fragment calculations. In addition, a way of initializing the ASC distribution for the very first calculation has to be specified. In this work, we explore two possibilities: one that requires a single full COSMO cycle for the initial density (pre-calc) and another that sets all charges initially to $q_{\text{mol}}/n^{\text{tot}}$ (build-up). This leads to four different modes (see Table 2), which will be examined in Section 4.

Table 2. Calculation Modes of the LoCOSMO Algorithm Explored in This Work

FT type	ASC initialization	mode
sequential	pre-calc	I
sequential	build-up	II
parallel	pre-calc	III
parallel	build-up	IV

3. COMPUTATIONAL DETAILS AND TEST SYSTEMS

3.1. Computational Details. If it is not specified otherwise, all calculations were carried out with a locally modified version of the ADF program^{40,41} and the FDE implementation from ref 44. The BP86 density functional^{45,46} was employed with a DZP basis featuring a 1s frozen core for C, H, N, and O and a 2p frozen core for S. The test systems were split in fragments consisting of one amino acid residue each, and the default parameters of the ADF 2013 release version were used for constructing the cavity around the solute. For the nonadditive kinetic energy functional in FDE calculations, we always applied the so-called PW91k approximation.^{47,48} The cap convergence threshold in 3-FDE calculations was 10^{-3} for the integrated density difference. The relative permittivity was set to 78 (for water) in all COSMO and LoCOSMO

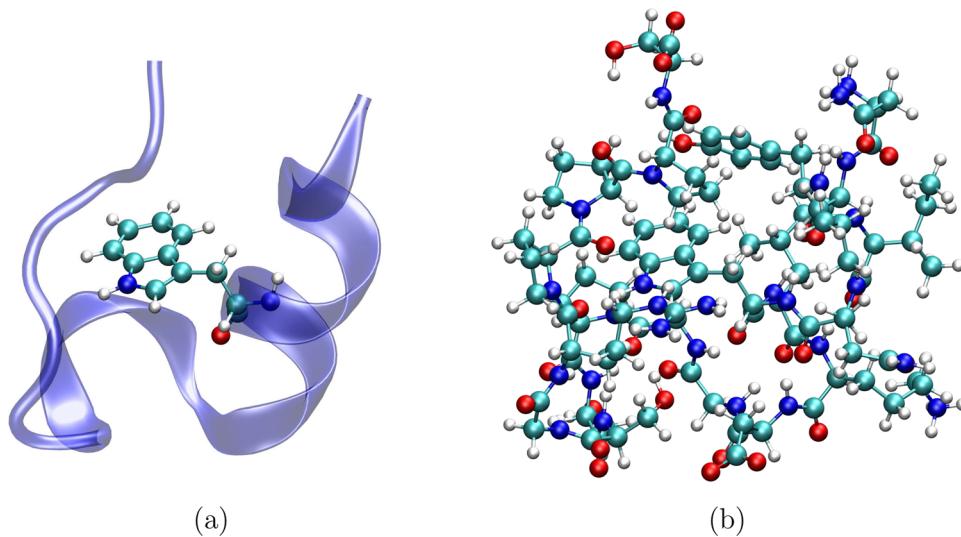


Figure 3. Structure of the TC5b protein used in this work: (a) cartoon representation and (b) atomistic CPK representation.

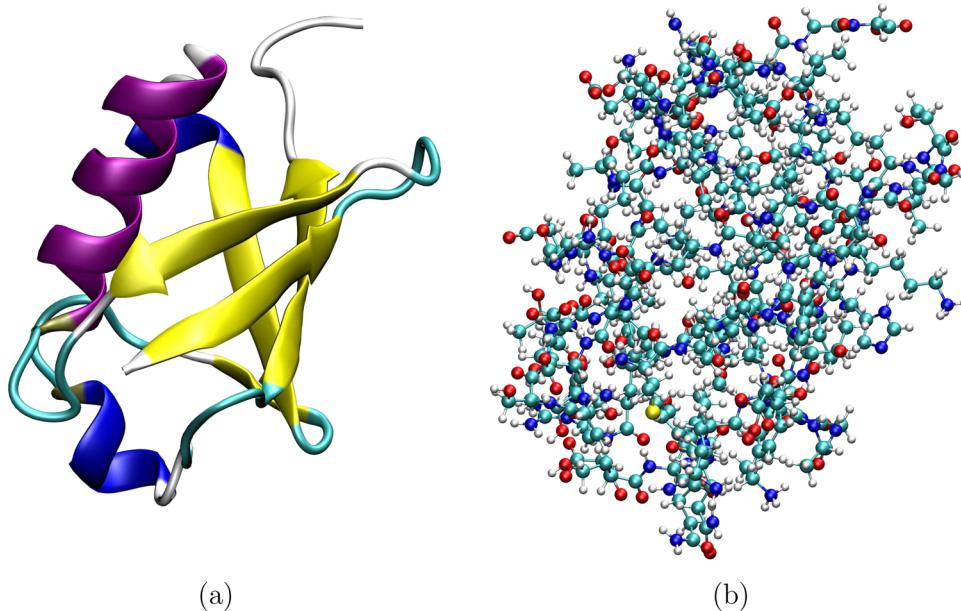


Figure 4. Structure of ubiquitin used in this work: (a) cartoon representation and (b) atomistic CPK representation.

calculations. All calculations were set up with the PYADF framework.⁴⁹

3.2. TC5b. TC5b is a synthetic protein with only 20 amino acid residues featuring the Trp-Cage motif. Due to its small size (304 atoms), it is well-suited for technical tests and method development. The structure studied in this work is derived from the original solution-state NMR data in the Protein Data Bank (ID: 1L2Y) obtained by Neidigh and co-workers.⁵⁰ While the heavy-atom coordinates were not altered, the positions of the hydrogen atoms were optimized with BP86/DZP and the regular COSMO implementation in a supermolecular ADF calculation. The resulting geometry, which was used in all subsequent calculations, is shown in Figure 3.

3.3. Ubiquitin. Ubiquitin is a popular test case for computational studies because, in spite of its relatively small size compared to other proteins (78 amino acid residues, 1231 atoms), it features the three most prominent secondary structure elements (α helix, β sheet, and 3_{10} helix). We used

a crystal structure from the Protein Data Bank (ID: 1UBQ) obtained with a resolution of 1.8 Å by Vijay-Kumar et al. in 1987.⁵¹ In contrast to the TC5b structure, no geometry optimization was carried out due to the size of the protein. The structure is presented in Figure 4.

4. RESULTS AND DISCUSSION

To assess the accuracy of the LoCOSMO approach, we chose three specific properties for comparison with a regular COSMO calculation. The first one is the electrostatic part of the COSMO solvation energy, which is directly dependent on the surface charge magnitudes, according to eq 4. In addition, we examined two properties more closely related to the molecular system in the solvent cavity. These were chosen to be the magnitudes of the fragment dipole moments as well as the total electron density on a grid.

For the first two quantities, we always calculate the relative errors with respect to the reference results and construct a root-

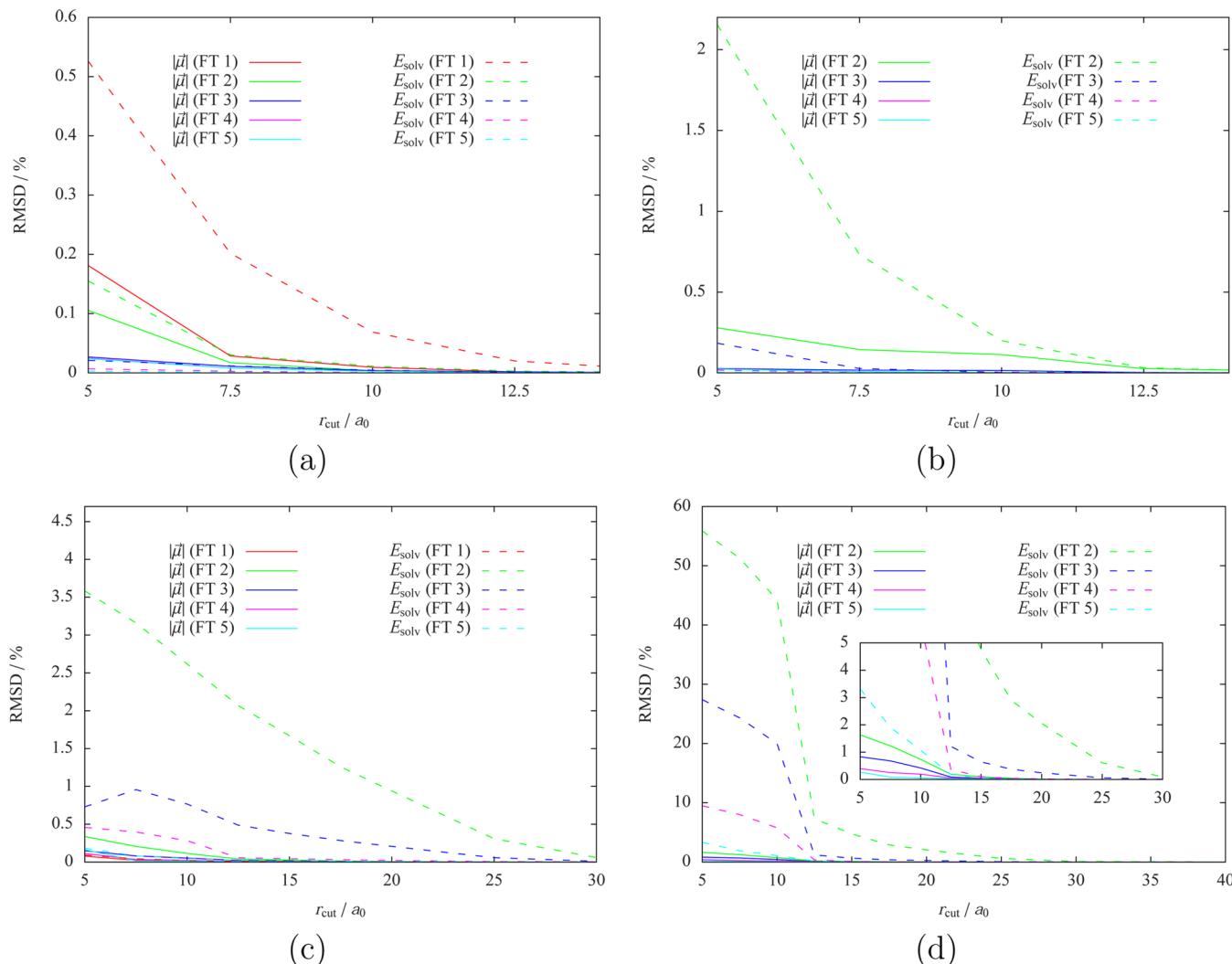


Figure 5. RMSD in dipole moment magnitude ($|\vec{\mu}|$) and solvation energy (E_{solv}) over all fragments with different cutoff radii for TCSb: (a) mode I, (b) mode II, (c) mode III, and (d) mode IV.

mean-square deviation (RMSD) over all (N_{frag}) fragments of the whole system

$$\Delta E_{\text{solv},\text{rel}}^{\text{RMS}} = \frac{1}{N_{\text{frag}}} \sqrt{\sum_A^K \left(\frac{E_{\text{solv},A}^{\text{ref}} - E_{\text{solv},A}^{\text{approx}}}{E_{\text{solv},A}^{\text{ref}}} \right)^2} \quad (15)$$

$$\Delta |\vec{\mu}|_{\text{rel}}^{\text{RMS}} = \frac{1}{N_{\text{frag}}} \sqrt{\sum_A^K \left(\frac{|\vec{\mu}|_A^{\text{ref}} - |\vec{\mu}|_A^{\text{approx}}}{|\vec{\mu}|_A^{\text{ref}}} \right)^2} \quad (16)$$

This facilitates a more rigorous evaluation of the errors compared to an assessment of the total solvation energy or dipole moment, where deviations for individual fragments might cancel each other. A technical problem could arise in the case of a vanishing reference value, as the denominator in eqs 15 and 16 would go to zero. In the case of our example systems, though, this situation is most unlikely. Furthermore, because many of the amino acids are charged, the dipole moments of the fragments are, of course, origin-dependent. However, since the reference calculations were carried out with exactly the same coordinates, this effect is identical in the reference and test calculations. To assess the effect of our algorithm on the direction of the dipole moment, we also calculated the angle

between the reference vector and the one obtained at different stages of our procedure. The resulting angles were never larger than 1.5° , except for the difficult case of mode IV (see Section 4.1.2), where angles of up to 2.0° were obtained. Therefore, we focus only on the magnitude of the dipole moment in the following sections.

Following earlier tests of the 3-FDE method,^{15,16} the density differences are expressed as integrated root-mean-square errors and normalized to the total number of electrons $N_{\text{el}}^{\text{tot}}$

$$\Delta \rho^{\text{RMS}} = \frac{1}{N_{\text{el}}^{\text{tot}}} \sqrt{\int (\rho^{\text{ref}}(\vec{r}) - \rho^{\text{approx}}(\vec{r}))^2 d\vec{r}} \quad (17)$$

Furthermore, we also use the RMSD over the discrete ASC distribution to directly assess the quality of the obtained charge magnitudes

$$\Delta q^{\text{RMS}} = \sqrt{\frac{1}{n_{\text{tot}}^{\text{tot}}} \sum_{\mu} (q_{\mu}^{\text{ref}} - q_{\mu}^{\text{approx}})^2} \quad (18)$$

We would like to emphasize that, unless specifically mentioned, we do not compare the results in this work to supermolecular calculations. Since the focus is only on the performance of the LoCOSMO scheme, the reference is usually a 3-FDE

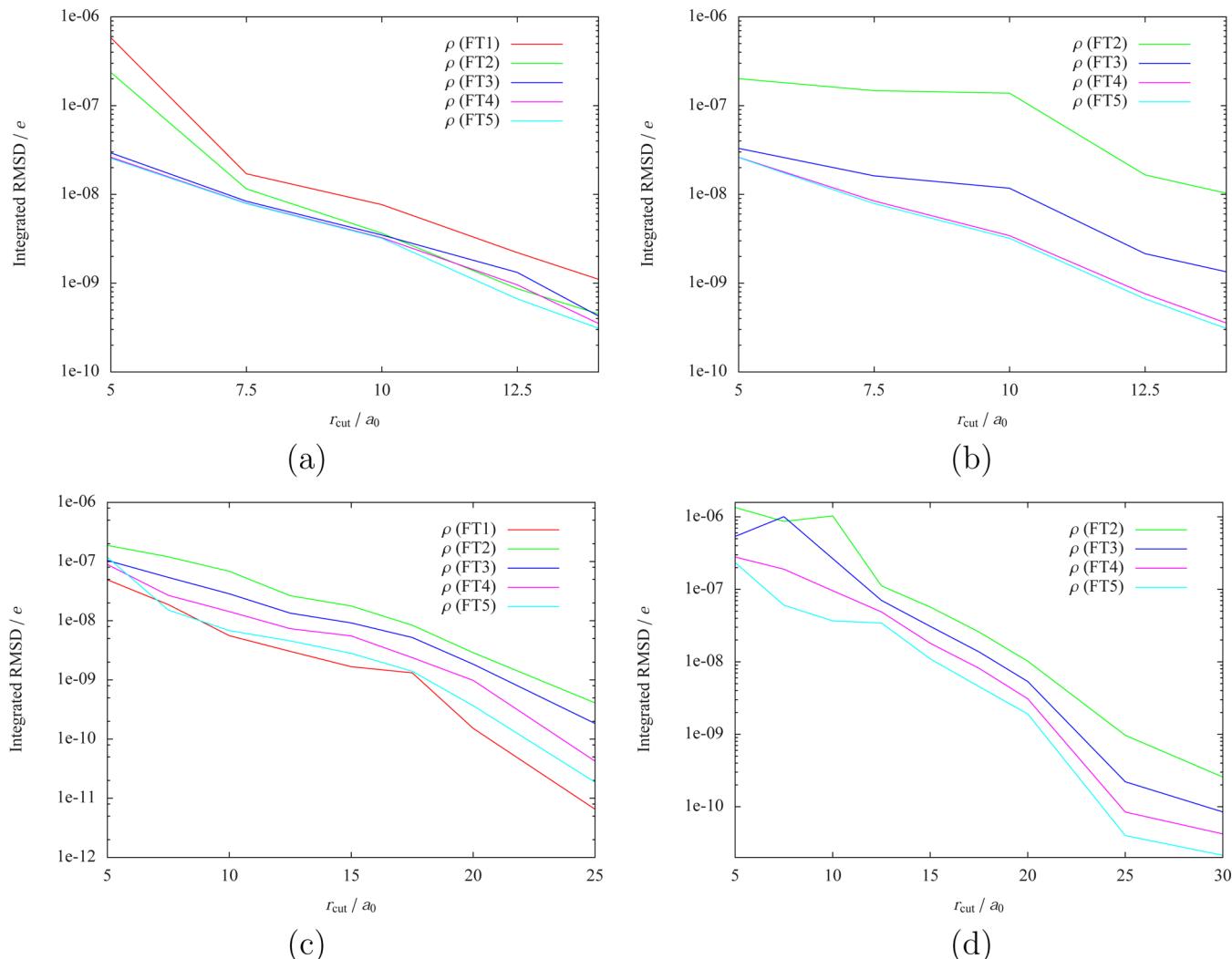


Figure 6. Integrated RMSD in the electron density of TCSb for different cutoff radii (logarithmic plot): (a) mode I, (b) mode II, (c) mode III, and (d) mode IV.

calculation carried out with the same density update scheme (sequential/parallel) but using the regular COSMO algorithm.

4.1. TC5b. 4.1.1. Sequential FT. In Figure 5a, the errors in solvation energy and dipole moment are plotted for different cutoff radii in a LoCOSMO calculation of mode I (sequential FT/precalc) for five consecutive FT cycles. It is apparent that the LoCOSMO scheme works very well in this case because, even for the smallest cutoff radius of $5a_0$, the errors for both quantities are in the subpercent area even in the first FT cycle. It is observed that the dipole moments of the individual fragments are less sensitive than the solvation energies. As expected, the errors are reduced even further through subsequent FT cycles, as the density, and thus also the surface charge distribution, converges.

The integrated deviation of the electron density is plotted in Figure 6a in logarithmic form. When the cutoff is increased, the error steadily decreases from 10^{-6} e downward, reaching magnitudes of below 10^{-9} e for larger values. To compare the error introduced by the LoCOSMO algorithm to the one generated by the usage of the 3-FDE scheme itself, we also computed the difference between a supermolecular density and the 3-FDE density achieved after five FT cycles (sequential FT) with regular COSMO. The obtained value of 2.2×10^{-4} e is 2

orders of magnitude larger than even the LoCOSMO error obtained with the smallest cutoff. The 3-FDE error is also in the same range as the ones observed by Kiewisch et al. in their assessment of the 3-FDE method.¹⁶ Hence, we can conclude that the use of the LoCOSMO method introduces only a negligible additional error.

The results for mode II are depicted in Figure 5b. Since this mode features the gradual update scheme, tremendous errors are observed in the first FT cycle (not plotted). This behavior is self-evident, since the calculations are started from a uniform ASC distribution that does not resemble the actual density in any way. Already in the second FT cycle, however, the updated charge distribution is accurate enough to produce errors of around 2% even for the smallest cutoff, and a value of $7.5a_0$ is sufficient to reach subpercent accuracy.

The integrated density errors in mode II behave similarly compared to those in mode I, as depicted in Figure 6b. Even for the smallest cutoff in the first production cycle the values are below 10^{-6} e and decrease with the number of FT cycles as well as with increasing cutoff radius, although convergence is somewhat slower than it is in mode I.

A graphical representation of the observed density errors for one example ($r_{\text{cut}} = 10a_0$) in modes I and II is given in Figure 7.

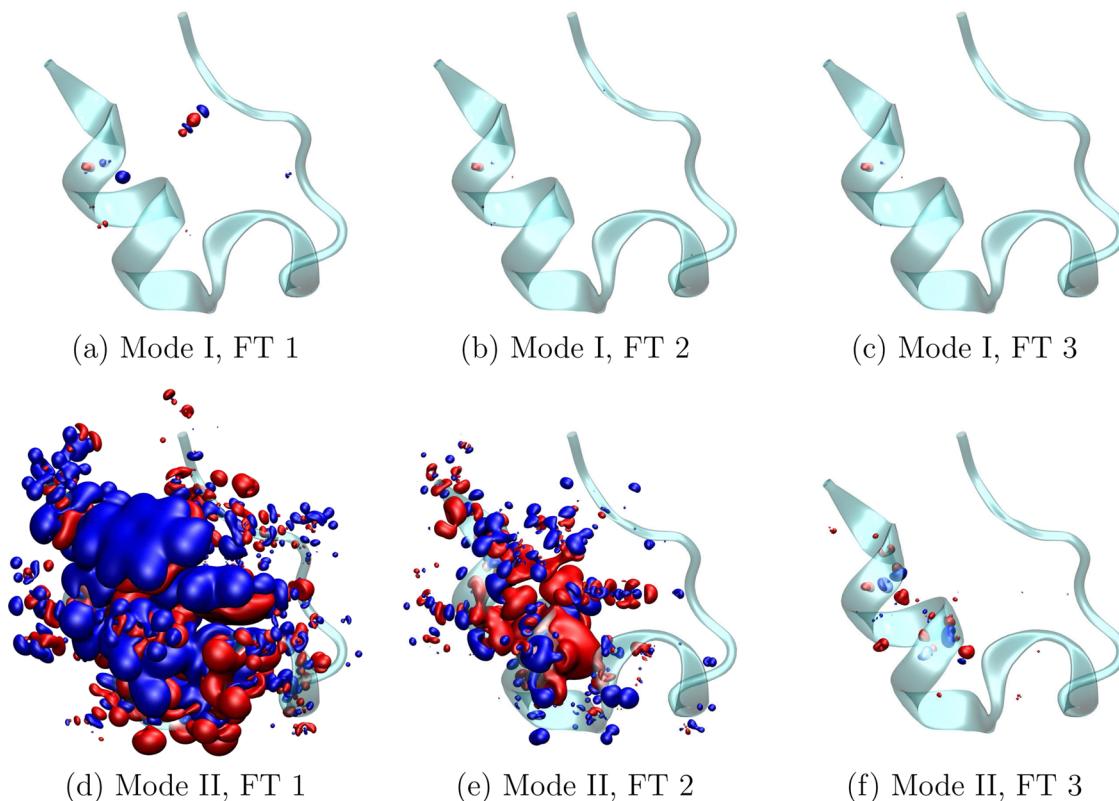


Figure 7. Isosurface representation of density errors for the first three FT cycles of LoCOSMO calculations in modes I and II on TCSb ($r_{\text{cut}} = 10a_0$; isovalue, 10^{-6}). All graphics were generated with VMD.⁵²

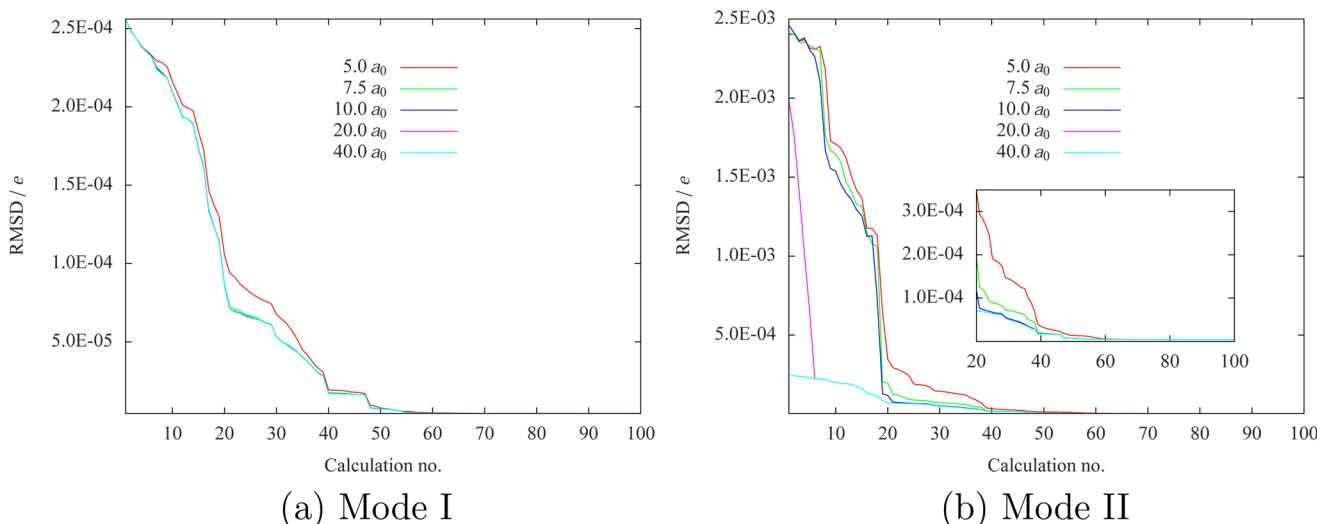


Figure 8. RMSD in ASC magnitudes over the course of five FT cycles for TCSb with different cutoff radii and calculation modes.

It can be seen how the errors in mode II gradually approach the ones resulting from a mode I calculation. Note that even though the errors in the first two FT cycles appear to be extremely large due to the chosen isovalue they are still at least an order of magnitude below the ones introduced by the underlying subsystem approach.

To directly observe the effect of our algorithm on the ASC distribution, the RMSD in the charge magnitudes (see eq 18) over the course of five FT cycles (20 calculations per cycle) is plotted for different cutoff radii in Figure 8. The reference distribution was the one obtained from the full COSMO

calculation after five sequential FT cycles. It can be seen that the distribution converges rapidly for both modes and that a residual error of only $4.3 \times 10^{-6} e$ remains. The latter can be attributed to numerical noise, as it is even obtained with the largest cutoff, where all charges are active.

4.1.2. Parallel FT. The situation is much more difficult for parallel FT calculations. This is due to the fact that after every complete FT cycle average magnitudes have to be generated for all charges that are active in more than one fragment calculation. For small cutoff radii, this seems to lead to discontinuities, which can add up over the subsequent cycles,

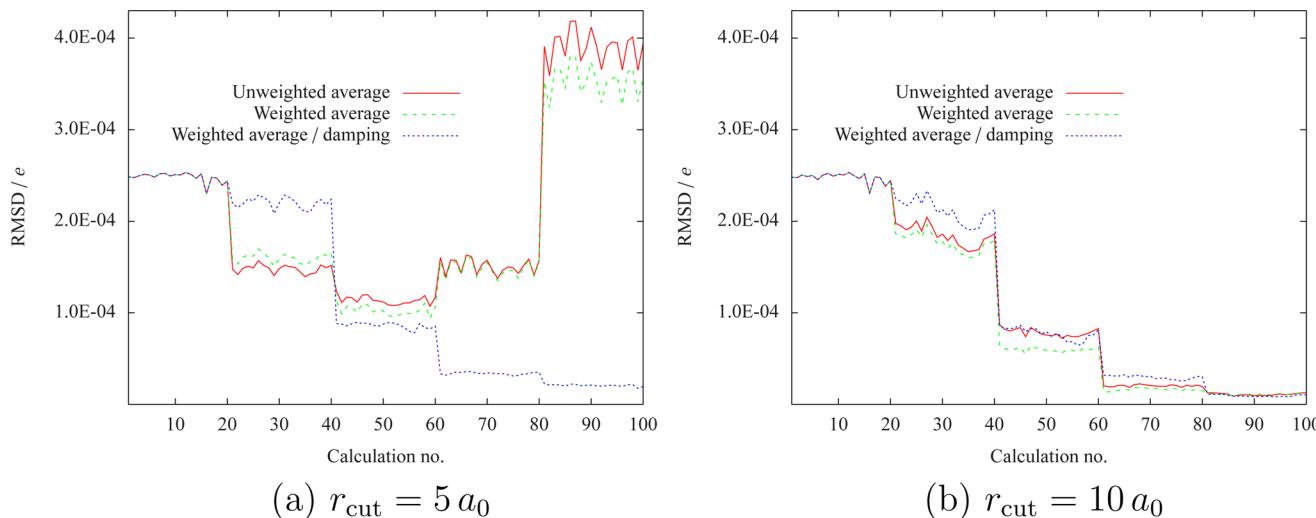


Figure 9. RMSD in ASC magnitudes over the course of five FT cycles for TC5b. Three different averaging schemes (unweighted, weighted by closest atom distance, and the latter in conjunction with damping) were used together with mode III.

resulting in a completely distorted ASC distribution. In Figure 9a, we plot the RMSD of all charge magnitudes for a calculation in mode III with a cutoff radius of $5a_0$. Here, the reference is the final charge distribution resulting from five FT cycles in a parallel FT calculation with all charges being active. The averaging procedure was employed after each complete FT cycle. The solid line was created with an unweighted averaging procedure, i.e., a simple arithmetic mean over all obtained values for each charge.

Even though the first FT cycle is initialized with a precalculated ASC distribution, it is evident that the results only worsen after the third FT cycle and go completely awry starting from the fifth. The situation slightly improves if we use a weighted average, where the distances to the closest atom of each associated fragment are used as the weighting criterion (dashed line). However, the qualitative features remain the same. Only when we introduce a damping factor that includes the previous charge distribution with a certain weight can the situation be controlled (dotted line). In the plot, we applied a damping factor of $0.7^{n_{\text{FT}}}$, where n_{FT} is the number of completed FT cycles (i.e., the damping decreases exponentially over the course of the whole calculation).

For larger cutoff radii, the size of the overlapping areas eliminates the problem. Already with a value of $10a_0$, no damping is required, as shown in Figure 9b. For TC5b, this corresponds to an average of about 30% of the charges being active. This might lead to the conclusion that it is not the increasing overlap but the large total percentage of active charges that is responsible for curing the problem. However, tests with the much larger ubiquitin show that the same cutoff radius is sufficient to eliminate the need for damping. Since with the same value only 9% of all charges are active in that case, we conclude that the reason is indeed the larger overlap of adjacent regions, not the total amount of active charges.

The property results for modes III and IV with different cutoff radii are presented in Figure 5, panels c and d, respectively. We used the mentioned damping scheme for all calculations with r_{cut} of up to $10a_0$, whereas no damping was applied for larger values.

It can clearly be seen that even with damping for small cutoff radii the property results are not nearly as accurate as those for the sequential FT calculations. In mode III, small errors are

observed for the first FT cycle, but, due to the damping procedure, the convergence is not as systematic as before. For instance, the solvation energy is much worse in the second FT cycle compared to that in the first one.

With mode IV, the behavior is even more irregular. Only with five FT cycles and a cutoff radius greater than $10a_0$ is it possible to reach subpercent accuracy for the solvation energy. A more sophisticated damping scheme might lead to a better balance between convergence speed and reliability. However, due to the fact that, in general, the sequential FT scheme is preferred anyway, we did not investigate this any further.

The trends are similar for the errors in the electron density, as depicted in Figure 6, panels c and d. The absolute values for the differences are in a similar range as the ones obtained with the sequential FT scheme. The curves also exhibit somewhat unsystematic convergence behavior with respect to the number of FT cycles, but they decrease steadily when the cutoff radius is enlarged.

As plotted in Figure 9, the RMSD in the ASC magnitudes also converges nicely with the damping scheme, although for very small cutoff radii ($5a_0$) errors of up to $2.0 \times 10^{-5} e$ can remain.

In conclusion, the results show that the LoCOSMO method works extremely well for TC5b in the case of the sequential FT scheme. In contrast, if parallel FT is used, then either the chosen cutoff radius has to be somewhat larger or a damping scheme has to be included. However, the latter introduces new degrees of freedom pertaining to the value of the damping factor in each cycle and can lead to less systematic convergence behavior.

4.1.3. Timings. The total time spent in COSMO routines for TC5b is plotted for all four modes in Figure 10. All timings are summed over five consecutive FT cycles and given in CPU hours. The test calculations were carried out on a single core of a 2.5 GHz Intel Xeon CPU. The timings were measured only for all subroutines pertaining to the solvation model. To guarantee a fair comparison, however, for the pre-calc modes (I and III) the total time required for preparing the initial ASC distribution (including density preparation processes) has been added, as this step is not necessary in a regular COSMO calculation.

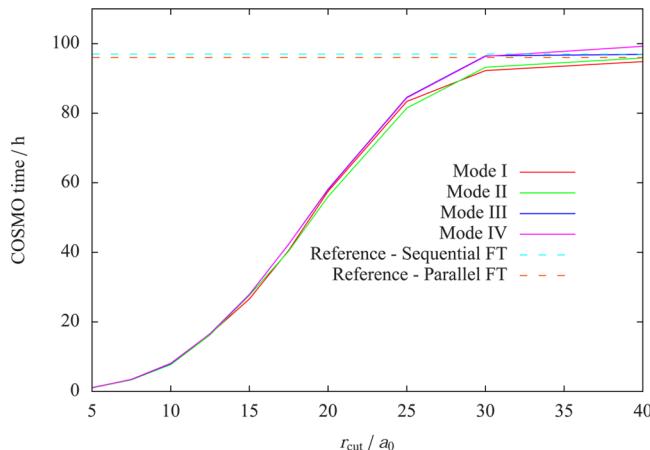


Figure 10. CPU hours spent in COSMO routines for TC5b during five FT cycles with different cutoff radii and calculation modes. The dashed lines represent full COSMO calculations.

It is evident at first glance that the LoCOSMO scheme can easily save more than 90% of the required computational time in the solvent routines if small cutoff radii are used. Naturally, when the cutoff radius is increased, the required time approaches the one for the full calculation. The small overshoot at radii above $30a_0$ can be explained by the fact that almost all charges are included in the active subset but that the determination of active charges (and, in the case of modes I and III, also the calculation of the initial ASC distribution) creates a slight overhead. Furthermore, due to the technical problems observed in modes II and IV, a slight increase in the number of COSMO iterations can change the total duration. Nonetheless, for smaller values of r_{cut} the savings clearly outweigh the cost of the additional operations.

4.2. Ubiquitin. Due to the problems with parallel FT calculations (see Section 4.1.2), we restricted our tests to modes I and II (i.e., the ones using the sequential FT scheme) for ubiquitin. Apart from that, the same tests as in Section 4.1 were performed.

4.2.1. Sequential FT. The results of the property calculations are presented in Figure 11, panels a and b for modes I and II,

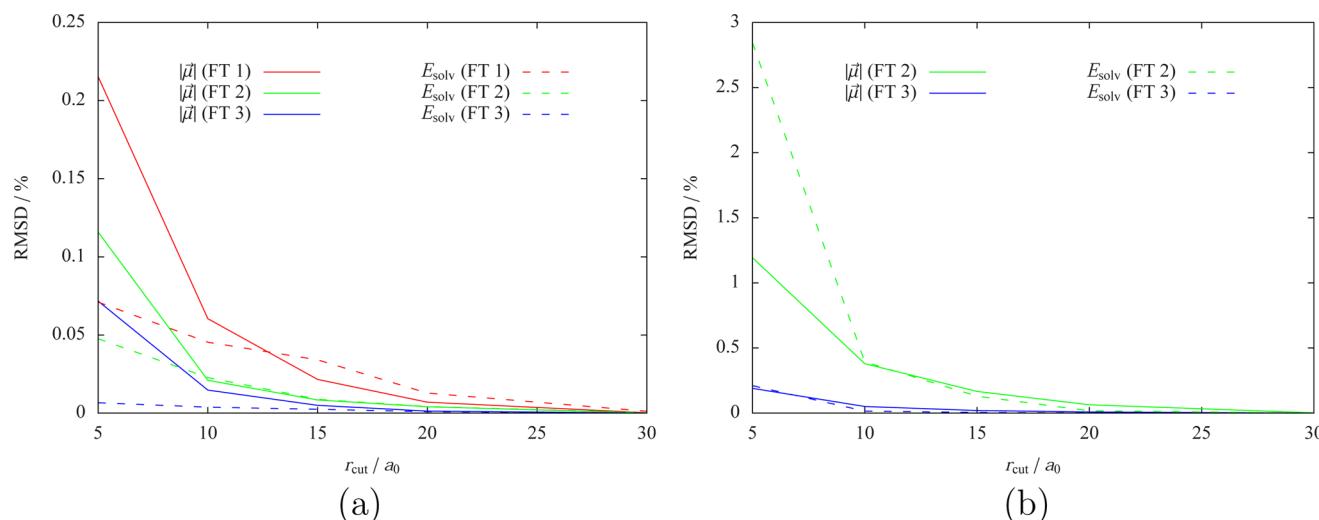


Figure 11. RMSD in dipole moment magnitude ($|\vec{\mu}|$) and solvation energy (E_{solv}) over all fragments with different cutoff radii for ubiquitin: (a) mode I and (b) mode II.

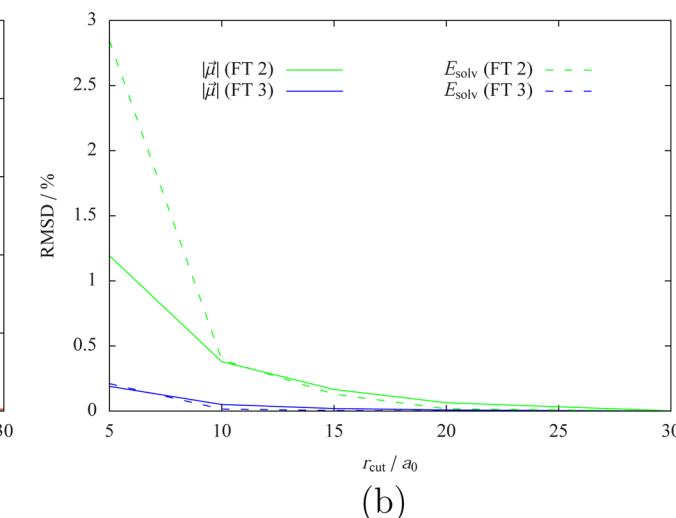
respectively. It can be directly seen that the situation is very similar to the calculations on the smaller TC5b protein. If a precalculated charge distribution is used, then even the smallest cutoff radius of $5.0a_0$ produces errors of less than 1% already in the first FT cycle. In mode II, the errors for the same cutoff radius are still on the order of a few percent in the first production cycle, but they drop below 1% when a value of $10a_0$ is used. Again, additional FT cycles bring the results even closer to the reference calculations.

For the electron density, some unsystematic behavior is observed for small cutoff radii, where the error increases from the second to the third FT cycle in mode I (Figure 12a). However, the absolute errors are still at least an order of magnitude lower than the method-inherent density errors of 3-FDE reported by Kiewisch et al. for a similar system (neutral ubiquitin).¹⁶ Akin to the TC5b results, the errors decrease roughly exponentially when the cutoff radius is enhanced. The trends are similar for mode II, as shown in Figure 12b.

The RMSD in the ASC magnitudes exhibits similar convergence behavior as that for TC5b (not shown) and also reaches values below $10^{-5} e$ even for moderate values of r_{cut} .

4.2.2. Timings. The timings for ubiquitin are presented in Figure 13. In contrast to the calculations for TC5b, the values were obtained on all 12 cores of the same 2.5 GHz Intel Xeon machines. Naturally, the percentage of time saved is much larger due to the size of the system. Even with a cutoff radius of $15a_0$, only 3.5% of the COSMO time for the reference calculation is required. Furthermore, for this system, the time required to construct the initial ASC distribution in mode I is negligible compared to the total time.

In terms of memory requirements, a similar gain is observed. Due to the quadratic scaling of the matrix A with the number of ASCs, see eq 5, even with a reasonable cutoff radius of $10a_0$ at most 2% of the memory required for the full matrix has to be allocated in any given fragment calculation. Although the absolute numbers are still modest for ubiquitin (33 MB and 2 GB, respectively), calculations on even larger systems might only become feasible due to this reduction. Furthermore, an additional speedup could be achieved by keeping the reduced matrix in memory instead of recalculating it in every SCF iteration.



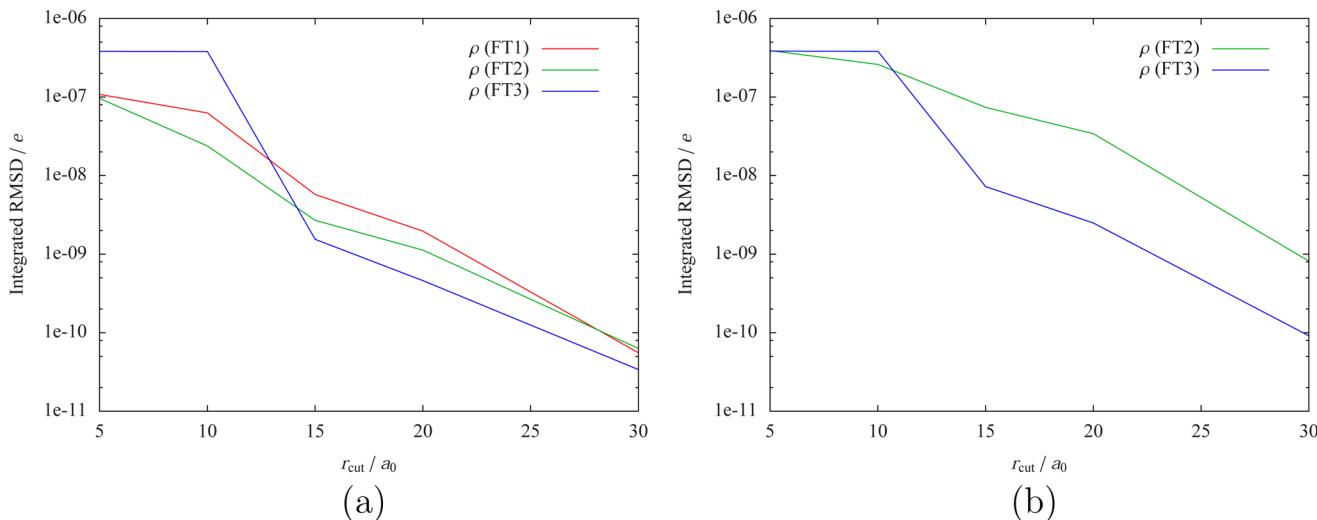


Figure 12. Integrated RMSD in the electron density of ubiquitin for different cutoff radii (logarithmic plot): (a) mode I and (b) mode II.

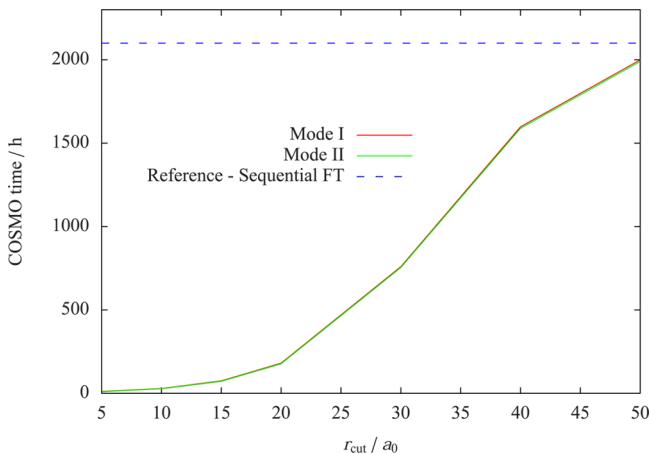


Figure 13. CPU hours spent in COSMO routines for ubiquitin during three FT cycles with different cutoff radii and calculation modes. The dashed line represents a full COSMO calculation.

In any case, it is expected that for ever larger systems the savings will become even more pronounced, as no larger cutoff is required, even if the molecular size is increased.

5. SUMMARY AND CONCLUSIONS

We have presented an alternative to the standard COSMO scheme, which can be used in conjunction with subsystem methods such as frozen density embedding. Our method exploits the fact that only the magnitudes of surface charges close to the active fragment have to be taken into account when minimizing the solvation energy. In our implementation, we determine the active charges via a cutoff radius based on the distance to the atoms of the currently active fragment.

An important choice concerns the density update scheme. When the so-called parallel FT scheme is used, severe difficulties arise due to the fact that we use overlapping subsets of active charges in different fragment calculations. Thus, after a complete FT cycle, several different values can exist for each charge and an averaging procedure is required. We found that with small cutoff radii (usually below $10 a_0$) a damping scheme has to be employed to keep discontinuities from building. This, in turn, requires at least one additional parameter (damping factor) and can also slow the overall convergence of the charge

distribution and electron density. This problem does not arise with the sequential FT scheme, where there is only one possibly discontinuous interface in every calculation instead of many. We therefore recommend the use of the latter scheme if massive parallelization over fragments is not essential.

Two methods for the initialization of the inactive charges have been explored. One of them depends on a single full COSMO calculation to generate charges adapted to the initial density (pre-calc), whereas the other one gradually builds the adapted charge distribution over the course of the first FT cycle (build-up). The advantage of the latter method is that the full A matrix (see eq 5) never has to be constructed and inverted. However, in this case, at least one additional production pass is required after the first FT cycle. If the molecular system is still small enough for a single calculation of the whole matrix to be fast compared to that for a whole (reduced) FT cycle, then we recommend to use the pre-calc scheme.

The solvation energy and fragment dipole moments converge very fast with respect to the cutoff radius. When using the recommended sequential FT scheme, a cutoff radius of $10 a_0$ is sufficient to generate errors of less than 1% in the first production cycle. The same holds for the error in the electron density, which turns out to be several orders of magnitude lower than the one already introduced by the 3-FDE approximation. The corresponding speedups depend on the size of the investigated system, but for the comparably small TCSb, 90% of the total COSMO time can already be saved. Naturally, for larger systems, the speedup will be even more favorable. The parallel FT scheme offers a slightly smaller gain due to the problems mentioned above. However, even if a somewhat larger cutoff has to be chosen to ensure convergence, the computational time can still decrease tremendously for larger systems.

In conclusion, the LoCOSMO scheme presents a straightforward way to drastically reduce the required time to compute the solvent influence for large systems described by a fragment-based method while retaining almost the same accuracy as that in a full calculation. This will pave the way for subsystem-based electronic-structure calculations on very large systems in solution.

AUTHOR INFORMATION

Corresponding Author

*E-mail: j.neugebauer@uni-muenster.de.

Funding

This work was supported by a VIDI grant (700.59.422) of The Netherlands Organisation for Scientific Research (NWO).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors would like to thank Lucas Visscher and Christoph Jacob for stimulating discussions.

REFERENCES

- (1) Akimov, A. V.; Prezhdo, O. V. Large-Scale Computations in Chemistry: A Bird's Eye View of a Vibrant Field. *Chem. Rev.* **2015**, *115*, 5797–5890.
- (2) van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. Biomolecular modeling: goals, problems, perspectives. *Angew. Chem., Int. Ed.* **2006**, *45*, 4064–4092.
- (3) Neugebauer, J. Subsystem-Based Theoretical Spectroscopy of Biomolecules and Biomolecular Assemblies. *ChemPhysChem* **2009**, *10*, 3148–3173.
- (4) König, C.; Neugebauer, J. Quantum Chemical Description of Absorption Properties and Excited-State Processes in Photosynthetic Systems. *ChemPhysChem* **2012**, *13*, 386–425.
- (5) Senatore, G.; Subbaswamy, K. R. Density dependence of the dielectric constant of rare-gas crystals. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *34*, 5754–5757.
- (6) Cortona, P. Self-consistently determined properties of solids without band-structure calculations. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1991**, *44*, 8454–8458.
- (7) Wesolowski, T. A.; Warshel, A. Frozen Density Functional Approach for ab Initio Calculations of Solvated Molecules. *J. Phys. Chem.* **1993**, *97*, 8050–8053.
- (8) Jacob, C. R.; Neugebauer, J. Subsystem density-functional theory. *WIREs Comput. Mol. Sci.* **2014**, *4*, 325–362.
- (9) Wesolowski, T. A.; Shedge, S.; Zhou, X. Frozen-Density Embedding Strategy for Multilevel Simulations of Electronic Structure. *Chem. Rev.* **2015**, *115*, 5891–5928.
- (10) König, C.; Neugebauer, J. First-principles calculation of electronic spectra of light-harvesting complex II. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10475–10490.
- (11) König, C.; Neugebauer, J. Protein Effects on the Optical Spectrum of the Fenna-Matthews-Olson Complex from Fully Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2013**, *9*, 1808–1820.
- (12) Goez, A.; Jacob, C. R.; Neugebauer, J. Modeling environment effects on pigment site energies: Frozen density embedding with fully quantum-chemical protein densities. *Comput. Theor. Chem.* **2014**, *1040-1041*, 347–359.
- (13) Jacob, C. R.; Visscher, L. Towards the Description of Covalent Bonds in Subsystem Density-Functional Theory. In *Recent Progress in Orbital-Free Density Functional Theory*, 1st ed.; Wesolowski, T. A., Wang, Y. A., Eds.; World Scientific Publishing: Hoboken, NJ, 2013; pp 297–322.
- (14) Casida, M. E.; Wesolowski, T. A. Generalization of Kohn-Sham Equations with Constrained Electron Density Formalism and Its Time-Dependent Response Theory Formulation. *Int. J. Quantum Chem.* **2004**, *96*, 577–588.
- (15) Jacob, C. R.; Visscher, L. A subsystem density-functional theory approach for the quantum chemical treatment of proteins. *J. Chem. Phys.* **2008**, *128*, 155102.
- (16) Kiewisch, K.; Jacob, C. R.; Visscher, L. Quantum-Chemical Electron Densities of Proteins and of Selected Protein Sites from Subsystem Density Functional Theory. *J. Chem. Theory Comput.* **2013**, *9*, 2425–2440.
- (17) 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.
- (18) Tronrud, D. E.; Wen, J.; Gay, L.; Blankenship, R. E. The structural basis for the difference in absorbance spectra for the FMO antenna protein from various green sulfur bacteria. *Photosynth. Res.* **2009**, *100*, 79–87.
- (19) Beglov, D.; Roux, B. An Integral Equation To Describe the Solvation of Polar Molecules in Liquid Water. *J. Phys. Chem. B* **1997**, *101*, 7821–7826.
- (20) Kovalenko, A.; Hirata, F. Three-dimensional density profiles of water in contact with a solute of arbitrary shape: a RISM approach. *Chem. Phys. Lett.* **1998**, *290*, 237–244.
- (21) Kaminski, J. W.; Gusarov, S.; Wesolowski, T. A.; Kovalenko, A. Modeling Solvatochromic Shifts Using the Orbital-Free Embedding Potential at Statistically Mechanically Averaged Solvent Density. *J. Phys. Chem. A* **2010**, *114*, 6082–6096.
- (22) Shedge, S. V.; Wesolowski, T. A. Nonuniform Continuum Model for Solvatochromism Based on Frozen-Density Embedding Theory. *ChemPhysChem* **2014**, *15*, 3291–3300.
- (23) Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the revision of solvent effects. *Chem. Phys.* **1981**, *55*, 117–129.
- (24) Cammi, R.; Tomasi, J. Remarks on the use of the apparent surface charges (ASC) methods in solvation problems: Iterative versus matrix-inversion procedures and the renormalization of the apparent charges. *J. Comput. Chem.* **1995**, *16*, 1449–1458.
- (25) 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., Perkin Trans. 2* **1993**, *2*, 799–805.
- (26) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99*, 2224–2235.
- (27) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. Refinement and Parametrization of COSMO-RS. *J. Phys. Chem. A* **1998**, *102*, 5074–5085.
- (28) Cancès, E.; Maday, Y.; Stamm, B. Domain decomposition for implicit solvation models. *J. Chem. Phys.* **2013**, *139*, 054111.
- (29) Lipparini, F.; Stamm, B.; Cancès, E.; Maday, Y.; Mennucci, B. Fast Domain Decomposition Algorithm for Continuum Solvation Models: Energy and First Derivatives. *J. Chem. Theory Comput.* **2013**, *9*, 3637–3648.
- (30) Lipparini, F.; Lagardère, L.; Scalmani, G.; Stamm, B.; Cancès, E.; Maday, Y.; Piquemal, J.-P.; Frisch, M. J.; Mennucci, B. Quantum Calculations in Solution for Large to Very Large Molecules: A New Linear Scaling QM/Continuum Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 953–958.
- (31) Lipparini, F.; Scalmani, G.; Lagardère, L.; Stamm, B.; Cancès, E.; Maday, Y.; Piquemal, J.-P.; Frisch, M. J.; Mennucci, B. Quantum, classical, and hybrid QM/MM calculations in solution: General implementation of the ddCOSMO linear scaling strategy. *J. Chem. Phys.* **2014**, *141*, 184108.
- (32) Greengard, L.; Rokhlin, V. A fast algorithm for particle simulations. *J. Comput. Phys.* **1987**, *73*, 325–348.
- (33) Scalmani, G.; Barone, V.; Kudin, K. N.; Pomelli, C. S.; Scuseria, G. E.; Frisch, M. J. Achieving linear-scaling computational cost for the polarizable continuum model of solvation. *Theor. Chem. Acc.* **2004**, *111*, 90–100.
- (34) 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.

- (35) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment molecular orbital method: an approximate computational method for large molecules. *Chem. Phys. Lett.* **1999**, *313*, 701–706.
- (36) Nakano, T.; Kaminuma, T.; Sato, T.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. Fragment molecular orbital method: application to polypeptides. *Chem. Phys. Lett.* **2000**, *318*, 614–618.
- (37) 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.
- (38) Raghavachari, K.; Saha, A. Accurate Composite and Fragment-Based Quantum Chemical Models for Large Molecules. *Chem. Rev.* **2015**, *115*, 5643–5677.
- (39) Pye, C. C.; Ziegler, T. An implementation of the conductor-like screening model of solvation within the Amsterdam density functional package. *Theor. Chem. Acc.* **1999**, *101*, 396–408.
- (40) ADF2014, SCM; Theoretical Chemistry, Vrije Universiteit: Amsterdam, The Netherlands, <http://www.scm.com>.
- (41) 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.
- (42) 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.
- (43) Dresselhaus, T.; Neugebauer, J. Part and whole in wavefunction/DFT embedding. *Theor. Chem. Acc.* **2015**, *134*, 97.
- (44) Jacob, C. R.; Neugebauer, J.; Visscher, L. A flexible implementation of frozen-density embedding for use in multilevel simulations. *J. Comput. Chem.* **2008**, *29*, 1011–1018.
- (45) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098–3100.
- (46) Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33*, 8822–8824.
- (47) Lembarki, A.; Chermette, H. Obtaining a gradient-corrected kinetic-energy functional from the Perdew-Wang exchange functional. *Phys. Rev. A: At., Mol., Opt. Phys.* **1994**, *50*, 5328–5331.
- (48) Wesolowski, T. A.; Chermette, H.; Weber, J. Accuracy of approximate kinetic energy functionals in the model of Kohn-Sham equations with constrained electron density: The FH···NCH complex as a test case. *J. Chem. Phys.* **1996**, *105*, 9182–9190.
- (49) 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.
- (50) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Designing a 20-residue protein. *Nat. Struct. Biol.* **2002**, *9*, 425–430.
- (51) 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.
- (52) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.