Article

Multilevel Density Functional Theory

Gioia Marrazzini,[§] Tommaso Giovannini,*,[§] Marco Scavino, Franco Egidi, Chiara Cappelli, and Henrik Koch*

Cite This: https://dx.doi.org/10.1021/acs.jctc.0c00940

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: Following recent developments in multilevel embedding methods, we introduce a novel density matrix-based multilevel approach within the framework of density functional theory (DFT). In this multilevel DFT, the system is partitioned in an active and an inactive fragment, and all interactions are retained between the two parts. The decomposition of the total system is performed upon the density matrix. The orthogonality between the two parts is maintained by solving the Kohn−Sham equations in the MO basis for the active part only, while keeping the inactive density matrix frozen. This results in the reduction of computational cost. We outline the theory and implementation and discuss the differences and similarities with state-of-the-art DFT embedding methods. We present applications to aqueous solutions of methyloxirane and glycidol.

$\rho_{total} = \rho_{active} + \rho_{inactive}$

Multilevel DFT - full DFT
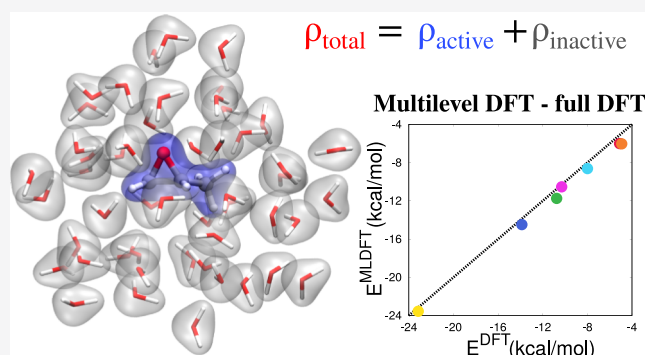
1. INTRODUCTION

The study of the energetics and physico-chemical properties of large molecular systems is one of the most challenging problems in quantum chemistry.[1] Many processes of chemical interest take place in solution,[2−6] in biological matrices,[7,8] or at the interfaces between different materials.[9−11] The large size of such systems poses theoretical and computational challenges because high-level correlated electronic structure methods are usually unfeasible because of their high computational cost and unfavorable scaling.[12,13] A good compromise between accuracy and computational cost is provided by density functional theory (DFT),[14,15] which accounts for electron correlation in an approximate way. Because of the proven reliability of the results that can be obtained at the DFT level, it has become the most widely used approach for describing the electronic structure of large systems.

DFT permits the investigation of much larger systems than offered by highly correlated methods. However, it cannot be routinely applied to systems constituted by more than 500 atoms, unless implementations through graphical processing units are exploited.[16] The practical limit of 500 atoms makes applications of DFT to biological matrices, interfaces, and solutions particularly cumbersome and, in some cases, even impossible. For these reasons, several approximations have been developed in the past. The nature of any approximation aimed at reducing the computational cost of a calculation is often rooted in the information that one seeks to extract from a simulation. If one is interested in the interaction between two complex medium-sized subsystems and how they influence each other, then, one might seek a partition method which accurately models both moieties with a computational cost comparable to treating both systems separately. More often than not, however, one is more interested in modeling only a small part of a larger chemical system, while the rest is only relevant in as much as it can alter the physical and chemical properties of the smaller region. This is the case of systems embedded in an external environment, whose properties are modified by the latter, the archetypal example being a molecule dissolved in a solvent. For these problems, the system may be partitioned into two or more regions which are then treated at different levels of accuracy, and different approaches may be more or less suitable depending on the specificities of the system/environment. In the special case of systems in solution, particular success has been enjoyed by methods belonging to the family of the so-called focused models,[17−20] which are extremely useful when dealing with the property of a moiety or a chromophore embedded in an external environment.

In focused models, the target molecule is described at a higher level of theory with respect to the environment, which acts as a perturbation on the target system. Among the different focused models that have been developed in the past, the large majority belongs to the family of quantum mechanics (QM)/classical approaches, in which the target is treated at the

QM level. The environment is instead described classically, by means of either continuous descriptions, such as the polarizable continuum model,[17,18] or by retaining its atomistic nature in the so-called QM/molecular mechanics (QM/MM) approaches.[21−24] In all these methods, however, the interaction between the two parts of the whole system is usually described by classical electrostatics[20,25−27] and very rarely by including the interactions of the quantum nature, such as Pauli repulsion and dispersion.[28−30] Also, QM/classical methods allow for the treatment of very large systems; however, their accuracy crucially depends on the quality of the parametrization of the classical fragments. In order to avoid such a variability, quantum embedding methods can be exploited.[31−48] In these approaches, the whole system is treated by resorting to a QM description; thus, Pauli repulsion effects are introduced in the modeling. The reduction in the computational cost is then achieved by partitioning the system in at least one active and one inactive part. The former is described at a higher level of accuracy, whereas the latter may be kept frozen or described at a lower level of theory. Different approaches have been proposed in the past, ranging from projection-based methods, such as DFT-in-DFT or HF-in-DFT,[39,46,47,49,50] to frozen density embedding (FDE).[48,51−57] Quantum embeddings have become a popular choice to overcome the aforementioned shortcomings of QM/classical approaches; however, the partitioning of the electronic degrees of freedom for the whole system into the two layers is theoretically far from trivial. In fact, all such partitioning methods can be thought of as being based upon the same basic idea. However, the partition of the electronic degrees of freedom may be performed on molecular orbitals, the density function, or the density matrix. Which basis set must be used to expand the density and wavefunction of the different regions is also an important factor. All these choices lead to drastically different methods which have their unique advantages and complications (*vide infra*). In addition, all quantum-embedding approaches based on DFT suffer from the complication of the exchange−correlation functionals being nonlinear, which leads to nonadditive energy terms in the partitioned equations.

In this paper, we are proposing a novel quantum embedding approach defined in a DFT framework. We denote this method multilevel DFT (MLDFT), due to its similarity with multilevel Hartree−Fock (MLHF),[58,59] and multilevel coupled-cluster[12,60] that we have recently developed. Similar to FDE, in MLDFT, the system is partitioned into an active and an inactive region, and all the interactions are retained between the two parts. The MLDFT conceptually differs from the aforementioned quantum embedding methods because the partitioning is performed on the density matrix instead of the density function as in FDE.[61] As a consequence, the nonadditive term in MLDFT does not contain any kinetic energy term, thus avoiding the theoretical complications present in FDE (*vide infra*). Also, the MLDFT equations are defined in the MO basis of the active fragment only. This feature automatically allows for savings in the computational cost because the inactive MOs are not involved in the self-consistent field (SCF) procedure. In this paper, we derive MLDFT, and we apply the method to ground state (GS) energies of aqueous solutes. The results are compared, in all cases, to full DFT, in order to assess the quality of the multilevel partition.

The manuscript is organized as follows. In the next section, DFT theory is formulated in the MO basis, and MLDFT equations are presented with a particular focus on the computational savings that can be expected. Theoretical comparison with other quantum embedding approaches is also discussed. Then, after a brief section reporting on the computational details of the method, MLDFT is applied to selected aqueous systems, with emphasis on comparison with full DFT results. Conclusions and perspectives of the present work end the manuscript.

2. THEORY

Our starting point is the DFT expression for the electronic energy of the system

$$
\begin{aligned}
E[\mathbf{D}] &= \mathrm{Tr}\,\mathbf{hD} + \frac{1}{2}\mathrm{Tr}\,\mathbf{DJ}(\mathbf{D}) - \frac{1}{2}c_x\mathrm{Tr}\,\mathbf{DK}(\mathbf{D}) \\
&\quad + (1 - c_x)E_x[\mathbf{D}] + E_c[\mathbf{D}] \\
&= \mathrm{Tr}\,\mathbf{hD} + \frac{1}{2}\mathrm{Tr}\,\mathbf{DJ}(\mathbf{D}) - \frac{1}{2}c_x\mathrm{Tr}\,\mathbf{DK}(\mathbf{D}) \\
&\quad + (1 - c_x)\int \rho(\mathbf{r})\varepsilon_x(\rho(\mathbf{r}))\mathrm{d}\mathbf{r} \\
&\quad + \int \rho(\mathbf{r})\varepsilon_c(\rho(\mathbf{r}))\mathrm{d}\mathbf{r}
\end{aligned}
\tag{1}
$$

$$
\rho(\mathbf{r}) = \sum_{\mu\nu} D_{\mu\nu}\chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r})
\tag{2}
$$

Here, $\mathbf{D}$ is the one-particle density matrix expanded in the atomic orbital (AO) basis $\{\chi_\mu\}$, and $\mathbf{h}$ is the one-electron operator, whereas $\mathbf{J}$ and $\mathbf{K}$ are Coulomb and exchange matrices, respectively. The E_x and E_c terms are DFT exchange and correlation energy functionals; $\rho(\mathbf{r})$ is the DFT density function, and ε_x and ε_c are the exchange and correlation energy densities per unit particle, respectively. The coefficient c_x defines whether pure DFT ($c_x = 0$) or hybrid DFT functionals ($c_x \neq 0$) is used. Note that in the second equality in eq 1, it is implied that we are restricting the treatment to semilocal possibly hybrid functionals that are commonly used in quantum chemistry.

The energy defined in eq 1 is usually minimized in the AO basis. It is also possible to reformulate the minimization in the MO basis, which is the strategy employed in this work, similarly to what was presented for the Hartree−Fock case by Sæther et al.[58] This can be accomplished by parametrizing the density matrix $\mathbf{D}$ in terms of an antisymmetric rotation matrix, in which only the nonredundant occupied-virtual rotations are considered.[58]

2.1. Multilevel DFT. The MLDFT method belongs to the family of the so-called focused models. The part of the system which is under investigation (active) is described accurately, whereas the remaining (inactive) part remains frozen during the optimization of the active fragment. The choice of the partitioning intimately depends on the specificities of the system, its chemical nature, and the properties one wishes to simulate. Like other QM/QM methods, the details of the partitioning largely define the nature of the method, with its advantages as well as the limitations that arise from the simplifications that bring about the sought-after computational savings. Within the MLDFT formalism, the separation of the system into two parts is based on the following decomposition of the density matrix $\mathbf{D}$, which in turn defines the separation of the density function $\rho(\mathbf{r})$ as well

$$
\mathbf{D} = \mathbf{D}^A + \mathbf{D}^B \Rightarrow \rho(\mathbf{r}) = \rho^A(\mathbf{r}) + \rho^B(\mathbf{r})
\tag{3}
$$

where A and B indicate the active and inactive fragments, respectively. As stated above, the active and inactive densities are usually defined on a physico-chemical basis. In case of a molecular system in solution, it is natural to define the solute as the active fragment, whereas the solvent molecules are treated as the inactive part. It is crucial to notice that the partitioning in eq 3 is mathematical and therefore has some level of arbitrariness; that is, there is no unique way to perform the separation purely based on physico-chemical properties of the system. Therefore, a choice has to be made at this point. In this work, the partitioning is performed by means of Cholesky decomposition of the total density matrix for the active occupied MOs, from which the active density matrix $\mathbf{D}^A$ is calculated, similarly to what was done in previous works presenting alternative multilevel methods.[12,58,60,62,63] One advantage of this procedure is that it ensures the all active and inactive orbitals are orthogonal and remain so during all the subsequent SCF procedure performed on the active subsystem.[64]

Now using eq 3, the total electronic energy in eq 1 can be written as

$$E[\mathbf{D}^A; \mathbf{D}^B] = \mathrm{Tr}\mathbf{h}\mathbf{D}^A + \mathrm{Tr}\mathbf{h}\mathbf{D}^B$$
$$+ \frac{1}{2}\mathrm{Tr}\mathbf{D}^A\mathbf{J}(\mathbf{D}^A) + \frac{1}{2}\mathrm{Tr}\mathbf{D}^B\mathbf{J}(\mathbf{D}^B) + \mathrm{Tr}\mathbf{D}^A\mathbf{J}(\mathbf{D}^B)$$
$$- c_x\left(\frac{1}{2}\mathrm{Tr}\mathbf{D}^A\mathbf{K}(\mathbf{D}^A) + \frac{1}{2}\mathrm{Tr}\mathbf{D}^B\mathbf{K}(\mathbf{D}^B) + \mathrm{Tr}\mathbf{D}^A\mathbf{K}(\mathbf{D}^B)\right)$$
$$+ (1 - c_x)\int\rho(\mathbf{r})\varepsilon_x(\rho(\mathbf{r}))\mathrm{d}\mathbf{r} + \int\rho(\mathbf{r})\varepsilon_c(\rho(\mathbf{r}))\mathrm{d}\mathbf{r}$$

$$(4)$$

where the symmetry of $\mathbf{J}$ and $\mathbf{K}$ matrices has been used. So far, the method is analogous to the recently proposed MLHF approach.[58] However, in MLDFT, the last two terms are not linear in the densities of the two subsystems; therefore, we cannot directly split them into distinct contributions arising from $\rho^A(\mathbf{r})$ and $\rho^B(\mathbf{r})$. In order to acquire a physical understanding of eq 4, we rewrite the last two terms by using this trivial identity for the exchange−correlation energy density ($\varepsilon_{xc} = \varepsilon_x + \varepsilon_c$)

$$\int[\rho^A(\mathbf{r}) + \rho^B(\mathbf{r})]\varepsilon_{xc}(\rho^A(\mathbf{r}) + \rho^B(\mathbf{r}))\mathrm{d}\mathbf{r}$$
$$= \int[\rho^A(\mathbf{r}) + \rho^B(\mathbf{r})]\varepsilon_{xc}(\rho^A(\mathbf{r}) + \rho^B(\mathbf{r}))\mathrm{d}\mathbf{r}$$
$$+ \int\rho^A(\mathbf{r})\varepsilon_{xc}(\rho^A(\mathbf{r}))\mathrm{d}\mathbf{r} - \int\rho^A(\mathbf{r})\varepsilon_{xc}(\rho^A(\mathbf{r}))\mathrm{d}\mathbf{r}$$
$$+ \int\rho^B(\mathbf{r})\varepsilon_{xc}(\rho^B(\mathbf{r}))\mathrm{d}\mathbf{r} - \int\rho^B(\mathbf{r})\varepsilon_{xc}(\rho^B(\mathbf{r}))\mathrm{d}\mathbf{r}$$
$$+ \int\rho^A(\mathbf{r})\varepsilon_{xc}(\rho^B(\mathbf{r}))\mathrm{d}\mathbf{r} - \int\rho^A(\mathbf{r})\varepsilon_{xc}(\rho^B(\mathbf{r}))\mathrm{d}\mathbf{r}$$
$$+ \int\rho^B(\mathbf{r})\varepsilon_{xc}(\rho^A(\mathbf{r}))\mathrm{d}\mathbf{r} - \int\rho^B(\mathbf{r})\varepsilon_{xc}(\rho^A(\mathbf{r}))\mathrm{d}\mathbf{r}$$

$$(5)$$

Substituting eq 5 into 4 and reorganizing the different terms we obtain

$$E[\mathbf{D}^A; \mathbf{D}^B] = \mathrm{Tr}\mathbf{h}\mathbf{D}^A + \frac{1}{2}\mathrm{Tr}\mathbf{D}^A\mathbf{J}(\mathbf{D}^A) - \frac{1}{2}c_x\mathrm{Tr}\mathbf{D}^A\mathbf{K}(\mathbf{D}^A)$$
$$+ (1 - c_x)\int\rho^A(\mathbf{r})\varepsilon_x(\rho^A(\mathbf{r}))\mathrm{d}\mathbf{r} + \int\rho^A(\mathbf{r})\varepsilon_c(\rho^A(\mathbf{r}))\mathrm{d}\mathbf{r}$$
$$+ \mathrm{Tr}\mathbf{h}\mathbf{D}^B + \frac{1}{2}\mathrm{Tr}\mathbf{D}^B\mathbf{J}(\mathbf{D}^B) - \frac{1}{2}c_x\mathrm{Tr}\mathbf{D}^B\mathbf{K}(\mathbf{D}^B)$$
$$+ (1 - c_x)\int\rho^B(\mathbf{r})\varepsilon_x(\rho^B(\mathbf{r}))\mathrm{d}\mathbf{r} + \int\rho^B(\mathbf{r})\varepsilon_c(\rho^B(\mathbf{r}))\mathrm{d}\mathbf{r}$$
$$+ \mathrm{Tr}\mathbf{D}^A\mathbf{J}(\mathbf{D}^B) - c_x\mathrm{Tr}\mathbf{D}^A\mathbf{K}(\mathbf{D}^B)$$
$$+ (1 - c_x)\left(\int\rho^A(\mathbf{r})\varepsilon_x(\rho^B(\mathbf{r}))\mathrm{d}\mathbf{r} + \int\rho^B(\mathbf{r})\varepsilon_x(\rho^A(\mathbf{r}))\mathrm{d}\mathbf{r}\right)$$
$$+ \int\rho^A(\mathbf{r})\varepsilon_c(\rho^B(\mathbf{r}))\mathrm{d}\mathbf{r} + \int\rho^B(\mathbf{r})\varepsilon_c(\rho^A(\mathbf{r}))\mathrm{d}\mathbf{r} + E^{AB}_{\text{non-add}}$$

$$(6)$$

where

$$E^{AB}_{\text{nonadd}}[\mathbf{D}^A; \mathbf{D}^B] = \int\rho(\mathbf{r})\varepsilon_c(\rho(\mathbf{r}))\mathrm{d}\mathbf{r}$$
$$- \int\rho(\mathbf{r})\varepsilon_c(\rho^A(\mathbf{r}))\mathrm{d}\mathbf{r}$$
$$- \int\rho(\mathbf{r})\varepsilon_c(\rho^B(\mathbf{r}))\mathrm{d}\mathbf{r}$$
$$+ (1 - c_x)\left(\int\rho(\mathbf{r})\varepsilon_x(\rho(\mathbf{r}))\mathrm{d}\mathbf{r}\right.$$
$$- \int\rho(\mathbf{r})\varepsilon_x(\rho^A(\mathbf{r}))\mathrm{d}\mathbf{r}$$
$$\left. - \int\rho(\mathbf{r})\varepsilon_x(\rho^B(\mathbf{r}))\mathrm{d}\mathbf{r}\right)$$

$$(7)$$

In eq 6, the first four lines define the energy of the active and inactive fragments, whereas the last three lines define the active−inactive interaction. In MLDFT, the density matrix of the inactive part $\mathbf{D}^B$ (and $\rho^B(\mathbf{r})$) is frozen, and therefore, it acts as an external field on the active fragment. Therefore, only the energy terms containing the B labels are fixed during the SCF procedure. In this regard, the MLDFT is similar in construction to FDE approaches.[52,55,61] However, as it can be noticed from eq 7, one crucial difference is that the nonadditive energy terms do not involve the kinetic energy functional, thus allowing for a computational advantage with respect to common FDE approaches, which need to specifically employ nonadditive kinetic terms. This advantage stems from the fact that in MLDFT, the separation of the system into two subsystems is performed on the total density matrix, rather than the density function. However, one disadvantage with respect to FDE is that for MLDFT, we are forced to select a suitable decomposition algorithm [Cholesky/projected atomic orbitals (PAOs) are used in this work for active occupied and virtual orbitals, respectively], and the results may depend on it, while FDE does not need to make this choice. The total DFT Fock matrix is given by

$$F_{\mu\nu} = h_{\mu\nu} + J_{\mu\nu}(\mathbf{D}) - c_x K_{\mu\nu}(\mathbf{D})$$
$$+ (1 - c_x) \int v_x(\rho(\mathbf{r})) \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r}$$
$$+ \int v_c(\rho(\mathbf{r})) \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r}$$

(8)

where $v_x(\rho(\mathbf{r}))$ and $v_c(\rho(\mathbf{r}))$ are the exchange and correlation potential densities, respectively. Using the partitioning in eqs 3 and 8 we get

$$F_{\mu\nu} = h_{\mu\nu} + J_{\mu\nu}(\mathbf{D}^A) + J_{\mu\nu}(\mathbf{D}^B)$$
$$- c_x(K_{\mu\nu}(\mathbf{D}^A) + K_{\mu\nu}(\mathbf{D}^B))$$
$$+ (1 - c_x) \int v_x(\rho^A(\mathbf{r}) + \rho^B(\mathbf{r})) \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r}$$
$$+ \int v_c(\rho^A(\mathbf{r}) + \rho^B(\mathbf{r})) \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r}$$

(9)

We exploit the same identity of eq 5 for the exchange–correlation potential density ($v_{xc} = v_x + v_c$). In this way, the last two terms in eq 9 become

$$\int v_{xc}(\rho_A(\mathbf{r}) + \rho_B(\mathbf{r})) \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r}$$
$$= \int v_{xc}(\rho_A(\mathbf{r}) + \rho_B(\mathbf{r})) \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r}$$
$$+ \int v_{xc}(\rho_A(\mathbf{r})) \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r} - \int v_{xc}(\rho_A(\mathbf{r})) \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r}$$
$$+ \int v_{xc}(\rho_B(\mathbf{r})) \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r} - \int v_{xc}(\rho_B(\mathbf{r})) \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r}$$

(10)

Reorganizing the terms in eq 9, we can obtain the working expression for the MLDFT Fock matrix

$$F_{\mu\nu} = h_{\mu\nu}$$
$$+ \underbrace{J_{\mu\nu}(\mathbf{D}^A) - c_x K_{\mu\nu}(\mathbf{D}^A) + (1 - c_x) \int v_x(\rho^A(\mathbf{r})) \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r} + \int v_c(\rho^A(\mathbf{r})) \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r}}_{2e_{A,\mu\nu}}$$
$$+ \underbrace{J_{\mu\nu}(\mathbf{D}^B) - c_x K_{\mu\nu}(\mathbf{D}^B) + (1 - c_x) \int v_x(\rho^B(\mathbf{r})) \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r} + \int v_c(\rho^B(\mathbf{r})) \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r}}_{2e_{B,\mu\nu}}$$
$$+ \underbrace{(1 - c_x) \int \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) [v_x(\rho^A(\mathbf{r}) + \rho^B(\mathbf{r})) - v_x(\rho^A(\mathbf{r})) - v_x(\rho^B(\mathbf{r}))] d\mathbf{r}}_{2e_{AB,\mu\nu}^{\text{non-add}}}$$
$$+ \underbrace{\int \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) [v_c(\rho^A(\mathbf{r}) + \rho^B(\mathbf{r})) - v_c(\rho^A(\mathbf{r})) - v_c(\rho^B(\mathbf{r}))] d\mathbf{r}}_{2e_{AB,\mu\nu}^{\text{non-add}}}$$

(11)

where the two-electron contributions of A and B fragments and the interaction term AB are highlighted as $2e_{X,\mu\nu}$, $\{X = A, B, AB\}$.

There are two main advantages of using MLDFT compared to full DFT. First, the HF exchange contribution is usually the most expensive term in most hybrid functionals. In MLDFT, only the active exchange term is to be computed at each SCF cycle, whereas the exchange integral of the inactive fragment is computed at the first SCF cycle only, as it is constant during the optimization. Second, the MLDFT SCF procedure can be performed in the MO basis of the active part only, thus intrinsically reducing the computational time as previously observed for the MLHF method.[58]

2.2. Comparison between MLDFT and Other Quantum-Embedding Methods. As stated above, in MLDFT, the density matrix of the whole system is partitioned into active and inactive fragment densities (see eq 3), from which the active and inactive density functions are calculated. The inactive density matrix is kept frozen during the SCF cycles and enters the active Fock matrix as a one-electron term. The MLDFT shares many similarities with FDE methods;[42−45,52,54,61] however, some relevant differences are present, some of which are mentioned in the previous section. First, due to the partitioning performed on the density matrix instead of the density function as in FDE, we avoid the problems arising in the definition of the nonadditive kinetic potential terms that enforce Pauli exclusion between the electrons of the various subsystems in FDE. This is clearly evident in the definition of the nonadditive energy terms in eq 7, which are only due to the intrinsic nonlinearity of the DFT exchange–correlation functionals. However, a choice needs to be made on the decomposition algorithm to partition the density matrix.

One other difference that is worth mentioning is that if the decomposition is applied to the fully converged DFT density matrix, then, the MLDFT energy corresponds to the exact DFT energy of the full system, and it is in fact completely independent on the decomposition algorithm. This feature is shared with projection-based approaches, though the philosophy of MLDFT conceptually differs from DFT-in-DFT methods.[35,39,41,46,47,49] Of course, the point of MLDFT is to avoid a full DFT calculation on the entire system and instead find a way to approximate the total density matrix, which is then partitioned and the density of the active subsystem is then optimized through a subsequent SCF procedure while the density matrix of the inactive part is kept frozen. In this work, this initial full density matrix is obtained through superposition of molecular densities[51,65] followed by Fock matrix diagonalization, although we should emphasize that the method is general and a different choice may be made. Within this procedure, the solute−solvent and solvent−solvent polarization energy terms are only introduced in the starting full Fock diagonalization step. In projection-based approaches, a DFT calculation on the whole system is performed, and the occupied orbitals are then localized. The localized MOs are then assigned to the active or inactive fragments, based on a population threshold.[50] Neither step is required in MLDFT, in which a starting Fock matrix is diagonalized and the obtained density matrix is decomposed into fragment densities. This explains why the SCF procedure on the active subsystem has to be performed in the MO basis of the active fragment only, which is defined in terms of the full AO basis set. We do not perform any basis set truncation which is instead usually
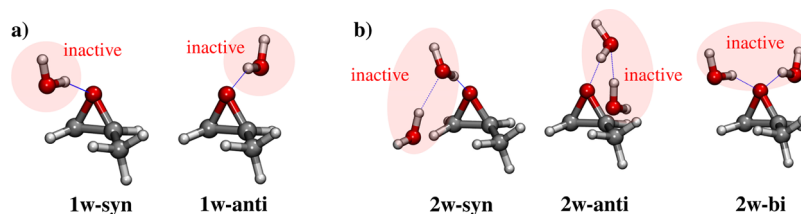
Figure 1. Structure of conformers methyloxirane/water clusters. MLDFT partition is constructed so that methyloxirane is the active part whereas the water molecules are the inactive fragments.

applied in projection-based approaches. This feature automatically allows for saving the computational cost because the inactive MOs are not involved in the SCF procedure.

3. COMPUTATIONAL DETAILS

The DFT and MLDFT are implemented in a development version of the electronic structure program e^T v.1.0.[66] In particular, the DFT grid is constructed using the widely employed Lebedev grid,[67] with the radial quadrature proposed in ref 68. All the calculations are performed using a quadrature of 25th order, and the radial threshold is set to 10^{-5}. The DFT functionals are implemented using the LibXC library.[69,70]

The MLDFT calculation follows this computational protocol:

1. Construction of the initial density matrix by means of superposition of molecular densities,[51,65] followed by the diagonalization of the initial Fock matrix.

2. Partitioning of the new density matrix into A and B densities, using Cholesky decomposition for the active occupied orbitals and PAOs for active virtual orbitals.[12,58,60,62−64,71,72] We refer the reader to ref 63 for the full details on this partitioning of the density matrix. The inactive density matrix is obtained by subtracting the active density matrix from the total one.

3. Calculation of the constant energy terms and the one-electron contributions due to the inactive density matrix B entering in eqs 6 and 11.

4. Minimization of the energy defined in eq 6 in the MO basis of the active part A only, until convergence is reached. All the equations reported in Section 2, which are expressed in the AO basis set, can be transformed in the active MO basis by using the active MO coefficients.

To show the robustness of MLDFT, three different functionals are used: LDA,[73] GGA (PBE[74]), and hybrid (B3LYP[75]). These are combined with three different basis sets: 6-31G, 6-31G*, and aug-cc-pVDZ.

4. NUMERICAL APPLICATIONS

In this section, the MLDFT is applied to some test cases to show the accuracy and the performance of the method. Solvation is one of the main physico-chemical phenomena in which such approaches can be exploited. We show the results of coupling MLDFT with two alternative, fully atomistic, strategies to model aqueous solutions. The first consists of a static modeling, which uses small clusters composed of the solute and a small number of surrounding water molecules. As an alternative, we apply MLDFT to snapshots extracted from a molecular dynamics (MD) simulation. In the latter framework, the dynamical aspects of the solvation phenomenon are retained, as those arising from the combination of conformational changes in the solute and the surrounding solvent. In

addition, long range interactions are taken into account. This latter modeling of the solvation phenomenon has been amply and successfully exploited by some of us within the framework of QM/MM approaches.[20,30,76−78]

In the following sections, the combination of MLDFT to the two aforementioned solvation approaches is tested, with application to two relatively small molecules, that is, methyloxirane and glycidol in aqueous solution, which have been studied in the literature both theoretically and experimentally.[27,79−89] Such systems are chosen not only for their simplicity but also because methyloxirane is a rigid molecule, whereas glycidol is not. Therefore, in the latter case, the results depend on the selected QM level, and the approach used to solvation and conformational flexibility, which is instead discarded in the case of methyloxirane. In this way, we can dissect the various effects and highlight the quality of the MLDFT approach in details.

4.1. Cluster Models. *4.1.1. Methyloxirane/Water Clusters.* The first studied solute is methyloxirane (MOXY), which is one of the smallest molecules that exhibits a chiral carbon. We have selected different clusters constituted by MOXY and one or two water molecules (see Figure 1) that have been previously studied by Su and Xu[86] to explain the unique characteristics of MOXY in aqueous solution.[86,89]

The two different conformers for the cluster composed of MOXY and one water molecule (MOXY + 1w) are depicted in Figure 1a. In the **1w-syn** structure, water interacts with MOXY through hydrogen bonding on the same side of the methyl group, whereas the opposite occurs for the **1w-anti** structure. In both cases, MOXY is the active fragment, and water is the inactive moiety in MLDFT calculations.

GS energy differences between DFT and MLDFT calculations are depicted in Figure 2, panel (a), left. Raw data are reported in Table S1 given in Supporting Information. We see that the error between MLDFT and full DFT is below 1 mHartree (<0.628 kcal/mol), irrespective of the combination of functional/basis set employed. The error due to the MLDFT partitioning is well below the chemical accuracy (i.e. 1 kcal/mol).

In the right panel of Figure 2a, DFT and MLDFT energy differences between **1w-anti** and **1w-syn** conformers are reported for all the considered combinations of the functional/basis set. The raw data are reported in Table S1 in the Supporting Information. We see that DFT and MLDFT values almost coincide. In particular, LDA and PBE functionals predict **1w-syn** to be the most stable conformer, both at DFT and MLDFT levels, independently of the selected basis set. Notice however that the energy difference between the two conformers decreases either as GGA functionals are employed or diffuse/polarization basis sets are used. The inclusion of HF exchange makes **1w-anti** the most stable conformer, if polarization/diffuse functions are considered. However, for

(a) MOXY+1w clusters

(b) MOXY+2w clusters
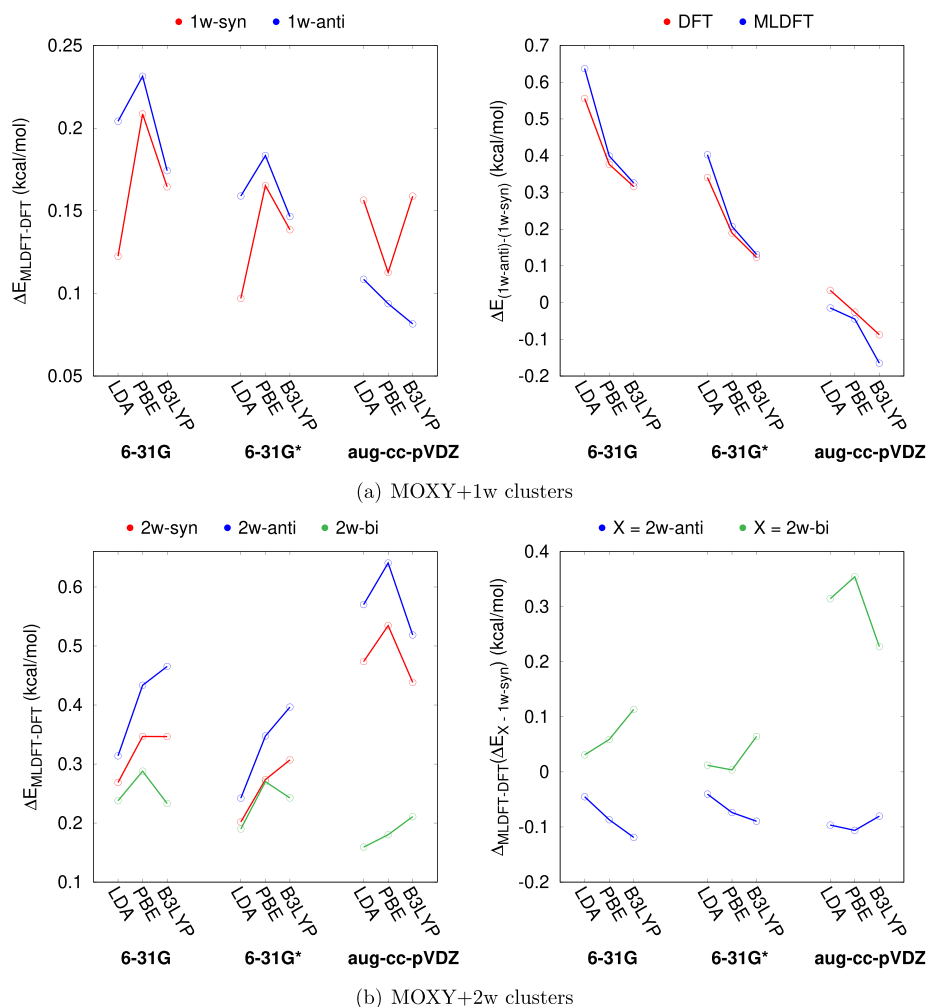
Figure 2. (a) (Left) **1w-syn** and **1w-anti** total energy differences between MLDFT and DFT. (Right) MLDFT and DFT energy difference between **1w-anti** and **1w-syn**. (b) (Left) **2w-syn**, **2w-anti**, and **2w-bi** total energy differences between MLDFT and DFT. (Right) MLDFT-DFT difference on relative energies of **2w-anti** and **2w-bi** with respect to **2w-syn**.

all the considered combinations of the functional/basis set, MLDFT and DFT values are almost perfectly in agreement, with the largest discrepancy being reported for B3LYP/aug-cc-pVDZ (0.08 kcal/mol). These findings clearly show that for this system, MLDFT is able to catch small energy differences, which are again well below the chemical accuracy.

We now turn to the clusters composed of MOXY and two water molecules (MOXY + 2w, Figure 1b). Three main conformers are considered, according to Su and Xu:[86] **2w-syn**, **2w-anti**, and **2w-bi**. The first two conformers differ from the position of water molecules, being both placed on the same side with respect to the methyl group in case of **2w-syn** or on the opposite side for **2w-anti**. In **2w-bi**, the two water molecules are instead placed on the opposite sides of the epoxyl oxygen atom. In all MLDFT calculations, MOXY is the active moiety, whereas the two water molecules are inactive.

In Figure 2b, left, GS energy differences between DFT and MLDFT for the three conformers are reported. The raw values associated with the data plotted in Figure 2b are given in Table S2 in the Supporting Information. The MLDFT and DFT results are, also in this case, in very good agreement with an absolute error below 1 kcal/mol for all combinations of functional/basis sets. However, the absolute deviation between DFT and MLDFT energies is larger than for the previous case

(see Figure 2a). In particular, the MLDFT error is larger for **2w-syn** and **2w-anti** than for **2w-bi**, for which it is in line with what we have shown above for MOXY + 1w clusters (∼0.1–0.3 kcal/mol). The increase in the error may be justified by the fact that **2w-syn** and **2w-anti** feature one water molecule that is linked to another water molecule by means of intermolecular hydrogen bonding. The density matrix of the inactive fragments (the two water molecules) is kept frozen; therefore, the water molecule that is not directly bonded to the solute remains in its frozen electronic configuration, resulting in a larger error in the total energy. Such an hypothesis is confirmed by the fact that the error increases when the diffuse aug-cc-pVDZ basis set is used, and the same does not occur for **2w-bi**, where both water molecules are directly linked to methyloxirane through hydrogen bonding interactions.

The MLDFT-DFT deviations in energy differences between each conformer and **2w-syn** are shown in Figure 2b, right. We note small discrepancies between MLDFT and full DFT; however, also in this case, they are below the chemical accuracy, with the maximum error reported by PBE/6-31G* (∼0.35 kcal/mol). The error in the energy differences between the conformers is lower than for the total GS energies reported in Figure 2b, left.

F

4.1.2. Glycidol/Water Clusters. MOXY is a rigid molecule, so the different solvated conformers mainly differ by the position of the water molecules. In this section, we show how MLDFT can treat flexible solutes, and to this end, we have selected glycidol (GLY), which is a derivative of MOXY, where one hydrogen of the methyl group is replaced by the OH group (see Figures 3–5). In all MLDFT calculations, the GLY
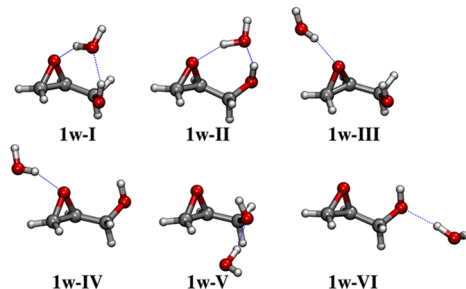
Figure 3. Structures of the six conformers of glycidol + 1 water clusters. In MLDFT calculations, the glycidol moiety is active whereas the water molecule is inactive.
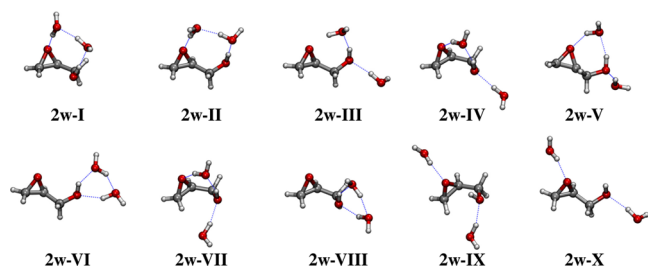
Figure 4. Structures of the ten conformers of glycidol + 2 waters clusters. In MLDFT calculations, glycidol is active whereas water molecules are inactive.
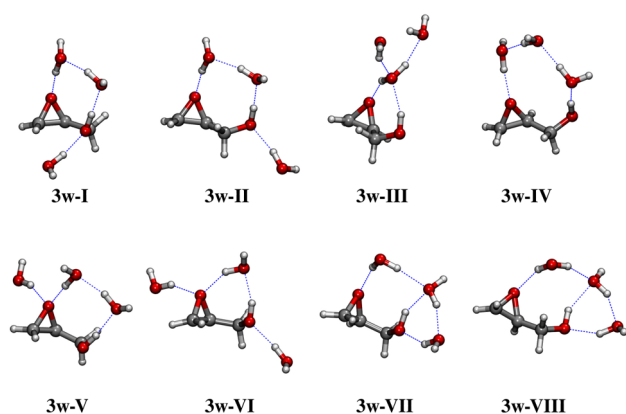
Figure 5. Structures of the eight conformers of glycidol + 3 waters clusters. In MLDFT calculations, glycidol is active whereas water molecules are inactive.

moiety is the active fragment and the water molecules are the inactive part. The presence of the hydroxyl group makes glycidol flexible up to the point that eight different conformers can be located in the gas phase potential energy surface (PES).[79,84]

To build up a glycidol/water clusters, different structures constituted by GLY and one, two, and three water molecules were constructed, by following the strategy reported in ref 84. Such structures are depicted in Figures 3–5. We note that the

different structures not only differ by the position of the water molecules but also by the conformation of glycidol. In particular, the six conformers constituted by GLY and one water (GLY + 1w) are characterized by a different position of the water molecule. The latter interacts via hydrogen bonding with both the hydroxyl and epoxyl groups (**1w-I** and **1w-II**), with the epoxyl group only (**1w-III** and **1w-IV**) or with only the oxygen atom of the hydroxyl group (**1w-V** and **1w-VI**). The inclusion of additional water (GLY + 2w) results in ten different conformers, which are shown in Figure 4. These contain three or four center bridges (conformers **2w-I**, **2w-II**, **2w-IV**, **2w-V**, **2w-VI**, **2w-VII,** and **2w-VIII**) or are conformers where the two water molecules interact via hydrogen bonding with the epoxyl and hydroxyl groups (conformers **2w-III**, **2w-IX,** and **2w-X**). If three explicit water molecules are added to GLY (GLY + 3w), the conformational search provides eight main conformers, which are graphically depicted in Figure 5. Similarly to the previous case, some of them contain three or four center bridges (conformers **3w-I**, **3w-II**, **3w-V**, and **3w-VI**), whereas in conformers **3w-IV**, **3w-VII,** and **3w-VIII**, a five center bridge is present. In all cases, water molecules that are not involved in bridges interact with GLY through hydrogen bonding interaction. Conformer **3w-III** is instead characterized by a three center bridge and by the remaining water molecules hydrogen bonded to the bridge water.

We now move to discuss GS energy differences between DFT and MLDFT (see Figure 6a, raw data are given in Tables S3–S5 in the Supporting Information).

In Figure 6, panel (a), MLDFT - DFT GS energy differences for all the different conformers of GLY + 1w, GLY + 2w, and GLY + 3w water clusters are shown. The error reported by MLDFT is below 0.1 mH (<0.627 kcal/mol) when applied to GLY + 1w, at all the selected levels of theory. In particular, energy differences are perfectly in line with what is shown in Figure 2a, left panel, in case of MOXY + 1w clusters. Moving to GLY + 2w conformers, the agreement between DFT and MLDFT is almost perfect at all levels of theory, being the energy difference below 0.8 kcal/mol in all cases. We also see that at the B3LYP/aug-cc-pVDZ level, for **2w-I** and **2w-II,** the difference between MLDFT and full DFT is larger than for the other conformers (>0.1 mH, 0.627 kcal/mol). This is due to the specific spatial arrangement of water molecules, which create a four-center bridge connecting GLY hydroxyl and epoxyl groups (see Figure 4).

As stated above for MOXY + 2w clusters, MLDFT accounts for all the interactions between active and inactive parts, with the exception of dispersion; however, the inactive fragment(s) are described by a frozen density matrix. Therefore, polarization and charge transfer (and dispersion) effects are neglected in the inactive region. For **2w-I** and **2w-II,** we can speculate that such interactions may play a relevant role because the two inactive water molecules are hydrogen bonded. Also, their role is clearly increased when diffuse and polarization functions are included in the basis set (aug-cc-pVDZ) because such functions enhance the effects of these interactions. This does not occur in case of other conformers because of the different spatial arrangement of the solvent molecules.

We now focus on GLY + 3w conformers. The agreement between MLDFT and the reference full DFT values is generally worse than in the previous cases (see the right panel of Figure 6a). However, the average error is of about 0.67 kcal/mol (∼0.1 mH), that is, again well beyond the
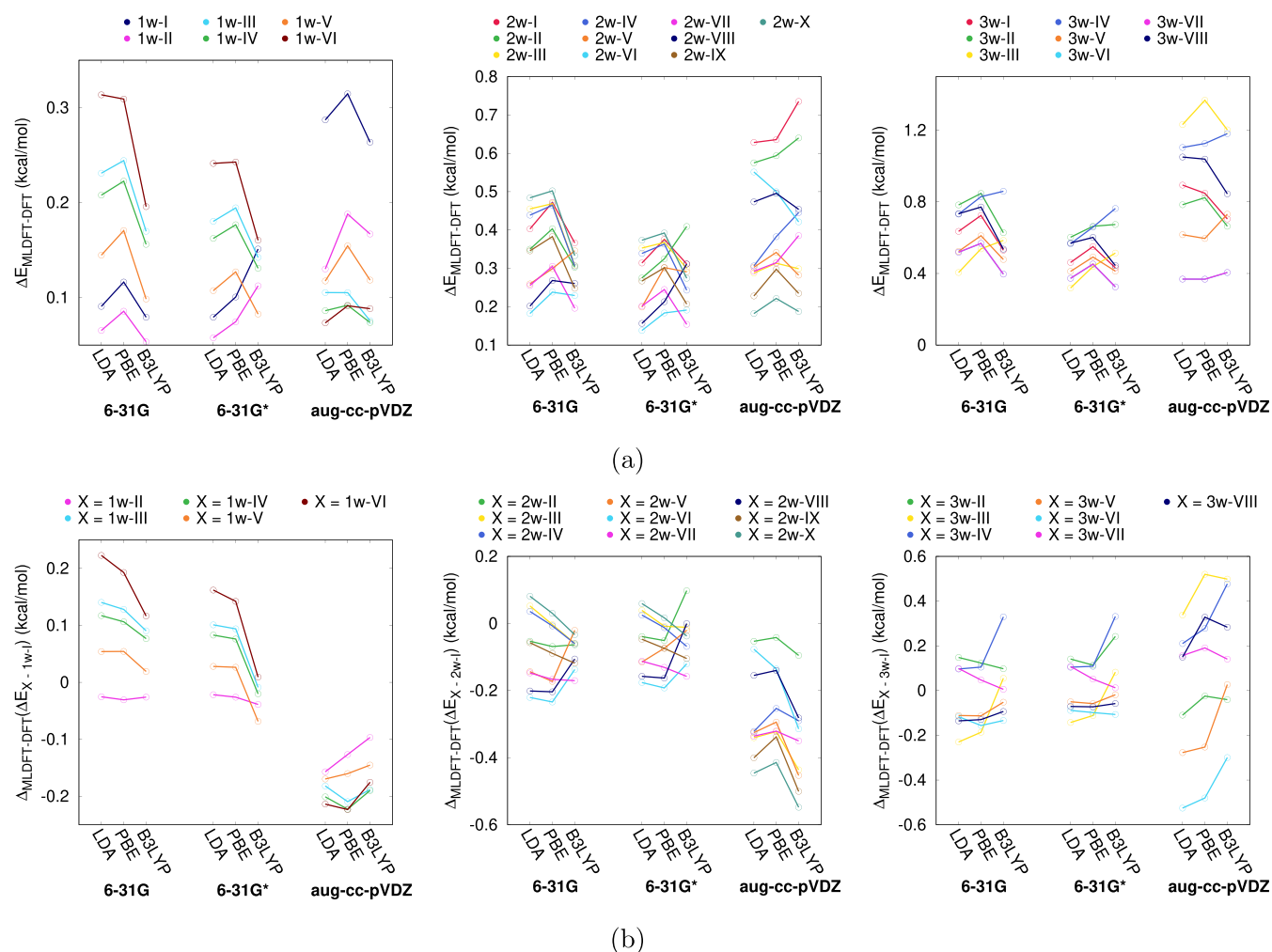
Figure 6. (a) GLY + 1w (left), GLY + 2w (middle), and GLY + 3w (right) GS energy differences between MLDFT and reference DFT values. (b) MLDFT-DFT energy deviations for the energy differences between each conformer of GLY + 1w conformers and **1w-I** (left), GLY + 2w conformers and **2w-I** (middle) and GLY + 3w conformers and **3w-I** (right). All values are reported in kcal/mol.

chemical accuracy. The largest discrepancy is shown by **3w-III** for all the functionals (LDA, PBE, or B3LYP) in combination with aug-cc-pVDZ (∼1.2 kcal/mol). Again, this can be explained by considering the spatial arrangement of water molecules around GLY (see Figure 5). Similar to **2w-I** and **2w-II**, the effect of charge transfer and polarization interactions, which are neglected by the partitioning of the inactive density matrix in MLDFT, may play a relevant role. Such effects are larger for **3w-III**; however, they affect also other conformers which are characterized by a four/five center bridge. It is also worth noticing that the MLDFT error is expected to increase with the size of the studied system because the energy is an extensive quantity. Such a trend is in fact reported for both MOXY and GLY clusters.

Let us now discuss the MLDFT-DFT energy deviations for the energy differences between each conformer of the GLY clusters and **1w-I**, **2w-I**, and **3w-I**, which are reported in Figure 6b. Raw data are given in Tables S3−S5 in the Supporting Information.

For the GLY + 1w system, both MLDFT and DFT predict **1w-I** to be the most stable at all levels of theory, whereas the relative populations of the other conformers strongly depend on the theory level (see Figure 6b, left panel). In particular, the energy differences of each conformer with respect to **1w-I**

decrease as larger basis sets are employed and also by moving from LDA to PBE and B3LYP. The error between MLDFT and DFT is instead almost constant (in absolute value) for all different combinations of the basis set and DFT functional, and in all cases, MLDFT correctly reproduces the trends obtained at the reference full DFT level.

The same considerations outlined above for GLY + 1w conformers also apply to GLY + 2w ones (see Figure 6b, middle panel). In fact, by moving from LDA to B3LYP and by including polarization and diffuse functions in the basis set, MLDFT errors with respect to DFT reference values decrease. The largest DFT-MLDFT discrepancy is reported for **2w-X** at the B3LYP/aug-cc-pVDZ level (−0.55 kcal/mol). This is due to the fact that the largest error is associated to the GS energy of the most stable conformer **2w-I** (see left panel of Figure 6a) for this combination of the DFT functional/basis set. However, as already reported for all the other studied systems, the error in the relative energies of the different conformers is always lower than the corresponding error in the total energies.

Finally, also in case of GLY + 3w clusters, the agreement between DFT and MLDFT is almost perfect, with errors ranging from −0.6 to 0.6 kcal/mol. The maximum error is observed for **3w-III** at the PBE/aug-cc-pVDZ level (0.53 kcal/mol), whereas the minimum is reported for **3w-VII** at the

Journal of Chemical Theory and Computation
pubs.acs.org/JCTC
Article

B3LYP/6-31G* level (error < 0.01 kcal/mol). Therefore, also for these systems, MLDFT provides a reliable description of the relative energies of the different conformers. The only notable exceptions are conformers **3w-III** and **3w-IV** at the LDA/6-31G and B3LYP/6-31G levels, respectively. As a final comment, we note that although the MLDFT error on total GS energy can be larger than 1 kcal/mol, relative energies of the different conformers are accurately predicted, with an error that is always below 1 kcal/mol.

4.2. Toward a Realistic Picture of Solvation. In the previous sections, we have presented and discussed solute–solvent structures obtained by modeling the solvation phenomenon in aqueous solution by means of the so-called cluster approach,[90] in which only the closest water molecules are explicitly treated at the QM level. However, this picture is not realistic, being a strongly approximate way of modeling solvation. In fact, any dynamical aspect of solvation is neglected as well as, more importantly, long range interactions which are especially relevant for polar environments such as water. In this section, we show how MLDFT may be coupled to approaches that have been developed to model solvation more realistically. In particular, we will apply MLDFT to a randomly selected structure extracted from a classical MD simulation performed on both MOXY and GLY in aqueous solution. In this way, the atomistic details of solvation are retained, and dynamical aspects could easily be introduced by repeating the calculations on several structures. A closer investigation of the latter aspect is beyond the scope of our first work on MLDFT and will be the topic of further studies.

Let us start with MOXY. We have selected one random snapshot extracted from a MD simulation, which was previously reported by some of the present authors.[82,91,92] Note that MOXY is a rigid molecule; therefore, a single snapshot well represents its conformational structure (Figure 7).
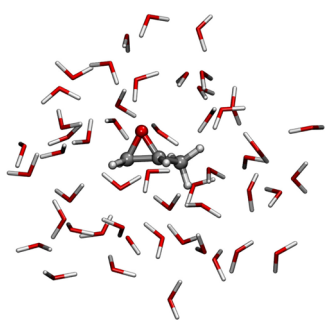
Figure 7. Selected structure of MOXY + 50 water molecules, as extracted from MD. In MLDFT calculations, MOXY is the active part and water molecules are inactive.

In MLDFT calculations, MOXY is the active fragment and it is treated at the B3LYP/6-31+G* level. The inactive part is constituted by the 50 closest water molecules, which are described at the B3LYP/6-31G level. The reference full DFT calculation is instead performed by using the B3LYP functional, in combination with the 6-31+G* basis set for MOXY and the 6-31G one for water molecules.

In order to quantify the accuracy of MLDFT, we compute the solvation energy E_{solv}, which is defined as

$$E_{solv} = E_{tot} - E_{MOXY} - E_w \qquad (12)$$

where E_{tot}, E_{MOXY}, and E_w are the total, MOXY, and water GS energies, respectively. Note that E_{MOXY} is calculated in the gas phase, and thus, it is the same in both full DFT and MLDFT calculations. E_{tot} and E_w are defined differently in the two approaches; in MLDFT, E_w is calculated at step 1 of the computational protocol (see Section 3), whereas in full DFT, it refers to the GS energy of the 50 water molecules.

Computed energy values for MOXY are reported in Table 1 for both DFT and MLDFT. We first notice that the MLDFT

Table 1. DFT and MLDFT Total GS Energies (E_{tot}) of MOXY + 50 Water Molecules Snapshot Depicted in Figure 7[a]

	DFT	MLDFT
E_{tot}	−4013.1956	−4013.1660
E_{MOXY}	−193.1079	−193.1079
E_w	−3820.0681	−3820.0382
E_{solv}	−0.0196	−0.0199

[a]E_{MOXY}, E_w, and E_{solv} are also reported. All values are given in Hartree.

error on the total energy E_{tot} is larger than what is found for clusters (see previous sections). This is not surprising because the error of the method scales with the number of the water molecules in the inactive part. Such discrepancies are primarily due to the neglect of polarization and charge-transfer interactions in the inactive solvent water molecules because their density matrix remains fixed in MLDFT. The largest contribution to the error on total energy is due to E_w. In fact, MLDFT E_w differs from full DFT of about the same extent as total energies. Such differences between MLDFT and DFT are reflected by the computed solvation energy, which can be taken as a measure of the accuracy of MLDFT. For the studied snapshot, the agreement between MLDFT and DFT is almost perfect (−12.30 vs −12.48 kcal/mol), and the error is of about 0.2 kcal/mol.

The same analysis may be applied to glycidol, for which the snapshots were extracted from MD simulations previously reported by some of us.[79] We recall that GLY is a flexible solute, of which the main conformers may be identified by means of two dihedral angles δ_1 and δ_2 [see Figure 8, panel (a)]. Seven most probable conformers have been selected [see Figure 8 panel (b)].

The MLDFT partition has been done so that GLY is the active fragment and treated at the B3LYP/6-31+G* level, whereas water molecules are inactive and described at the B3LYP/6-31G level. All the reference full DFT calculations are performed by using the B3LYP functional in combination with the 6-31+G* basis set for the solute and the 6-31G one for the water molecules.

The DFT and MLDFT energies (E_{tot}, E_{GLY}, E_w, and E_{solv}) are reported in Table S6 in the Supporting Information. Overall, MLDFT total energies are higher than DFT values of about 0.02−0.03 Hartree. The reasons of this discrepancy are the same as reported for MOXY.

The DFT and MLDFT solvation energies are graphically compared in Figure 9. We observe that all MLDFT values are almost in perfect agreement with the reference full DFT data. The average discrepancy is of about 0.7 kcal/mol (∼1 mH), with the largest discrepancy reported for conformer 5 (1.1 kcal/mol). Notice that in this study, we only include the GLY moiety in the active part. Similar calculations performed at the
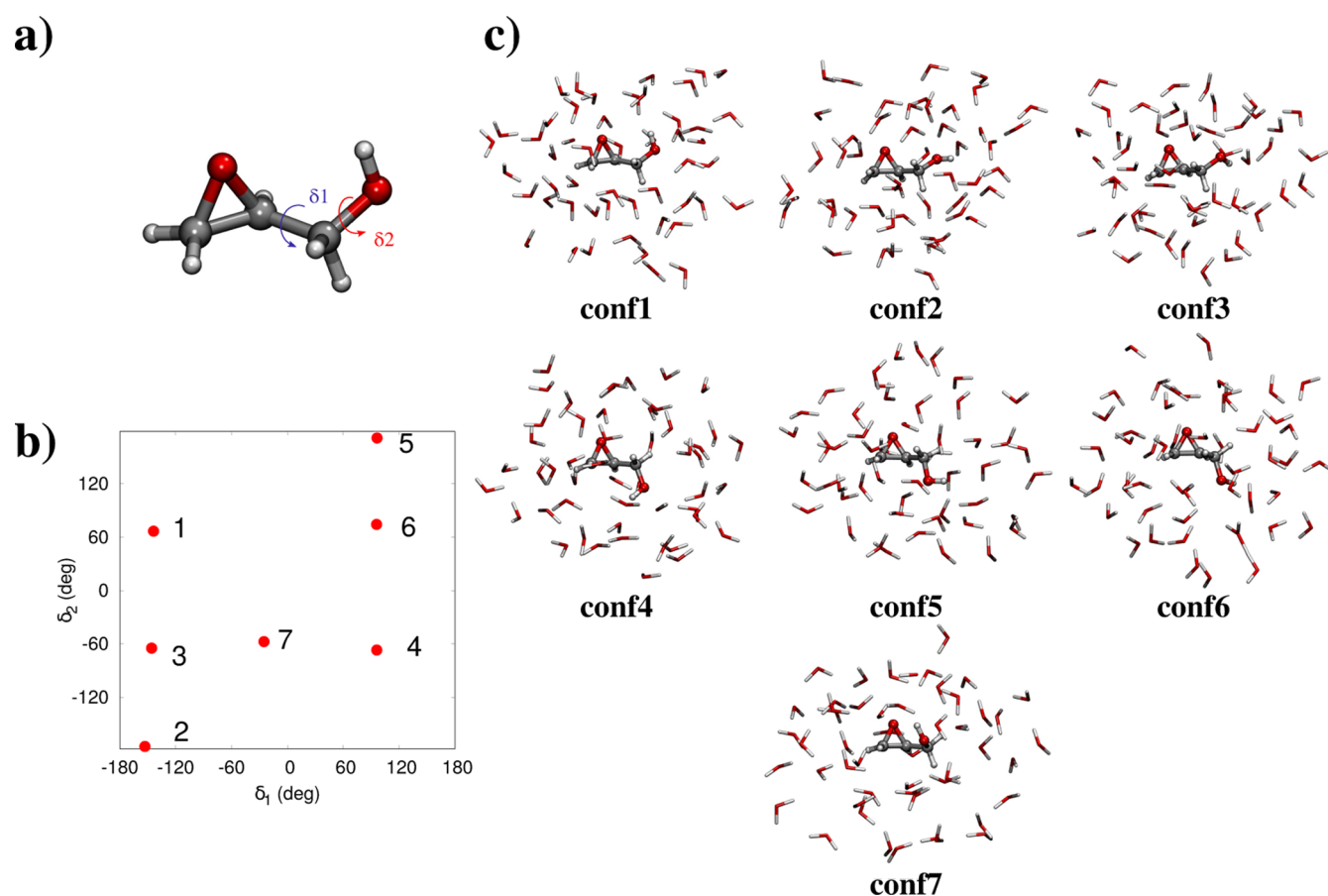
Figure 8. (a) Definition of δ_1 and δ_2 dihedral angles of GLY; (b) δ_1 and δ_2 values for the selected GLY + 50 water molecules snapshots extracted from MD; (c) molecular structures of the seven selected snapshots. In MLDFT calculations, GLY is the active part, whereas water molecules are inactive.
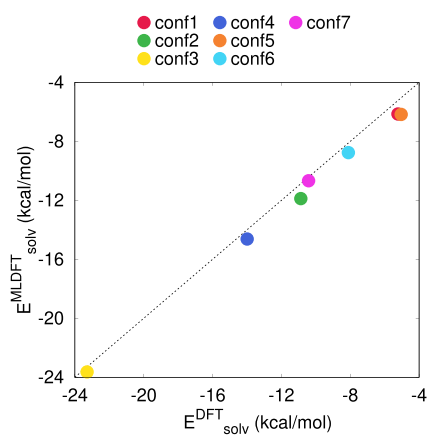
Figure 9. DFT and MLDFT solvation energies (E_{solv}) for the different conformers graphically depicted in Figure 8. Values are given in kcal/mol.

MLHF level[58] needed to insert at least five water molecules in the active fragment to reach the same level of accuracy.

5. SUMMARY AND CONCLUSIONS

In this work, we report a novel density-based multilevel approach based on the DFT treatment of the electronic structure problem. In MLDFT, the full system is partitioned in two layers, one active and one inactive. MLDFT stands apart from methods based on a similar paradigm because of the choice of partitioning an initial density matrix into active and inactive parts, instead of the density function. The MLDFT SCF procedure is then performed in the MO basis of the active subsystem only. This is the source of the reduction in computational cost because the density matrix of the inactive fragments is kept frozen during the optimization of the density.

The MLDFT was applied to aqueous methyloxirane and glycidol, for which two different approaches to solvation were discussed. First, the so-called cluster approach is employed, which models solvation in terms of minimal clusters composed of the solute and a small number of water molecules. Second, a more realistic picture is considered, which focuses on randomly selected snapshots extracted from MD simulations. For all studied structures, the computed data confirm that MLDFT is able to correctly reproduce reference full DFT values, with errors which are always ≤1 kcal/mol. Because of its favorable computational scaling, MLDFT can be coupled to more realistic approaches to solvation, that is, it can treat a large number of representative snapshots extracted from MD simulations, so to effectively take into account the dynamical aspects of solvation.

In this first presentation of the approach, we have limited the analysis to GS energies. However, MLDFT has the potential to be extended to the calculation of molecular properties and spectra. In particular, the analytical evaluation of molecular gradients will allow for MLDFT geometry optimizations. However, this extension is not trivial because in order to guarantee the continuity of the PES, the same pivots must be

imposed in the Cholesky decomposition. The calculation of molecular geometries and properties will be the topic of future communications. Also, the active and inactive parts can be described at two different levels of theory, for instance, using two different DFT functionals. Similarly to projection-based approaches, the active and inactive regions can also be treated at the HF and DFT levels, respectively, thus allowing for post-HF calculations on the active part only.[50] Such extensions will be the topic of future communications.

The method will also be further developed by focusing on some technical aspects, which are worth being improved. For instance, in the current implementation, the DFT grid is homogeneous in the whole space. However, it is reasonable to assume that the grid can be downgraded further away from the active part, because within the focused model paradigm, we only seek to accurately model the effect of the environment on the properties of the active subsystem, rather than the intrinsic properties of the environment itself. Technical refinements of the current implementation are in progress and will be discussed in future communications.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.0c00940.

Data related to Figures 2, 6, and 9 (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Tommaso Giovannini − *Department of Chemistry, Norwegian University of Science and Technology, 7491 Trondheim, Norway;* ⓞ orcid.org/0000-0002-5637-2853; Email: tommaso.giovannini@ntnu.no

Henrik Koch − *Scuola Normale Superiore, 56126 Pisa, Italy;* ⓞ orcid.org/0000-0002-8367-8727; Email: henrik.koch@sns.it

Authors

Gioia Marrazzini − *Scuola Normale Superiore, 56126 Pisa, Italy*

Marco Scavino − *Scuola Normale Superiore, 56126 Pisa, Italy*

Franco Egidi − *Scuola Normale Superiore, 56126 Pisa, Italy;* ⓞ orcid.org/0000-0003-3259-8863

Chiara Cappelli − *Scuola Normale Superiore, 56126 Pisa, Italy;* ⓞ orcid.org/0000-0002-4872-4505

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.0c00940

Author Contributions

§G.M. and T.G. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Dykstra, C.; Frenking, G.; Kim, K.; Scuseria, G. *Theory and Applications of Computational Chemistry: The First Forty Years*; Elsevier, 2011.

(2) Reichardt, C. Solvatochromism, thermochromism, piezochromism, halochromism, and chiro-solvatochromism of pyridinium N-phenoxide betaine dyes. *Chem. Soc. Rev.* **1992**, *21*, 147−153.

(3) Buncel, E.; Rajagopal, S. Solvatochromism and solvent polarity scales. *Acc. Chem. Res.* **1990**, *23*, 226−231.

(4) Reichardt, C. Solvatochromic dyes as solvent polarity indicators. *Chem. Rev.* **1994**, *94*, 2319−2358.

(5) Cannelli, O.; Giovannini, T.; Baiardi, A.; Carlotti, B.; Elisei, F.; Cappelli, C. Understanding the interplay between the solvent and nuclear rearrangements in the negative solvatochromism of a push−pull flexible quinolinium cation. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32544−32555.

(6) Carlotti, B.; Cesaretti, A.; Cannelli, O.; Giovannini, T.; Cappelli, C.; Bonaccorso, C.; Fortuna, C. G.; Elisei, F.; Spalletti, A. Evaluation of Hyperpolarizability from the Solvatochromic Method: Thiophene Containing Push- Pull Cationic Dyes as a Case Study. *J. Phys. Chem. C* **2018**, *122*, 2285−2296.

(7) Cupellini, L.; Calvani, D.; Jacquemin, D.; Mennucci, B. Charge transfer from the carotenoid can quench chlorophyll excitation in antenna complexes of plants. *Nat. Commun.* **2020**, *11*, 1−8.

(8) Bondanza, M.; Cupellini, L.; Lipparini, F.; Mennucci, B. The multiple roles of the protein in the photoactivation of Orange Carotenoid Protein. *Chem* **2020**, *6*, 187−203.

(9) Lomize, A. L.; Pogozheva, I. D.; Lomize, M. A.; Mosberg, H. I. Positioning of proteins in membranes: a computational approach. *Protein Sci.* **2006**, *15*, 1318−1333.

(10) Furse, K. E.; Corcelli, S. A. The dynamics of water at DNA interfaces: Computational studies of Hoechst 33258 bound to DNA. *J. Am. Chem. Soc.* **2008**, *130*, 13103−13109.

(11) Zhao, S.; Wei, G. W. High-order FDTD methods via derivative matching for Maxwell's equations with material interfaces. *J. Comput. Phys.* **2004**, *200*, 60−103.

(12) Myhre, R. H.; Koch, H. The multilevel CC3 coupled cluster model. *J. Chem. Phys.* **2016**, *145*, 044111.

(13) Høyvik, I.-M.; Myhre, R. H.; Koch, H. Correlated natural transition orbitals for core excitation energies in multilevel coupled cluster models. *J. Chem. Phys.* **2017**, *146*, 144109.

(14) Parr, R. G. *Horizons of Quantum Chemistry*; Springer, 1980; pp 5−15.

(15) Burke, K. Perspective on density functional theory. *J. Chem. Phys.* **2012**, *136*, 150901.

(16) Sisto, A.; Stross, C.; van der Kamp, M. W.; O'Connor, M.; McIntosh-Smith, S.; Johnson, G. T.; Hohenstein, E. G.; Manby, F. R.; Glowacki, D. R.; Martinez, T. J. Atomistic non-adiabatic dynamics of the LH2 complex with a GPU-accelerated ab initio exciton model. *Phys. Chem. Chem. Phys.* **2017**, *19*, 14924−14936.

(17) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999−3094.

(18) Mennucci, B. Polarizable Continuum Model. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 386−404.

(19) Tomasi, J.; Persico, M. Molecular interactions in solution: an overview of methods based on continuous distributions of the solvent. *Chem. Rev.* **1994**, *94*, 2027−2094.

(20) Cappelli, C. Integrated QM/Polarizable MM/Continuum Approaches to Model Chiroptical Properties of Strongly Interacting Solute-Solvent Systems. *Int. J. Quantum Chem.* **2016**, *116*, 1532−1542.

(21) Warshel, A.; Karplus, M. Calculation of ground and excited state potential surfaces of conjugated molecules. I. Formulation and parametrization. *J. Am. Chem. Soc.* **1972**, *94*, 5612−5625.

(22) 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.

(23) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.

(24) Lin, H.; Truhlar, D. G. QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* **2007**, *117*, 185−199.

(25) Curutchet, C.; Muñoz-Losa, A.; Monti, S.; Kongsted, J.; Scholes, G. D.; Mennucci, B. Electronic energy transfer in condensed phase studied by a polarizable QM/MM model. *J. Chem. Theory Comput.* **2009**, *5*, 1838−1848.

(26) Olsen, J. M. H.; Kongsted, J. Molecular properties through polarizable embedding. *Adv. Quantum Chem.* **2011**, *61*, 107−143.

(27) Giovannini, T.; Puglisi, A.; Ambrosetti, M.; Cappelli, C. Polarizable QM/MM approach with fluctuating charges and fluctuating dipoles: the QM/FQFμ model. *J. Chem. Theory Comput.* **2019**, *15*, 2233−2245.

(28) Giovannini, T.; Lafiosca, P.; Cappelli, C. A General Route to Include Pauli Repulsion and Quantum Dispersion Effects in QM/MM Approaches. *J. Chem. Theory Comput.* **2017**, *13*, 4854−4870.

(29) Giovannini, T.; Lafiosca, P.; Chandramouli, B.; Barone, V.; Cappelli, C. Effective yet Reliable Computation of Hyperfine Coupling Constants in Solution by a QM/MM Approach: Interplay Between Electrostatics and Non-electrostatic Effects. *J. Chem. Phys.* **2019**, *150*, 124102.

(30) Giovannini, T.; Ambrosetti, M.; Cappelli, C. Quantum Confinement Effects on Solvatochromic Shifts of Molecular Solutes. *J. Phys. Chem. Lett.* **2019**, *10*, 5823−5829.

(31) Gordon, M. S.; Smith, Q. A.; Xu, P.; Slipchenko, L. V. Accurate first principles model potentials for intermolecular interactions. *Annu. Rev. Phys. Chem.* **2013**, *64*, 553−578.

(32) Gordon, M. S.; Slipchenko, L.; Li, H.; Jensen, J. H. The effective fragment potential: a general method for predicting intermolecular interactions. *Annu. Rep. Comput. Chem.* **2007**, *3*, 177−193.

(33) Sun, Q.; Chan, G. K.-L. Quantum embedding theories. *Acc. Chem. Res.* **2016**, *49*, 2705−2712.

(34) Knizia, G.; Chan, G. K.-L. Density matrix embedding: A strong-coupling quantum embedding theory. *J. Chem. Theory Comput.* **2013**, *9*, 1428−1432.

(35) Chulhai, D. V.; Goodpaster, J. D. Projection-based correlated wave function in density functional theory embedding for periodic systems. *J. Chem. Theory Comput.* **2018**, *14*, 1928−1942.

(36) Chulhai, D. V.; Goodpaster, J. D. Improved accuracy and efficiency in quantum embedding through absolute localization. *J. Chem. Theory Comput.* **2017**, *13*, 1503−1508.

(37) Wen, X.; Graham, D. S.; Chulhai, D. V.; Goodpaster, J. D. Absolutely Localized Projection-Based Embedding for Excited States. *J. Chem. Theory Comput.* **2020**, *16*, 385−398.

(38) Ding, F.; Manby, F. R.; Miller, T. F., III Embedded mean-field theory with block-orthogonalized partitioning. *J. Chem. Theory Comput.* **2017**, *13*, 1605−1615.

(39) Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller, T. F., III Density functional theory embedding for correlated wavefunctions: Improved methods for open-shell systems and transition metal complexes. *J. Chem. Phys.* **2012**, *137*, 224113.

(40) Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller, T. F., III Accurate and systematically improvable density functional theory embedding for correlated wavefunctions. *J. Chem. Phys.* **2014**, *140*, 18A507.

(41) Manby, F. R.; Stella, M.; Goodpaster, J. D.; Miller, T. F., III A simple, exact density-functional-theory embedding scheme. *J. Chem. Theory Comput.* **2012**, *8*, 2564−2568.

(42) Goodpaster, J. D.; Ananth, N.; Manby, F. R.; Miller, T. F., III Exact nonadditive kinetic potentials for embedded density functional theory. *J. Chem. Phys.* **2010**, *133*, 084103.

(43) Zhang, K.; Ren, S.; Caricato, M. Multi-state QM/QM Extrapolation of UV/Vis Absorption Spectra with Point Charge Embedding. *J. Chem. Theory Comput.* **2020**, *16*, 4361−4372.

(44) Ramos, P.; Papadakis, M.; Pavanello, M. Performance of frozen density embedding for modeling hole transfer reactions. *J. Phys. Chem. B* **2015**, *119*, 7541−7557.

(45) Pavanello, M.; Neugebauer, J. Modelling charge transfer reactions with the frozen density embedding formalism. *J. Chem. Phys.* **2011**, *135*, 234103.

(46) Tamukong, P. K.; Khait, Y. G.; Hoffmann, M. R. Density differences in embedding theory with external orbital orthogonality. *J. Phys. Chem. A* **2014**, *118*, 9182−9200.

(47) Tamukong, P. K.; Khait, Y. G.; Hoffmann, M. R. Accurate dissociation of chemical bonds using DFT-in-DFT embedding theory with external orbital orthogonality. *J. Phys. Chem. A* **2017**, *121*, 256−264.

(48) Wesolowski, T. A. Hydrogen-bonding-induced shifts of the excitation energies in nucleic acid bases: an interplay between electrostatic and electron density overlap effects. *J. Am. Chem. Soc.* **2004**, *126*, 11444−11445.

(49) Bennie, S. J.; Curchod, B. F. E.; Manby, F. R.; Glowacki, D. R. Pushing the limits of EOM-CCSD with projector-based embedding for excitation energies. *J. Phys. Chem. Lett.* **2017**, *8*, 5559−5565.

(50) Lee, S. J. R.; Welborn, M.; Manby, F. R.; Miller, T. F., III Projection-based wavefunction-in-DFT embedding. *Acc. Chem. Res.* **2019**, *52*, 1359−1368.

(51) 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.

(52) Wesolowski, T. A.; Shedge, S.; Zhou, X. Frozen-density embedding strategy for multilevel simulations of electronic structure. *Chem. Rev.* **2015**, *115*, 5891−5928.

(53) Fux, S.; Jacob, C. R.; 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.

(54) 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.

(55) Jacob, C. R.; Visscher, L. Calculation of nuclear magnetic resonance shieldings using frozen-density embedding. *J. Chem. Phys.* **2006**, *125*, 194104.

(56) 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.

(57) Wesołowski, T. A. Embedding a multideterminantal wave function in an orbital-free environment. *Phys. Rev. A: At., Mol., Opt. Phys.* **2008**, *77*, 012504.

(58) Sæther, S.; Kjærgaard, T.; Koch, H.; Høyvik, I.-M. Density-Based Multilevel Hartree−Fock Model. *J. Chem. Theory Comput.* **2017**, *13*, 5282−5290.

(59) Høyvik, I.-M. Convergence acceleration for the multilevel Hartree−Fock model. *Mol. Phys.* **2020**, *118*, 1626929.

(60) Myhre, R. H.; Sánchez de Merás, A. M. J.; Koch, H. Multi-level coupled cluster theory. *J. Chem. Phys.* **2014**, *141*, 224105.

(61) Wesolowski, T. A.; Warshel, A. Frozen density functional approach for ab initio calculations of solvated molecules. *J. Phys. Chem.* **1993**, *97*, 8050−8053.

(62) Folkestad, S. D.; Koch, H. Equation-of-motion MLCCSD and CCSD-in-HF oscillator strengths and their application to core excitations. *J. Chem. Theory Comput.* **2020**, *16*, 6869.

(63) Giovannini, T.; Koch, H. Energy-Based Molecular Orbital Localization in a Specific Spatial Region. *J. Chem. Theory Comput.* **2020**, *17*, 139−150.

(64) Sánchez de Merás, A. M. J.; Koch, H.; Cuesta, I. G.; Boman, L. Cholesky decomposition-based definition of atomic subsystems in electronic structure calculations. *J. Chem. Phys.* **2010**, *132*, 204105.

(65) He, X.; Merz, K. M., Jr. Divide and conquer Hartree- Fock calculations on proteins. *J. Chem. Theory Comput.* **2010**, *6*, 405−411.

(66) Folkestad, S. D.; Kjønstad, E. F.; Myhre, R. H.; Andersen, J. H.; Balbi, A.; Coriani, S.; Giovannini, T.; Goletto, L.; Haugland, T. S.;

Hutcheson, A.; Høyvik, I.-M.; Moitra, T.; Paul, A. C.; Scavino, M.; Skeidsvoll, A. S.; Tveten, Å. H.; Koch, H. eT 1.0: An open source electronic structure program with emphasis on coupled cluster and multilevel methods. *J. Chem. Phys.* **2020**, *152*, 184103.

(67) Lebedev, V. I.; Laikov, D. *Doklady Mathematics*; Pleiades Publishing, Ltd., 1999; Vol. 59, pp 477−481.

(68) Krack, M.; Köster, A. M. An adaptive numerical integrator for molecular integrals. *J. Chem. Phys.* **1998**, *108*, 3226−3234.

(69) Marques, M. A. L.; Oliveira, M. J. T.; Burnus, T. Libxc: A library of exchange and correlation functionals for density functional theory. *Comput. Phys. Commun.* **2012**, *183*, 2272−2281.

(70) Lehtola, S.; Steigemann, C.; Oliveira, M. J. T.; Marques, M. A. L. Recent developments in libxc—A comprehensive library of functionals for density functional theory. *SoftwareX* **2018**, *7*, 1−5.

(71) Aquilante, F.; Boman, L.; Boström, J.; Koch, H.; Lindh, R.; de Merás, A. S.; Pedersen, T. B. *Linear-Scaling Techniques in Computational Chemistry and Physics*; Springer, 2011; pp 301−343.

(72) Koch, H.; Sánchez de Merás, A.; Pedersen, T. B. Reduced scaling in electronic structure calculations using Cholesky decompositions. *J. Chem. Phys.* **2003**, *118*, 9481−9484.

(73) Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965**, *140*, A1133.

(74) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(75) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(76) Giovannini, T.; Egidi, F.; Cappelli, C. Molecular Spectroscopy of Aqueous Solutions: A The- oretical Perspective. *Chem. Soc. Rev.* **2020**, *49*, 5664−5677.

(77) Giovannini, T.; Egidi, F.; Cappelli, C. Theory and algorithms for chiroptical properties and spectroscopies of aqueous systems. *Phys. Chem. Chem. Phys.* **2020**, *22*, 22864−22879.

(78) Giovannini, T.; Riso, R. R.; Ambrosetti, M.; Puglisi, A.; Cappelli, C. Electronic transitions for a fully polarizable qm/mm approach based on fluctuating charges and fluctuating dipoles: linear and corrected linear response regimes. *J. Chem. Phys.* **2019**, *151*, 174104.

(79) Giovannini, T.; Del Frate, G.; Lafiosca, P.; Cappelli, C. Effective computational route towards vibrational optical activity spectra of chiral molecules in aqueous solution. *Phys. Chem. Chem. Phys.* **2018**, *20*, 9181−9197.

(80) Lipparini, F.; Cappelli, C.; Barone, V. A gauge invariant multiscale approach to magnetic spectroscopies in condensed phase: General three-layer model, computational implementation and pilot applications. *J. Chem. Phys.* **2013**, *138*, 234108.

(81) Giovannini, T.; Grazioli, L.; Ambrosetti, M.; Cappelli, C. Calculation of ir spectra with a fully polarizable qm/mm approach based on fluctuating charges and fluctuating dipoles. *J. Chem. Theory Comput.* **2019**, *15*, 5495−5507.

(82) Giovannini, T.; Olszówka, M.; Egidi, F.; Cheeseman, J. R.; Scalmani, G.; Cappelli, C. Polarizable Embedding Approach for the Analytical Calculation of Raman and Raman Optical Activity Spectra of Solvated Systems. *J. Chem. Theory Comput.* **2017**, *13*, 4421−4435.

(83) Losada, M.; Nguyen, P.; Xu, Y. Solvation of propylene oxide in water: Vibrational circular dichroism, optical rotation, and computer simulation studies. *J. Phys. Chem. A* **2008**, *112*, 5621−5627.

(84) Yang, G.; Xu, Y. Probing chiral solute-water hydrogen bonding networks by chirality transfer effects: a vibrational circular dichroism study of glycidol in water. *J. Chem. Phys.* **2009**, *130*, 164506.

(85) Merten, C.; Bloino, J.; Barone, V.; Xu, Y. Anharmonicity Effects in the Vibrational CD Spectra of Propylene Oxide. *J. Phys. Chem. Lett.* **2013**, *4*, 3424−3428.

(86) Su, Z.; Xu, Y. Hydration of a Chiral Molecule: The Propylene Oxide⋯ (Water)$_2$ Cluster in the Gas Phase. *Angew. Chem., Int. Ed.* **2007**, *119*, 6275−6278.

(87) Yu, Z.; Xu, L.; Wei, Y.; Wang, Y.; He, Y.; Xia, Q.; Zhang, X.; Liu, Z. A new route for the synthesis of propylene oxide from bio-glycerol derivated propylene glycol. *Chem. Commun.* **2009**, 3934−3936.

(88) Su, Z.; Borho, N.; Xu, Y. Chiral self-recognition: Direct spectroscopic detection of the homochiral and heterochiral dimers of propylene oxide in the gas phase. *J. Am. Chem. Soc.* **2006**, *128*, 17126−17131.

(89) Su, Z.; Wen, Q.; Xu, Y. Conformational Stability of the Propylene Oxide- Water Adduct: Direct Spectroscopic Detection of O- H⋯ O Hydrogen Bonded Conformers. *J. Am. Chem. Soc.* **2006**, *128*, 6755−6760.

(90) Perera, A. S.; Thomas, J.; Poopari, M. R.; Xu, Y. The clusters-in-a-liquid approach for solvation: new insights from the conformer specific gas phase spectroscopy and vibrational optical activity spectroscopy. *Front. Chem.* **2016**, *4*, 9.

(91) Lipparini, F.; Egidi, F.; Cappelli, C.; Barone, V. The optical rotation of methyloxirane in aqueous solution: a never ending story? *J. Chem. Theory Comput.* **2013**, *9*, 1880−1884.

(92) Giovannini, T.; Olszòwka, M.; Cappelli, C. Effective Fully Polarizable QM/MM Approach To Model Vibrational Circular Dichroism Spectra of Systems in Aqueous Solution. *J. Chem. Theory Comput.* **2016**, *12*, 5483−5492.