First-principle molecular dynamics with ultrasoft pseudopotentials: Parallel implementation and application to extended bioinorganic systems

P. Giannozzi[a)]
Department of Chemistry and Princeton Materials Institute, Princeton University, Princeton, New Jersey 08544 and NEST-INFM, Scuola Normale Superiore di Pisa, I-56126 Pisa, Italy

F. De Angelis
Department of Chemistry and Princeton Materials Institute, Princeton University, Princeton, New Jersey 08544 and Istituto CNR di Scienze e Tecnologie Molecolari (ISTM), Dipartimento di Chimica, Università di Perugia, Perugia, Italy

R. Car
Department of Chemistry and Princeton Materials Institute, Princeton University, Princeton, New Jersey 08544

(Received 14 November 2003; accepted 8 January 2004)

We present a plane-wave ultrasoft pseudopotential implementation of first-principle molecular dynamics, which is well suited to model large molecular systems containing transition metal centers. We describe an efficient strategy for parallelization that includes special features to deal with the augmented charge in the contest of Vanderbilt's ultrasoft pseudopotentials. We also discuss a simple approach to model molecular systems with a net charge and/or large dipole/quadrupole moments. We present test applications to manganese and iron porphyrins representative of a large class of biologically relevant metalorganic systems. Our results show that accurate density-functional theory calculations on systems with several hundred atoms are feasible with access to moderate computational resources. © *2004 American Institute of Physics.* [DOI: 10.1063/1.1652017]

I. INTRODUCTION

There is increasing interest in studying the electronic structure of complex biological molecules. This is an essential step to understand, e.g., enzymatic and/or biomimetic catalysis. Modeling biocatalytic systems is however very challenging, because a proper description of the active site needs the inclusion of a large number of atoms (from several tens to a few hundreds) treated at a high level of quantum chemical theory. In this respect a good compromise in terms of accuracy and computational cost is provided by density-functional theory (DFT),[1,2] whose use to model the electronic structure of protein active sites is becoming increasingly popular. In combination with Car–Parrinello (CP)[3] first-principle molecular dynamics (MD), DFT allows us to optimize molecular structures, study dynamical and finite-temperature properties, and model reaction paths.

In most standard implementations, the CP method employs a plane-wave (PW) basis set. An advantage of PWs is that they do not depend on atomic positions and are free of basis-set superposition errors. Total energies and forces on the atoms can be calculated using computationally efficient fast Fourier transform (FFT) techniques. Finally, the convergence of a calculation can be controlled in a simple way, since it depends only upon the number of PWs included in the expansion of the electron density. The dimension of a PW basis set is controlled by a cutoff in the kinetic energy of the PWs, which is usually measured in Ry units. A disadvantage of PWs is their extremely slow convergence in describing core states. To deal with this difficulty, one usually employs norm-conserving (NC) pseudopotentials (PPs)[4] to model the interaction of the valence electrons with the ionic core (nucleus+core electrons). Parallel implementations of PW calculations based on NC PPs are well documented in the literature (see, e.g., Refs. 5–7).

For many elements such as Al, Si, and P, it is possible to build accurate ''soft'' NC PPs, i.e., PPs whose pseudo-orbitals are slowly varying in real space. Typical soft PPs can be adequately described by PWs up to a cutoff of 20 Ry or less. For other elements, however, this is not possible: very large PW basis sets are needed to accurately represent the contracted p orbitals of the first-row elements O, N, F, and the $3d$ orbitals of the transition metal block. These orbitals belong to elements whose NC PPs are necessarily ''hard,'' typically requiring cutoffs of more than 70 Ry in order to yield sufficiently converged results. A consequence of the delocalized nature of the PWs is that the presence of a single hard PP in a system requires the use of a correspondingly high cutoff for all the other PPs. This difficulty is particularly serious for metalorganic systems containing one or more transition metal centers. The high cutoff required for such atoms translates into a very large number of PWs, which in turn implies long execution times and large memory requirements.

An approach that drastically reduces the PW cutoff needed for elements that would be described by hard NC PPs was proposed by Vanderbilt,[8] who introduced ''ultrasoft'' (US) PPs. In this approach the normalized charge density is the sum of two terms: a soft part represented in terms of

a)Electronic mail: giannozz@nest.sns.it

smooth orbitals and a hard part which is treated as an augmented charge. A closely related approach, the ''projector-augmented wave'' (PAW) method introduced by Blöchl,[9] is an all-electron rather than a PP electronic structure method. The PAW approach provides a simple and effective algorithm for reconstructing all-electron orbitals from pseudo-orbitals.[10] An efficient serial implementation of the CP scheme with US PPs is described in Ref. 11.

In this paper we present in detail a parallel implementation of the CP scheme using US PPs. We also provide further details on the procedures used in serial implementation. We compare the relative efficiency of US and NC PPs in realistic calculations for large molecules, performed on parallel machines. Our test molecules are a reduced and an extended model of the active site of myoglobin, containing the iron–porphyrin motif. We focus on metalloporphyrin systems because they are representative of a large class of biomolecules which can be modeled efficiently with US PPs. Our benchmarks indicate that US calculations are at least 2–3 times less expensive than NC calculations of comparable accuracy.

The use of a PW basis set implies that periodic boundary conditions (PBCs) are imposed, i.e., an isolated molecule has to be placed into a periodically repeated box (a ''supercell''). The supercell must be large enough to ensure that the total potential is vanishingly small at the box boundary, thus minimizing spurious interactions between periodic replicas. For neutral systems with small dipole/quadrupole moments, supercells of reasonable size can be safely used. For charged molecules, or molecules with large dipole/quadrupole moments, however, the error induced by PBC may be rather large unless exceedingly large supercells are used.

We follow here a technique introduced by Makov and Payne (MP)[12] to eliminate the spurious electrostatic interactions in the latter case. The MP technique is approximate because it is not self-consistent and takes into account only moments up to quadrupole. To check the accuracy of the MP technique, we compare CP calculations on highly charged manganese porphyrins, performed using PWs and US PPs, with calculations on the same systems using localized basis sets which do not require PBC. The comparison shows that the MP correction yields results that to all practical effects are indistinguishable from results obtained without PBC.

The paper is organized as follows: In Sec. II, we recall the main aspects of US–PP implementation in the serial case. In Sec. III, we describe our parallel implementation. In Sec. IV, we compare the computer performances of US and NC PPs for a reduced and an extended model of the myoglobin active site. In Sec. V, we compare CP calculations with localized basis-set calculations not requiring PBC. The test systems are highly charged isomeric *meso*-substituted manganese porphyrins. Section VI contains our conclusions.

II. PLANE-WAVE ULTRASOFT PSEUDOPOTENTIAL IMPLEMENTATION

A. Kohn–Sham equations with ultrasoft pseudopotentials

The implementation of CP molecular dynamics within a US–PP framework is described in Ref. 11. Here we briefly note the main formulas, using the same notation of Ref. 11.

The total energy of a system of N_v valence electrons, having one-electron Kohn–Sham (KS) orbitals ϕ_i, is given by

$$E_{\rm tot}[\{\phi_i\},\{\mathbf{R}_I\}]=\sum_i \langle\phi_i|-\frac{\hbar^2}{2m}\nabla^2+V_{\rm NL}|\phi_i\rangle+E_H[n] +E_{\rm xc}[n]+\int d\mathbf{r}V_{\rm loc}^{\rm ion}(\mathbf{r})n(\mathbf{r})+U(\{\mathbf{R}_I\}), \tag{1}$$

where $n(\mathbf{r})$ is the electron density, $E_H[n]$ is the Hartree energy

$$E_H[n]=\frac{e^2}{2}\int\int d\mathbf{r}d\mathbf{r}'\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}, \tag{2}$$

$E_{\rm xc}[n]$ is the exchange and correlation energy, $U(\{\mathbf{R}_I\})$ is the ion–ion interaction energy, and $\mathbf{R}_I$ indicate atomic positions. In the following, potentials have energy dimensions. The PP is composed of a local part $V_{\rm loc}^{\rm ion}$, given by a sum of atom-centered radial potentials

$$V_{\rm loc}^{\rm ion}(\mathbf{r})=\sum_I V_{\rm loc}^I(|\mathbf{r}-\mathbf{R}_I|) \tag{3}$$

and a nonlocal part $V_{\rm NL}$

$$V_{\rm NL}=\sum_{nm,I} D_{nm}^{(0)}|\beta_n^I\rangle\langle\beta_m^I|, \tag{4}$$

where the functions β_n^I and the coefficients $D_{nm}^{(0)}$ characterize the PP and are specific for each atomic species. For simplicity, we consider a single atomic species only in what follows. The β_n^I functions, centered at site $\mathbf{R}_I$, depend on the ionic positions through

$$\beta_n^I(\mathbf{r})=\beta_n(\mathbf{r}-\mathbf{R}_I). \tag{5}$$

β_n is an angular momentum eigenfunction in the angular variables, times a radial function which vanishes outside the core region; the indices n and m in Eq. (4) run over the total number N_β of these functions.

The electron density in Eq. (1) is given by

$$n(\mathbf{r})=\sum_i\left[|\phi_i(\mathbf{r})|^2+\sum_{nm,I}Q_{nm}^I(\mathbf{r})\langle\phi_i|\beta_n^I\rangle\langle\beta_m^I|\phi_i\rangle\right], \tag{6}$$

where the sum runs over occupied KS orbitals. The augmentation functions $Q_{nm}^I(\mathbf{r})=Q_{nm}(\mathbf{r}-\mathbf{R}_I)$ are localized in the core. The ultrasoft PP is fully determined by the quantities $V_{\rm loc}^I(r)$, $D_{nm}^{(0)}$, $Q_{nm}(\mathbf{r})$, and $\beta_n(\mathbf{r})$. The functions $Q_{nm}(\mathbf{r})$ are defined in terms of atomic orbitals as: $Q_{nm}(\mathbf{r})=\psi_n^{ae*}(\mathbf{r})\psi_m^{ae}(\mathbf{r})-\psi_n^{ps*}(\mathbf{r})\psi_m^{ps}(\mathbf{r})$, where ψ^{ae} are atomic one-electron orbitals (not necessarily bound), and ψ^{ps} are the corresponding pseudo-orbitals. The $Q_{nm}(\mathbf{r})$ are pseudized as described in Ref. 11. This enables us to treat the $Q_{nm}(\mathbf{r})$ with Fourier transform techniques.

The KS orbitals obey generalized orthonormality conditions

$$\langle\phi_i|S(\{\mathbf{R}_I\})|\phi_j\rangle=\delta_{ij}, \tag{7}$$

where S is a Hermitian overlap operator given by

$$S=1+\sum_{nm,I} q_{nm}|\beta_n^I\rangle\langle\beta_m^I|, \tag{8}$$

and

$$q_{nm}=\int d\mathbf{r}Q_{nm}(\mathbf{r}). \tag{9}$$

The orthonormality condition (7) is consistent with the conservation of the charge $\int d\mathbf{r}n(\mathbf{r})=N_v$. Note that the overlap operator S depends on ionic positions through the $|\beta_n^I\rangle$.

The ground-state orbitals ϕ_i minimize the total energy Eq. (1) subject to the constraints Eq. (7)

$$\frac{\delta E_{\rm tot}}{\delta\phi_i^*(\mathbf{r})}=\epsilon_i S\phi_i(\mathbf{r}), \tag{10}$$

where the ϵ_i are Lagrange multipliers. Equation (11) is the KS equation

$$H|\phi_i\rangle=\epsilon_i S|\phi_i\rangle, \tag{11}$$

where

$$H=-\frac{\hbar^2}{2m}\nabla^2+V_{\rm eff}+\sum_{nm,I} D_{nm}^I|\beta_n^I\rangle\langle\beta_m^I|. \tag{12}$$

$V_{\rm eff}$ is a screened effective local potential

$$V_{\rm eff}(\mathbf{r})=V_{\rm loc}^{\rm ion}(\mathbf{r})+V_H(\mathbf{r})+\mu_{\rm xc}(\mathbf{r}). \tag{13}$$

$\mu_{\rm xc}(\mathbf{r})$ is the exchange-correlation potential

$$\mu_{\rm xc}(\mathbf{r})=\frac{\delta E_{\rm xc}[n]}{\delta n(\mathbf{r})}, \tag{14}$$

and $V_H(\mathbf{r})$ is the Hartree potential

$$V_H(\mathbf{r})=e^2\int d\mathbf{r}'\frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}. \tag{15}$$

The ''screened'' coefficients D_{nm}^I appearing in Eq. (12) are defined as

$$D_{nm}^I=D_{nm}^{(0)}+\int d\mathbf{r}V_{\rm eff}(\mathbf{r})Q_{nm}^I(\mathbf{r}). \tag{16}$$

They depend on the KS orbitals through $V_{\rm eff}$, Eq. (13), and the charge density, Eq. (6).

B. Molecular dynamics with ultrasoft pseudopotentials

In the CP approach,[3] the electronic orbitals and the ionic coordinates evolve according to a classical Lagrangian

$$\mathcal{L}=\mu\sum_i\int d\mathbf{r}|\dot\phi_i(\mathbf{r})|^2+\frac{1}{2}\sum_I M_I\dot{\mathbf{R}}_I^2-E_{\rm tot}(\{\phi_i\},\{\mathbf{R}_I\}), \tag{17}$$

subject to a set of constraints

$$\mathcal{N}_{ij}(\{\phi_i\},\{\mathbf{R}_I\})=\langle\phi_i|S|\phi_j\rangle-\delta_{ij}=0. \tag{18}$$

Here μ is a fictitious mass parameter for the electronic degrees of freedom, M_I are the masses of the atoms, and $E_{\rm tot}$ and S are given in Eqs. (1) and (8), respectively. The holonomic orthonormality constraints Eq. (18) do not cause energy dissipation in a MD run. They may be incorporated in the Euler equations of motion by introducing Lagrange multipliers Λ_{ij}:

$$\mu\ddot\phi_i=-\frac{\delta E_{\rm tot}}{\delta\phi_i^*}+\sum_j \Lambda_{ij}S\phi_j, \tag{19}$$

$$\mathbf{F}_I=M_I\ddot{\mathbf{R}}_I=-\frac{\partial E_{\rm tot}}{\partial\mathbf{R}_I}+\sum_{ij}\Lambda_{ij}\langle\phi_i|\frac{\partial S}{\partial\mathbf{R}_I}|\phi_j\rangle. \tag{20}$$

At equilibrium, Eq. (19) reduces to the electronic KS equations (10) or (11). A unitary rotation brings the Λ matrix into diagonal form: $\Lambda_{ij}=\epsilon_i\delta_{ij}$. The equilibrium for the ions is achieved when the ionic forces $\mathbf{F}_I$ in Eq. (20) vanish.

In deriving explicit expressions for the forces, Eq. (20), one should keep in mind that the electron density also depends on $\mathbf{R}_I$ through Q_{nm}^I and β_n^I. Introducing the quantities

$$\rho_{nm}^I=\sum_i\langle\phi_i|\beta_n^I\rangle\langle\beta_m^I|\phi_i\rangle, \tag{21}$$

and

$$\omega_{nm}^I=\sum_{ij}\Lambda_{ij}\langle\phi_j|\beta_n^I\rangle\langle\beta_m^I|\phi_i\rangle, \tag{22}$$

we arrive at the expression

$$\begin{aligned}\mathbf{F}_I=&-\frac{\partial U}{\partial\mathbf{R}_I}-\int d\mathbf{r}\frac{\partial V_{\rm loc}^{\rm ion}}{\partial\mathbf{R}_I}n(\mathbf{r})\\&-\int d\mathbf{r}V_{\rm eff}(\mathbf{r})\sum_{nm}\frac{\partial Q_{nm}^I(\mathbf{r})}{\partial\mathbf{R}_I}\rho_{nm}^I-\sum_{nm}D_{nm}^I\frac{\partial\rho_{nm}^I}{\partial\mathbf{R}_I}\\&+\sum_{nm}q_{nm}\frac{\partial\omega_{nm}^I}{\partial\mathbf{R}_I},\end{aligned} \tag{23}$$

where D_{nm}^I and $V_{\rm eff}$ have been defined in Eqs. (16) and (13), respectively. The last term of Eq. (23) gives the constraint contribution to the forces. Since the PW basis set does not depend on atomic positions, Pulay-type corrections[13] do not appear in the expression for the forces.

C. Discretization of the equation of motion and orthonormality constraints

The equations of motion Eqs. (19) and (20) are usually discretized using the Verlet or the velocity-Verlet algorithms. The following discussion, including the treatment of the $\mathbf{R}_I$ dependence of the orthonormality constraints, applies to the Verlet algorithm when using the Fourier acceleration scheme of Ref. 14. In this framework the fictitious electron mass is represented by an operator Θ, whose matrix elements between PWs are given by

$$\Theta_{\mathbf{G},\mathbf{G}'}=\max\left(\mu,\mu\frac{\hbar^2G^2}{2mE_c}\right)\delta_{\mathbf{G},\mathbf{G}'}, \tag{24}$$

where $\mathbf{G}$, $\mathbf{G}'$ are the wave vector of PWs, E_c is a cutoff (typically a few Ry) which defines the threshold for Fourier acceleration. The fictitious electron mass depends on G as

the kinetic energy for large G, it is constant for small G. This scheme allows us to use larger time steps with negligible computational overhead.

The electronic orbitals at time $t+\Delta t$ are given by

$$\phi_i(t+\Delta t)=2\phi_i(t)-\phi_i(t-\Delta t)-(\Delta t)^2\Theta^{-1}\left[\frac{\delta E_{\rm tot}}{\delta\phi_i^*}-\sum_j \Lambda_{ij}(t+\Delta t)S(t)\phi_j(t)\right], \tag{25}$$

where Δt is the time step, and $S(t)$ indicates the operator S evaluated for ionic positions $R_I(t)$. Similarly the ionic coordinates at time $t+\Delta t$ are given by

$$\mathbf{R}_I(t+\Delta t)=2\mathbf{R}_I(t)-\mathbf{R}_I(t-\Delta t)-\frac{(\Delta t)^2}{M_I}\left[\frac{\partial E_{\rm tot}}{\partial \mathbf{R}_I}-\sum_{ij}\Lambda_{ij}(t+\Delta t)\langle\phi_i(t)|\frac{\partial S(t)}{\partial \mathbf{R}_I}|\phi_j(t)\rangle\right]. \tag{26}$$

The orthonormality conditions must be imposed at each time step

$$\langle\phi_i(t+\Delta t)|S(t+\Delta t)|\phi_j(t+\Delta t)\rangle=\delta_{ij}, \tag{27}$$

leading to the following matrix equation:

$$A+\lambda B+B^\dagger\lambda^\dagger+\lambda C\lambda^\dagger=1, \tag{28}$$

where the unknown matrix λ is related to the matrix of Lagrange multipliers Λ at time $t+\Delta t$ via $\lambda=(\Delta t)^2\Lambda^*(t+\Delta t)$. In Eq. (28) the dagger indicates Hermitian conjugate ($\lambda=\lambda^\dagger$). The matrices A, B, and C are given by

$$\begin{aligned}A_{ij}&=\langle\overline{\phi_i}|S(t+\Delta t)|\overline{\phi_j}\rangle,\\ B_{ij}&=\langle\Theta^{-1}S(t)\phi_i(t)|S(t+\Delta t)|\overline{\phi_j}\rangle,\\ C_{ij}&=\langle\Theta^{-1}S(t)\phi_i(t)|S(t+\Delta t)|\Theta^{-1}S(t)\phi_j(t)\rangle,\end{aligned} \tag{29}$$

with

$$\overline{\phi_i}=2\phi_i(t)-\phi_i(t-\Delta t)-(\Delta t)^2\Theta^{-1}\frac{\delta E_{\rm tot}(t)}{\delta\phi_i^*}. \tag{30}$$

The solution of Eq. (28) in the ultrasoft PP case is not obvious, because Eq. (26) is not a closed expression for $\mathbf{R}_I(t+\Delta t)$. The problem is that $\Lambda(t+\Delta t)$ appearing in Eq. (26) depends implicitly on $\mathbf{R}_I(t+\Delta t)$ through $S(t+\Delta t)$. Consequently, it is in principle necessary to solve iteratively for $\mathbf{R}_I(t+\Delta t)$ in Eq. (26).

A simple solution to this problem is given in Ref. 11. $\Lambda(t+\Delta t)$ is extrapolated using two previous values

$$\Lambda_{ij}^{(0)}(t+\Delta t)=2\Lambda_{ij}(t)-\Lambda_{ij}(t-\Delta t). \tag{31}$$

Equation (26) is used to find $\mathbf{R}_I^{(0)}(t+\Delta t)$, which is correct to $O(\Delta t^4)$. From $\mathbf{R}_I^{(0)}(t+\Delta t)$ we can obtain a new set $\Lambda_{ij}^{(1)}(t+\Delta t)$ and repeat the procedure until convergence is achieved. It turns out that in most practical applications the procedure converges at the very first iteration. Thus, the operations described above are generally executed only once per time step.

The solution of Eq. (28) is found using a modified version[7,11] of the iterative procedure of Ref. 15. The matrix B is decomposed into hermitian (B_h) and antihermitian (B_a) parts

$$B=B_h+B_a, \tag{32}$$

and the solution is obtained by iteration

$$\lambda^{(n+1)}B_h+B_h\lambda^{(n+1)}=1-A-\lambda^{(n)}B_a-B_a^\dagger\lambda^{(n)}-\lambda^{(n)}C\lambda^{(n)}. \tag{33}$$

The initial guess $\lambda^{(0)}$ can be obtained from

$$\lambda^{(0)}B_h+B_h\lambda^{(0)}=1-A. \tag{34}$$

Here the B_a- and C-dependent terms are neglected because they are of higher order in Δt (B_a vanishes for vanishing Δt). Equations (34) and (33) have the same structure:

$$\lambda B_h+B_h\lambda=X, \tag{35}$$

where X is a Hermitian matrix. Equation (35) can be solved exactly by finding the unitary matrix U that diagonalizes B_h: $U^\dagger B_h U=D$, where $D_{ij}=d_i\delta_{ij}$. The solution is obtained from

$$(U^\dagger\lambda U)_{ij}=(U^\dagger X U)_{ij}/(d_i+d_j). \tag{36}$$

When $X=1-A$ Eq. (36) yields the starting $\lambda^{(0)}$, while $\lambda^{(n+1)}$ is obtained from $\lambda^{(n)}$ by solving Eq. (36) with X given by Eq. (33). This iterative procedure usually converges in ten steps or less.

D. Ultrasoft pseudopotential implementation in the serial case

1. Plane-wave expansion

Let $\{\mathbf{R}\}$ be the translation vectors of the periodically repeated supercell. The corresponding reciprocal lattice vectors $\{\mathbf{G}\}$ obey the conditions $\mathbf{R}_i\cdot\mathbf{G}_j=2\pi n$, with n an integer number.

The KS orbitals can be expanded in a PW basis up to a kinetic energy cutoff $E_c^{\rm wf}$

$$\phi_{j,\mathbf{k}}(\mathbf{r})=\frac{1}{\sqrt{\Omega}}\sum_{\mathbf{G}\in\{\mathbf{G}_c^{\rm wf}\}}\phi_{j,\mathbf{k}}(\mathbf{G})e^{-i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}, \tag{37}$$

where Ω is the volume of the cell, and $\{\mathbf{G}_c^{\rm wf}\}$ is the set of $\mathbf{G}$ vectors satisfying the condition

$$\frac{\hbar^2}{2m}|\mathbf{k}+\mathbf{G}|^2<E_c^{\rm wf}, \tag{38}$$

and $\mathbf{k}$ is the Bloch vector of the electronic states. In crystals, one must use a grid of $\mathbf{k}$ points dense enough to sample the Brillouin zone (the unit cell of the reciprocal lattice). In molecules, liquids, and in general if the simulation cell is large enough, the Brillouin zone can be sampled using only the $\mathbf{k}=0$ (Γ) point. An advantage of this choice is that the orbitals can be taken to be real in $\mathbf{r}$ space. In the following we will drop the $\mathbf{k}$ vector index. Functions in real space and their Fourier transforms will be denoted by the symbols, if this does not originate in ambiguity.

The $\phi_j(\mathbf{G})$s are the electronic variables. The calculation of $H\phi_j$ and of the forces acting on the ions are the basic ingredients of the computation. Scalar products $\langle \phi_j|\beta_n^I\rangle$ and their spatial derivatives are typically evaluated in **G** space. An important advantage of working in **G** space is that atom-centered functions like β_n^I and Q_{nm}^I are easily evaluated at any atomic position, for example

$$\beta_n^I(\mathbf{G})=\beta_n(\mathbf{G})e^{-i\mathbf{G}\cdot\mathbf{R}_I}. \tag{39}$$

Thus

$$\langle \phi_j|\beta_n^I\rangle=\sum_{\mathbf{G}\in\{\mathbf{G}_c^{\mathrm{wf}}\}}\phi_j^*(\mathbf{G})\beta_n(\mathbf{G})e^{-i\mathbf{G}\cdot\mathbf{R}_I}, \tag{40}$$

and

$$\left\langle \phi_j\left|\frac{\partial\beta_n^I}{\partial\mathbf{R}_I}\right.\right\rangle=-i\sum_{\mathbf{G}\in\{\mathbf{G}_c^{\mathrm{wf}}\}}\mathbf{G}\phi_j^*(\mathbf{G})\beta_n(\mathbf{G})e^{-i\mathbf{G}\cdot\mathbf{R}_I}. \tag{41}$$

The kinetic energy term is diagonal in **G** space and is easily calculated

$$-(\nabla^2\phi_j)(\mathbf{G})=G^2\phi_j(\mathbf{G}). \tag{42}$$

In summary, the kinetic and nonlocal PP terms in $H\phi_j$ are calculated in **G** space.

2. Dual space technique

The local potential term $V_{\mathrm{eff}}\phi_j$ could be calculated in **G** space, but it is more convenient to use a different (''dual space'') technique. The idea is to switch from **G** to **r** space, back and forth, using FFT, and to perform the calculation in the space where it is more convenient. The KS orbitals are first Fourier transformed to **r** space; then $(V_{\mathrm{eff}}\phi_j)(\mathbf{r})=V_{\mathrm{eff}}(\mathbf{r})\phi_j(\mathbf{r})$ is calculated in **r** space, where V_{eff} is diagonal; finally $(V_{\mathrm{eff}}\phi_j)(\mathbf{r})$ is Fourier transformed back to $(V_{\mathrm{eff}}\phi_j)(\mathbf{G})$.

In order to use FFT, one discretizes the **r** space by a uniform grid spanning the unit cell

$$\begin{aligned} &f(m_1,m_2,m_3)\equiv f(\mathbf{r}_{m_1,m_2,m_3}),\\ &\mathbf{r}_{m_1,m_2,m_3}=m_1\frac{\mathbf{a}_1}{N_1}+m_2\frac{\mathbf{a}_2}{N_2}+m_3\frac{\mathbf{a}_3}{N_3}, \end{aligned} \tag{43}$$

where $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ are lattice basis vectors, the integer index m_1 runs from 0 to N_1-1, and similarly for m_2 and m_3. In the following we will assume for simplicity that N_1, N_2, N_3 are even numbers. The FFT maps a discrete periodic function in real space $f(m_1,m_2,m_3)$ into a discrete periodic function in reciprocal space $\tilde{f}(n_1,n_2,n_3)$ (where n_1 runs from 0 to N_1-1, and similarly for n_2 and n_3), and vice versa.

The link between **G** space components and FFT indices is

$$\begin{aligned} &\tilde{f}(n_1,n_2,n_3)\equiv f(\mathbf{G}_{n_1',n_2',n_3'}),\\ &\mathbf{G}_{n_1',n_2',n_3'}=n_1'\mathbf{b}_1+n_2'\mathbf{b}_2+n_3'\mathbf{b}_3, \end{aligned} \tag{44}$$

where $n_1=n_1'$ if $n_1'\geqslant 0$, $n_1=n_1'+N_1$ if $n_1'<0$, and similarly for n_2 and n_3. The FFT dimensions N_1, N_2, N_3 must be big enough to include all non-negligible Fourier components of the function to be transformed: ideally the Fourier component corresponding to $n_1'=N_1/2$, and similarly for n_2' and n_3', should vanish. In the following, we will refer to the set of indices n_1, n_2, n_3 and to the corresponding Fourier components as the ''FFT grid.''

The soft part of the charge density: $n_{\mathrm{soft}}(\mathbf{r})=\Sigma_j|\phi_j(\mathbf{r})|^2$, contains Fourier components up to a kinetic energy cutoff $E_c^{\mathrm{soft}}=4E_c^{\mathrm{wf}}$. This is evident from the formula

$$n_{\mathrm{soft}}(\mathbf{G})=\sum_{\mathbf{G}'\in\{\mathbf{G}_c^{\mathrm{wf}}\}}\sum_j\phi_j^*(\mathbf{G}-\mathbf{G}')\phi_j(\mathbf{G}'). \tag{45}$$

In the case of NC PPs, the entire charge density is given by $n_{\mathrm{soft}}(\mathbf{r})$.

V_{eff} should be expanded up to the same E_c^{soft} cutoff since all the Fourier components of $V_{\mathrm{eff}}\phi_j$ up to E_c^{wf} are required. Let us call $\{\mathbf{G}_c^{\mathrm{soft}}\}$ the set of **G** vectors such that

$$\frac{\hbar}{2m}G^2<E_c^{\mathrm{soft}}. \tag{46}$$

The soft part of the charge density is conveniently calculated in **r** space, by Fourier transforming $\phi_j(\mathbf{G})$ into $\phi_j(\mathbf{r})$ and summing over the occupied states.

The exchange-correlation potential $\mu_{\mathrm{xc}}(\mathbf{r})$, Eq. (14), is a function of the local charge density and—for gradient-corrected functionals—of its gradient at point **r**

$$\mu_{\mathrm{xc}}(\mathbf{r})=V_{\mathrm{xc}}(n(\mathbf{r}),|\nabla n(\mathbf{r})|). \tag{47}$$

The gradient $\nabla n(\mathbf{r})$ is conveniently calculated from the charge density in **G** space, using $(\nabla n)(\mathbf{G})=-i\mathbf{G}n(\mathbf{G})$. The Hartree potential $V_H(\mathbf{r})$, Eq. (15), is also conveniently calculated in **G** space

$$V_H(\mathbf{G})=\frac{4\pi}{\Omega}\frac{n(\mathbf{G})^*}{G^2}. \tag{48}$$

Thus in the NC–PP case, a single FFT grid, large enough to accommodate the $\{\mathbf{G}_c^{\mathrm{soft}}\}$ set, can be used for orbitals, charge density, and potential.

The use of FFT is mathematically equivalent to a pure **G**-space description (we neglect here a small inconsistency in exchange–correlation potential and energy density, due to the presence of a small amount of components beyond the $\{\mathbf{G}_c^{\mathrm{soft}}\}$ set). This has important consequences: working in **G** space means that translational invariance is exactly conserved and that forces are analytical derivatives of the energy (apart from the effect of the small inconsistency mentioned above). Forces that are analytical derivatives of the energy ensure that the constant of motion (i.e., the sum of kinetic and potential energy of the ions in Newtonian dynamics) is conserved during the evolution.

3. Double-grid technique

Let us focus on US PPs. In **G** space the charge density is

$$n(\mathbf{G})=n_{\mathrm{soft}}(\mathbf{G})+\sum_{i,nm,I}Q_{mn}^I(\mathbf{G})\langle\phi_i|\beta_n^I\rangle\langle\beta_m^I|\phi_i\rangle. \tag{49}$$

If E_c^{wf} is the cutoff for the KS orbitals, the cutoff for the soft part of the charge density is $E_c^{\mathrm{soft}}=4E_c^{\mathrm{wf}}$. The augmentation

term often requires a cutoff higher than $E_c^{\rm soft}$, and as a consequence a larger set of **G** vectors. Let us call $\{\mathbf{G}_c^{\rm dens}\}$ the set of **G** vectors that arc needed for the augmented part

$$\frac{\hbar^2}{2m}G^2 < E_c^{\rm dens}. \tag{50}$$

In typical situations, using pseudized augmented charges, $E_c^{\rm dens}$ ranges from $E_c^{\rm soft}$ to $\sim 2-3E_c^{\rm soft}$.

The same FFT grid could be used for both the augmented charge density and for KS orbitals. This however would imply using an oversized FFT grid in the most expensive part of the calculation, dramatically increasing computer time. A better solution is to introduce two FFT grids:

(i) a coarser grid (in **r** space) for the KS orbitals and the soft part of the charge density. The FFT dimensions N_1, N_2, N_3 of this grid are big enough to accommodate all **G** vectors in $\{\mathbf{G}_c^{\rm soft}\}$; and
(ii) a denser grid (in **r** space) for the total charge density and the exchange–correlation and Hartree potentials. The FFT dimensions $M_1 \geqslant N_1$, $M_2 \geqslant N_2$, $M_3 \geqslant N_3$ of this grid are big enough to accommodate all **G** vectors in $\{\mathbf{G}_c^{\rm dens}\}$.

In this framework, the soft part of the electron density $n_{\rm soft}$, is calculated in **r** space using FFTs on the coarse grid and transformed in **G** space using a coarse-grid FFT on the $\{\mathbf{G}_c^{\rm soft}\}$ grid. The augmented charge density is calculated in **G** space on the $\{\mathbf{G}_c^{\rm dens}\}$ grid, using Eq. (49) as described in the next section. $n(\mathbf{G})$ is used to evaluate the Hartree potential, Eq. (48). Then $n(\mathbf{G})$ is Fourier transformed in **r** space on the dense grid, where the exchange–correlation potential, Eq. (47), is evaluated.

In real space, the two grids are not necessarily commensurate. Whenever the need arises to go from the coarse to the dense grid, or vice versa, this is done in **G** space. For instance, the potential $V_{\rm eff}$, Eq. (13), is needed both on the dense grid to calculate quantities such as the D_{nm}^I, Eq. (16), and on the coarse grid to calculate $V_{\rm eff}\phi_j$, Eq. (11). The connection between the two grids occurs in **G** space, where Fourier filtering is performed: $V_{\rm eff}$ is first transformed in **G** space on the dense grid, then transferred to the coarse **G** space grid by eliminating components incompatible with $E_c^{\rm soft}$, and then backtransformed in **r** space using a coarse-grid FFT.

We remark that for each time step only a few dense-grid FFT are performed, while the number of necessary coarse-grid FFTs is much larger, proportional to the number of KS states $N_{\rm ks}$.

4. Augmentation boxes

Let us consider the augmentation functions Q_{nm}, which appear in the calculation of the electron density, Eq. (49), in the calculation of D_{nm}^I, Eq. (16), and in the integrals involving $\partial Q_{nm}^I/\partial \mathbf{R}_I$ needed to compute the ionic forces, Eq. (23). The calculation of the Q_{nm} in **G** space has a large computational cost because the cutoff for the Q_{nm} is the large cutoff $E_c^{\rm dens}$. The computational cost can be significantly reduced if we take advantage of the localization of the Q_{nm} in the core region.

We call ''augmentation box'' a fraction of the supercell, containing a small portion of the dense grid in real space. An augmentation box is defined only for atoms described by US PPs. The augmentation box for atom I is centered at the point of the dense grid that is closer to the position $\mathbf{R}_I$. During a MD run, the center of the Ith augmentation box makes discontinuous jumps to one of the neighboring grid points whenever the position vector $\mathbf{R}_I$ gets closer to such grid point. In a MD run, the augmentation box must always completely contain the augmented charge belonging to the Ith atom; otherwise, the augmentation box must be as small as possible. Augmentation boxes of different sizes for different atoms could in principle be used, but in our implementation the same box size is chosen for all the atoms. Thus the atomic species having the less localized augmented charge determines the size of all the augmentation boxes.

The volume of the augmentation box is much smaller than the volume of the supercell. The number of **G** vectors in the reciprocal space of the augmentation box is smaller than the number of **G** vectors in the dense grid by the ratio of the volumes of the augmentation box and of the supercell. As a consequence, the cost of calculations on the augmentation boxes increases linearly with the number of atoms described by US PPs.

Augmentation boxes are used twice in the calculation:

(i) to construct the augmented charge density, Eq. (6) and
(ii) to calculate the self-consistent contribution to the coefficients of the nonlocal PP, Eq. (16).

In case (i), the augmented charge is conveniently calculated in **G** space, following Ref. 11, and Fourier transformed in **r** space. All these calculations are done on the augmentation box grid. Then the calculated contribution at each **r** point of the augmentation box grid is added to the charge density at the same point in the dense grid. In case (ii), it is convenient to calculate D_{nm}^I as follows: for every US atom, take the Fourier transform of $V_{\rm eff}(\mathbf{r})$ on the corresponding augmentation box grid and evaluate the integral of Eq. (16) in **G** space.

III. PARALLEL ULTRASOFT PSEUDOPOTENTIAL IMPLEMENTATION

Various parallelization strategies for PW–PP calculations have been described in the literature. A strategy that ensures excellent scalability in terms of both computer time and memory consists of distributing the PW basis set and the FFT grid points in real and reciprocal space across processors. A crucial issue for the success of this approach is the FFT algorithm, which must be capable of performing three-dimensional FFT on data shared across different processors with good load balancing.[5] This algorithm can be generalized to the US case as described in the following subsection.

A. Parallel FFT in the US case

Partitioning a real-space FFT grid across processors is straightforward. The FFT grid, Eq. (43), is subdivided in a number of slices equal to the number of processors, so that each processor can take care of a different slice. The slices are cut along planes orthogonal to a chosen crystallographic direction. We label the crystallographic directions by 1, 2, 3. For instance, let us consider a FFT grid with N_3 planes along direction 3, which is distributed across N_p processors. If N_p is a divisor of N_3, good load balancing is achieved if each slice contains the same number (N_3/N_p) of planes. Processor p contains planes with m_3 values such that: $(p-1)(N_3/N_p) \leqslant m_3 \leqslant p(N_3/N_p)-1$. If N_p is not a divisor of N_3, all the slices cannot be equal. In this case their dimension is chosen in such a way as to minimize load imbalance. If N_p exceeds the number of planes N_3, this strategy has to be refined.

The partition of the **G** space grid is more involved. The Fourier components of the quantities of interest (e.g., the orbitals, the charge density, etc.) are stored as vectors (one-dimensional arrays): $f(i) \equiv f(\mathbf{G}_i)$, where the index i spans one of the three sets of **G** vectors defined above, namely the set $\{\mathbf{G}_c^{\mathrm{wf}}\}$, the set $\{\mathbf{G}_c^{\mathrm{soft}}\}$, and the set $\{\mathbf{G}_c^{\mathrm{dens}}\}$. When a FFT is needed, the Fourier components have to be transferred to one of the two grids (three-dimensional arrays), defined by Eq. (44). The two grids are either the coarse grid, with dimensions N_1, N_2, N_3, or the dense grid, with dimensions M_1, M_2, M_3. The Fourier components must be evenly distributed across processors in order to achieve optimal load balancing for operations like scalar products. At the same time, their distribution across processors should achieve good load balancing in the FFTs *and* minimize the amount of data communication needed to perform the FFTs.

For each pair n_1', n_2' in Eq. (44) we define a "column" in **G** space, including all $\mathbf{G}_{n_1',n_2',n_3'}$ with $-M_3/2 \leqslant n_3' \leqslant M_3/2$. Since the KS orbitals have nonzero Fourier component only for **G** vectors belonging to the set $\{\mathbf{G}_c^{\mathrm{wf}}\}$, only a subset of all the columns contributes to a one-dimensional FFT of a KS orbital in the direction 3. We call these columns "active columns" for the set $\{\mathbf{G}_c^{\mathrm{wf}}\}$. In general, the number of nonzero Fourier components is different for each active column. Ideally, we would like to distribute the active columns across the processors, so that each processor receives the same number of active columns and the same number of Fourier components. Although not possible in general, this can be achieved to a good extent with a simple algorithm:[16] (1) create a list of columns, ordered by decreasing number of nonzero Fourier components; (2) assign the column to the processors, following the order in the list; (3) when all the processors contain at least one column, assign the following column in the list to the processor with the smallest number of nonzero Fourier components. This algorithm works nicely when the number of columns per processor is large enough.

After assigning to the processors all the columns that are active for the set $\{\mathbf{G}_c^{\mathrm{wf}}\}$, we distribute across the processors the remaining columns, that are active for the set $\{\mathbf{G}_c^{\mathrm{soft}}\}$, using the same algorithm. Finally we distribute across the processors the remaining columns, that are active for the set $\{\mathbf{G}_c^{\mathrm{dens}}\}$, again using the same algorithm. The remaining columns are not active for any set of **G** vectors and play no role.

After distributing all the columns across the processors, a one-dimensional FFT along direction 3 is done on local data (on a single processor). However, the data on the planes orthogonal to direction 3 are distributed across the processors. In order to perform FFTs in each of these planes, the corresponding data must be made local to a processor. This is achieved by a parallel transpose operation, performed with a single call to the appropriate MPI[17] library routine. Two-dimensional FFTs can then be performed on the planes, with each processor operating on local data. Nonzero contributions are present only for (n_1', n_2') pairs corresponding to active columns. This fact can be exploited to reduce the number of FFT operations, by performing only the FFTs along direction 1 (or 2) that include nonzero contributions. The strategy for parallel three-dimensional FFT that we have presented requires the number of processors N_p be smaller than or equal to the number of planes N_3. The FFT from **r** to **G** space uses the same algorithm in reversed sequence.

In calculations using only the Γ point ($\mathbf{k}=0$), the KS orbitals can be chosen to be real functions in **r** space, so that $\phi(\mathbf{G}) = \phi^*(-\mathbf{G})$. This allows us to store only half of the Fourier components. Moreover, two real FFTs can be performed as a single complex FFT. To this end the auxiliary function Φ is introduced:

$$\Phi(\mathbf{r}) = \phi_j(\mathbf{r}) + i\,\phi_{j+1}(\mathbf{r}), \tag{51}$$

whose Fourier transform $\Phi(\mathbf{G})$ yields

$$\phi_j(\mathbf{G}) = \frac{\Phi(\mathbf{G}) + \Phi^*(\mathbf{G})}{2}, \tag{52}$$

$$\phi_{j+1}(\mathbf{G}) = \frac{\Phi(\mathbf{G}) - \Phi^*(\mathbf{G})}{2i}. \tag{53}$$

A side effect on parallelization is that **G** and $-\mathbf{G}$ must reside on the same processor. As a consequence, pairs of columns with $\mathbf{G}_{n_1',n_2',n_3'}$ and $\mathbf{G}_{-n_1',-n_2',n_3'}$ (with the exception of the case $n_1'=n_2'=0$), must be assigned to the same processor.

B. Scalar products

All scalar products $\langle f|g\rangle = \Sigma_i f_i^* g_i$, $i=1, n$ where i runs on a distributed grid, can be calculated by calling standard optimized library routines (like BLAS from NetLib[18]) on each processor, and subsequently by summing the partial results of all processors, using a call to standard MPI[17] libraries. Scalar products between vectors for which only half of the Fourier components are stored require a special treatment. Let n_p be the number of Fourier components stored on processor p. The contribution of this processor to the scalar product is $\langle f|g\rangle_p = 2\Sigma_{i=1,n_p} f_i^* g_i$ if the $\mathbf{G}=0$ components are not within the set of n_p components. If instead the $\mathbf{G}=0$ components, identified by $i=1$, are stored on processor p, the contribution of processor p to the scalar product is $\langle f|g\rangle_p = f_1 g_1 + 2\Sigma_{i=2,n_p} f_i^* g_i$.

C. Iterative orthonormalization

The scalar products in the matrix elements, Eq. (29), needed to compute the Lagrange multipliers are calculated in parallel, following the procedure of the previous subsection. In the present implementation, the solution of the matrix equation (35), involving square matrices of dimension equal to the number $N_{\rm ks}$ of KS orbitals, is not parallelized but replicated on all the processors. Usually the time spent in the nonparallelized part of the iterative orthonormalization is only a small fraction of the total time of the calculation.

To efficiently perform calculations on very large systems, using a large number of processors, the solution of Eq. (35) should also be parallelized. The time consuming steps are matrix–matrix multiplication and the diagonalization of the B matrix. Both calculations require $\mathcal{O}(N_{\rm ks}^3)$ floating-point operations. A convenient parallelization approach is described in Ref. 7.

D. Augmentation boxes

The parallelization of the calculations performed on the augmentation boxes is not obvious for two reasons: (1) each augmentation box has a grid which is a portion of the dense grid and is distributed across processors; and (2) the boxes follow the atoms in the MD evolution, causing the portion of the dense grid to change with time. In the present implementation, we deal with these difficulties as follows. We keep on all processors a copy of all the quantities defined on the augmentation boxes. Calculations on the grid of a given augmentation box are performed only in the processors that contain at least a fraction of the given augmentation-box grid. This causes some replication of the calculations. FFTs on the augmentation box grid are performed locally on each processor. In order to reduce the amount of replication, in the FFTs from **G** to **r** space, the two-dimensional FFTs along planes orthogonal to direction 3 are performed only in the planes belonging to the slice of the dense grid that is local to a given processor. No communication is needed to copy the augmented charge in **r** space from the augmentation-box grid to the dense grid (see Sec. II D 4). In the calculation of D_{nm}^I, we evaluate Q_{nm}^I in **G** space, transform it in **r** space using augmented-box FFT, evaluate the integral of Eq. (16) in **r** space and sum the final result over all processors. This approach keeps communications to a minimum, at the expense of a number of augmentation-box FFTs larger than in the serial case.

Augmentation-box grid related calculations constitute a very small part of the overall computational cost, both in computer time and in memory. Therefore the simple approach that we have presented is convenient even if some calculations are replicated on few processors and the load balance is not optimal.

IV. TEST CASE: IRON PORPHYRINS

We report here a comparison of computer performances for US and standard PPs in CP calculations. Our test systems—prototypes of systems containing the iron–porphyrin motif—are a reduced and an extended model of the active site of myoglobin.

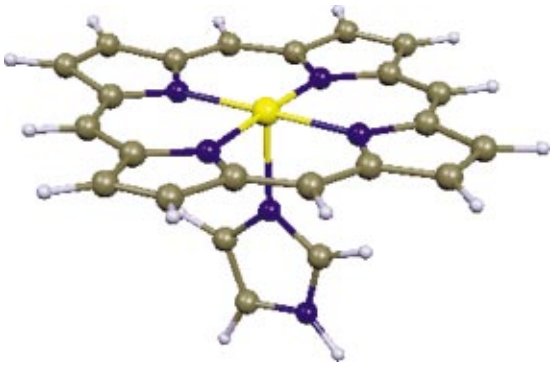

FIG. 1. (Color) Reduced model: the iron–porphyrin–imidazole complex: (yellow) Fe, (dark gray) C, (blue) N, (light gray) H.

A. Models and computational details

The reduced model is composed of an iron–porphyrin–imidazole complex, already investigated using the CP method by Rovira *et al.*;[19] the metallic pentacoordinated center is bound to the four planar porphyrin nitrogens, with the imidazole nitrogen occupying one of the axial sites, binding approximately orthogonal to the porphyrin plane (see Fig. 1). The chemical formula is $[FeN_6C_{23}H_{16}]$. A simple cubic cell of size 15.875 Å, containing a total of 46 atoms and 154 electrons, is used. For the reduced model, we compare both spin-restricted and unrestricted ($S=2$) calculations.

The extended model is composed of a large portion of the myoglobin active site, defined by the full heme group (same coordination as for the reduced model) plus the 13 surrounding residues which were comprised within a sphere of 8 Å radius centered on the iron atom (see Fig. 2). The initial geometry has been taken from the x-ray experimental structure of the O_2–myoglobin complex[20] and the included residues have been terminated by NH_2 groups, resulting in a total of 332 atoms and 902 electrons. The chemical formula is $[FeO_{19}N_{35}C_{106}H_{173}]$. A simple cubic cell of size 25.4 Å has been used, ensuring a minimum separation of 5 Å between periodic replicas. Such distance has been deemed sufficient based on previous experience with neutral molecules. For the extended model we discuss only the performances of the more computationally demanding spin-unrestricted ($S=2$) calculations.

For a correct comparison of performances we need to compare data of similar quality, in terms of accuracy of chemical properties, obtained with algorithms of comparable quality in terms of serial speed and parallel speedup. In order to satisfy the first requirement, we need to determine a set of cutoffs for the US and standard calculations yielding comparable structural properties. To this end we compared the optimized geometry of the triplet ground state for the reduced model from our US calculations and from published NC results, performed at $E_c^{\rm wf}=70$ Ry.[19] Our geometrical parameters are converged to the same extent as those in Ref. 19 at $E_c^{\rm wf}=25$ Ry and $E_c^{\rm dens}=200$ Ry; in particular the critical Fe–N distances in the porphyrin and in the imidazole are found to be 2.00 and 2.12 Å, respectively, versus 2.00 and 2.14 Å of Ref. 19, with the out-of-plane displacement of the iron atom computed to be 0.14 versus 0.15 Å. The overall

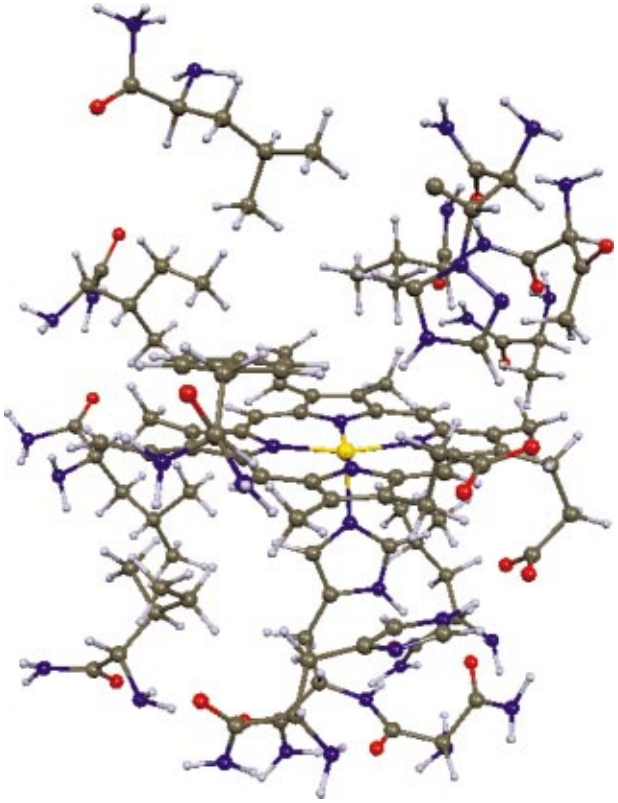

FIG. 2. (Color) Extended model of the myoglobin active site: (red) O, other colors as in Fig. 1.

agreement between the two sets of results is excellent, and the residual difference can be attributed to the different functionals used: BP86[21,22] in Ref. 19, and PW91[23] in our calculations. We estimate that we can safely compare US–PP calculations performed at $E_c^{\mathrm{wf}}=25$ Ry to standard calculations at $E_c^{\mathrm{wf}}=70$ Ry. For a fair comparison we use the same cutoff for the charge density in the US case as in the standard case ($E_c^{\mathrm{dens}}=280$ Ry). We also performed calculations at $E_c^{\mathrm{wf}}=35$ Ry for the US case, at $E_c^{\mathrm{wf}}=100$ Ry for the standard case ($E_c^{\mathrm{dens}}=400$ Ry in both cases).

In order to compare algorithms of similar quality, all calculations were performed using the same code[24] (standard PPs are just a special case of US PPs). The PPs used in the standard case tests were generated using the technique of Troullier and Martins.[25] In the US case, we use US PPs for all atoms, including H.[26] The PW91 functional[23] is used in all calculations.

TABLE I. Performances of the US and NC calculations, for the spin-restricted case. For US PPs: $E_c^{\mathrm{wf}}=25$ Ry, $E_c^{\mathrm{dens}}=280$ Ry, FFT grid size: 160, 96, 16 for the dense, coarse, and augmentation box grids, respectively. For NC PPs: $E_c^{\mathrm{wf}}=70$ Ry, $E_c^{\mathrm{dens}}=280$ Ry, FFT grid size: 160. N_p is the number of processors; M_r is an estimate of the RAM needed per processor in Mb; T_e is the execution time per electronic time step (at fixed atoms), in s; T_i is the same as T_e per CP time step (atoms moving), in s.

	US			NC		
N_p	M_r	T_e	T_i	M_r	T_e	T_i
4	190	49.5	56.2	357	136.1	159.7
8	100	23.6	26.1	139	44.6	49.8
16	57	10.7	12.1	77	21.0	22.9

TABLE II. Performances of the US calculations, spin-restricted case at higher cutoff. For US PPs: $E_c^{\mathrm{wf}}=35$ Ry, $E_c^{\mathrm{dens}}=400$ Ry, FFT grid size: 192, 120, 20 for the dense, coarse, and augmentation box grids, respectively. For NC PPs: $E_c^{\mathrm{wf}}=100$ Ry, $E_c^{\mathrm{dens}}=400$ Ry, FFT grid size: 192. The meaning of the various columns is the same as in Table I.

	US			NC		
N_p	M_r	T_e	T_i	M_r	T_e	T_i
8	167	44.7	50.1	307	118.7	143.7
16	93	20.9	23.0	157	50.5	57.2

B. Results

The results for the reduced models were obtained on a 32-node IBM SP3 (4×375 MHz power3 processors per node), while the extended model calculations were performed on a 64-processor SGI Origin (64×300 MHz MIPS R12 000 processors), both at the Keck Materials Science Laboratory, Princeton University.

The reported execution times are an average over 20 time steps of the measured wall time (the sum of CPU and system time, differing by only a few percent from pure CPU time).

The parallelization performances of the US versus NC–PPs implementation for the spin-restricted case of the reduced model are shown in Table I. The execution times for a CP step (including calculation of forces and time evolution of atomic positions) are about 15% larger than for a purely electronic step (orbital time evolution only), as expected. The small superlinear speedup observed both for US and NC PPs is a consequence of caching: since the memory per processor decreases almost linearly with the number of processors, better caching can be achieved with an increased number of processors, thus increasing the serial speed of the code. It is worth noting that US calculations are faster by a factor of ~2.5 with respect to the NC case and require half RAM memory and 1/4 disk space with respect to standard calculations.

The performances of US versus NC–PPs calculations at higher cutoff are shown in Table II. The number of PWs is approximately $(35/25)^{3/2}\simeq 1.65$ times larger than in the preceding case. Execution times should approximately be proportional to the same factor. The factor is actually somewhat larger ($\simeq 1.9$), but the cache effects mentioned above and the effect of the discreteness of the FFT grid explain the difference. Again, US calculations are faster by a factor of ~2.5 with respect to the NC case and require half RAM memory with respect to standard calculations.

In Table III we report spin-unrestricted results, showing an approximate doubling of execution time and of memory requirements, in line with expectations.

Table IV contains the performances of US calculations on the quintet state ($S=2$), which is experimentally known to be the ground spin state in myoglobin, of our extended model of myoglobin at $E_c^{\mathrm{wf}}=25$ Ry and $E_c^{\mathrm{dens}}=200$ Ry. The scaling with the number of processors is excellent in this

TABLE III. Performances of the US calculations: spin-unrestricted calculations, same cutoffs as in Table I.

N_p	M_r	T_e	T_i
4	263	92.5	104.5
8	139	44.6	49.8
16	77	21.0	22.9

case too, with a slight superlinear scaling up to 32 processors. The ratio between execution times for electronic and CP time steps is almost the same in the extended and in the reduced models. A geometry optimization time step requires less than 20 min in the spin-unrestricted case on 48 processors, with a total RAM usage of ~21 Gb. The total size of the file containing the KS orbitals is 6.7 Gb. In this case no standard calculation was attempted; indeed, an extrapolation from the results of Tables I, III, and IV points to a total memory requirement of more than 40 Gb.

A typical local geometry optimization requires ~250 time steps. An execution time of 20 min per time step thus translates into less than four days for the optimization to complete. This is a perfectly feasible calculation, not even requiring a state-of-the art massive parallel computer. On the other hand, a typical MD run requires no less than 10 000 time steps, corresponding in the present case to a few ps of simulation time. A true dynamical simulation would therefore become accessible on a state-of-the art massive parallel computer.

The relevance of simulating extended portions of myoglobin active site can be understood by considering the spin density distribution of the quintet spin state of the extended model, reported in Fig. 3 together with selected integrated spin density values. The spin density is mainly localized on the iron atom (~88%), even though a sizable contribution (~12%) is computed to be delocalized over the rest of the system, with the largest contributions arising from the propionate groups bound to the porphyrin ring. This finding is of particular interest, considering that CO rebinding in myoglobin has been recently related to a spin crossover from the quintet spin state, corresponding to unbound CO +myoglobin, to the singlet spin state characterizing the bound configuration of the CO–myoglobin complex.[27] Since explicit inclusion of the protein environment alters the spin distribution of the quintet state in our extended model, an effect on the relative energy of the different spin states can be expected.

TABLE IV. Performances of the US calculations: spin-unrestricted calculations for the extended model, $E_c^{\mathrm{wf}}=25$ Ry, $E_c^{\mathrm{dens}}=200$ Ry, FFT grid size: 224, 160, 20 for the dense, coarse, and augmentation box grids, respectively.

N_p	M_r	T_e	T_i
16	1067	3011	3764
32	656	1407	1690
48	441	992	1190

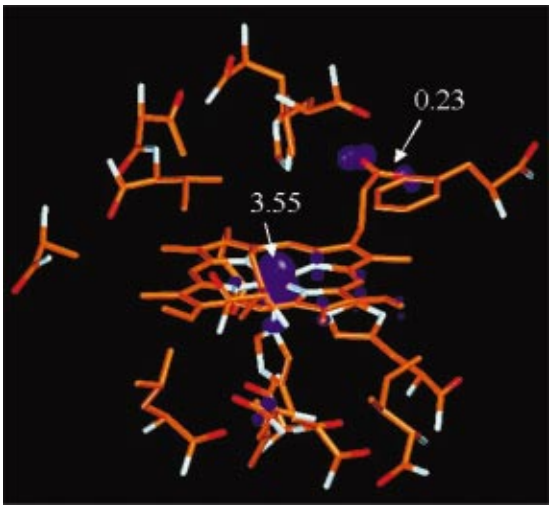

FIG. 3. (Color) Isodensity contour plot (contour value 0.02) and selected integrated values of the spin density for the quintet state of myoglobin extended model. Hydrogen atoms have been omitted for clarity.

V. ACCURACY OF PERIODIC BOUNDARY CONDITIONS

The use of PBC to describe molecules is perfectly appropriate for neutral molecules with small dipole/quadrupole moments, provided that the chosen supercell is large enough to minimize the spurious interactions between periodic replicas. This goal can usually be reached with supercells that leave a few Å of empty space between periodic replicas.

Charged molecules should be described by charged supercells, but these have infinite electrostatic energy. A finite energy is obtained by setting to zero the divergent $G=0$ contributions to the energy, as if the system were neutral. This is equivalent to adding a neutralizing background. Energies obtained in this way will be referred to as ''uncorrected.'' The direct comparison of uncorrected energies between different charge states is usually meaningless, because the error induced by PBC is large in this case. The long-range character of Coulomb interactions would require unpractically large supercells. Uncorrected energies may also be affected by a large error in molecules with large dipole/quadrupole moments.

Several techniques have been devised to overcome such limitation. The Hockney technique[28] yields an exact treatment of charged species using PWs without imposing PBC. This is achieved by cutting the Coulomb potential in real space beyond a suitably chosen cutoff that excludes all spurious interactions between periodic replica, still taking into account intramolecular interactions. This technique is rather expensive, since it requires the definition of an enlarged FFT grid for the Coulomb potential. A similar technique[29] where the cutoff acts in reciprocal space allows for faster execution with minor loss of accuracy.

A much simpler and approximate technique, due to Makov and Payne (MP),[12] consists of calculating the leading electrostatic correction terms and removing them from the uncorrected total energy. The MP correction is performed *a posteriori* on the energy only: the effect of the charge on the potential and on atomic forces is therefore neglected.

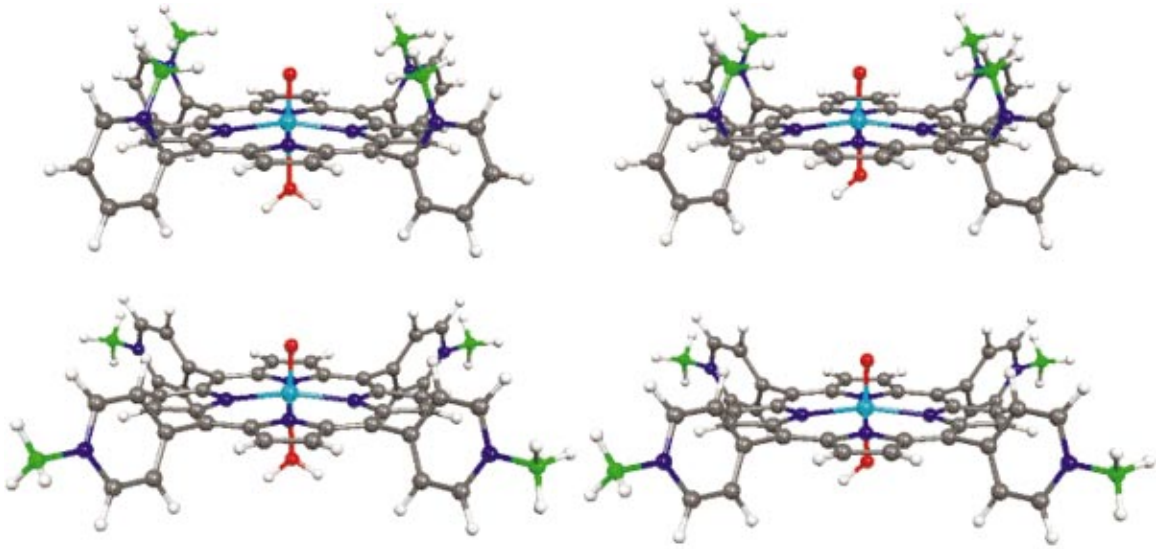

FIG. 4. (Color) 2-Pyp (upper panel) and 4-Pyp (lower panel) isomers of the oxo-aquo (left panel) and oxo-hydroxo (right panel) Mn(V) porphyrins. Mn is the light blue atom, the green atom signals the C of the methyl group bound to a N in the aromatic ring, whose position differs in the two isomers. Other colors are as in Fig. 2.

The Makov–Payne corrected energy E_{MP} for a cubic supercell has the form

$$E_{\mathrm{MP}}=E+\frac{q^2\alpha}{2L}-\frac{2\pi qQ}{3L^3}, \tag{54}$$

where E is the uncorrected energy, q is the net charge, Q is the quadrupole moment, α is the Madelung constant, and L is the supercell side. This is Eq. (15) of Ref. 12 with the correct sign. When applying Eq. (54), the origin has to be translated so that the dipole moment vanishes.[12] Therefore, the calculation of the Makov–Payne correction is straightforward since it requires only calculation of the dipole and quadrupole moments.

A. Models and computational details

We want to verify the ability of PW–PP calculations with the MP correction to reproduce the electronic and structural properties of highly charged species. We compute the energy difference between two *meso*-substituted Mn(V) porphyrins: the oxo-aquo-Mn(V)TM-2-Pyridyl (2-Pyp) and -Mn(V)TM-4-Pyridyl (4-Pyp) porphyrins, **I** (see Fig. 4), and between the corresponding oxo-hydroxo species, **II**, in which the axial water molecule has been replaced by a OH^- ligand. We compare our results to calculations employing a localized basis set of Slater type orbitals (STOs). The oxo-aquo and oxo-hydroxo porphyrins, recently experimentally characterized as mimics of the halide oxidation reaction catalyzed by haloperoxidases,[30,31] have a charge $+5$ and $+4$, respectively, with four positive charges approximately localized on the aryl nitrogens and, in the case of the oxo-aquo species, the residual positive charge located at the metal center; the two isomeric porphyrins differ, both in the oxo-aquo and oxo-hydroxo form, for the position of the methyl-substituted nitrogen in the aromatic ring attached to the *meso* porphyrin carbons, which should lead to a considerable increase of the quadrupole moment from the 2-Pyp to the 4-Pyp isomer. Due to charge differences between oxo-aquo and oxo-hydroxo species, to the high total charge and to the large expected difference in the quadrupole moment between the 2-Pyp and 4-Pyp porphyrins, we believe that the calculation of the relative energies of the two isomeric porphyrins represents a severe test for PW calculations within PBC.

We consider a reduced model of the real systems, in which the methyl groups bound to the aryl nitrogens are replaced by hydrogens. The US–PP results are compared to those obtained by using STOs in the frozen core approximation. US–PP calculations were performed by using a cutoff of 25–200 Ry for the KS orbitals and density, respectively, and a cubic cell of side 19.05 Å, ensuring a minimum separation of 8 Å between periodic replicas, without any symmetry constraints.

STOs results were obtained using the Amsterdam density-functional (ADF) program;[32–36] the frozen cores include $1s-2p$ states for Mn, $1s$ states for O, N, and C. The KS orbitals were expanded in an uncontracted Double-Zeta STO, standard basis set II[37] for all atoms with the exception of the transition metal for which we used a standard basis set IV,[37] of Triple-Zeta plus Polarization quality. STO calculations were performed within C_{2v} and C_S symmetry constraints for species **I** and **II**, respectively.

B. Results

In Table V we compare the relative energy of the singlet ground states[30] of the two isomeric 2-Pyp and 4-Pyp porphyrins, for both the oxo-aquo and oxo-hydroxo species. For the PP calculations we report both uncorrected and Makov–Payne corrected energy differences. Table VI contains the calculated values of the quadrupole moment used in Eq. (54).

Results obtained with STOs localized basis sets compute the oxo-aquo 4-Pyp system to be ~26 kcal/mol more stable than the 2-Pyp one. The Makov–Payne corrected US–PP results are in excellent agreement with STOs results, while

TABLE V. Energy differences ΔE_{2-4} (kcal/mol) between the two isomeric porphyrins, for the oxo-aquo (first row) and oxo-hydroxo (second row) species, computed with localized STOs basis and with US PPs, with Makov—Payne (MP) correction and uncorrected.

			US PPs	
		STOs	MP	no-MP
I	ΔE_{2-4}	26.4	24.5	9.5
II	ΔE_{2-4}	11.5	11.1	−7.0

uncorrected PP results indicate instead that the 4-Pyp isomer is more stable than the 2-Pyp one by only 9.5 kcal/mol. Moreover, in the case of the oxo-hydroxo species, the uncorrected results yield an incorrect energy ordering, with the 2-Pyp isomer computed to be more stable than the 4-Pyp one by 7.0 kcal/mol. The discrepancy is resolved upon correcting the total energies with the Makov–Payne term, resulting again in an excellent agreement with the STOs energy differences. Interestingly, the geometrical structures of the investigated charged systems calculated using the US–PP approach with PBC, turn out to quantitatively compare with results obtained using localized basis sets (see Table VII) for a comparison of main optimized geometrical parameters of species **I**. The main discrepancy (0.03 Å) is computed for the formally triple Mn≡O bond, probably because of the lack of polarization functions in the O STO basis set, which in turn leads to an overestimate of such parameter. The agreement between US PPs and STOs results suggests that the error introduced on the electrostatic potential by the presence of a charge in PBC does not significantly affect the structural properties. On the other hand, the effect on the total energy is sizable but mostly corrected by the use of Eq. (54).

C. Conclusions

We believe that our results demonstrate that the Car–Parrinello approach in conjunction with ultrasoft pseudopotentials represents a valuable and relatively cheap tool to describe the electronic and geometrical properties of complex bioinorganic systems, including highly charged and open-shell species. The study of the electronic and geometrical properties of such systems can now be achieved at a reasonable computational cost on conventional parallel machines with a limited number of processors. Moreover, a simple correction allows us to calculate energy differences in charged systems with an accuracy that is comparable to that of localized basis-set calculations not using periodic boundary conditions. First-principle molecular dynamics simulations of extended bioinorganic systems may still be feasible only on a state-of-the art massive parallel computer. The key to reach such a goal is the availability of parallel machines with increased performances and number of processors *and* highly optimized scalable algorithms. We believe that the present parallel implementation of the Car–Parrinello method using ultrasoft pseudopotentials provides such an algorithm that will allow the simulation of the dynamical properties of such complex systems in the near future.

TABLE VI. Quadrupole moments (a.u.) calculated with US PPs for the 2-Pyp (Q_2) and 4-Pyp (Q_4) isomers of the oxo-aquo (first row) and oxo-hydroxo (second row) species.

	Q_2	Q_4
I	518.1	627.2
II	412.8	522.8

TABLE VII. Main optimized geometrical parameters (Å, degree) for the 2-Pyp and 4-Pyp oxo-aquo isomers (**I**), computed with localized STOs basis and with US PPs. $\angle$N–Mn–N_{par} denotes the angle parallel to the Mn–OH_2 plane, $\angle$N–Mn–N_{perp} denotes the angle perpendicular to the Mn–OH_2 plane.

	2-Pyp		4-Pyp	
	US PPs	STOs	US PPs	STOs
$r_{Mn\equiv O}$	1.54	1.57	1.54	1.57
r_{Mn-OH_2}	2.21	2.20	2.22	2.21
r_{Mn-N} (average)	2.04	2.02	2.04	2.02
$\angle$N–Mn–N_{par}	163.3	163.5	163.4	163.8
$\angle$N–Mn–N_{perp}	167.8	166.6	166.9	165.4

ACKNOWLEDGMENTS

Calculations were performed at Keck computational facility of the Princeton Materials Institute and at ISTM Support from the NSF (Grant No. CHE-0121432) is acknowledged. We wish to thank Professor J. T. Groves, Professor T. G. Spiro, and Dr. A. Jarzecki for helpful discussions. F.D.A. thanks CNR (Progetto Finalizzato ''Materiali Speciali per Technologie Avanzate II'') for financial support. P.G. thanks MIUR, Grant No. PRIN 2001-028432 for partial support.

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[2] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[3] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[4] D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).
[5] L. J. Clarke, I. Ştich, and M. C. Payne, Comput. Phys. Commun. **72**, 14 (1993).
[6] D. Marx and J. Hutter, in *Modern Methods and Algorithms of Quantum Chemistry* (John von Neumann Institute for Computing, FZ Jülich, 2000), pp. 301–449.
[7] C. Cavazzoni and G. L. Chiarotti, Comput. Phys. Commun. **123**, 56 (1999).
[8] D. Vanderbilt, Phys. Rev. B **41**, 7892 (1990).
[9] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
[10] B. Hetényi, F. De Angelis, P. Giannozzi, and R. Car, J. Chem. Phys. **115**, 5791 (2001).
[11] K. Laasonen, A. Pasquarello, R. Car, C. Lee, and D. Vanderbilt, Phys. Rev. B **47**, 10142 (1993).
[12] G. Makov and M. C. Payne, Phys. Rev. B **51**, 4014 (1995).
[13] P. Pulay, Mol. Phys. **17**, 197 (1969).
[14] F. Tassone, F. Mauri, and R. Car, Phys. Rev. B **50**, 10561 (1994).
[15] R. Car and M. Parrinello, in *Simple Molecular Systems at Very High Density* (Plenum, New York, 1989), p. 455.
[16] S. de Gironcoli (private communication).
[17] Message Passing Interface Forum: *MPI: A message-passing interface standard*, Int. J. Supercomput. Appl. **8**, 3/4 (1994).
[18] A large collection of highly optimized libraries for linear algebra can be found at the URL ⟨http://www.netlib.org⟩
[19] C. Rovira, K. Kunc, J. Hutter, P. Ballone, and M. Parrinello, J. Phys. Chem. A **101**, 8914 (1997).
[20] S. E. V. Phillips and B. P. Schoenborn, Nature (London) **292**, 81 (1981).
[21] A. D. Becke, J. Chem. Phys. **84**, 4524 (1986).

[22] J. P. Perdew, Phys. Rev. B **33**, 8822 (1986).
[23] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, D. J. Singh, and C. Fiolhais, Phys. Rev. B **46**, 6671 (1992).
[24] Code CP, available at the URL ⟨http://www.democritos.it/scientific.php⟩
[25] N. Troullier and J. L. Martins, Phys. Rev. B **46**, 1754 (1992).
[26] A computer code for US-PP generation can be downloaded at the URL ⟨http://www.physics.rutgers.edu/~dhv/uspp/⟩
[27] J. N. Harvey, J. Am. Chem. Soc. **123**, 142 (2001).
[28] R. W. Hockney and J. W. Eastwood, in *Computer Simulation Using Particles* (McGraw–Hill, New York, 1981), p. 211.
[29] G. J. Martyna and M. E. Tuckerman, J. Chem. Phys. **110**, 2810 (1999).
[30] N. Jin and J. T. Groves, J. Am. Chem. Soc. **121**, 2923 (1999).
[31] N. Jin, J. L. Bourassa, S. Tizio, and J. T. Groves, Angew. Chem., Int. Ed. Engl. **39**, 3849 (2000).
[32] T. Ziegler, V. Tshinke, E. J. Baerends, J. G. Snijders, and W. Ravenek, J. Phys. Chem. **93**, 3050 (1989).
[33] E. J. Baerends, D. E. Ellis, and P. Ros, Chem. Phys. **2**, 42 (1973).
[34] E. J. Baerends and P. Ros, Chem. Phys. **2**, 51 (1973).
[35] E. J. Baerends and P. Ros, Chem. Phys. **8**, 41 (1975).
[36] E. J. Baerends and P. Ros, Int. J. Quantum Chem. **S12**, 169 (1978).
[37] ADF STO's basis set database is available at the URL ⟨http://tc.chem.vu.nl/SCM/Doc/atomicdatabase⟩