

Semi-Empirical Born–Oppenheimer Molecular Dynamics (SEBOMD) within the Amber Biomolecular Package

Antoine Marion, Hatice Gökcan, Gerald Monard

► **To cite this version:**

Antoine Marion, Hatice Gökcan, Gerald Monard. Semi-Empirical Born–Oppenheimer Molecular Dynamics (SEBOMD) within the Amber Biomolecular Package. *Journal of Chemical Information and Modeling*, American Chemical Society, 2018, 59 (1), pp.206-214. 10.1021/acs.jcim.8b00605 . hal-02188215

HAL Id: hal-02188215

<https://hal.univ-lorraine.fr/hal-02188215>

Submitted on 18 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SemiEmpirical Born-Oppenheimer Molecular Dynamics (SEBOMD) within the Amber biomolecular package

Antoine Marion,^{†,‡} Hatice Gokcan,^{†,¶} and Gerald Monard^{*,†}

[†]*Université de Lorraine, CNRS, LPCT, F-54000 Nancy, France*

[‡]*Middle East Technical University, Department of Chemistry, 06800, Ankara, Turkey*

[¶]*University of North Texas, Department of Chemistry, Denton, TX 76201, USA*

E-mail: Gerald.Monard@univ-lorraine.fr

Abstract

Semiempirical quantum methods from the Neglect of Differential Diatomic Overlap (NDDO) family such as MNDO, AM1, or PM3 are fast albeit approximate quantum methods. By combining them with linear scaling methods like the Divide & Conquer (D&C) method, it is possible to quickly evaluate the energy of systems containing hundreds to thousands of atoms. We here present our implementation in the Amber biomolecular package of a SEBOMD module that provides a way to run SemiEmpirical Born-Oppenheimer Molecular Dynamics. At each step of a SEBOMD molecular dynamics, a fully converged SCF calculation is performed to obtain the semiempirical quantum potential energy of a molecular system encaged or not in periodic boundary conditions. We describe the implementation and the features of our SEBOMD implementation. We show the requirements to conserve the total energy in NVE simulations, and how to accelerate SCF convergence through density matrix extrapolation. Specific ways of handling periodic boundary conditions using mechanical embedding or electrostatic embedding through a tailored quantum Ewald summation is developed. The parallel performance of SEBOMD simulations using the D&C scheme are presented for liquid water systems of various sizes, and a comparison between the traditional full diagonalization scheme and the D&C approach for the reproduction of the structure of liquid water illustrates the potentiality of SEBOMD to simulate molecular systems containing several hundreds of atoms for hundreds of picoseconds with a quantum mechanical potential in a reasonable amount of CPU time.

Introduction

Modeling reactive systems like enzymes at the molecular level is still a challenging problem for computation chemistry today. First, these systems are large. They can contain thousands of atoms, including not only a protein, its ligand, and sometimes a co-factor, but also solvent molecules and counterions.^{1,2} Understanding their physical and chemical properties requires to sample their (free) energy landscape using techniques like Monte Carlo (MC) or Molecular Dynamics (MD), which intrinsically request millions of successive energy and force computations.³ Second, these systems are reactive. While there exists tentative development of reactive molecular mechanics (MM) force fields,^{4,5} the forming and breaking of chemical bonds in these systems need a certain level of quantum mechanics (QM) in the general case to properly describe the electronic changes. A full description at the ab initio or DFT level using approaches like Born-Oppenheimer Molecular Dynamics (BOMD)⁶⁻⁸ or Car-Parrinello Molecular Dynamics (CPMD)^{9,10} is possible but limited to small systems, up to a few hundreds of atoms, and to short simulation times, up to a few tens of picoseconds,¹¹ since these methods are quite demanding in terms of CPU resources.^{12,13} Recently, the use of accelerators like Graphical Processing Units (GPU) by Ufimtsev et al. has permitted the molecular dynamics simulation of the small protein BPTI solvated by a couple of water layers for 8.8 picoseconds at a full DFT level.¹⁴ The mixture of MM and QM approaches in QM/MM methods has also been very popular during the last twenty years or so.¹⁵⁻¹⁹ In the QM/MM approach, only a portion of the system, the reactive one, is described by quantum mechanics while the rest that acts as a perturbation to the reactive part is described by a cheaper molecular mechanics description. The QM/MM approach has been used to describe the reactivity of many biomolecular systems in the literature.^{1-3,20-25} However, it still suffers from technical drawbacks that arise from the handling of the interactions between the QM and the MM parts. First, the QM part is often a small portion of a bigger molecule like a protein. Chemical bonds are therefore cut at the QM/MM border and need a special treatment like the addition of link atoms²⁶⁻²⁹ or the freezing of some molecular orbitals.³⁰⁻³⁴ Second, the way in which the QM and MM parts should interact is not straightforward.¹⁶ Multiple technical suggestions to handle the way the MM charges influence the

QM part^{26,35-37} and specific van der Waals parameters to represent non-electrostatic QM/MM interactions³⁸⁻⁴⁰ have been suggested. Third, the choice of the atoms that are included in the QM region can greatly influence the chemical mechanisms that can be modeled.^{20-22,41,42}

Between the expensive ab initio molecular dynamics and the delicate QM/MM approach, we suggest a third way that allows the full description of relatively large molecular systems at the quantum level while retaining fast computations: the SemiEmpirical Born-Oppenheimer Molecular Dynamics (SEBOMD) method. In a SEBOMD simulation, a Born-Oppenheimer molecular dynamics calculation is run, meaning that only the positions of the nuclei are defined and, at each step of the molecular dynamics, a fully converged self-consistent field (SCF) calculation is performed to obtain molecular orbitals, energy, and forces. The full quantum mechanical description of the system is obtained by using fast albeit approximate Neglect of Differential Diatomic Overlap (NDDO) semiempirical quantum mechanical method.⁴³⁻⁴⁵ To enhance the speed of the QM calculation, a linear scaling Divide & Conquer (D&C) algorithm can be used for large systems.^{46,47}

We have implemented the SEBOMD approach within the Amber biomolecular package,⁴⁸ more particularly in the open source AmberTools package that is part of Amber. Here, we highlight some specific features of our implementation. First, we briefly describe the implementation of SEBOMD as a module in the sander program. Second, we present how the total energy can be conserved in NVE simulations. Third, a simple way to accelerate SEBOMD simulations is given. Fourth, different Periodic Boundary Conditions (PBC) methods to handle long range electrostatics interactions are introduced. Fifth, the speed-up of SEBOMD simulations using the linear scaling Divide & Conquer algorithm with parallel computers is presented. Finally, we provide a comparison of SEBOMD simulations using either standard full Fock matrix diagonalization or approximate linear scaling D&C method.

The SEBOMD module

Implementation in the sander program

SEBOMD has been implemented in the sander program as a component of the free open source AmberTools package that is part of Amber.^{48,49} The sander program is in charge of performing molecular dynamics simulations. So far, it could run full MM and QM/MM MD. With SEBOMD, we have added the capability of running full QM simulations with or without periodic boundary conditions. At each step of a molecular dynamics simulation, the energy and forces calculations for a given set of coordinates are directed to a SEBOMD module that will compute these quantities at the semiempirical NDDO level. Currently available NDDO hamiltonians are MNDO,⁵⁰ AM1,⁵¹ PM3,⁵² RM1,⁵³ PM3/PDDG,⁵⁴ AM1/d-PhoT,⁵⁵ and AM1/d-CB1.⁵⁶ MAIS^{57,58} and PIF⁵⁸⁻⁶² corrections to the PM3 Hamiltonian are also provided. SEBOMD simulations can be run with sander using usual thermodynamical ensembles: NVE, NVT, or NPT.

At each step of an MD simulation, an MM energy calculation is still computed and is mixed with the SEBOMD energy if requested, as controlled by a λ parameter (keyword: lambda):

$$E_{\text{pot}} = (1 - \lambda)E_{\text{MM}} + \lambda E_{\text{QM}} \quad (1)$$

where E_{pot} is the total potential energy, E_{MM} is the MM potential energy, and E_{QM} is the semiempirical QM potential energy. When invoking SEBOMD simulation, the lambda keyword is set to 1.0 by default. Using a λ factor different than 1.0 can be highly relevant for equilibration purpose. Often, SEBOMD simulations are started from systems equilibrated at the MM level. However semiempirical NDDO potential energy surfaces are sometimes quite different from their MM counterparts.⁶³ This can yield unrealistically large atomic forces at the beginning of SEBOMD simulations, leading to large velocities and temperatures, and eventually bond breaking since bonds are no longer represented by harmonic springs. By using successive equilibration steps with λ gradually varying from 0 to 1, one can ensure a smooth transition between a classical MM

simulation and a SEBOMD simulation.

Energy conservation

The determination of the NDDO wavefunction in SEBOMD simulations can be performed either using a so-called "standard" calculation scheme or via a linear scaling Divide & Conquer (D&C) algorithm. In the standard method, full diagonalization of the entire Fock matrix is performed at each step of the SEBOMD. The CPU timing in such cases therefore scales as $O(N^3)$ where N is roughly the size of the system in number of atoms because the diagonalization of the Fock matrix is the bottleneck in semiempirical NDDO computations.^{46,47,64-66} We do not provide here a detailed description of the Divide & Conquer method and let the reader refer to appropriate publications in the literature instead.^{46,47,67} We only recall that the D&C method is a way to circumvent the diagonalization bottleneck by dividing a molecular system into smaller overlapping subsystems. Each subsystem is composed of a *core* and one or two *buffer* regions. The density matrix of each subsystem is obtained by solving a localized set of Roothan-Hall equations for which the local Fock matrix is extracted from the full Fock matrix. Buffer regions are used to determine what information from the local density matrices should be used to recover the entire, global density matrix. In our implementation that is originally based on the DivCon99 software,⁶⁸ we use the dual buffer region scheme as suggested by Dixon and Merz, Jr.⁴⁷ Because the subsystems are mostly equal in terms of size, the CPU times corresponding to the diagonalizations of the local Fock matrices are similar and linear scaling is achieved by the fact that the number of subsystems scales linearly with the size of the system.

When no external force is applied to a system, a molecular dynamics simulation should conserve the total energy. This corresponds to the NVE thermodynamical ensemble where the number of particles N , the volume V and the total energy E being the sum of the kinetic energy and the potential energy are kept constant along the simulation. In contrast to classical MM molecular dynamics, conservation of the total energy in Born-Oppenheimer molecular dynamics not only depends on the time step that is used but also on the precision with which the SCF is converged

at each step of the simulation. This is due to the fact that forces are not directly derived from the potential energy equations but make use of the SCF convergence in their calculation to avoid the costly evaluation of the derivatives of the density matrix elements. Therefore, to ensure a proper energy conservation during a SEBOMD simulation, one must pay a special attention to the convergence of the SCF at each step of the molecular dynamics.

In many quantum chemistry software, the SCF convergence is achieved when the electronic energy between two steps of the SCF cycle does not change for more than a certain threshold. For example, in MOPAC⁶⁹ this threshold is normally set to 1×10^{-4} kcal/mol. In SEBOMD, such a way to check for SCF convergence becomes unsuitable when a molecular system becomes very large (i.e., several hundreds or thousands of atoms). For example, a system with 1 water molecule has roughly a semiempirical electronic energy of about -1×10^4 kcal/mol (kcal/mol is the energy unit used in sander). A system with 216 water molecule has an electronic energy of about -7×10^6 kcal/mol and a solvated protein of about 5,700 atoms has an electronic energy of about -2×10^9 kcal/mol. The electronic energy in a typical SEBOMD simulation is thus very large and a "standard" convergence criterion would exceed the precision that can be expected from using double precision floating point numbers.

In SEBOMD, the SCF convergence criterion is handled by the `dpmax` keyword. For two successive iterations, if any density matrix element has not been changed by more than `dpmax`, then the energy of the system is stated as converged and forces are computed from the current density. Figure 1 shows how the total energy is conserved for a set of 64 water molecules in gas phase depending on the choice of the SCF algorithm (i.e., "standard" method vs. D&C), the molecular dynamics time step, and the value of `dpmax`. For the sake of clarity, the variations of total energy along time are represented relative to the initial energy as $[E(t) - E(t = 0)] / E(t = 0)$. Using a full diagonalization of the Fock matrix at each step of the MD (Figure 1a; "Standard"), the energy is always conserved with `dpmax` = 10^{-7} a.u., regardless of the chosen time step. Even for a time step of 1.0 fs, which is relatively large to describe the dynamics of quantum mechanical systems, the MD integrator in sander still delivers stable simulations. Linear regression analysis of the en-

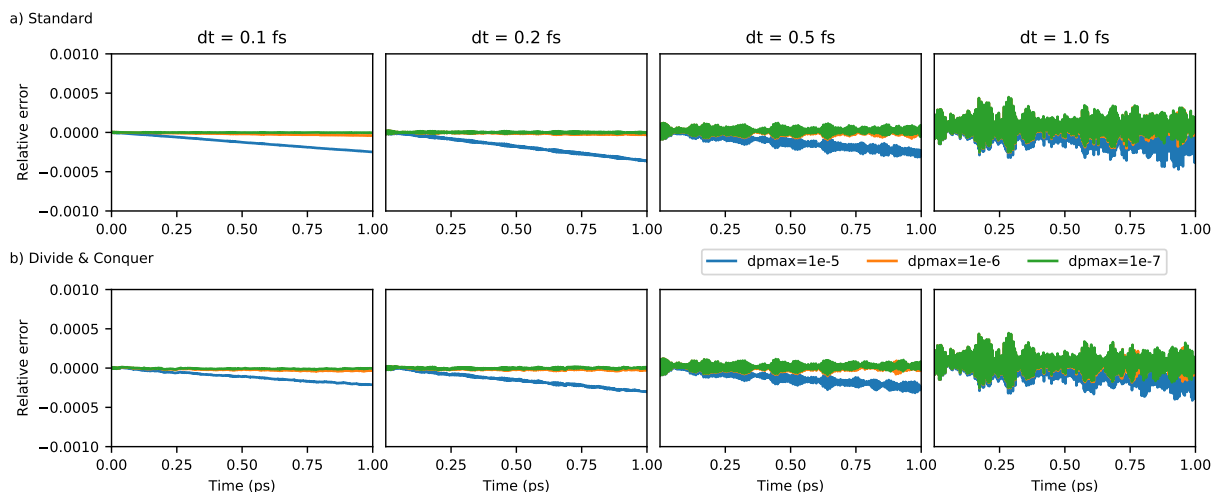


Figure 1: Conservation of the total energy for a set of 64 water molecules in gas phase using the PM3-PIF hamiltonian. Relative error compared to the initial total energy is computed for a 1 picosecond simulation using different time steps and different dp_{\max} threshold values. a) *Standard*: full diagonalization of the Fock matrix; b) *Divide & Conquer*: D&C scheme applied with 1 water molecule per core, a first buffer of 6 Å and no second buffer. From left to right: time steps of 0.1, 0.2, 0.5, and 1.0 fs, respectively. *green line*: $dp_{\max} = 10^{-7}$, *orange line*: $dp_{\max} = 10^{-6}$, *blue line*: $dp_{\max} = 10^{-5}$.

ergy variations gives a slope that is mostly independent of the time step used. For $dp_{\max} = 10^{-7}$ a.u., the (absolute) slope in relative energy is about 10^{-6} per picosecond, meaning that the total energy drifts by about 10^{-4} % every picosecond. In the case of this system comprised of 64 water molecules, it corresponds to an energy drift of approximately -0.01 kcal/mol/ps for a potential energy of about -3650 kcal/mol. For $dp_{\max} = 10^{-6}$ a.u., the slope of the total relative energy is in the range of 10^{-5} per picosecond, and for $dp_{\max} = 10^{-5}$ a.u., the slope is about 10^{-4} per picosecond. The total energy conservation is therefore in direct relationship with the convergence of the density matrix elements. When using the Divide & Conquer scheme (Figure 1b), a similar conclusion can be drawn: total energy can be conserved for time steps up to 1.0 fs and the small energy drift is directly dependent on the value of dp_{\max} . Here the Divide & Conquer scheme is set such that the system is divided into 64 subsystems with 1 water molecule per core, a first buffer of 6 Å and no second buffer. The inclusion of atoms in the first buffer region is molecular-based, meaning that if one atom of a water molecule is positioned less than 6 Å away from one atom of the core region,

then the entire water molecule is included in the buffer region. It is worth noting here that, at each step of the molecular dynamics, the Divide & Conquer scheme updates the composition of each subsystem depending on the instantaneous positions of all molecules: for each water molecule that defines one core region other water molecules are included or not in the buffer region depending on their distances to the core molecule. The number of water molecules and the composition of each subsystem can therefore vary along the SEBOMD simulation. Nonetheless, the molecules that are dynamically included or excluded from subsystems physically lay away enough from the core region and therefore have a negligible influence on the total energy and its conservation.

SCF acceleration

The convergence of Self Consistent Field is faster (i.e., requires fewer iteration cycles) when the initial guess density is close to the final, converged electronic density. By default in SEBOMD simulations, at each step of the molecular dynamics, the converged density matrix of the previous MD step is used as a guess for the current MD step. As a way to decrease the number of SCF cycles at each molecular dynamics step and therefore to accelerate the overall SEBOMD simulation, we have also implemented a polynomial extrapolation of the guess density matrix based on the last three previous SEBOMD steps. This is triggered by the `ipolyn` keyword. When polynomial extrapolation is activated, starting at the fourth SEBOMD step, all guess density matrix elements are constructed as a polynomial extrapolation of the corresponding elements of the last three converged density matrices. In Table 1, we have averaged the number of SCF cycles that are needed for each MD step of a one picosecond SEBOMD simulation of 64 water molecules in gas phase using the "standard" or the D&C method for different values of `dpmax` and timesteps (Δt). For any combination of `dpmax` and Δt , the number of SCF steps required to obtain a converged density matrix is lower when using the extrapolation scheme for the guess density matrix. The ratio of acceleration ranges from 7.0, for $\Delta t = 0.1\text{fs}$ and `dpmax` = 10^{-5} a.u., to 1.4, for $\Delta t = 1.0\text{fs}$ and `dpmax` = 10^{-7} a.u.. Overall, the smaller the Δt the fewer the number of SCF cycles. A tighter convergence criterion (`dpmax`) leads to a larger number of SCF cycles for any guess density matrix

Table 1: Average number of SCF cycles per MD step for a 1 ps NVE simulation of 64 water molecules in gas phase using the full diagonalization scheme (Standard) or the Divide & Conquer scheme (D&C: 1 water molecule per core, first buffer zone = 6Å, no second buffer zone). *previous*: density matrix initial guess is copied from the last converged SCF step; *extrapolate*: all density matrix initial guess elements are extrapolated from the corresponding density matrix elements of the last three MD steps.

		Average number of SCF cycles per MD step			
		Standard		D&C	
Δt (fs)	d _p max	previous	extrapolate	previous	extrapolate
0.1	10 ⁻⁵	11.2 ± 0.4	1.6 ± 0.5	11.1 ± 0.4	2.0 ± 1.0
	10 ⁻⁶	15.9 ± 0.4	3.3 ± 0.5	15.9 ± 0.4	3.7 ± 0.9
	10 ⁻⁷	20.6 ± 0.5	6.4 ± 0.5	20.5 ± 0.5	6.4 ± 0.8
0.2	10 ⁻⁵	12.7 ± 0.5	3.1 ± 0.4	12.6 ± 0.5	3.3 ± 0.6
	10 ⁻⁶	17.2 ± 0.4	6.0 ± 0.5	17.2 ± 0.4	6.1 ± 0.5
	10 ⁻⁷	21.9 ± 0.5	9.3 ± 0.5	22.0 ± 0.5	9.3 ± 0.6
0.5	10 ⁻⁵	14.3 ± 0.5	6.6 ± 0.6	14.5 ± 0.5	6.6 ± 0.6
	10 ⁻⁶	19.0 ± 0.4	10.0 ± 0.5	19.1 ± 0.4	10.0 ± 0.5
	10 ⁻⁷	23.8 ± 0.5	14.1 ± 0.6	23.8 ± 0.5	14.0 ± 0.6
1.0	10 ⁻⁵	15.8 ± 0.4	9.5 ± 0.6	15.8 ± 0.5	9.5 ± 0.6
	10 ⁻⁶	20.5 ± 0.5	13.5 ± 0.7	20.4 ± 0.5	13.5 ± 0.7
	10 ⁻⁷	25.3 ± 0.5	18.0 ± 0.7	25.3 ± 0.6	18.0 ± 0.7

scheme, but using the extrapolation option always reduces the number of SCF cycles needed for full convergence. Also, the number of SCF cycles is nearly independent of the energy scheme that is used, either the full diagonalization scheme (Standard) or the Divide & Conquer scheme.

Long range electrostatics

By default, SEBOMD handles periodic boundary conditions (PBC) using the minimum image (MI) convention. When the Fock matrix is built, all its elements are either mono- or di-atomic terms. This is intrinsically due to the Zero Differential Overlap (ZDO) approximation used in all NDDO methods.⁴³ Consequently, diatomic terms are dependent on the distance between two atoms in the system. SEBOMD uses the MI convention for the calculation of such distances. This way there is no real border for the electronic wavefunctions and any particle of the systems, core or electron, is surrounded by all others as if they were at the center of the simulation box.

Long-range electrostatics, beyond the simulation box, are considered as a perturbation to the MI wavefunctions. There are two ways in SEBOMD to handle long-range electrostatics: mechanical embedding or electrostatic embedding, via the `longrange` keyword.

In mechanical embedding, a classical Particle Mesh Ewald (PME)^{70,71} correction is added to the QM minimum image potential energy:

$$E_{\text{pot}} = E_{\text{MI}}[\mathbf{P}] + \Delta E_{\text{PME}}[\mathbf{q}_{\text{ext}}] = E_{\text{MI}}[\mathbf{P}] + E_{\text{PME}}[\mathbf{q}_{\text{ext}}] - E_{\text{MI}}[\mathbf{q}_{\text{ext}}] \quad (2)$$

$E_{\text{MI}}[\mathbf{P}]$ is the semiempirical quantum energy of the system computed using the minimum image convention. $E_{\text{PME}}[\mathbf{q}_{\text{ext}}]$ is the classical PME energy computed by sander using the set of atomic point charges \mathbf{q}_{ext} provided by the Amber topology file of the simulation. $E_{\text{MI}}[\mathbf{q}_{\text{ext}}]$ is the classical electrostatic energy within the box, which must be removed to avoid double counting as it is already included in the quantum energy. This mechanical embedding scheme provides an implicit polarization of the quantum wavefunction via an external perturbation of the simulation box by an infinite set of surrounding classical atomic point charges.

In the electrostatic embedding, a quantum Ewald correction is applied to directly polarize the semiempirical quantum wavefunction:⁷²

$$E_{\text{pot}} = E_{\text{MI}}[\mathbf{P}] + \Delta E_{\text{Ewald}}[\mathbf{P}] \quad (3)$$

with:

$$\Delta E_{\text{Ewald}}[\mathbf{P}] = \frac{1}{2} \sum_{\alpha} \sum_{\beta} q_{\alpha} q_{\beta} \Delta \Psi_{\text{Ewald}}(\mathbf{R}_{\alpha}, \mathbf{R}_{\beta}) \quad (4)$$

$$= \frac{1}{2} \sum_{\alpha} \sum_{\beta} q_{\alpha} q_{\beta} [\Psi_{\text{dir}}(\mathbf{R}_{\alpha}, \mathbf{R}_{\beta}) + \Psi_{\text{rec}}(\mathbf{R}_{\alpha}, \mathbf{R}_{\beta}) - \Psi_{\text{box}}(\mathbf{R}_{\alpha}, \mathbf{R}_{\beta})] \quad (5)$$

where q_{α} are atomic charges derived from the quantum electronic density \mathbf{P} , Ψ_{dir} and Ψ_{rec} are the conventional direct and reciprocal Ewald field, respectively,^{73,74} and Ψ_{box} represents the charge-charge interaction field within the box:

$$\Psi_{dir}(\mathbf{R}_\alpha, \mathbf{R}_\beta) = \sum_{\mathbf{n}}' \frac{\operatorname{erfc}(\kappa|\mathbf{R}_\alpha - \mathbf{R}_\beta + \mathbf{n}|)}{|\mathbf{R}_\alpha - \mathbf{R}_\beta + \mathbf{n}|} - \frac{2\kappa}{\sqrt{\pi}} \delta_{\alpha\beta} \quad (6)$$

$$\Psi_{rec}(\mathbf{R}_\alpha, \mathbf{R}_\beta) = \frac{4\pi}{V} \sum_{\mathbf{k} \neq 0} \frac{\exp\left(\frac{-\mathbf{k}^2}{4\kappa^2}\right)}{\mathbf{k}^2} \cos(\mathbf{k} \cdot (\mathbf{R}_\alpha - \mathbf{R}_\beta)) \quad (7)$$

$$\Psi_{box}(\mathbf{R}_\alpha, \mathbf{R}_\beta) = \frac{1}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|} \quad (0 \text{ if } \alpha = \beta) \quad (8)$$

Then, any Fock matrix element can be written as:

$$F_{\mu\nu} = \frac{\partial E_{\text{pot}}}{\partial P_{\mu\nu}} = F_{\mu\nu}^{\text{MI}} + \Delta F_{\mu\nu}^{\text{Ewald}} \quad (9)$$

$$= F_{\mu\nu}^{\text{MI}} + \frac{\partial \Delta E_{\text{Ewald}}[\mathbf{P}]}{\partial P_{\mu\nu}} \quad (10)$$

$$= F_{\mu\nu}^{\text{MI}} + \sum_{\alpha} \frac{\partial q_{\alpha}}{\partial P_{\mu\nu}} \frac{\partial \Delta E_{\text{Ewald}}}{\partial q_{\alpha}} \quad (11)$$

In the case of Mulliken atomic charges, the computation of $\frac{\delta q_{\alpha}}{\delta P_{\mu\nu}}$ is straightforward and the perturbation to the minimum image Fock matrix elements can be written as:

$$\Delta F_{\mu\nu}^{\text{Ewald}} = - \sum_{\alpha \in \mu, \nu} \delta_{\mu\nu} \sum_{\gamma} q_{\gamma} \Delta \Psi_{\text{Ewald}}(\mathbf{R}_{\alpha}, \mathbf{R}_{\gamma}) \quad (12)$$

The electrostatic embedding using Mulliken atomic charges is usually preferred over mechanical embedding since it provides a direct polarization of the quantum wavefunction.

Parallel performance

Matrix diagonalization is a linear algebra operation with a formal complexity of $O(N^3)$. Its parallel efficiency is very low and therefore does not scale very well with the number of processors. In SEBOMD, the standard full diagonalization scheme is available only for the serial version of sander, i.e., it cannot be used in parallel. In contrast, the advantage of using the Divide & Conquer approach is that the division into subsystems can be spread among processors to, thereby, achieve

a proper parallel scaling. We show in Figure 2 the average speed (in ps/day) of NVT SEBOMD simulations for different sizes of pure water molecular systems. From left to right, the systems respectively contain 64, 125, 216, or 512 water molecules, encaged in a periodic box with a density of 1.0 g/cm^3 . The D&C algorithm uses 1 water molecule per core, a first buffer of 6 \AA and no second buffer. Different schemes are used according to our findings in the energy conservation section. A *safe* scheme, in orange, employs a Δt of 0.5 fs with a dpmax of 10^{-6} a.u.. This achieves a good energy conservation as depicted in Figure 1. The *slow* scheme, in green, is set with $\Delta t = 0.5$ fs and $\text{dpmax} = 10^{-7}$ a.u. while the *fast* scheme, in blue, is set with $\Delta t = 1.0$ fs and $\text{dpmax} = 10^{-5}$ a.u..

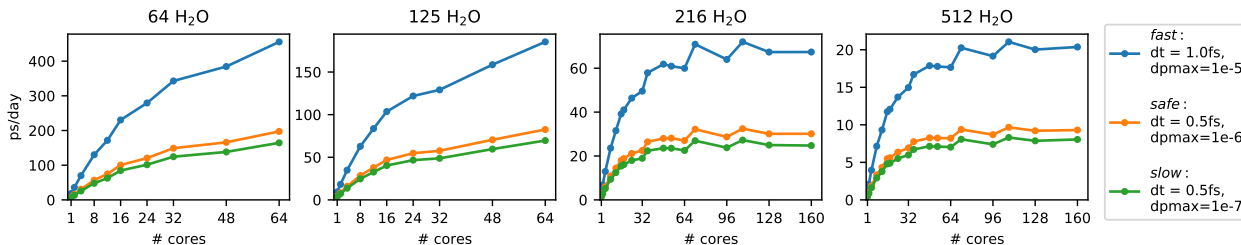


Figure 2: Average speed of SEBOMD simulations, in ps/day, for different sizes of molecular systems. From left to right: boxes of 64, 125, 216, and 512 water molecules, respectively. Hardware specs: 2x Intel Gold 6130 (32 cores/node), 192 Gbytes RAM, Intel OmniPath 100 Gbs interconnect.

First, we observe no significant difference between the *slow* and *safe* schemes in terms of speed and parallel scalability. Regardless of the system size, simulations using the *safe* scheme are about 15-20% faster than with the *slow* scheme. Similarly, using the *fast* scheme gives simulations that are always 2.1–2.3 times faster than with the *safe* scheme. It is also noteworthy that the D&C scheme can be used in parallel before the so-called cross-over point,⁶⁴ that is to say when the system is small enough that the standard scheme gives a better performance than the D&C scheme on one processor. For example, with 64 water molecules, the serial version of sander with the *fast* setup provides a speed of ~ 41 ps/day with the standard method and only ~ 18 ps/day using the D&C algorithm. However, if one uses 2 CPUs, the D&C speeds is then ~ 36 ps/day. With 4 CPUs, the D&C speed goes to ~ 70 ps/day, beyond the serial speed of the full diagonalization

scheme. With 64 CPUs, the D&C scheme even reaches a speed of ~ 455 ps/day, way beyond the performance that the full diagonalization scheme can achieve. For other molecular systems with larger size, the advantage of using the D&C scheme is even more potent. With the *fast* setup on 125 water molecules, the standard and the D&C scheme give about the same speed on one processor: ~ 10 ps/day. But the D&C scheme can scale nicely, giving speed of 18, 35, and 185 ps/day for 2, 4, and 64 CPUs, respectively. Beyond 125 water molecules, the D&C scheme is always faster than the full diagonalization scheme, even on one processor.

Standard diagonalization vs. Divide & Conquer: the example of liquid water

The Divide & Conquer method, as suggested by Dixon and Merz, Jr.,^{46,47} is an approximate method to obtain the total semiempirical quantum energy of a molecular system. Two main approximations are made in D&C. First, density matrix elements $P_{\mu\nu}$ are considered as negligible if the two atoms bearing the μ and ν atomic orbitals are well-separated. In practice, this is the purpose of buffers: atoms located beyond the buffer regions of a given subsystem do not interact with the electronic density of the atoms in that sub-system. Second, because the diagonalization of all the Fock matrices defining all the subsystems generates overlapping molecular orbitals, the molecular orbital occupancy is determined via the use, and the determination, of a Fermi energy under the constraint of the conservation of the total number of electrons in the system. To assess the difference in using either a full diagonalization scheme or the D&C method, we have performed three different SEBOMD simulations of liquid water in the NVT ensemble at 300K using the PM3-PIF hamiltonian.^{52,58-62} The *slow* simulation corresponds to our previously defined *slow* scheme: $\text{dpmax} = 10^{-7}$ a.u., $\Delta t = 0.5\text{fs}$. It is 40 ps long and can serve as a reference simulation. The *safe* simulation is also 40 ps long with $\Delta t = 0.5\text{fs}$ and a relaxed $\text{dpmax} = 10^{-6}$ a.u.. Both *slow* and *safe* simulations use the full diagonalization scheme. The *fast* simulation is 100 ps long with $\text{dpmax} = 10^{-5}$ a.u. and $\Delta t = 1.0\text{fs}$. It uses the Divide & Conquer scheme: 1 water molecule per core, a first buffer of 6 Å and no second buffer. For all simulations, the NVT ensemble was enforced using Langevin dynamics as a thermostat with a weak coupling (collision frequency of 1ps^{-1}).

Long-range electrostatics was taken into account using quantum Ewald summation of Mulliken atomic charges extracted from the semiempirical wavefunction at each step of the simulations (i.e., electrostatic embedding scheme). Polynomial extrapolation of the guess density matrix was used to accelerate the convergence of the SCF calculations. SEBOMD analysis was performed using the `cpptraj`⁷⁵ module of Amber.

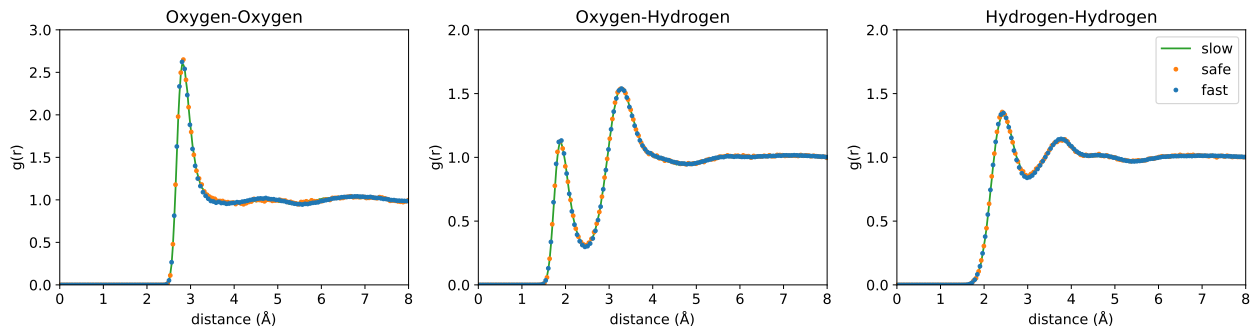


Figure 3: Radial distribution functions for liquid water using the PM3-PIF hamiltonian in an NVT SEBOMD simulation of 216 water molecules at 300K. *slow*: full diagonalization scheme, $\text{dpmax} = 10^{-7}$ a.u., $\Delta t = 0.5\text{fs}$; *safe*: full diagonalization scheme, $\text{dpmax} = 10^{-6}$ a.u., $\Delta t = 0.5\text{fs}$; *fast*: divide & conquer scheme (1 water molecule per core, 6 Å buffer), $\text{dpmax} = 10^{-5}$ a.u., $\Delta t = 1.0\text{fs}$.

The radial distribution functions (RDFs) for water obtained with the three SEBOMD simulations are represented in Figure 3. As it is not the scope of the present study, no comparison with any experimental values are provided here. All simulations present nearly identical curves for the oxygen-oxygen, hydrogen-oxygen, and hydrogen-hydrogen RDFs. As a result, we can safely conclude that using the Divide & Conquer scheme does not impact the structure of liquid water compared to using a full diagonalization scheme. Similarly, using a relaxed convergence setup ($\text{dpmax} = 10^{-5}$ a.u.) does not influence the structuration of water molecules. However, in terms of speed of simulations, the differences are rather critical, notably because the full diagonalization scheme can only be run with a single CPU in the current SEBOMD implementation. On a 3.5 Ghz Intel Xeon E5-2637 v4, the *slow* simulation runs at ~ 1.44 ps/day while the *safe* simulation runs at ~ 1.75 ps/day. In contrast, the *fast* simulation runs at ~ 34 ps/day using 32 cores on a 2.1 Ghz Intel Xeon E5-2683 v4.

Conclusions

Semiempirical methods are fast approximate quantum methods that are suitable for long molecular dynamics runs. In the Amber biomolecular package, we have implemented a module that is capable of running SEBOMD simulations: SemiEmpirical Born-Oppenheimer Molecular Dynamics. At each step of a SEBOMD simulation, the semiempirical quantum wavefunction is fully converged through a Self Consistent Field process. In our implementation, we have taken care of the SCF convergence to ensure the conservation of the total energy when no external forces or restraints are applied (i.e., in the NVE thermodynamical ensemble). This is triggered by the `dpmax` keyword. A safe setting for `dpmax` is 10^{-6} a.u. but SEBOMD simulations can be made faster using a `dpmax` value of 10^{-5} a.u. without modifying results such as the structural properties of liquid water. An additional way of accelerating SEBOMD simulations is to use guess density matrix extrapolation which reduces the number of SCF cycles made at every MD step. Finally, the Divide & Conquer scheme provides a fast and parallel way of finding the NDDO energy of "large" molecular systems in SEBOMD simulations without compromising the results obtained with this approximate linear scaling quantum method as compared to the traditional full Fock matrix diagonalization scheme. With the SEBOMD module implemented in the Amber biomolecular package, it is now possible to routinely perform full-QM molecular dynamics simulations of systems containing hundreds to thousands of atoms in reasonable amounts of CPU time for simulation lengths up to the nanosecond timescale using parallel computers.

Acknowledgement

High Performance Computing resources were partially provided by the EXPLOR centre hosted by the University de Lorraine, the ROMEO centre hosted by the Université de Reims Champagne-Ardenne, and GENCI (allocation number A0020807496). H. G., A. M. and G. M. are grateful to Campus France and Tubitak for travel grants (PHC Bosphore 31862YG and 35673VB). H. G. thanks the French Embassy in Ankara, Turkey for *co-tutelle de thèse* grant, and TUBITAK-Science

Fellowships and Grant Programmes Department for financial support.

References

- (1) Wei, W.; Gault, J. W.; Monard, G. Pretransfer Editing in Threonyl-tRNA Synthetase: Roles of Differential Solvent Accessibility and Intermediate Stabilization. *ACS Catal.* **2017**, *7*, 3102–3112.
- (2) Wei, W.; Monard, G.; Gault, J. W. Computational insights into substrate binding and catalytic mechanism of the glutaminase domain of glucosamine-6-phosphate synthase (GlmS). *RSC Adv.* **2017**, *7*, 29626–29638.
- (3) Ugur, I.; Marion, A.; Aviyente, V.; Monard, G. Why Does Asn71 Deamidate Faster Than Asn15 in the Enzyme Triosephosphate Isomerase? Answers from Microsecond Molecular Dynamics Simulation and QM/MM Free Energy Calculations. *Biochemistry* **2015**, *54*, 1429–1439.
- (4) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396–9409.
- (5) Chenoweth, K.; van Duin, A. C. T.; Goddard, W. A. r. ReaxFF reactive force field for molecular dynamics simulations of hydrocarbon oxidation. *J. Phys. Chem. A* **2008**, *112*, 1040–1053.
- (6) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comp. Phys. Comm.* **2005**, *167*, 103–128.
- (7) Guidon, M.; Schiffmann, F.; Hutter, J.; VandeVondele, J. Ab initio molecular dynamics using hybrid density functionals. *J. Chem. Phys.* **2008**, *128*, 214104.

- (8) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5*, 2619–2628.
- (9) Car, P.; Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.
- (10) Grossman, J. C.; Schwegler, E.; Draeger, E. W.; Gygi, F.; Galli, G. Towards an assessment of the accuracy of density functional theory for first principles simulations of water. *J. Chem. Phys.* **2004**, *120*, 300–311.
- (11) Duboue-Dijon, E.; Mason, P. E.; Fischer, H. E.; Jungwirth, P. Hydration and Ion Pairing in Aqueous Mg(2+) and Zn(2+) Solutions: Force-Field Description Aided by Neutron Scattering Experiments and Ab Initio Molecular Dynamics Simulations. *J. Phys. Chem. B* **2018**, *122*, 3296–3306.
- (12) Hutter, J.; Curioni, A. Car-Parrinello molecular dynamics on massively parallel computers. *ChemPhysChem* **2005**, *6*, 1788–1793.
- (13) Andermatt, S.; Cha, J.; Schiffmann, F.; VandeVondele, J. Combining Linear-Scaling DFT with Subsystem DFT in Born-Oppenheimer and Ehrenfest Molecular Dynamics Simulations: From Molecules to a Virus in Solution. *J. Chem. Theory Comput.* **2016**, *12*, 3214–3227.
- (14) Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. Charge Transfer and Polarization in Solvated Proteins from Ab Initio Molecular Dynamics. *J. Phys. Chem. Lett.* **2011**, *2*, 1789–1793.
- (15) Monard, G.; Merz, K. M. Combined Quantum Mechanical/Molecular Mechanical Methodologies Applied to Biomolecular Systems. *Acc. Chem. Res.* **1999**, *32*, 904–911.
- (16) Monard, G.; Prat-Resina, X.; González-Lafont, A.; Lluch, J. M. Determination of enzymatic reaction pathways using QM/MM methods. *Int. J. Quant. Chem.* **2003**, *93*, 229–244.

- (17) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 1198–1229.
- (18) 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.
- (19) Acevedo, O.; Jorgensen, W. L. Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions. *Acc. Chem. Res.* **2010**, *43*, 142–151.
- (20) Hermann, J. C.; Ridder, L.; Mulholland, A. J.; Holtje, H.-D. Identification of Glu166 as the general base in the acylation reaction of class A beta-lactamases through QM/MM modeling. *J. Am. Chem. Soc.* **2003**, *125*, 9590–9591.
- (21) Hermann, J. C.; Hensen, C.; Ridder, L.; Mulholland, A. J.; Holtje, H.-D. Mechanisms of antibiotic resistance: QM/MM modeling of the acylation reaction of a class A beta-lactamase with benzylpenicillin. *J. Am. Chem. Soc.* **2005**, *127*, 4454–4465.
- (22) Hermann, J. C.; Pradon, J.; Harvey, J. N.; Mulholland, A. J. High level QM/MM modeling of the formation of the tetrahedral intermediate in the acylation of wild type and K73A mutant TEM-1 class A beta-lactamase. *J. Phys. Chem. A* **2009**, *113*, 11984–11994.
- (23) Swiderek, K.; Arafet, K.; Kohen, A.; Moliner, V. Benchmarking Quantum Mechanics/Molecular Mechanics (QM/MM) Methods on the Thymidylate Synthase-Catalyzed Hydride Transfer. *J. Chem. Theory Comput.* **2017**, *13*, 1375–1388.
- (24) Delgado, M.; Görlich, S.; Longbotham, J. E.; Scrutton, N. S.; Hay, S.; Moliner, V.; Tuñón, I. Convergence of Theory and Experiment on the Role of Preorganization, Quantum Tunneling, and Enzyme Motions into Flavoenzyme-Catalyzed Hydride Transfer. *ACS Catal.* **2017**, *7*, 3190–3198.
- (25) Saura, P.; Suardíaz, R.; Masgrau, L.; González-Lafont, À.; Rosta, E.; Lluch, J. M. Understanding the Molecular Mechanism of the Ala-versus-Gly Concept Controlling the Product

Specificity in Reactions Catalyzed by Lipoygenases: A Combined Molecular Dynamics and QM/MM Study of Coral 8R-Lipoxygenase. *ACS Catal.* **2017**, *7*, 4854–4866.

- (26) Field, M. J.; Bash, P. A.; Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* **1990**, *11*, 700–733.
- (27) Reuter, N.; Dejaegere, A.; Maignet, B.; Karplus, M. Frontier Bonds in QM/MM Methods: A Comparison of Different Approaches. *J. Phys. Chem. A* **2000**, *104*, 1720–1735.
- (28) Antes, I.; Thiel, W. Adjusted Connection Atoms for Combined Quantum Mechanical and Molecular Mechanical Methods. *J. Phys. Chem. A* **1999**, *103*, 9290–9295.
- (29) Zhang, Y. Pseudobond ab initio QM/MM approach and its applications to enzyme reactions. *Theor. Chem. Acc.* **2006**, *116*, 43–50.
- (30) Théry, V.; Rinaldi, D.; Rivail, J.-L.; Maignet, B.; Ferenczy, G. G. Quantum mechanical computations on very large molecular systems: The local self-consistent field method. *J. Comput. Chem.* **1994**, *15*, 269–282.
- (31) Assfeld, X.; Rivail, J.-L. Quantum chemical computations on parts of large molecules: the ab initio local self consistent field method. *Chem. Phys. Lett.* **1996**, *263*, 100–106.
- (32) Monard, G.; Loos, M.; Théry, V.; Baka, K.; Rivail, J.-L. Hybrid classical quantum force field for modeling very large molecules. *Int. J. Quant. Chem.* **1996**, *58*, 153–159.
- (33) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. A Generalized Hybrid Orbital (GHO) Method for the Treatment of Boundary Atoms in Combined QM/MM Calculations. *J. Phys. Chem. A* **1998**, *102*, 4714–4721.
- (34) Pu, J.; Gao, J.; Truhlar, D. G. Combining Self-Consistent-Charge Density-Functional Tight-Binding (SCC-DFTB) with Molecular Mechanics by the Generalized Hybrid Orbital (GHO) Method. *J. Phys. Chem. A* **2004**, *108*, 5454–5463.

- (35) Luque, F. J.; Reuter, N.; Cartier, A.; Ruiz-López, M. F. Calibration of the Quantum/Classical Hamiltonian in Semiempirical QM/MM AM1 and PM3 Methods. *J. Phys. Chem. A* **2000**, *104*, 10923–10931.
- (36) Wang, Q.; Bryce, R. A. Improved Hydrogen Bonding at the NDDO-Type Semiempirical Quantum Mechanical/Molecular Mechanical Interface. *J. Chem. Theory Comput.* **2009**, *5*, 2206–2211.
- (37) Gokcan, H.; Kratz, E.; Darden, T. A.; Piquemal, J.-P.; Cisneros, G. A. QM/MM Simulations with the Gaussian Electrostatic Model: A Density-based Polarizable Potential. *J. Phys. Chem. Lett.* **2018**, *9*, 3062–3067.
- (38) Tu, Y.; Laaksonen, A. On the effect of Lennard-Jones parameters on the quantum mechanical and molecular mechanical coupling in a hybrid molecular dynamics simulation of liquid water. *J. Chem. Phys.* **1999**, *111*, 7519–7525.
- (39) Freindorf, M.; Shao, Y.; Furlani, T. R.; Kong, J. Lennard-Jones parameters for the combined QM/MM method using the B3LYP/6-31G*/AMBER potential. *J. Comput. Chem.* **2005**, *26*, 1270–1278.
- (40) Koyano, Y.; Takenaka, N.; Nakagawa, Y.; Nagaoka, M. On the Importance of Lennard-Jones Parameter Calibration in QM/MM Framework: Reaction Path Tracing via Free Energy Gradient Method for Ammonia Ionization Process in Aqueous Solution. *Bull. Chem. Soc. Jap.* **2010**, *83*, 486–494.
- (41) Pitarch, J.; Pascual-Ahuir, J.-L.; Silla, E.; Tuñón, I. A quantum mechanics/molecular mechanics study of the acylation reaction of TEM1 β -lactamase and penicillanate. *J. Chem. Soc., Perkin Trans. 2* **2000**, 761–767.
- (42) Meroueh, S. O.; Fisher, J. F.; Schlegel, H. B.; Mobashery, S. Ab initio QM/MM study of class A beta-lactamase acylation: dual participation of Glu166 and Lys73 in a concerted base promotion of Ser70. *J. Am. Chem. Soc.* **2005**, *127*, 15397–15407.

- (43) Pople, J. A.; Beveridge, D. L. *Approximate Molecular Orbital Theory*; McGraw-Hill Book Company: New York, 1970.
- (44) Thiel, W. Semiempirical quantum-chemical methods: Semiempirical quantum-chemical methods. *WIREs Comput. Mol. Sci.* **2014**, *4*, 145–157.
- (45) Christensen, A. S.; Kubar, T.; Cui, Q.; Elstner, M. Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications. *Chem. Rev.* **2016**, *116*, 5301–5337.
- (46) Dixon, S. L.; Merz Jr., K. M. Semiempirical molecular orbital calculations with linear system size scaling. *J. Chem. Phys.* **1996**, *104*, 6643–6649.
- (47) Dixon, S. L.; Merz, K. M. Fast, accurate semiempirical molecular orbital calculations for macromolecules. *J. Chem. Phys.* **1997**, *107*, 879.
- (48) Case, D. A.; Brozell, S. R.; Cerutti, D. S.; T. E. Cheatham, I.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. AMBER 2018. University of California: San Francisco, 2018.
- (49) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *WIREs Comput. Mol. Sci.* **2013**, *3*, 198–210.
- (50) Dewar, M. J. S.; Thiel, W. Ground states of molecules. 39. MNDO results for molecules containing hydrogen, carbon, nitrogen, and oxygen. *J. Am. Chem. Soc.* **1977**, *99*, 4907–4917.

- (51) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- (52) Stewart, J. J. P. Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.* **1989**, *10*, 209–220.
- (53) Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. RM1: a reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *J. Comput. Chem.* **2006**, *27*, 1101–1111.
- (54) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. PDDG/PM3 and PDDG/MNDO: improved semiempirical methods. *J. Comput. Chem.* **2002**, *23*, 1601–1622.
- (55) Nam, K.; Cui, Q.; Gao, J.; York, D. M. Specific Reaction Parametrization of the AM1/d Hamiltonian for Phosphoryl Transfer Reactions: H, O, and P Atoms. *J. Chem. Theory Comput.* **2007**, *3*, 486–504.
- (56) Govender, K.; Gao, J.; Naidoo, K. J. AM1/d-CB1: A Semiempirical Model for QM/MM Simulations of Chemical Glycobiology Systems. *J. Chem. Theory Comput.* **2014**, *10*, 4694–4707.
- (57) Bernal-Uruchurtu, M. I.; Ruiz-López, M. F. Basic ideas for the correction of semiempirical methods describing H-bonded systems. *Chem. Phys. Lett.* **2000**, *330*, 118–124.
- (58) Arillo-Flores, O. I.; Ruiz-López, M. F.; Bernal-Uruchurtu, M. I. Can semi-empirical models describe HCl dissociation in water? *Theor. Chem. Acc.* **2007**, *118*, 425–435.
- (59) Bernal-Uruchurtu, M. I.; Martins-Costa, M. T. C.; Millot, C.; Ruiz-López, M. F. Improving description of hydrogen bonds at the semiempirical level: water–water interactions as test case. *J. Comput. Chem.* **2000**, *21*, 572–581.
- (60) Harb, W.; Bernal-Uruchurtu, M. I.; Ruiz-Lpez, M. F. An improved semiempirical method for hydrated systems. *Theor. Chem. Acc.* **2004**, *112*.

- (61) Marion, A.; Monard, G.; Ruiz-Lopez, M. F.; Ingrosso, F. Water interactions with hydrophobic groups: assessment and recalibration of semiempirical molecular orbital methods. *J. Chem. Phys.* **2014**, *141*, 034106.
- (62) Thiriot, E.; Monard, G. Combining a genetic algorithm with a linear scaling semiempirical method for protein-ligand docking. *J. Molec. Struct. (THEOCHEM)* **2009**, *898*, 31–41.
- (63) Monard, G.; Bernal-Uruchurtu, M. I.; van der Vaart, A.; Merz, K. M. J.; Ruiz-Lopez, M. F. Simulation of liquid water using semiempirical Hamiltonians and the divide and conquer approach. *J. Phys. Chem. A* **2005**, *109*, 3425–3432.
- (64) Goedecker, S. Linear scaling electronic structure methods. *Rev. Mod. Phys.* **1999**, *71*, 1085–1123.
- (65) Daniels, A. D.; Millam, J. M.; Scuseria, G. E. Semiempirical methods with conjugate gradient density matrix search to replace diagonalization for molecular systems containing thousands of atoms. *J. Chem. Phys.* **1997**, *107*, 425.
- (66) Daniels, A. D.; Scuseria, G. E. What is the best alternative to diagonalization of the Hamiltonian in large scale semiempirical calculations? *J. Chem. Phys.* **1999**, *110*, 1321.
- (67) van der Vaart, A.; Suárez, D.; Merz Jr., K. M. Critical assessment of the performance of the semiempirical divide and conquer method for single point calculations and geometry optimizations of large chemical systems. *J. Chem. Phys.* **2000**, *113*, 10512–10523.
- (68) Dixon, S. L.; van der Vaart, A.; Gogonea, V.; Vincent, J. J.; Brothers, E. N.; Westerhoff, L. M.; Merz, K. M., Jr. The DivCon99 program. The Pennsylvania State University: Department of Chemistry, 152 Davey Laboratory, The Pennsylvania State University, University Park, PA-16802 U.S.A., 1999.
- (69) Stewart, J. J. P. MOPAC: A semiempirical molecular orbital program. *J. Comput. Aided Mol. Des.* **1990**, *4*, 1–103.

- (70) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (71) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577.
- (72) Nam, K.; Gao, J.; York, D. M. An Efficient Linear-Scaling Ewald Method for Long-Range Electrostatic Interactions in Combined QM/MM Calculations. *J. Chem. Theory Comput.* **2005**, *1*, 2–13.
- (73) Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* **1921**, *369*, 253–287.
- (74) Toukmaji, A. Y.; Board Jr., J. A. Ewald summation techniques in perspective: a survey. *Comp. Phys. Comm.* **1996**, *95*, 73–92.
- (75) Roe, D. R.; Cheatham III, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.