

Combining Iteration-Free Polarization with Large Time Step Stochastic-Isokinetic Integration

Alex Albaugh^{1,2}, Mark Tuckerman⁵⁻⁷, and Teresa Head-Gordon^{1-4*}

¹*Pitzer Theory Center, Departments of* ²*Chemical and Biomolecular Engineering,* ³*Chemistry, and*

⁴*Bioengineering, University of California, Berkeley, CA 94720*

⁵*Department of Chemistry and* ⁶*Courant Institute of Mathematical Sciences,*

New York University, New York, NY 10003, USA

⁷*NYU-ECNU, Center for Computational Chemistry at NYU, Shanghai, Shanghai 200062, China*

In order to accelerate molecular dynamics simulations using polarizable force fields, we combine a new extended Lagrangian approach that eliminates the self-consistent field step (iEL/0-SCF) with a stochastic integration scheme that allows for a long time step using a multi-time stepping algorithm (SIN(R)). We consider different algorithms for the combined scheme that places different components of the non-bonded forces into different timescales, as well as splitting individual non-bonded forces across timescales, to demonstrate that the combined method works well for bulk water as well as for a concentrated salt solution, aqueous peptide, and solvated protein. Depending on system and desired accuracy, the iEL/0-SCF and SIN(R) combination yields lower bound computational speed-ups of ~6-8 relative to a molecular dynamics Verlet integration using a standard SCF solver implemented in the reference program TINKER 8.1. The combined approach embodies a significant advance for equilibrium simulations in the canonical ensemble of many-body potential energy surfaces for condensed phase systems with speed-ups that exceed what is possible by either method alone.

*corresponding author: thg@berkeley.edu

INTRODUCTION

Over the last 50 years molecular simulation has relied on the most tractable classical model for condensed phase simulation which assumes pairwise additivity of molecular interactions.¹ However, pairwise additivity breaks down when one considers “asymmetric environments”² such as the heterogeneity of solutions and interfaces³⁻⁴, electric fields in protein active sites⁵⁻⁶, environments used to compute hydration free energies of amino acid analogues⁷⁻⁹, solvation structures around peptides¹⁰⁻¹¹, and structural ensembles for intrinsically disordered proteins¹². This has led to the development of more advanced force fields¹³ that include many-body effects such as polarizability¹⁴⁻⁴⁵. In order to gain the full advantage of these advanced potential energy surfaces, better algorithms⁴⁶⁻⁴⁹ and software implementations⁵⁰ must be developed so as to lower the barrier for their use for large molecular systems simulated on long timescales.

Recently we introduced the iEL/SCF method to classical polarization that reduces the number of SCF iterations by dynamically integrating a set of auxiliary dipoles that serve as a time-reversible initial guess for the SCF solver for classical polarization⁴⁶, and extended its use for Born-Oppenheimer molecular dynamics (BOMD)⁵¹. We subsequently built on that work by developing the iEL/0-SCF method that eliminates the SCF iterations for the classical polarizable induced dipoles altogether.⁴⁸⁻⁴⁹ The iEL/0-SCF method gives excellent results that match the standard SCF solution and thus physical properties for water, dilute to concentrated salt solutions, and large solvated proteins for polarization models that use induced point dipoles⁴⁸ as well as Drude particles⁴⁸ and fluctuating charges⁴⁹, all of which have been implemented in TINKER⁵². This work therefore reduced the cost of full mutual polarization to the two-body cost of just the chosen level of pairwise permanent electrostatics.⁴⁸

Tuckerman and colleagues have developed a numerical integration scheme for molecular dynamics that allows use of very large time steps^{47, 53-55} that would otherwise be unstable within standard integration approaches such as velocity Verlet. The stochastic-isokinetic integration (SII) algorithm works by constraining the amount of kinetic energy in each degree of freedom in a simulation so that unphysical flow of energy between fast and slow modes, known as resonance, is prevented. By combining SII with the reversible reference system propagator algorithm (RESPA) method⁵⁶, which integrates fast and slow modes on different time scales, the resulting SIN(R) method is able to integrate the equations of motion stably with much longer time steps to increase the computational efficiency.^{47, 53-54} The SIN(R) method was recently applied to the integration of the

equations of motion for the classical polarizable model AMOEBA, which used the standard SCF solution for the induced dipoles.⁴⁷

In this work, we combine iEL/0-SCF, which removes the inefficiency of solving the SCF calculation at each time step, with SIN(R), which removes resonance effects in any integrator applied to either a standard SCF or an iEL/0-SCF solution for polarization. We consider the tradeoffs of accuracy and computational speed-ups for different SII-RESPA schemes that place different components of the forces into different timescales for increased computational benefits. We also consider the case where individual forces are split between timescales, also known as force splitting, to further increase accuracy and computational performance. Using the two approaches together, we show that we can further increase the efficiency and acceleration of molecular dynamics simulations using polarizable force fields, not only for water but also for more complicated systems for which polarization is important, such as concentrated salt solutions, aqueous zwitterionic peptides, and large solvated protein systems.

THEORY

We briefly review the iEL/0-SCF⁴⁸ and SIN(R)^{47, 53-54} approaches and present the modifications necessary for these methods to be combined.

iEL/0-SCF. The iEL/0-SCF method for induced dipoles, $\boldsymbol{\mu}$, performs a polarization calculation by introducing a set of auxiliary induced dipoles $\boldsymbol{a}$ which are driven dynamically by an equation of motion (Eq. 1a) along with the atomic degrees of freedom, $\boldsymbol{r}$ (Eq. 1b)

$$\ddot{\boldsymbol{a}}_i = \omega^2 \gamma (\boldsymbol{\mu}_i - \boldsymbol{a}_i) \quad (1a)$$

$$m_i \ddot{\boldsymbol{r}}_i = - \left. \frac{dU(\boldsymbol{r}, \boldsymbol{a})}{d\boldsymbol{r}_i} \right|_{\boldsymbol{a}} \quad (1b)$$

where $\boldsymbol{r}_i$, $\boldsymbol{\mu}_i$ and $\boldsymbol{a}_i$ are the position, induced dipole, and associated auxiliary induced dipole, of the i -th atom. Eq. (1a) drives the auxiliary dipoles according to a harmonic potential that seeks to keep the auxiliaries close to a ground state solution, where ω is the frequency of the harmonic potential and is set to its maximum stable value for velocity Verlet integration $\sqrt{2}/\Delta t$ where Δt is the simulation time step.⁵⁷ γ is a tunable parameter that comes from estimating a ground state SCF solution from a simple linear mixing of real and auxiliary dipoles as discussed in previous work⁴⁸.

$$\boldsymbol{\mu}_{SCF,i} \approx \gamma \boldsymbol{\mu}_i + (1 - \gamma) \boldsymbol{a}_i \quad (1c)$$

γ is set to 0.9 for all test systems, although we examine its behavior for the large protein system. To be clear Eq. (1c) is not used for the calculation of the true polarization energy and forces, but only

applies to the derivation of the auxiliary equation of motion (Eq. (1a)). We couple the integration of the auxiliary dipoles to an auxiliary thermostat such that the auxiliary pseudo temperature, $T_{aux} = 1/3\langle\dot{\mathbf{a}}_i^2\rangle$, is controlled to a set point as described previously⁴⁸.

Eq. (1b) is the familiar Newtonian equation of motion for the atom positions evolving under the potential, $U(\mathbf{r}, \mathbf{a})$

$$U(\mathbf{r}, \mathbf{a}) = U^{other}(\mathbf{r}) + U^{polar}(\mathbf{r}, \mathbf{a}) = U^{other}(\mathbf{r}) + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \boldsymbol{\mu}_i^T \mathbf{C}_{ij} \boldsymbol{\mu}_j - \sum_{i=1}^N \boldsymbol{\mu}_i^T \mathbf{E}_i \quad (2a)$$

$$\boldsymbol{\mu}_i = \alpha_i \mathbf{E}_i + \alpha_i \sum_{j=1}^N \mathbf{T}'_{ij} \mathbf{a}_j \quad (2b)$$

where $U^{other}(\mathbf{r})$ can include bonds, angles, van der Waals, and permanent electrostatic terms, and $U^{polar}(\mathbf{r}, \mathbf{a})$ is a general form of the polarization potential that does not assume an SCF solution; in Eq. (2) $\mathbf{E}_i = \sum_j \mathbf{T}_{ij} \mathbf{M}_j$ is the permanent electrostatic field due to the permanent moments in the system, $\mathbf{M}_i$, interacting through the full interaction tensor, $\mathbf{T}_{ij}$ between the i^{th} and j^{th} sites, α_i is the polarizability of the i -th atom, while $\mathbf{T}'_{ij}$ is the induced dipole-induced dipole interaction matrix between sites i and j , and $\mathbf{C}_{ij} = \alpha_j^{-1} \delta_{ij} - \mathbf{T}'_{ij}$. From Eq. (2a), the polarization potential is solved for the real dipoles using Eq. (2b), i.e. a single update needed to define the real dipoles at each time step.

Stochastic-Isokinetic Integration. SII constrains the total kinetic energy in each degree of freedom to be a fixed quantity that can then transfer between the real degrees of freedom and the thermostat variables. The equations of motion under the isokinetic scheme are given in Eq. (3)⁴⁷.

$$dr_{i,\alpha} = v_{i,\alpha} dt \quad \alpha \in x, y, z \quad (3a)$$

$$dv_{i,\alpha} = \left[\frac{F_{i,\alpha}(\mathbf{r})}{m_i} - \lambda_{i,\alpha} v_{i,\alpha} \right] dt \quad (3b)$$

Where $r_{i,\alpha}$ is the α component of the i -th atom's position and $v_{i,\alpha}$ and $F_{i,\alpha}(\mathbf{r})$ are the associated velocity and force, respectively, and m_i is the mass. The Lagrange multiplier for each degree of freedom, $\lambda_{i,\alpha}$ is introduced to enforce an isokinetic constraint between the physical velocity $v_{i,\alpha}$ and a stochastic Nosé-Hoover (NH) thermostat⁵⁸ employing two additional sets of velocities $v_{1,i,\alpha}^{(k)}$ and $v_{2,i,\alpha}^{(k)}$, $k = 1, \dots, L$ is the number of stochastic NH thermostats per degree of freedom. The stochastic NH equations of motion are

$$dv_{1,i,\alpha}^{(k)} = -\lambda_{i,\alpha} v_{1,i,\alpha}^{(k)} dt - v_{2,i,\alpha}^{(k)} v_{1,i,\alpha}^{(k)} dt \quad (4a)$$

$$dv_{2,i,\alpha}^{(k)} = \frac{Q_1(v_{1,i,\alpha}^{(k)})^2 - k_B T}{Q_2} dt - \gamma_{SI} v_{2,i,\alpha}^{(k)} dt + \sigma dw_{i,\alpha}^{(k)} \quad (4b)$$

where a friction term with friction constant γ_{SI} is employed, and Q_1 and Q_2 are mass parameters with associated time scales τ_1 and τ_2 via the usual expressions $Q_i = k_B T \tau_i^2$. The last term in Eq. (4b) is a stochastic Ornstein-Uhlenbeck (OU) process $dw_{i,\alpha}^{(k)}$ to ensure ergodicity, where $\sigma = \sqrt{k_B T \gamma_{SI} / Q_2}$. Eqs. (4a) and (4b) are coupled to the physical velocity $v_{i,\alpha}$ via the a kinetic energy constraint of the form

$$m_i v_{i,\alpha}^2 + \frac{L}{L+1} \sum_{k=1}^L Q_1(v_{1,i,\alpha}^{(k)})^2 = L k_B T \quad (5)$$

From this constraint condition, an analytical expression for the Lagrange multiplier $\lambda_{i,\alpha}$ is derived by differentiating Eq. (5) with respect to time, substituting in Eqs. (3b) and (4a) for the time derivatives, and solving for the multiplier. The resulting expression is

$$\lambda_{i,\alpha} = \frac{v_{i,\alpha} F_{i,\alpha}(\mathbf{r}) - \frac{L}{L+1} \sum_{k=1}^L Q_1(v_{1,i,\alpha}^{(k)})^2 v_{2,i,\alpha}^{(k)}}{m_i v_{i,\alpha}^2 + \frac{L}{L+1} \sum_{k=1}^L Q_1(v_{1,i,\alpha}^{(k)})^2} \quad (6)$$

Eq. (6) is then substituted back into the equations of motion to obtain the full set of SII equations.

Combining iEL/0-SCF and SII. When combining iEL/0-SCF with SII, the evolution of the auxiliary induced dipoles needs to be formulated in a manner similar to Eq. (3) to yield Eq. (7)

$$da_{i,\alpha} = v_{i,\alpha}^a dt \quad (7a)$$

$$dv_{i,\alpha}^a = [\gamma \omega^2 (\mu_{i,\alpha} - a_{i,\alpha}) - \lambda_{i,\alpha}^a v_{i,\alpha}] dt \quad (7b)$$

where we have used the force due to the harmonic potential, and using analogous constraint equations using the analytical expression for the Lagrange multiplier given by Eq. (8).

$$\lambda_{i,\alpha}^a = \frac{v_{i,\alpha}^a \gamma \omega^2 (\mu_{i,\alpha} - a_{i,\alpha}) - \frac{L}{L+1} \sum_{k=1}^L Q_1^a (v_{1,i,\alpha}^{a(k)})^2 v_{2,i,\alpha}^{a(k)}}{(v_{i,\alpha}^a)^2 + \frac{L}{L+1} \sum_{k=1}^L Q_1 (v_{1,i,\alpha}^{a(k)})^2} \quad (8)$$

With the Lagrange multiplier, the pseudo isokinetic constraint for each auxiliary degree of freedom is given by Eq. (9)

$$(v_{i,\alpha}^a)^2 + \frac{L}{L+1} \sum_{k=1}^L Q_1^a (v_{1,i,\alpha}^{a(k)})^2 = L T_{aux} \quad (9)$$

which couples the auxiliary dipoles to an associated set of stochastic NH thermostat variables that evolve according to

$$dv_{1,i,\alpha}^{a(k)} = -\lambda_i^a v_{1,i,\alpha}^{a(k)} dt - v_{2,i,\alpha}^{a(k)} v_{1,i,\alpha}^{a(k)} dt \quad (10a)$$

$$dv_{2,i,\alpha}^{a(k)} = \frac{Q_1^a v_{1,i,\alpha}^{a(k)2} - T_{aux}}{Q_2^a} dt - \gamma_{SI}^a v_{2,i,\alpha}^{a(k)} dt + \sigma dw_{i,\alpha}^{a(k)} \quad (10b)$$

Again, for each i, α auxiliary degree of freedom, we have introduced L thermostat velocities $v_{1,i,\alpha}^{a(k)}$ and $v_{2,i,\alpha}^{a(k)}$, with k running from 1 to L . The first set, $v_{1,i,\alpha}^{a(k)}$ (Eq. 10a), couple to the isokinetic constraint, and the second set couple to a driving force between the actual and set point auxiliary temperature T_{aux} , a friction term, and an OU stochastic process. Here $Q_1^a = T_{aux}(\tau_1^a)^2$ and $Q_2^a = T_{aux}(\tau_2^a)^2$, γ_{SI}^a controls the friction, and $\sigma = \sqrt{T_{aux}\gamma_{SI}^a/Q_2}$.

Multiple Time Stepping. SII eliminates unphysical energy flow between fast and slow modes. To take full advantage of the SII approach, it is combined with a multiple time-scale RESPA integration⁴⁷, where different components of the force field are integrated with different time steps⁵⁶. More specifically, all of the bonded forces $\mathbf{F}_s$ are evaluated at the shortest time step of Δt_s , and the non-bonded interactions are then partitioned between a single longer time step, or further partitioned into an intermediate timescale integrated at a time step of Δt_m , and a longer (outer) time scale with a time step of Δt , with corresponding forces $\mathbf{F}_m$ and $\mathbf{F}_l$, respectively. The longer time scale time step must be an integer multiple of the shorter time step(s), that is $\Delta t = m\Delta t_m = ms\Delta t_s$ for integers m and s . Tuckerman and co-workers refer to this resulting algorithm as the Stochastic Isokinetic Nosé-Hoover (RESPA) or SIN(R) scheme, which can involve two time steps (Δt_s and Δt) or three time steps (Δt_s , Δt_m , and Δt), both of which we consider in the results section.

The three non-bonded interactions for AMOEBA, polarization - including both real and auxiliary dipoles- as well as permanent electrostatics and van der Waals interactions, are placed in the intermediate and/or long time scales in various combinations to determine optimum efficiency and accuracy, as discussed in the Results section. We first consider the case of partitioning individual but complete nonbonded forces into a long timescale using a 2-step SIN(R). The second case considers the splitting of the individual non-bonded forces into short-ranged and long-ranged components using a spatial cutoff, which are evaluated at Δt_m and Δt , respectively under a 3-step SIN(R). In order to smooth the split forces at the boundary, the following switching function between sites, S_{ij} , is used

$$S_{ij}(r_{ij}; r_c, \lambda) = \begin{cases} 1 & r_{ij} \leq r_c - \lambda \\ g_{ij}(r_{ij}; r_c, \lambda) & r_c - \lambda \leq r \leq r_c \\ 0 & r \geq r_c \end{cases} \quad (11a)$$

$$g(r_{ij}; r_c, \lambda) = 1 + u_{ij}^3(15u_{ij} - 6u_{ij}^2 - 10) \quad (11b)$$

$$u_{ij} = \frac{1}{\lambda}(r_{ij} - r_c + \lambda) \quad (11c)$$

where r_{ij} is the distance between the i^{th} and j^{th} sites, r_c is the cutoff distance between short and long range, and λ is the healing length over which the switch is applied. The van der Waals forces can be split straightforwardly into short- and long-ranged components using equation (11) with an $r_c = 7.5$ Å and $\lambda = 0.5$ Å. For the electrostatic and polarization interactions, which are calculated using the particle-mesh Ewald method⁵⁹, their real space contributions are split into short and long range forces with the applied cutoff $r_c = 5.0$ Å and $\lambda = 0.5$ Å, while the reciprocal space contributions for the permanent and induced multipole and induced dipole interactions fall entirely into the long range forces.

For the many-body polarization, the full force is calculated at the outer time step and then the short-range contribution is subtracted to give the resulting long-range contribution. In order to maintain stability with large outer time steps, a full SCF polarization calculation is done at the outer time step and iEL/0-SCF is used to treat the short-range real space polarization forces only. While this does introduce SCF iterations into our method, it becomes a negligible amount for large outer time steps for which updates occur much less frequently. Therefore the auxiliary dipoles only represent short-range interactions, $\mathbf{a}_{SR}$, and create short-range real dipoles according to $\boldsymbol{\mu}_{SR} = \alpha \mathbf{E}_{SR} + \alpha \mathbf{T}' \mathbf{a}_{SR}$. For completeness, Eq. (12) is the short-range polarization gradient of the iEL/0-SCF method which accounts for the switching operator in Eq. (11).

$$\begin{aligned} \left. \frac{dU_{SR}^{pol}(\mathbf{r}^N, \mathbf{a}^N)}{d\mathbf{r}_i} \right|_{\mathbf{a}^N} = & -\frac{1}{2} \sum_{j,k} S_{jk} [\boldsymbol{\mu}_{SR,j}^d]^T \frac{\partial \mathbf{T}'_{jk}}{\partial \mathbf{r}_i} \boldsymbol{\mu}_{SR,k}^p + \\ & \frac{1}{2} \sum_{j,k} S_{jk} \left[\boldsymbol{\mu}_{SR,j}^d - \alpha_j \sum_l \mathbf{T}'_{lj} \boldsymbol{\mu}_{SR,j}^d - \alpha_j \mathbf{E}_{SR,j}^d \right]^T \frac{\partial \mathbf{T}'_{jk}}{\partial \mathbf{r}_i} \mathbf{a}_{SR,k}^p - \\ & \frac{1}{2} \sum_j \left[\alpha_j \sum_l \mathbf{T}'_{lj} \boldsymbol{\mu}_{SR,j}^d + \alpha_j \mathbf{E}_{SR,j}^d \right]^T \frac{\partial \sum_k S_{jk} \mathbf{T}_{jk}^p \mathbf{M}_k}{\partial \mathbf{r}_i} + \\ & \frac{1}{2} \sum_{j,k} S_{jk} \left[\boldsymbol{\mu}_{SR,j}^p - \alpha_j \sum_l \mathbf{T}'_{lj} \boldsymbol{\mu}_{SR,j}^p - \alpha_j \mathbf{E}_{SR,j}^p \right]^T \frac{\partial \mathbf{T}'_{jk}}{\partial \mathbf{r}_i} \mathbf{a}_{SR,k}^d - \\ & \frac{1}{2} \sum_j \left[\alpha_j \sum_l \mathbf{T}'_{lj} \boldsymbol{\mu}_{SR,j}^p + \alpha_j \mathbf{E}_{SR,j}^p \right]^T \frac{\partial \sum_k S_{jk} \mathbf{T}_{jk}^d \mathbf{M}_k}{\partial \mathbf{r}_i} \end{aligned} \quad (12)$$

where the permanent electrostatic field at site i th due to other sites within r_c of i is $\mathbf{E}_{SR}$. The d and p superscripts denote d - and p -scaling of the electrostatic and polarization terms to account for excluded bonded interactions which depend on the polarization group (d) and the bond separation (p), as has been described elsewhere¹⁹.

Simulation Details. The combined iEL/0-SCF and SIN(R) methodology was applied to test systems of 512 water molecules, a highly concentrated 4.66 M MgCl₂ solution, a zwitterionic glycine molecule solvated with 256 water molecules, and the dihydrofolate reductase (DHFR) protein in water. The water, salt, and proteins test cases used the latest AMOEBA force field parameterizations⁶⁰⁻⁶¹; the glycine system used a recent reparameterization of the AMOEBA force field in its zwitterionic state.⁶² For the systems in which individual forces are not split across timescales, the simulation conditions were as follows. All electrostatics were treated via particle-mesh Ewald summation⁵⁹ with a real space cutoff of 7.0 Å. For the scheme employing force splitting, the overall cutoff for real space electrostatic and polarization forces and the van der Waals force was 10.4 Å. The atomic system set point temperature T was 298.0 K and the auxiliary set point pseudo temperature was $5.3 \text{ e}^2 \text{Å}^2/\text{ps}^2$.

For the auxiliary SIN(R), we found that the best auxiliary thermostat time scale parameters were 0.1 ps and 0.001 ps for τ_1^a and τ_2^a , respectively. The auxiliary integration also used a friction parameter, γ_{SI}^a , of 100.0 ps⁻¹. For the atomic integration a value of 0.1 ps was used for τ_1 and τ_2 and a value of 0.01 ps⁻¹ for γ_{SI} . Atomic and auxiliary integrations used either $L = 1$ or $L = 4$ stochastic Nosé-Hoover thermostats, and the thermostat equations of motion were integrated with a 3rd-order Suzuki-Yoshida decomposition and time step reduced by a factor $n_c=5$ ⁶³⁻⁶⁴; when the stochastic NH thermostat is applied at the intermediate time step, the resulting scheme is termed XM-SIN(R), and in this case, we investigate both $L = 1$ and $L = 4$; when it is applied at the outer time step, the scheme is termed XO-SIN(R), and in this case, we use $L = 4$. We integrate with XM-SIN(R) when we use force splitting and XO-SIN(R) otherwise. In either case while the values of the intermediate and outer time steps are varied for analysis purposes, the shortest timescale time step was fixed at $\Delta t_s = 0.25$ fs for $L = 4$ and $\Delta t_s = 0.5$ fs for $L = 1$. Simulations that do not use SII integration for comparison purposes instead employ the velocity Verlet integrator⁶⁵ and standard Nosé-Hoover chain thermostats for temperature control⁶⁶ in the canonical ensemble.

RESULTS

Figure 1 summarizes the polarization properties collected from bulk water using a standard SCF and iEL/0-SCF, both integrated with velocity Verlet (henceforth referred to as the reference or standard method), as well as our new scheme that combines iEL/0-SCF with SII using a single 1 fs time step. It is clear that the equilibrium induced dipole distributions match well between the methods for water oxygen and water hydrogen (Fig. 1a and 1b), and thus the system potential energy is well reproduced (Table 1). The induced dipole time autocorrelations for both water atom types under the SII scheme exhibit deviations from the reference, and the diffusion constants are found to be too slow by a factor of ~ 5 .

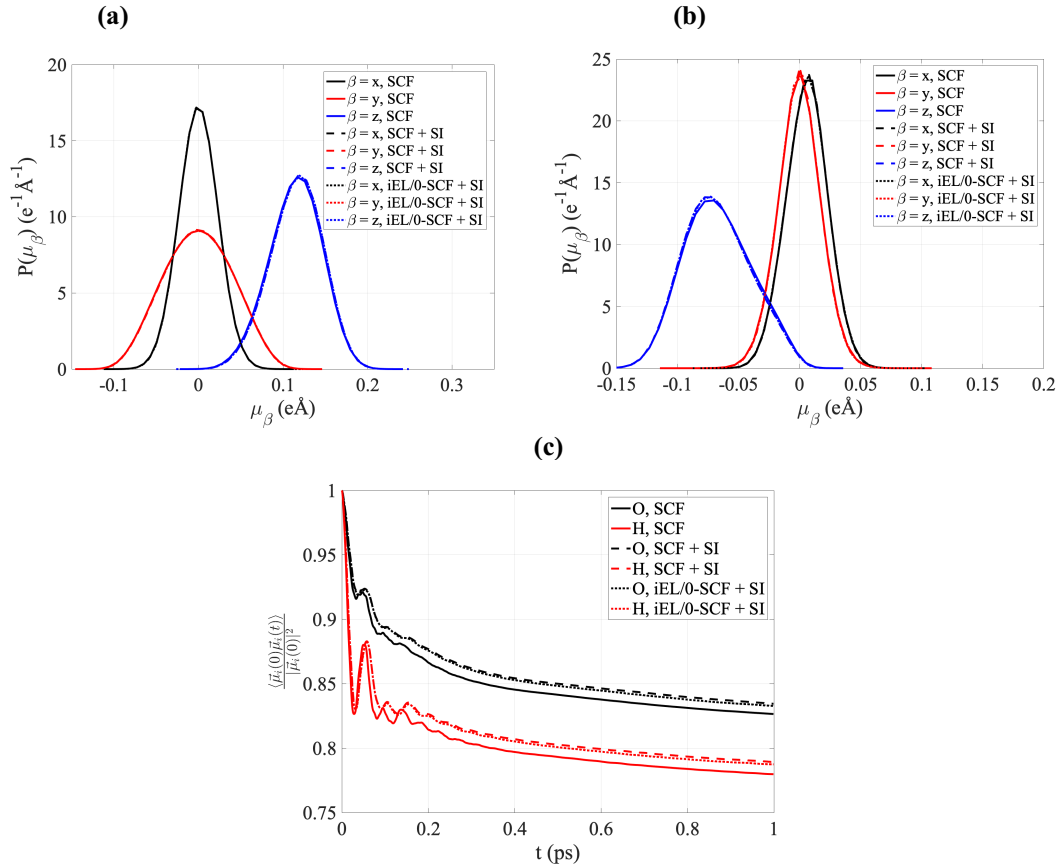


Figure 1. Comparison of standard SCF, SCF+SII, and iEL/0-SCF+SII for water. Induced dipole probability distributions for (a) water oxygen, (b) water hydrogen, (c) water induced dipole autocorrelations. All simulations were performed with a 1.0 fs time step and at a temperature of 298.0 K. The x -, y -, and z -components correspond to internal molecular axes (see ¹⁹ for details).

This is not surprising since SII is only canonical in positions but not velocities⁵⁴, so one can expect deviations in dynamical properties, placing it within the same class of thermodynamic methods such as Monte Carlo, Replica Exchange calculations, or the use of Langevin or Anderson thermostats. For the first time, we present the formulation and results of SII for many-body polarization for systems

other than water, including a 4.66 M MgCl_2 salt solution and a solvated zwitterionic glycine peptide (and later we show results for a protein in water). We find that the iEL/0-SCF and SII combination yields excellent results for the concentrated 4.66 M MgCl_2 salt solutions, as given in Table 1 and Fig. S1. We find more disagreement using SII for the detailed polarization properties for certain atoms of the glycine peptide (Fig. S2), but they are relatively small and the resulting potential energy for the aqueous peptide solution is within the uncertainties of the standard iEL/0-SCF solution (Table 1).

Table 1. *The average total potential energy of bulk water, a concentrated salt solution, and a solvated glycine peptide as calculated from the different methods evaluated in this work.* The iEL/0-SCF method is used as the reference calculation in all reported figures given its excellent agreement with a standard SCF solver. The remaining entries are combining iEL/0-SCF with SII using different RESPA schemes integrated with XO-SIN(R) and XM-SIN(R), with and without force splitting, and using different intermediate Δt_m and outer time steps Δt .

Potential energy $\langle U \rangle$ (kcal/mol)				
Method	Water	4.66 M MgCl_2	Glycine	
Standard SCF	-4621.9 +/- 40.7	-43188 +/- 48	-2431.7 +/- 30.3	
iEL/0-SCF	-4604.6 +/- 42.0	-43175 +/- 51	-2424.7 +/- 30.4	
iEL/0-SCF + SII ($\Delta t_s = \Delta t_m = \Delta t = 1.0$ fs)	-4648.8 +/- 41.9	-43108 +/- 55	-2443.3 +/- 30.0	
iEL/0-SCF + XO-SIN(R) ($L = 4, \Delta t_m = \Delta t = 6.0$ fs)	-4399.0 +/- 55.4	-42876 +/- 64	-2315.9 +/- 39.4	
iEL/0-SCF + XM-SIN(R) (split, $L = 4, \Delta t_m = 3.0$ fs, $\Delta t = 90.0$ fs)	-4483.9 +/- 45.4	-	-	
iEL/0-SCF + XM-SIN(R) (split, $L = 1, \Delta t_m = 3.0$ fs, $\Delta t = 90.0$ fs)	-4577.1 +/- 44.0	-43108 +/- 55	-2163.7 +/- 42.0	
iEL/0-SCF + XM-SIN(R) (split, $L = 1, \Delta t_m = 5.0$ fs, $\Delta t = 90.0$ fs)	-4491.0 +/- 44.2	-4300.9 +/- 57	-	
iEL/0-SCF + XM-SIN(R) (split, $L = 1, \Delta t_m = 3.0$ fs, $\Delta t = 18.0$ fs)	-4643.0 +/- 39.9	-	-2439.6 +/- 30.4	
iEL/0-SCF + XM-SIN(R) (split, $L = 1, \Delta t_m = 5.0$ fs, $\Delta t = 10.0$ fs)	-	-	-2410.0 +/- 30.0	
iEL/0-SCF + XM-SIN(R) ($L = 1, \Delta t_m = \Delta t = 6.0$ fs)	-4525.5 +/- 41.4	-43060 +/- 47	-2382.5 +/- 30.3	
iEL/0-SCF + XM-SIN(R) ($L = 1, \Delta t_m = \Delta t = 8.0$ fs)	-4328.9 +/- 43.8	-42795 +/- 55	-2285.0 +/- 31.8	

Ultimately, the purpose of using SII is to eliminate resonances between fast and slow modes, which in turn empowers SII when combining it with RESPA to separate and maximize the integration timestep of the more expensive but more slowly varying non-bonded forces, leading to a significant savings in computational time within the SIN(R) approach. The actual integration performed with SIN(R) involves a choice as to whether to apply the NH thermostats at the outer time step (XO-SIN(R)), which requires more ($L=4$) thermostat variables versus application of the NH thermostats at an intermediate time step (XM-SIN(R)), or inner most time step (XI-SIN(R)), which we show requires only $L = 1$ set of thermostat variables on each degree of freedom.

We first consider the combination of iEL/0-SCF with a 2-step XO-SIN(R) scheme where the bonded forces are updated with a time step of $\Delta t_s = 0.25$ fs, while all of the non-bonded forces (electrostatic, polarization, van der Waals) are then treated at a variable longer time scale time step, Δt . Figure 2 tracks the average value of the induced dipole molecular moment, $\langle \mu_{\text{mol}} \rangle$ of water; we see that the larger time step can be increased to $\Delta t = 6.0$ fs with only a small error in this property.

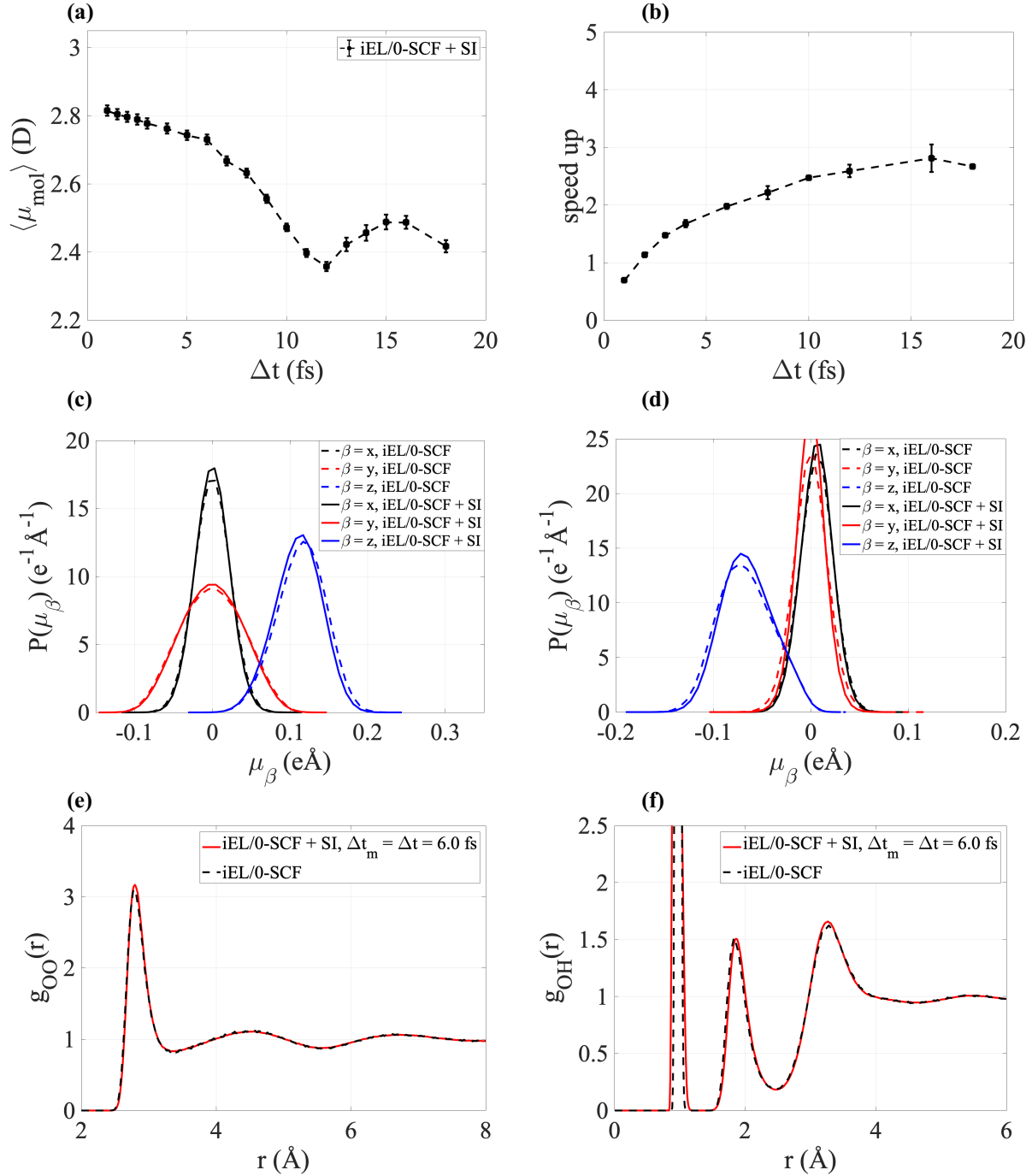


Figure 2: Two-step XO-SIN(R)+iEL/0-SCF compared to standard iEL/0-SCF for water. All bonded forces are in the inner time step ($\Delta t_s = 0.25$ fs) and all non-bonded forces in the outer time step. (a) $\langle \mu_{\text{mol}} \rangle$ and (b) algorithmic speed-ups relative to iEL/0-SCF as a function of the outer RESPA time step, Δt . Probability distributions of the vector components of the induced dipole on (c) oxygen and (d) hydrogen, and radial distribution functions (e) $g_{\text{OO}}(r)$ and (f) $g_{\text{OH}}(r)$ using an outer time step of $\Delta t = 6.0$ fs. Timings were performed with 16 CPU cores and 10000 water molecules. The x -, y -, and z -components correspond to internal molecular axes.¹⁹

A computational speed-up of ~ 2 is achieved relative to standard iEL/0-SCF; the latter is, in turn, ~ 1.6 times faster than an SCF solution using the preconditioned conjugate gradient solver available in TINKER 8.1⁵². To put this in perspective, typical TINKER timings using standard integrators and SCF solvers reported for DHFR in water on 16 cpu cores allows for sampling of ~ 2 ns/day. Thus, we would expect the sampling time to increase to ~ 6.8 ns/day with similar accuracy using the combined 2 step XO-SIN(R) and iEL/0-SCF method with the TINKER 8.1 reference code.

When fixing the outer time step to a value of $\Delta t = 6.0$ fs, more sensitive polarization properties such as the distributions of the vector components of the water dipole show overall good agreement with only a small degradation relative to iEL/0-SCF and an error in the average potential energy, $\langle U \rangle$, of ~ 4 -5% (Table 1); an insensitive property such as the radial distribution function for water maintains good agreement with the standard SCF solution under this 2-step XO-SIN(R) scheme. Furthermore, a single outer time step of $\Delta t = 6$ fs for the 2 step XO-SIN(R) is excellent across all properties for the concentrated MgCl_2 system (Fig. S3 and Table 1), whereas the solvated glycine peptide is comparable to the pure water result, with small degradation of the detailed induced dipole properties and $\langle U \rangle$, and little effect on structural properties (Fig. S4 and Table 1).

In order to fully exploit the SIN(R) approach, we next split the individual non-bonded terms of AMOEBA into their own short- and long-ranged contributions as was done by Margul and Tuckerman⁴⁷, thereby creating a 3-time step SIN(R) scheme that permits substantially larger outer time steps. In order to maintain stability, we used the iEL/0-SCF method for all short-ranged polarization forces at the intermediate time step and perform a full SCF optimization at the outer time step, as described in Methods. Figure 3a shows how the polarization property $\langle \mu_{\text{mol}} \rangle$ changes with different combinations of Δt_m and Δt using the 3 step XM-SIN(R) with the force splitting scheme with $L = 4$. This basic property is well reproduced with $\Delta t_m = 1$ or 3 fs, and for outer time steps reaching up to ~ 120 fs; it is evident that for $\Delta t_m = 5$ fs, $\langle \mu_{\text{mol}} \rangle$ accrues more error at any value of the outer time step.

Overall, we find that the best tradeoff between large outer time step and accuracy using iEL/0-SCF with the 3-step XM-SIN(R) and force splitting with $L = 4$ algorithm is $\Delta t_m = 3$ fs and $\Delta t = \sim 90$ -100 fs for liquid water, with good reproduction of the probability distributions of the vector components of the induced dipole (Figs. 3b and 3c), structural metrics such as $g_{\text{OO}}(r)$ and $g_{\text{OH}}(r)$ (Figs. 3d and 3e), and an acceptable ~ 2 -3% error in $\langle U \rangle$ (Table 1). However, the computational speed-up is no better than a 2-step XO-SIN(R) scheme at any value of the outer time step (Fig. 3f).

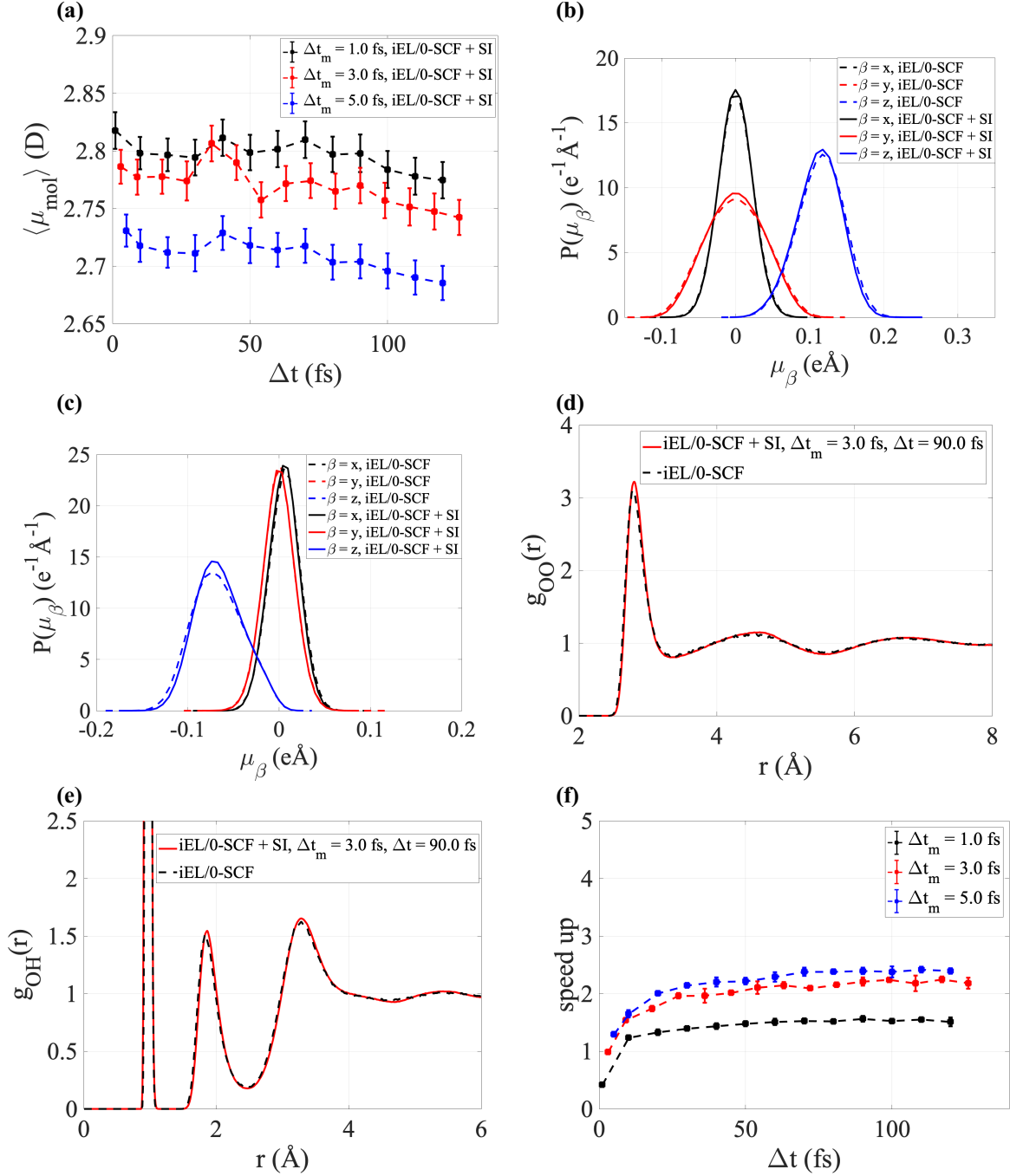


Figure 3: Three-time scale XM-SIN(R)+iEL/0-SCF with $L = 4$ using force splitting compared to standard iEL/0-SCF for pure water. (a) $\langle \mu_{\text{mol}} \rangle$ as a function of the outer time step for various values of the intermediate time step, Δt_m . Probability distributions of the vector components of the induced dipole on (b) oxygen and (c) hydrogen, and radial distribution functions (d) $g_{\text{OO}}(r)$ and (e) $g_{\text{OH}}(r)$ for $\Delta t_m = 3.0$ fs and $\Delta t = 90.0$ fs. (f) speed-ups for bulk water as a function of the outer time step for various values of the intermediate time step, Δt_m measured relative to an iEL/0-SCF single time scale integration at 1.0 fs. Further details are given in Figure 2.

This is not due to the cost of the full SCF update in the outer time step for long-ranged polarization, which is performed infrequently but is a result of the high computational overhead associated with the NH-Langevin thermostats for both the real and auxiliary degrees of freedom when $L = 4$, as these are evaluated every 1-3 fs at the intermediate timestep. Hence, efficiency can be gained by reducing the value of L to 1, which brings with it the added benefit of increased accuracy.

Therefore, we examine an XM-SIN(R) with $L = 1$ to examine the performance and timings of a combined iEL/0-SCF and SII method with a reduced number of thermostat variables. Figure 4a shows how $\langle\mu_{\text{mol}}\rangle$ changes with different combinations of Δt_m and Δt under the 3-step XM-SIN(R) scheme with $L = 1$ and force splitting and the corresponding computational speed-ups for water. We find that there is negligible change in $\langle\mu_{\text{mol}}\rangle$ for all combinations of values of Δt_m and Δt , and intermediate timesteps can be as large as $\Delta t_m = 5$ fs. When fixing $\Delta t_m = 3.0$ or 5.0 fs and $\Delta t = 90.0$ fs, detailed polarization properties such as the induced dipole vector components, as well as RDFs, are excellent (Fig. S5), and the more careful integration using XM-SIN(R) improves the $\langle U \rangle$ error for water to be within the uncertainties of the standard SCF calculation (Table 1). Under the same simulation conditions, we find very good agreement with the standard iEL/0-SCF approach for the concentrated salt solution as well (Fig. S6 and Table 1). Fig. 4b shows that molecular fluids and solutions can be accurately simulated when iEL/0-SCF is combined with the 3-step XM-SIN(R) and force splitting to yield speed-ups that are ~ 4 -5 over the standard iEL/0-SCF calculation, and therefore a factor of ~ 7 -8 relative to a standard SCF solver and Verlet integration in TINKER 8.1.

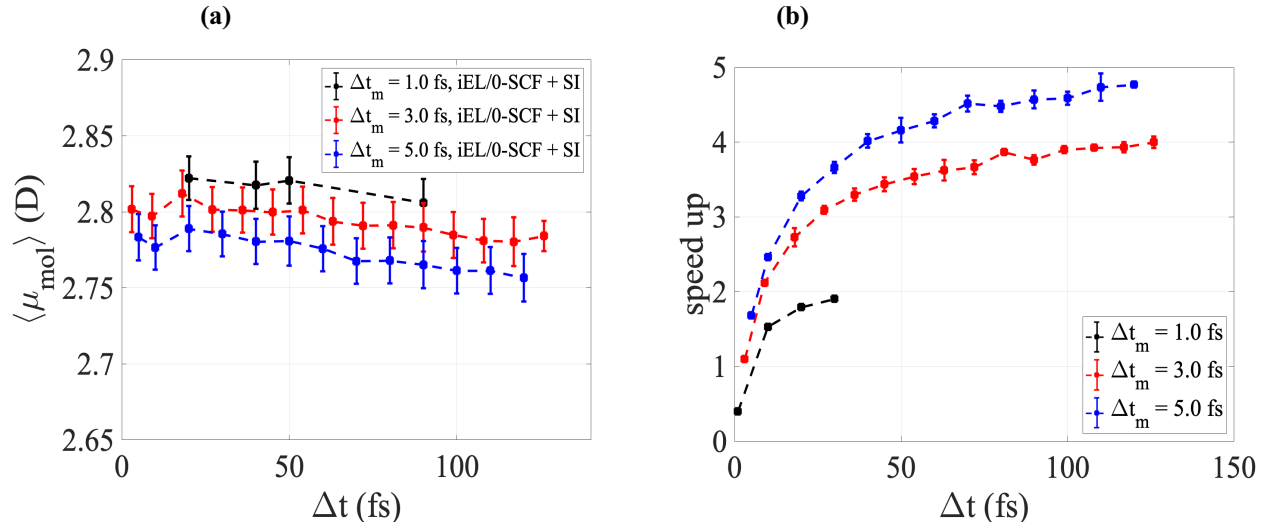


Figure 4: Three-time scale XM-SIN(R)+iEL/0-SCF with $L = 1$ using force splitting compared to standard iEL/0-SCF for pure water. (a) Accuracy in $\langle\mu_{\text{mol}}\rangle$ and (b) speed-ups as a function of the outer time step for various values of the intermediate time step, Δt_m . Timings used 16 cores and 10,000 water molecule simulations with an inner time step of $\Delta t_s = 0.5$ fs.

We also find that the three-time scale XM-SIN(R) and iEL/0-SCF with force splitting method and an outer time step of 90 fs performs well for the dihydrofolate reductase (DHFR) protein in water; Figure 5 compares the root mean square deviation between iEL/0-SCF alone and when combined with the SIN(R) scheme. In our previous iEL/0-SCF study⁴⁸ we found that the γ value of 0.9 (see Eq. (1)) did not work well for the large protein, requiring us to vary γ to a new value of 0.7 or to assign different γ values for protein and water to achieve stability comparable to a standard SCF solution (Figure 5a). However, when iEL/0-SCF is combined with the SI integration, we reach a far more satisfactory result that any γ variation now performs similarly—thus making it insensitive to the value of γ —and the RMSD plateaus at a significantly lower value than found using Verlet integration (Figure 5b). Hence we have shown for the first time that SI integration not only allows for much larger time steps, but it results in a more accurate solution by removing undesired energy partitioning in solvated protein systems using many-body potentials.

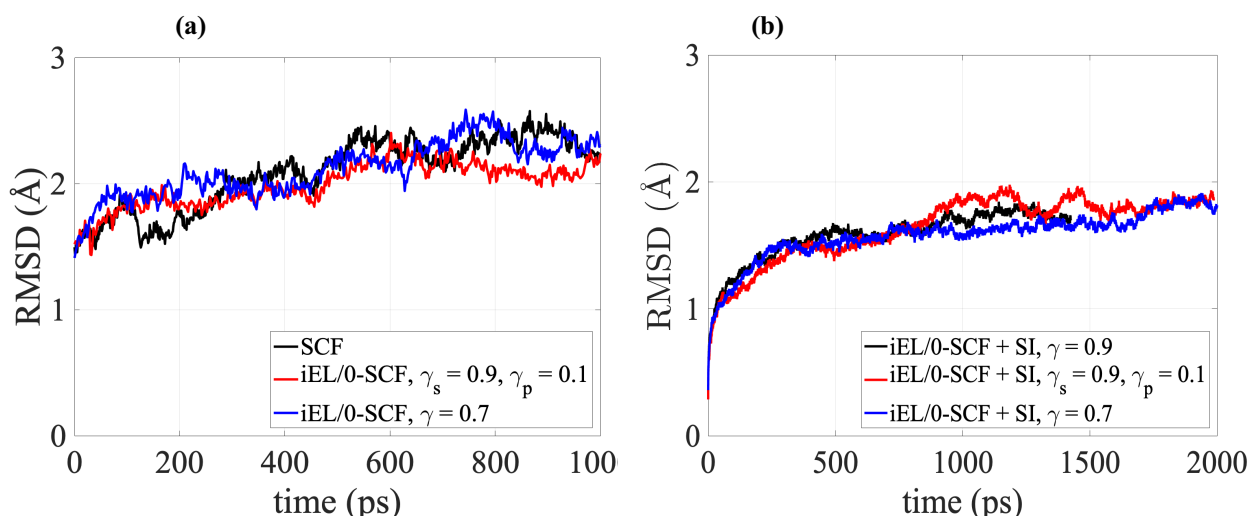


Figure 5: *RMSD of dihydrofolate reductase (DHFR) using different integration schemes.* (a) the standard SCF method and iEL/0-SCF using a single value of $\gamma = 0.7$ and dual values of $\gamma_s=0.9$ and $\gamma_p=0.1$ were used for the local kernel definition for solvent and protein (reproduced with permission from [48]). (b) three-time scale XM-SIN(R)+iEL/0-SCF with force splitting using $\Delta t_s = 0.5$, $\Delta t_m = 3.0$ fs, and $\Delta t = 90.0$ fs with $L = 1$ thermostat variables along with several variations γ values.

However, from a much more detailed analysis we find that the 90 fs outer time step does degrade the accuracy of the polarization properties, shown for the glycine peptide system in Figure 6 and Figure S7, and $\langle U \rangle$ is in error in excess of 10% (Table 1). We note that previous work using SIN(R) for small solvated peptides⁵⁵ or large gas phase biomolecules⁵³, showed little changed in accuracy over a wide range in time steps compared to standard integration methods when using

simpler fixed charge force fields. This discrepancy with smaller outer timesteps is not due to iEL/0-SCF but is due to the complexity of the AMOEBA model, in particular a combination of higher order poles, geometric flexibility, and intramolecular polarization, that reduces the allowed time steps due to rapid variations in force at the short-ranged; these more complex features are not part of the simpler fixed charge force fields.

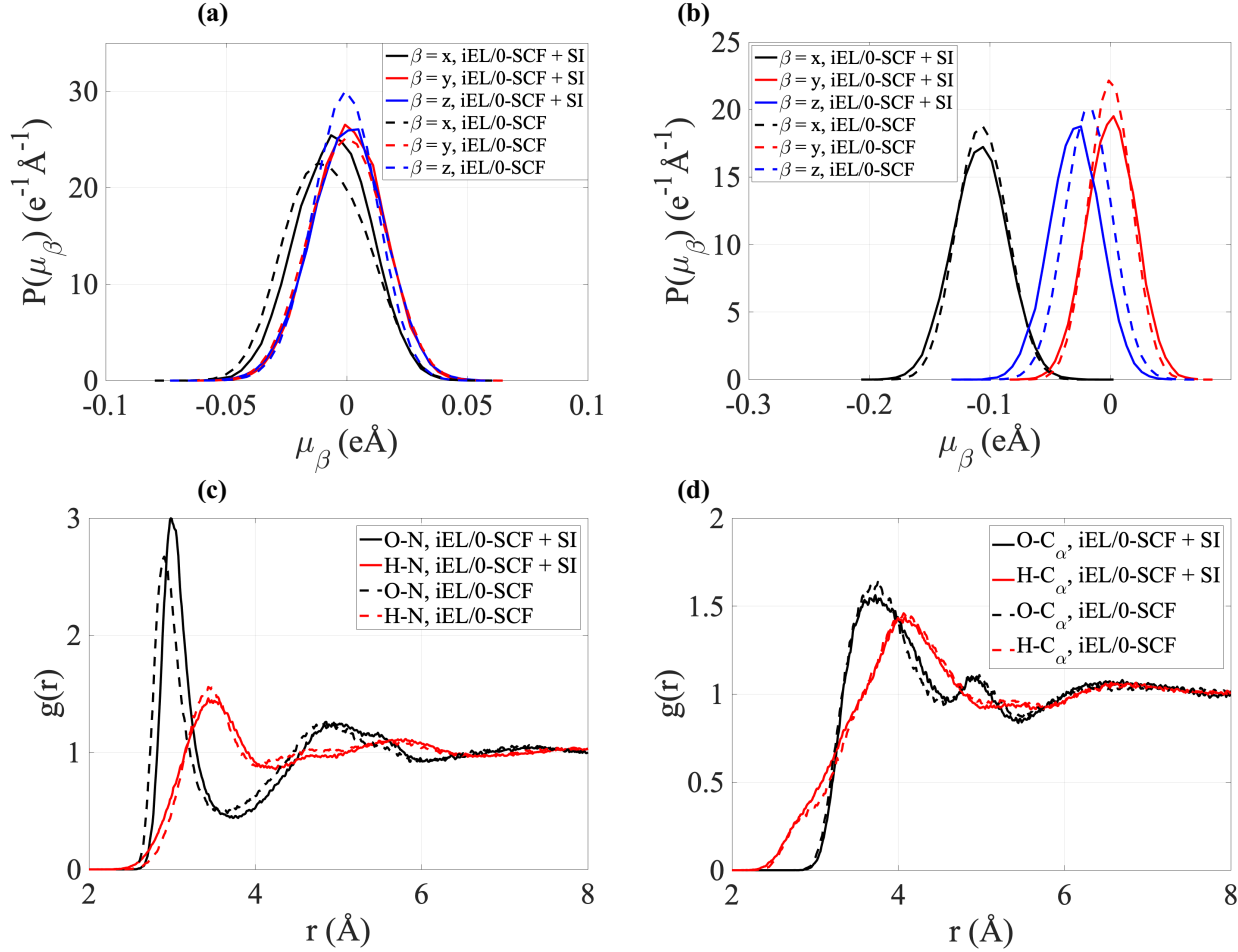


Figure 6: 3-step XM-SIN(R)+ iEL/0-SCF with $L = 1$ using force splitting relative to standard iEL/0-SCF for glycine in water. All bonded forces use $\Delta t_s = 0.5$ fs and non-bonded forces split into short-ranged and long-ranged evaluated at $\Delta t_m = 3.0$ fs $\Delta t = 90.0$ fs, respectively. Probability distributions for the induced dipole vector components of the backbone for (a) N and (b) $C\alpha$. Radial distribution functions for water interacting with for (c) N and (d) $C\alpha$. Data was collected from 5.0 ns simulations.

To illustrate the consequences, we have previously shown that with our standard iEL/0-SCF method we can take 6-7 fs timesteps using a simple Drude model for water, PSpC, which is rigid and uses point charges and Drude particle on heavy atoms only⁴⁸, whereas we could not go beyond ~ 1 -2 fs timesteps for the corresponding evaluation with the AMOEBA water model. Like the iEL/0-SCF

method, the SIN(R) outer time step will also be dependent on the underlying force field. We find that $\langle U \rangle$ values improve substantially for the peptide system if we reduce the outer time step to 18 fs, as seen in Table 1 and for properties as reported in Fig. S8; changing the intermediate time step to $\Delta t_m = 5$ fs and outer time step to $\Delta t = 10$ fs also yields good results. Although both variations reduce the speed-up by half (see Fig. 4b) for bonded systems such as peptides and proteins in the present Tinker 8.1 implementation, it is nonetheless a computational win for the combined iEL/0-SCF and SIN(R) methods.

Since the size of the intermediate time step ultimately controls the computational gains, we return to the simpler non-bonded force decomposition that uses no force splitting, but this time, we employ the XM-SIN(R) with $L = 1$ for the water, 4.66 M MgCl_2 , and the aqueous glycine systems. Figure 7a shows how $\langle \mu_{\text{mol}} \rangle$ for water changes with Δt under the 2-step XM-SIN(R) scheme. Very acceptable levels of accuracy in properties is evident at $\Delta t = 6$ fs for water (Figs. S9) and the concentrated salt solution (Fig. S10), while accuracy degrades noticeably using the larger 8 fs time step, evident from the $\langle U \rangle$ values reported in Table 1. Although the 3-step XM-SIN(R) integration with force splitting provides a faster simulation for water (Fig. 4b), the simpler 2-step XM-SIN(R) integration still yields speed-ups of close to a factor of ~ 4 relative to iEL/0-SCF, or a factor of ~ 6 relative to a standard SCF solver (Figure 7b) in Tinker 8.1.

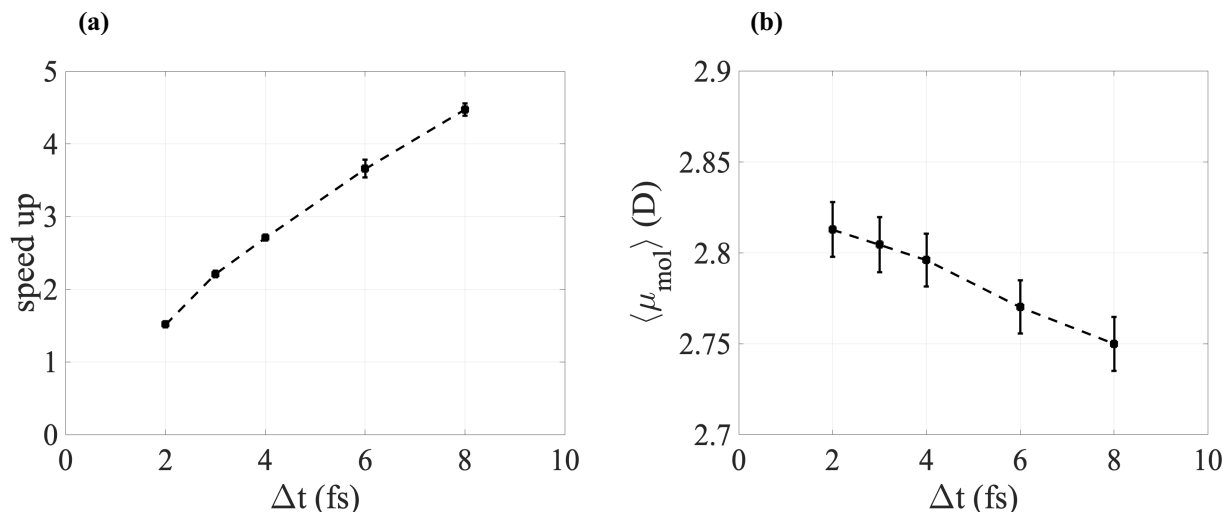


Figure 7: Accuracy and improvements in computational costs with a 2 step XM-SIN(R) with $L = 1$ using no force splitting and iEL/0-SCF relative to standard iEL/0-SCF for pure water. (a) $\langle \mu_{\text{mol}} \rangle$ and (b) speed-ups, as a function of the outer time step. Timings used 16 cores and 10,000 water molecule simulations. All data referenced to an iEL/0-SCF single time scale integration at 1.0 fs.

The primary benefit of the simpler 2-step XM-SIN(R) is that it is more accurate than the 3-step XM-SIN(R) for the glycine peptide (comparing Figs. 6 and 8 and Figs. S7 and S11) and hence for general protein systems. Furthermore the 2-step XM-SIN(R) with the $L = 1$ algorithm at $\Delta t = 6$ fs has computational efficiency that is nearly as good as the 3-step XM-SIN(R) with force splitting using $\Delta t_m = 3.0$ fs and $\Delta t = 18.0$ fs, with a speed-up relative to a standard SCF solution of 6X in the present Tinker 8.1 implementation.

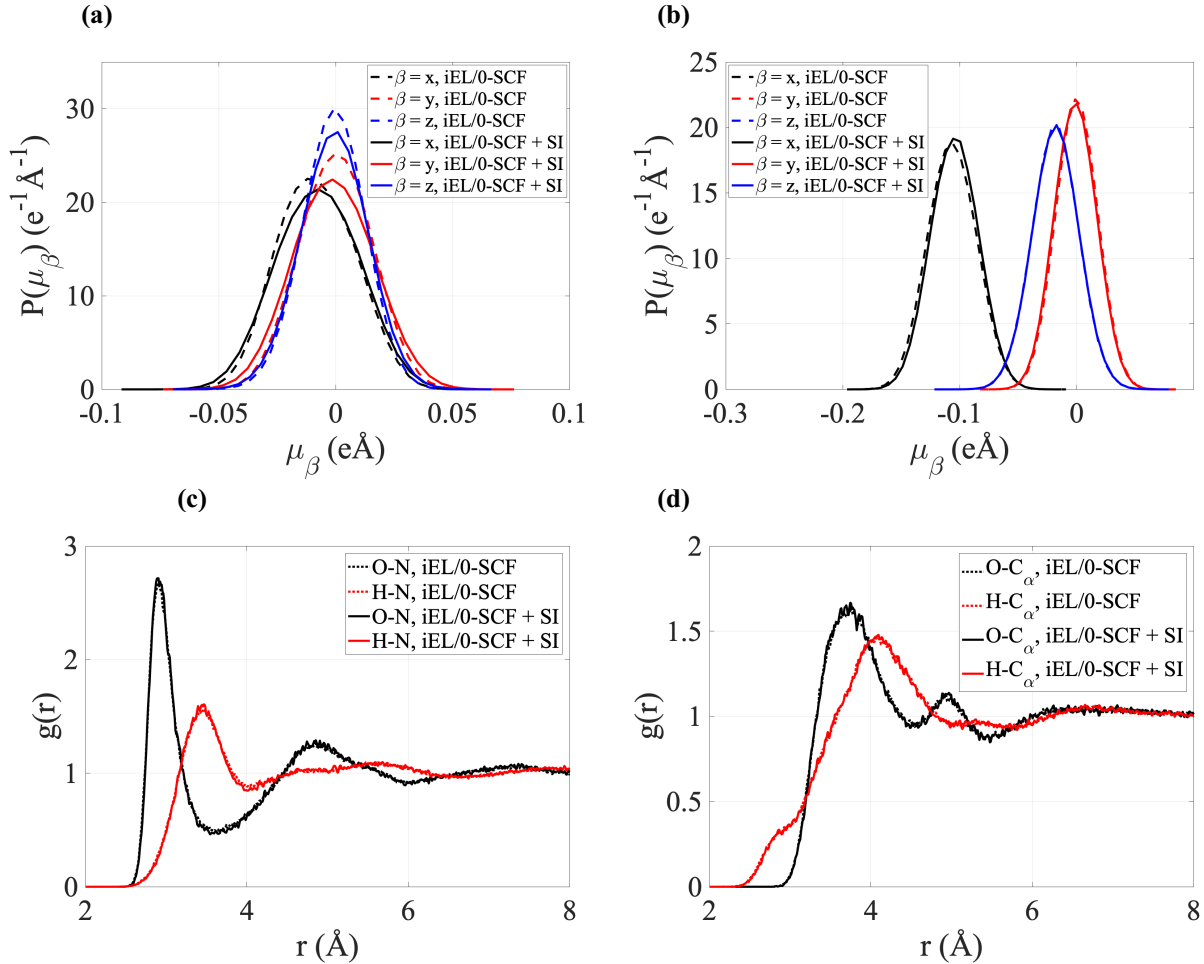


Figure 8: 2-step XM-SIN(R) and iEL/0-SCF with $L = 1$ relative to standard iEL/0-SCF for glycine in water. All bonded forces are evaluated at $\Delta t_s = 0.5$ fs, and all non-bonded forces are evaluated at $\Delta t = 6.0$ fs. Probability distributions for the induced dipole vector components of the backbone (a) N and (b) C α . Radial distribution functions for water interacting with backbone (c) N and (d) C α . Data was collected from 5.0 ns simulations.

CONCLUSION

In this work we have combined an SCF-less solution to classical polarization⁴⁸⁻⁴⁹ with an integration scheme that eliminates the resonances that inhibit long integration time steps^{47, 53-55} to increase

computational efficiency for condensed phase systems including water, salt solutions, and more complicated bonded systems such as peptides and proteins. We achieve excellent accuracy when iEL/0-SCF is combined with the XM-SIN(R) stochastic isokinetic integration algorithm, that classifies forces into bonded (evaluated at 0.5 fs) and nonbonded components (evaluated at 6.0 fs) with $L = 1$ thermostat variables, with very good computational speed-ups $\sim 6X$ compared to a conventional AMOEBA calculation in Tinker 8.1 for all systems investigated here, including water, 4.66M $MgCl_2$ solution, a zwitterionic glycine peptide, and DHFR protein in water. Greater computational efficiencies for water and the concentrated salt solution can be found using the same XM-SIN(R) approach, by further splitting the non-bonded forces into a short-ranged component evaluated at an intermediate time scale of 3-5 fs, and a long-ranged and more smoothly varying force evaluated at ~ 90 -100 fs, to give speed-ups close to a factor of $\sim 8X$. In both cases, better optimized code could improve the performance over what could be achieved in the comparisons done here using the TINKER 8.1 software release⁵². We will reserve such a comparison for future work. In summary, the iEL/0-SCF and XM-SIN(R) combination can achieve both accuracy and meaningful speed-ups that represent a significant advance for equilibrium simulations in the canonical ensemble for many-body potential energy surfaces.

ACKNOWLEDGMENTS. AA and THG thank the National Science Foundation Grant No. CHE-1665315 for support of this work. M.T. acknowledges support from the National Science Foundation Grant No. CHE-1565980. This research used computational resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

REFERENCES

1. Lifson, S.; Warshel, A. Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkane and n-Alkane Molecules. *J. Chem. Phys.* **1968**, *49*, 5116-5129.
2. Remsing, R. C.; Baer, M. D.; Schenter, G. K.; Mundy, C. J.; Weeks, J. D. The Role of Broken Symmetry in Solvation of a Spherical Cavity in Classical and Quantum Water Models. *J. Phys. Chem. Lett.* **2014**, *5* (16), 2767-2774.
3. Jungwirth, P.; Winter, B. Ions at aqueous interfaces: From water surface to hydrated proteins. *Ann. Rev. Phys. Chem.* **2008**, *59*, 343-366.
4. Vazdar, M.; Pluharova, E.; Mason, P. E.; Vacha, R.; Jungwirth, P. Ions at Hydrophobic Aqueous Interfaces: Molecular Dynamics with Effective Polarization. *J. Phys. Chem. Lett.* **2012**, *3* (15), 2087-2091.

5. Fried, S. D.; Wang, L. P.; Boxer, S. G.; Ren, P.; Pande, V. S. Calculations of the electric fields in liquid solutions. *J. Phys. Chem. B* **2013**, *117* (50), 16236-48.
6. Thompson, M. A.; Schenter, G. K. Excited-States of the Bacteriochlorophyll-B Dimer of Rhodospseudomonas-Viridis - a Qm/Mm Study of the Photosynthetic Reaction-Center That Includes Mm Polarization. *J. Phys. Chem.* **1995**, *99* (17), 6374-6386.
7. Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. B. Affinities of amino-acid side chains for water. *Biochem.* **1981**, *20*, 849-855.
8. Shirts, M. R.; Pitner, J. W.; Swope, W. C.; Pande, V. S. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.* **2003**, *119*, 5740-5761.
9. Abrams, J. B.; Rosso, L.; Tuckerman, M. E. Efficient and precise solvation free energies via alchemical adiabatic molecular dynamics. *J. Chem. Phys.* **2006**, *125*, 074115.
10. Johnson, M.; Malardier-Jugroot, C.; Murarka, R.; Head-Gordon, T. Hydration water dynamics near biological interfaces. *J. Phys. Chem. B* **2009**, *113*, 4082-4092.
11. Johnson, M. E.; Malardier-Jugroot, C.; Head-Gordon, T. Effects of co-solvents on peptide hydration water structure and dynamics. *Phys. Chem. Chem. Phys.* **2010**, *12*, 393-405.
12. Bhowmick, A.; Brookes, D. H.; Yost, S. R.; Dyson, H. J.; Forman-Kay, J. D.; Gunter, D.; Head-Gordon, M.; Hura, G. L.; Pande, V. S.; Wemmer, D. E.; Wright, P. E.; Head-Gordon, T. Finding Our Way in the Dark Proteome. *J. Amer. Chem. Soc.* **2016**, *138* (31), 9730-9742.
13. Demerdash, O.; Wang, L.-P.; Head-Gordon, T. Advanced models for water simulations. *WIREs: Comput. Mol. Sci.* **2017**, *8* (1), e1355.
14. Warshel, A.; Kuwajima, S. Incorporating Electric Polarizabilities in Water-Water Interaction Potentials. *J. Phys. Chem.* **1990**, *94*, 460-466.
15. Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. An effective fragment method for modeling solvent effects in quantum mechanical calculations. *J. Chem. Phys.* **1996**, *105* (5), 1968-1986.
16. Gao, J. L. Toward a molecular orbital derived empirical potential for liquid simulations. *J. Phys. Chem. B* **1997**, *101* (4), 657-663.
17. Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R. H.; Berne, B. J.; Friesner, R. A. Fluctuating charge, polarizable dipole, and combined models: Parameterization from ab initio quantum chemistry. *J. Phys. Chem. B* **1999**, *103* (22), 4730-4737.
18. Cieplak, P.; Caldwell, J.; Kollman, P. Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: Aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *J. Comp. Chem.* **2001**, *22* (10), 1048-1057.
19. Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X. X.; Murphy, R. B.; Zhou, R. H.; Halgren, T. A. Development of a polarizable force field for proteins via ab initio quantum chemistry: First generation model and gas phase tests. *J. Comput. Chem.* **2002**, *23* (16), 1515-1531.
20. Lamoureux, G.; MacKerell, A. D.; Roux, B. A simple polarizable model of water based on classical Drude oscillators. *J. Chem. Phys.* **2003**, *119* (10), 5185-5197.
21. Piquemal, J. P.; Williams-Hubbard, B.; Fey, N.; Deeth, R. J.; Gresh, N.; Giessner-Prettre, C. Inclusion of the ligand field contribution in a polarizable molecular mechanics: SIBFA-LF. *J. Comput. Chem.* **2003**, *24* (16), 1963-1970.

22. Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. Development of an accurate and robust polarizable molecular mechanics force field from ab initio quantum chemistry. *J. Phys. Chem. A* **2004**, *108*, 621-627.
23. Patel, S.; Brooks, C. L. CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J. Comp. Chem.* **2004**, *25* (1), 1-15.
24. Donchev, A. G.; Ozrin, V. D.; Subbotin, M. V.; Tarasov, O. V.; Tarasov, V. I. A quantum mechanical polarizable force field for biomolecular interactions. *Proc. Natl. Acad. Sci. USA* **2005**, *102* (22), 7829-7834.
25. Harder, E.; Kim, B. C.; Friesner, R. A.; Berne, B. J. Efficient simulation method for polarizable protein force fields: application to the simulation of BPTI in liquid. *J. Chem. Theo. Comp.* **2005**, *1*, 169-180.
26. Wang, Z. X.; Zhang, W.; Wu, C.; Lei, H. X.; Cieplak, P.; Duan, Y. Strike a balance: Optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides. *J. Comput. Chem.* **2006**, *27* (6), 781-790.
27. Fanourgakis, G. S.; Xantheas, S. S. The Bend Angle of Water in Iced Ih and Liquid Water: The Significance of Implementing the Nonlinear Monomer Dipole Moment Surface in Classical Interaction Potentials. *J. Chem. Phys.* **2006**, *124*, 174504.
28. Benjamin, K. M.; Schultz, A. J.; Kofke, D. A. Virial Coefficients of Polarizable Water Applications to Thermodynamic Properties and Molecular Clustering. *J. Phys. Chem. C* **2007**, *111*, 16021-16027.
29. Rasmussen, T. D.; Ren, P. Y.; Ponder, J. W.; Jensen, F. Force field modeling of conformational energies: Importance of multipole moments and intramolecular polarization. *Int. J. Quant. Chem.* **2007**, *107* (6), 1390-1395.
30. Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J. P. Anisotropic, polarizable molecular mechanics studies of inter- and intramolecular interactions and ligand-macromolecule complexes. A bottom-up strategy. *J. Chem. Theo. Comp.* **2007**, *3* (6), 1960-1986.
31. Geerke, D. P.; van Gunsteren, W. F. Calculation of the free energy of polarization: Quantifying the effect of explicitly treating electronic polarization on the transferability of force-field parameters. *J. Phys. Chem. B* **2007**, *11*, 6425-6436.
32. Jorgensen, W. L.; Jensen, K. P.; Alexandrova, A. N. Polarization effects for hydrogen-bonded complexes of substituted phenols with water and chloride ion. *J. Chem. Theo. Comp.* **2007**, *3* (6), 1987-1992.
33. Salanne, M.; Vuilleumier, R.; Madden, P. A.; Simon, C.; Turq, P.; Guillot, B. Polarizabilities of Individual Molecules and Ions in Liquids from First Principles. *J Phys-Condens Mat* **2008**, *20*, 494207.
34. Cieplak, P.; Dupradeau, F. Y.; Duan, Y.; Wang, J. M. Polarization effects in molecular mechanical force fields. *J Phys-Condens Mat* **2009**, *21* (33), 333102.
35. Patel, S.; Davis, J. E.; Bauer, B. A. Exploring Ion Permeation Energetics in Gramicidin A Using Polarizable Charge Equilibration Force Fields. *J. Amer. Chem. Soc.* **2009**, *131* (39), 13890-13891.
36. Tafipolsky, M.; Engels, B. Accurate Intermolecular Potentials with Physically Grounded Electrostatics. *J. Chem. Theo. Comput.* **2011**, *7*, 1791-1803.
37. Cisneros, G. A. Application of Gaussian Electrostatic Model (GEM) Distributed Multipoles in the AMOEBA Force Field. *J. Chem. Theo. Comp.* **2012**, *8* (12), 5072-5080.

38. Wang, L. P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. Systematic improvement of a classical molecular model of water. *J. Phys. Chem. B* **2013**, *117* (34), 9956-9972.
39. Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D. Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator. *J. Chem. Theo. Comput.* **2013**, *9* (12), 5430-5449.
40. Demerdash, O.; Yap, E. H.; Head-Gordon, T. Advanced Potential Energy Surfaces for Condensed Phase Simulation. *Ann. Rev. Phys. Chem.* **2014**, *65*, 149-174.
41. Shi, Y.; Ren, P.; Schnieders, M.; Piquemal, J. P. Polarizable Force Fields for Biomolecular Modeling. *Rev. Comp. Chem.* **2015**, *28*, 51-86.
42. Arismendi-Arrieta, D. J.; Riera, M.; Bajaj, P.; Prosimiti, R.; Paesani, F. i-TTM Model for Ab Initio-Based Ion–Water Interaction Potentials. 1. Halide–Water Potential Energy Functions. *J. Phys. Chem. B* **2016**, *120* (8), 1822-1832.
43. Kohagen, M.; Lepsik, M.; Jungwirth, P. Calcium Binding to Calmodulin by Molecular Dynamics with Effective Polarization. *J. Phys. Chem. Lett.* **2014**, *5* (22), 3964-3969.
44. Kohagen, M.; Mason, P. E.; Jungwirth, P. Accounting for Electronic Polarization Effects in Aqueous Sodium Chloride via Molecular Dynamics Aided by Neutron Scattering. *J. Phys. Chem. B* **2016**, *120* (8), 1454-1460.
45. Baker, C. M. Polarizable force fields for molecular dynamics simulations of biomolecules. *WIREs Comput. Mol. Sci.* **2015**, *5* (2), 241-254.
46. Albaugh, A.; Demerdash, O.; Head-Gordon, T. An efficient and stable hybrid extended Lagrangian/self-consistent field scheme for solving classical mutual induction. *J. Chem. Phys.* **2015**, *143* (17), 174104.
47. Margul, D. T.; Tuckerman, M. E. A Stochastic, Resonance-Free Multiple Time-Step Algorithm for Polarizable Models That Permits Very Large Time Steps. *J. Chem. Theo. Comput.* **2016**, *12*, 2170-2180.
48. Albaugh, A.; Head-Gordon, T. A New Method for Treating Drude Polarization in Classical Molecular Simulation. *J. Chem. Theo. Comp.* **2017**, *13* (11), 5207-5216.
49. Albaugh, A.; Head-Gordon, T.; Niklasson, A. M. N. Higher-Order Extended Lagrangian Born–Oppenheimer Molecular Dynamics for Classical Polarizable Models. *J. Chem. Theo. Comp.* **2018**, *14* (2), 499-511.
50. Albaugh, A.; Boateng, H. A.; Bradshaw, R. T.; Demerdash, O.; Dziedzic, J.; Mao, Y.; Margul, D. T.; Swails, J.; Zeng, Q.; Case, D. A.; Eastman, P.; Essex, J. W.; Head-Gordon, M.; Pande, V. S.; Ponder, J. W.; Shao, Y.; Skylaris, C.-K.; Todorov, I. T.; Tuckerman, M. E.; Head-Gordon, T. Advanced Potential Energy Surfaces for Molecular Simulation. *J. Phys. Chem. B (Feature article)* **2017**, *13* (11), 5207-5216.
51. Vitale, V.; Dziedzic, J.; Albaugh, A.; Niklasson, A.; Head-Gordon, T.; Skylaris, C.-K. An evaluation of extended Lagrangian schemes for molecular dynamics simulations with polarisable force fields and DFT. *J. Chem. Phys.* **2017**, *146* (12), 124115.
52. Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. Tinker 8: Software Tools for Molecular Design. *J. Chem. Theo. Comput.* **2018**, *14* (10), 5273–5289.
53. Minary, P.; Tuckerman, M. E.; Martyna, G. J. Long Time Molecular Dynamics for Enhanced Conformational Sampling in Biomolecular Systems. *Phys. Rev. Lett.* **2004**, *93*, 150201.
54. Leimkuhler, B.; Margul, D. T.; Tuckerman, M. E. Stochastic, Resonance-Free Multiple Time-Step Algorithm for Molecular Dynamics with Very Large Time Steps. *Mol. Phys.* **2013**, *111*, 3579.

55. Chen, P. Y.; Tuckerman, M. E. Molecular dynamics based enhanced sampling of collective variables with very large time steps. *J. Chem. Phys.* **2018**, *148*, 024106.
56. Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible Multiple Time Scale Molecular Dynamics. *J. Chem. Phys.* **1992**, *97*, 1990-2001.
57. Niklasson, A. M.; Cawkwell, M. J. Generalized extended Lagrangian Born-Oppenheimer molecular dynamics. *J. Chem. Phys.* **2014**, *141* (16), 164123.
58. Leimkuhler, B.; Noorizadeh, F.; Theil, F. A gentle stochastic thermostat for molecular dynamics. *J. Stat. Phys.* **2009**, *135*, 261-277.
59. Toukmaji, A.; Sagui, C.; Board, J.; Darden, T. Efficient particle-mesh Ewald based approach to fixed and induced dipolar interactions. *J. Chem. Phys.* **2000**, *113* (24), 10913-10927.
60. Laury, M. L.; Wang, L. P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W. Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *J. Phys. Chem. B* **2015**, *119*, 9423.
61. Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. The Polarizable Atomic Multipole-Based Amoeba Force Field for Proteins. *J. Chem. Theo. Comput.* **2013**, *9* (9), 4046-4063.
62. Esser, A.; Belsare, S.; Marx, D.; Head-Gordon, T. Mode specific THz spectra of solvated amino acids using the AMOEBA polarizable force field. *Phys. Chem. Chem. Phys.* **2017**, *19* (7), 5579-5590.
63. Yoshida, H. Construction of higher order symplectic integrators. *Phys. Lett. A* **1990**, *150* (5-7), 262-268.
64. Suzuki, M. General theory of fractal path integrals with applications to many-body theories and statistical physics. *J. Math. Phys.* **1991**, *32* (2), 400-407.
65. Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *J. Chem. Phys.* **1982**, *76*, 637.
66. Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. Explicit Reversible Integrators for Extended Systems Dynamics. *Mol. Phys.* **1996**, *87*, 1117-1157.

