Noname manuscript No.
(will be inserted by the editor)

Ewald–based methods for Gaussian Integral Evaluation: Application to a New Parametrization of GEM*

Robert E. Duke · G. Andrés Cisneros

Received: date / Accepted: date

Abstract The development of accurate potentials for computational simulations has been an active area of research. Our group has been involved in the development of the Gaussian electrostatic model (GEM), a force field based on molecular densities. The philosophy of GEM is based on the pioneering work of N. Gresh and co–workers of the reproduction of individual inter–molecular interaction components obtained from quantum mechanical (QM) energy decomposition analysis (EDA). The molecular densities used in GEM are represented by fitting accurate QM molecular densities using auxiliary basis sets (comprised of Hermite Gaussians). The use of these molecular densities results in the need to evaluate a large number of Gaussian integrals. We have previously shown that the particle–mesh Ewald (PME), and fast Fourier Poison (FFP) methods can be used for efficiently evaluating these types of integrals. Here, we present the latest parametrization of GEM* and its application for an extensive study of PME and FFP for molecular dynamics (MD) simulations using a hybrid version of our potential, GEM*. The temperature dependence of various bulk properties is presented and discussed, as well as the effect of various parameters affecting the performance/accuracy of both methods.

Keywords Gaussian Electrostatic Model · Particle mesh Ewald · Fast Fourier Poison · Polarizable Force Field

Robert E. Duke
Department of Chemistry, University of North Texas, Denton, TX 76202

G. Andrés Cisneros
Department of Chemistry, University of North Texas, Denton, TX 76202
E-mail: andres@unt.edu

1 Introduction

Force field (FF) development continues to receive significant attention. [1–6]. The need for extensive phase–space sampling results in the need to make various approximations to increase computational efficiency. In particular, many FF separate the calculation of the energy in bonded and non–bonded terms. The non–bonded interactions are usually approximated by Coulomb and Van der Waals (VdW) terms using atom–centered point charges and atom–specific parameters for VdW. The use of these functions provides improved sampling efficiency. However, one problem with this approximation is the result of a loss in accuracy because of the loss of charge density anisotropy and the failure to account for penetration effects [7].

Some of these issues can be ameliorated by including a better description of the charge density distribution and/or a more detailed description of the intermolecular interactions. For example, force fields like AMOEBA, SIBFA, EFP, X–Pol, and NEMO employ distributed multipoles and use explicit polarization [8–14]. The use of distributed multipoles results in an improved treatment of electrostatics [7, 15–19]. However, multipoles still suffer from charge penetration errors at close range [7, 20], although this shortcoming may be ameliorated by the use of damping functions to correct the electrostatic interactions at close distances [20–26].

The explicit description of the molecular charge density avoids the charge penetration issue. This feature has been exploited by several methods [27–34]. This is a key feature of the Gaussian Electrostatic Model (GEM) [35–39]. GEM uses the density fitting (DF) formalism [40–42] to expand the molecular density using Hermite Gaussian auxiliary basis sets (ABSs). Additionally, following the philosophy promoted by Gresh and co–workers, GEM relies on the use of *ab initio* QM intermolecular interaction results from energy decomposition analysis (EDA) [43–52].

The GEM densities are fitted to linear combinations of Hermite Gaussians and used to calculate each of the components of the QM intermolecular interaction (Coulomb, exchange–repulsion, polarization, charge–transfer and dispersion) [36, 37, 39]. The use of continuous functions provides a more accurate description of molecular properties compared to conventional point charges [37]. We have also shown that GEM results in very accurate energies and forces for a range of systems including homo–dimers, hetero–dimers and molecular clusters, and it can be used in multi–scale implementations. [36, 37, 53–56].

The use of Hermite Gaussians results in the need to evaluate a large number of Coulomb and overlap integrals. Various reciprocal space methods have been developed for the efficient evaluation of integrals [57–61]. We have previously shown that Ewald–based methods can be extended for Gaussian distributions for the efficient evaluation of these integrals [37]. In particular, we have extended the smooth particle mesh Ewald (PME) [62] and fast Fourier Poisson (FFP) [63] methods to allow the evaluation of continuous Gaussian functions in periodic boundary conditions (PBC) [37, 64].

We previously reported a water potential that included the Coulomb and exchange–repulsion terms from GEM, combined with the polarization and a modified Halgren function from AMOEBA, this hybrid water potential was termed GEM* [65]. This potential employed a molecular electronic density fitted at the MP2/aug–cc–pVTZ level of theory for the water monomer, and was parametrized to reproduce BSSE–corrected MP2/aug-cc-pVTZ intermolecular interaction energies, as well as individual EDA terms calculated at the RVS (HF/aug–cc–pVTZ) level for dimers and selected oligomers.

In this contribution we present an update for GEM* by fitting the molecular electronic density and parametrizing the individual terms to higher levels of theory (vide infra). This new paremetrization is employed to test the latest implementation of the PME and FFP algorithms in pmemd.gem as released in the AMBER18 suite of programs. Thermodynamic, structural and dynamical properties for this new GEM* parametrization are presented and discussed. The remainder of the paper is organized as follows, Section 2 presents the theory and details of the computational simulations, followed by presentation of the performance of the reciprocal space algorithms using various parameters and bulk property calculations in Section 3, followed by concluding remarks.

2 Computational Methods

This section presents the details of the GEM* force field in Subsection 2.1, followed in Subsection 2.2 by a description of the mathematical details for the extension of PME and FFP for Gaussian distributions in PBC. Finally, Subsection 2.3 provides a description of the new parametrization and MD simulations.

2.1 The GEM* force field

The initial implementation of the full GEM potential involved the use of spherical type Hermites only, resulting in what was termed GEM–0 [36]. This version, along with the extension to Hermite Gaussians of arbitrary angular momentum were shown to be highly accurate for the calculation of single energies and forces [37]. The only drawback with these potentials was that MD simulations were not possible in an efficient manner since the analytical form of one of the terms (charge–transfer) was not available. Therefore, a compromise to enable efficient MD simulations was realized by combining terms from GEM with terms from AMOEBA, which resulted in the development of the GEM* force field [65].

In brief, GEM* combines the Coulomb and exchange–repulsion terms from GEM with the polarization, van der Waals (modified) and bonded terms from AMOEBA. The functional form for GEM* is thus:

$$\begin{aligned} E_{Total}^{GEM*} = & E_{Coulomb}^{GEM} + E_{exch-rep}^{GEM} + E_{polarization}^{AMOEBA} \\ & + E_{Halgren}^{modified} + E_{bonded}^{AMOEBA}. \end{aligned} \tag{1}$$

The Coulomb and exchange–repulsion terms for GEM* are obtained by evaluating the two center Coulomb and overlap integrals involving the fitted densities [65]. Since GEM* includes an explicit term for exchange, it was necessary to modify the original van der Waals function implemented in AMOEBA. In this case, we have modified the buffered Halgren function (modHalgren) by removing the repulsive term as follows:

$$E_{modHalgren} = -\epsilon_{ij}\left[\frac{1.07R^*_{ij}}{(R_{ij}+0.07R^*_{ij})}\right]^7. \quad (2)$$

The polarization and bonded terms are the same as those in the original AMOEBA implementation [66]. The only difference in the polarization is that the permanent electric fields for the calculation of the induced dipoles are calculated with the distributed multipoles obtained from the fitted Hermites for consistency between the Coulomb and polarization terms [67].

The initial implementation of GEM* was tested by fitting parameters for a water potential [**?**]. These parameters were compared to reference *ab initio* values for total intermolecular interactions corrected for basis set superposition error via the counterpoise correction . The reference data was calculated at the MP2(full)/aug–cc–pVTZ level to match the original AMOEBA parametrization [66,68,10,69]. The molecular density used to obtain the fitting coefficients for GEM* was calculated at the same level of theory as above for a water molecule at the AMOEBA equilibrium geometry.

Three parametrizations were investigated, termed models 1–3. The difference among the three models involves the use of different ABSs, A1 or A2 [70,71], and/or the dataset of water oligomers used for the parametrization. Model 1 was fitted using the A2 ABS to reproduce intermolecular interaction energies for the canonical water dimer, several random dimers, and selected water clusters from [72]. Models 2 and 3 were parametrized to reproduce intermolecular energies for the canonical water dimer only using the A2 (model 2) and A1 (model 3) ABSs. All calculations for GEM* were performed with a modified pmemd.amoeba version in the AMBER16 suite of programs [1].

2.2 Ewald–based methods for Evaluation of Coulomb Integrals

This subsection describes the extension of the Ewald, PME, and FFP formalisms for Gaussian distributions. To begin, let U denote a unit cell whose edges are given by the vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$. An idealized infinite crystal can be generated by periodic translations $\mathbf{n} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$ for all integer triples (n_1, n_2, n_3), with n_1, n_2, n_3 not all zero. Now, consider a collection of N normalized spherical Gaussian charge densities $\rho_1 \ldots \rho_N$ centered at $\{\mathbf{R}_1 \ldots \mathbf{R}_N\} \in U$ with exponents α_i. Initially, we will consider only spherical Gaussian functions, as was the case with GEM–0 [36]: $\rho_i(\mathbf{r}) =$

$q_i(\alpha_i/\pi)^{3/2}exp(-\alpha_i(\mathbf{r}-\mathbf{R}_i)^2)$, and let $q_1+\ldots+q_N=0$. Note that the Gaussian densities N do not need to be centered exclusively on atomic positions. A case in point is that GEM–0 includes sites on the bisector line between the hydrogen atoms and at the lone pairs of the oxygen atom [36]. The Coulomb energy of the central unit cell within a large spherical crystal, due to the interactions of the Gaussian charge distributions with each other and all periodic images within the crystal can be calculated using Ewald methods.

One key obstacle to overcome is the fact that it is necessary to grid the Gaussian densities to determine the reciprocal part in the Ewald sum. However, this can become intractable for Gaussian functions with large exponents (compact Gaussians). In our initial implementations, the charge densities were classified as compact or diffuse Hermites based on a single Ewald exponent β. Thus, if the exponent of a given Hermite was above the cutoff it was considered compact, and diffuse ($\alpha_i < \beta$) otherwise. With this approach, the contributions involving diffuse Hermites can be calculated in reciprocal space exclusively [37].

This was later improved by realizing that the Ewald exponent, β can be different for each Hermite pair ij [64]. In this way, β can be chosen to be infinite for Hermite pairs where at least one of the Gaussians is diffuse. Thus, all pairs that involve diffuse Hermites can be efficiently evaluated in reciprocal space. That is, given $\theta > 0$ a Gaussian distribution $q_i\rho_i$ is classified as compact ($i \in c$) if $\alpha_i \geq 2\theta$ and diffuse ($i \in d$) otherwise. Subsequently, for $i,j \in \{c\}$, a β is selected such that $1/\theta = 1/\alpha_i + 1/\alpha_j + 1/\beta$, otherwise $\beta = \infty$. In the case of GEM, the fitted densities are expanded in a linear combination of Hermite Gaussians $\Lambda_{tuv}(\mathbf{r},\alpha,\mathbf{R})$. Thus, the charge distribution is given by $\rho_i(\mathbf{r},\alpha,\mathbf{R}) = \sum_{l=1}^{L}\sum_{tuv}\mathbf{c}_{i,l,tuv}\Lambda_{tuv}(\mathbf{r},\alpha_l,\mathbf{R_i})$, where $\mathbf{c}_{i,l,tuv}$ are the Hermite coefficients and L denotes the different ABS exponents on center i. With this, the Ewald expression becomes

$$
\begin{aligned}
E(\rho^{\{N\}}) = \frac{1}{2}\sum_{\mathbf{n}}{}' \sum_{(i,j)\in c\times c} q_iq_j \left\{ \frac{erfc(\theta^{1/2}|\mathbf{R}_{ij}-\mathbf{n}|) - erfc(\mu_{ij}^{1/2}|\mathbf{R}_{ij}-\mathbf{n}|)}{|\mathbf{R}_{ij}-\mathbf{n}|} \right\} \\
+\frac{1}{2\pi V}\sum_{\mathbf{m}\neq 0}\sum_{(i,j)\in c\times c}^{N} q_iq_j \frac{exp(-\pi^2\mathbf{m}^2/\theta)}{\mathbf{m}^2} exp(-2\pi i\mathbf{m}\cdot\mathbf{R}_{ij}) \\
+\frac{1}{2\pi V}\sum_{\mathbf{m}\neq 0}\sum_{(i,j)\notin c\times c}^{N} q_iq_j \frac{exp(-\pi^2\mathbf{m}^2/\mu_{ij})}{\mathbf{m}^2} exp(-2\pi i\mathbf{m}\cdot\mathbf{R}_{ij}) \qquad (3) \\
-\frac{\pi}{2V}\sum_{i,j=1}^{N} q_iq_j\left(\frac{1}{\theta}+\frac{1}{\alpha_i}+\frac{1}{\alpha_j}\right) - \sum_{i=1}^{N} q_i^2\left(\frac{\theta}{\pi}\right)^{1/2} \\
-\sum_{i=1}^{N} q_i^2\left(\frac{\alpha_i}{\pi}\right)^{1/2} + \frac{2\pi\mathbf{D}^2}{3V} + \varepsilon(K),
\end{aligned}
$$

where the first term corresponds to the direct part of the Ewald sum, the second and third terms to the reciprocal part, $\mathbf{R}_{ij} = \mathbf{R}_i - \mathbf{R}_j$, the term involving the unit cell dipole $\mathbf{D} = q_1\mathbf{R}_1 + \ldots + q_N\mathbf{R}_N$ is the surface term, $\varepsilon(K)$ denotes a quantity that converges to 0 as $K \longrightarrow \infty$, $\mathbf{m}$ denotes the reciprocal lattice vectors, and $1/\mu_{ij} = 1/\alpha_i + 1/\alpha_j$. Equation 3 describes the special case where the Gaussian distribution is composed exclusively by spherical ($l = 0$) Gaussians, as is the case with the GEM–0 potential [36].

This expression can be further generalized for the case when the auxiliary and optimized bases used for the fitting include Gaussians with $l > 0$ [37]. In that case, the Gaussian charge distribution is given by $\rho_i(\mathbf{r}, \mathbf{R}_i, \alpha) : \sum_{l=1}^{L}\sum_{tuv} \mathbf{c}_{i,l,tuv}\Lambda_{tuv}(\mathbf{r}, \alpha_i.\mathbf{R}_i)$, where $\mathbf{c}_{i,l,tuv}$ are the Hermite coefficients and L denotes the different exponents in the auxiliary basis set on center i. Given a distribution of this form, the Coulomb energy is given by:

$$
\begin{aligned}
E(\rho^{\{N\}}) = \frac{1}{2}\sum_{\mathbf{n}}{}' \sum_{i=1}^{N}\sum_{l_i\in c}\sum_{t_i u_i v_i} & \mathbf{c}_{i,l_i,t_i u_i v_i} \\
\sum_{j=1}^{N}\sum_{l_j\in c}\sum_{t_j u_j v_j} & (-1)^{(t_j+u_j+v_j)}\mathbf{c}_{j,l_j,t_j u_j v_j} \\
\times & \left(\frac{\partial}{\partial \mathbf{R}_{ijx}}\right)^{t_i+t_j}\left(\frac{\partial}{\partial \mathbf{R}_{ijy}}\right)^{u_i+u_j}\left(\frac{\partial}{\partial \mathbf{R}_{ijz}}\right)^{v_i+v_j} \\
\times & \left\{\frac{erfc(\theta^{1/2}|\mathbf{R}_{ij}-\mathbf{n}|) - erfc(\mu_{l_i l_j}^{1/2}|\mathbf{R}_{ij}-\mathbf{n}|)}{|\mathbf{R}_{ij}-\mathbf{n}|}\right\} \\
+\frac{1}{2\pi V}\sum_{\mathbf{m}\neq 0}\frac{1}{\mathbf{m}^2} & \exp(-\pi^2\mathbf{m}^2/2\theta)\sum_{l_1\in c} S_{l_1}(\mathbf{m}) \\
\times\exp(-\pi^2\mathbf{m}^2/2\theta)\sum_{l_2\in c} & S_{l_2}(-\mathbf{m}) \qquad (4) \\
+\frac{1}{2\pi V}\sum_{\mathbf{m}\neq 0}\frac{1}{\mathbf{m}^2}\sum_{(l_1,l_2)\notin c\times c}^{N} & \exp(-\pi^2\mathbf{m}^2/\alpha_{l_1}) \\
\times\exp(-\pi^2\mathbf{m}^2/\alpha_{l_2}) & S_{l_1}(\mathbf{m})S_{l_2}(-\mathbf{m}) \\
-\frac{\pi}{2V}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l_1\in c}\sum_{l_2\in c} & \mathbf{c}_{i,l_1,000}\mathbf{c}_{j,l_2,000}\left(\frac{1}{\theta}-\frac{1}{\alpha_{l_1}}-\frac{1}{\alpha_{l_2}}\right) \\
-\sum_{i=1}^{N} E_{self}(\rho_i) + \frac{2\pi\mathbf{D}^2}{3V} & + \varepsilon(K),
\end{aligned}
$$

where the first term corresponds to the direct part of the Ewald sum, the second and third terms to the reciprocal part, $\mathbf{R}_{ij} = \mathbf{R}_i - \mathbf{R}_j$, the term involving the unit cell dipole $\mathbf{D} = q_1\mathbf{R}_1 + \ldots + q_N\mathbf{R}_N$ is the surface term, $\varepsilon(K)$ denotes a quantity that converges to 0 as $K \longrightarrow \infty$, $\mathbf{m}$ denotes the

reciprocal lattice vectors, and $1/\mu_{l_i l_j} = 1/\alpha_{l_i} + 1/\alpha_{l_j}$ [37,64]. The structure factors $S_l(\mathbf{m})$ involve derivatives of the Fourier exponential with respect to the Hermite centers as given by:

$$S_l(\mathbf{m}) = \sum_{i=1}^{N} \sum_{tuv} \mathbf{c}_{i,l,tuv} \left(\frac{\partial}{\partial \mathbf{R}_{ijx}}\right)^{t_i+t_j} \left(\frac{\partial}{\partial \mathbf{R}_{ijy}}\right)^{u_i+u_j} \left(\frac{\partial}{\partial \mathbf{R}_{ijz}}\right)^{u_i+u_j} exp(-2\pi i \mathbf{m} \cdot (\mathbf{R}_i)), \tag{5}$$

and $E_{self}(\rho_i)$ is the correction due to the self energy of each Hermite interacting with its replicate:

The direct space contributions can be efficiently evaluated using the McMurchie-Davidson (MD) recursion since the ABSs involve Hermites with $l > 0$, thus, only vertical recursions need to be employed [73]. This recursion has been used to calculate the required erfc and higher derivatives for multipole interactions [74]. This approach was also employed for the Hermite Gaussians [37], where it was shown that the MD recursion is applicable to other types of operators besides $1/r$. For the reciprocal sums three methods were implemented: full Ewald [75], sPME [76] and FFP [77].

The smooth PME (sPME) mehtod is based on the fact that the complex exponential in the Eq. 5 can be approximated by a well behaved function with continuous derivatives. In sPME, the functions are B–splines [76], which can be efficiently calculated using finite support, resulting in the evaluation of the structure factors in O(N(log(N)) time. The structure factors for sPME are approximated by:

$$\begin{aligned} S_l(\mathbf{m}) \approx \lambda(z_1)\lambda(z_2)\lambda(z_3) & \sum_{k_1=-\frac{M_1}{2}+1}^{\frac{M_1}{2}} \sum_{k_2=-\frac{M_2}{2}+1}^{\frac{M_2}{2}} \sum_{k_3=-\frac{M_3}{2}+1}^{\frac{M_3}{2}} \sum_{i=1}^{N} \sum_{tuv} \mathbf{d}_{i,l,tuv} \\ & \times \sum_{l_1=-\infty}^{\infty} \left(\frac{\partial}{\partial w_{1i}}\right)^{t} B(w_{1i} - k_1 - l_1 M_1) \\ & \times \sum_{l_2=-\infty}^{\infty} \left(\frac{\partial}{\partial w_{2i}}\right)^{u} B(w_{2i} - k_2 - l_2 M_2) \\ & \times \sum_{l_3=-\infty}^{\infty} \left(\frac{\partial}{\partial w_{3i}}\right)^{v} B(w_{3i} - k_3 - l_3 M_3) \\ & exp\left(-2\pi i \left(\frac{m_1 k_1}{M1} + \frac{m_2 k_2}{M2} + \frac{m_3 k_3}{M3}\right)\right), \end{aligned} \tag{6}$$

where $\mathbf{d}_{i,l,tuv}$ are the transformed Hermite coefficients obtained from $\mathbf{c}_{i,l,tuv}$ by a change of variable, and $B(w)$ are the B–splines [37,64]. In this way, the structure factors $S_l(\mathbf{m})$ can be approximated as a three–dimensional discrete Fourier transform (3DFFT).

Finally, the FFP method uses a similar method to FFT–based approaches used to accelerate structure factor and density map calculations in macromolecular structure determinations. In this way, FFP is based on the realization that $S_l(\mathbf{m})$ can be approximated by re–writing the reciprocal sum such that $S_l(\mathbf{m})$ are re–expressed as the 3DFFT evaluated at $\mathbf{m}$ of a Gaussian density $\rho(\mathbf{r}) = \sum_{i=1}^{N} q_i \rho_{2\alpha_0}(\mathbf{r} - \mathbf{r_i})$, where $\rho_{2\alpha_0}$ is a normalized Gaussian with exponent $2\alpha_0$.

2.3 Parametrization and MD Simulation Details

The original parametrization of GEM* employed a fitted density for the water monomer calculated using a single orientation of a reference water molecule, with a molecular electronic density obtained at the MP2/aug–cc–pVTZ level. The inter–molecular interactions were parametrized using total inter–molecular interaction energies at the MP2/aug–cc–pVTZ level, combined with energy decomposition analysis using the RVS method with the aug–cc–pVTZ as reference [65].

In the present work, we have updated the parameters by fitting single monomer densities for 500 water molecules in random orientations using the A2 ABS only, with reference monomer densities obtained at the CCSD/aug–cc–pVTZ level of theory. The calculated GEM coefficients were optimized and averaged to obtain one single set of average coefficients. These coefficients were used to update the parameters of the different terms of the GEM* potential. The new parametrization has been performed by optimizing the fitting parameters to reproduce the interactions of 500 water dimers selected from the water dimer surface calculated at the CCSD(T)/CBS level reported by Babin *et al.* [78,79], coupled with EDA results obtained from SAPT2+3/aug–cc–pVTZ as reported in [55] using the SAPT implementation in Psi4 [80]. Additionally, the energies and forces of six hexamers have been employed to ensure accurate reproduction of many–body effects as well as the value of the density and heat of vaporization at 300K to optimize the modified Halgren parameters, similarly to the initial GEM* implementation [65]. The optimized parameters were tested against the 43500 dimers from the water dimer potential energy surface both for total intermolecular energies, as well as individual EDA components (see SI), in addition to various bulk properties (vide infra).

The optimized GEM* parameters were employed to perform MD simulations for water boxes of various sizes in the NVT, NPT and NVE ensembles to test the accuracy and performance of the new parameters and optimized implementation of pmemd.gem. For the constant temperature and pressure simulations, the Berendsen thermostat [81] and Monte Carlo barostat were employed. Simulations for the determination of bulk properties were performed using boxes of 1024 water molecules for at least 1 ns at various temperatures. NVE simulations were performed on water boxes of 512 water molecules. In all cases the polarization tolerance is 10^{-6}, with a non–bonded cutoff of 8 Å and similar cutoff for the direct space cutoff for the Ewald–based methods. For the

integral evaluation, the B–spline order is 6, with FFT grids of 64^3, exchange factor of 6.6899 and exchange cutoff of 6.0 Å. All bulk–property calculations were calculated with a split exponent of 0.15 (i.e. all Hermites assumed as compact). For the Gaussian split performance test, the split exponent was also set to 0.5, which results in one of the Gaussians being diffuse. All MD simulations were carried out using the Beeman integrator with a 1 fs time–step.

3 Results and Discussion

This section presents the results and discussion of the new GEM* parametrization with respect to calculation of bulk properties in Subsection 3.1 as well as the performance of the new GEM* implementation in pmemd.gem and testing of the two Ewald–based methods using two different Gaussian cutoffs in Subsection 3.2.

3.1 Bulk properties

Radial and structural distribution functions were calculated from a 1024 water box at 300 K run under NPT conditions. Figures 1 and 2 show that the new parametrization results in an overall better agreement of the structural features for GEM* compared with experimental data, in particular for the O–O rdf. Interestingly, the new parameters results in slight over–stabilization of the O–H and H–H interactions, in particular for the first solvation shell.

Following the work for the initial GEM* implementation, the new parameters have been evaluated by calculating the density and heat of vaporization for a range of temperatures between 250 and 320 K. The heat of vaporization can be calculated from the difference in potential energies between the liquid and gas phases. The average potential energy of the gas is 0.9 kcal/mol at 300 K as reported previously [65] and follows a similar trend as the bulk. As shown in Figure 3 the new parametrization shows good accuracy for the temperature range close to 300K, especially for the calculated heats of vaporization. However, deviations are exhibited, especially at low temperatures, in particular for the calculated densities, which show errors as large as 4% for the lower temperatures. The observed deviation for the calculated thermodynamic properties is due to the inadequate functional form for the description of the induction and charge transfer terms, as well as other appsoximations such as the lack of nuclear quantum effects (in particular for $\Delta \mathrm{H}_{vap}$). One more possible effect that could affect the accuracy of the calculated densities is the relatively short sampling times (1ns).

Another liquid property that can be used to validate the quality of the potential is the self–diffusion coefficient, $D_{\pm}$. In the present work, the $D_{\pm}$ have been calculated by means of Einstein's relation: $D_{\pm} = lim_{t\to\infty}\langle MSD(t)_{\pm}\rangle(6t)^{-1}$, where $\langle MSD(t)_{\pm}\rangle$ is the mean squared displacement of the water molecule's center of mass and t is the time. The calculated diffusion coefficients for the

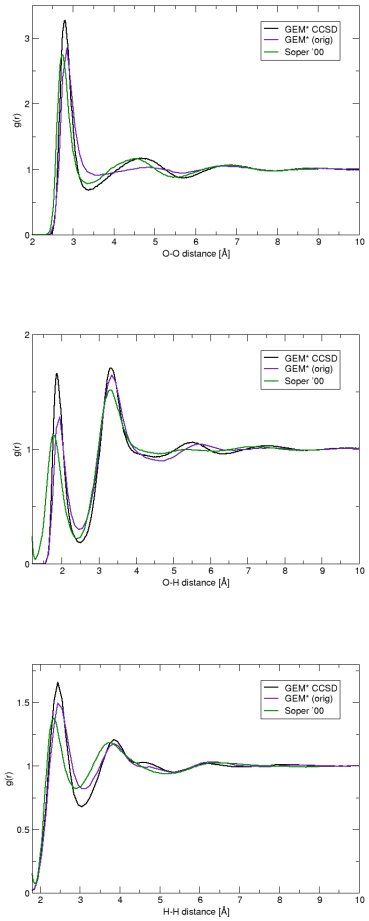

Fig. 1 Radial distribution functions ($g(r)$) for O–O (top), O–H (middle), and H–H (bottom) for GEM* compared with previous GEM* parametrization and experimental data [65, 82]

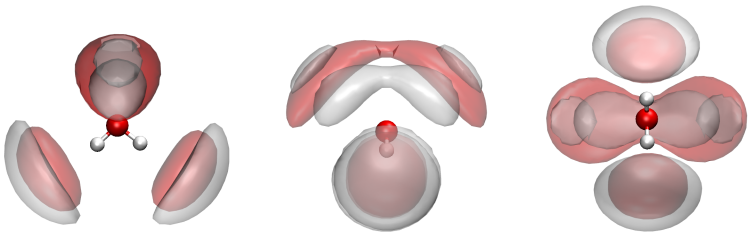

Fig. 2 Structural distribution functions in three orientations calculated with the new GEM* parameters

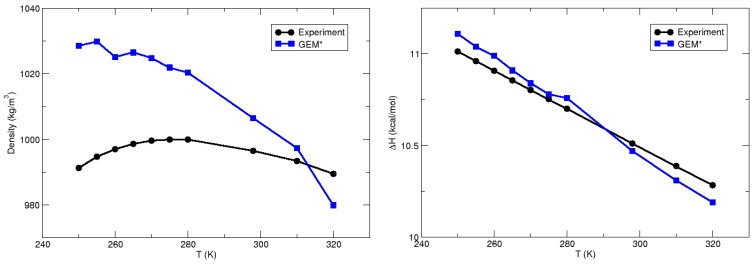

Fig. 3 Calculated densities and heats of vaporization for a range of temperatures

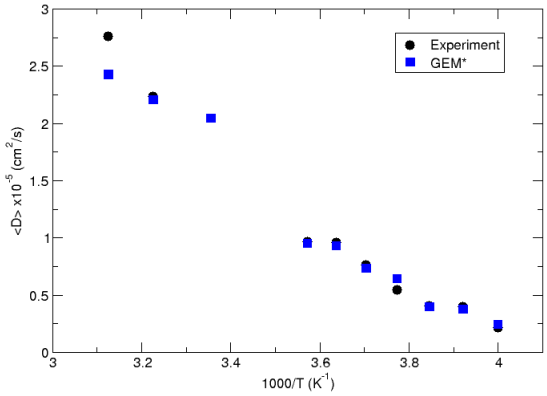

Fig. 4 Calculated diffusion coefficients for a range of temperatures

temperature range under consideration shows very good agreement at both low and high temperatures (Figure 4), and predict faster diffusion at high temperatures and low diffusion rates at low temperatures.

3.2 Ewald–based Methods Performance Analysis

The A2 auxiliary basis sets employed for the current version of GEM* comprise 16 Gaussian primitives with exponents ranging from 0.30 to 2566.00. Timings for for the single energy and force calculation of the Coulomb energy for various water boxes under PBC showed significant improvement of PME and FFP compared with full Ewald sum results [37,64]. The new implementation of the various terms in Equation 6 for improved parallel performance maintains this trend as shown in Table 1. The trends observed in the previous implementation are maintained in pmemd.gem . Additionally, the use of different C–D split exponent (C–D exp), θ in Eqs. 5 and 6, shows the possibility of increased performance. In most cases, it is necessary to use a relatively dense grid for the accurate evaluation of the Coulomb interaction of the entire system if a C–D split exponent is chosen such that some integrals involving diffuse functions are evaluated exclusively in reciprocal space (grid $\geq$ 100 for this example). The

best performance/accuracy is observed for sPME with a C–D split exponent of 0.5, which results in two primitives from the A2 set designated as diffuse, using a grid of 100^3. This combination results in the evaluation of the energies and forces for this system in 1.96 s, with an error of 0.007 kcal mol^1 (0.01 μH) in the total Coulomb energy. It should also be noted that both PME and FFP are almost 30 times faster than the fastest setup for the full Ewald sum.

Table 1 Accuracy and performance of Ewald, sPME and FFP for single Coulomb energy/force calculations of a 512 GEM* water box using a B–spline order of 8. All calculations were performed on an i7-5820K@3.3GHz using a single core.

			grid size = 200^3			
C–D exp	PME		FFP		Ewald	
	E^{Coul}	t (s)	E^{Coul}	t (s)	E^{Coul}	t (s)
0.100	-7761.784660	6.72	-7761.784660	156.62	–	–
0.250	-7761.784659	6.04	-7761.784659	52.52	–	–
0.500	-7761.784724	6.42	-7761.784718	21.99	–	–
1.250	-7761.775239	6.51	-7538.336333	12.26	–	–
2.000	-7761.760290	6.46	6817.189360	10.17	–	–
			grid size = 150^3			
	PME		FFP		Ewald	
	E^{Coul}	t (s)	E^{Coul}	t (s)	E^{Coul}	t (s)
0.100	-7761.784660	4.56	-7761.784660	62.55	–	–
0.250	-7761.784661	4.37	-7761.784659	21.74	–	–
0.500	-7761.784238	4.24	-7761.784720	10.17	–	–
1.250	-7761.658504	4.26	-7619.653416	5.71	–	–
2.000	-7761.692765	4.83	2518.484768	4.85	–	–
			grid size = 100^3			
	PME		FFP		Ewald	
	E^{Coul}	t (s)	E^{Coul}	t (s)	E^{Coul}	t (s)
0.100	-7761.784659	2.50	-7761.784660	19.42	-7761.784660	410.56
0.250	-7761.784643	2.13	-7761.784659	7.59	-7761.784659	408.89
0.500	-7761.791994	1.96	-7761.784723	3.82	-7761.784677	406.25
1.250	-7757.544636	1.93	-7709.917195	2.47	-7758.809650	410.10
2.000	-7787.095605	1.92	-2837.159878	2.53	-7787.734594	387.86
			grid size = 50^3			
	PME		FFP		Ewald	
	E^{Coul}	t (s)	E^{Coul}	t (s)	E^{Coul}	t (s)
0.100	-7761.784675	1.95	-7761.784660	4.22	-7761.784660	64.51
0.250	-7761.788669	1.46	-7761.784888	2.21	-7761.784659	64.05
0.500	-7734.891895	1.18	-7753.682565	1.42	-7747.938468	65.00
1.250	57367.505736	1.26	-41549.335491	1.22	45434.658996	64.72
2.000	-7761.692765	1.16	-13677.969316	1.17	9852.239609	61.69

The stability of the present implementation of these methods was tested by performing a short (200 ps) run of the 512 box under NVE conditions. Figure 5 shows that performing the evaluation of integrals involving diffuse Gaussians exclusively in reciprocal space results in energy conservation for the simulation time. For comparison, an NVE simulation with a C–D exp of 0.15 (all Gaussians considered as compact) is presented to show energy conservation in this regime as well.

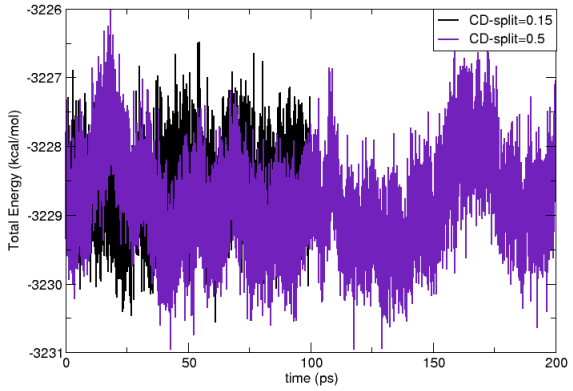

Fig. 5 Total energy vs. time for a 512 water box using the NVE ensemble with two different C–D exp (0.15 and 0.5).

Finally, Figure 6 shows a comparison between the performance of the original implementation of GEM* in the modified pmemd.amoeba implemented in AMBER16, with the new pmemd.gem implementation released in AMBER18. As can be observed, good scaling is observed for a water box of 1024 molecules,. In particular, the evaluation of 100 MD steps running on Stampede2 shows a 2–fold improvement for the new pmemd.gem implementation, with a top performance of 22 MD steps per second on 432 cores.

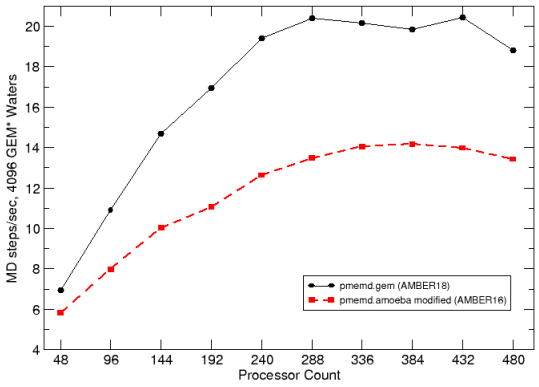

Fig. 6 Performance comparison between the initial and current implementations of GEM*.

4 Conclusions

This contribution presents the implementation and testing of two Ewald–based methods compared with full Ewald sums for the evaluation of integrals involving Gaussian distributions as applied for the GEM* potential. A new parametrization of GEM* employing very high accuracy data for intermolecular interactions for total and EDA contributions is also presented and tested by determining various bulk properties. Our results show that the new GEM* implementation at the higher level of theory provides an accurate reproduction of the water dimer surface and gives accurate results for bulk properties around 300K, although deviations are observed especially at low temperature. These deviations are likely due to the very approximate functional form employed to reproduce the dispersion and charge–transfer contributions, a new version of GEM with a more accurate functional form is currently under development. The use of Ewald–based methods is shown to provide a viable approach for the evaluation of a large number of Gaussian distributions under PBC with high accuracy, in particular when evaluating some of the Gaussian primitives exclusively in reciprocal space.

5 Supplementary Information

SAPT2+3/aug–cc–pVTZ components compared with individual GEM*, and total intermolecular interaction energies for the full 43500 water dimers are provided as a csv file.

Acknowledgements This work was supported by NIH/NIGMS via grant R01GM108583. Computing time from CASCaM at UNT and XSEDE (Stampede2) project #CHE160044, is gratefully acknowledged, with partial support from NSF Grant No. CHE1531468. The authors thank Mr. E.A. Vázquez–Montelongo for assistance with preparing the SDF images.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. D.A. Case, T.E. Cheatham III, T.A. Darden, H. Gohlke, R. Luo, K.M. Merz Jr., A. Onufirev, C. Simmerling, B. Wang, R.J. Woods, J. Comput. Chem. **26**, 1668 (2005)
2. W.L. Jorgensen, J. Tirado-Rives, J. Comput. Chem. **26**, 1669 (2005)
3. D. van der Spoel, E. Lindahl, B. Hess, G. Groenhoff, A.E. Mark, H.J.C. Berensen, J. Comput. Chem. **26**, 1701 (2005)
4. M. Christen, P.H. Hünenberger, D. Bakowies, R. Baron, R. Brúgl, D.P. Geerke, T.N. Heinz, M.A. Kastenholz, V. Kräutler, C. Oostenbrink, C. Peter, D. Trzesniak, W.F. van Gunsteren, J. Comput. Chem. **26**, 1719 (2005)

5. A. D. MacKerrell Jr., B. Brooks, C. L. Brooks III, N.B. Roux, Y. Won, M. Karplus, *"CHARMM: The energy function and its parametrization with an overview of the program" in Encyclopedia of computational Chemistry* (John Wiley & Sons Ltd., New York, NY, 1998)
6. R. Salomon-Ferrer, D.A. Case, R.C. Walker, Wiley Interdisciplinary Reviews: Computational Molecular Science **3**(2), 198 (2013)
7. A.J. Stone, *The theory of intermolecular forces* (Oxford University Press, Oxford, UK, 2000)
8. J.M. Hermida-Ramón, S. Brdarski, G. Karlström, U. Berg, J. Comput. Chem. **24**(2), 161 (2003)
9. N. Gresh, G.A. Cisneros, T.A. Darden, J.P. Piquemal, J. Chem. Theo. Comp. **3**, 1960 (2007)
10. J.W. Ponder, C. Wu, P. Ren, V.S. Pande, J.D. Chodera, M.J. Schnieders, I. Haque, D.L. Mobley, D.S. Lambrecht, Robert A. DiStasio Jr., M. Head-Gordon, G.N.I. Clark, M.E. Johnson, T. Head-Gordon, J. Phys. Chem. B **114**, 2549 (2010)
11. P.N. Day, J.H. Jensen, M.S. Gordon, S.P. Webb, W.J. Stevens, M. Krauss, D. Garmer, H. Basch, D. Cohen, J. Chem. Phys. **105**, 1968 (1996)
12. W. Xie, J. Gao, J. Chem. Theo. Comp. **3**(6), 1890 (2007)
13. W. Xie, M. Orozco, D.G. Truhlar, J. Gao, J. Chem. Theo. Comp. **5**(3), 459 (2009)
14. M.S. Shaik, S.Y. Liem, t..P.j..j.y...v...n...p... Popelier, Paul L. A.",
15. S. Price, in *Reviews in Computational Chemistry*, vol. 14, ed. by K. Lipkowitz, D.B. Boyd (VCH Publishers, New York, NY, 1999)
16. P. Popelier, *Atoms in Molecules: An Introduction* (Prentice Hall, Harlow, England, 2000)
17. D.S. Kosov, P.L.A. Popelier, J. Phys. Chem. A **104**, 7339 (2000)
18. P.L.A. Popelier, L. Joubert, D.S. Kosov, J. Phys. Chem. A **105**, 8254 (2001)
19. P.L.A. Popelier, D.S. Kosov, J. Chem. Phys. **114**, 6539 (2001)
20. M.A. Freitag, M.S. Gordon, J.H. Jensen, W.J. Stevens, J. Chem. Phys. **112**, 7300 (2000)
21. V. Kairys, J.H. Jensen, Chem. Phys. Lett. **315**(1-2), 140 (1999)
22. J.P. Piquemal, N. Gresh, C. Giessner-Prettre, J. Phys. Chem. A **107**, 10353 (2003)
23. G.A. Cisneros, S.N.I. Tholander, O. Parisel, T.A. Darden, D. Elking, L. Perera, J.P. Piquemal, Int. J. Quantum Chem. **108**, 1905 (2008)
24. B. Wang, D.G. Truhlar, J. Chem. Theo. Comp. **6**(11), 3330 (2010)
25. A.J. Stone, J. Phys. Chem. A **115**(25), 7017 (2011)
26. C. Liu, J.P. Piquemal, P. Ren, J. Chem. Theo. Comp. **0**(0), null (0). DOI 10.1021/acs.jctc.9b00261
27. R. Wheatley, Mol. Phys. **7**(3), 761 (2011). DOI 10.1021/ct100530r
28. A. Gavezzotti, J. Phys. Chem. B **106**, 4145 (2002)
29. C.J. Eckhardt, A. Gavezzotti, J. Phys. Chem. B **111**(13), 3430 (2007)
30. A. Volkov, P. Coppens, J. Comput. Chem. **25**, 921 (2004)
31. P. Coppens, A. Volkov, Acta Cryst. A **60**(5), 357 (2004)
32. P. Paricaud, M. Predota, A.A. Chialvo, P.T. Cummings, J. Chem. Phys. **122**(24), 244511 (2005)
33. T.J. Giese, H. Chen, T. Dissanayake, G.M. Giambasu, H. Heldenbrand, M. Huang, E.R. Kuechler, T.S. Lee, M.T. Panteva, B.K. Radak, D.M. York, J. Chem. Theo. Comp. **9**(3), 1417 (2013)
34. H. Hu, Z. Lu, M. Elstner, J. Hermans, W. Yang, J. Phys. Chem. A **111**(26), 5685 (2007)
35. G.A. Cisneros, J.P. Piquemal, T.A. Darden, J. Chem. Phys. **123**, 044109 (2005)
36. J.P. Piquemal, G.A. Cisneros, P. Reinhardt, N. Gresh, T.A. Darden, J. Chem. Phys. **124**, 104101 (2006)
37. G.A. Cisneros, J.P. Piquemal, T.A. Darden, J. Chem. Phys. **125**, 184101 (2006)
38. G.A. Cisneros, D.M. Elking, J.P. Piquemal, T.A. Darden, J. Phys. Chem. A **111**, 12049 (2007)
39. G.A. Cisneros, T.A. Darden, N. Gresh, P. Reinhardt, O. Parisel, J. Pilmé, J.P. Piquemal, *"Design of next generation polarizable force fields from ab initio computations: beyond point charges", in "Multi-scale Quantum Models for Biocatalysis: Modern Techniques and Applications"*. Challenges and Advances in Computational Chemistry and Physics (Springer–Verlag, 2009)

40. S.F. Boys, I. Shavit, *A Fundamental Calculation of the Energy Surface for the System of Three Hydrogen Atoms* (AD212985, NTIS, Springfield, VA, 1959)
41. B.I. Dunlap, J.W.D. Connolly, J.R. Sabin, J. Chem. Phys. **71**, 4993 (1979)
42. A.M. Köster, P. Calaminici, Z. Gómez, U. Reveles, *"Density functional theory calculation of transition metal clusters", in Reviews of Modern Quantum Chemistry, A Celebration of the Contribution of Robert G. Parr* (World Scientific, Singapore, 2002)
43. K. Kitaura, K. Morokuma, Int. J. Quantum Chem. **10**, 325 (1976)
44. P.S. Bagus, K. Hermann, C. W. Bauschlicher Jr., J. Chem. Phys. **80**, 4378 (1984)
45. W.J. Stevens, W.H. Fink, Chem. Phys. Lett. **139**, 15 (1987)
46. B. Jeziorski, R. Moszynski, K. Szalewicz, Chem. Rev. **94**, 1887 (1994)
47. E.D. Glendening, J. Am. Chem. Soc. **118**, 2473 (1994)
48. Y. Mo, J. Gao, S.D. Peyerimhoff, J. Chem. Phys. **112**, 5530 (2000)
49. A. Heßelmann, G. Jansen, M. Schütz, J. Chem. Phys. **122**, 14103 (2005)
50. J.P. Piquemal, A. Marquez, O. Parisel, C. Giessner-Prettre, J. Comput. Chem. **26**, 1052 (2005)
51. R.Z. Khaliullin, M. Head-Gordon, A.T. Bell, J. Chem. Phys. **124**(20), 204105 (2006)
52. Z. Lu, N. Zhou, Q. Wu, Y. Zhang, J. Chem. Theo. Comp. **7**(12), 4038 (2011)
53. G.A. Cisneros, J.P. Piquemal, T.A. Darden, J. Phys. Chem. B **110**, 11571 (2006)
54. R. Chaudret, S. Ulmer, M.C. van Severen, N. Gresh, O. Parisel, G.A. Cisneros, T.A. Darden, J.P. Piquemal, AIP Conf. Proc. **1102**, 185 (2009)
55. H. Gokcan, E. Kratz, T.A. Darden, J.P. Piquemal, G.A. Cisneros, J. Phys. Chem. Lett. **9**(11), 3062 (2018)
56. H. Gokcan, E.A. Vázquez-Montelongo, G.A. Cisneros, J. Chem. Theo. Comp. **15**(15), 3056 (2019)
57. L. Fusti-Molnar, P. Pulay, J. Chem. Phys. **116**(18), 7795 (2002)
58. L. Fü-Molnár, J. Kong, J. Chem. Phys. **122**(7), 074108 (2005)
59. J. Kong, S.T. Brown, L. Fusti-Molnar, J. Chem. Phys. **124**(9), 094109 (2006)
60. C.M. Chang, Y. Shao, J. Kong, J. Chem. Phys. **136**(11), 114112 (2012)
61. L. Exl, N.J. Mauser, Y. Zhang, J. Comput. Phys. **327**, 629 (2016)
62. U. Essmann, L. Perera, M. Berkowitz, T.A. Darden, H. Lee, L.G. Pedersen, J. Chem. Phys. **103**, 8577 (1995)
63. D. York, W. Yang, J. Chem. Phys. **101**(4), 3298 (1994)
64. T.A. Darden, *"Dual bases in crystallographic computing" in International Tables of Chrystallography*, vol. B (Kluwer Academic Publishers, Dordrecht, The Netherlands, 2007)
65. R.E. Duke, O.N. Starovoytov, J.P. Piquemal, G.A. Cisneros, J. Chem. Theo. Comp. **10**, 1361 (2014)
66. P. Ren, J.W. Ponder, J. Phys. Chem. B **107**, 5933 (2003)
67. G.A. Cisneros, J. Chem. Theo. Comp. **12**, 5072 (2012)
68. P. Ren, J.W. Ponder, J. Comput. Chem. **23**, 1497 (2002)
69. P. Ren, C. Wu, J.W. Ponder, J. Chem. Theo. Comp. **7**(10), 3143 (2011). DOI 10.1021/ct200304d
70. J. Andzelm, E. Wimmer, J. Chem. Phys. **96**, 1280 (1992)
71. N. Godbout, J. Andzelm, *DGauss Version 2.0, 2.1, 2.3, 4.0: the file that contains the A1, A2 and P1 auxiliary basis sets can be obtained from the CCL WWW site at http://www.ccl.net/cca/data/basis-sets/DGauss/basis.v3.html* (Computational Chemistry List, Ltd., Ohio, 1999)
72. B. Temelso, K.A. Archer, G.C. Shields, J. Phys. Chem. A **115**(43), 12034 (2011)
73. L. McMurchie, E. Davidson, J. Comput. Phys **26**, 218 (1978)
74. C. Sagui, L.G. Pedersen, T.A. Darden, J. Chem. Phys. **120**, 73 (2004)
75. P. Ewald, Ann. Phys. **64**, 253 (1921)
76. U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, J. Chem. Phys. **103**, 8577 (1995)
77. D. York, W. Yang, J. Chem. Phys. **101**, 3298 (1994)
78. V. Babin, C. Leforestier, F. Paesani, J. Chem. Theo. Comp. **9**(12), 5395 (2013). DOI 10.1021/ct400863t. URL http://pubs.acs.org/doi/abs/10.1021/ct400863t
79. V. Babin, G.R. Medders, F. Paesani, J. Phys. Chem. Lett. **3**(24), 3765 (2012). DOI 10.1021/jz3017733. URL http://pubs.acs.org/doi/abs/10.1021/jz3017733

80. R.M. Parrish, L.A. Burns, D.G.A. Smith, A.C. Simmonett, A.E. DePrince, E.G. Hohenstein, U. Bozkaya, A.Y. Sokolov, R. Di Remigio, R.M. Richard, J.F. Gonthier, A.M. James, H.R. McAlexander, A. Kumar, M. Saitow, X. Wang, B.P. Pritchard, P. Verma, H.F. Schaefer, K. Patkowski, R.A. King, E.F. Valeev, F.A. Evangelista, J.M. Turney, T.D. Crawford, C.D. Sherrill, J. Comput. Chem. **13**(7), 3185 (2017)
81. H.J. Berendsen, J.v. Postma, W.F. van Gunsteren, A. DiNola, J. Haak, J. Chem. Phys. **81**(8), 3684 (1984)
82. A. Soper, P. Rossky, Chem. Phys. **258**, 107 (2000)