

Supplementary Information

Effective Force Coarse-Graining

Yanting Wang, W. G. Noid, Pu Liu, and Gregory A. Voth

Center for Biophysical Modeling and Simulation and Department of Chemistry, University of Utah, 315

S. 1400 E., Rm. 2020, Salt Lake City, Utah 84112-0850, USA

SI. Derivation of Eq (4)

Equation (4) employs the identity:

$$-\frac{\partial u_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_I^{n_I})}{\partial M_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_I^{n_I})} = \hat{\mathbf{M}}_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_I^{n_I}) \cdot \mathbf{f}_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_I^{n_I}) \quad (\text{S.1})$$

where $u_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J})$ and $\mathbf{f}_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J})$ are the atomistic potential and associated force between the atomic groups associated with CG sites I and J defined in eqs (3) and (6), and

$$\begin{aligned} \mathbf{M}_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}) &= \mathbf{M}_{\mathbf{R}I}(\mathbf{r}_I^{n_I}) - \mathbf{M}_{\mathbf{R}J}(\mathbf{r}_J^{n_J}) \\ M_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}) &= \left| \mathbf{M}_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}) \right| \\ \hat{\mathbf{M}}_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}) &= \mathbf{M}_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}) / M_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}) \quad . \end{aligned} \quad (\text{S.2})$$

The present appendix explicitly derives identity (S.1).

It follows from the chain rule that

$$-\frac{\partial u_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_I^{n_I})}{\partial M_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_I^{n_I})} = -\sum_{i \in I_I} \frac{\partial u_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_I^{n_I})}{\partial \mathbf{r}_i} \cdot \frac{\partial \mathbf{r}_i}{\partial M_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_I^{n_I})} - \sum_{j \in I_J} \frac{\partial u_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_I^{n_I})}{\partial \mathbf{r}_j} \cdot \frac{\partial \mathbf{r}_j}{\partial M_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_I^{n_I})} \quad (\text{S.3})$$

The first gradient in the first sum may be directly evaluated from eq (3):

$$-\frac{\partial u_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_I^{n_I})}{\partial \mathbf{r}_i} = \sum_{j \in I_J} \mathbf{f}_{ij}(\mathbf{r}_i, \mathbf{r}_j), \quad (\text{S.4})$$

and similarly for the corresponding gradient in the second summation.

The remaining challenge is to evaluate the partial derivative of each coordinate with respect to the inter-site distance $M_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J})$. To evaluate this derivative it is convenient to represent the Cartesian coordinates for the atoms associated with CG site I , $\mathbf{r}_I^{n_I} = \{\mathbf{r}_i | i \in I_I\}$, with a second set of $3n_I$ coordinates, $\{\mathbf{R}_I, \mathbf{Q}_I^{n_I-1}\}$, such that the coordinates $\mathbf{R}_I = \mathbf{M}_{\mathbf{R}_I}(\mathbf{r}_I^{n_I})$ correspond to the center of mass mapping for the CG site according to eq (2) and the remaining $3(n_I - 1)$ coordinates $\mathbf{Q}_I^{n_I-1}$ (in combination with \mathbf{R}_I) form a linearly independent set of generalized coordinates for the n_I atoms, but are otherwise arbitrary.¹ Because \mathbf{R}_I corresponds to the center of mass for the atomic group it follows that for each atom i involved in CG site I ^{2,3}

$$\mathbf{r}_i = \mathbf{R}_I + \sum_{i=1}^{n_I-1} v_{ii} \mathbf{Q}_{ii} \quad (\text{S.5})$$

for each $i \in I_I$, where the v_{ii} are constant coefficients. It then follows that for the Cartesian components, α, β ,

$$\left(\frac{\partial r_i^\alpha}{\partial R_I^\beta} \right)_{R_I^\gamma, \mathbf{Q}_I^{n_I-1}} = \delta_{\alpha\beta} \quad (\text{S.6})$$

for each $i \in I_I$, where the subscript R_I^γ denotes differentiation with respect to R_I^β while holding fixed the remaining Cartesian components of \mathbf{R}_I . Similarly, the atomic coordinates associated with CG site J may be represented by a set of independent generalized coordinates, $\{\mathbf{R}_J, \mathbf{Q}_J^{n_J-1}\}$, such that

$\mathbf{R}_J = \mathbf{M}_{\mathbf{R}_J}(\mathbf{r}_j^{n_j})$ is the center of mass for the atomic group associated with CG site J . Consequently, for each atom j involved in site J , and for Cartesian components α, β , it follows that

$$\left(\frac{\partial r_j^\alpha}{\partial R_J^\beta} \right)_{R_J^\gamma, \mathbf{Q}_J^{n_j-1}} = \delta_{\alpha\beta}. \quad (\text{S.7})$$

Therefore, the set of $n_I + n_J$ Cartesian coordinates $\{\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J}\}$ may be represented by the set of linearly independent generalized coordinates $\{\mathbf{R}_I, \mathbf{R}_J, \mathbf{Q}_I^{n_I-1}, \mathbf{Q}_J^{n_J-1}\}$. The proof of eq (4) may be completed by employing a third set of generalized coordinates obtained by representing the Cartesian coordinates corresponding to the centers of mass for the two atomic groups, $\{\mathbf{R}_I, \mathbf{R}_J\}$, with sum and difference variables, $\{\mathbf{X}_{IJ}, R_{IJ}, \Omega_{IJ}\}$, where Ω_{IJ} denotes the two angle variables that define the direction of the unit vector between the two groups, $\hat{\mathbf{R}}_{IJ}(\Omega_{IJ})$, such that

$$\begin{aligned} \mathbf{X}_{IJ} &= \frac{1}{2}(\mathbf{R}_I + \mathbf{R}_J) \\ \mathbf{R}_{IJ} &= \mathbf{R}_I - \mathbf{R}_J \end{aligned} \quad (\text{S.8})$$

and

$$\begin{aligned} R_{IJ} &= |\mathbf{R}_{IJ}| \\ \hat{\mathbf{R}}_{IJ}(\Omega_{IJ}) &= \mathbf{R}_{IJ} / R_{IJ} \end{aligned} \quad (\text{S.9})$$

When expressed in terms of Cartesian coordinates, these definitions correspond to eqs (S.2) and (7),

such that $R_{IJ} = M_{IJ}(\mathbf{r}_I^{n_I}, \mathbf{r}_J^{n_J})$. The reverse transformation may be expressed as:

$$\begin{aligned} \mathbf{R}_I &= \mathbf{X}_{IJ} + \frac{1}{2} R_{IJ} \hat{\mathbf{R}}_{IJ}(\Omega_{IJ}) \\ \mathbf{R}_J &= \mathbf{X}_{IJ} - \frac{1}{2} R_{IJ} \hat{\mathbf{R}}_{IJ}(\Omega_{IJ}) \end{aligned} \quad (\text{S.10})$$

Therefore, it follows that

$$\begin{aligned} \left(\frac{\partial \mathbf{R}_I}{\partial R_{IJ}} \right)_{\mathbf{x}_{IJ}, \Omega_{IJ}} &= \frac{1}{2} \hat{\mathbf{R}}_{IJ}(\Omega_{IJ}) \\ \left(\frac{\partial \mathbf{R}_J}{\partial R_{IJ}} \right)_{\mathbf{x}_{IJ}, \Omega_{IJ}} &= -\frac{1}{2} \hat{\mathbf{R}}_{IJ}(\Omega_{IJ}) \end{aligned} \quad (\text{S.11})$$

The desired partial derivative may be evaluated by employing both sets of transformed coordinates:

$$\left(\frac{\partial r_i^\alpha}{\partial R_{IJ}} \right)_{\mathbf{x}_{IJ}, \Omega_{IJ}, \mathbf{Q}_I^{n_I-1}, \mathbf{Q}_J^{n_J-1}} = \sum_{\beta} \left(\frac{\partial r_i^\alpha}{\partial R_I^\beta} \right)_{\mathbf{Q}_I^{n_I-1}} \left(\frac{\partial R_I^\beta}{\partial R_{IJ}} \right)_{\mathbf{x}_{IJ}, \Omega_{IJ}} = \frac{1}{2} \hat{R}_{IJ}^\alpha \quad (\text{S.12})$$

for each $i \in I_I$ and, similarly,

$$\left(\frac{\partial r_j^\alpha}{\partial R_{IJ}} \right)_{\mathbf{x}_{IJ}, \Omega_{IJ}, \mathbf{Q}_I^{n_I-1}, \mathbf{Q}_J^{n_J-1}} = \sum_{\beta} \left(\frac{\partial r_j^\alpha}{\partial R_J^\beta} \right)_{\mathbf{Q}_J^{n_J-1}} \left(\frac{\partial R_J^\beta}{\partial R_{IJ}} \right)_{\mathbf{x}_{IJ}, \Omega_{IJ}} = -\frac{1}{2} \hat{R}_{IJ}^\alpha \quad (\text{S.13})$$

for each $j \in I_J$. It then follows from eqs (S.3), (S.4), (S.12), and (S.13), that

$$-\left(\frac{\partial u_{IJ}}{\partial R_{IJ}} \right)_{\mathbf{x}_{IJ}, \Omega_{IJ}, \mathbf{Q}_I^{n_I-1}, \mathbf{Q}_J^{n_J-1}} = \frac{1}{2} \sum_{i \in I_I} \sum_{j \in I_J} \mathbf{f}_{ij} \cdot \hat{\mathbf{R}}_{IJ} - \frac{1}{2} \sum_{j \in I_J} \sum_{i \in I_I} \mathbf{f}_{ji} \cdot \hat{\mathbf{R}}_{IJ} = \hat{\mathbf{R}}_{IJ} \cdot \mathbf{f}_{IJ}, \quad (\text{S.14})$$

which, when expressed in terms of Cartesian coordinates, corresponds to eq (S.1).

SIII. Comparisons of the EF-CG and MS-CG Models

The simulation results with the EF-CG and MS-CG models for one-site neopentane, two-site methanol, and 5-site EMIM+/NO3- ionic liquid are compared below to show that, in general, the EF-CG method is less accurate but generally more transferable than the MS-CG method. The EF-CG models were constructed in the way described in the paper. The MS-CG models were constructed based on the same all-atom trajectories as used for the EF-CG models, and the bin size is 0.01 Å for all systems. The force cutoff is 12 Å for neopentane, 10 Å for methanol, and for 12 Å for ionic liquid. The

MS-CG and EF-CG CG pair force forces for neopentane are compared in Figure S1 where it is seen that the EF-CG force field contains less statistical noise than the MS-CG case due to the simpler averaging procedure of the former. The CG MD runs with the MS-CG force fields were the same as those with the EF-CG force fields, except the effective forces.

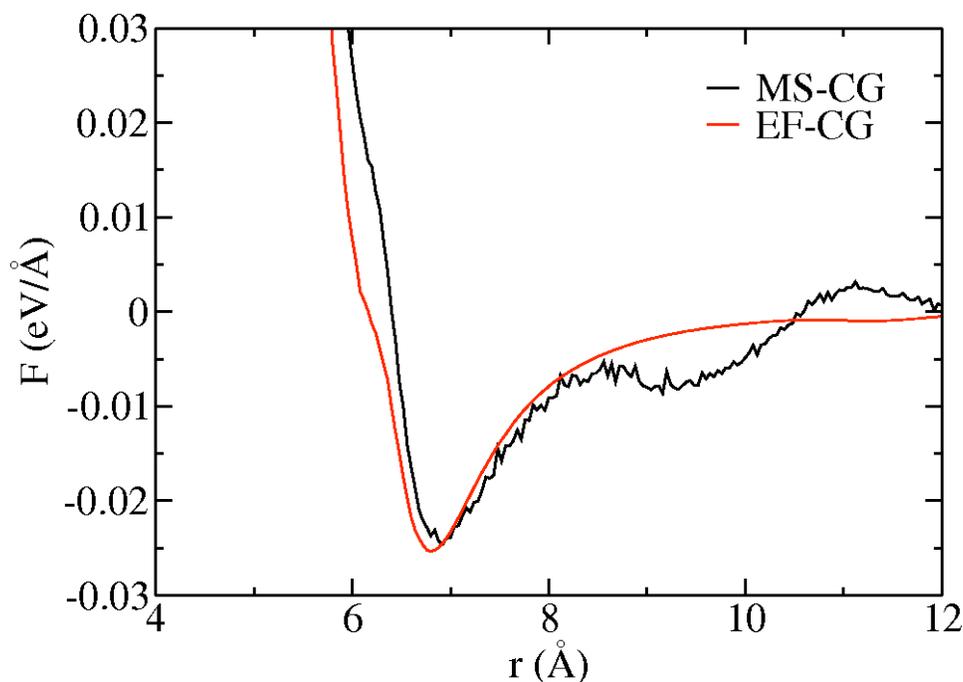


Figure S1. The MS-CG and EF-CG forces for neopentane.

a) Interfacial Properties

Figure S2 compares the running averages of the surface tension for 1000 methanol molecules computed from simulations of an all-atom model and from simulations of two-site EF-CG and MS-CG models. The EF-CG model semiquantitatively reproduces the surface tension of the atomistic model. In contrast, the surface tension generated by the MS-CG model is quite different from the surface tension of the all-atom model. This indicates that MS-CG force fields constructed from bulk may not accurately

model interfaces without additional improvements, but the EF-CG force fields constructed from bulk systems may provide a semiquantitatively accurate description of such interfaces.

b) Pair distributions

The RDFs for the B-B sites of methanol at $T = 300$ K are shown in Figure S3. It can be seen that, in comparison to the EF-CG model, the MS-CG model better reproduces the pair distribution for B-B sites. This is because the MS-CG method incorporates the many-body effects into the effective pair forces to provide an optimal approximation to the many-body PMF, and thus it provides an improved description of the average pair structure. Similar results can be seen for the ionic liquid by comparing Figure 12 in the paper and Figure 6 in ref 4.

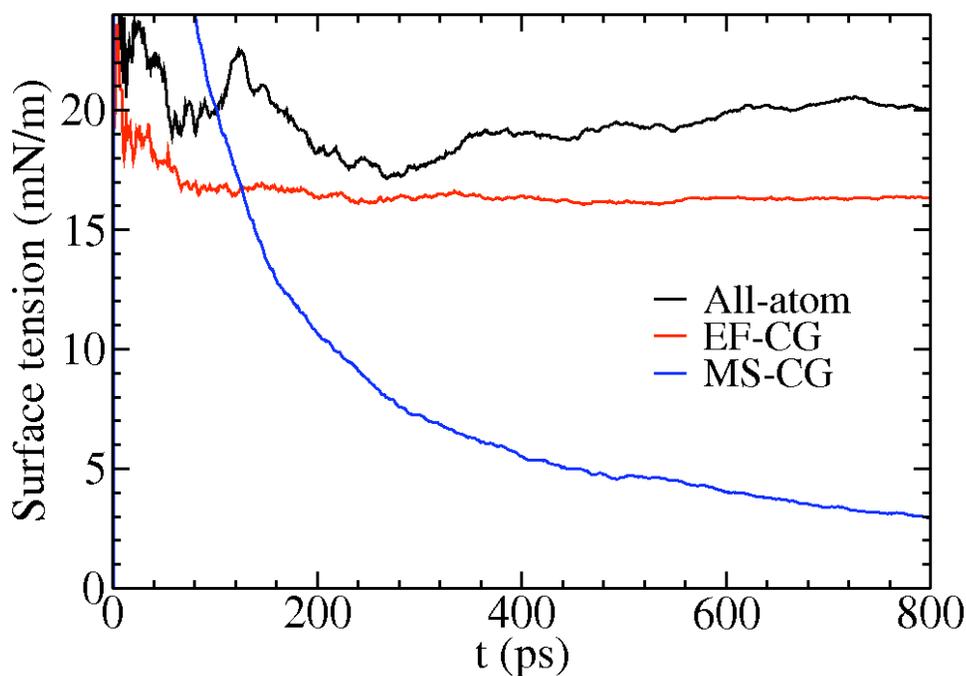


Figure S2 Running averages of the surface tension for 1000 two-site methanol.

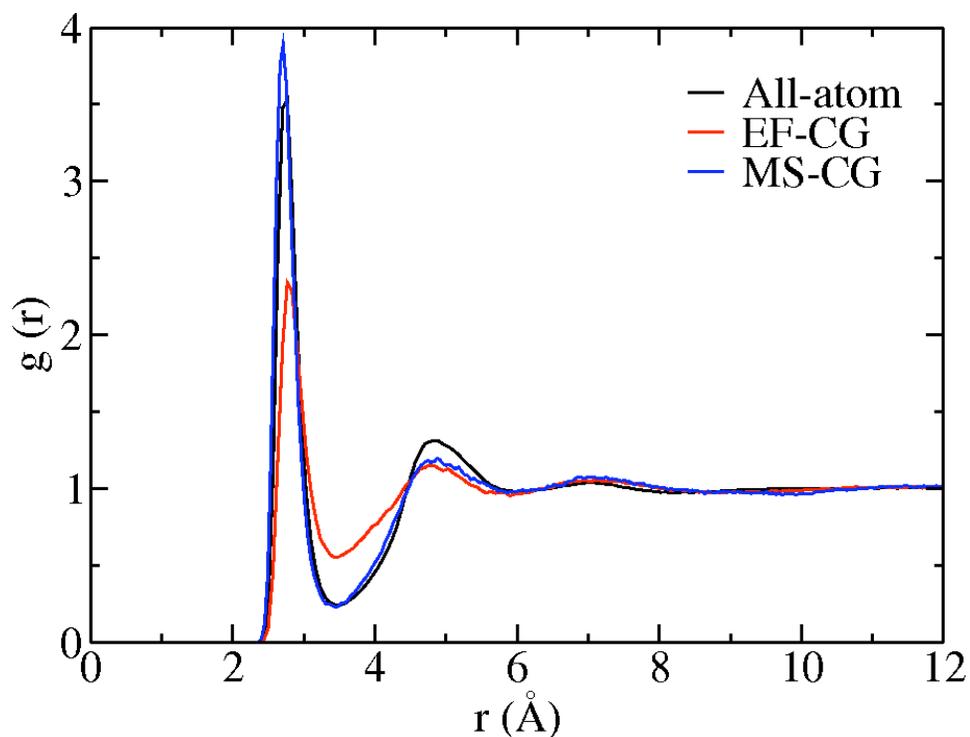


Figure S3. RDFs for the B-B sites of methanol at $T = 300$ K.

c) Transferability between temperatures

The MS-CG and EF-CG models for methanol constructed at $T = 300$ K are used to simulate the same system at $T = 1000$ K. Figure S4 compares the B-B RDFs computed from these high temperature simulations with the B-B RDF computed from all-atom MD simulations at the same temperature. In comparison to the MS-CG model, the EF-CG model provides a somewhat better description of the B-B pair structure generated by the atomistic model.

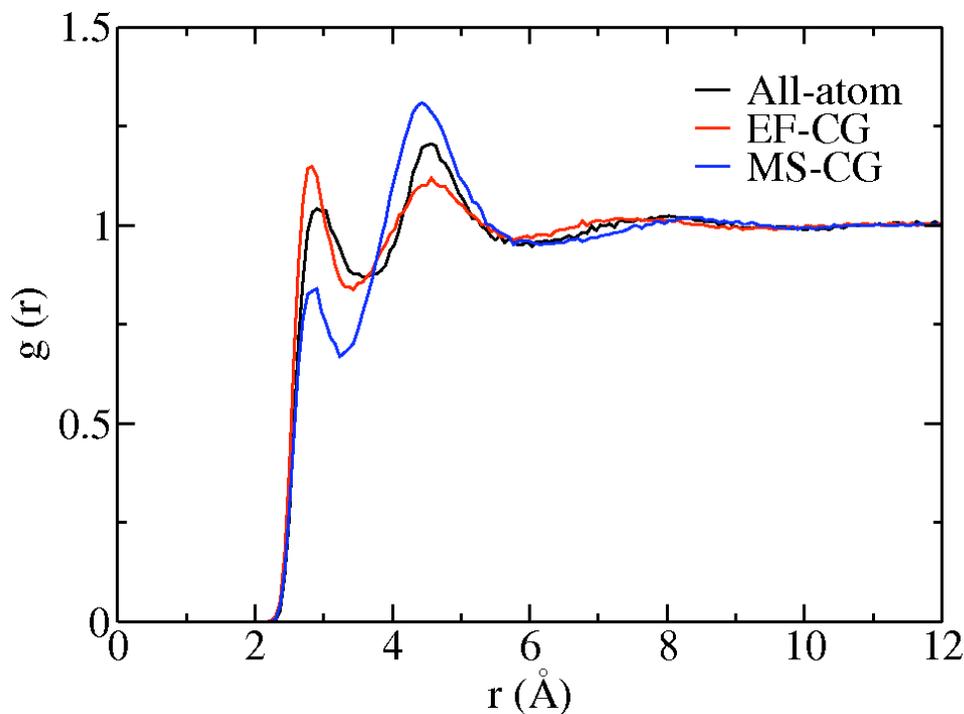


Figure S4. RDFs for the B-B sites of methanol at $T=1000$ K.

A similar comparison for the D-D RDF of the ionic liquid system is given in Figure S5. In this case, the MS-CG model provides a better description of the D-D pair structure. Therefore, it is generally not possible to draw a universal conclusion which method has better temperature transferability. Indeed, we anticipate that the origins of the errors are different for these two methods. The errors in the MS-CG method are mainly introduced by the temperature-dependent many-body entropic effects incorporated into the effective forces. In contrast, the errors in the EF-CG method result primarily from the integration over spatial orientations, which we anticipate should be less temperature dependent. However, more extensive studies will be necessary to more fully understand the temperature transferability of both methods.

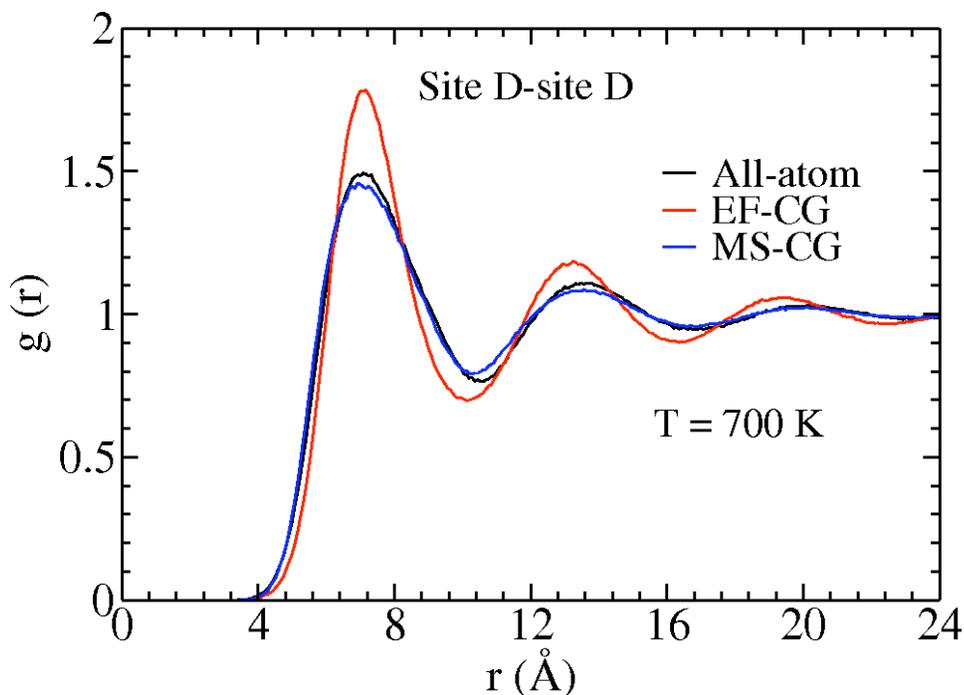


Figure S5. RDFs for the D-D site of ionic liquid at $T = 700$ K.

d) Computational Complexity

The computational complexity of both methods depends upon system sizes, force calculation parameters, and computer hardware. However, simple considerations demonstrate that calculations of the MS-CG force field are significantly more complex than calculations of the EF-CG force field. Consider a system that is described by N CG sites of N_T distinct types with $N_T(N_T + 1)/2$ distinct types of pair nonbonded interactions, each of which is described by N_D parameters and for which n_i configurations have been sampled. For a typical complex system there may be $N \sim 10^3$ CG sites of $N_T \sim 10$ different types, each different pair interaction may be represented by $N_D \sim 10^3$ different parameters and $n_i \sim 10^3$ or more configurations may be necessary to determine an accurate estimate of the MS-CG force field.

The parameters for the MS-CG force field are determined by solving a high-dimensional linear least squares problem. Iterative matrix decomposition techniques, such as QR decomposition, may be employed, but these methods require iteration with a matrix of roughly $3n_i N \sim 10^6$ rows and $N_T(N_T + 1)N_D / 2 \sim 10^4$ columns, which can require considerable computational memory. Alternatively, the MS-CG force field parameters may be determined by directly solving the associated system of normal equations, involving a normal matrix of roughly $N_T(N_T + 1)N_D / 2 \sim 10^4$ rows and an equal number of columns. Although this approach requires substantially less memory, the evaluation of the normal matrix requires the computation of many-body correlation functions, which are time-intensive to evaluate and which also require extensive sampling. Moreover, the resulting system of equations is relatively strongly coupled and may be quite ill-conditioned. The above considerations demonstrate that the numerical difficulty of the MS-CG calculation, by either iterative or direct methods, rapidly grows with increasing size and complexity.

In comparison, the EF-CG approach calculates each of the $N_T(N_T + 1) / 2 \sim 10^2$ types of pair interactions independently by evaluating the relatively simple correlation function expressed in Eq (4) separately for each pair of site types. Moreover, such a calculation may be simply parallelized.

Here we only give one example for a relatively small system. Both the MS-CG and EF-CG calculations were performed on the same computer for a 64 ion-pair EMIM⁺/NO₃⁻ system with 4000 sampled all-atom configurations. The MS-CG calculation with the bin width of 0.04 Å and the cutoff of 12 Å took 137 CPU minutes. The EF-CG calculation with the bin width of 0.01 Å took 85 CPU minutes. However, for larger systems and finer bins the computational cost of the MS-CG approach grows faster than the EF-CG approach.

References

1. Goldstein, H., Poole, C. & Safko, J. *Classical Mechanics*, (Addison-Wesley, Reading, MA, 2002).
2. Ciccotti, G. & Ryckaert, J.P. *Comput. Phys. Rep.* 4, 345 (1986).
3. Fixman, M. *Proc. Natl. Acad. Sci. U.S.A.* 71, 3050 (1974).
4. Noid, W.G. et al. The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models. *J. Chem. Phys.* 128, 244115 (2008).