Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Bottom-Up Derived Flexible Water Model with Dipole and Quadrupole Moments for Coarse-Grained Molecular Simulations

Chen Li,[†] Zhongyuan Qin,[†] Wei Han[*†]

[†]State Key Laboratory of Chemical Genomics, School of Chemical Biology and Biotechnology, Peking University Shenzhen Graduate School, Shenzhen, 518055, China

*Corresponding author: email: hanw@pkusz.edu.cn

Electronic Supplementary Information

S1. Clustering of water clusters based on both energy and geometry

In the first step of the parameterization of our CG model, we need to partition water molecules in an arbitrary configuration into clusters of four based on the atomistic coordinates. We solved this partitioning problem by employing a MC-based algorithm to minimize the partition energy Θ_{par} defined in **Eq. (3)**, a quantity which assesses the quality of the partitioning. Θ_{par} consists of two terms, one pertaining to the internal energy (U_{intra}) of clustered water molecules and the other pertaining to the compactness (A) of the cluster indicated by the surface area of the tetrahedron formed by these water molecules. The relative importance of the two terms is controlled by the factor λ.

The internal energy (U_{intra}) intuitively would be the most natural indicator of the formation of water clusters since water molecules in the same cluster are supposed to interact strongly with each other. Hence, we first attempted to cluster waters based purely on the internal energy (λ=0). For comparison, a distance-based clustering approach described previously (see also **Eq. 2**) was conducted using the same sets of AA coordinates. This approach evaluates the quality of the clustering results based on the compactness of the resulting clusters. Clearly, the energy-based approach yielded much lower internal energy per water cluster than the distance-based approach (-72 kJ/mol *versus* -49 kJ/mol) (**Table S1**). However, the analysis of the RDF of COMs of water clusters revealed that the water clusters generated by the energy-based approach could come very close and overlap with each other (**Figure S1a**). This is problematic for MD simulations as a small or zero distance between interacting particles tends to cause instability in the simulations. On the other hand, the distance-based clustering method has no such issue (data not shown). The RDF of water clusters derived with this method displayed noticeable distribution only beyond R = 2 Å.

To deduce how the clusters obtained with the energy-based method could come so close to each other, we analyzed the structures of those water clusters that were separated by distances of < 2 Å. **Figure S1b** illustrates one such representative structure. To our surprise, the water molecules in these clusters were rather scattered, which explains why it was possible for the COMs of the clusters to be positioned in close proximity to one another. Using the average surface area A of a tetrahedron formed by four water molecules as a measure of cluster compactness (see the Method section), we assessed this property for water clusters separated at various distances. The calculated A values obtained with the energy-based method were much larger than those obtained with the distance-based method (60-180 Å *versus* ~20 Å), especially for those water clusters with short separations, suggesting that the energy-based

method can sometimes generate clusters composed of scattered water molecules. We attributed this deficiency to how U_{intra} is defined (see the Method section). U_{intra} is sensitive to subtle configurational differences between compact clusters but imposes little penalty on clusters with scattered members. It is likely that the energy-based clustering algorithm allowed some water clusters with scattered water molecules to exist in order to ensure that other clusters have excellent internal geometry.

We thus gradually increased the λ factor, attempting to find a minimum value with which U_{intra} was still able to play a significant role in clustering while the clusters with scattered members could be avoided. It turned out that a value of 0.1 is a reasonable choice (**Figure S1a**). The resulting clusters exhibited similar a cluster size to that of distance-based water clusters (**Figures S1a** and **S1c**), but their internal energy was considerably lower (-57 kJ/mol *versus* -49 kJ/mol). The representative cluster structures (**Figure S1d**) displayed a good tetrahedral geometry, an essential feature that needs to be captured in coarse-graining of water, as suggested previously.[1] Hence, this λ value was adopted in the present study.

S2. Cubic spline functions for short-range nonbonded interactions

The short-range interactions $u_{IJ}(R)$ are approximated by a cubic spline that connects a series of data points that are distributed at $\{R_k\}$ with function values of $\{u_k\}$. For a R between R_k and R_{k+1}, the corresponding $u(R)$ can be expressed as

$$u(R,\{u_k,u_k''\}) = (1-\eta)u_k + \eta u_{k+1} + \Delta R^2\left(\frac{1}{2}\eta^2 - \frac{1}{6}\eta^3 - \frac{1}{3}\eta\right)u_k'' + \Delta R^2\left(\frac{1}{6}\eta^3 - \frac{1}{6}\eta\right)u_{k+1}'' \quad \text{(S1)}$$

where ΔR is the spacing between data points, η is $(R - R_k)/\Delta R$ with $R_k = k \times \Delta R + R_0$, and u_k'' is the secondary derivative of $u(R)$ at R_k. $\{u_k\}$ and $\{u_k''\}$ are the parameters of $u(R)$ and can be determined using target values of the function at each R_k with conditions that the spline is differentiable at these positions. The spline is capped at R_0=2 Å and R_C=12 Å. Below R_0, the reference data for spline parameterization is normally difficult to obtain as atoms rarely approach each other so closely. As such, $u(R<R_0)$ is modeled as $u(R) = a_u/R^{12} + b_u$ where a_u and b_u are parametrized to ensure continuity and differentiability at R_0. Beyond R_C, short-range interactions are thought to disappear. Finally, the natural boundary condition, i.e., $u''(R_0) = u''(R_C) = 0$, is applied. Given that there are two types of interaction sites in our model, three cubic splines, are needed to represent the short-range interactions.

S3. Implementation of iterative Boltzmann inversion method

U^{CG} can be determined using the iterative Boltzmann inversion (IBI) method.[2,3] For instance, the nonbonded interaction $u_{AB}(R)$ between sites A and B (**Eq. 10**) can be parameterized with IBI to match the RDFs between the two types of sites. In the IBI procedure, the interaction potential is gradually improved through multiple rounds of optimization. In the i^{th} round, the potential energy function $u_{AB}^{(i)}$ is updated according to the results of the last round by following:

$$u_{AB}^{(i)}(R) = u_{AB}^{(i-1)}(R) + \xi \ln\left(\frac{g_{AB}^{(i-1)}(R)}{g_{AB}^{\mathrm{ref}}(R)}\right) \quad \text{(S2)}$$

where ξ is a tunable parameter of the Boltzmann inversion method to control the update speed and usually takes a value of k_BT, and $g_{AB}^{(i-1)}(R)$ and $g_{AB}^{\mathrm{ref}}(R)$ are the pair of correlation functions (or RDFs) derived, respectively, either from the last round of CG simulation or from the mapping results obtained from AA simulation as described in the previous section. An initial guess for u_{AB}^{0} was made based on $u_{AB}^{0}(R) = -k_B\mathrm{T}\ln g_{AB}^{\mathrm{ref}}(R)$. We monitored the deviation of the RDFs between the CG and AA systems with a functional $f[g^{(i)}]$ defined as:[4]

$$f\left[g^{(i)}\right] = \frac{\int dR(g_{AB}^{(i)}(R) - g_{AB}^{\mathrm{ref}}(R))^2}{\int dR(1 - g_{AB}^{\mathrm{ref}}(R))^2}. \quad \text{(S3)}$$

The optimization was thought to converge if the values of $f[g^{(i)}]$ in two sequential rounds were smaller than 10^{-3}.

The parameterization of bonded parameters shown in **Eq.** (**11**) is similar. Of note, as a harmonic form of potential function was used to model potential energy of bond stretching and angle bending, the IBI scheme illustrated in **Eq.** (**S2**) cannot be applied directly to the fitting of these bonded parameters, and a modified IBI was adopted here. The AA simulations showed that the distributions of bond length and bond angle were apparently consistent with a normal distribution and thus can be characterized by the peak position and height. Based on this, we assumed that it is possible to use these features as targets to optimize the force constant k_b/k_θ and the reference position R_0/θ_0. For a simple system whose Hamiltonian is a function of a scalar x, such as R_b or θ and has a form of $k_x(x\text{-}x_0)^2$, its maximum probability density is proportional to $(\pi k_B\mathrm{T}k_x)^{1/2}$. This relationship suggests that in each modified IBI round, the bonded parameter k_x can be updated according to:

$$k_x^{(i)} = k_x^{(i-1)} + \xi_k\left(P_{(i-1)}^2 - P_{\mathrm{AA}}^2\right) \quad \text{(S4)}$$

where ξ_k is again a scalar tuning the speed of update of the force constant and $P_{(i\text{-}1)}$ and P_{AA} are the maximum probabilities of x observed in the i-1th round of CG simulation and the AA

simulation, respectively. The reference position x_0 can be updated according to the difference between the peak positions of x distributions obtained from the previous round of CG simulation ($x_{peak}^{(i-1)}$) and the AA simulation (x_{peak}^{AA}), i.e.:

$$x_0^{(i)} = x_0^{(i-1)} + \xi_x \left(x_{\text{peak}}^{(i-1)} - x_{\text{peak}}^{\text{AA}} \right). \quad \text{(S5)}$$

The iteration stops if the following conditions are met: (1) R_0 - $R_0^{AA} \leq 0.1$ Å; (2) θ_0 -$\theta_0^{AA} \leq 1^o$; (3) P_{peak}^{CG} - $P_{peak}^{ref} \leq 0.05$.

S4. Dielectric constant calculation

The dielectric constant was calculated from the potential of mean force between ions in FlexDQ water. For atomistic system, one can calculate dielectric constant by following the standard Kirkwood scheme directly. While for CG models, there is a small difference in its definition of the permittivity. A dielectric screening coefficient, ε_r=2.4, is introduced into the FlexDQ water model. Therefore, the dielectric properties has contribution from both this screening coefficient and the configurations of the CG multipoles. Thus, the oginial definition of dielectric permeability in the Kirkwood scheme may not be applicable for the case investigated here. Instead, we followed Zhe et al. [5]to calculate the permittivity from the PMF between charged ions. The basic idea is that according to the general definition of ε_r, it can be calculated as the ratio of the work required to bring two charges at a distance R in vacuum to the work required in a given medium, i.e., $\varepsilon_r = q_1 q_2 / 4\pi\varepsilon_0 G(R)$. Here G(R) is the PMF of separating two ions at R in water, which can be obtained with the umbrella sampling method. $G(R)$ will deviate from the Coulomb law for short R due to the contribution from short-range nonbonded interaction. However, the long-range part of the PMF (for R ranging from 1.2 nm to 1.4 nm) where the short-range interactions vanish can be fitted to obtain the permittivity for the FlexDQ model by minimizing

$$\sum (G(R) - \frac{q_1 q_2}{4\pi\varepsilon R})^2. \quad \text{(S6)}$$

S5. Parameterization of FlexDQ in tandem with MARTINI

We optimized the interactions of wn/wp sites with MARTINI particles by reproducing hydration free energy of each MARTINI particle type. Both the original MARTINI water model and its PMW version were parameterized in this way to ensure that they can be used to model correctly partitioning behavior of biomolecules between water and other organic

solvent.[6] Short-range nonbonded interaction between CG sites in MARTINI is described with the LJ potential, i.e., $\frac{C_{12}}{R^{12}} - \frac{C_6}{R^6}$, where the coefficients C_{12} and C_6 determine the strength of repulsive and attractive forces between CG sites. Following Yesylevskyy *et al.*, we only consider the interaction of the wn site with the MARITNI particles while the wp site is not visible to the MARTINI particles except for those charged types "Q" (see the next paragraph). For non-charged MARTINI types, we first attempted to use the original parameters optimized for the PMW model. The hydration free energy was calculated with the thermodynamic integration method. The resulting hydration free energy turned out to be systematically overestimated (data not shown), indicating the interaction between the wn site and the MARTINI particles was too strong. We then gradually scaled down interactions by multiplying all the C_{12} and C_6 coefficients with a factor and eventually found that a factor of 0.8 could reproduce hydration free energy with a reasonable accuracy (**Table S2**).

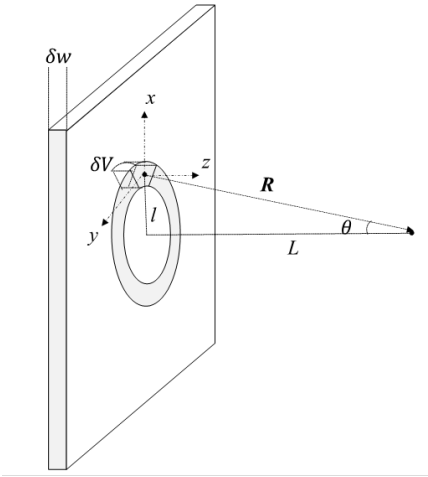
A key advantage of the MARTINI model is that a large time step can be used to boost simulation speed. The time step for the PMW model is normally 20 fs. Although the time step that permits stable simulation with our model can be 18 fs for pure water system and 16 fs when non-charged MARTINI particles are included. The simulation with our model in tandem with charged MARTINI particle was unstable. This largely resulted from the lack of short-range repulsion between the wp sites and the charged MARTINI particles. As our water model is flexible, the wp site may not be always kept within repulsive radius of the central wn site, which is different from the distal sites of the PMW model that are attached to the central site with fixed bonds. Hence, there is a chance that the wp sites overlap with the non-charged MARTINI particles through fluctuation, causing instability of simulations. To avoid this, we added a short-range repulsive interaction between the wp sites and the charged MARTINI particles. For this interaction, its C_6 coefficient was set to zero and C_{12} was set to a small number. We tried the value of C_{12} in the range of 10^{-10}~10^{-6} and found that when C_{12} was not less than 10^{-7}, simulations with a time step of 14 fs were stable. We also found that the resulting hydration free energy of the charged particles in the FlexDQ water remained invariant when C_{12} took values ranging from 10^{-10} to 10^{-6} (**Table S2**). Based on these findings, we chose a C_{12} = 10^{-7} that was sufficiently small but still allowed us to use a time step of 14 fs. Of note, our calculation gave rise to very similar hydration free energy for various particle types with either a 2 fs time step or a 14 fs time step, suggesting that a large time step of 14 fs does not incur any noticeable error in numeric integration in our cases.

As pointed out previously, the original MARTINI water model, due to the lack of description

electrostatic interactions, underestimated its interaction with charged solutes, causing the unphysical presence of these solutes in nonpolar environment.[7] The embedded charges in the PMW model improved description of interactions between charged particles and water, yielding more negative hydration free energy of ions and thus resulting in more realistic behaviors.[6] Our model, which was also designed to capture electrostatic elements of solute-solvent, predicted hydration free energy of the charged particles even more negative than those obtained with the PMW model (**Table S2**). As the charged particle types ("Q" types) represent monovalent ions plus its first solvation shell, the hydration free energy of particles of these types was difficult to conceive conceptually and to measure experimentally. Thus, there is no ground truth for us to determine which of the results is correct. Nonetheless, these hydration free energy results are still well below the known hydration free energy of bared monovalent ions (e.g., -365 kJ/mol for Na^+ and -340 kJ/mol for Cl^-)[8], which should not be against physical intuition. Taken together, these results suggest that our model may furnish a reasonable description of solvation of charged solutes, preventing their presence in nonpolar environment although a more systematic evaluation of the feasibility of the combination of MARTINI and FlexDQ is still needed.

S6. Derivation of Eqs. 17 and 18

Let us consider an infinite slab of neutral molecules (**Scheme S1**). The slab has an infinitesimal width δw. Let the molecule have an arbitrary charge distribution $\rho_C(\boldsymbol{r})$ where $\boldsymbol{r}$ is a position vector. Let the system be anisotropic in the direction of the slab normal but isotropic in the perpendicular directions. We seek to learn the electrostatic potential at a given point $\boldsymbol{R}$ whose distance to the slab is L.



Scheme S1. Calculation of electrostatic potential at $\boldsymbol{R}$ arising from a slab of molecules with dipole and quadrupole moments.

We first must consider the electrostatic potential of a single molecule at $\boldsymbol{R}$. Let $\boldsymbol{\mu}$ and $\boldsymbol{Q}$ be dipole moment vector and quadrupole moment tensor of the molecule. For a discrete charge distribution, $\boldsymbol{\mu}$ and $\boldsymbol{Q}$ can be expressed as in **Eqs. 4** and **5**. When the origin is properly chosen such that $|\boldsymbol{r}| \ll |\boldsymbol{R}|$, the electrostatic potential can be decomposed through multipole expansion as follows:

$$\phi_{\mathrm{mol}}(\boldsymbol{R}) = \frac{1}{4\pi\varepsilon}\left(q_{\mathrm{net}} R^{-1} + \boldsymbol{\mu}\cdot\hat{\boldsymbol{R}} R^{-2} + \frac{1}{2} R^{-3} \sum_{i,j} Q_{ij}\hat{R}_i\hat{R}_j \right) + \mathrm{O}(R^{-4}) \quad (S7)$$

where $R=|\boldsymbol{R}|$, $\hat{\boldsymbol{R}}$ is a unit vector in the direction of $\boldsymbol{R}$, and $\hat{R}_i$ and $\hat{R}_j$ are its components. The first term in **Eq. S7** is zero as the molecule is neutral. The second and the third terms represent the contribution of dipole moment and quadrupole moment to the potential, respectively. If the chance to find a specific dipole moment $\boldsymbol{\mu}$ for a molecule in the slab is $P^D(\boldsymbol{\mu})$, the average dipole potential of the molecule is given by:

$$\langle \phi^D_{\mathrm{mol}}(\boldsymbol{R}) \rangle = \frac{1}{4\pi\varepsilon} \int d\boldsymbol{\mu}\, P^D(\boldsymbol{\mu}) \boldsymbol{\mu}\cdot\hat{\boldsymbol{R}} R^{-2} = \frac{\langle \boldsymbol{\mu} \rangle \cdot \hat{\boldsymbol{R}}}{4\pi\varepsilon R^2} \quad (S8)$$

where $\langle \boldsymbol{\mu} \rangle \equiv \langle \mu_x \rangle \boldsymbol{i} + \langle \mu_y \rangle \boldsymbol{j} + \langle \mu_z \rangle \boldsymbol{k}$ and $\langle \mu_\alpha \rangle = \int d\boldsymbol{\mu}\, P^D(\boldsymbol{\mu})\, \mu_\alpha$. In this calculation, the result should not depend on the choice of the coordinate system as long as $|\boldsymbol{r}| \ll |\boldsymbol{R}|$. We thus set the coordinate system as shown in **Scheme S1** to ease the calculation. With this choice of coordinate system and on the basis of the fact that the system is isotropic in the directions perpendicular to the slab normal, it will be found that $\langle \phi^D_{\mathrm{mol}}(\boldsymbol{R}) \rangle = \frac{\cos\theta \langle \mu_z \rangle}{4\pi\varepsilon}$ where θ is the angle between $\boldsymbol{R}$ and the slab normal.

Next, we consider the dipole potential of the molecules in the ring shown in **Scheme S1**. The ring is centered at the projection of $\boldsymbol{R}$ onto the slab, $2l$ in diameter and dl in width. The molecules in the ring should give rise to the same average dipole potential at $\boldsymbol{R}$. Hence, the dipole potential of the entire slab at $\boldsymbol{R}$ is the summation of all the rings varying in radius l, which can be expressed as:

$$\phi^D_{\mathrm{slab}}(L) = 2\pi\rho\delta w \int_0^\infty dl\, l \langle \phi^D_{\mathrm{mol}}(\boldsymbol{R}) \rangle$$

$$= \frac{\rho\delta w}{2\varepsilon} \int_0^\infty dl\, \frac{l\cos\theta\, \langle \mu_z \rangle}{L^2 + l^2} = \frac{\rho L \langle \mu_z \rangle \delta w}{2\varepsilon |L|} \quad (S9)$$

where ρ is the number density of the molecules in the slab. Let L be the difference between the z coordinates of $\boldsymbol{R}$ and the slab. Summation of the dipole potential arising from the slabs at different z positions leads to the expression of the dipole potential at $\boldsymbol{R}$ shown in **Eq. 17**.

Following similar arguments, the quadrupole potential of molecules in the slab at $\boldsymbol{R}$ can be

expressed as:

$$\phi^{Q}_{\text{slab}}(L) = \frac{\rho\delta w}{4\varepsilon}\sum_{i,j}\left(\langle Q_{ij}\rangle \int_0^{\infty} dl\ \frac{l\hat{R}_i\hat{R}_j}{R^3}\right) \quad (S10)$$

Using the same coordinate system that is used for the calculation of dipole potential, one has $\hat{R}_x = \sin\theta$, $\hat{R}_y = 0$ and $\hat{R}_z = \cos\theta$. All the terms involving $\hat{R}_y$ in **Eq. S10** vanish. After integration of the nonzero terms, we can obtain the following expression of quadrupole potential of the slab at $\boldsymbol{R}$ as:

$$\phi^{Q}_{\text{slab}}(L) = \frac{\rho\delta w}{4\varepsilon|L|}\left(\frac{2}{3}\langle Q_{xx}\rangle + \frac{1}{3}\langle Q_{zz}\rangle + \frac{2|L|}{3L}\langle Q_{xz}\rangle\right). \quad (S115)$$

Using conditions $\langle Q_{xx}\rangle = \langle Q_{yy}\rangle$ and $Q_{xx} + Q_{yy} + Q_{zz} = 0$ and summing $\phi^{Q}_{\text{slab}}(L)$ over all slabs eventually led to **Eq. 18**.

Table S1. Average internal energy of water clusters obtained with different clustering methods

	Energy-based	Distance-based	Energy plus surface area
Average internal energy (kJ/mol)	-71.9±0.2	-48.7±0.2	-56.7±0.3

Table S2. Hydration free energy (in a unit of kJ/mol) of MARTINI CG particles in water modeled with the FlexDQ model, the original MARTINI water model and the PMW model.

	FlexDQ (2fs)[a]	FlexDQ (14fs)[a]	Original MARTINI[d]	PMW[d]
P1	-14.2	-14.8	-13.6	-13.5
P2	-14.1	-14.7	-13.6	-13.5
P3	-18.6	-19.5	-18.5	-18.7
P4	-18.6	-19.4	-18.5	-18.7
P5	-23.4	-24.4	-24.7	-24.5
C1	8.0	8.2	11.6	10.6
C3	2.0	1.9	4.7	4
C4	1.9	1.9	4.7	4
Na	-9.7	-10.1	-7.8	-8.3
Nd	-9.5	-10.2	-7.8	-8.3
Nda	-9.6	-10.1	-7.8	-8.3
N0	-4.4	-5.0	-2.8	-3.2
Qa	-209.2[b]	-199.5 (ND, -197.0)[c]	-24.7	-67.5
Qd	-137.6	-139.5 (-139.6, -139.5)	-24.7	-67.5
Qda	-143.3	-145.3 (-145.0, -145.1)	-24.7	-73.8
Q0	-132.8	-134.6 (-134.5, -134.4)	-24.7	-62.8

a) All the results were obtained with the TI method. The results shown in the first and second columns were obtained with a time step of 2 fs and 14 fs, respectively. b) For "Q" types, a correction was added to remove ion-ion self-interaction due to the combined use of the Eward method and the periodic boundary condition.[9] This contribution was calculated as $\frac{q^2}{8\pi}\left(\frac{L}{L'}-\frac{1}{\varepsilon}\right)\frac{\xi}{L}$ where L and L' are the edge length of simulation boxes of fully coupled and decoupled systems, respectively, ε is the dielectric permeability of water and ξ= -2.84. This correction was similar for all the systems tested and equal to ~ -48~-50 kJ/mol. Standard errors: 0.2-0.4 kJ/mol. c) All the results for "Q" types were calculated with $C_{12} = 10^{-7}$. The values in the parenthesis denote the data obtained with $C_{12}=10^{-10}$ (left) and 10^{-6} (right). "ND": the result is unavailable because the simulation was not stable. d) From Ref 6.

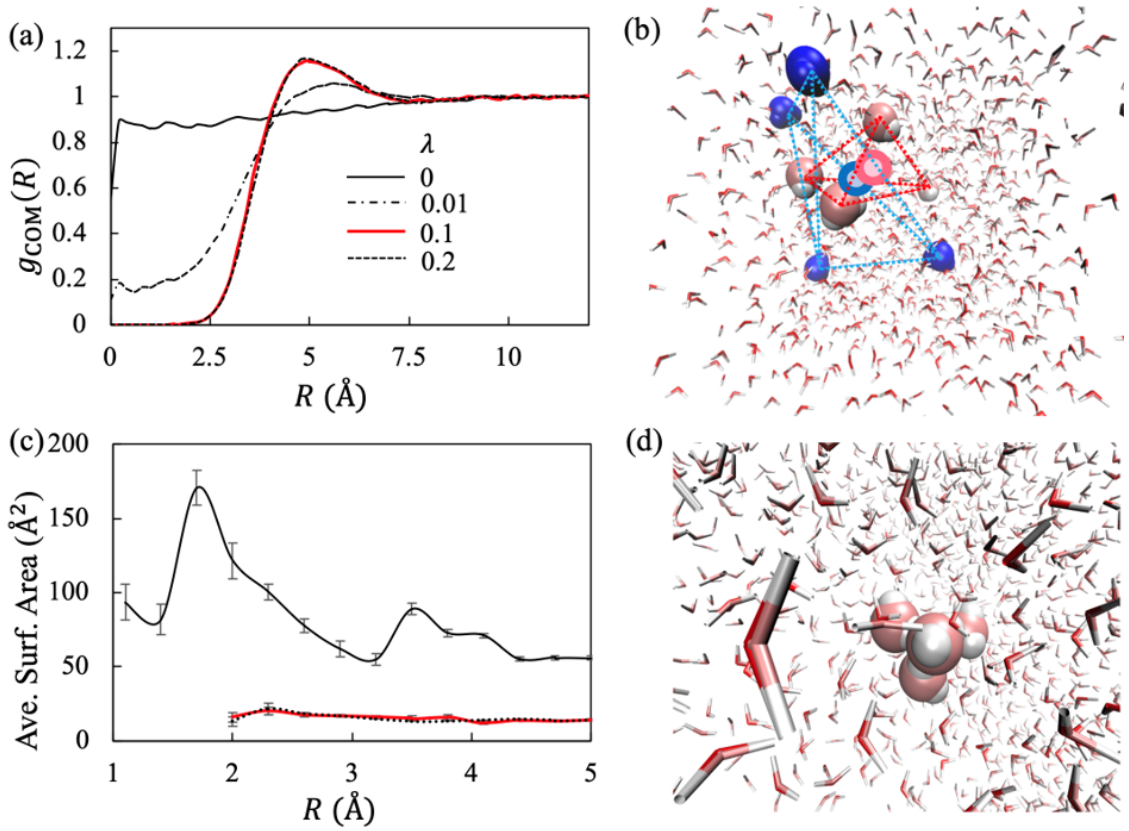


Figure S1. Dependence of clustering results on the choice of λ used in the score function Θ_{par} for clustering. (a) RDFs of COMs of water clusters obtained by clustering water molecules with different λ. (c) Average surface areas of water clusters that are found to be separated at a distance of R. The results obtained with our clustering algorithm using λ =0 or 0.1 are shown as black and red curves, respectively. Those obtained with the geometry-based clustering method are shown as a dotted black curve. The representative grouping of a water cluster with λ =0 or 0.1 is shown in (b) and (d), respectively. The water molecules grouped in the same cluster are shown as spheres of the same color. In (b), dashed lines outline the tetrahedron of clusters and circles denote their COMs.

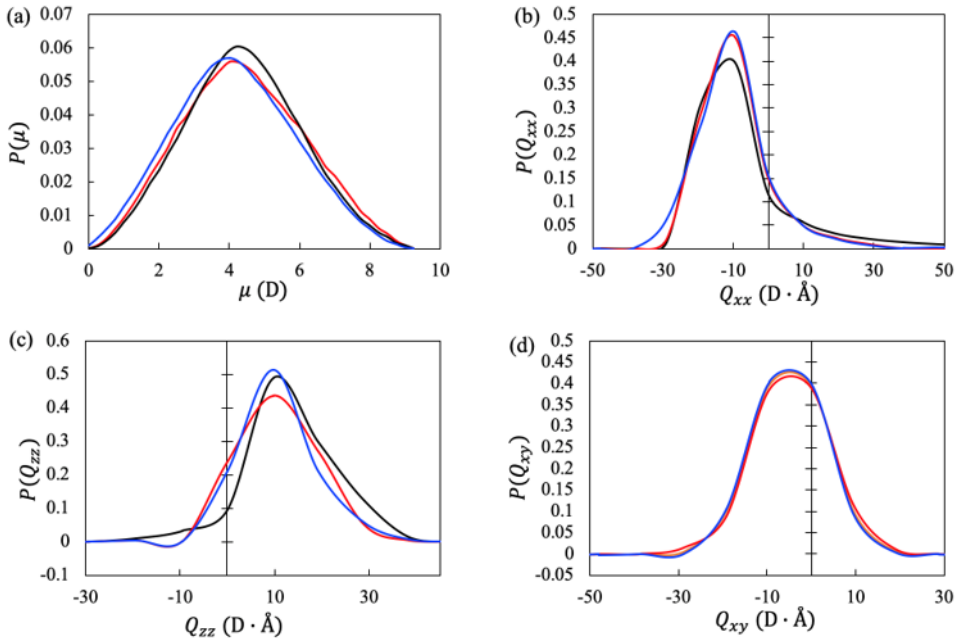


Figure S2. Water clusters derived with different atomic water models share similar distributions of dipole moment (a) and quadrupole moment (b-d). Black, red and blue curves are reference data generated by mapping the results obtained with TIP4P-ew, SPC/E and TIP3P, respectively. In the calculation of quadrupole moment, the center of negative charges was set to the origin. The dipole vector was aligned with *x* axis. The three mapped sites were in the *xy* plane.

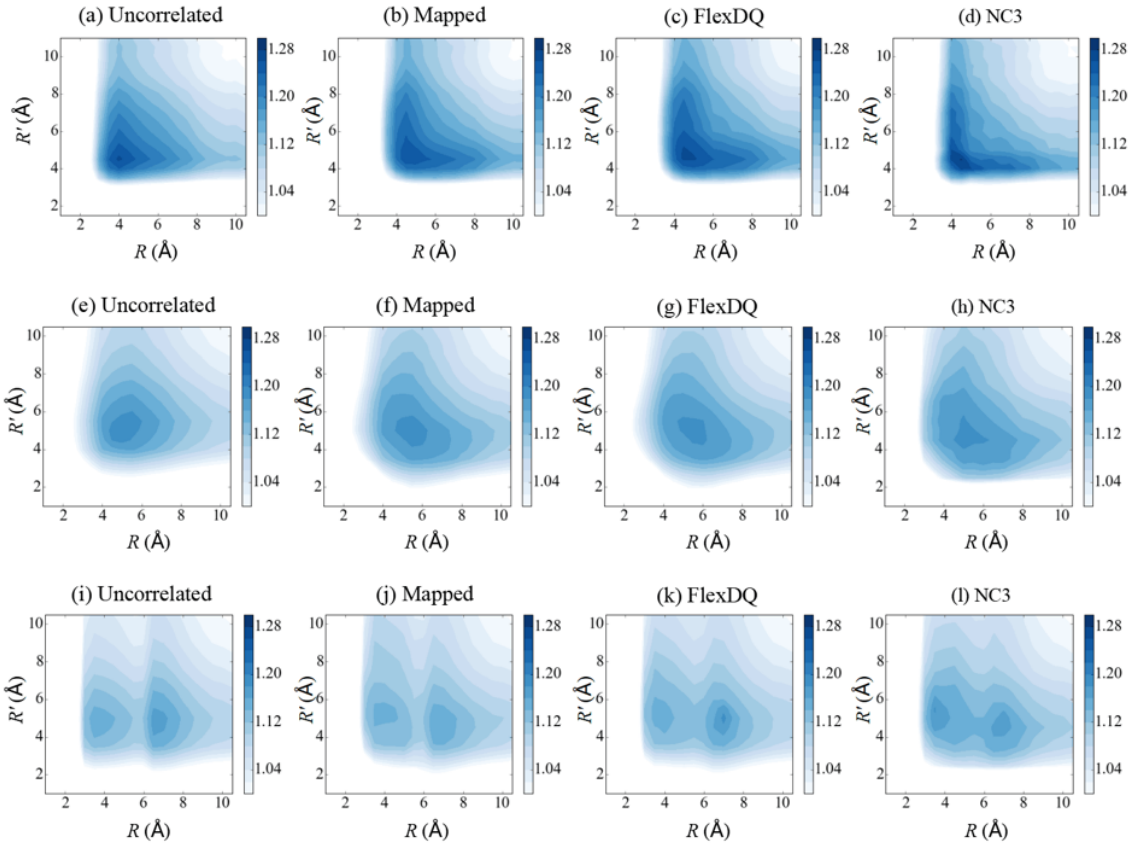


Figure S3. Intensity map of three-body correlation function $g_{\mathrm{wp-wp-wp}}\left(R, R^{'}\right)$ as function of the distances (R and $R^{'}$). Shown from left to right are the plots for hypothetical ensemble without correlation (a), the mapped ensemble (b) and the ensembles obtained with FlexDQ (c) and NC3 (d), respectively. Rows from top to bottom correspond to the triples of wn-wn-wn, wp-wn-wn, and wp-wp-wn respectively. Shown as color bars are the scale of intensity.

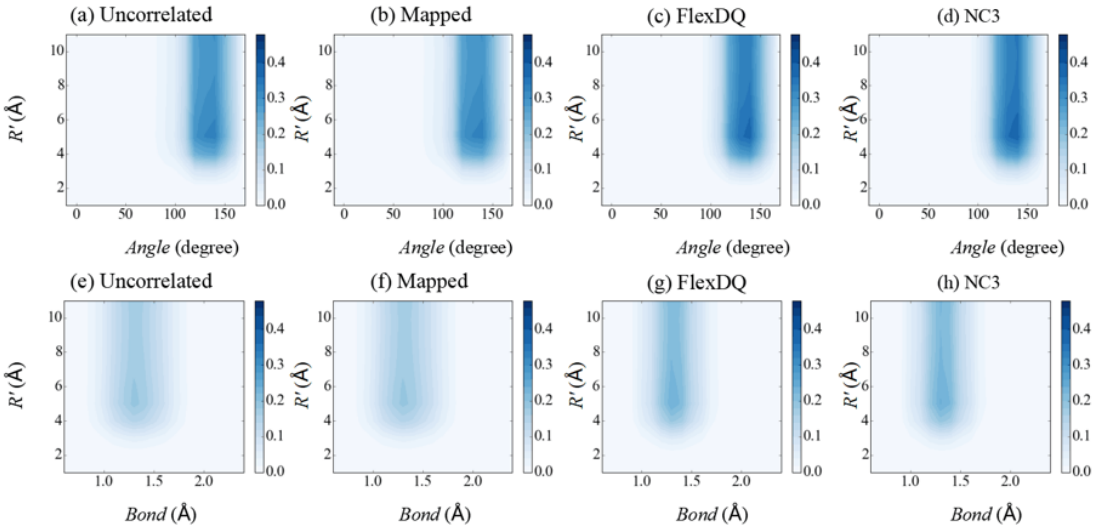


Figure S4. Intensity map of three-body correlation function $g(R_\theta, R')$ (top) and $g(R_b, R')$ (bottom) as function of the intramolecular degree and intermolecular degree (R_b/R_θ and R'). Shown from left to right are the plots for hypothetical ensemble without correlation (a), the mapped ensemble (b) and the ensembles obtained with FlexDQ (c) and NC3 (d), respectively. Shown as color bars are the scale of intensity.

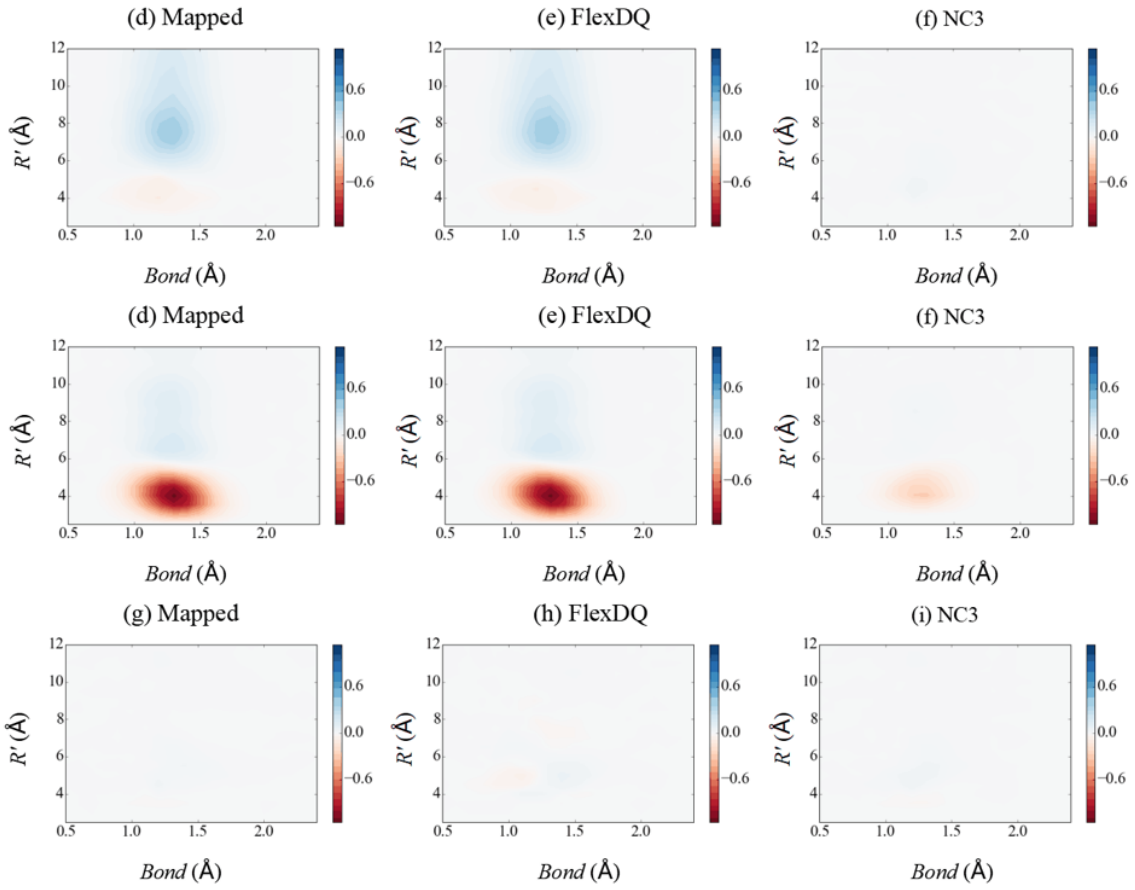


Figure S5. Intensity map of three-body correlation function $G(R_b, R')$ as function of the distances (R_b, R') and their included angle. Shown from top to bottom are the plots for wp-wn-wp, wp-wn-wn and wn-wp-wn, respectively. Shown from left to right are the plots for the mapped ensemble (a, d, g) and the ensembles obtained with FlexDQ (b,e,h) and NC3 (c, f, i), respectively. Shown as color bars are the scale of G function with blue colors for positive values and red colors for negative values.

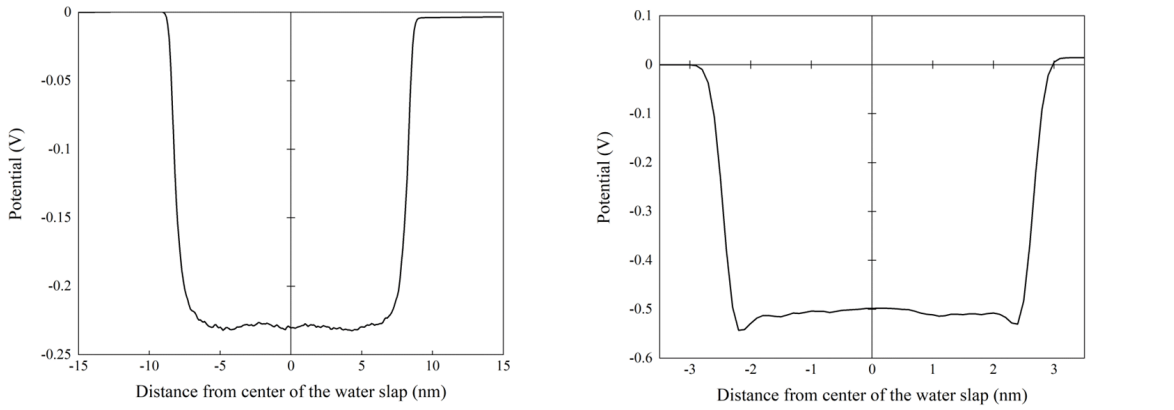


Figure S6. The water-air electrostatic potential profile along the z axis obtained with (a) TIP4P-ew and (b) FlexDQ. The center of the water slab is located at 0 nm.

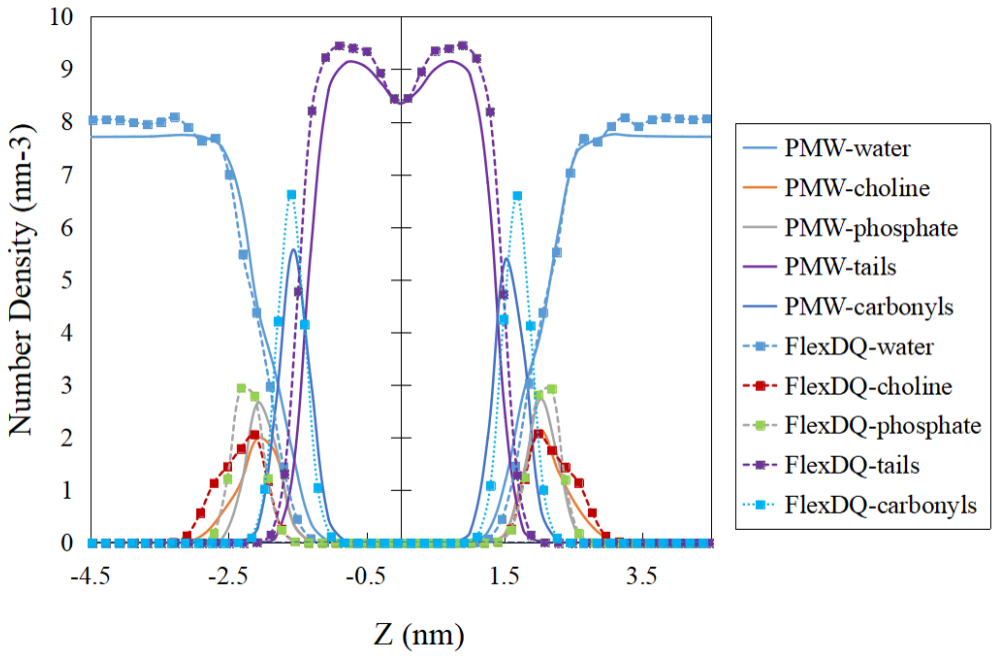


Figure S7. Distribution of the particle density for different CG groups of DPPC bilayer, with respect to the bilayer center (Z=0). Solid line is PMW's result, dash line with square dot is FlexDQ's result. The parameters of FlexDQ with MARTINI membrane is fit to reproduce the hydration free energy.

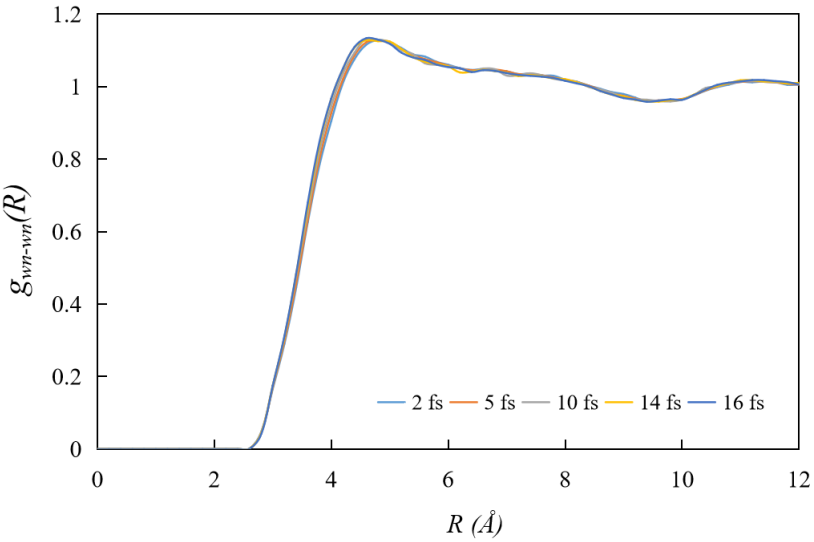


Figure S8. Radial distribution functions of wn sites of CG water particles obtained from simulations at 300 K with a time step varying from 2 fs to 16 fs.

References

1. Wang, H.; Junghans, C.; Kremer, K. Comparative Atomistic and Coarse-Grained Study of Water: What Do We Lose by Coarse-Graining? *Eur. Phys. J. E* **2009,** *28*, 221.
2. Reith, D.; Pütz, M.; Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* **2003,** *24*, 1624.
3. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G. Mark, A. E.; Berendsen. H. J. C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005,** *26*, 1701–1718.
4. van Hoof, B.; Markvoort, A. J.; van Santen, R. A.; Hilbers, P. A. J. The CUMULUS Coarse Graining Method: Transferable Potentials for Water and Solutes. *J. Phys. Chem. B* **2011,** *115*, 10001−10012.
5. Wu, Z.; Cui, Q.; Yethiraj, A. A New Coarse-Grained Model for Water: The Importance of Electrostatic Interactions. *J. Phys. Chem. B* **2010,** *114*, 10524.
6. Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput. Biol*. **2010,** *6*, No. e1000810.
7. Marrink, S. J.; de Vries, A. H.; Mark, A. E. Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **2004,** *108*, 750.
8. Marcus, Y. *Ion Properties*, 1997, Taylor & Francis.
9. Warren, G. L.; Patel, S. Hydration free energies of monovalent ions in transferable intermolecular potential four point fluctuating charge water: An assessment of simulation methodology and force field performance and transferability. *J. Chem. Phys.* **2007,** *127*, 064509.