Supporting Information for:

Assimilating Radial Distribution Functions to Build Water Models with Improved Structural Properties

Alexander D. Wade [a], *Lee-Ping Wang* [b], [*]*David J. Huggins* [a, c]

Affiliations:

[a]University of Cambridge, TCM Group, Cavendish Laboratory, 19 J J Thomson Avenue, Cambridge CB3 0HE, United Kingdom

[b]Department of Chemistry, University of California, Davis, Davis, California 95616, United States

[c]University of Cambridge, Department of Chemistry, Lensfield Road, Cambridge, UK CB2 1EW, United Kingdom

*Current Addresses:

Tri-Institutional Therapeutics Discovery Institute, Belfer Research Building, 413 East 69th Street, 16th Floor, Box 300, New York, USA

Department of Physiology and Biophysics, Weill Cornell Medicine, 1300 York Avenue, New York, NY, 10065, USA

Corresponding Author: David J. Huggins and Lee-Ping Wang

Email: djh210@cam.ac.uk, leeping@ucdavis.edu

Density Curves.

Here we present plots of density curves for the original and optimized parameterizations of the water models considered in this work. Figure (1) shows plots for models with three interaction sites and Figure (2) shows plots for models with four interaction sites.

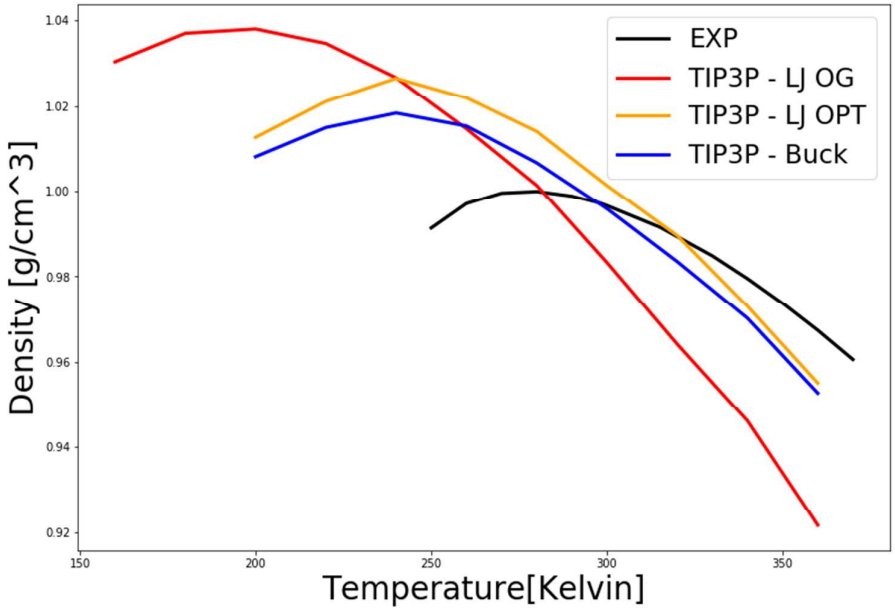

Figure (S1): Plot of density curve for 3P like models. EXP is experimental data taken from Abascal et al [1]. TIP3P is the original TIP3P model [2]. TIP3P – LJ OPT is the model still using the LJ optimized in this work. TIP3P – Buck is the Buckingham type model optimized in this work.

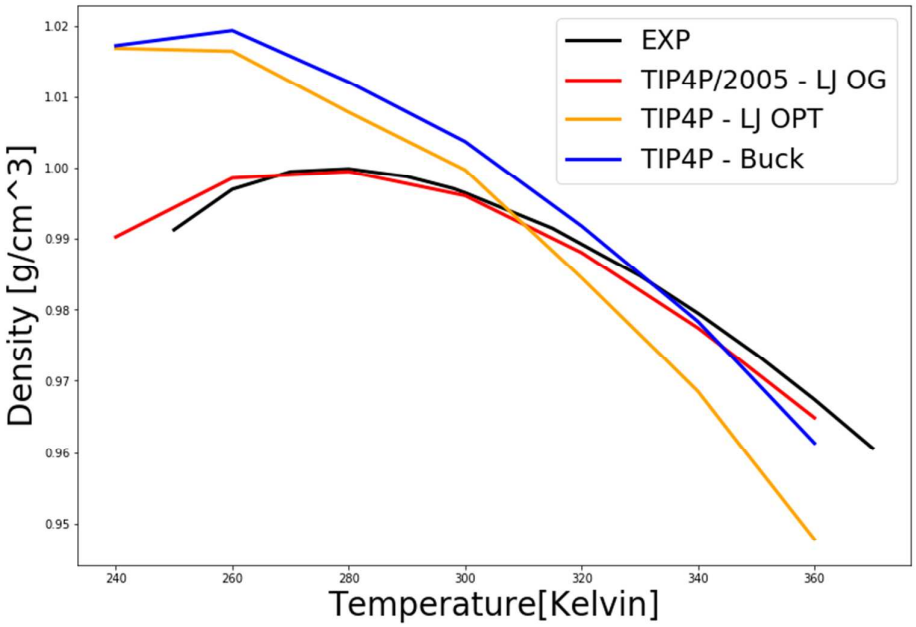

Figure (S2): Plot of density curve for 4P like models. EXP is experimental data taken from Abascal et al [1]. TIP4P/2005 is the original TIP4P/2005 model [1]. TIP4P – LJ OPT is the model still using the LJ optimized in this work. TIP4P – Buck is the Buckingham type model optimized in this work.

Difference plots

Here we present difference plots for the RDF of the optimized models and experimental data.

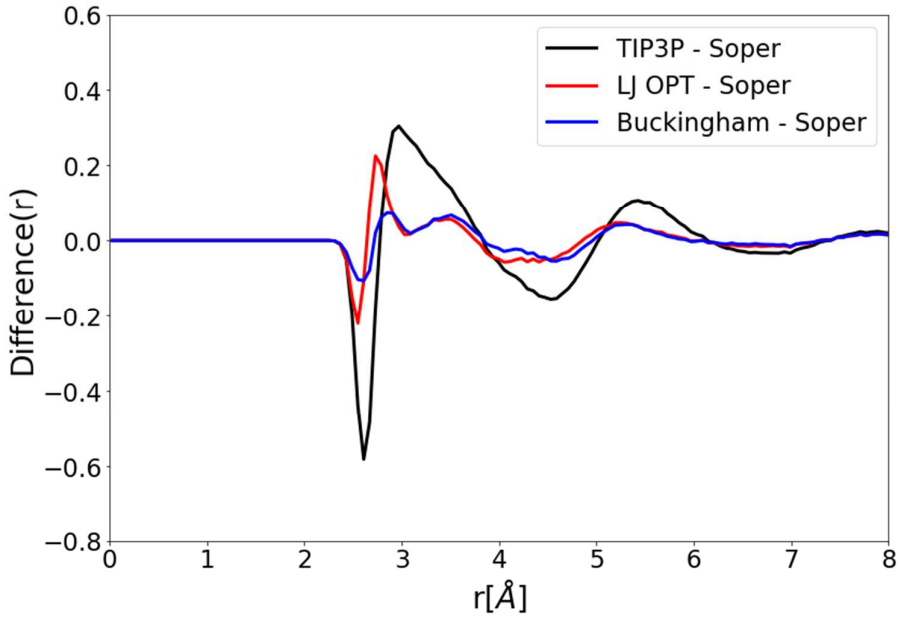

Figure (S3): Plot of difference between calculated and experimental OO RDFs. Experimental data for OO RDF taken from Soper et al [3]. TIP3P is the original TIP3P model [2]. LJ OPT is the model still using the LJ optimized in this work. Buckingham is the Buckingham type model optimized in this work. This graph shows 3P like models.

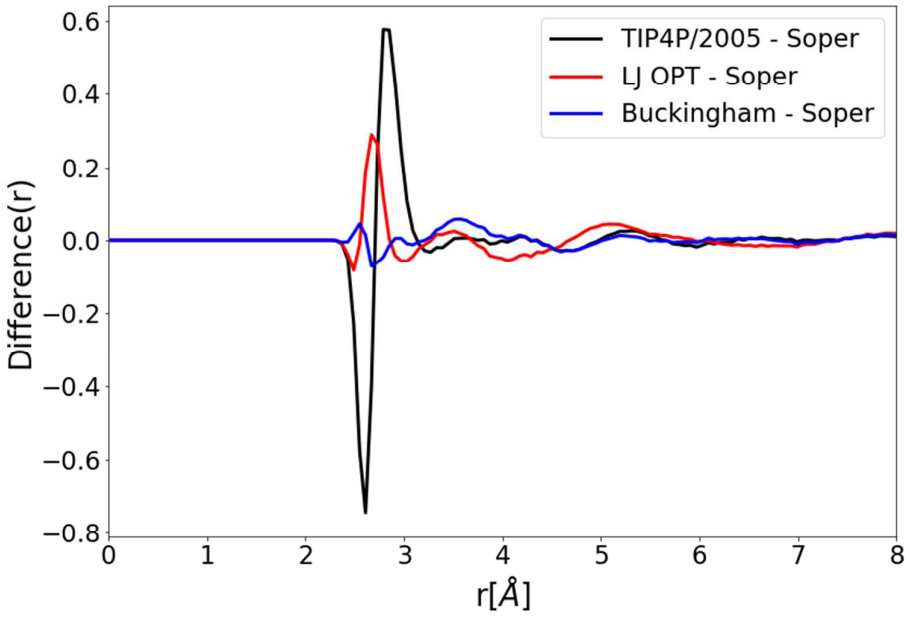

Figure (S4): Plot of difference between calculated and experimental OO RDFs. Experimental data for OO RDF taken from Soper et al [3]. TIP4P is the original TIP3P model [1]. LJ OPT is the model still using the LJ optimized in this work. Buckingham is the Buckingham type model optimized in this work. This graph shows 4P like models.

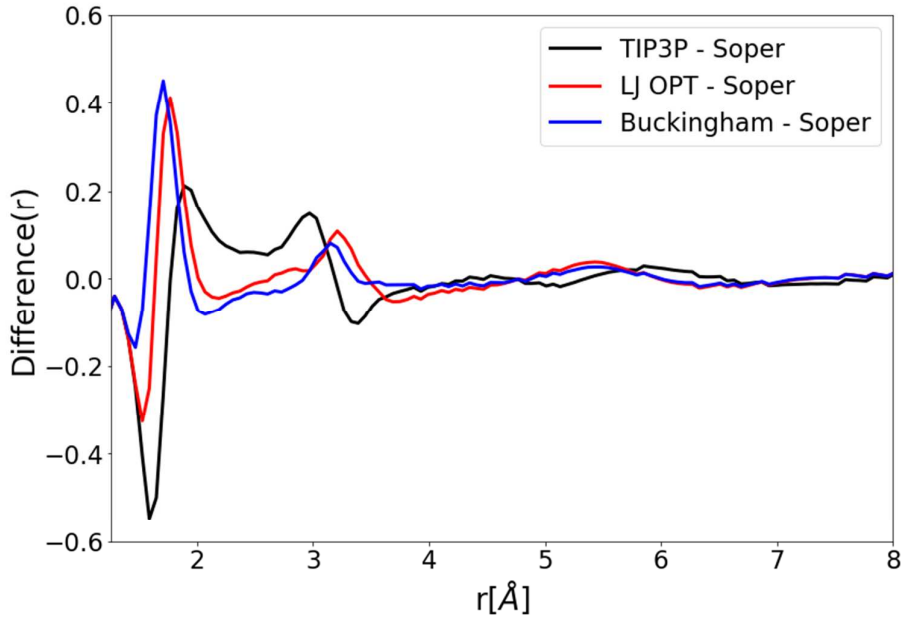

Figure (S5): Plot of difference between calculated and experimental OH RDFs. Experimental data for OH RDF taken from Soper et al [3]. TIP3P is the original TIP3P model [2]. LJ OPT is the model still using the LJ optimized in this work. Buckingham is the Buckingham type model optimized in this work. This graph shows 3P like models.

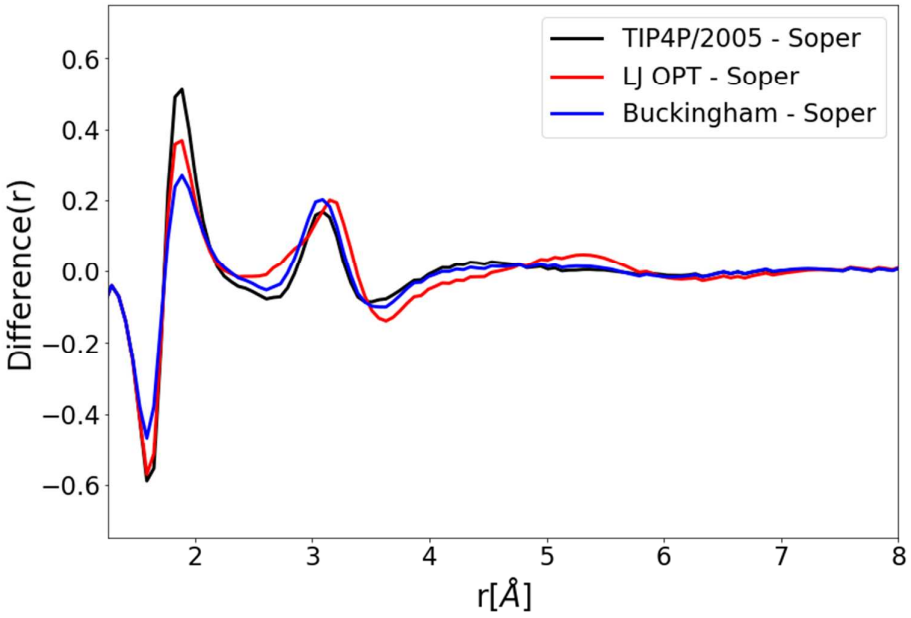

Figure (S6): Plot of difference between calculated and experimental OH RDFs. Experimental data for OH RDF taken from Soper et al [3]. TIP4P is the original TIP3P model [1]. LJ OPT is the model still using the LJ optimized in this work. Buckingham is the Buckingham type model optimized in this work. This graph shows 4P like models.

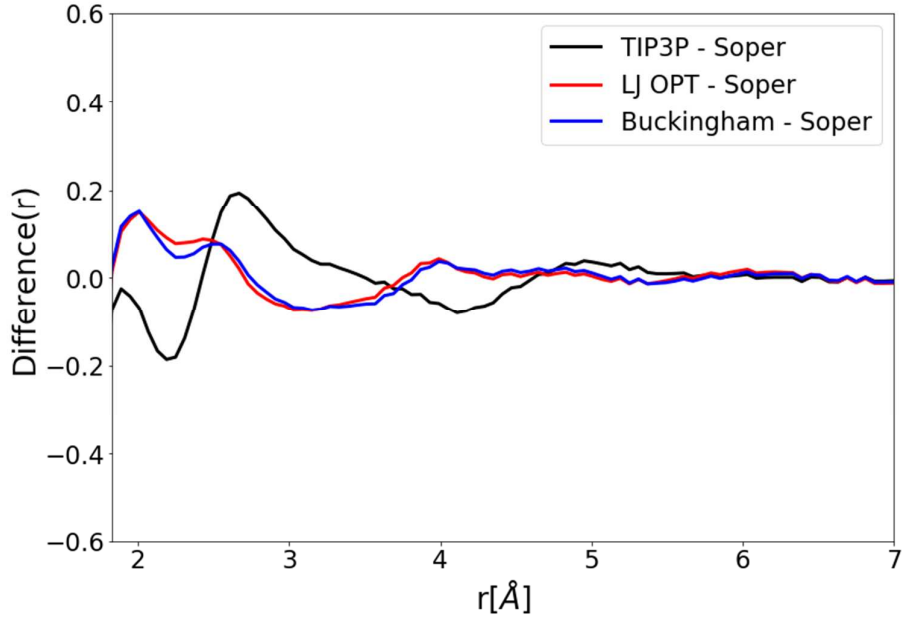

Figure (S7): Plot of difference between calculated and experimental HH RDFs. Experimental data for HH RDF taken from Soper et al [3]. TIP3P is the original TIP3P model [2]. LJ OPT is the model still using the LJ optimized in this work. Buckingham is the Buckingham type model optimized in this work. This graph shows 3P like models.

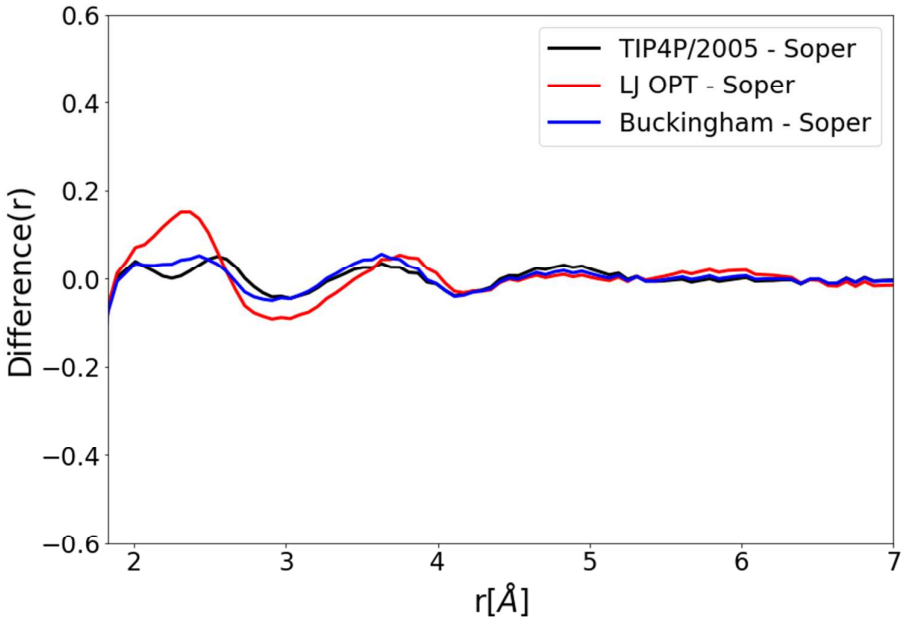

Figure (S8): Plot of difference between calculated and experimental HH RDFs. Experimental data for OO RDF taken from Soper et al [3]. TIP4P is the original TIP3P model [1]. LJ OPT is the model still using the LJ optimized in this work. Buckingham is the Buckingham type model optimized in this work. This graph shows 4P like models.

Free energy calculation

To calculate the free energies in this work the annihilation of a single water in a 25 Å box of water was performed with YANK [4] using Hamiltonian replica exchange [5] for 15 replicas with swapping made between neighbouring replicas every 1 *ps* for a total of 5000 iterations of swapping, giving 5 ns of sampling per replica. These calculations where performed three times for every water model, and an analysis was performed by YANK using the MBAR method [6] to calculate three free energy values which are given a polarization correction. Table (1) shows these unaverage calculations and their polarization corrections for the free energies calculated with YANK.

	3P - LJ OG	3P - LJ OPT	3P - Buck	4P - LJ OG	4P - LJ OPT	4P - Buck
$G[kcal/mol] - 1$	-6.301	-6.732	-7.324	-7.317	-6.640	-7.004
$G[kcal/mol] - 2$	-6.412	-6.68	-7.328	-7.451	-6.548	-7.127
$G[kcal/mol] - 3$	-6.475	-6.698	-7.372	-7.311	-6.605	-7.027
$G[kcal/mol] - AVG$	-6.396	-6.703	-7.341	-7.360	-6.598	-7.053
$E_{pol}[kcal/mol]$	1.184	0.584	1.156	0.975	0.440	0.771
$G_{excess}[kcal/mol]$	-5.212	-6.120	-6.186	-6.385	-6.158	-6.282

Table (S1): G is the free energy calculated by YANK, $G - $ VG is the average of these free energies, E_{pol} is the polarization correction applied and G_{excess} is the finial free energy calculated and used in the main body of this work.

Computational Performance

To assess the computational impact of using the Buckingham potential on computational performance calculations for the speed $[ns/day]$ of the Buckingham and LJ potentials used in three and four site models across several computational platforms were made and presented in Table (4). These calculations were made for a 30 $Å^3$ box of water with CPU calculations performed on a quad-core Intel 6th generation chip (i5-6300HQ) and GPU calculation performed on a Nvidia Maxwell card (Quadro M1000M). The software used for this calculation was CUDA 6.0, OpenCL 1.2 and OpenMM 7.1.1

REFERENCES

(1) Abascal, J. L.; Vega, C. A General Purpose Model for the Condensed Phases of Water: TIP4P/2005. *J. Chem. Phys*. **2005**, 123, 234505.

(2) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** 79, 926–935.

(3) Soper, A. K. The Radial Distribution Functions of Water and Ice from 220 to 673 K and at Pressures up to 400 MPa. *Chem. Phys*. **2000**, 258, 121–137.

(4) Rizzi, A.; Grinaway, P.B.; Parton, D.L.; Shirts, M.R.; Wang, K.; Eastman, P.; Friedrichs, M.; Pande, V.S.; Branson, K.; Mobley, D.L.; Chodera, J.D. YANK: A GPU-Accelerated Platform for Alchemical Free Energy Calculations [Internet]. http://getyank.org

(5) Chodera, J.D.; Shirts, M.R. Replica Exchange and Expanded Ensemble Simulations as Gibbs Sampling: Simple Improvements for Enhanced Mixing. *J. Chem. Phys.* **2011**, 135, 194110

(6) Shirts, M.R.; Chodera, J.D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, 129, 124105