Unveiling Enigmatic Phase Transitions of Water in the Supercooled Region and No man's Land

Saeed Pourasad[1] **and Kwang S. Kim**[1✉]

1 Materials and Methods

1.1 Simulation Model

We performed simulations using the TIP4P/2005 water model,[1] a re-parametrized version of the 'classical' TIP4P model. Widely utilized, this model has demonstrated exceptional reliability and accuracy in capturing various experimental characteristics of water.

1.2 Simulation Methods

We conducted molecular dynamics simulations of water using the LAMMPS package, employing both the isothermal-isochoric (NVT) and the isothermal isobaric (NPT) ensembles. In both cases, we utilized the Verlet integrator with a time step of 1 fs. Electrostatic interactions were computed using the Particle-Particle Particle-Mesh (PPPM) method with a real space cutoff distance of 1.2 nm. Similarly, a 1.2 nm cutoff distance was applied for van der Waals interactions. For both NVT and NPT simulations, we employed the Nose´–Hoover thermostat and barostat. Cubic systems (with periodic boundary conditions) of size 1024 and cubic systems (with periodic boundary conditions) of sizes 512, 768, and 1024 were studied in NVT and NPT Simulations, respectively.

1.3 Generalized Replica Exchange Method

The generalized replica exchange method (gREM) is tailored for simulating systems undergoing first-order phase transitions. In a NVT simulation, the probability distribution of the potential energy, $P(E)$ is defined as $P(E) = W(E)\ \Omega(E)/Z$, where $W(E) \equiv exp(-w(E))$ is the sampling weight, $\Omega(E)$ is the density of states (linked to entropy, $S(E)$, through $\Omega(U) = exp(-S(E)/k_B)$), and $Z = \int dU W(E)\ \Omega(E)$ represents the generalized configurational integral.

The generalized ensemble (g) introduces a generalized temperature function, $dw(E)/dE = 1/(k_B T^g(E))$ with $w = \beta E$, $\beta = 1/(k_B T)$, and T^g is the usual constant temperature T. Computing the extrema, E^* of $P(E)$ yields $T^g(E^*) = T_S(E^*)$, where $T_S(E) = 1/(dS(E)/dE)$ is the statistical temperature. At a first-order phase transition, the statistical temperature exhibits an S-shaped loop, attributed to the "convex intruder" in entropy stemming from the reduced density of states in the phase coexistence range due to surface effects.

In the vicinity of the transition region, within a constant temperature ensemble, $T_S(E^*)$ has three solutions corresponding to two pure phases at E_1 and E_2, and the barrier at E_{bar}. These points align with two minima and

[1]Center for Superfunctional Materials, Department of Chemistry, Ulsan National Institute of Technology and Science (UNIST), Ulsan 44919, Republic of Korea. ✉ email: kimks@unist.ac.kr

one maximum in the Gibbs free-energy profile at the specified temperature. States around the barrier are seldom visited, with the system predominantly residing in one of the free-energy minima (E_1 or E_2).

The gREM conceptually involves employing an ensemble with a temperature function that intersects T_S only once. This approach facilitates the sequential exploration of the unstable energy region between two stable states. This generalized temperature function as the linear effective temperature is defined as follows:

$$T_\alpha(E) = \lambda_\alpha + \gamma_0(E - E_0) \tag{1}$$

where the slope, γ_0, should be sufficiently negative to ensure a single intersection with T_S despite the S loop. By varying the intercept λ_α, the intersection point with T_S can be systematically adjusted through the coexistence range, enabling thorough sampling with a series of unimodal, overlapping energy distributions. Replica exchanges are periodically attempted, with acceptance probabilities chosen to preserve detailed balance. Thermodynamic quantities can be computed using the statistical temperature, accessible from energy histograms in the replicas through the statistical temperature-weighted histogram analysis method (ST-WHAM).

In this particular case, 16 generalized replicas (ensembles) of water molecules (N=1024) at a density of $\rho = 1.0$ g/cm^3 were utilized. A swap between adjacent ensembles was attempted every 1000 time steps. Each production run spanned 60-80 ns, during which the potentials of the configurations were sampled every 10 time steps for subsequent analysis.

1.4 Umbrella Sampling Method

This method[2] allows us to control a specified collective variable using a potential well set to a predetermined target value. Subsequently, multiple simulations are conducted with varying target values, ensuring that the statistical data from adjacent "umbrellas" overlap. The term "umbrella sampling" derives from the bias potential's role in energetically linking distinct regions in phase space. As such, umbrella sampling, an enhanced sampling technique, facilitates the exploration of state space regions that might otherwise lack sufficient sampling. In this scenario, thanks to the known bias potential, unbiased potential statistics can be reconstructed from the biased statistics obtained in the simulations.

To implement this method, the PLUMED package[3] is integrated with the LAMMPS package. NPT simulations on water molecules (N=512) are conducted across a wide range of temperatures and pressures. The system's volume serves as the collective variable, with the number of windows (grids) used varying from 30 at high temperatures to 65 at low temperatures. For larger systems (N=1024) at T=185K, more than 100 windows were employed. Prior to the production run, each window in these simulations undergoes equilibration at the target volume for over 80 ns.

The constant *k* in the harmonic restraint on the collective variable remains consistent across all calculations (s: collective variables, s_0= collective variable reference value, $V(s)$: harmonic restraint).

$$V(s) = k/2(s - s_0)^2 \tag{2}$$

To obtain the free energy curves and error bars, we utilize the weighted histogram analysis method (WHAM)[4] and the bootstrap error analysis method.

1.5 Bootstrap Analysis

Bootstrap analysis[5] is a statistical method utilized to estimate the sampling distribution and ascertain its level of uncertainty. This technique generates numerous resamples by randomly selecting observations from the initial sample, with replacement. Consequently, an observation may be chosen multiple times within the same bootstrap

sample, while others may not be selected at all. Each bootstrap sample maintains the same size as the original sample, meaning that if there are "n" observations in the original data set, each bootstrap sample also encompasses "n" observations. We then calculate the average of these re-sampled data sets, repeat this process numerous times, and subsequently determine the standard deviation of the averages obtained. This provides us with an estimate of the statistical uncertainty associated with the average computed using actual data.

1.6 Radial Distribution Function

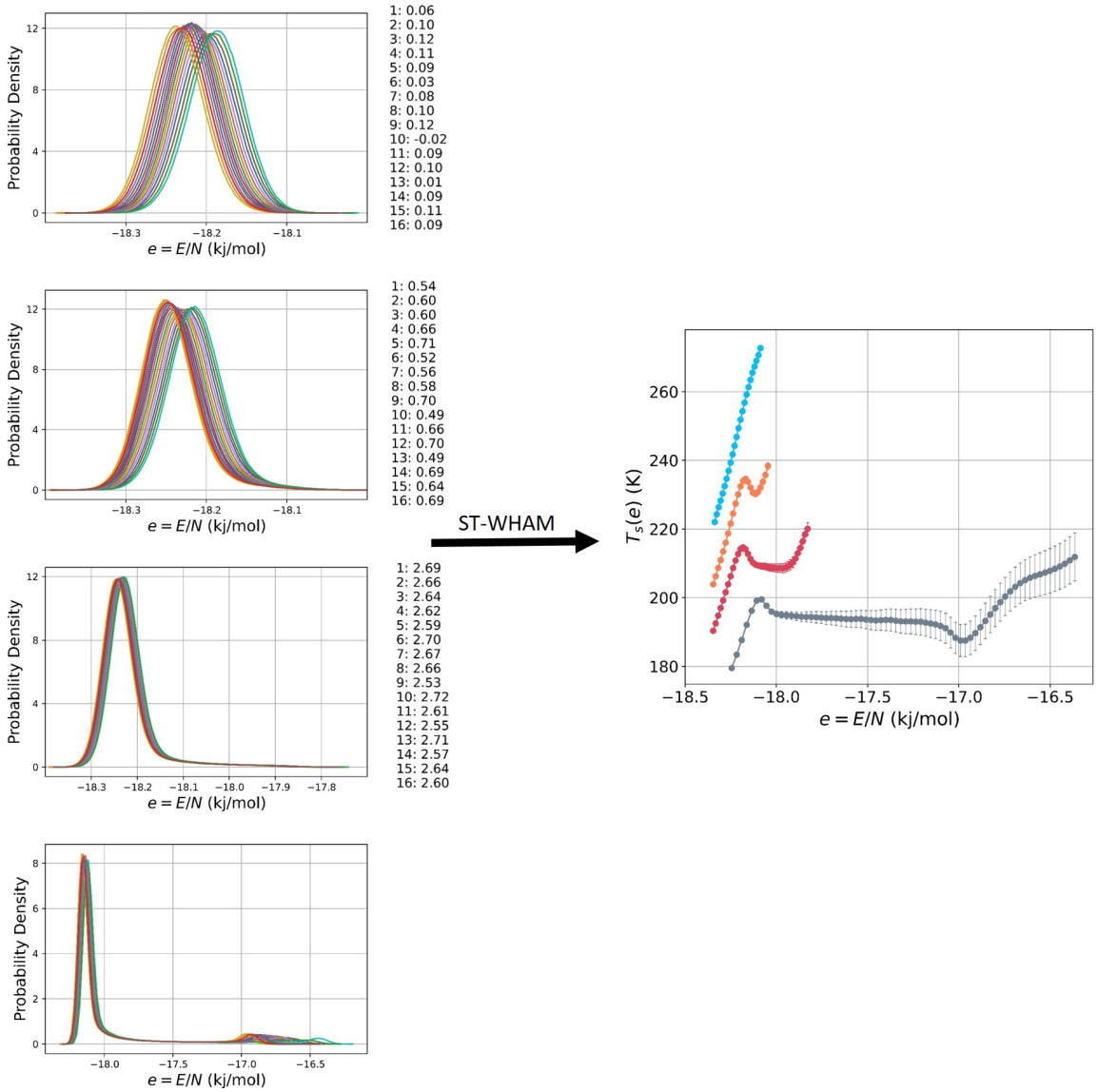
The structural distinctions among the two or three states are illustrated in the Extended Data Figs.1 and 2 through the radial distribution function.

1.7 Orientational Tetrahedral Order parameter

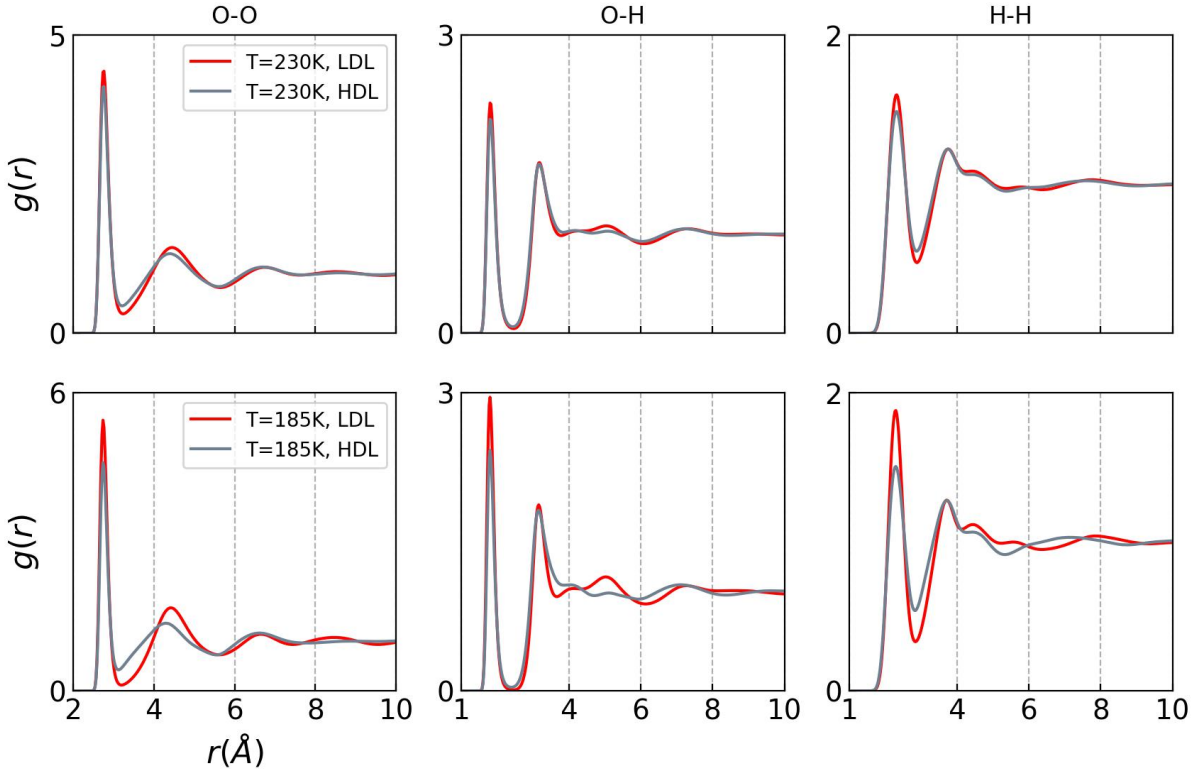
The orientational tetrahedral order parameter,[6] denoted as q, is defined by the equation:

$$q = 1 - \frac{3}{8}\sum_{j=1}^{3}\sum_{k=j+1}^{4}(\cos\varphi_{jk} + \frac{1}{3})^2$$

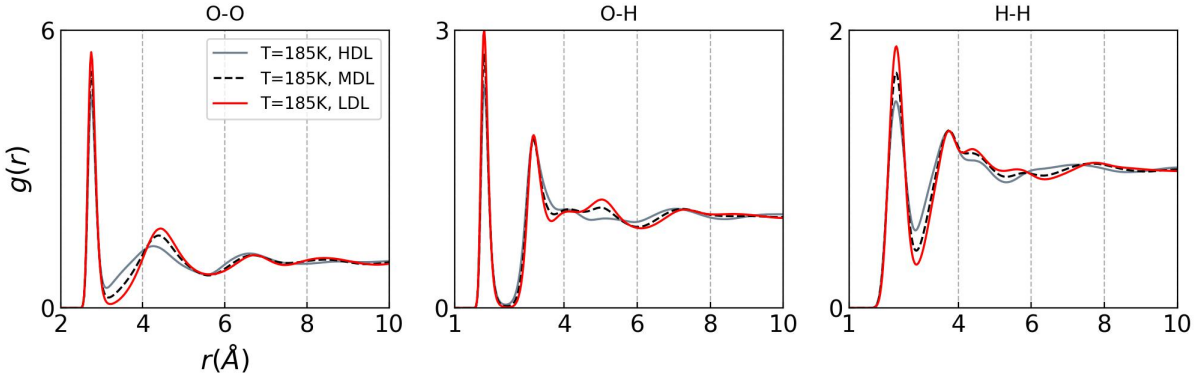
here, φ_{jk} represents the angle formed by the lines connecting the oxygen atom of the water molecule in question with its closest neighboring oxygen atoms j and k. The value of q ranges between 0 to 1, where 0 signifies an ideal gas and 1 corresponds to a regular tetrahedron. Variation in the q value unveils the presence of two or three distinct structures exhibited by water in the supercooled region.



Supplementary Fig. 1 | Total energy histograms of 16 replicas with different temperature ranges T, T+2,..., T+30K. The uppermost graph corresponds to the higher temperature. The skewness of each replica at each graph is presented in the column next to the graphs.



Supplementary Fig. 2 | Radial distribution functions for high and low-density states near the second critical point (T =230K) and far from it (T =185K).



Supplementary Fig. 3 | Radial distribution functions of three different states (HDL, MDL, LDL) at constant temperature.

References

[1] Abascal, J. L. F. & Vega, C. A general purpose model for the condensed phases of water: Tip4p/2005. *J. Chem. Phys.* **123**, 234505 (2005).

[2] Torrie, G. & Valleau, J. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **23**, 187–199 (1977).

[3] Bonomi, M. *et al.* Plumed: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications* **180**, 1961–1972 (2009).

[4] Grossfield, A. Wham: the weighted histogram analysis method (2012).

[5] Hub, J. S., de Groot, B. L. & van der Spoel, D. g_wham—a free weighted histogram analysis implementation including robust error and autocorrelation estimates. *Journal of Chemical Theory and Computation* **6**, 3713–3720 (2010).

[6] Errington, J. R. & Debenedetti, P. G. Relationship between structural order and the anomalies of liquid water. *Nature* **409**, 318–321 (2001).