Supplementary Information for
"Revealing the three-component structure of water with principal component analysis on X-ray spectrum"

Zhipeng Jin[1], Jiangtao Zhao[2], Gang Chen[3], Guo Chen[4*], Zhenlin Luo[2*] and Lei Xu[1*]

[1] *Department of Physics, The Chinese University of Hong Kong, Hong Kong, China*

[2] *National Synchrotron Radiation Laboratory, University of Science and Technology of China, Hefei 230026, China*

[3] *School of Physical Science and Technology, ShanghaiTech University, Shanghai 201210, China*

[4] *Chongqing Key Laboratory of Soft Condensed Matter Physics and Smart Materials, College of Physics, Chongqing University, Chongqing 400044, China*

1 Experimental details

To explore bulk water's structure at different temperatures, we designed a transmission XRD experiment. The setup at BL19U2 station of Shanghai Synchrotron Radiation Facility (SSRF) is shown in Fig. S1(a). The deionized water was sealed in a copper ring (inner radius 8 mm, thickness 1 mm), and two thin Kapton films were used as the windows to make the X-ray get through. A small hole is drilled in the side wall of the copper ring for water injection and temperature detection. The setup (i.e., water inside copper ring) was attached to a thermal stage (Linkam, HFSX350) which can adjust the temperature between -196 °C to 350 °C. The thermal stage can also provide nitrogen atmosphere to prevent the Kaptpom film window from fogging when the sample is cooled to low temperatures. A thermistor (TE connectivity, Micro-BetaCHIP) was inserted into the water to detect the real-time temperature. The distance between the thermistor and the X-ray area is less than 0.5 mm to make sure that the measured temperature is accurate. An area detector (Pilatus 1M, 169*179 mm^2) is used to collect the scattered X-ray. The distance between the detector and the container is calibrated by the silver behenate powder, which is 234.6 mm. The X-ray beam has the spot size 320*43 μm^2 and energy 12.000 ± 0.002 keV, and the q range collected by the detector is 0.3∼4.5 Å^{-1}. As shown in Fig. S1(b): besides the DI water in the container, the X-ray beam also goes through other materials in other regions, such as air and Kapton film. Therefore, the detector also collects X-ray scattered by these materials and a careful data correction is required for the diffraction raw images. The correction details will be discussed in the next section.

Compared with the temperature set by the thermal stage, the temperature measured by the
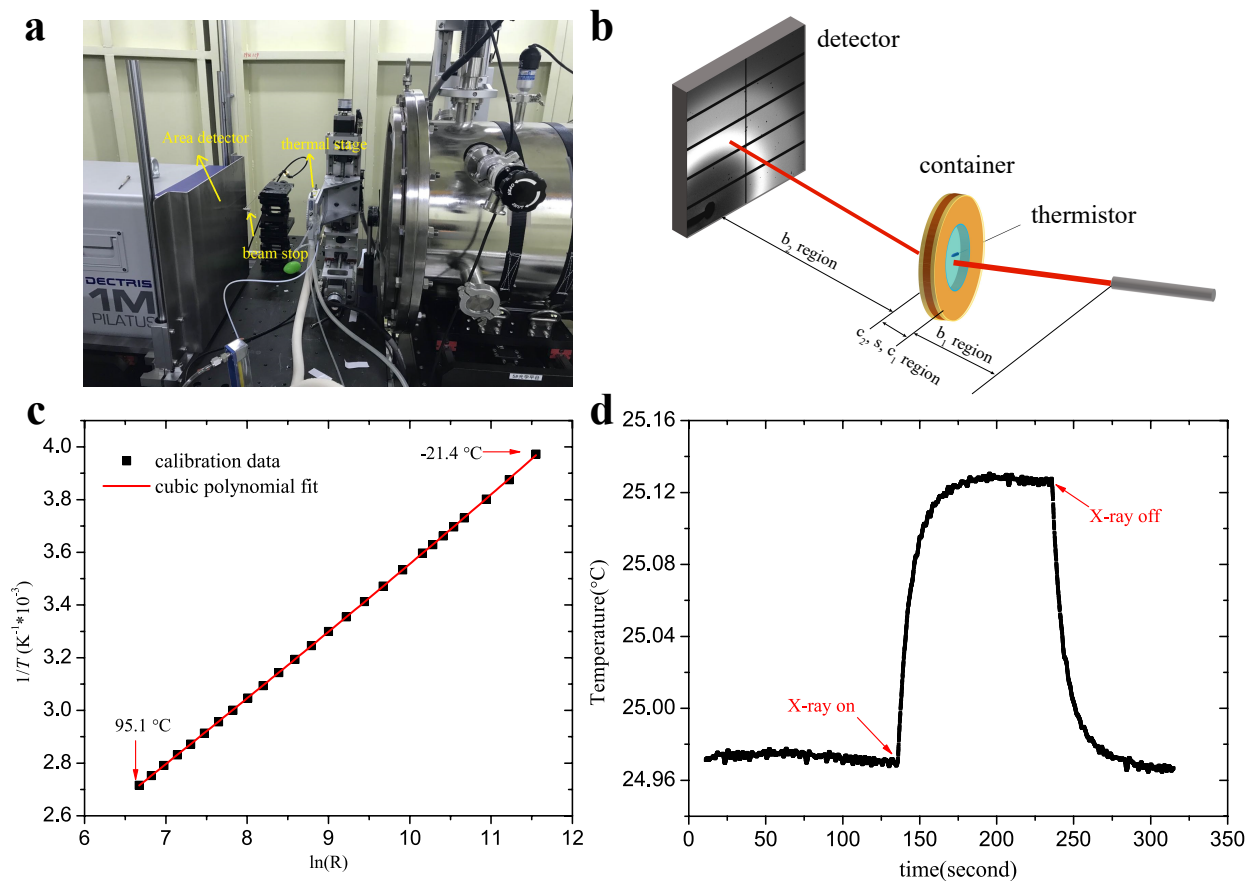
Figure S1: (a) The experimental setup at SSRF. (b) The simplified schematics of experimental setup. The sample was sealed in a copper ring container with a thermistor detecting the real-time temperature. The area detector can collect scattered X-ray from different regions. (c) The calibration curve between $1/T$ and $\ln(R)$ for the temperature probe (thermistor). The curve is excellently fitted by a cubic polynomial between -20 °C to 95 °C. (d) The real-time temperature of the sample. The laser heating effect is clearly observed indicating the sensitivity of the thermistor.

thermistor inserted into the water is more accurate. The thermistor was calibrated in advance in the circulating bath system (PolyScience, PD15R-40) between -20 °C to 95 °C. The relationship between $1/T$ and $\ln(R)$ (R is thermistor's resistance) can be fitted with a cubic polynomial (show in Fig. S1(c)). We used a mutimeter (Keithley, model 2700) to measure the real-time resistance of the thermistor with a frequency up to 30 Hz, and the temperature can be obtained with the calibration curve Fig. S1(c). A typical temperature measurement curve is shown in Fig. S1(d): there is a slight heating effect below 0.2 °C after the X-ray is turned on. Apparently, we can monitor the sample temperature in real time with a high accuracy below 0.1 °C.

2 Experiment data correction

When water was measured in the copper-ring container, as shown in Fig. S1(b), the scattering intensity collected can be represented as the following (multiple scattering was neglected):

$$
I_{scb}(\theta) = I_{b1}(\theta) + I_{c1}(\theta) \cdot att_c(\theta, \mu_c, t_{c1}) \cdot e^{-(\mu_s t_s + \mu_c t_{c2} + \mu_b t_{b2})/cos\theta}
$$

$$
+ I_s(\theta) \cdot T_{c1} \cdot att_s(\theta, \mu_s, t_s) \cdot e^{-(\mu_c t_{c2} + \mu_b t_{b2})/cos\theta} + I_{c2}(\theta) \cdot T_{c1} \cdot T_s \cdot att_c(\theta, \mu_c, t_{c2}) \cdot e^{-\mu_b t_{b2}/cos\theta}
$$

$$
+ I_{b2}(\theta) \cdot T_{c1} \cdot T_s \cdot T_{c2} \cdot att_b(\theta, \mu_b, t_{b2})
$$

$$
\tag{1}
$$

where the subscripts s, c and b represent contributions from sample (i.e., water), container windows (i.e., Kapton film) and background air respectively. 1 and 2 represent different regions along the light path (see Fig. S1(b)). μ is the attenuation coefficient. t is the thickness of the corresponding material. θ is the scattering angle. $att(\theta, \mu, t)$ is the angle dependence of the attenuation when the scattering X-ray goes through the corresponding material. $att(\theta, \mu, t)$ in general also depends on

the incident X-ray energy, but in our experiment the energy is fixed. T is the $\theta = 0$ transmission factor, typically represented as $e^{-\mu t}$. The final exponential part in relevant terms represents the absorption by the following up materials after the current scattering material: the scattered light needs to go through these following up materials and gets absorbed before reaching the detector. When there was no water in the container and the container space (labelled by $b3$) was filled with air, the corresponding background scattering intensity is:

$$I_{cb}(\theta) = I_{b1}(\theta) + I_{c1}(\theta) \cdot att_c(\theta, \mu_c, t_{c1}) \cdot e^{-(\mu_b t_s + \mu_c t_{c2} + \mu_b t_{b2})/cos\theta}$$

$$+I_{b3}(\theta) \cdot T_{c1} \cdot att_b(\theta, \mu_b, t_s) \cdot e^{-(\mu_c t_{c2} + \mu_b t_{b2})/cos\theta} + I_{c2}(\theta) \cdot T_{c1} \cdot T_{b3} \cdot att_c(\theta, \mu_c, t_{c2}) \cdot e^{-\mu_b t_{b2}/cos\theta}$$

$$+I_{b2}(\theta) \cdot T_{c1} \cdot T_{b3} \cdot T_{c2} \cdot att_b(\theta, \mu_b, t_{b2})$$

$$(2)$$

In our experiment, the I_{b1} term was prevented from entering the detector and thus can be neglected. The attenuation coefficient for sample water, container windows (Kapton film) and background air at 20 °C are 2.9079 cm^{-1}, 2.454 cm^{-1} and 3.349*10^{-3} cm^{-1} respectively. Their thicknesses are 0.1 cm, 0.005 cm and 23.46 cm respectively. We can find that $\mu_s t_s \gg \mu_c t_c, \mu_s t_s \gg \mu_b t_b$, after combining Eq. (1) and Eq. (2), the intensity contributed by the sample can be calculated as [45]:

$$I_s(\theta) \approx \frac{I_{scb}(\theta) - T_s \cdot I_{cb}(\theta)}{att_s(\theta, \mu_s, t_s)} \qquad (3)$$

Because all the materials the X-ray goes through are slab shaped with uniform thickness, the attenuation for this geometry is:

$$att(\theta, \mu, t) = \frac{\exp(\frac{-\mu t}{\sin(\theta + \alpha)}) \left[\exp(\frac{\mu t}{\sin(\theta + \alpha)} - \frac{\mu t}{\sin \alpha}) - 1\right]}{\frac{\mu t}{\sin(\theta + \alpha)} - \frac{\mu t}{\sin \alpha}} \qquad (4)$$

where α is the angle between the incident light and the slab surface, which is $\pi/2$ in our experiment, and thus Eq. (4) simplifies to:

$$att(\theta, \mu, t) = \frac{\exp(-\mu t) - \exp(-\mu t/\cos\theta)}{\mu t(1/cos\theta - 1)} \tag{5}$$

Because of the geometry of the area detector and the polarization of the X-ray beam, the geometric and polarization [46] corrections are required:

$$I_s^c(\theta) = \frac{I_s(\theta)geo(\theta)}{pol(\theta)}$$

$$geo(\theta) = cos^3(\theta) \tag{6}$$

$$pol(\theta) = \tfrac{1}{2}(1 + \cos^2\theta - f \cos\varphi \sin^2\theta)$$

where θ is the scattering angle, φ is the azimuthal angle, and f is the polarization factor of the X-ray beam, which is 0.99 in our experiment. The geometric and polarization corrections were performed by the software Fit2D [47], which was used to integrate the X-ray diffraction images into $I(q)$ curves.

However, Eq. (3) cannot be directly used to subtract the background, because the incident light intensity is always fluctuating with time and a correction factor based on the actual intensity is required. To obtain this correction factor, we notice that there is a sharp peak around $q = 0.4$ Å^{-1} in the intensity curve I_{scb} (see Fig. S2(a)), which comes from the Kapton film instead of liquid water. Therefore, we multiply a correction factor T_s to make sure that after background subtraction, the true intensity curve around $q = 0.4$ Å^{-1} is completely flat without any peak feature [48]. The profiles of the intensity curves before and after corrections are presented in Fig. S2(b).

The fully corrected intensity can be normalized by the number of water molecules, the inci-
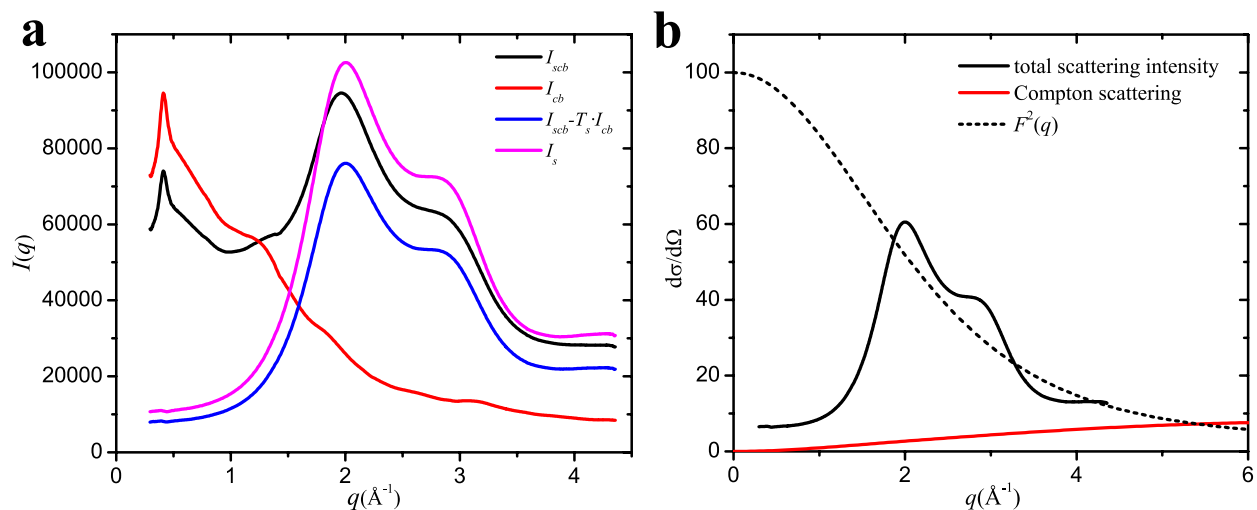
6

Figure S2: (a) The integrated intensity before and after data correction: the original raw data without corrections (I_{scb}, black), intensity from the background air and sample container (I_{cb}, red), intensity after subtracting the background ($I_{scb} - T_s \cdot I_{cb}$, blue), the final intensity after slab absorption correction (I_s, pink). (b) The necessary data needed to calculate the molecular structure factor $S(q)$. The fully corrected scattering intensity at 25 °C (black solid). The square of water molecular form factor calculated by quantum mechanics [51] (black dash). Compton scattering in theory (red).

dent beam flux and the exposure time. In practice the KroghMoe [49] and Norman [50] methods are generally used because of convenience. The scale factor α is defined as:

$$\alpha = \frac{-2\pi^2 z^2 \rho + \int_0^{q_{max}} \left[F^2(q) + I_{incoherent}(q) \right] q^2 \mathrm{d}q}{\int_0^{q_{max}} I_{experiment}(q) q^2 \mathrm{d}q} \tag{7}$$

where ρ is the number density of molecules (in molecules/Å^3), z is the number of electrons per water molecule, $F^2(q)$ is the square of water molecule's form factor, q_{max} is the maximum wavenumber we can measure in the experiment, $I_{incoherent}(q)$ is the incoherent or Compton scattering from the sample, and $I_{experiment}(q)$ is the fully corrected intensity measured from our experiment. Both $F^2(q)$ and $I_{incoherent}(q)$ can be calculated theoretically by quantum mechanics [51]. In our experiment, the minimum q is 0.3 Å^{-1} because of the beam stop, and thus the lower limit of the integration is 0.3 Å^{-1} instead of zero. After normalized by α, the intensity is re-scaled onto a universal scale in the electron units, and then the Compton scattering [51] is subtracted. One typical example of the normalized intensity after subtracting Compton scattering is illustrated in Fig. S2(b) by the pink I_s curve. Using $S(q) - 1 = [I^{nor}(q) - F^2(q)]/[2f_H(q) + f_O(q)]^2$, the molecular structure factor $S(q)$ - 1 at different temperatures can be calculated [52]. Following the work in the previous research [45], we start from the independent atom approximation (IAA) form factors of hydrogen and oxygen atoms using the 6-Gaussian fitting functions [53]. After considering the electron density change caused by the chemical bonding, the modified atomic form factors (MAFF), $f_\alpha(q) = f_\alpha^0(q) \left[1 - (a_\alpha/z_\alpha) \cdot \exp(-q^2/2\delta^2) \right]$, are then calculated, where $f_\alpha^0(q)$ is the IAA form factor, z_α is the atomic number for atom α, and a_α represents the electron transfer, which is -1 for oxygen atom and +0.5 for hydrogen atom according to the previous literature [52]. The value of δ is 2.0 Å^{-1} for both oxygen and hydrogen atoms.

3 PCA analysis

The PCA analysis is done on the structure factor curves $S(q)$-1 at different temperatures. We discretize every $S(q)$-1 curve into 915 discrete values with equal partition in our experimental range between 0.3 to 4.5 Å^{-1}. Thus each $S(q)$-1 curve at one specific temperature is converted into a 1(row)*915(column) vector. We measured 18 different temperatures between -5 °C to 80 °C in our experiment, and all the 18 curves are converted to a 18*915 data matrix. Then this data matrix's covariance matrix (915*915 in dimension) is constructed and the covariance matrix's eigenmodes and eigenvalues are calculated [54,55], as shown in Fig. 1(e) and (f) in the main text. These calculations are performed by the professional mathematical software Matlab [56]. Note that the partition number (915 in our case) chosen to discretize $S(q) - 1$ curves has no significant influence on the final PCA results, as long as the partition number is large enough to reflect the curve's true profile. We also tried other partition numbers and the PCA results are the same.

In addition, more data sets across a broader temperature range can increase the accuracy of PCA results. Therefore, we also performed the PCA analysis on the structure factor curves from numerical simulations, which cover a much broader temperature range between -60 °C and 100 °C with 33 temperatures in between. Consistent with the experimental results, the fraction of LDL, s, has a good linear relationship with $p1$, and the 1st eigenmode has very similar profile as the structure factor difference between LDL and HDL, as shown in Fig. S3.
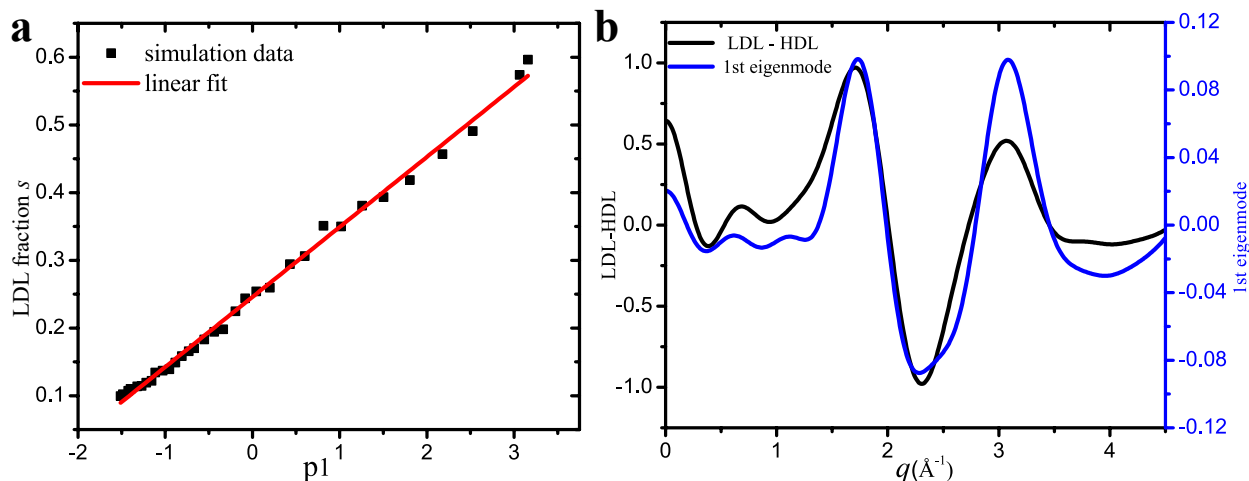
Figure S3: PCA analysis on our numerical simulations, which covers a much broader temperature range (-60 °C to 100 °C) than the experiment (-5 °C to 80 °C). (a) The LDL fraction shows an excellent linear relation with the projection pre-factor $p1$. (b) The first eigenmode agrees very well with the curve LDL minus HDL.

4   Simulation details

Many different water models exist. According to the number of interaction points, there are 3-site models SPC [57], TIP3P [58], SPC/E [59], 4-site models TIP4P [58], TIP4P-Ew [60], TIP4P/2005 [61], 5-site models ST2 [62], and TIP5P [63], etc.. In general, the electrostatic interaction is calculated with Coulomb's law, and the dispersion and repulsion forces are calculated with the Lennard-Jones potential [64]. The main differences between these models are the charges carried by the sites, the bond lengths, and the parameters in the Lennard-Jones potential besides the number of the sites. We tried five models, SPC/E, TIP3P, TIP4P, TIP4P-Ew and TIP5P in our MD simulations, with

the software Gromacs (v.2016.5) [65]. Consistent with experimental conditions, the simulations were performed under the isothermal-isobaric NPT ensemble. The isotropic Parrinello-Rahman barostat [66,67] and the Nose-Hoover thermostat [68,69] were used to guarantee the simulations were performed at the set pressure and temperature. 512 water molecules were put in a cubic box (side length $\approx$ 25 Å) and the pressure was kept fixed at the atmospheric pressure 1.01 bar and the temperature varied from -60 °C to 100 °C in every five degrees. Periodical boundary condition (PBC) was used to eliminate the influence from boundary. The simulation time for water at low (below -20 °C) and high (above -20 °C) temperatures last 20 and 10 ns respectively, which are long enough for the system to reach the equilibrium state. Besides the normal pressure, the simulation time for water at extreme HDL and LDL conditions (super-high and super-low pressure) was 20 ns.

5 Simulated scattering intensity and structure factor

In general, the total X-ray differential scattering cross section dσ/dΩ of molecular liquids can be divided into two parts, contributions from individual molecules (self-scattering) and from inter-molecular correlations defined by the function $H_{ij}(q)$,

$$I(q) = \sum_{ij} x_i x_j f_i(q) f_j(q) \frac{\sin(q r_{ij})}{q r_{ij}} + \sum_{i \leq j} x_i x_j f_i(q) f_j(q) H_{ij}(q) \tag{8}$$

$$H_{ij}(q) = 4\pi\rho \int_0^\infty r^2 dr (g_{ij}(r) - 1) \frac{\sin(qr)}{qr} \tag{9}$$

where x_i is the atomic fraction of atom type i, $f_i(q)$ is the atomic scattering factor for atom type i and r_{ij} are the intramolecular distances between atom centers. ρ is the atomic density and $g_{ij}(r)$
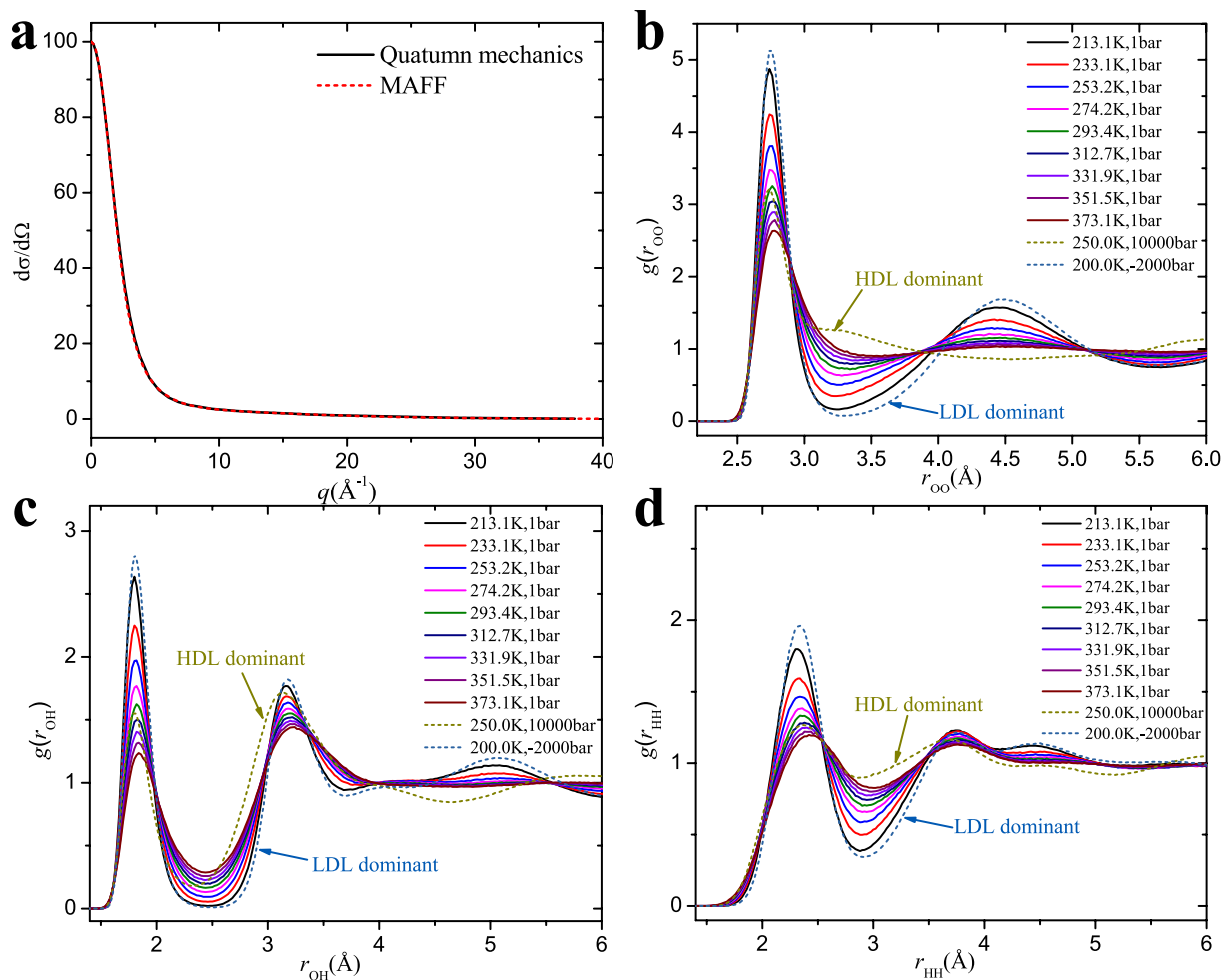
Figure S4: (a) The $F^2(q)$ calculated using quantum mechanics and MAFF agree very well within a broad q range. (b), (c) and (d) The related radial distribution functions, $g(r_{OO})$, $g(r_{OH})$ and $g(r_{HH})$, which are required to calculate the simulated scattering intensity at different conditions.

is the radial distribution function (RDF) between atom i and j. In our experiment, the scattering intensity $I(q)$ can be represented as,

$$
\begin{aligned}
I(q) &= I_{self-intra}^{nor}(q) + I_{inter}^{nor}(q) \\
&= f_O^2(q) + 2f_H^2(q) + 4f_O(q)f_H(q)\frac{\sin(qr_{OH})}{qr_{OH}} + 2f_H^2(q)\frac{\sin(qr_{HH})}{qr_{HH}} \\
&+ \tfrac{4}{3}H_{HH}^{inter}(q)f_H^2(q) + \tfrac{4}{3}H_{OH}^{inter}(q)f_H(q)f_O(q) + \tfrac{1}{3}H_{OO}^{inter}(q)f_O^2(q)
\end{aligned}
\tag{10}
$$

The first term in Eq. (10) is $F^2(q)$ mentioned in *Experiment data correction* section, which can be calculated using quantum mechanics [51]. Of course, it can also be calculated using the formula in Eq. (10). In fact, the difference between the two methods can be neglected when appropriate parameters are used in Eq. (10), as shown in Fig. S4(a). The bond lengths r_{OH} = 0.9572 Å and r_{HH} = 1.5139 Å are used. Our simulated results are calculated using the quantum mechanics data. The second term in Eq. (10) (or Eq. (8)) represents the intermolecular correlations: H_{ij} term can be calculated using Eq. (9), the atomic density is 512*3 divided by volume, and the upper limit of integral is set to 10 Å. The related $g_{ij}(r)$ (shown in Fig. S4(b), (c) and (d)) can be calculated after the system reaches equilibrium, and then the scattering intensity can be calculated using Eq. (9) and (10). The corresponding molecular structure factor $S(q)$-1 based on $I(q)$ can then be obtained.

To calculate the molecular structure factor of the 5-hydrogen-bond component, $S_{5H}(q)$-1, we first pick out the molecules surrounded by five hydrogen bonds, and set them as origin to obtain $g_{ij}(r)$. Then the intensity can be calculated using Eq. (9) and (10), which further gives $S_{5H}(q)$-1.

Among all the five popular water models tested, we found that the results from the TIP4P-Ew model agree with our experiments the best (shown in the main text Fig. 2(e) and (f)), which is used
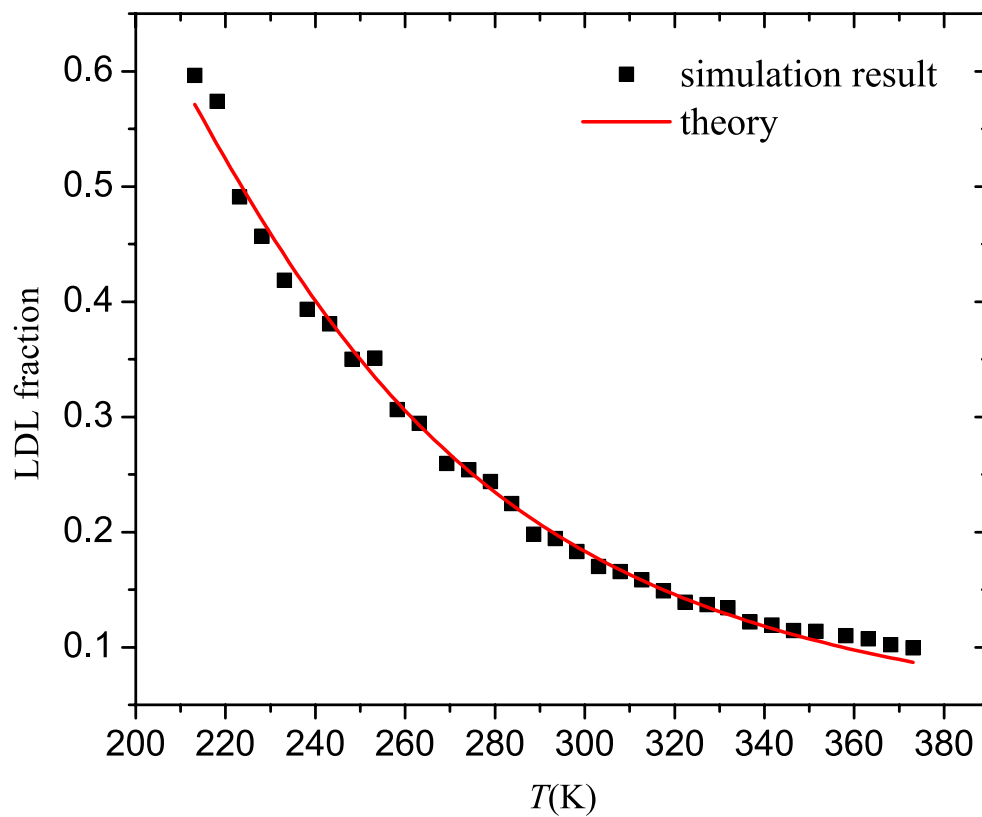
throughout our simulation.

Figure S5: Verifying the two-component model with simulation. The simulation results (square symbols) agree well with the two-component model (solid curve), in a temperature range much broader than the experiment.

6 Two-component model

For a liquid composed by two components, such as LDL and HDL, its free energy G can be represented as [70]:

$$G = G_{LDL} + s\Delta G + k_B T \left[s \ln s + (1-s) \ln(1-s) \right] + Js(1-s) \tag{11}$$

where s is the fraction of LDL, G_{LDL} is the free energy of pure LDL, $\Delta G = G_{LDL} - G_{HDL} = \Delta E - T\Delta\sigma + P\Delta V$, and J is the coupling between the two components. As the two components convert into each other, the equilibrium will be reached under the following condition:

$$\frac{\partial G}{\partial s} = \Delta G + k_B T \ln(\tfrac{s}{1-s}) + J(1 - 2s) = 0 \tag{12}$$

Under our experiment and simulation conditions, the terms J and $P\Delta V$ in ΔG can be neglected [41,71], and then the relation between s and T can be obtained:

$$s = \frac{1}{1 + e^{\frac{\Delta E}{k_B T} - \frac{\Delta\sigma}{k_B}}} \tag{13}$$

In fact, this relationship of Eq. (13) is applicable in a temperature range much broader than our experiment. To verify it, we apply it to the simulations that can reach down to -60 °C, and find that the two-component model agrees excellently with the simulations, as shown in Fig. S5. The fitting gives that $\Delta E/k_B$ = -1311.6 K and $\Delta\sigma/k_B$ = -5.87, which are close to our main text results.

7 Calculating the fractions of LDL and HDL

Following the method in ref [38], we decompose the order parameter distribution $P(\zeta)$ into the combination of two Gaussian functions,

$$
\begin{aligned}
P(\zeta) &= (1-s)P_{\mathrm{HDL}}(\zeta) + sP_{\mathrm{LDL}}(\zeta) \\
&= \frac{P(0)}{\exp(-\frac{m_{\mathrm{HDL}}^2}{2\sigma_{\mathrm{HDL}}^2})}\exp(-\frac{(\zeta-m_{\mathrm{HDL}})^2}{2\sigma_{\mathrm{HDL}}^2}) + \left(1 - \frac{\sigma_{\mathrm{HDL}}\sqrt{2\pi}P(0)}{\exp(-\frac{m_{\mathrm{HDL}}^2}{2\sigma_{\mathrm{HDL}}^2})}\right)\frac{\exp(-\frac{(\zeta-m_{\mathrm{LDL}})^2}{2\sigma_{\mathrm{LDL}}^2})}{\sqrt{2\pi}\sigma_{\mathrm{LDL}}}
\end{aligned}
\tag{14}
$$

where s is the fraction of LDL, and this formula has assumed that $P_{LDL}(0) = 0$, which is reasonable because of the open structures of LDL. The distribution of order parameter, $P(\zeta)$, under different temperatures are shown in Fig. S6(a). For comparison, the HDL and LDL dominant samples are also plotted. As temperature increases, the data curve shifts from LDL-like to HDL-like, as a result of their mutual conversion. We can further obtain the exact fractions of LDL and HDL in each sample, by fitting its $P(\zeta)$ curve with two Gaussian curves, as shown in Fig. S6(b): the two Gaussian curves represent thermally broadened HDL and LDL components respectively, and the weight of each Gaussian gives the fraction of each component. Using this formula to fit HDL and LDL dominant systems' $P(\zeta)$, we can get their fractions as shown in Fig. S6(c) and (d).

Due to the existence of the 3rd component, the fractions of LDL and HDL need to be recalculated. We use a similar method as the two-component model, except that we identify the 3rd component and exclude such molecules first. After eliminating the third component and renormalizing the rest into unity, we obtained the renormalized order parameter distribution $P(\zeta)$ and decomposed it into two Gaussian curves to get the new renormalized fractions of LDL and HDL at different temperatures. One typical fitting with two Gaussian curves is shown in Fig. S6(b) and
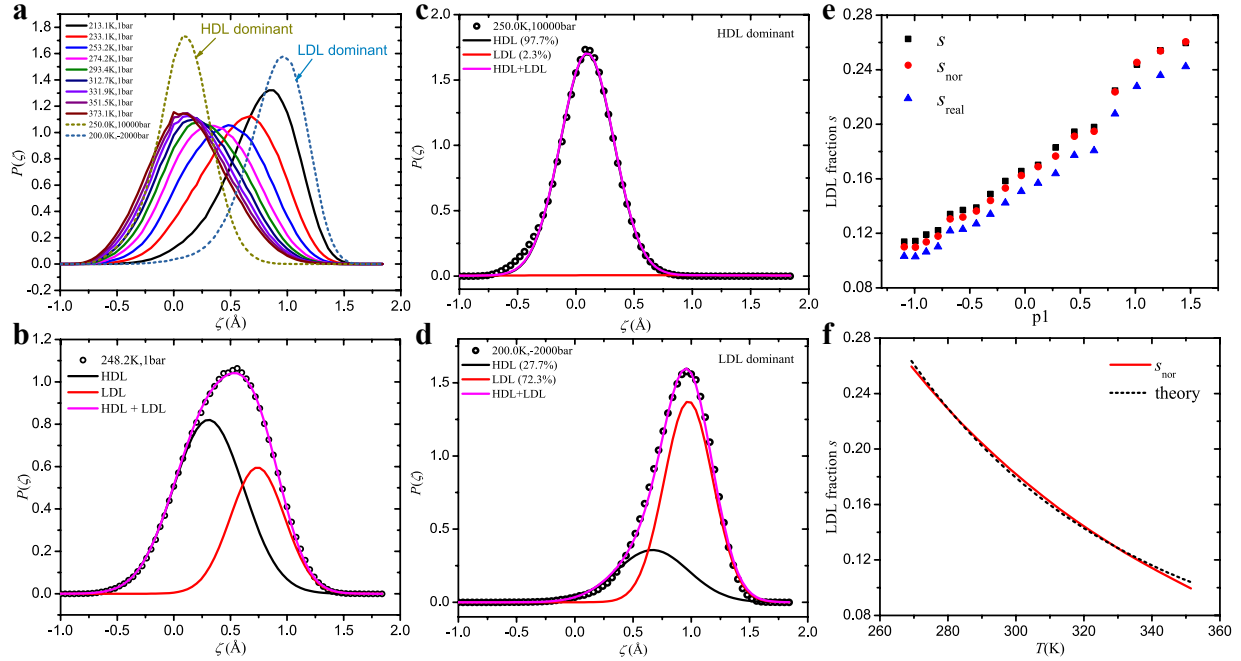
Figure S6: (a) The distribution of the order parameter, $P(\zeta)$, under different conditions. (b) Fitting $P(\zeta)$ excellently with two Gaussian curves. The left Gaussian corresponds to HDL and the right one corresponds to LDL. Their weights give the fractions of HDL and LDL. The two-Gaussian fitting method also works well after the 3rd component is considered and excluded. (c) and (d) The two-Gaussian decomposing result of our HDL-dominant and LDL-dominant systems. The result shows that we have obtained a HDL dominant (97.7%) system (c) and an LDL dominant (72.3%) system (d). (e) The fraction of LDL obtained in three different methods: (1) the previous two-component method without considering the 3rd component (black), (2) after considering and excluding the 3rd component and then renormalizing the LDL fraction among the rest (red), and (3) the real LDL fraction in the system after considering the 3rd component (blue). The result shows that the fraction of LDL changes slightly (a few percent) due to the low fraction of the 3rd component (blue). (f) Fitting the re-normalized LDL result with the two-component model, Eq. (13). The red solid curve comes from the data points of the re-normalized method (red) in (e) and the agreement with theory (dashed curve) is very well.

17

the agreement is excellent. Due to the low fraction of the 3rd component, the new renormalized

fractions of LDL (and HDL) only change very little in comparison to the two-component model,

as shown by the red disc and black square symbols in Fig. S6(e). The real fractions of LDL after

considering the existence of the 3rd component decreases by a few percent in comparison to the

previous pure two-component result, which are shown as the blue triangles in Fig. S6(e). Fitting

the theory of Eq. (13) with the renormalized s curve in Fig. S6(f), we can obtain the fitting param-

eters of $\Delta E/k_B$ = -1296.6 K and $\Delta\sigma/k_B$ = -5.84, which only deviate slightly (around 5%) from

the previous two-component model ($\Delta E/k_B$ = -1238.3 K, $\Delta\sigma/k_B$ = -5.63). Therefore, the third

component does not cause significant changes and the two-component model still describes the

main feature of water's structure.

8   Structures of LDL and HDL

According to the fitting parameters in Eq. (13), we know that LDL has lower energy and entropy

compared with HDL. Therefore, LDL tends to dominate at low temperatures. Following previous

research [38,39], we construct LDL dominant system at P = -2000 bar, T = 200.0 K and HDL domi-

nant system at P = 10000 bar, T = 250.0 K in simulation. Using Eq. (10), their scattering intensity

$I(q)$ can be calculated from simulation, and the corresponding molecular structure factor $S(q)$ is

obtained, both are shown in Fig. S7(a). Obviously, the two systems exhibit different features in

their $I(q)$ and $S(q)$ curves. In their $I(q)$ curves, the HDL dominant system has one main peak at

q = 2.28 Å^{-1} while the main peak for the LDL dominant system is at 1.68 Å^{-1}. In a recent exper-

iment [31], similar peaks at q = 2.15 Å^{-1} for HDL and q = 1.7 Å^{-1} for LDL were observed during
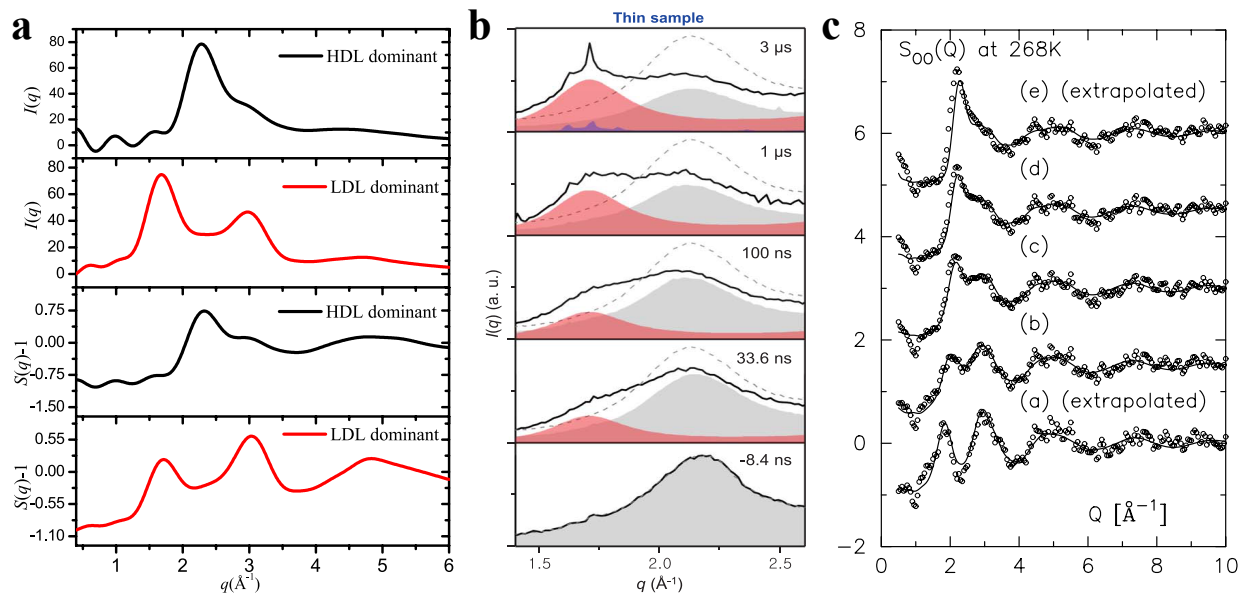
18

Figure S7: Comparison between our simulated LDL and HDL system in (a) with the previous experiment data in (b) [31] and (c) [42]. (a) Our simulated scattering intensity ($I(q)$) and molecular structure factor ($S(q)$) of HDL-dominant and LDL-dominant systems. (b) The contribution of HDL and LDL to the scattering intensity in ref [31]. The LDL curve (red) shows a peak at q = 1.7 $Å^{-1}$ while the HDL curve (gray) has a peak at q = 2.15 $Å^{-1}$, which are close to the main peaks in our simulated $I(q)$. (c) The extrapolated structure factor of HDL and LDL in ref [42]. The HDL curve (e) exhibits a peak and a shoulder while the LDL curve (a) shows two well separated peaks between 1~4 $Å^{-1}$, which are similar to our simulated $S(q)$.

the liquid-liquid transition process (shown in Fig. S7(b)), which agrees well with our simulation results. In addition, our HDL dominant system's $S(q)$ exhibits a main peak at $q = 2.32$ Å^{-1} and a shoulder at $q = 2.97$ Å^{-1} while the LDL dominant system exhibits two peaks at $q = 1.71$ Å^{-1} and 3.03 Å^{-1}. These properties also agree well with the results in a previous neutron diffraction experiment [42] (shown in Fig. S7(c)). The two well-separated peaks in the $S(q)$ of LDL dominant system indicate the separation of its first and second shells at molecular level. The separation between the two shells results in a low-density and ordered structure with low energy and entropy. By contrast, for the $S(q)$ of HDL dominant system, the main peak and the shoulder-like peak on its right are not well separated, due to the collapse of the first and second shells in the molecular structure. This results in a high-density and disordered structure with high energy and entropy. The agreement between experiment and simulation provides a strong support for our simulation.

9   Structure of the 5-hydrogen-bond component

To explore the structure of the 5-hydrogen-bond component in real space, we calculate its angle distribution P(θ) as the following: we first connect the central oxygen atom with the five surrounding oxygen atoms to form five straight lines, the angle between any two lines is then defined as θ and the distribution of these angles gives P(θ). P(θ) at different temperatures are shown in the main text Fig. 4(d). Obviously, P(θ) is not a random distribution and some specific structure should exist. To figure out this structure, we note that all five hydrogen bonds have similar lengths, and approximately assume that the five surrounding oxygen atoms are uniformly distributed on a spherical surface. How to distribute N points uniformly on a spherical surface? This is the
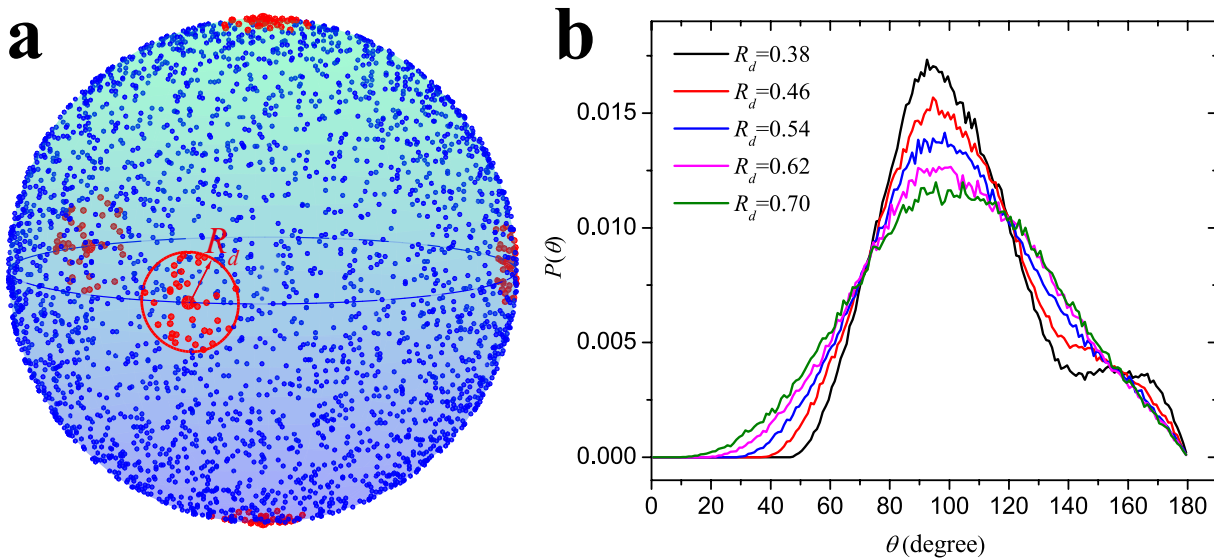
20

Figure S8: Random perturbations to the uniform distribution of five points on a spherical surface. (a) First we put many (10^5) random points uniformly distributed on the spherical surface whose radius is 1. All the points are generated using the method in the ref [74]. The red points represent the points within the distance R_d from the original ideal positions. Five such red points plus the sphere center form one specific configuration mimicing the 5-hydrogen-bond structure. Numerous such configurations ($3 * 10^4$) are then obtained for good statistics. (b) The distribution P(θ) under different R_d perturbations. The distribution agrees well with the simulated 5-hydrogen-bond structure when R_d equals to 0.46 (shown in the main text Fig. 4(d)).

so called *Spherical Code* or *Thomson problem*. For our $N = 5$ situation, putting two points at north and south pole respectively, and the rest three on the equator with the angle $120°$ is the most uniform configuration [72,73]. To mimic thermal effect, we add random perturbations to the five points' positions as the following. We first generate 10^5 points randomly and uniformly distributed on a spherical surface [74], as shown in Fig. S8(a). Then we find all the points within the distance of R_d from the five ideal points. We randomly choose one point for each center and form one specific configuration and calculate the corresponding angles θ. For each R_d, we calculate $3*10^4$ random configurations to get the angle distribution as shown in Fig. S8(b). We find that the angle distribution agrees the best with the actual simulation results when R_d equals to 0.46 of sphere radius (shown in the main text Fig. 4(d)), which indicates that 0.46 best matches the thermal fluctuations around our simulation temperature range. Therefore, the 5-hydrogen-bond structure can be considered as five water molecules uniformly distributed around the central molecule on a spherical surface. Due to the high symmetry [73] and low potential energy of this configuration, the 5-hydrogen-bond component exhibits a very stable fraction with temperature change, which is different from LDL and HDL.

10 Differentiating three components and calculating their local densities

The fractions of the three components can be calculated by the order parameter distribution (LDL and HDL) and the unique 5-H-bond structure. To calculate their local densities, however, a well-defined criterion must be defined to differentiate the three components microscopically. The 5-H-bond component can be easily picked out because of its unique structure, and a convenient method

to differentiate the LDL and HDL in the rest molecules is to set a cut-off order parameter ζ to satisfy the fractions calculated by the previous two-Gaussian decomposing method (shown in Fig. S9(a)). Water molecules with the order parameter less than the cut-off ζ belong to the HDL component and the rest belong to the LDL component. This method requires the cut-off ζ to change with the temperature as shown in Fig. S9(b).
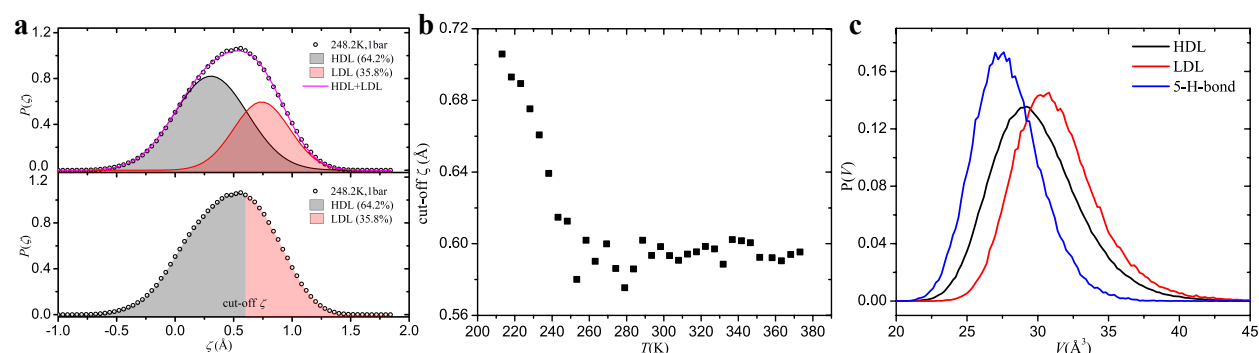
Figure S9: (a) Our approach to differentiate the LDL and HDL at the molecular level. Upper panel: the two-Gaussian decomposing method to calculate the fractions of HDL and LDL. Lower panel: a cut-off ζ is chosen to separate HDL from LDL. The cut-off ζ is determined such that the fractions of HDL and LDL matches the two-Gaussian decomposing method. (b) The cut-off ζ is quite stable above 260K (-13 °C) while it changes significantly below this temperatures. (c) The Voronoi cell volume distributions of the three components in one typical sample at 25 °C.

Then we calculated the Voronoi cell volume of each water molecule based on the positions of the oxygen atoms. For the oxygen atoms at the boundary of simulation box, their neighboring oxygen atoms should be counted based on the periodic boundary condition, and then their Voronoi

cells are constructed. The Voronoi cell calculation is performed with the software Matlab [75]. The distribution of the Voronoi cell volume is shown in Fig. S9(c), and apparently the third component has the smallest volume while the LDL component has the largest volume. Their densities are calculated by their mean Voronoi cell volumes, as shown in the main text Fig. 4(f): The 5-H-bond or 3rd component exhibits an ultra-high density significantly larger than both HDL and LDL.

11 PCA on $I(q)$ curves for water system

In general the XRD scattering intensity $I(q)$ is the quantity obtained in actual scattering experiments, which has contributions from both intramolecular and intermolecular correlations, as Eq. (10) shows for the water system. The relationship between the normalized intensity $I^{nor}(q)$ and $S(q)$ is described by the formula $S(q) - 1 = [I^{nor}(q) - F^2(q)]/[2f_H(q) + f_O(q)]^2$. We note that in our water samples the normalized intensity $I^{nor}(q)$ and $S(q)$ are linearly correlated because the factors $F(q)$, $f_H(q)$ and $f_O(q)$ in the formula are only related to the structure of one single water molecule and thus they keep unchanged with temperature and component variations. Therefore, in water system all the results of PCA on $I(q)$ curves should be equivalent to $S(q)$ curves. We can directly perform PCA on water's $I(q)$ curves at different temperatures and compare the results with $S(q)$ counterparts. First, there are two large eigenvalues meaning that there are two main reasons for water structure's evolution at different temperatures, as show in Fig. S10(a). The dominant reason is mutual conversion between HDL and LDL, which results in the linear relationship between the LDL fraction s and p1 (the projection pre-factor to the 1st eigenmode), as show in Fig. S10(b). Second, according to the analysis in the main text, the 1st eigenmode should correspond to the dif-
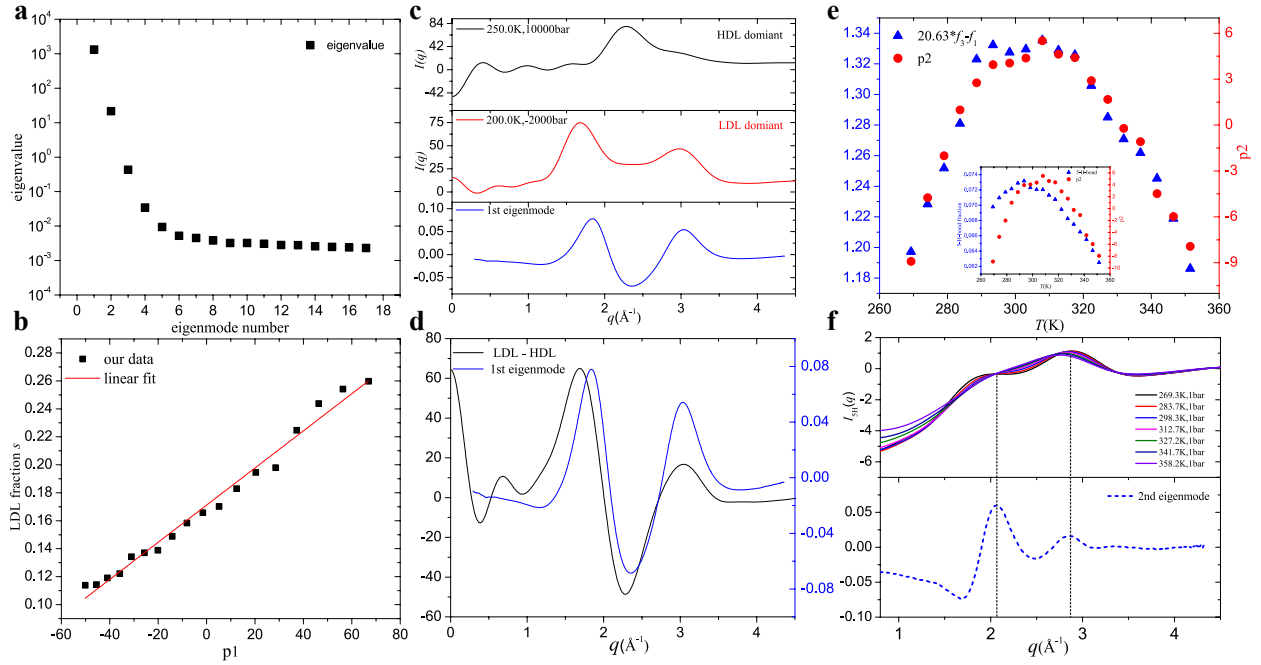
Figure S10: (a) The eigenvalues from large to small. The first two are significantly larger than the rest. (b) LDL fraction has a linear relation with the projection pre-factor $p1$. (c) Top to bottom panels show HDL system's $I(q)$, LDL system's $I(q)$ and the first eigenmode. (d) The first eigenmode agrees well with the curve obtained by the LDL curve minus the HDL curve. (e) Inset: the 5-H-bond fraction f_3 from simulation and the projection pre-factor p2 from experiment are both non-monotonic and having similar proile. Main panel: the non-monotonic $p2$ matches $20.63 \times f_3 - f_1$ nicely. (f) The main peaks in the second eigenmode correspond nicely to the isosbestic point and the main peak of $I_{5H}(q)$.

ference of HDL and LDL, and the results of PCA on $I(q)$ curves also support this interpretation, as show in Fig. S10(c) and (d). Third, the 2nd eigenvalue is related to the fraction change of the 3rd component, 5-H-bond structure. Therefore, $p2$ directly correlates to the linear combination of the fractions, f_3 and f_1: $p2 \propto 20.63 \times f_3 - f_1$, as show in Fig. S10(e). This formula is identical to the one for $S(q)$'s PCA analysis in the main text and quantitatively proves the equivalence of $I(q)$ and $S(q)$ results. At last, the intensity curves calculated from the 5-H-bond molecules as origins also exhibit main features that match the 2nd eigenmode, as shown in Fig. S10(f). This agrees with the $S(q)$ result in Fig. 4(b) in the main text. To conclude, all the reproducible results unambiguously show that the PCA analysis of $I(q)$ and $S(q)$ curves in our water samples are equivalent.

1. M. Vedamuthu, S. Singh, G. W. Robinson, Properties of liquid water: origin of the density anomalies, *The Journal of Physical Chemistry* **98**, 2222 (1994).

2. C. H. Cho, S. Singh, G. W. Robinson, Understanding all of water's anomalies with a nonlocal potential, *The Journal of chemical physics* **107**, 7979 (1997).

3. H. Tanaka, Simple physical explanation of the unusual thermodynamic behavior of liquid water, *Physical review letters* **80**, 5750 (1998).

4. C. H. Cho, S. Singh, G. W. Robinson, An explanation of the density maximum in water, *Physical review letters* **76**, 1651 (1996).

5. C. Angell, W. Sichina, M. Oguni, Heat capacity of water at extremes of supercooling and superheating, *The Journal of Physical Chemistry* **86**, 998 (1982).

6. G. S. Kell, Isothermal compressibility of liquid water at 1 atm., *Journal of Chemical and Engineering Data* **15**, 119 (1970).

7. R. Speedy, C. Angell, Isothermal compressibility of supercooled water and evidence for a thermodynamic singularity at -45 °C, *The Journal of Chemical Physics* **65**, 851 (1976).

8. R. J. Speedy, Stability-limit conjecture. an interpretation of the properties of water, *The Journal of Physical Chemistry* **86**, 982 (1982).

9. P. H. Poole, F. Sciortino, U. Essmann, H. E. Stanley, Phase behaviour of metastable water, *Nature* **360**, 324 (1992).

10. P. H. Poole, F. Sciortino, T. Grande, H. E. Stanley, C. A. Angell, Effect of hydrogen bonds on the thermodynamic behavior of liquid water, *Physical review letters* **73**, 1632 (1994).

11. C. A. Angell, Insights into phases of liquid water from study of its unusual glass-forming properties, *Science* **319**, 582 (2008).

12. S. Sastry, P. G. Debenedetti, F. Sciortino, H. E. Stanley, Singularity-free interpretation of the thermodynamics of supercooled water, *Physical Review E* **53**, 6144 (1996).

13. G. M. Schütz, C. Domb, J. Lebowitz, *Phase transitions and critical phenomena* (2001).

14. D. Fuentevilla, M. Anisimov, Scaled equation of state for supercooled water near the liquid-liquid critical point, *Physical review letters* **97**, 195702 (2006).

15. V. Holten, M. Anisimov, Entropy-driven liquid-liquid separation in supercooled water, *Scientific reports* **2**, 1 (2012).

16. O. Mishima, H. E. Stanley, The relationship between liquid, supercooled and glassy water, *Nature* **396**, 329 (1998).

17. O. Mishima, Y. Suzuki, Propagation of the polyamorphic transition of ice and the liquid-liquid critical point, *Nature* **419**, 599 (2002).

18. J. C. Palmer, *et al.*, Metastable liquid-liquid transition in a molecular model of water, *Nature* **510**, 385 (2014).

19. P. G. Debenedetti, F. Sciortino, G. H. Zerze, Second critical point in two realistic models of water, *Science* **369**, 289 (2020).

20. J. A. Sellberg, *et al.*, Ultrafast x-ray probing of water structure below the homogeneous ice nucleation temperature, *Nature* **510**, 381 (2014).

21. K. H. Kim, *et al.*, Maxima in the thermodynamic response and correlation functions of deeply supercooled water, *Science* **358**, 1589 (2017).

22. B. Cwilong, Sublimation in a wilson chamber, *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **190**, 137 (1947).

23. B. Mason, The supercooling and nucleation of water, *Advances in Physics* **7**, 221 (1958).

24. D. Thomas, L. Staveley, 889. a study of the supercooling of drops of some molecular liquids, *Journal of the Chemical Society (Resumed)* pp. 4569–4577 (1952).

25. S. Cerveny, F. Mallamace, J. Swenson, M. Vogel, L. Xu, Confined water as model of super-cooled water, *Chemical reviews* **116**, 7608 (2016).

26. L. Liu, S.-H. Chen, A. Faraone, C.-W. Yen, C.-Y. Mou, Pressure dependence of fragile-to-strong transition and a possible second critical point in supercooled confined water, *Physical review letters* **95**, 117802 (2005).

27. P. Gallo, *et al.*, Water: A tale of two liquids, *Chemical reviews* **116**, 7463 (2016).

28. D. Corradini, S. V. Buldyrev, P. Gallo, H. E. Stanley, Effect of hydrophobic solutes on the liquid-liquid critical point, *Physical Review E* **81**, 061504 (2010).

29. S. Woutersen, B. Ensing, M. Hilbers, Z. Zhao, C. A. Angell, A liquid-liquid transition in supercooled aqueous solution related to the HDA-LDA transition, *Science* **359**, 1127 (2018).

30. L. Kringle, W. A. Thornley, B. D. Kay, G. A. Kimmel, Reversible structural transformations in supercooled liquid water from 135 to 245 K, *Science* **369**, 1490 (2020).

31. K. H. Kim, *et al.*, Experimental observation of the liquid-liquid transition in bulk supercooled water under pressure, *Science* **370**, 978 (2020).

32. E. B. Moore, V. Molinero, Structural transformation in supercooled water controls the crystallization rate of ice, *Nature* **479**, 506 (2011).

33. H. Tanaka, A self-consistent phase diagram for supercooled water, *Nature* **380**, 328 (1996).

34. F. Smallenburg, L. Filion, F. Sciortino, Erasing noman's land by thermodynamically stabilizing the liquid-liquid transition in tetrahedral particles, *Nature physics* **10**, 653 (2014).

35. L. Xu, S. V. Buldyrev, C. A. Angell, H. E. Stanley, Thermodynamics and dynamics of the two-scale spherically symmetric jagla ramp model of anomalous liquids, *Physical Review E* **74**, 031108 (2006).

36. P. Tan, N. Xu, A. B. Schofield, L. Xu, Understanding the low-frequency quasilocalized modes in disordered colloidal systems, *Physical review letters* **108**, 095501 (2012).

37. H. Shen, P. Tan, L. Xu, Probing the role of mobility in the collective motion of nonequilibrium systems, *Physical review letters* **116**, 048302 (2016).

38. J. Russo, H. Tanaka, Understanding water's anomalies with locally favoured structures, *Nature communications* **5**, 1 (2014).

39. R. Shi, H. Tanaka, Direct evidence in the scattering function for the coexistence of two types of local structures in liquid water, *Journal of the American Chemical Society* **142**, 2868 (2020).

40. H. Tanaka, Simple physical model of liquid water, *The Journal of Chemical Physics* **112**, 799 (2000).

41. R. Shi, J. Russo, H. Tanaka, Common microscopic structural origin for water's thermodynamic and dynamic anomalies, *The Journal of chemical physics* **149**, 224502 (2018).

42. A. K. Soper, M. A. Ricci, Structures of high-density and low-density water, *Physical review letters* **84**, 2881 (2000).

43. V. H. Segtnan, Š. Šašić, T. Isaksson, Y. Ozaki, Studies on the structure of water using two-dimensional nearinfrared correlation spectroscopy and principal component analysis, *Analytical chemistry* **73**, 3153 (2001).

44. C. Yang, C. Zhang, F. Ye, X. Zhou, Ultra-high-density local structure in liquid water, *Chinese Physics B* **28**, 116104 (2019).

45. L. B. Skinner, C. J. Benmore, J. B. Parise, Area detector corrections for high quality synchrotron x-ray structure factor measurements, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **662**, 61 (2012).

46. R. Kahn, *et al.*, Macromolecular crystallography with synchrotron radiation: photographic data collection and polarization correction, *Journal of Applied Crystallography* **15**, 330 (1982).

47. A. Hammersley, Fit2d: a multi-purpose data reduction, analysis and visualization program, *Journal of Applied Crystallography* **49**, 646 (2016).

48. G. Hura, *et al.*, Water structure as a function of temperature from x-ray scattering experiments and ab initio molecular dynamics, *Physical Chemistry Chemical Physics* **5**, 1981 (2003).

49. J. Krogh-Moe, A method for converting experimental x-ray intensities to an absolute scale, *Acta Crystallographica* **9**, 951 (1956).

50. N. Norman, The fourier transform method for normalizing intensities, *Acta Crystallographica* **10**, 370 (1957).

51. J. Wang, A. N. Tripathi, V. H. Smith Jr, Chemical binding and electron correlation efeects in x-ray and high energy electron scattering, *The Journal of chemical physics* **101**, 4842 (1994).

52. L. B. Skinner, *et al.*, Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide q-range, *The Journal of chemical physics* **138**, 074506 (2013).

53. Z. Su, P. Coppens, Relativistic x-ray elastic scattering factors for neutral atoms z= 1-54 from multiconfiguration dirac-fock wavefunctions in the 0-12 Å^{-1} $\sin\theta/\lambda$ range, and six-gaussian analytical expressions in the 0-6 Å^{-1} range, *Acta Crystallographica Section A: Foundations of Crystallography* **53**, 749 (1997).

54. S. Henkes, C. Brito, O. Dauchot, Extracting vibrational modes from fluctuations: a pedagogical discussion, *Soft Matter* **8**, 6092 (2012).

55. K. Chen, *et al.*, Low-frequency vibrations of soft colloidal glasses, *Physical review letters* **105**, 025501 (2010).

56. Principal component analysis of raw data, `https://www.mathworks.com/help/stats/pca.html`.

57. H. Berendsen, J. Postma, W. Van Gunsteren, a. J. Hermans, Intermolecular forces (1981).

58. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water, *The Journal of chemical physics* **79**, 926 (1983).

59. H. Berendsen, J. Grigera, T. Straatsma, The missing term in effective pair potentials, *Journal of Physical Chemistry* **91**, 6269 (1987).

60. H. W. Horn, *et al.*, Development of an improved four-site water model for biomolecular simulations: Tip4p-ew, *The Journal of chemical physics* **120**, 9665 (2004).

61. J. L. Abascal, C. Vega, A general purpose model for the condensed phases of water: Tip4p/2005, *The Journal of chemical physics* **123**, 234505 (2005).

62. F. H. Stillinger, A. Rahman, Improved simulation of liquid water by molecular dynamics, *The Journal of Chemical Physics* **60**, 1545 (1974).

63. M. W. Mahoney, W. L. Jorgensen, A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions, *The Journal of Chemical Physics* **112**, 8910 (2000).

64. M. P. Allen, D. J. Tildesley, *Computer simulation of liquids* (Oxford university press, 2017).

65. H. J. Berendsen, D. van der Spoel, R. van Drunen, Gromacs: a message-passing parallel molecular dynamics implementation, *Computer physics communications* **91**, 43 (1995).

66. S. Nosé, M. Klein, Constant pressure molecular dynamics for molecular systems, *Molecular Physics* **50**, 1055 (1983).

67. M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, *Journal of Applied physics* **52**, 7182 (1981).

68. W. G. Hoover, Canonical dynamics: Equilibrium phase-space distributions, *Physical review A* **31**, 1695 (1985).

69. S. Nosé, A molecular dynamics method for simulations in the canonical ensemble, *Molecular physics* **52**, 255 (1984).

70. H. Tanaka, Thermodynamic anomaly and polyamorphism of water, *EPL (Europhysics Letters)* **50**, 340 (2000).

71. R. Shi, J. Russo, H. Tanaka, Origin of the emergent fragile-to-strong transition in supercooled water, *Proceedings of the National Academy of Sciences* **115**, 9444 (2018).

72. L. Whyte, Unique arrangements of points on a sphere, *The American Mathematical Monthly* **59**, 606 (1952).

73. J. Edmundson, The distribution of point charges on the surface of a sphere, *Acta Crystallographica Section A: Foundations of Crystallography* **48**, 60 (1992).

74. Weisstein, Eric W. Sphere Point Picking. From MathWorld–A Wolfram Web Resource.
    `https://mathworld.wolfram.com/SpherePointPicking.html`.

75. N-D Voronoi diagram, `https://www.mathworks.com/help/matlab/ref/`
    `voronoin.html`.