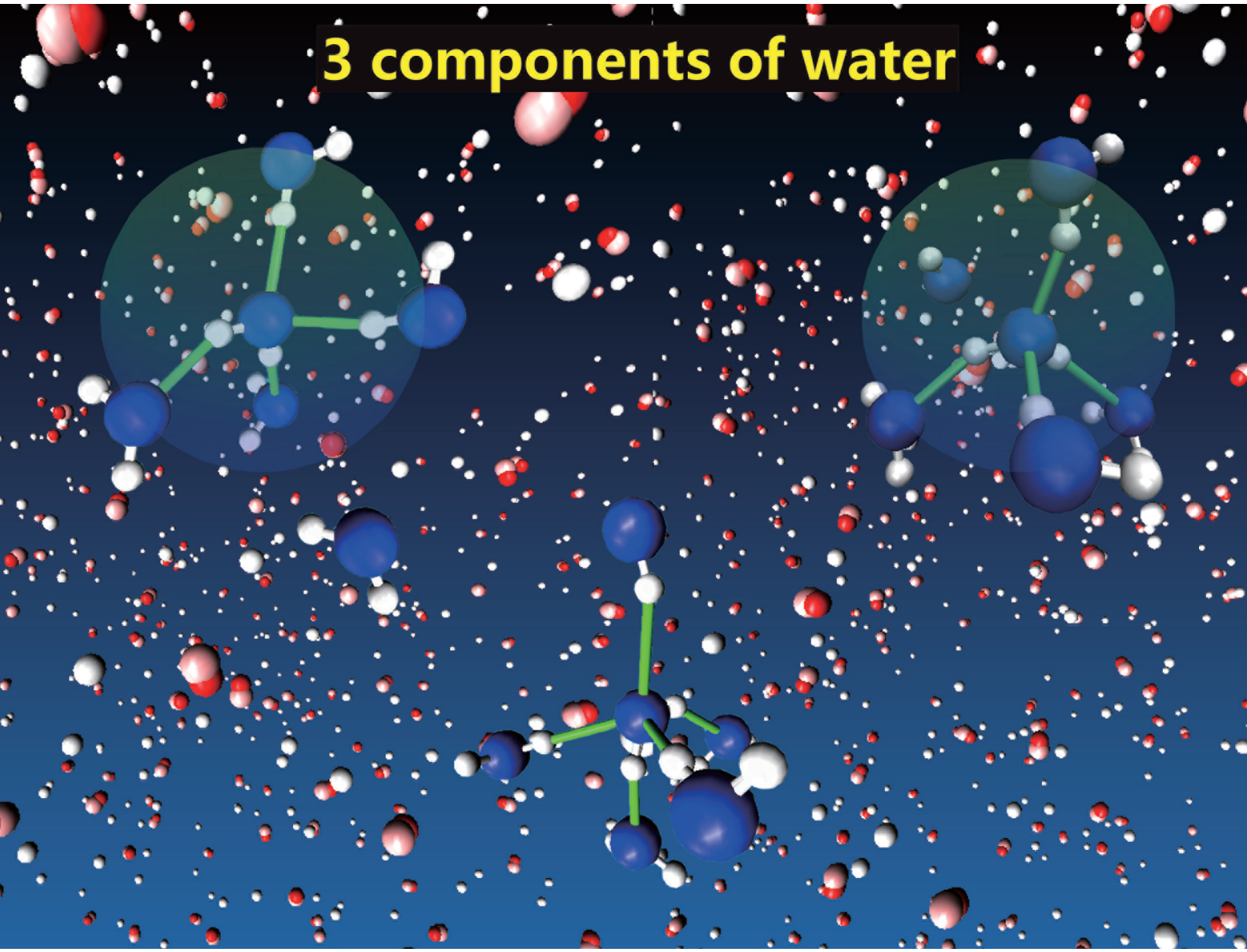
Volume 18
Number 39
21 October 2022
Pages 7455–7690

Soft Matter

rsc.li/soft-matter-journal



ISSN 1744-6848

ROYAL SOCIETY OF CHEMISTRY

PAPER
Guo Chen, Zhenlin Luo, Lei Xu *et al.*
Revealing the three-component structure of water with principal component analysis (PCA) of X-ray spectra

Soft Matter

PAPER

View Article Online
View Journal | View Issue

Cite this: *Soft Matter*, 2022, **18**, 7486

Revealing the three-component structure of water with principal component analysis (PCA) of X-ray spectra†

Zhipeng Jin,[a] Jiangtao Zhao,[b] Gang Chen,[c] Guo Chen,*[d] Zhenlin Luo*[b] and Lei Xu*[a]

Received 4th May 2022,
Accepted 30th July 2022

DOI: 10.1039/d2sm00576j

rsc.li/soft-matter-journal

Combining principal component analysis (PCA) of X-ray spectra with MD simulations, we experimentally reveal the existence of three basic components in water. These components exhibit distinct structures, densities, and temperature dependencies. Among the three, the two major components correspond to the low-density liquid (LDL) and the high-density liquid (HDL) predicted by the two-component model, and the third component exhibits a unique 5-hydrogen-bond configuration with ultra-high local density. As the temperature increases, the LDL component decreases and the HDL component increases, while the third component varies non-monotonically with a peak around 20 °C to 30 °C. The 3D structure of the third component is further illustrated as the uniform distribution of five hydrogen-bonded neighbors on a spherical surface. Our study reveals experimental evidence for water's possible three-component structure, which provides a fundamental basis for understanding water's special properties and anomalies.

1 Introduction

As the most common and important liquid on Earth, water is not typical at all; it exhibits rather special properties and anomalies. For example, some properties exhibit non-monotonic behavior, such as the maximum density at 4 °C,[1–4] the minimum heat capacity at 35 °C,[5] and the minimum isothermal compressibility at 46 °C,[6,7] which are very different from most other liquids. These special properties originate from water's unique structure, which is still a big mystery under active debate. The main difficulty lies in the experimental front and the lack of direct experimental proof which prevents a clear understanding of water's structure.

At the theoretical front, several different models and scenarios have been proposed, including the stability-limit conjecture,[8] the second critical point hypothesis,[9] the critical-point-free scenario,[10,11] and the singularity-free scenario.[12] In particular, researchers have found that multiple quantities of water seem to diverge when extrapolated to $T_c = -45$ °C,[7] which is a signature of the critical point.[13] As a result, the second critical point hypothesis has gained prevailing support.[14–20] This hypothesis proposes a liquid–liquid critical point (LLCP) for water at $T_c = -45$ °C: below this the water phase separates into a low-density liquid (LDL) and high-density liquid (HDL); and above this water is a mixture of the two (note that the critical point also requires a pressure much higher than the atmospheric pressure and cannot be realized under normal conditions[20–22]). Recent experiments show that water exhibits a maximum in both isothermal heat capacity[14] and compressibility[22] at 1 bar, consistent with the existence of a LLCP at lower temperatures ($P > 0$).

However, it is very difficult to realize the critical point experimentally, because the predicted critical temperature, $T_c = -45$ °C, is far below the freezing temperature of water and even below the homogeneous nucleation temperature at −41 °C.[23–25] Consequently, water will spontaneously crystallize far above T_c and it is very difficult to observe LDL and HDL in their bulk phases or their transition, until a recent experiment that was performed at very rapid speed and high pressures.[26] Although crystallization can be inhibited through strong confinement,[27–29] adding solutes[30,31] or some other special treatment,[21,26,32] the results under such special conditions may not be representative for pure bulk water at normal

[a] *Department of Physics, the Chinese University of Hong Kong, Hong Kong, China. E-mail: xuleixu@cuhk.edu.hk*
[b] *National Synchrotron Radiation Laboratory, University of Science and Technology of China, Hefei 230026, China. E-mail: zlluo@ustc.edu.cn*
[c] *School of Physical Science and Technology, ShanghaiTech University, Shanghai 201210, China*
[d] *Chongqing Key Laboratory of Soft Condensed Matter Physics and Smart Materials, College of Physics, Chongqing University, Chongqing 400044, China. E-mail: wezer@cqu.edu.cn*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d2sm00576j

pressure. As a result, it is more practical to numerically probe the components of water with molecular dynamics (MD) simulations,[19,20,33–36] which however give very different conclusions with different water models. Therefore, an objective and systematic experimental analysis, which can extract direct and unambiguous experimental evidence, is currently the most critical aim in this research field.

2 X-ray diffraction experiment

To address this critical issue, we turn to a powerful mathematical tool, principal component analysis (PCA). It is a mature and robust approach for identifying crucial components or dominant factors from their complex combinations. We apply this powerful analysis to the X-ray spectra of liquid water at different temperatures (−5 °C to 80 °C), and combine it with numerical simulations to illustrate the microscopic structure of water. As expected, the experimental data unambiguously reveals two major components with LDL-like and HDL-like structures, which convert into each other as temperature changes. More strikingly, our data further uncovers a third component, which exhibits distinct structure, density, and temperature dependency from the other two. Combining experiment with simulation, our study elucidates the unique three-component structure of water, and provides a fundamental basis to understand water's special properties and anomalies.

To probe water's microscopic structure, we perform X-ray scattering experiments at the BL19U2 station of Shanghai Synchrotron Radiation Facility (SSRF). Our X-ray beam has a spot size of 320 × 43 μm^2 and the energy of 12.000 ± 0.002 keV. After the incident X-rays interact with the deionized water sample, the scattered X-ray is collected by a Pilatus 1M area detector. The schematics of the setup are shown in Fig. 1(a) and a typical raw scattering image is shown in Fig. 1(b). To maximize the measurement range in q, the detector only collects about one-quarter of the scattered light, as shown in Fig. 1(b). The temperature of the deionized water is systematically varied by a thermal stage (Linkam HFSX350) between −5 °C to 80 °C, and we record the X-ray scattering image after the sample temperature is stabilized at designated values. The water temperature is measured in real time, by a Micro-BetaCHIP Thermistor probe placed adjacent to the incident area (within 0.5 mm). Further experimental details can be found in the ESI.†

After the data processing of image integration, background subtraction, absorption correction, geometry and polarization correction (see details in ESI†), we obtain the scattering intensity curves $I(q)$ under different temperatures, as plotted in Fig. 1(c). The structure factor $S(q)$ can be further obtained from $I(q)$ (see details in ESI†), as plotted in Fig. 1(d), which is consistent with previous measurements.[37] Based on these $S(q)$ curves, we perform the principal component analysis (PCA).

3 Principal component analysis

PCA is a powerful mathematical approach (see Methods) widely used in various fields, such as statistics, data science, quantitative finance, neuroscience, and physics.[38,39] Analogous to finding the principal axes of a rigid body, PCA identifies the most critical

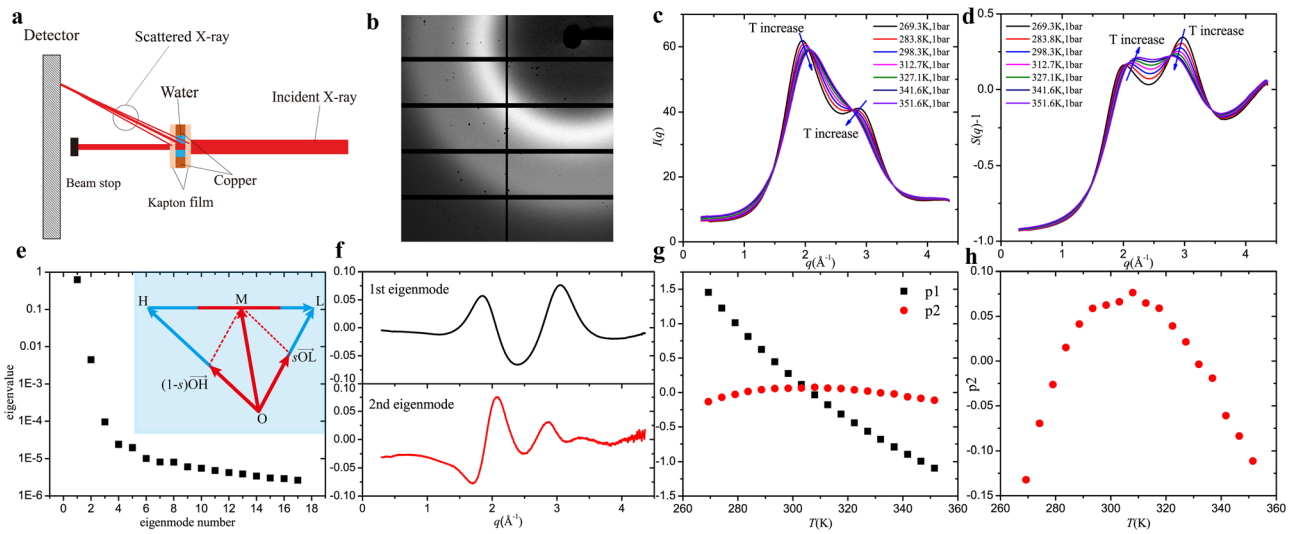


Fig. 1 (a) The schematics of the experimental setup. (b) A typical X-ray scattering image. The horizontal and vertical dark regions are gap regions without active pixels. (c) Scattering intensity $I(q)$ under different temperatures (−5 °C to 80 °C). (d) Structure factor $S(q)$ under different temperatures. (e) Main panel: eigenvalues from large to small. The first two are significantly larger than the rest. Inset: vector analysis of the first eigenmode. $\overrightarrow{\mathrm{OL}}$, $\overrightarrow{\mathrm{OH}}$ and $\overrightarrow{\mathrm{HL}}$ represent LDL, HDL, and the first eigenmode, respectively. $\overrightarrow{\mathrm{OM}}$ represents an actual sample with a fraction of s LDL and $(1-s)$ HDL. According to vector addition, point M locates on the straight line of $\overrightarrow{\mathrm{HL}}$:$\overrightarrow{\mathrm{OM}} = s\overrightarrow{\mathrm{OL}} + (1-s)\overrightarrow{\mathrm{OH}} = s\left(\overrightarrow{\mathrm{OL}} - \overrightarrow{\mathrm{OH}}\right) + \overrightarrow{\mathrm{OH}}$, which gives $\overrightarrow{\mathrm{OM}} - \overrightarrow{\mathrm{OH}} = s\left(\overrightarrow{\mathrm{OL}} - \overrightarrow{\mathrm{OH}}\right)$ and thus $\overrightarrow{\mathrm{HM}} = s\overrightarrow{\mathrm{HL}}$. As s varies with temperature, point M sweeps through the red segment on $\overrightarrow{\mathrm{HL}}$ and each point on this red segment corresponds to one actual sample we can experimentally measure. (f) The first and second eigenmodes calculated from $S(q)$ curves. (g) Each $S(q)$ curve measured in (d) is projected onto the first and second eigenmodes, with $p1$ and $p2$ the projection pre-factors. $p1$ decreases significantly while $p2$ is quite stable. (h) Zoomed-in plot of $p2$ shows non-monotonic behavior.

eigenmodes of the covariance matrix built from different measurements, and illustrates the data with these critical modes. For a complex system composed of multiple components, the eigenmodes of PCA can reveal valuable information from basic components, and the projections along these eigenmodes can provide the component weight information. Compared with previous analysis on water structure,[40,41] our PCA approach has the advantage of being completely objective without any fitting, systematically combining all data from different measurements, and thus can provide unambiguous and robust experimental evidence.

Based on the different $S(q)$ measurements in Fig. 1(d), we construct the covariance matrix and obtain its eigenvalues and eigenmodes (see ESI†). Fig. 1(e) main panel shows the eigenvalues arranged from large to small: apparently, the first and second eigenvalues are significantly larger than the rest, which are small and at noise level. Thus, most physics in our data can be represented by the first two eigenmodes, which indicate two main reasons for water structure evolution under different temperatures. We will show that the first reason is due to the mutual conversion between two dominant water components, and the second reason is related to the third component of water.

The first eigenvalue accounts for the majority of the total sum, indicating that the physics is dominated by the first eigenmode and thus essentially one-dimensional (1D) (note that the eigenvalues are the variance of raw data projected along each eigenmode but not the fraction of each component). This 1D behavior is consistent with a mutual conversion between two basic components as proven below. Without loss of generality, the basic components and their superpositions can be represented by vectors in the eigenmode space. Therefore, we use two vectors – $\overrightarrow{\mathrm{OH}}$ for HDL and $\overrightarrow{\mathrm{OL}}$ for LDL – to represent the two basic components of water as shown in the inset of Fig. 1(e). Linear combinations of $\overrightarrow{\mathrm{OH}}$ and $\overrightarrow{\mathrm{OL}}$, which represent mixtures of HDL and LDL, can in general spread in a 2D plane. However, the mutual conversion means that one changes into another, but their sum remains as a constant. This constraint eliminates one degree of freedom and collapses the 2D combinations into a 1D straight line: assuming that an experimental system contains LDL with a fraction of s and HDL with a fraction of $(1 - s)$, its vector representation, $\overrightarrow{\mathrm{OM}} = s\overrightarrow{\mathrm{OL}} + (1 - s)\overrightarrow{\mathrm{OH}}$, must lie on the straight line of vector $\overrightarrow{\mathrm{HL}}$ for an arbitrary s, as shown by the vector analysis in the inset of Fig. 1(e). As we vary the temperature from -5 °C to 80 °C, the fraction s will change correspondingly and $\overrightarrow{\mathrm{OM}}$ will sweep across the red segment on $\overrightarrow{\mathrm{HL}}$. Therefore, if water is indeed mainly composed of two components, their mutual conversion should generate 1D behavior with one dominant eigenvalue and eigenmode, consistent with our observation. Furthermore, this dominant eigenmode has the physical meaning of vector $\overrightarrow{\mathrm{HL}}$ (after renormalization), which means the difference between LDL and HDL: $\overrightarrow{\mathrm{HL}} = \overrightarrow{\mathrm{OL}} - \overrightarrow{\mathrm{OH}}$. We plot this first eigenmode in Fig. 1(f) (upper panel) and its physical meaning as vector $\overrightarrow{\mathrm{HL}}$ will be experimentally confirmed later.

Besides the dominant first eigenvalue, the second eigenvalue also exceeds the rest by a significant amount, indicating new physics beyond the two-component picture. The second eigenmode is plotted in Fig. 1(f) lower panel, which exhibits more characteristic peaks and valleys than the first mode. Because the second eigenmode is orthogonal to the first, it must be independent of the two components' conversion and indicates the existence of a third component we will verify later. We emphasize that despite the discovery of the third component, the two-component model does describe the major water structure, and the third component is a small yet significant ingredient. Its discovery demonstrates the high sensitivity of our PCA analysis for picking up small yet significant signals.

To understand the water structure based on the first and second eigenmodes, we project the measured $S(q)$ curves onto the two modes with the projection prefactors $p1$ and $p2$, as shown in Fig. 1(g). Later we will show that $p1$ and $p2$ correlate to the various component weights in the actual samples. Clearly, $p1$ reduces significantly with T, corresponding to the decrease of the LDL component as verified later. For the second eigenmode, $p2$ is relatively stable and non-monotonic, suggesting the distinct behavior of the third component. An enlarged picture for $p2$ is plotted in Fig. 1(h), which reveals a peak around 30 °C. To summarize, PCA reveals intriguing clues from experimental data; however a complete understanding requires molecular level information, which will be achieved by combining experimental data with molecular dynamics (MD) simulations.

4 Molecular dynamics simulations and two-component model

We conduct simulations using multiple popular water models – SPC/E, TIP3P, TIP4P, TIP4P-Ew and TIP5P – and find that the TIP4P-Ew model fits our experiment the best and will be used throughout this work (see more simulation details in ESI†). A snapshot of 512 water molecules equilibrating at 25 °C is shown in Fig. 2(a), and we average over 2000 such snapshots to achieve good statistics. With the locations of every water molecule obtained from simulations, we plot the radial distribution function of water molecules, $g(r_{\mathrm{OO}})$, under various temperatures in Fig. 2(b). For comparison, we also simulate the LDL-dominant and HDL-dominant samples based on the previous literature,[40,41] and plot them as dashed curves. Due to the small system size and short time scale, the simulated samples can reach down to -60 °C without freezing, and clearly the structure approaches LDL as the temperature lowers. After Fourier transform, we can further obtain the scattering intensity $I(q)$ and the structure factor $S(q)$ (see ESI† for details), as plotted in Fig. 2(c) and (d), respectively. Even over the much broader temperature range explored, the simulations exhibit a very similar trend as the actual measurements in Fig. 1(c) and (d). Fig. 2(e) and (f) demonstrate the direct comparison between simulation and experiment at 4 different temperatures, and the agreement between experiment and simulation is excellent. Therefore, the water structures obtained in the simulations

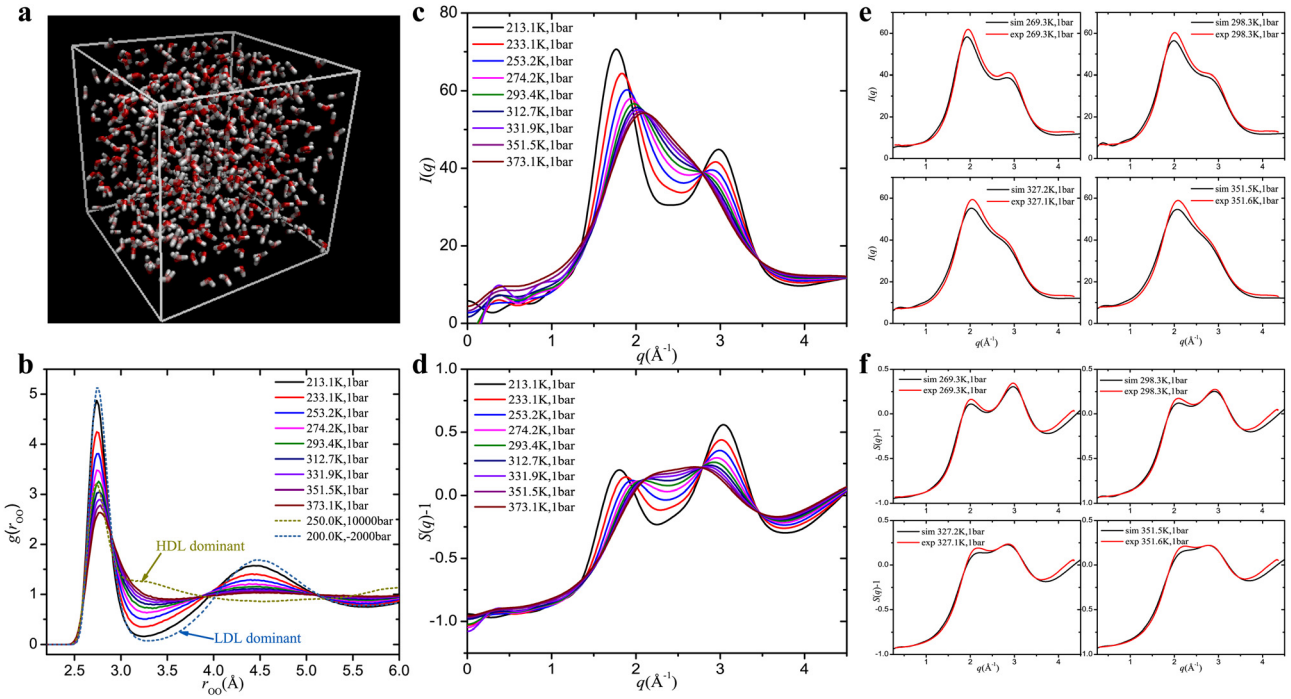


Fig. 2 (a) A snapshot of 512 water molecules at 25 °C simulated by the TIP4P-Ew model. We average over 2000 snapshots to obtain the simulation data. (b) The radial distribution function of oxygen atoms simulated at different temperatures. The LDL and HDL dominant systems are shown by dashed curves. (c) The simulated scattering intensity $I(q)$ under different conditions. (d) The simulated structure factor $S(q)$ under different conditions. (e and f) The comparison of $I(q)$ and $S(q)$ between simulation and experiment at 4 different temperatures.

can be considered as a reliable representation of the actual samples in the experiment.

Next, we verify the two-component model which is well established theoretically but has very rare direct experimental proof.[41] Based on previous research,[40] we distinguish the high-density liquid (HDL) and the low-density liquid (LDL) with the order parameter, $\zeta = r_2 - r_1$; for an arbitrary water molecule, r_1 is the distance to its furthermost hydrogen-bonded neighbor, and r_2 is the distance to its closest non-hydrogen-bonded neighbor, as shown in Fig. 3(a). This definition differentiates LDL and HDL by comparing the hydrogen-bonded neighbors with the non-hydrogen-bonded ones. For the case of LDL, the hydrogen-bonded neighbors (typically 4 of them) form the nearest shell within r_1, while the non-hydrogen-bonded neighbors are far away from this shell, *i.e.*, $\zeta > 0$ by a significant amount. In the HDL situation, however, the non-hydrogen-bonded neighbors are very close to or even enter into, this r_1 shell, producing a larger local density as shown in Fig. 3(a), right panel. These two components are indeed commonly observed in our simulations, as shown in Fig. 3(b): the relevant molecules are enlarged and colored in blue, and the hydrogen bonds are drawn as green long connections. Clearly, for LDL the closest non-hydrogen-bonded neighbor is far away from the r_1 shell while for HDL it is inside the r_1 shell.

Based on this definition of LDL and HDL, we can find their fractions at different temperatures in simulations (see ESI†). We then combine the fraction information from the simulation with experimental data, and verify the two-component model experimentally. According to the model, any actual system we can experimentally measure is a mixture of LDL and HDL: $\overrightarrow{OM} = s\overrightarrow{OL} + (1-s)\overrightarrow{OH} = s\left(\overrightarrow{OL} + \overrightarrow{OH}\right) + \overrightarrow{OH}$, with s and $(1-s)$ the fractions of LDL and HDL, and vectors $\overrightarrow{OM}$, $\overrightarrow{OL}$ and $\overrightarrow{OH}$ representing the actual mixture system, LDL and HDL, respectively (see Fig. 1(e)). According to this vector equation, $\overrightarrow{OM}$ depends linearly on s and so does the point M coordinate on $\overrightarrow{HL}$. Because the M coordinate on $\overrightarrow{HL}$ can be represented by the projection prefactor $p1$, the two-component model thus predicts a linear relation between $p1$ and s. To test this, we plot s *versus* $p1$ in Fig. 3(c) and obtain good linear dependence. As $p1$ and s are independently obtained from experiment and simulation, this linear relation shows clear evidence for the two-component model, as well as good agreement between our experiment and simulation.

We further compare this essential result, *i.e.*, the linear line in Fig. 3(c), with theory. Because all samples, both in simulation and experiment, are measured under specific temperatures, T is thus a hidden variable in Fig. 3(c). Therefore, we can change variables and plot the straight line of Fig. 3(c) in terms of the variables s and T, as shown in Fig. 3(d). For comparison, the theoretical prediction from the two-component model[40,42,43] is also plotted as the dashed curve, which agrees well with our result. According to this theory, we have the following relation between the LDL fraction s and the temperature T:

$$s = \frac{1}{1 + e^{\frac{\Delta E}{k_B}\frac{1}{T} - \frac{\Delta\sigma}{k_B}}} \quad (1)$$

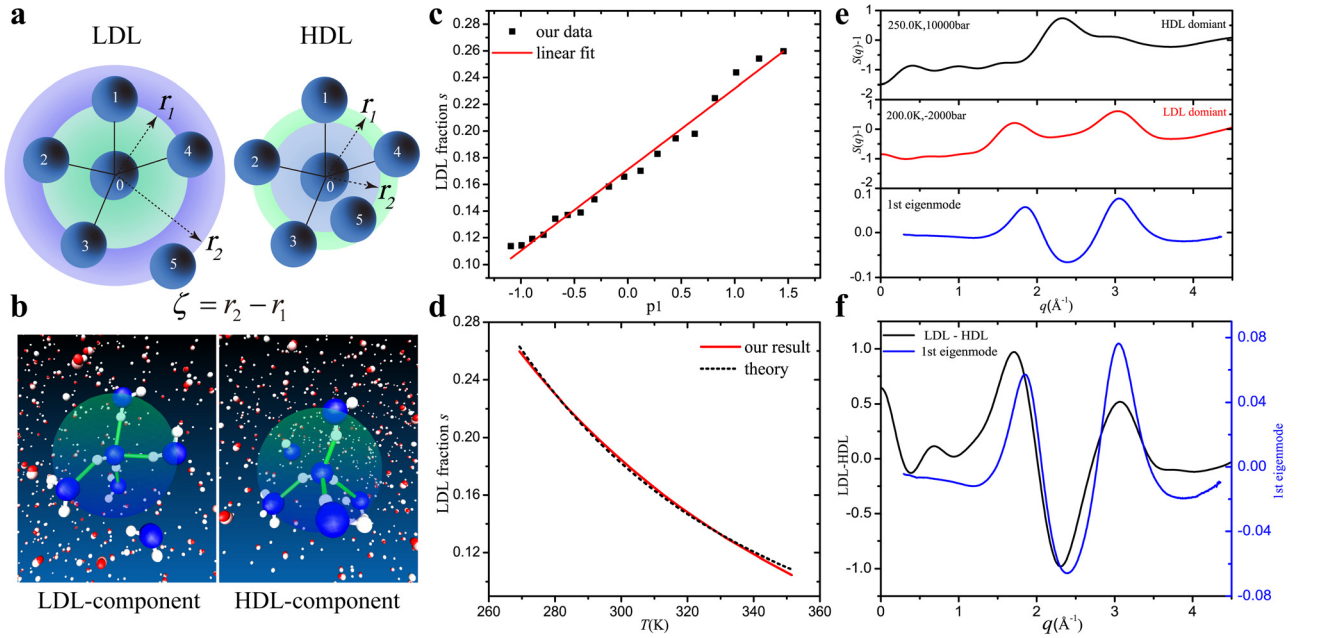


Fig. 3 (a) Schematics illustrating LDL and HDL. Molecules 1 to 4 are hydrogen bonded to 0 while 5 is not. r_1 is the distance to the furthermost hydrogen-bonded molecule, and r_2 is the distance to the closest non-hydrogen-bonded molecule. Following previous research,[40] the order parameter is defined as: $\zeta = r_2 - r_1$. For LDL, r_2 is well beyond the r_1 shell, while for HDL r_2 is very close to or even enters the r_1 shell. (b) Typical LDL and HDL components observed in our simulation. The relevant molecules are enlarged and shown in blue, and the hydrogen bonds are indicated by the green connections. (c) LDL fraction has a linear relation with the projection pre-factor $p1$. (d) Our result agrees well with the theoretical fitting. (e) Top to bottom panels show the HDL system's $S(q)$, the LDL system's $S(q)$ and the first eigenmode. The locations of the three main features in the eigenmode agree well with the HDL and LDL features: the two peaks correspond to the two LDL peaks, and the valley corresponds to the HDL peak. (f) The first eigenmode agrees well with the curve obtained by the LDL curve minus the HDL curve.

where $\Delta E = E_{\mathrm{LDL}} - E_{\mathrm{HDL}}$ and $\Delta\sigma = \sigma_{\mathrm{LDL}} - \sigma_{\mathrm{HDL}}$ are the energy and entropy difference between LDL and HDL, respectively. By fitting the theoretical dashed curve to our solid curve, we get: $\Delta E/k_{\mathrm{B}} = -1238.3$ K and $\Delta\sigma/k_{\mathrm{B}} = -5.63$, which are consistent with the previous study.[41] Apparently, LDL has a lower energy and entropy than HDL, because LDL typically exhibits four hydrogen bonds per molecule (3.98 in our simulation) with a relatively-ordered structure, while HDL contains a mixture of three and four hydrogen bonds (averaged at 3.79 in our simulation) with a more disordered structure. As a result, LDL is preferred at low temperatures due to its lower energy, while HDL is preferred at high temperatures due to its higher entropy. Moreover, the Schottky temperature[43] at which LDL and HDL components have an equal amount, $T\,(s = 0.5) = \Delta E/\Delta\sigma = 220.0$ K (*i.e.*, −53 °C), is close to the critical two-phase separation temperature 228.1 K (*i.e.*, −45 °C), as we naturally expect. Also the relationship of eqn (1) works nicely in simulations over a much broader temperature range (−60 °C to 100 °C) than our experiment, as shown in the ESI.†

Besides the theoretical agreement, we further illustrate the physical meaning of the first eigenmode, as promised earlier. According to the vector analysis in Fig. 1(e), the first eigenmode should correspond to the vector $\overrightarrow{\mathrm{HL}}$ (after renormalization), which is the difference between LDL and HDL: $\overrightarrow{\mathrm{HL}} = \overrightarrow{\mathrm{OL}} - \overrightarrow{\mathrm{OH}}$. We directly test it in q space, by comparing the first eigenmode from experiment with the difference of LDL and HDL from simulation. We first numerically construct HDL and LDL dominant systems under special conditions: the HDL-dominant (97.7%) system is realized at very high pressure ($P = 10\,000$ bar, $T = 250.0$ K), and the LDL-dominant (72.3%) system is achieved at very low pressure and temperature ($P = -2000$ bar, $T = 200.0$ K). Both systems agree well with previous studies.[26,44] Due to the very short time scale of simulations (10 to 20 ns), the LDL system can remain as liquid even at very low temperatures. The HDL and LDL dominant systems are plotted in the top two panels of Fig. 3(e), and the first eigenmode is plotted in the bottom panel for direct comparison.

Apparently, there are three main features in the bottom panel (2 peaks and 1 valley), whose locations or characteristic sizes match nicely with the main features of the top two panels (2 peaks in LDL and 1 peak in HDL). We then directly subtract the top two curves, *i.e.*, LDL minus HDL, and compare this difference against the first eigenmode, as shown in Fig. 3(f). The main features and trend match very well, confirming the picture that the first eigenmode corresponds to the difference between LDL and HDL. Once again, two independent data sets, the eigenmode from experiment and the LDL minus HDL curve from simulation, agree well with each other. This provides further strong experimental evidence for the two-component model.

5 The third component of the three-component picture of water

More interestingly, our study goes beyond the two-component picture and reveals the experimental evidence for the third component. According to our PCA analysis, besides the dominant first eigenvalue, the second eigenvalue is also significantly

larger than the rest and contains important physical information (see Fig. 1(e)). Because different PCA modes are orthogonal to each other, the second mode must be independent of the first, and reveals valuable information beyond the two-component picture. A previous study has uncovered some clues about the existence of a third component,[45] and a pioneering numerical study has proposed a possible candidate with five hydrogen bonds (5-H-bond).[46] We now reveal direct experimental evidence for its existence and illustrate the 3D structure.

We again illustrate the third component with independent data sets from experiment and simulation: the projection prefactors along the second eigenmode, $p2$, are obtained experimentally, and the fractions of 5-H-bond component under different temperatures, f_3, are obtained numerically. The two sets of data are plotted together in the Fig. 4(a) inset and their profiles resemble each other. Moreover, further analysis (shown later in Fig. 5(h) and Methods) indicates that $p2$ linearly correlates to the fractions of both the first component f_1 (*i.e.*, LDL fraction) and the third component f_3 (*i.e.*, 5-H-bond fraction): $p2 \propto 20.63 \times f_3 - f_1$. In the main panel of Fig. 4(a) we plot them together and observe an almost precise match. Such excellent agreement unambiguously proves that the third component is indeed the 5-H-bond structure.

To gain even more evidence, we compare the characteristic sizes in q space. Due to the lack of knowledge on the formation conditions, we cannot construct a system dominated by the 5-H-bond component, like the LDL or HDL dominant system. However, we can extract the 5-H-bond information from normal systems: we only pick out the molecules surrounded by five hydrogen bonds, and set such molecules as the origin to calculate the radial distribution function, $g_{5H}(r)$. We then Fourier transform $g_{5H}(r)$ to get $S_{5H}(q)$, and compare it with the second eigenmode from the experiment, as shown in Fig. 4(b). Clearly, there are two pronounced peaks in the second eigenmode curve at $q = 2.06$ Å^{-1} and $q = 2.88$ Å^{-1}, which correspond exactly to the isosbestic point[45] and the main peak of the $S_{5H}(q)$ curves as shown in the upper panel. This match once again confirms the 5-H-bond structure as the third component. In addition to peaks, there are also three major valleys in the second eigenmode, which come from LDL and HDL as they are close to the three main peaks in LDL and HDL (see Fig. 3(e)) (although the second eigenmode is orthogonal to the difference of LDL and HDL, it still contains features from LDL and HDL).

Next we illustrate the 3D structure of the third component, *i.e.*, how do the five hydrogen bonds locate spatially. Because the five hydrogen bonds have similar lengths, we can approximate the five hydrogen-bonded molecules as locating on a

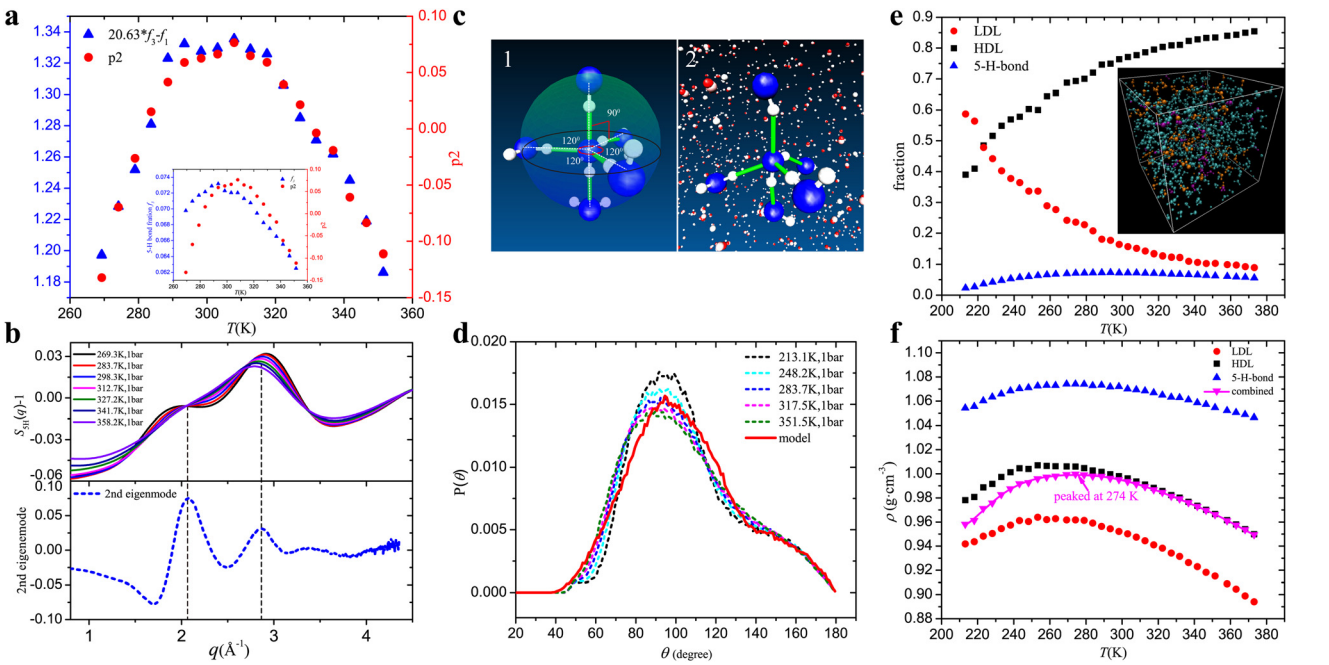


Fig. 4 (a) Inset: The 5-H-bond fraction f_3 from simulation and the projection pre-factor $p2$ from experiment, are both non-monotonic and resemble each other. Main panel: the non-monotonic $p2$ matches $20.63 \times f_3 - f_1$. (b) The main peaks in the second eigenmode (bottom panel) at $q = 2.06$ Å^{-1} and $q = 2.88$ Å^{-1} correspond exactly to the isosbestic point and the main peak of $S_{5H}(q)$ (top panel). $S_{5H}(q)$ is the structure factor calculated from 5-H-bond molecules. (c) Left panel: Schematics showing the uniform distribution of five hydrogen-bonded molecules on a spherical surface, with two molecules at the north and south poles, and three on the equator with 120° angle between them. Right panel, a typical 5-H-bond structure in an actual simulation exhibits a very similar configuration as the left panel, differing only by some thermal distortions. (d) The distribution of angles between 5-H-bond oxygen atoms, $P(q)$. The solid curve from our uniform distribution model plus random fluctuations (mimicking thermal fluctuations) agrees well with the actual simulations at different temperatures (dashed curves). (e) Main panel: the fractions of three components at various T. Inset: A snapshot at 293 K showing a perfect mix of three components at molecular level. Orange: LDL, cyan: HDL, purple: 5H-bond. (f) The densities of the three components and their combined density. The 5-H-bond component is significantly denser than both HDL and LDL. The combined density peaks at 274 K, close to the density maximum of water at 277 K.

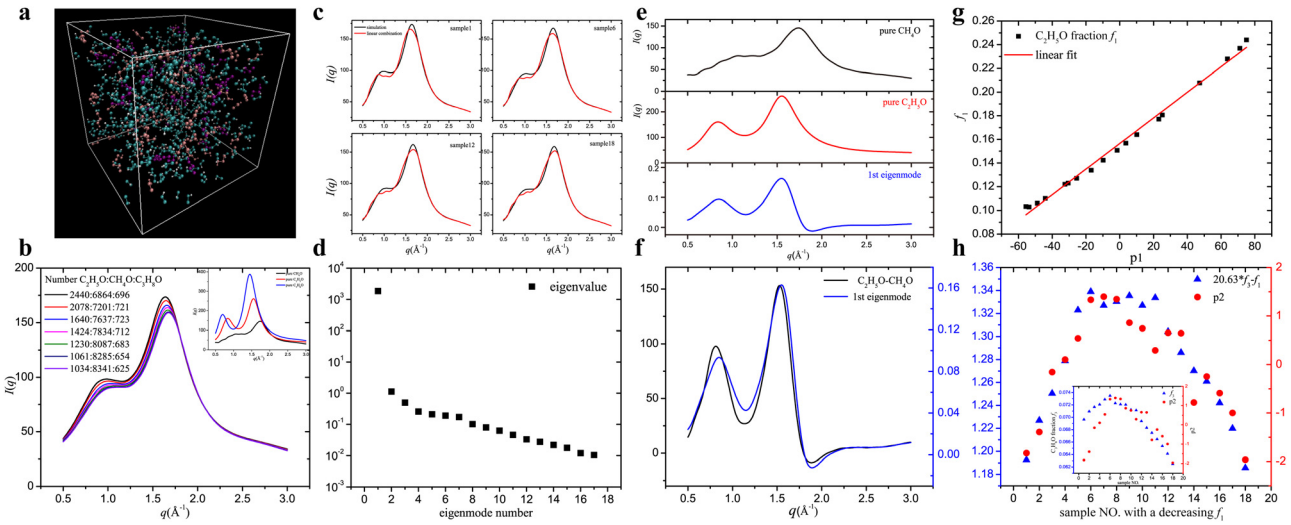


Fig. 5 (a) A snapshot of 512 alcohol molecules (5% of the total system) at 25 °C simulated with GROMOS 54A7 force field. The cyan, orange and purple represent CH_4O (methanol), C_2H_6O (ethanol) and C_3H_8O (1-propanol), respectively. (b) Several typical scattering intensity curves, $I(q)$, of the system with different fractions of the three alcohols. Inset: $I(q)$ curves for the three pure alcohols. (c) In a mixture, the linear superposition of three pure alcohols' $I(q)$ based on their fractions agree well with the mixture's actual $I(q)$. (d) The eigenvalues arranged from large to small. (e) Bottom panel: The 1st eigenmode from PCA analysis on $I(q)$. Middle panel: pure ethanol's $I(q)$. Top panel: pure methanol's $I(q)$.(f) The comparison between the difference of the two dominant components and the 1st eigenmode. They agree well with each other. (g) $p1$ has a nice linear relationship with ethanol's fraction f_1. (h) Inset: Non-monotonic profiles of $p2$ and f_3 resemble each other, indicating their close correlation. Main panel: $p2 \propto 20.63 \times f_3 - f_1$ is verified well.

spherical surface. We consider the most uniform distribution under which the molecules are separated to the largest degree. As shown in Fig. 4(c) left panel: two molecules locate on the north and south poles respectively, and three on the equator with a 120° angle between their connections. Indeed, typical 5-H-bond structures in our simulations exhibit very similar configurations as this assumption (except with some thermal fluctuations), as shown in the right panel of Fig. 4(c).

To statistically verify this most uniform distribution model, we collect a large amount of 5-H-bond molecules from actual simulations, connect their oxygen atoms, and statistically calculate the distribution of angles between these connections, $P(q)$. For comparison, we also calculate $P(q)$ generated from our uniform distribution model. In the ideal situation, there are only three possible angles: 90° (60%), 120° (30%), and 180° (10%) (left panel of Fig. 4(c)), which can however broaden into a continuous distribution after random perturbations that mimic thermal fluctuations (see ESI†). The curves from the actual simulation (dashed) and our model calculation (solid) are compared in Fig. 4(d) and excellent agreement is observed. This agreement confirms that the 3D structure of the third component is the most uniform distribution of five hydrogen-bonded molecules around the center, as demonstrated in Fig. 4(c).

To illustrate the overall three-component picture, we plot their individual fractions in Fig. 4(e) which add up to unity. As temperature increases, LDL decreases and HDL increases significantly, as a result of their mutual conversion. However, the third component with the 5-H-bond is much more stable, with distinct non-monotonic behavior peaking around 300 K. This indicates a fundamentally different thermal response of the third component. In addition, the fraction of the third component is around 5% to 7%, which is significant, although the other two components are more important. Correspondingly, the third component causes a few percent modifications to the two-component model and the theoretical predictions of eqn (1), which are shown in the ESI.† The inset shows a snapshot at 293 K with perfect mixing of three components at the molecular level. Our study thus provides robust evidence for the two-component model, and at the same time reveals valuable information for the third component.

We further calculate the densities of the three components with Voronoi cells obtained from each molecule, as shown in Fig. 4(f) (see ESI† for details). Interestingly, the third component exhibits ultrahigh density that is significantly higher than both HDL and LDL, agreeing with the previous study.[46] When the three densities are combined according to their fractions, $\frac{1}{\rho} = \frac{f_1}{\rho_1} + \frac{f_2}{\rho_2} + \frac{f_3}{\rho_3}$, we obtain the bulk water density ρ from the molecular level. As shown by the connected data curve in Fig. 4(f), ρ has a peak at 274 K, which is close to the maximum water density at 277 K. Therefore, our three-component picture at the microscopic level agrees with the macroscopic density anomaly of bulk water.

6 Comparison with the three-alcohol system

To ensure that our PCA method reliably reveals the three components of water, and more importantly to establish it as a general approach to distinguish the components of various system mixtures, we precisely design a well-defined three-component system and check whether it reproduces all the

PCA results of water. We construct a distinct system with three types of alcohol: C_2H_6O (ethanol, 0.789 g mL^{-1}), CH_4 O (methanol, 0.792 g mL^{-1}), and C_3H_8O (1-propanol, 0.804 g mL^{-1}), which mimic water's LDL, HDL and 5-H-bond structures, respectively, according to their low-to-high densities. Previously our water sample i was measured at 18 different temperatures and thus 18 combinations of its three component fractions were obtained in Fig. 4(e); correspondingly, 18 mixtures of the three alcohols with identical molar fractions are numerically constructed. These precisely-designed mixtures are then analyzed with PCA and compared against water's results. In these MD simulations, 10 000 molecules in total are put inside a cubic box, and the mixtures reach equilibrium at 25 °C and 1 bar. One example is shown in Fig. 5(a), which shows perfect mixing of the three alcohols just like water's three components. Several typical scattering intensity curves, $I(q)$, are shown in Fig. 5(b) and the inset shows the $I(q)$ curves of the three pure alcohols.

First we use this system to address a fundamental open issue: the vector analysis in Fig. 1(e) assumes the mixture's signal as a linear superposition of the pure components' signals, which has not been verified. With the three-alcohol system whose components' fractions are precisely known, we check this fundamental assumption directly. For any mixture we can linearly add up its three pure components' $I(q)$ based on their fractions and compare the sum with the mixture's actual $I(q)$, as shown in Fig. 5(c): the excellent match in several typical samples unambiguously proves this basic linear assumption as a good approximation. Thus all our previous linear analyses are fundamentally self-consistent.

Next we perform PCA analysis on the three-alcohol system. We directly analyze the $I(q)$ instead of $S(q)$ curves because the three alcohols have different molecular form factors and thus the mixtures only have $I(q)$ but not $S(q)$. In water we PCA analyze $S(q)$ curves because of the convenient comparison with previous studies that frequently use $S(q)$; however the PCA analysis on $I(q)$ is more general and valid for both water and other mixture systems.[49] The analysis on water's $I(q)$ and $S(q)$ also give the same results (see ESI†) and thus they are essentially equivalent.

All the water's PCA results are reproduced in the three-alcohol system: there is a dominant first eigenmode correlated to the fraction variations between the two major components, CH_4O and C_2H_6O, and a second eigenmode correlated to the third component, C_3H_8O (see Fig. 5(d) for eigenvalues). In Fig. 5(e) bottom panel, the first eigenmode exhibits two peaks and one valley which correspond to the main features in C_2H_6O (middle) and CH_4O (top), respectively, and in Fig. 5(f) the first mode agrees well with the two curves' difference, C_2H_6O–CH_4O (similar to LDL–HDL in Fig. 3(e) and (f)). For mode projections, the projection pre-factor on the first mode, $p1$, again exhibits excellent linear dependence with the first component's fraction f_1, as shown in Fig. 5(g). For the second mode, $p2$ also has a similar non-monotonic profile as the third component's fraction f_3, indicating their close correlation (see Fig. 5(h) inset). Theoretical analysis further finds that $p2$ directly correlates to the linear combination of f_3 and f_1: $p2 \propto 20.63 \times f_3 - f_1$ (see Methods), and the two data sets match each other very well as shown in Fig. 5(h). Moreover, this exact expression also works well in the water system, as shown in Fig. 4(a). One identical expression, $p2 \propto 20.63 \times f_3 - f_1$, works simultaneously in both the three-alcohol mixture and the water system, unambiguously proving the three-component picture of water and the universal validity of our PCA method.

7 A general model

As our PCA analysis on water and the three-alcohol simulations reveal reproducible results for a mixture of three components, we would like to test these results in a more general system. Because in general many features are peaks or valleys which can be approximated by Gaussian peaks, without loss of generality, we use Gaussian peaks or their superpositions as basic components, and then add three such components together to form a three-component mixture. A few types of basic components are shown in Fig. 6(a), which are either Gaussian peaks (type 1) or their superpositions (type 2 and 3). To compare with water, we use the identical fractions as water's three components shown in Fig. 4(e), and construct general three-component mixtures with these Gaussian peaks. Applying the PCA method to these mixture systems, we can reproduce all results from the water system. First, there are only two large eigenvalues as shown in Fig. 6(b). This indicates that there are two main reasons that cause the system evolution. The first reason is due to the mutual conversion between the two dominant components, as proved by the nice linear relationship between $p1$ and f_1 in Fig. 6(c), and the overlap of the first eigenmode and the difference between the two dominant components in Fig. 6(d). The second reason is related to the variation in the third component revealed by the second eigenmode. Once again, we observe a nice match between $p2$ and the linear combination of f_3 and f_1 in Fig. 6(e), which is similar to water and three-alcohol systems. Moreover, the pre-factor of f_3 is also very close to water and three-alcohol systems, regardless of the very different basic component shapes. To verify this further, we change the Gaussian peaks' height and width randomly for each component type in Fig. 6(a) and calculate the corresponding pre-factors of f_3. We find that all the pre-factors are around 21, as shown in Fig. 6(f), which agrees well with water's value of 20.63. Therefore, the quantitative agreement of this pre-factor in water, three-alcohol system, and the Gaussian peak system provides unambiguous evidence that our PCA analysis is robust and universal.

8 Conclusions

To conclude, in this work we apply PCA analysis to X-ray spectra, and obtain two significant eigenmodes that correspond to three components of water. The dominant first mode corresponds to the conversion between two major components, LDL and HDL, and the second eigenmode corresponds to a third component with five hydrogen bonds and ultra-high density. The 5-H-bond structure makes a separate third component as manifested by its

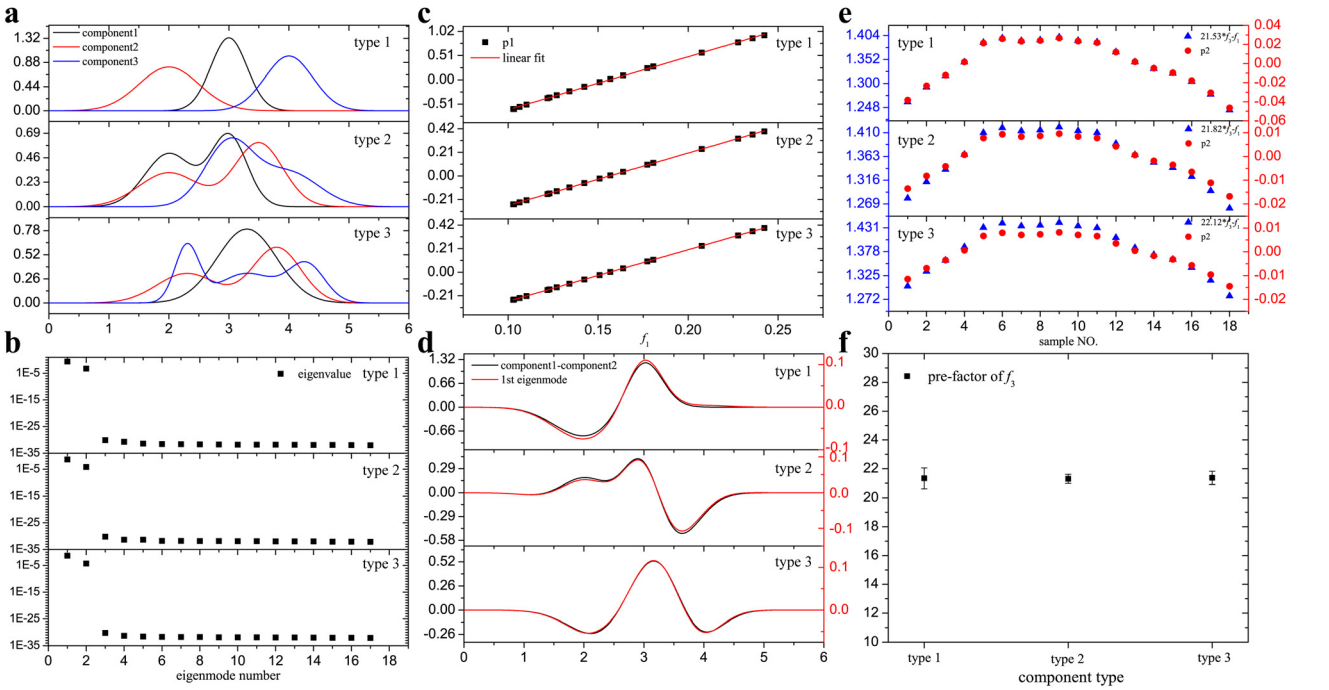


Fig. 6 (a) Three different types of the basic components used to generate the data set. Top panel: all three components are single Gaussian functions. Middle panel: all the three components are the sum of two Gaussian functions. Bottom panel: the first component is one Gaussian function, the second component is the sum of two Gaussian functions, and the third component is the sum of three Gaussian functions. (b) The eigenvalues of the three different data sets. Obviously, there are only two large eigenvalues. (c) All three data sets have a linear relationship between $p1$ and f_1. (d) The difference between two dominant components overlap well with the 1st eigenmode for all three tested systems. (e) The linear combination of f_1 and f_3 reproduce the same profile as $p2$, as predicted by eqn (3) in the section Methods, and the pre-factors of f_3 are around 21. (f) For more randomly generated data sets, the pre-factor of f_3 is also around 21.

unique non-monotonic response with the external condition, *i.e.*, temperature in our experiment, which is distinct from the monotonic behavior of LDL and HDL. The experimental evidence for the third component makes a breakthrough beyond the prevailing two-component picture, and opens new research directions such as a possible new phase separation for this third component as well as its special properties and interactions with the other two components. The PCA analysis also demonstrates itself as a powerful tool for identifying important components in complex systems.

Methods

PCA is a powerful mathematical tool for analyzing data with intrinsic connections. The main idea of PCA is to reduce data dimensionality by projecting each data point onto only the first few principal components or eigenmodes and obtain lower-dimensional data while preserving as much of the data's variation information as possible. Therefore, the eigenmodes in fact represent the directions towards which the data set vary most substantially. For a system with three different components such as water, the main reason for the system variation is the mutual conversion between the two major components, LDL and HDL. The second reason for the system variation is related to the fraction change in the third component, which has a fraction less than 10%. To simplify the question theoretically, we can approximate the mixture's signal as a linear combination of three different components' (see Fig. 5(c)), and the mixture's signal change is due to the fraction variations of these components. In terms of vector expression, the signal of an arbitrary mixture sample i can be expressed as: $\vec{I}_i(q) = f_{1i}\vec{\alpha} + f_{2i}\vec{\beta} + f_{3i}\vec{\gamma}$, with $\vec{\alpha}$, $\vec{\beta}$, $\vec{\gamma}$ the signals of pure components and f's their corresponding fractions. In the water system, LDL and HDL dominate and thus $f_{1i} + f_{2i} \approx 1$ and $f_{3i} \ll 1$. To get the relation between projection pre-factors and component fractions, we can project each sample curve to the eigenmodes, that is $p_{ij} = \vec{I}_i(q)\cdot\vec{e}_j$, where p_{ij} is the projection pre-factor of the ith sample onto the jth eigenmode. According to previous analysis in the Fig. 1(e) inset, the 1st eigenmode can be approximated as $\left(\vec{\alpha} - \vec{\beta}\right)$ and thus its unit vector can be approximated as: $\vec{e_1} \approx \left(\vec{\alpha} - \vec{\beta}\right)\Big/\left|\vec{\alpha} - \vec{\beta}\right|$. Next we derive the linear relationship between $p1$ and f_1:

$$\begin{aligned} p_{i1} &= \vec{I_i}(q)\cdot\vec{e_1} \approx \left(f_{1i}\vec{\alpha} + f_{2i}\vec{\beta} + f_{3i}\vec{\gamma}\right)\cdot\left(\vec{\alpha} - \vec{\beta}\right)\Big/\left|\vec{\alpha} - \vec{\beta}\right| \\ &= \left|\vec{\alpha} - \vec{\beta}\right| f_{1i} + f_{3i}\left(\vec{\gamma} - \vec{\beta}\right)\cdot\left(\vec{\alpha} - \vec{\beta}\right)\Big/\left|\vec{\alpha} - \vec{\beta}\right| \\ &\quad + \vec{\beta}\cdot\left(\vec{\alpha} - \vec{\beta}\right)\Big/\left|\vec{\alpha} - \vec{\beta}\right| \\ &= \left|\vec{\alpha} - \vec{\beta}\right| f_{1i} + C_1 + C_2 \end{aligned} \tag{2}$$

The latter two terms do not influence the linear relationship because $f_{3i} \ll 1$ and the last term is a constant. This linear relation is confirmed well by our experiment and simulation.

We further derive the relationship between $p2$ and f_3, by projecting the data curve onto the 2nd eigenmode:

$$\begin{aligned} p_{i2} &= \overrightarrow{I_i}(q)\cdot\overrightarrow{e_2} = \left(f_{1i}\overrightarrow{\alpha} + f_{2i}\overrightarrow{\beta} + f_{3i}\overrightarrow{\gamma}\right)\cdot\overrightarrow{e_2} \\ &= \left(\overrightarrow{\alpha} - \overrightarrow{\beta}\right)\cdot\overrightarrow{e_2}f_{1i} + \left(\overrightarrow{\gamma} - \overrightarrow{\beta}\right)\cdot\overrightarrow{e_2}f_{3i} + \overrightarrow{\beta}\cdot\overrightarrow{e_2} \\ &= \left[\frac{\left(\overrightarrow{\gamma} - \overrightarrow{\beta}\right)\cdot\overrightarrow{e_2}}{\left(\overrightarrow{\beta} - \overrightarrow{\alpha}\right)\cdot\overrightarrow{e_2}}f_{3i} - f_{1i}\right]C + C' \end{aligned} \tag{3}$$

Here C and C' are non-important constants and we only focus on the terms in the square brackets. Because all eigenmodes are orthogonal to each other, *i.e.*, $\overrightarrow{e_1}\cdot\overrightarrow{e_2} = 0$, and $\overrightarrow{e_1} \approx \left(\overrightarrow{\alpha} - \overrightarrow{\beta}\right)\Big/\left|\overrightarrow{\alpha} - \overrightarrow{\beta}\right|$, then the denominator part of the pre-factor of f_{3i}, $\left(\overrightarrow{\beta} - \overrightarrow{\alpha}\right)\cdot\overrightarrow{e_2}$, is close to zero, resulting in this pre-factor being much larger than 1. Thus $p2$'s behavior will strongly depend on f_3 due to this large pre-factor, consistent with our experiment and simulation results in the Fig. 4(a) and 5(h) insets. Moreover, this pre-factor is a universal constant in both the water system and three-alcohol system, which gives a universal relation: $p2 \propto 20.63 \times f_3 - f_1$ in both systems. This result is confirmed by the main panels of Fig. 4(a) and 5(h).

To test the robustness of our PCA approach even further, we theoretically construct three-component systems with different Gaussian peaks and their superpositions, and obtain the same results as previous water and three-alcohol systems, as shown in Fig. S11 (ESI†). All these tests unambiguously indicate that the PCA analysis is robust and universal. Therefore, PCA's conclusion that water is a mixture of three components, and their fractions vary with temperature, is reliable and robust, which should help to settle the long-time debate of water structure in the literature.[47,48]

Author contributions

L. X. conceived this research, Z. J. performed all the experiments, PCA analysis and simulation, Guo Chen and Z. L. helped with the experiments and supervised, J. Z. and Gang Chen helped in the experiment and data analysis, Z. J. and L. X. wrote the manuscript, and L. X. supervised the project.

Conflicts of interest

The authors declare that they have no competing financial interests.

Acknowledgements

The *in situ* synchrotron XRD was performed at the BL19U2 station of the Shanghai Synchrotron Radiation Facility (SSRF). We are thankful for the valuable suggestions from Dr Na Li, Dr Guangfeng Liu, Dr Wentao Li, Dr Rui Sun and Dr Jizhou Li. Funding: L. X. acknowledges financial support from GRF14306518, GRF14306920, CRF-C1018-17G, NSFC12074325, CUHK direct grant 4053354; Z. L. acknowledges the financial support from the National Key R & D program (2016YFA0300102) and the National Nature Science Foundation of China (11675179); Guo Chen acknowledges the financial support from the Natural Science Foundation Project of CQ CSTC (Grant No. cstc2020jcyj-msxmX0106), and the Fundamental Research Funds for the Central Universities.

Notes and references

1 M. Vedamuthu, S. Singh and G. W. Robinson, *J. Phys. Chem.*, 1994, **98**, 2222–2230.
2 C. H. Cho, S. Singh and G. W. Robinson, *J. Chem. Phys.*, 1997, **107**, 7979–7988.
3 H. Tanaka, *Phys. Rev. Lett.*, 1998, **80**, 5750.
4 C. H. Cho, S. Singh and G. W. Robinson, *Phys. Rev. Lett.*, 1996, **76**, 1651.
5 C. Angell, W. Sichina and M. Oguni, *J. Phys. Chem.*, 1982, **86**, 998–1002.
6 G. S. Kell, *J. Chem. Eng. Data*, 1970, **15**, 119–122.
7 R. Speedy and C. Angell, *J. Chem. Phys.*, 1976, **65**, 851–858.
8 R. J. Speedy, *J. Phys. Chem.*, 1982, **86**, 982–991.
9 P. H. Poole, F. Sciortino, U. Essmann and H. E. Stanley, *Nature*, 1992, **360**, 324–328.
10 P. H. Poole, F. Sciortino, T. Grande, H. E. Stanley and C. A. Angell, *Phys. Rev. Lett.*, 1994, **73**, 1632.
11 C. A. Angell, *Science*, 2008, **319**, 582–587.
12 S. Sastry, P. G. Debenedetti, F. Sciortino and H. E. Stanley, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 1996, **53**, 6144.
13 G. M. Schütz, C. Domb and J. Lebowitz, *Phase transitions and critical phenomena*, 2001.
14 H. Pathak, *et al.*, *Proc. Natl. Acad. Sci. U. S. A.*, 2021, **118**, e2018379118.
15 D. Fuentevilla and M. Anisimov, *Phys. Rev. Lett.*, 2006, **97**, 195702.
16 V. Holten and M. Anisimov, *Sci. Rep.*, 2012, **2**, 1–7.
17 O. Mishima and H. E. Stanley, *Nature*, 1998, **396**, 329–335.
18 O. Mishima and Y. Suzuki, *Nature*, 2002, **419**, 599–603.
19 J. C. Palmer, *et al.*, *Nature*, 2014, **510**, 385–388.
20 P. G. Debenedetti, F. Sciortino and G. H. Zerze, *Science*, 2020, **369**, 289–292.
21 J. A. Sellberg, *et al.*, *Nature*, 2014, **510**, 381–384.
22 K. H. Kim, *et al.*, *Science*, 2017, **358**, 1589–1593.
23 B. Cwilong, *Proc. R. Soc. A*, 1947, **190**, 137–143.
24 B. Mason, *Adv. Phys.*, 1958, **7**, 221–234.
25 D. Thomas and L. Staveley, *J. Chem. Soc.*, 1954, 4569–4577.
26 K. H. Kim, *et al.*, *Science*, 2020, **370**, 978–982.
27 S. Cerveny, F. Mallamace, J. Swenson, M. Vogel and L. Xu, *Chem. Rev.*, 2016, **116**, 7608–7625.
28 L. Liu, S.-H. Chen, A. Faraone, C.-W. Yen and C.-Y. Mou, *Phys. Rev. Lett.*, 2005, **95**, 117802.

29 P. Gallo, *et al.*, *Chem. Rev.*, 2016, **116**, 7463–7500.
30 D. Corradini, S. V. Buldyrev, P. Gallo and H. E. Stanley, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2010, **81**, 061504.
31 S. Woutersen, B. Ensing, M. Hilbers, Z. Zhao and C. A. Angell, *Science*, 2018, **359**, 1127–1131.
32 L. Kringle, W. A. Thornley, B. D. Kay and G. A. Kimmel, *Science*, 2020, **369**, 1490–1492.
33 E. B. Moore and V. Molinero, *Nature*, 2011, **479**, 506–508.
34 H. Tanaka, *Nature*, 1996, **380**, 328–330.
35 F. Smallenburg, L. Filion and F. Sciortino, *Nat. Phys.*, 2014, **10**, 653–657.
36 L. Xu, S. V. Buldyrev, C. A. Angell and H. E. Stanley, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2006, **74**, 031108.
37 L. B. Skinner, C. J. Benmore, J. C. Neuefeind and J. B. Parise, *J. Chem. Phys.*, 2014, **141**, 214507.
38 P. Tan, N. Xu, A. B. Schofield and L. Xu, *Phys. Rev. Lett.*, 2012, **108**, 095501.
39 H. Shen, P. Tan and L. Xu, *Phys. Rev. Lett.*, 2016, **116**, 048302.
40 J. Russo and H. Tanaka, *Nat. Commun.*, 2014, **5**, 1–11.
41 R. Shi and H. Tanaka, *J. Am. Chem. Soc.*, 2020, **142**, 2868–2875.
42 H. Tanaka, *J. Chem. Phys.*, 2000, **112**, 799–809.
43 R. Shi, J. Russo and H. Tanaka, *J. Chem. Phys.*, 2018, **149**, 224502.
44 A. K. Soper and M. A. Ricci, *Phys. Rev. Lett.*, 2000, **84**, 2881.
45 V. H. Segtnan, Š. Šašić, T. Isaksson and Y. Ozaki, *Anal. Chem.*, 2001, **73**, 3153–3161.
46 C. Yang, C. Zhang, F. Ye and X. Zhou, *Chin. Phys. B*, 2019, **28**, 116104.
47 C. Huang, K. T. Wikfeldt and T. Tokushima, *et al.*, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 15214–15218.
48 G. N. Clark, G. L. Hura and J. Teixeira, *et al.*, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 14003–14007.
49 J. Blobel, P. Bernado, D. I. Svergun, R. Tauler and M. Pons, *J. Am. Chem. Soc.*, 2009, **131**, 4378–4386.

