Supporting Information

Deciphering Charge Transfer and Hydrogen Bonding Characteristics from Liquid Water XAS Spectra

Alekos Segalina[a,b], Taehwan Jang[a,c], Minho M. Kim[a], Hyotcherl Ihee,[a,b] Jae Woo Park[d,], William A. Goddard III[c,*], and Hyungjun Kim[a,*]*

[a]Department of Chemistry, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea

[b]Center for Advanced Reaction Dynamics (CARD), Institute for Basic Science (IBS), Daejeon, 34141, Republic of Korea

[c]Materials and Process Simulation Center, California Institute of Technology (Caltech), Pasadena, California 91125, United States

[d]Department of Chemistry, Chungbuk National University (CBNU), Cheongju 28644, Republic of Korea

*Correspondence: (JWP) jaewoopark@cbnu.ac.kr ; (WAG) wag@caltech.edu; (HK) linus16@kaist.ac.kr

Computational Methods

Our proposed computational procedure for calculating the XAS spectra of water consists of: (i) conducting MD simulations of bulk water systems using various models (AIMD, RexPoN, and MB-pol(2023) in this case); (ii) sampling snapshots along the MD trajectories to achieve accurate and representative sampling of the structural manifold of room temperature bulk water; (iii) performing electronic structure calculations using multireference techniques to compute XAS spectra for individual snapshots; (iv) averaging these individual spectra over the ensemble of snapshots to obtain the thermal-averaged XAS spectrum; and finally, (v) analyzing the XAS spectra to establish structure-to-spectra relationships and evaluate the quality of the MD simulations. These procedures are described in detail in the following subsections.

(i) MD simulations and trajectories analysis

MB-pol(2023), hereafter referred to as MB-pol, is a water model that employs many-body interactions derived through the many-body expansion of the potential energy surface. Simulations for the MB-pol water model were performed using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) package with the MBX library embedded for MB-pol calculations. The MBX library parameters are provided in the mbx.json file. In the MB-pol simulations, 256 water molecules were included in the simulation cell. The pressure and temperature were maintained at 1 atm and 298 K, respectively, using the Parrinello–Rahman barostat and the Nosé–Hoover thermostat. Damping constants of 400 and 100 fs were used to control the pressure and temperature, respectively. Two-body interactions were considered within a 9 Å cutoff, whereas a 6 Å cutoff was employed for the three-body interactions. A total of 1 ns of simulation was performed with a 1 fs timestep, with the last 0.5 ns of the simulation used for analysis.

RexPoN is a water model that accurately describes intermolecular interactions by independently accounting for van der Waals (vdW) interactions, electrostatic interactions with polarization, and hydrogen bonding. The RexPoN simulations were conducted using LAMMPS. In the simulation of the RexPoN water model, 216 water molecules with a density of 0.9965 g/cm were included in the simulation cell. The temperature was maintained at 300 K using the Nosé-Hoover thermostat with a damping time of 100 fs. The O-H bond length and the H-O-H angle were constrained to their equilibrium values (0.9572 and 104.52°, respectively). The vdW and electrostatic interactions were evaluated with a 12 Å cutoff, whereas hydrogen bonding

interactions were calculated with a 4.6 Å cutoff. A total of 2 ns of simulation was performed with a 1 fs timestep, with the last 1 ns of the simulation used for analysis.

AIMD simulations of liquid water were performed using the CP2K package. The electronic structure and atomic forces were calculated using the self-consistent Kohn-Sham density functional theory (DFT) approach, with the van der Waals interactions incorporated via the rVV10 nonlocal functional with the *b*-parameter set to 9.3. This functional is an updated formulation by Vydrov and Van Voorhies that provides an accurate description of the structural properties of water under near-ambient conditions. In CP2K, the simulations employed a mixed Gaussian-plane-wave (GPW) approach, with the Kohn-Sham orbitals represented using a Gaussian basis set and the electron density represented by an auxiliary plane-wave basis set. Core-valence interactions were modeled using Goedecker–Teter–Hutter pseudopotentials, and a triple-ζ Gaussian basis set with polarization functions (TZV2P) was used. An 800 Ry cutoff was applied to the auxiliary plane-wave basis for electron density representation. The system comprised 64 water molecules enclosed in a cubic periodic simulation cell with an edge length of 12.42 Å. MD simulations were performed in the microcanonical (NVE) ensemble with canonical sampling through a velocity-rescaling scheme using a time constant of 50 fs to control the temperature. A time step of 0.5 fs was used to accurately capture the high-frequency O–H stretching mode. Out of the 20 ps of simulation performed, the final 15 ps were used for analysis.

The MD trajectories for the MB-pol(2023), RexPoN, and AIMD simulations, are publicly available on Materials Cloud at https://doi.org/10.24435/materialscloud:mh-6e.

(ii) Water cluster model sampling

After collecting the MD trajectories obtained using the different methods, we sampled snapshots according to the following procedure: First, 50 water molecules were selected randomly from each trajectory. For these water molecules, we extracted 20 random snapshots throughout the production run, including the selected water molecule and its first solvation shell, resulting in 1,000 snapshots per trajectory. The first solvation shell contained a variable number of water molecules as it included all molecules found within the average radius of the first solvation shell obtained from the MD simulation analysis. Because three different simulation methods (MB-pol, RexPoN, and AIMD) were employed in this study, the selection of the first solvation shell was based on the peak position of the $g_{O\text{-}O}$ obtained from each method. Accordingly, we used distinct average radii

of 3.47, 3.49, and 3.44 Å for MB-pol, RexPoN, and AIMD, respectively, to accurately reflect the structural characteristics captured by each approach. The extracted snapshots were categorized according to the number of molecules in each cluster and stored for the subsequent XAS spectral calculations.

This procedure ensured representative sampling of each trajectory, as the average number of water molecules in the first solvation shell from the selected snapshots closely matched the number of water molecules obtained by integration over the entire trajectory (Table S1). Furthermore, the ratios of the DD, SD, and ND water clusters (DD:SD:ND) obtained from the snapshots were consistent with those computed from the full trajectories (Table S3 and S4).

In-house Python scripts exploiting the MDAnalysis library were employed for snapshot extraction and hydrogen bond analyses. The extracted snapshots employed for the XAS spectral calculations are likewise publicly accessible at https://doi.org/10.24435/materialscloud:mh-6e.

(iii) Multireference calculations

All calculations were performed using the BAGEL code.[1] The final calculations utilized the aug-cc-pCVTZ-dk basis set.

To obtain accurate energy and oscillator strengths for the final XAS spectra, multireference calculations were carried out using the following procedure. First, we selected a random snapshot for each class, defined by the number of water molecules in the chosen water cluster. Initial Hartree–Fock (HF) calculations were performed, followed by orbital localization using the Pipek–Mezey scheme. From the localized orbitals, the orbitals related to the central water molecules were selected for each cluster class. The initial complete active space self-consistent field (CASSCF) (10,13) was conducted using the orbitals primarily localized on the central water molecule, incorporating all occupied orbitals and the first eight unoccupied orbitals. CASSCF calculations for all water clusters sampled using the three different MD methods were performed using the CASSCF wave function of the respective class as the initial estimate. Subsequently, we conducted state-averaged (SA) (8)-CASSCF(2,8) calculations, selecting an active space defined by the 1a1 orbitals and electrons as well as the $2b_2$, $2b_1$, $4a_1$, $5a_1$, $3b_2$, $3b_1$, and $6a_1$ virtual orbitals. Subsequently, the shapes of $2b_2$, $2b_1$, $4a_1$, and in some snapshots, the $5a_1$ orbitals remained recognizable, albeit with significant delocalization over the surrounding water molecules (Figure

S3). By contrast, the last three orbitals are predominantly localized on the molecules within the first solvation shell.

After this step, the SA-CASSCF wavefunctions were projected onto the larger complete active space configuration interaction (CASCI) (8,11) space, which included all occupied orbitals of the central water molecule, except the $4a_1$ orbital. During this step, the oscillator strengths of the roots related to core-to-valence excitations were also evaluated. Finally, the ground state roots and those corresponding to the core-to-valence excitations were corrected using state-specific NEVPT2 calculations. In this case, state-specific corrections are justified because the roots involved in the XAS spectra of water are well-separated in the energy space.

(iv) XAS spectra

The theoretical stick spectra derived from approximately 1000 snapshots for each MD simulation were broadened using Gaussian functions with a standard deviation of 0.20 eV. The resulting thermally averaged spectra were normalized to a maximum intensity of 1, and an energy shift was applied to align the pre-edge peak with the experimental spectrum. Experimental data for comparison were obtained from Ref. 2.

(v) XAS spectra analysis

The contribution of CT in XAS spectra was quantified by analyzing the electron density difference between the excited state (Ψ_1) and the ground state (Ψ_0). Both eigenvectors were derived from the CASCI wavefunctions. After defining the density difference in the molecular orbital (MO) basis, we applied Löwdin orthonormalization to transform it into the orthogonal atomic orbitals (OAO) basis. The resulting matrix is block-diagonal, with each block corresponding to a single water molecule. The trace of each block reflects the number of electrons removed or added in each water molecule during the transition from Ψ_0 to Ψ_1. This trace is then used to determine the charge transfer associated with the analyzed transition.

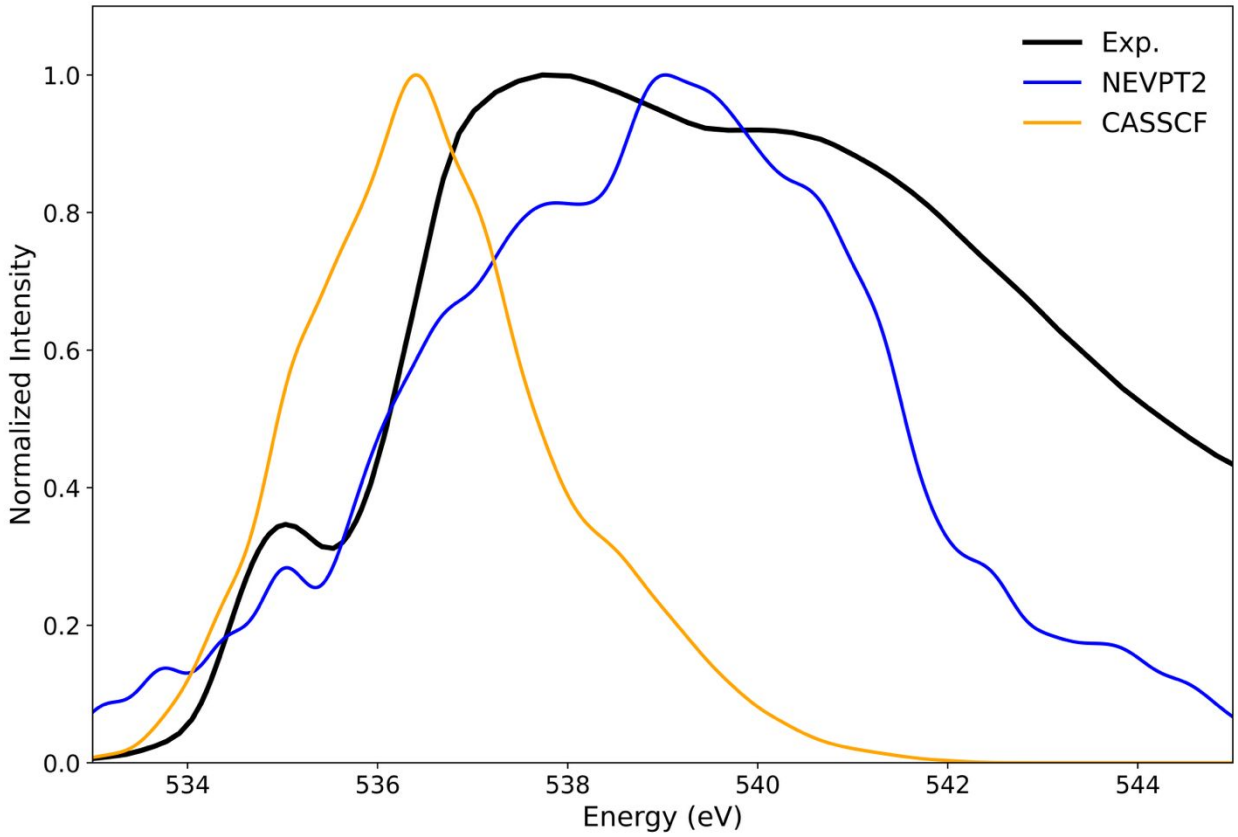

Figure S1. Comparison of XAS spectra computed from MB-pol trajectories at the CASSCF and NEVPT2 levels alongside the experimental spectrum (Exp.) The calculated spectra are based on 1000 snapshots, with the CASSCF spectrum shifted by 12 eV and the NEVPT2 spectrum by 6 eV. A Gaussian broadening of 0.20 eV was applied to both.

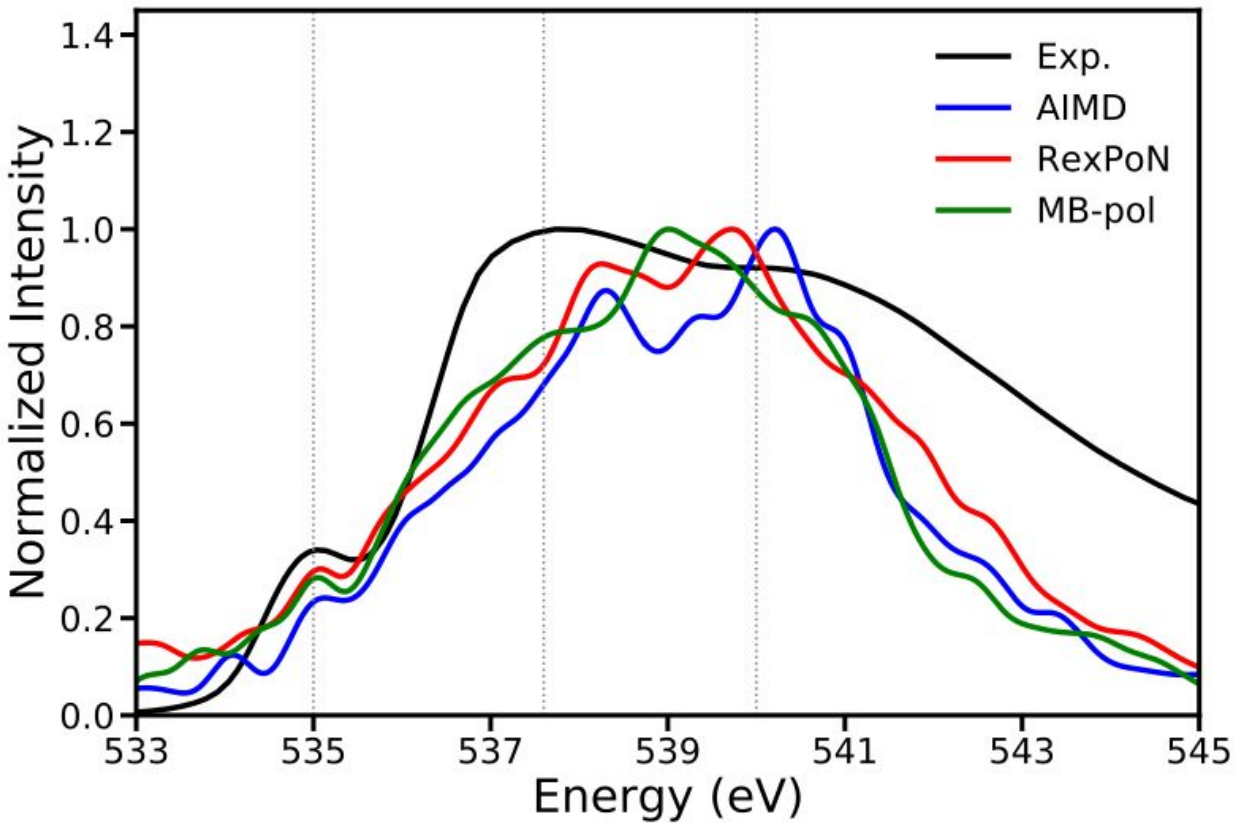

Figure S2. X-ray absorption spectra (XAS) of liquid water. Spectra calculated from snapshots sampled from the AIMD (blue), RexPoN (red), and MB-pol (green) trajectories are shown together with the experimental spectrum (black). The vertical dotted lines mark the characteristic features of the experimental spectrum. Each calculated spectrum was generated by convoluting vertical transitions with Gaussian functions using a full width at half-maximum (FWHM) of 0.20 eV.

XYZ Coordinates of DD:NA and ND:DA Water Clusters Extracted from Crystalline Ice Ih

9
DD:NA

O	0.000000	0.000000	0.000000
H	0.000000	0.997110	0.000000
H	0.480931	-0.345481	0.799836
O	0.014377	2.762505	0.023911
H	-0.466554	3.107985	-0.775925
H	-0.428723	3.080809	0.858522
O	1.321232	-0.949118	2.197339
H	1.655792	-0.192342	2.753746
H	0.710852	-1.507758	2.753746

9
ND:DA

O	0.000000	0.000000	0.000000
H	0.000000	0.997110	0.000000
H	0.480931	-0.345481	0.799836
O	-2.613376	-0.895491	0.023911
H	-1.668436	-0.577187	0.023912
H	-2.613375	-1.892602	0.023912
O	1.246578	-0.895491	-2.297030
H	0.803479	-0.577187	-1.462419
H	2.191520	-0.577186	-2.297030

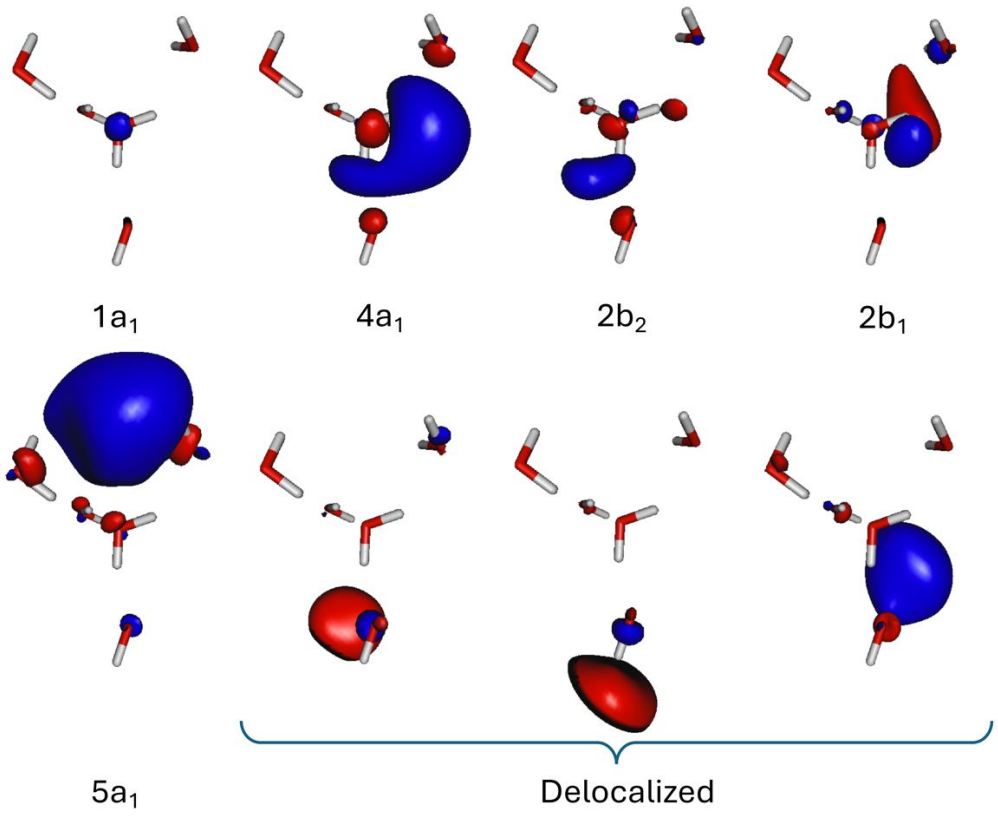

Figure S3. SA-CASSCF natural orbitals of the SA(8)-CASSCF(2,8) active space for a representative water cluster, extracted from MB-pol dynamics simulations comprising five water molecules.

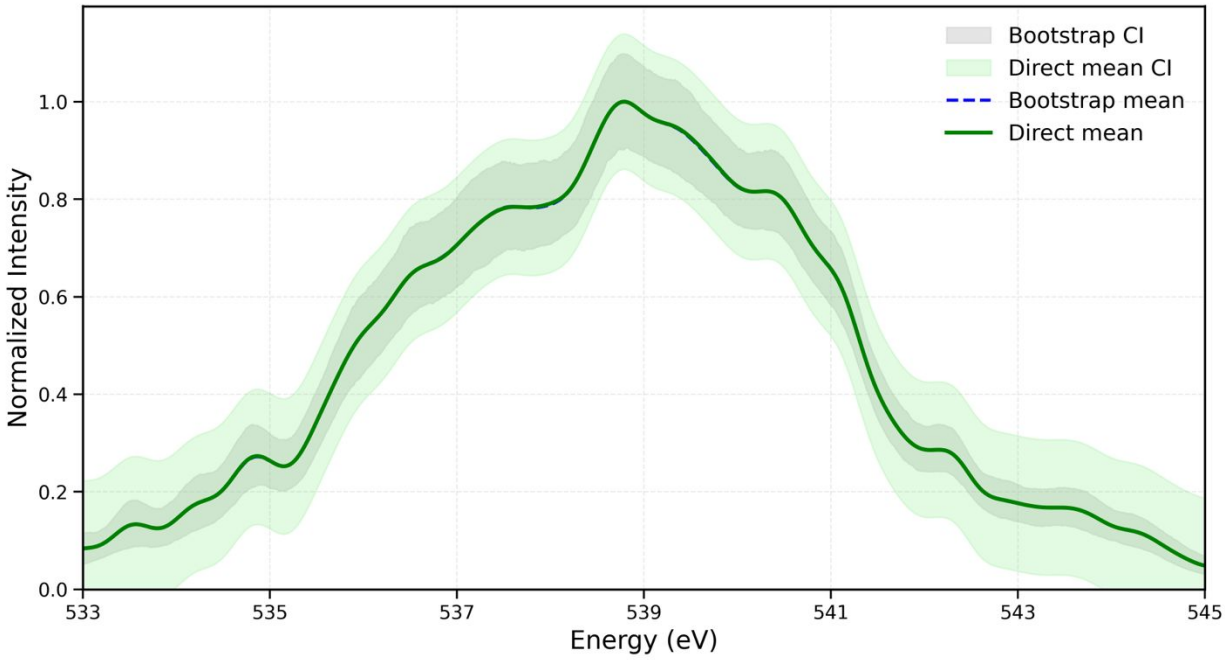

Figure S4. Comparison between spectra obtained from NEVPT2 calculations on MB-pol molecular dynamics snapshots (same dataset as in Figure 2 of the main text). The blue curve (bootstrap mean) and the green curve (direct average) nearly overlap, as expected, since bootstrap resampling does not alter the central estimate but instead quantifies its statistical uncertainty. Shaded regions represent 95% confidence intervals derived from bootstrap resampling, highlighting the robustness of the spectral features. Both spectra are normalized to unit peak intensity.

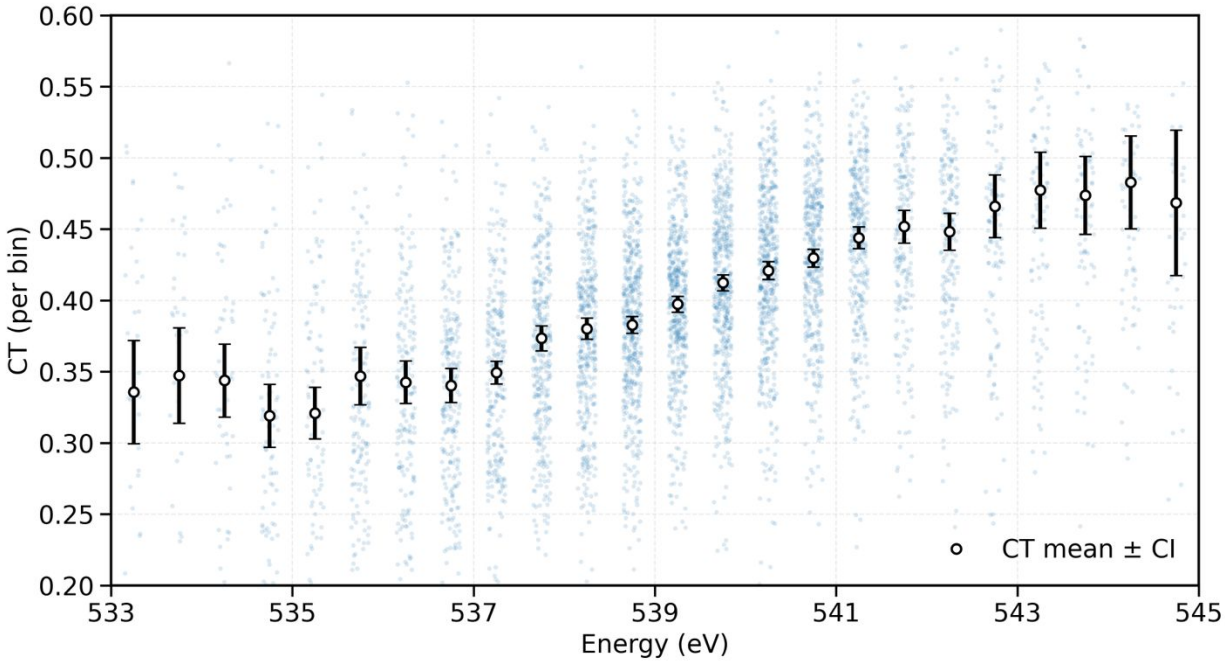

Figure S5. Charge-transfer (CT) values as a function of excitation energy. Light shaded points indicate individual transitions, while bold markers show the mean CT value within 0.5 eV bins (the same bin width used in Figure 4 of the main text). Vertical error bars represent 95% confidence intervals obtained from resampling, quantifying the statistical uncertainty of the averages.

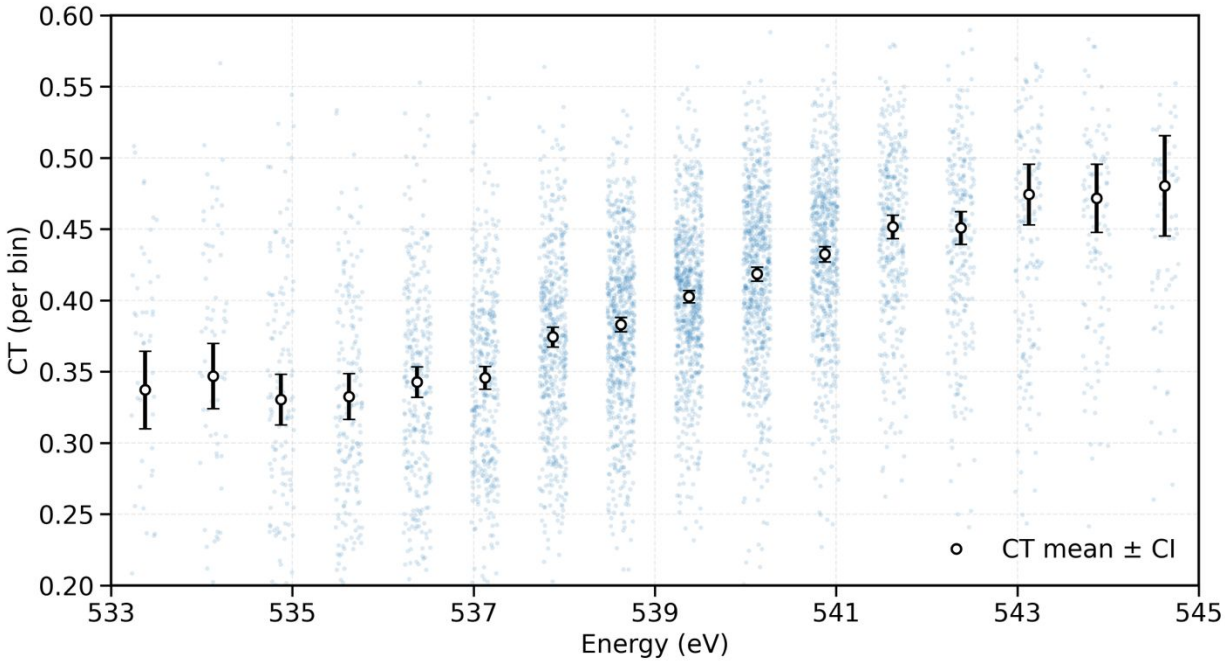

Figure S6. Charge-transfer (CT) values as a function of excitation energy. Light shaded points indicate individual transitions, while bold markers show the mean CT value within 0.75 eV bins. Vertical error bars represent 95% confidence intervals obtained from resampling, quantifying the statistical uncertainty of the averages.

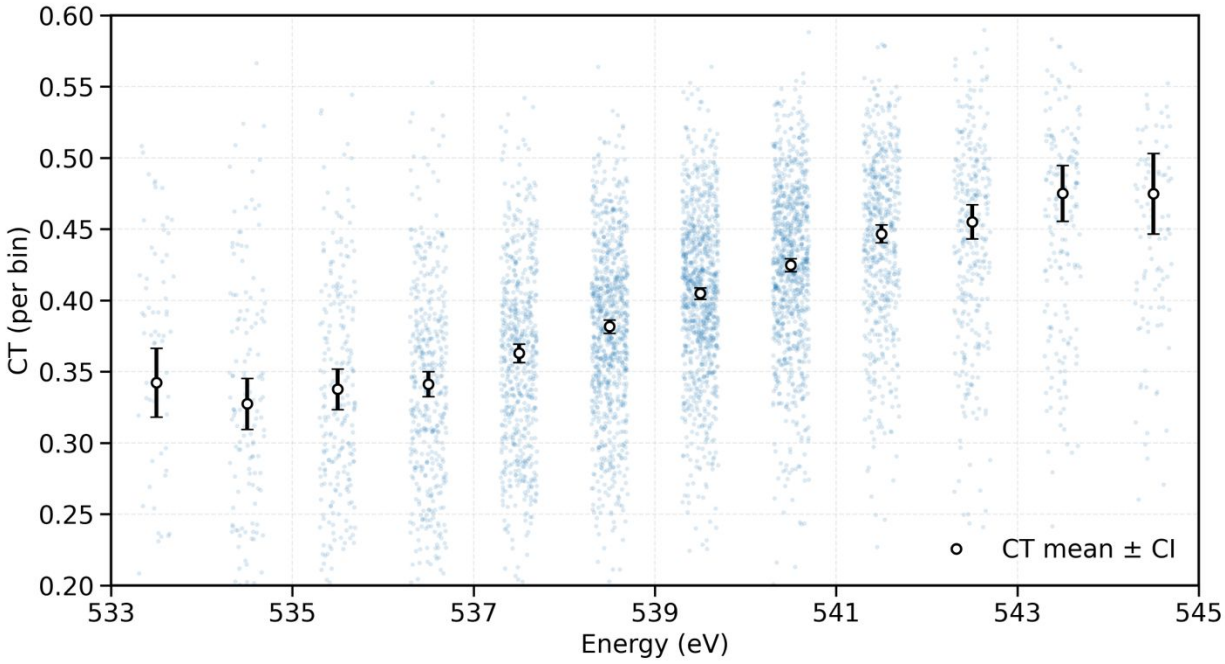

Figure S7. Charge-transfer (CT) values as a function of excitation energy. Light shaded points indicate individual transitions, while bold markers show the mean CT value within 1 eV bins. Vertical error bars represent 95% confidence intervals obtained from resampling, quantifying the statistical uncertainty of the averages.

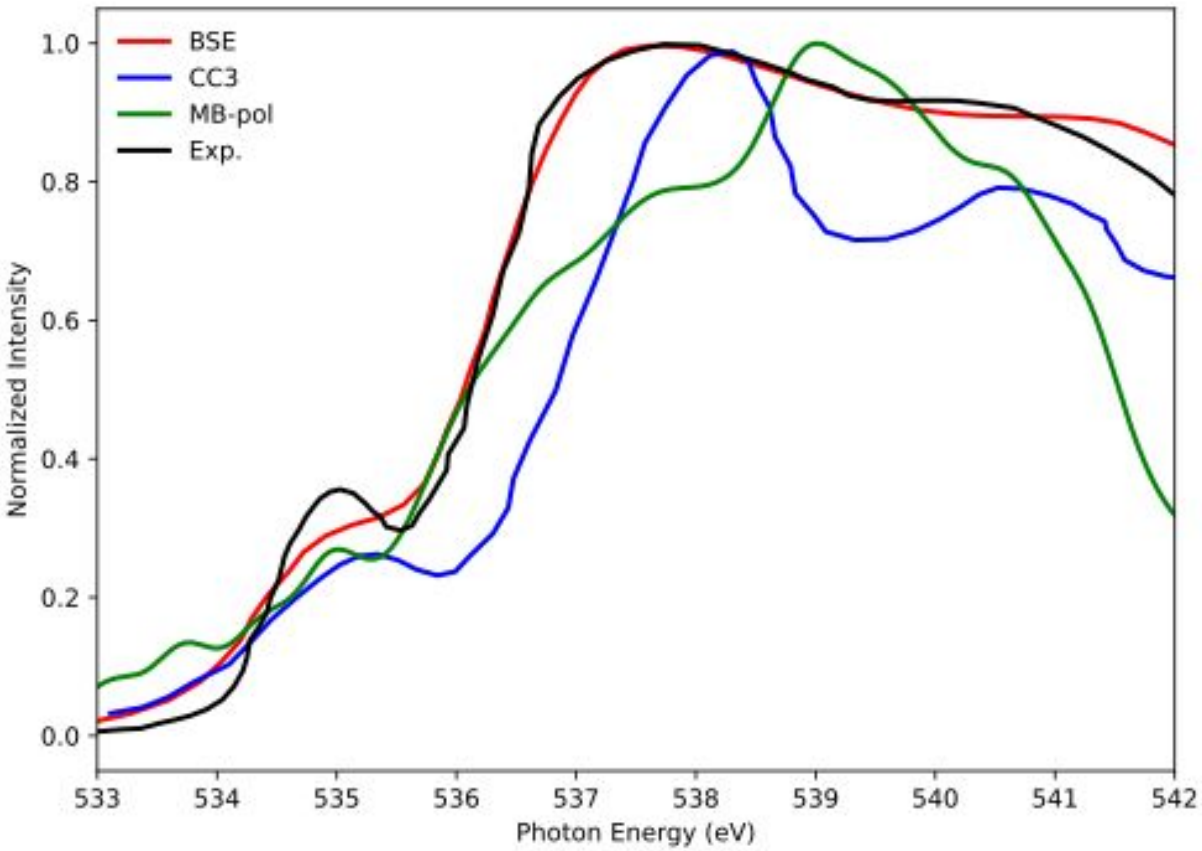

Figure S8. The spectra obtained at the GW-BSE[3] (red) and MLCC3[4] (blue) levels of theory, together with the NEVPT2 spectrum from our work based on MB-pol sampling (green), are shown alongside the experimental spectrum (black). All reference spectra were normalized to their respective maximum intensities and smoothed with a Gaussian filter ($\sigma = 1.0$ for the theoretical spectra and $\sigma = 0.5$ for the experiment) to reduce numerical noise and highlight spectral features. The energy window is restricted to 533–542 eV in order to focus on the pre-edge, main-edge, and post-edge regions.

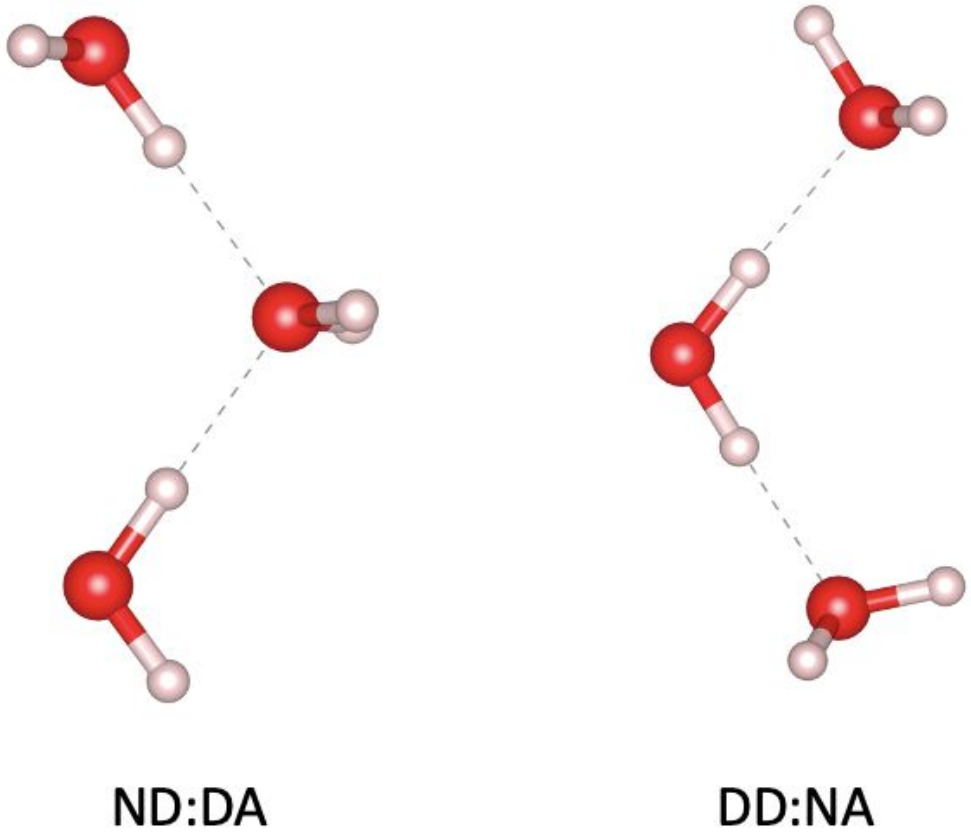

Figure S9. Water clusters extracted from Ice Ih, illustrating two distinct hydrogen bond arrangements: the Non-Donor Double Acceptor (ND:DA) arrangement in the left-hand panel and the Double Donor Non-Acceptor (DD:NA) arrangement in the right-hand panel.

Table S1. Comparison of the first-shell coordination number of water obtained from full trajectories versus those obtained from snapshots used in XAS spectra calculations together with the experimental value.

	Average over full trajectory	**Average over the 1000 sampled snapshots**
MB-pol	4.20	4.18
RexPoN	4.51	4.32
AIMD	3.90	4.00
Exp.	~4.50	

Table S2. Comparison of the O-O radial distance (Å) and intensity ($g_{O\text{-}O}$) for the 2nd peak obtained from three different MD simulations. The computed O-O radial distances (R) are given in Å together with the corresponding $g_{O\text{-}O}$ values as obtained from AIMD, RexPoN, and MB-pol simulations.

2nd solvation shell	**R (Å)**	**$g_{O\text{-}O}$**
AIMD	4.24	1.13
RexPoN	4.20	1.20
MB-pol	4.39	1.11
Exp.	4.65	1.12

Table S3. Percentages of hydrogen bond (HB) categories observed in the analyzed trajectories, covering all possible combinations of Double Donor (DD), Single Donor (SD), Non-Donor (ND), Double Acceptor (DA), Single Acceptor (SA), and Non-Acceptor (NA). Categories are aggregated based on their donor (DD, SD, ND) status (highlighted in bold).

	MB-pol	**RexPoN**	**AIMD**
DD:DA	37.05	23.47	30.36
DD:SA	20.76	20.27	20.74
DD:NA	1.23	4.84	1.74
TOT DD	**59.04**	**48.58**	**52.84**
SD:DA	16.29	17.96	17.50
SD:SA	16.06	17.07	19.78
SD:NA	1.60	4.38	2.36
TOT SD	**33.95**	**39.41**	**39.64**
ND:DA	1.42	3.83	1.68
ND:SA	2.49	3.87	3.29
ND:NA	0.42	1.06	0.63
TOT ND	**4.33**	**8.76**	**5.60**
Others	2.69	3.25	1.92

Table S4. Percentages of hydrogen bond (HB) categories observed in the snapshots used for the XAS spectra calculations. The analysis includes all possible combinations of Double Donor (DD), Single Donor (SD), Non-Donor (ND), Double Acceptor (DA), Single Acceptor (SA), and Non-Acceptor (NA). Categories are aggregated based on their donor (DD, SD, ND) status (highlighted in bold).

	MB-pol	**RexPoN**	**AIMD**
DD:DA	33.97	22.85	28.84
DD:SA	14.44	19.00	20.10
DD:NA	1.25	5.43	1.71
TOT DD	**49.66**	**47.28**	**50.65**
SD:DA	22.06	18.10	18.99
SD:SA	20.74	18.33	21.51
SD:NA	2.20	3.62	2.01
TOT SD	**45.00**	**40.05**	**42.51**
ND:DA	0.23	3.96	1.21
ND:SA	1.27	5.09	2.71
ND:NA	1.60	1.13	0.90
TOT ND	**3.10**	**10.18**	**4.82**
Others	2.24	2.49	2.02

Table S5. XAS lineshape calculated for a gas-phase water molecule (1 water molecule in vacuum) based on its optimized structure. The calculations were performed at the NEVPT2//SA-CASSCF level, utilizing two different active space configurations. After the initial SA-CASSCF procedure, CASCI calculations involving 8 electrons in 9 orbitals were carried out in both cases before applying NEVPT2 corrections. Experimental values from Ref. 5 are also provided for comparison (Exp.). Energies are given in eV, oscillator strengths are reported in parentheses, and the type of excitation is indicated in square brackets.

	NEVPT2//SA-CASSCF(2, 5)	**NEVPT2//SA-CASSCF(2, 8)**	**Exp.**
1st peak	531.96 (0.011) [A1 (1a1 → 4a1)]	530.76 (0.019) [A1 (1a1 → 4a1)]	534.0 (0.014)
2nd peak	532.45 (0.034) [1B2 (1a1 → 2b2)]	532.49 (0.022) [1B2 (1a1 → 2b2)]	535.9 (0.016)
3rd peak	538.83 (0.028) [1B1 (1a1 → 2b1)]	533.41 (0.018) [1B1 (1a1 → 2b1)]	537.1 (0.012)
2nd peak -1st peak	0.49 (0.023)	1.73 (0.003)	1.9 (0.002)
3rd peak -2nd peak	6.38 (-0.006)	0.92 (-0.004)	1.2 (-0.004)

Table S6. Spectral parameters from EXP, AIMD, RexPoN, and MB-pol

Summary of spectral parameters for the pre-edge, main edge, and post-edge regions extracted from Exp., AIMD, RexPoN, and MB-pol spectra. For each region, the peak energy (PeakE), full width at half maximum (FWHM), second central moment (M2), and the corresponding root-mean-square width (σE) are reported.

Method	Region	PeakE (eV)	FWHM (eV)	M2	σE
	pre-edge	535.040	0.489	0.427	0.690
Exp.	main edge	537.820	1.596	1.765	0.776
	pre-edge	534.092	0.333	0.556	0.602
AIMD	main edge	538.291	0.485	1.424	0.735
	pre-edge	535.012	0.167	0.715	0.643
RexPoN	main edge	538.086	0.793	1.296	0.735
	pre-edge	535.005	0.265	0.562	0.620
MB-pol	main edge	537.684	0.305	1.126	0.754
	post-edge	538.880	1.137	0.846	0.580

Table S7. Relative intensity ratio I(535)/I(537.8) as a function of the broadening factor σ (eV). For each method (AIMD, RexPoN, MB-pol) the values obtained using Gaussian (G), Lorentzian (L), and Voigt (V) line-shape profiles are reported. Values for experimental spectra are shown for reference. The reported σ values correspond to the broadening factor used for the Gaussian profile. For the Lorentzian profile, the full width at half maximum (FWHM = γ) was defined as $\gamma = 2.35482 \times \sigma$. For the Voigt profile, both parameters were set consistently to (σ, γ) using the same relation for γ.

σ (eV)	**Exp.**	**AIMD-G**	**AIMD-L**	**AIMD-V**	**RexPoN-G**	**RexPoN-L**	**RexPoN-V**	**MB-pol-G**	**MB-pol-L**	**MB-pol-V**
0.15	**0.363**	0.365	0.370	0.358	0.462	0.478	0.423	0.380	0.393	0.383
0.20		0.349	0.360	0.347	0.410	0.443	0.407	0.363	0.388	0.378
0.25		0.334	0.355	0.342	0.388	0.427	0.404	0.352	0.387	0.380
0.30		0.319	0.354	0.341	0.377	0.421	0.407	0.345	0.388	0.385
0.35		0.308	0.355	0.345	0.371	0.420	0.411	0.342	0.391	0.392

Table S8. Relative intensity ratio I(537.8)/I(540) as a function of the broadening factor σ (eV). For each method (AIMD, RexPoN, MB-pol) the values obtained using Gaussian (G), Lorentzian (L), and Voigt (V) line-shape profiles are reported. Values for experimental spectra are shown for reference. The reported σ values correspond to the broadening factor used for the Gaussian profile. For the Lorentzian profile, the full width at half maximum (FWHM = γ) was defined as $\gamma = 2.35482 \times \sigma$. For the Voigt profile, both parameters were set consistently to (σ, γ) using the same relation for γ.

σ (eV)	**Exp.**	**AIMD-G**	**AIMD-L**	**AIMD-V**	**RexPoN-G**	**RexPoN-L**	**RexPoN-V**	**MB-pol-G**	**MB-pol-L**	**MB-pol-V**
0.15	**0.859**	1.399	1.368	1.357	1.234	1.179	1.186	1.065	1.064	1.086
0.20		1.409	1.346	1.335	1.223	1.176	1.165	1.080	1.075	1.094
0.25		1.400	1.331	1.310	1.204	1.168	1.148	1.092	1.081	1.098
0.30		1.379	1.317	1.287	1.186	1.160	1.135	1.103	1.084	1.098
0.35		1.356	1.305	1.268	1.170	1.151	1.125	1.111	1.086	1.096

Table S9. Calculated unshifted excitation energies for the first six roots of water clusters extracted from ice Ih. Two configurations, namely ND:DA and DD:NA, are shown. Values in parentheses represent the oscillator strength, and those in square brackets denote the calculated CT value. The water clusters used for these calculations are depicted in Figure S2.

	ND:DA	**DD:NA**
1st root	531.93 (0.031) [0.35]	532.45 (0.048) [0.67]
2nd root	532.78 (0.047)[0.28]	536.02 (0.001)[0.64]
3rd root	534.42 (0.009)[0.43]	534.37 (0.001)[0.63]
4th root	538.91 (0.001)[0.37]	537.81 (0.019)[0.62]
5th root	540.21 (0.010)[0.54]	534.54 (0.001)[0.71]
6th root	548.13 (0.000)[0.51]	548.73 (0.001)[0.68]

References

(1) Shiozaki, T. BAGEL: Brilliantly Advanced General Electronic-Structure Library. *WIREs Comput. Mol. Sci.* **2018**, *8* (1), e1331. https://doi.org/10.1002/wcms.1331.

(2) Wernet, Ph.; Nordlund, D.; Bergmann, U.; Cavalleri, M.; Odelius, M.; Ogasawara, H.; Näslund, L. Å.; Hirsch, T. K.; Ojamäe, L.; Glatzel, P.; Pettersson, L. G. M.; Nilsson, A. The Structure of the First Coordination Shell in Liquid Water. *Science* **2004**, *304* (5673), 995–999. https://doi.org/10.1126/science.1096205.

(3) Tang, F.; Li, Z.; Zhang, C.; Louie, S. G.; Car, R.; Qiu, D. Y.; Wu, X. Many-Body Effects in the X-Ray Absorption Spectra of Liquid Water. *Proc. Natl. Acad. Sci.* **2022**, *119* (20), e2201258119. https://doi.org/10.1073/pnas.2201258119.

(4) Folkestad, S. D.; Paul, A. C.; Paul (Née Matveeva), R.; Coriani, S.; Odelius, M.; Iannuzzi, M.; Koch, H. Understanding X-Ray Absorption in Liquid Water Using Triple Excitations in Multilevel Coupled Cluster Theory. *Nat. Commun.* **2024**, *15* (1), 3551. https://doi.org/10.1038/s41467-024-47690-x.

(5) Ishii, I.; McLaren, R.; Hitchcock, A. P.; Robin, M. B. Inner-shell Excitations in Weak-bond Molecules. *J. Chem. Phys.* **1987**, *87* (8), 4344–4360. https://doi.org/10.1063/1.452893.