PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2024, **26**, 1845

Core-ionization spectrum of liquid water†

Sourav Dey, [a] Sarai Dery Folkestad, [a] Alexander C. Paul, [b] Henrik Koch [b] and Anna I. Krylov *[a]

Received 30th May 2023,
Accepted 10th November 2023

DOI: 10.1039/d3cp02499g

rsc.li/pccp

We present state-of-the-art calculations of the core-ionization spectrum of water. Despite significant progress in procedures developed to mitigate various experimental complications and uncertainties, the experimental determination of ionization energies of solvated species involves several non-trivial steps such as assessing the effect of the surface potential, electrolytes, and finite escape depths of photoelectrons. This provides a motivation to obtain robust theoretical values of the intrinsic bulk ionization energy and the corresponding solvent-induced shift. Here we develop theoretical protocols based on coupled-cluster theory and electrostatic embedding. Our value of the intrinsic solvent-induced shift of the $1s_O$ ionization energy of water is −1.79 eV. The computed absolute position and the width of the $1s_O$ peak in photoelectron spectrum of water are 538.47 eV and 1.44 eV, respectively, agreeing well with the best experimental values.

I. Introduction

Water is the most important substance on Earth. It is essential for life to exist – the working definition of a planet capable of sustaining life includes the presence of liquid water.[1] Water is a natural environment for biochemical, geophysical, environmental, and many technological processes. Hence, understanding properties of water on a molecular level is a prerequisite for understanding how water influences and drives chemistry. Yet, despite a plethora of experimental and theoretical studies, water continues to puzzle scientists.[2–4]

In this contribution, we focus on the most basic property of water, that is, its electronic structure. The key element of electronic structure is the shapes and energies of molecular orbitals, which describe the states of electrons and ultimately determine chemical properties of a substance. Molecular orbitals can be probed by photoelectron spectroscopies, connecting theory with the experiment.[5,6] An important question is how molecular orbitals are affected by the environment, that is, in which way a water molecule in bulk water differs from an isolated water molecule.[7] Because molecular orbitals are sensitive to the intermolecular interactions, their energies can also provide information about local solvent structure and its fluctuations.[8]

Photoelectron spectroscopy using microjets is a tool to interrogate electronic structure of solvated species, including that of the bulk solvent itself.[9] It has been applied to study various aqueous solutions, both in the UV-Vis and X-ray regimes.[7,10–16] Because the light beam has a finite width, it ionizes both molecules in the microjets and gaseous molecules around it, giving rise to the spectra containing two well-separated features – a narrow gas-phase peak and a broader band corresponding to the liquid. Superficially, these experiments appear to be a straightforward extension of the gas-phase photoelectron experiments[17–21] in which the ionization energy (IE) is given by the difference between the energy of the photon ($h\nu$) and a measured kinetic energy (KE) of the ejected electrons:

$$\mathrm{IE}_{\mathrm{gas}} = h\nu - \mathrm{KE}. \quad (1)$$

However, the quantitative interpretation of these experiments is more difficult, as explained in detail by Olivieri *et al.*[13] and more recently by Thürmer *et al.*[16] The essential difference is that in bulk measurements, the kinetic energy of ejected photoelectrons is affected by the interactions with the surface. In addition, the interpretation of experimental spectra is affected by various parasitic fields, which are not present in the gas-phase experiments, and uncertainties in calibration.[16]

The energy diagram in Fig. 1 explains this issue. Whereas the Fermi level (E_F) of the liquid sample (*e.g.*, microject) and the analyzer are matched by design, their vacuum levels (defined as zero kinetic energy of photoelectrons) do not match. The net result is that in order to extract the true IE from microjet experiments,[13] one needs to know the difference between the workfunctions of the liquid and the analyzer, ϕ_{wat} and ϕ_{ana},

$$\mathrm{IE}_{\mathrm{wat}} = h\nu - \mathrm{KE} + (\phi_{\mathrm{wat}} - \phi_{\mathrm{ana}}). \quad (2)$$

[a] *Department of Chemistry, University of Southern California, Los Angeles, CA 90089, USA. E-mail: krylov@usc.edu*
[b] *Department of Chemistry, Norwegian University of Science and Technology, 7491 Trondheim, Norway*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3cp02499g

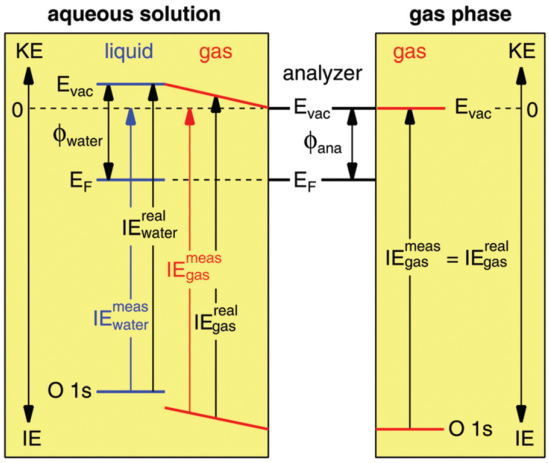

Fig. 1 Energy-level diagram for photoelectron experiments in a gas phase (right panel) and a liquid jet (left panel). Abbreviations: E_{vac}, vacuum level; E_F, Fermi energy; ϕ, work function; ana, analyzer; KE, kinetic energy; IE, ionization energy. The vacuum level (*i.e.*, the level corresponding to the ejected electrons with zero kinetic energy) in the liquid jet is higher than the vacuum level of an isolated molecule in gas phase due to the presence of the field created by the water surface. This field also affects the gas-phase molecules in the vicinity of the jet. Whereas in the gas-phase experiment the vacuum level of the gas matches that of the analyzer, in the aqueous solution experiment the vacuum level of the gas is pinned to the vacuum levels of the liquid and of the analyzer, which results in an electric field between the two (depicted as sloped energy levels of the gas). The Fermi levels of the jet and the analyzers match by design (they are both grounded), but their vacuum levels do not because of the difference in the respective workfunctions. Zero kinetic energy is defined for both experiments as the vacuum level of the analyzer. Reproduced from ref. 13 with permission from the Royal Society of Chemistry.

Proper accounting for this difference is not trivial[22] and different approaches have been developed to address referencing and calibration issues.

Importantly, the jet also affects the energy levels of the gas-phase molecules near it, so the measured IEs of the gas-phase molecules in microjet experiments differ from the true gas-phase IEs:[13]

$$\mathrm{IE}_{\mathrm{gas}} = h\nu - \mathrm{KE} - c(\phi_{\mathrm{wat}} - \phi_{\mathrm{ana}}). \quad (3)$$

Here c is a geometric factor, which depends on the details of the experimental setup – *i.e.*, which area around the microjet is probed by the beam – usually, c varies between 1 and 0.5.[13]

This contribution due to the workfunction differences, which gives rise to the vacuum levels mismatch between the jet and the analyzer and which was neglected in early experiments, is responsible for noticeable discrepancies in the reported IEs (as illustrated by the water core-level IEs below). Olivieri *et al.*[13] introduced a procedure designed to extract the true $\mathrm{IE}_{\mathrm{wat}}$ from the microjet experiments by applying a variable bias to the microjet. The bias shifts the energy levels of the liquid and by measuring the dependence of the measured IEs on the applied bias the true IEs can be extracted. By applying this technique to core ionization of water (specifically, 0.05 M solution of NaCl), the authors[13] determined that in their experiment ($\phi_{\mathrm{wat}} - \phi_{\mathrm{ana}}$) equals +0.57 (±0.07) eV and the geometric factor is c = 0.70 (±0.05) eV V^{-1}. This means that the bulk and gas-phase peak positions from the zero-bias measurement need to be shifted by 0.57 eV and by 0.40 eV, respectively. The value of 0.57 eV is consistent with the optimal bias of +0.5 eV determined in ref. 13.

Thus, the workfunctions difference affects not only the absolute value of the bulk IE, but also the shift of the bulk IE relative to the gas-phase IE (ΔIE). Therefore, the true ΔIE cannot be taken as a difference between the gas-phase and liquid peaks' maxima from an unbiased microjet spectrum. This effect on gas-phase molecules around the jet is clearly seen from the shift of the gas-phase peak relative to the true gas-phase IE – in ref. 13 this shift equals 0.41 eV; the microjet value is red-shifted, as expected from Fig. 1 and from the theoretical value of the interface potential[23] (illustrated in Fig. 2).

Thürmer *et al.* developed an alternative approach to eliminate referencing the bulk IEs to the perturbed gas-phase solvent peak[16] – they determine the absolute bulk IEs as the difference between the peaks in the photoelectron spectrum (*i.e.*, measured photoelectron kinetic energy) relative to the cutoff value defined as the slowest photoelectrons emerging from the liquid jet. These low-energy photoelectrons are photoelectrons that lost nearly all their kinetic energy due to inelastic scattering and have minimal energy to escape the liquid. To reveal this intrinsic onset of the photoelectron spectrum, a negative accelerating bias is applied.[16] This approach does not rely on the gas-phase peak and does not require quantification of various parasitic fields. Thürmer *et al.* also introduced an additional procedure aiming at determining the solution workfunction.[16] They reported[16] the following values of the $1s_O$ IE and ϕ_{wat}: 538.10 ± 0.05 eV and 4.73 ± 0.09 eV, to be compared with 538.21 ± 0.07 eV and 4.65 ± 0.09 eV by Olivieri *et al.*[13] Despite using different protocols, the two sets of values agree with each other within the reported error bars.

A recent paper[24] focused on another technique for measuring bulk IEs based on ambient-pressure experiments using

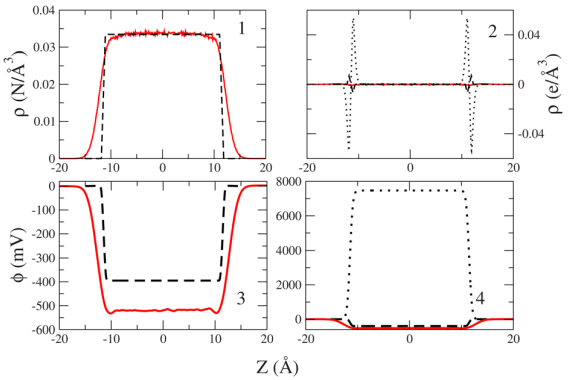

Fig. 2 Molecular density ($\rho(N)$), charge density ($\rho(e)$), and potential (denoted as ϕ in this figure and as φ_{in} in the text) along the interface normal of the vacuum–water system computed using several water models. Reprinted from ref. 23 with permission of AIP Publishing.

water films (dip-and-pull method) instead of microjets and discussed the application of bias for determining the true bulk IEs from the photoelectron measurements. The authors estimated bias required to align the vacuum levels of the liquid and the analyzer to be equal +0.435 V (*versus* the standard hydrogen electrode), which is close to the estimated optimal bias in the microjet experiment.[13] However, the extracted values of the core IEs of water were different.

In microscopic terms, the real IE of solvated species includes contribution of the interface potential, φ_{in} (Galvani potential). The difference between φ_{in} and φ_{out} (potential outside the sample called Volta potential) is called surface potential χ. For idealized neutral solutions, the outer potential is zero, such that $\chi = \varphi_{in}$; however, the exact value of φ_{out} depends on the position of the surface defining the interface. The net result is that in order to obtain the intrinsic, bulk IE, one should account for the interface potential:

$$IE_{wat} = h\nu - KE - \varphi_{in}. \quad (4)$$

Experimentally, the real IE is measured, *i.e.*, including contributions of φ_{in}. Thus, a special care is needed to make meaningful comparisons with the theoretical values (which may or may not include the interface potential).

Fig. 2 shows the results of simulations of water illustrating the electrostatic potential arising solely due to the interface. Hence, in order to reach the vacuum level (lowest free-electron state), the electrons need to overcome the electric field created by the liquid/gas interface. Surface potential also contributes to the solvation free energy, *e.g.*, for a particle with charge q

$$\Delta G_{real} = \Delta G_{intr} - q\varphi_{in}, \quad (5)$$

where ΔG_{real} and ΔG_{intr} are real and intrinsic (*i.e.*, bulk value without interface) Gibbs free energies of solvation (we use the label "real" for consistency with previous work[23]).

The microscopic origin of the interface potential and different ways to compute it have been extensively discussed.[23,25–28] Importantly, contrary to earlier proposals, surface potential does not arise due to ordered dipoles, but is dominated by quadrupole terms. For example, by using atomistic simulations Harder and Roux[23] have illustrated that the interface potential arises not due to orientational ordering of molecules on the surface, but due to the intramolecular asymmetry of the charge distribution. By using molecular dynamics (MD) simulations, they estimated the interface potential of water to be equal to −510 meV. An *ab initio* simulation by Mundy and co-workers[25] yielded a smaller value, −18 meV. The negative sign means that the solvation energy of anions becomes less negative and the solvation energy of cations becomes more negative due to the interface (as per eqn (5)), and that the apparent IE of the solvated neutral species is red-shifted relative to the intrinsic bulk IE (as per eqn (4)).

We note that the definition and the value of the interfacial potential depend on the type of experiment, as has been explained in ref. 26: *i.e.*, the magnitude and even the sign of water's surface potential differs between electrochemical experiments (which are relevant to the photoionization experiments) and high-energy electron holography measurements. Despite these difficulties faced by theory and experiment, the literature seems to converge on the existence of a negative effective potential for a single ion moving from the gas phase into liquid water of roughly −0.4 V[27] or less.

In addition to the effect of the interface, the measured IEs of bulk water can also be affected by the presence of solutes. Both microjet and dip-and-pull experiments are carried out using electrolyte solutions (*i.e.*, NaCl, KOH) to mitigate the effect of creating charges in the jet (due to streaming through the capillary and ionization) that can affect the subsequent ionization. The possible effect of ions present in these solutions on the IE of the bulk solvent was shown to be small[29] but non-negligible. Importantly, the ions affect both the intrinsic bulk IE and the surface potential. Finally, the measured IEs always contain contributions from both the surface and the bulk species because the photoelectrons have finite escape depth (determined by the inelastic mean free path). The contribution of the bulk increases at higher photon energies. The simulations[24] suggest that the apparent peak position of liquid water can vary by as much as 0.3 eV as the inelastic mean-free path changes from 3 to 50 Å.

In this contribution, we focus on the intrinsic $1s_O$ IE of bulk water. The gas-phase value is well-known:[13,30,31] 539.82 (±0.02) eV (value taken from ref. 13). The question is then what is the magnitude of the shift of the $1s_O$ level in the bulk relative to the gas-phase IE (ΔIE). As in the valence domain, the IEs of the solvated neutral species are expected to red-shift relative to the gas phase due to the strong solvent stabilization of the resulting cationic state. The magnitude of the shift quantifies the overall effect of the solvent stabilization of the core-hole state, whereas the width of the peak gives a measure of the solvent's structural fluctuations. In contrast to the valence domain, the photoelectron spectra of the core-level ionization do not contain multiple overlapping bands and, therefore, are somewhat easier to interpret. Core-level states are also known to be very sensitive to the local environment, which can provide a handle for connecting the spectroscopic measurements with the structure.

Table 1 summarizes experimental values of ΔIE for $1s_O$ ionization of liquid water from different experimental setups (microjets, clusters, dip-and-pull). These papers, which span the time range from 1986 till 2023, report values from −1.3 to −2.8 eV. To further illustrate the discrepancies between different experiments, Fig. 3 shows three spectra from ref. 12, 13 and 24. These discrepancies are quite substantial, providing the illustration of the challenges in experimental determination of IEs due to the factors discussed above. They also illustrate the progress made by the experimental community in addressing these challenges. In particular, the best experimental estimates derived from microjet experiments are −1.61 eV (ref. 13) and −1.72 eV (ref. 16) agree with each other within the specified error bars. In contrast to the microjet experiments, the best estimate from "dip-and-pull" experiments is −2.2 eV (ref. 24).

Given the challenges of the experimental determination of the absolute values of bulk IEs and the surface potential, and persistent disagreements[13,16,46,47] about the details of

Table 1 Experimental values of the $1s_O$ level shift (ΔIE) in bulk water. The best values are shown in bold (see text)

ΔIE, eV	Photon energy, eV	Details	Source
−1.3	560	Large water cluster (~1000 molecules)	32
−1.3	720	Water layer on SiO_2 film	33
−1.5	800	Liquid jet, 3 M $NaNO_2$/$NaNO_2$	34
−1.5	560	Liquid nanoparticles, 0.01 M glycerol	35
−1.58	420	Liquid jet, 0.05 M NaCl, zero bias	13
−1.61	**420**	**Liquid jet, 0.05 M NaCl, corrected**	13
−1.6	750	Liquid jet, 10^{-4} M NaCl, 0.003 mbar, 25 °C	12
−1.72	**650**	**Liquid jet, 0.05 M NaCl, from absolute IEs**[a]	16
−1.77	600	Liquid jet	11
−1.8	735	Ice	36
−1.9	1486	7 mol% LiCl, thin film	37
−1.9	1253	Liquid jet, 1 M KCl	14
−1.9	750	Liquid jet, 10^{-4} M NaCl, 1.4 mbar, 4 °C	12
−1.91	1012	Liquid jet, 0.14 M NaCl	15
−2.0	4000	Dip-and-pull, Pt/1 M KOH	38
−2.0	4000	Dip-and-pull, Pt/1 M KOH and 0.1 M KF	39
−2.0	4000	Dip-and-pull, Pt/0.1 M KOH	40
−2.0	4000	Dip-and-pull, Co/0.1 M KOH	41
−2.2	4000	Dip-and-pull, Co_xO/1 M KOH	42
−2.2	**5400**	**Dip-and-pull, Pt/1 M KF**	24
−2.3	4000	Dip-and-pull, NiFe/0.1 M KOH	43
−2.4	4000	Dip-and-pull, NiFecoCeO_x/1 M KOH and 0.1 M KF	44
−2.8	4000	Dip-and-pull, Pt/6 M KF	45

[a] Computed using the reported absolute $1s_O$ IE (538.10 eV) of liquid water and the gas-phase value (539.82 eV).

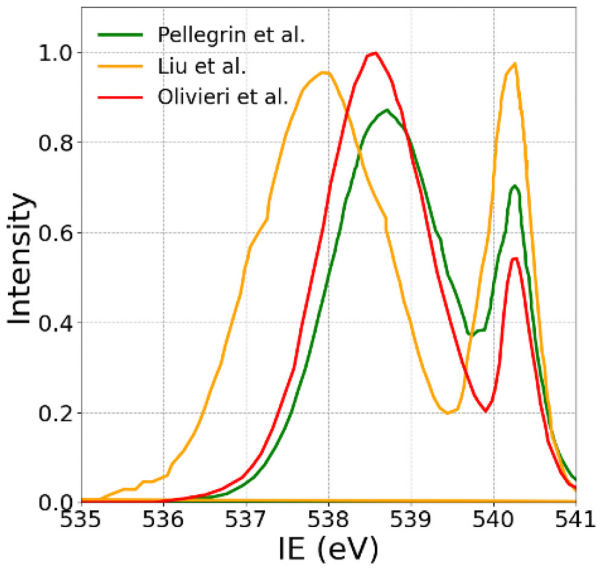

Fig. 3 Experimental spectra from Pellegrin *et al.* (at 0.003 mbar and room temperature),[12] Liu *et al.*,[24] and Olivieri *et al.* (at 0 V bias).[13] The spectra were aligned by the position of the gas-phase peak (the narrow feature) by applying global shifts of +0.59 eV, +4.9 eV, and +0.87 eV, respectively.

experimental protocols, accurate theoretical modeling of the core-ionization spectrum of water is important for providing a robust theoretical reference.

Previous theoretical calculations of core-level ionization of liquid water have been limited to density functional theory (DFT) and Hartree–Fock methods,[11,24,34,48] and varied greatly in terms of model structures and sampling of equilibrium dynamics (many were carried out on model clusters rather than bulk).

Here we employ high-level quantum chemistry methods to compute intrinsic core-level IE of bulk water. In addition to providing the best theoretical estimate of the bulk IE, we also aim to carefully investigate the convergence of the spectrum with respect to the details of computational protocol, to aid future theoretical studies. We use MD simulations with classical forcefields and *ab initio* potentials to simulate bulk water, and then use the snapshots from the MD simulations to carry out QM/MM (quantum mechanics/molecular mechanics) calculations of IEs using equation-of-motion coupled-cluster (EOM-CC) methods[49] adapted for calculations of core-level states[50,51] by core-valence separation (CVS).[52] EOM-CC is a state-of-the art technique capable of treating electronically excited and ionized species.[49] Here we go beyond EOM-CC with single and double excitations (EOM-CCSD) and also evaluate the effect of triple excitations by using the MLCC3 method (multilevel coupled-cluster method with triple excitations).[53–56] The main challenge in applying these methods to modeling condensed-phase phenomena is how to properly account for the effect of the solvent *via* embedding. Here we show that in calculations of core-level IEs simple electrostatic embedding (QM/MM)[57,58] converges slowly with respect to the size of the QM system, which illustrates the high sensitivity of the core-level states to the environment. We also investigate the contribution of different types of structures present in bulk water in the overall spectrum. Our calculations represent the most ambitious simulations that provide a reliable *ab initio* estimate of the core-ionization spectrum of liquid water.

The structure of the paper is as follows: the next section describes the details of computational protocols and Section III presents the results of the simulations of core-level IE of bulk water. In Conclusions, we outline the limitations of the current simulations and provide suggestions for future studies.

II. Computational details

The simulations of core-ionization spectra include two steps: equilibrium simulations of bulk water and the calculations of the IEs using the snapshots from the equilibrium simulations. Electronic structure and AIMD calculations were carried out using the Q-CHEM and eT electronic structure packages,[59–61] and MD simulations were carried out using GROMACS.[62]

A. Equilibrium simulations of bulk water

The equilibrium simulations were carried out using classical MD with TIP3P[63] waters and with *ab initio* molecular dynamics

(AIMD). In AIMD simulations, we used QM/MM scheme with the QM waters described by ωB97X-D/6-31G* and MM waters described by TIP3P.

The MD simulations were set up as follows. First, we optimized the structure of a single water molecule with ωB97M-V/aug-cc-pVTZ. We then used this structure to create a cubic water box 22 Å × 22 Å × 22 Å to serve as the starting structure for MD simulations. The simulation box consisted of 392 water molecules, giving rise to density of 997 kg m^{-3}. The system was then equilibrated using the *NVT* ensemble at 300 K for 2 ns (the time step for the thermostat was 100 fs). Following the equilibration, we ran a 3 ns production trajectory (with time step of 2 fs) from which snapshots for the spectra calculations were collected. This trajectory was also used to compute structural parameters of bulk water.

The QM/MM AIMD simulations were initiated from 40 structures taken from the equilibrium MD trajectory. The QM region included all waters within 6.5 Å radius from the central water (this selection criterion resulted in the 35–40 QM water molecules, depending on a snapshot). The QM part was treated by ωB97X-D/6-31G* and the MM part by TIP3P. From each starting structure, a 2 ps long trajectory was propagated with a time step of 42 a.u. (1.016 fs) using the *NVT* ensemble at T = 300 K (the thermostat was applied every 100 fs). The first picosecond of each trajectory was treated as equilibration and the second ps was treated as a production run. Hence, the total simulation time in AIMD was 40 ps; these trajectories were used to collect snapshots for the QM/MM simulations of the ionization spectrum and to compute structural parameters of bulk water. We note that this time is shorter than in the MD simulations, however, the convergence of the radial distribution function, $g_{OO}(r)$, shows that this simulation time is sufficient (*e.g.*, $g_{OO}(r)$ computed from the full MD trajectory of 3 ns and from a 40 ps segment look the same).

The link to downloadable tar files with 3000 MD and 400 AIMD snapshots is given in the ESI.†

The simulation of bulk water properties is known to be difficult. As discussed in several recent papers,[64–66] the results are sensitive to the interaction potentials used, the size of the simulation box, the exact details of dynamics (*e.g.*, thermostat), as well as on whether nuclear quantum effects are included.

Although we used one of the best functionals (ωB97X-D,[67] which includes long-range Coulomb exchange and dispersion correction) in our AIMD simulations, the analysis of structural parameters ($g_{OO}(r)$, the number of hydrogen bonds formed) shows that our AIMD water is somewhat over-structured relative to TIP3P and state-of-the-art simulations,[65,66] similar to the simulations using less accurate density functionals. However, the structure of the second solvation shell appears to be reproduced better with AIMD than with TIP3P, as compared to the experimental $g_{OO}(r)$. The detailed analysis of structural parameters extracted from the MD and AIMD simulations as well as comparison with other simulations are given in the ESI.† Overall, we find that the differences between the spectra computed with MD and AIMD snapshots are small and are washed out by the statistical averaging. However, it is desirable to improve the sampling in future work by using, for example, path-integral simulations with accurate many-body potentials, as was done in ref. 66.

Here we define hydrogen bonds by the criterion of Luzar and Chandler,[68] *i.e.*, when the O–O distance $R_{O-O} < 3.5$ Å and $\angle O\cdots O{-}H < 30^\circ$. We use this definition in the analysis of structures from the equilibrium simulations and in comparisons between different protocols of building up QM.

B. Calculations of core IEs

IEs were computed using the CVS-EOM-IP-CCSD method[50,69,70] and the 6-311+G(3df) basis set fully uncontracted on oxygen, denoted below as u6-311+G(3df). Uncontracted Pople's basis sets have been shown to be effective in describing strong orbital relaxation effects common in core-hole states.[71] The effect of triple excitations was accounted by additional CVS-EOM-IP-MLCC3 calculations for a smaller number of snapshots.

There are two variants of CVS-EOM-CC approach, one in which the core is frozen at the CCSD step (fc-CVS-EOM-CCSD[50]) and the one in which the core is active (CVS-EOM-CCSD[69,70]). Table 2 summarizes the results for the isolated water molecule (using ωB97X-D/aug-cc-pVTZ optimized structure). As one can see, the two versions of CVS-EOM-IP-CCSD differ by about 1 eV. At the CCSD level, fc-CVS-EOM-IP-CCSD is closer to the experiment due to fortuitous cancellation of errors, however, at the CC3 level, CVS-EOM-IP-CC3 yields a smaller error than fc-CVS-EOM-IP-CC3. The effect of triple excitations is −2.6 eV for both methods. Remarkably, even when triples are included (at the CC3 level), the computed IEs are red-shifted with respect to the experiment by 1–2 eV. This large discrepancy can be attributed to the slow convergence of the core IEs with respect to the correlation treatment. According to a detailed benchmark study,[72] for molecules comprising first-row elements, quantitative agreement with experimental IEs is achieved at the CVS-EOM-IP-CCSDTQ level whereas CVS-EOM-IP-CCSDT is within 0.3 eV from the experimental values. The results for water from this study are shown in Table 2 – as one can see, full inclusion of triple excitations brings the computed IE within 0.01 eV from the experiment. Whereas this calculation used a very large basis set and included scalar relativistic correction, the comparison between the respective CVS-EOM-IP-CCSD and CVS-EOM-IP-CCSDT values shows that the main source of the errors in our CC3 calculations is due to an insufficient correlation treatment. Despite these discrepancies in the absolute value of gas-phase

Table 2 $1s_O$ IE of isolated water molecule

Method	IE, eV	Δ *vs.* exp.
fc-CVS-EOM-IP-CCSD[a]	540.26	+0.44
CVS-EOM-IP-CCSD[a]	541.32	+1.50
fc-CVS-EOM-IP-CC3[a]	537.68	−2.14
CVS-EOM-IP-CC3[a]	538.72	−1.10
CVS-EOM-IP-CCSD/AC5Z[b]	541.78	+1.96
CVS-EOM-IP-CCSDT/AC5Z[b]	539.81	−0.01
Exp.[c]	539.82 ± 0.02	

[a] u6-311+G(3df). [b] From ref. 72, with scalar relativistic corrections; AC5Z denotes aug-cc-pCV5Z. [c] From ref. 13.

IE computed with CVS-EOM-IP-CC3, we anticipate much higher accuracy in ΔIE, as the errors should cancel out. In the simulations of the spectrum, we use fc-CVS-EOM-IP-CCSD and evaluate the effect of triple excitations to the shift by computing the difference between the ΔIE from CVS-EOM-IP-CCSD and CVS-EOM-IP-MLCC3.

In order to estimate the effect of the structure on the IE, we also computed IEs (with fc-CVS-EOM-IP-CCSD) for the structures optimized with ωB97X-D/6-31G* (the level at which the AIMD simulation was performed) and the TIP3P water structure. The respective IEs are 540.26 eV and 540.25 eV, respectively, which is close to 540.26 eV (for the ωB97X-D/aug-cc-pVTZ structure). Hence, small differences of water structures due to different levels of theory used to simulate the bulk are not expected to affect the computed IEs.

To compute bulk spectra, we considered several protocols designed to mitigate potential issues due to the description of waters on the boundary between the QM and MM parts. Our results indicate that electrostatic embedding in which the MM waters are described by point charges is not sufficient for describing core-level IEs and that the IEs of the water molecules on the boundary are not accurate. Therefore, we used the following multi-layer scheme for the calculation of the spectra. We first computed all core IEs in the QM part – for example, for a calculation with 5 waters in the QM part, we computed 5 IEs. We then analyzed the respective Dyson orbitals[51] to assign the computed IEs to particular water molecules. Fig. 4 shows a structure of 5 water molecules embedded in the MM part with the Dyson orbital on the central water molecule. The assignment of IEs can also be accomplished by considering the Mulliken charges from the natural orbital analysis[73] (we used this approach in the production simulations). We then constructed the bulk spectra by taking the IEs corresponding to a specified number of water molecules. For example, in the calculations with 5 QM waters, one can construct the spectrum by taking all 5 IEs or by taking only the IE of the central water molecule. The overall spectra were constructed as histograms (with 0.05 eV bins) by collecting the IEs from the snapshots and then convoluted with gaussians. The spectra shown in the manuscript were produced using gaussians with 0.2 eV width (FWHM). Discarding some of the computed IEs in the construction of the spectra slows down the convergence with respect to the number of snapshots, but removes the artifacts due to waters on the QM/MM interface. As we show below, the best protocol (in terms of balancing accuracy of IEs and convergence with respect to sampling) entails using 20 QM waters and assembling the spectra by taking the IEs of the 5 central waters.

Triple excitations are important for obtaining accurate core excitation and ionization energies within the CC/EOM-CC framework.[74–76] The CC3 model includes the effects of triple excitations in a perturbative manner.[53] This method generally scales as $\mathcal{O}(N^7)$. However, for core excitations using the CVS scheme, the scaling is $\mathcal{O}(N^6)$ in an optimized implementation.[55] Nevertheless, the ground-state calculation is rather expensive, even for systems where the CCSD calculation is routine. The costs can

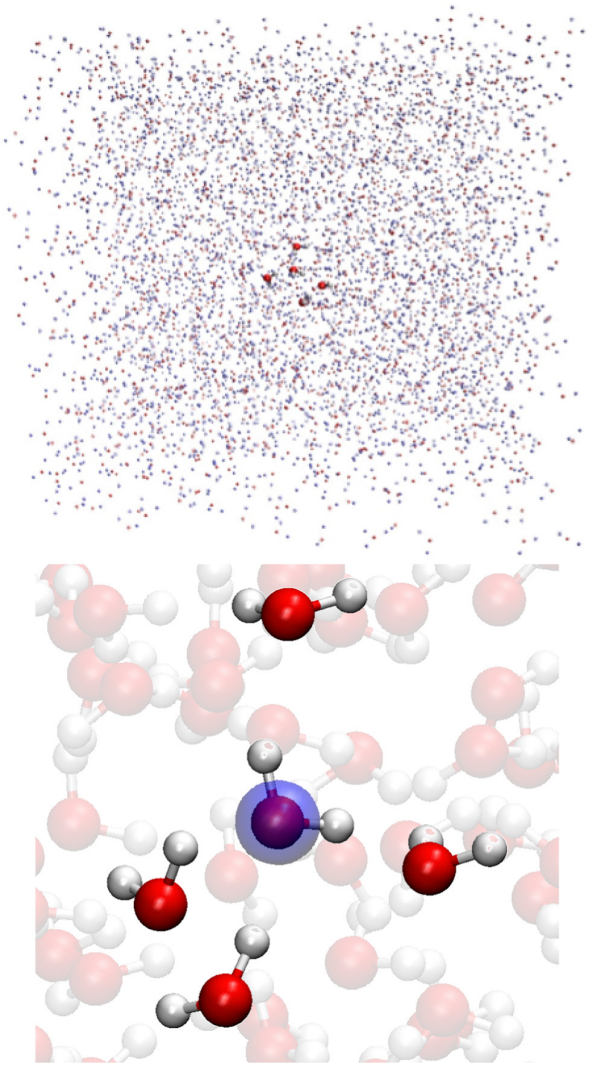

Fig. 4 Model system for liquid water. Top: Simulation box with 392 water molecules. Bottom: A cluster from the center of the box of one water molecule and its first solvation shell representing a minimal QM subsystem. The Dyson orbital associated with the $1s_O$ ionization of the central water is shown in blue.

be significantly reduced using the multilevel coupled-cluster approach (MLCC) in which the higher-order excitations in the cluster operator are restricted to an active orbital space. In the MLCC3 model, the triple excitation operator is restricted while the single and double excitation operators are unrestricted and act in the entire orbital space. The MLCC models target intensive properties, such as excitation or ionization energies. Additional reduction in cost can be achieved by restricting the entire cluster operator to an active space. In this way, one obtains the CC-in-HF models, where the inactive orbitals are not correlated but contribute to the energy *via* the Fock matrix.[61,77,78] This approach, which can be described as a type of electronic embedding, has been demonstrated to work well for modeling solvent effects in spectroscopy.[79]

The multilevel models rely on a physically appropriate selection of the active orbital space. Localized orbitals are an obvious option when the property of interest is localized. Core IEs is an example of such local properties. Another option is to use correlated natural transition orbitals (CNTOs),[80] which use information from the excitation amplitudes of a lower-level model to generate an active space for the MLCC calculation, similarly to other approaches using virtual orbital spaces truncated on the basis of lower-level natural orbitals.[81,82]

Here, we use a hybrid active orbital selection strategy where CNTOs are used for the occupied orbital space and projected atomic orbitals (PAOs)[83] on the five central water molecules determine the virtual orbital space. In the present MLCC calculations, the active orbital space contains 25 occupied and 245 virtual orbitals. A detailed description of the orbital selection procedure with CNTOs and PAOs for the MLCC models can be found in ref. 84.

C. Protocols for selecting the QM subsystem for IE calculations

The first solvation shell of a water molecule comprises 4–5 water molecules[85–87] (Fig. 4 shows a typical structure from the equilibrium simulations). Thus, our minimal QM system for computing bulk IEs comprises 5 water molecules. As we show below, the convergence of the core IEs with respect to the QM size is slow and much larger QM systems are needed for converged results.

To determine an optimal protocol for building up a larger QM system, we compared two different approaches. In the first approach, we used the distance from the oxygen or hydrogen of the central water $O_{cen}(H_{cen})\cdots H(O)$ to select the next water molecule to be added to the QM system. In the second approach, we used the distance from the waters in the first solvation shell $O_{1stshell}(H_{1stshell})\cdots H(O)$ to select the next water. Fig. 5(c) and (d) shows the shifts in the IE of the central water (δ) upon increasing the QM size from 5 to 6 waters (with the rest of the waters described by point charges) for two MD snapshots (the 5 waters are chosen to represent minimal QM in the center of the box and the 6th water is taken at random from the simulation box). For each snapshot, we determine whether the 6th water molecule acts as a donor or acceptor (or both). As expected, the shifts in IE (δ) are mostly positive when the added water molecule acts as hydrogen-bond donor and negative when the added water molecule acts as hydrogen-bond acceptor. For the structures used to construct Fig. 5, the δIE for hydrogen-bond acceptor structures, hydrogen-bond donor structures, and non-hydrogen bonded waters range between −0.016 and −0.150 eV, −0.024 and 0.109 eV and −0.032 and 0.039 eV, respectively. The waters that act as both hydrogen bond acceptor and donor result in δ between 0 and −0.05 eV. The magnitude of the shifts shown *versus* minimum $O_{cen}(H_{cen})\cdots H(O)$ distance (Fig. 5c) exhibit no systematic trend. In contrast, when δ are shown against the minimum $O_{1stshell}(H_{1stshell})\cdots H(O)$ distance (Fig. 5d), we observe a smoother behavior – as the distance increases, the shifts become less negative for the structures with hydrogen-bond acceptors and less positive for the structures with hydrogen-bond acceptors. Thus, water molecules that form hydrogen bonds with the waters in the first solvation shell have a greater impact on the IE of the central water molecule.

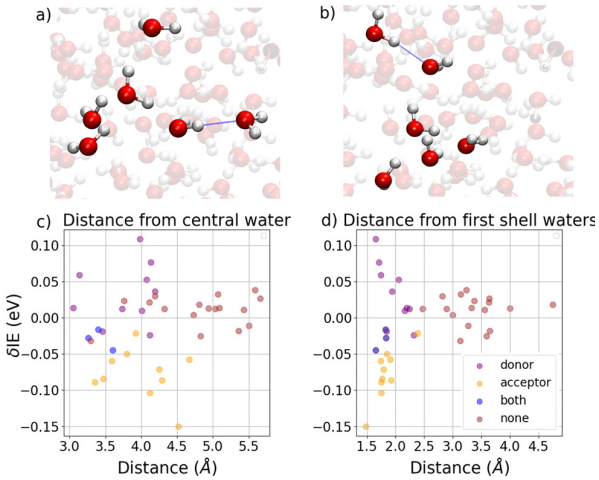

Fig. 5 Panels (a) and (b) show in bright colors 6-water QM systems comprising the central water, its first solvation shell, and a 6th water selected from the second solvation shell. In panel (a), the 6th water molecule accepts a hydrogen bond from a water molecule from the first solvation shell. In panel (b), the 6th water molecule donates a hydrogen bond to a water molecule from the first solvation shell. The points in panels (c) and (d) represent the shift in IE of the central water molecule (δ) due to adding the 6th water. Each point corresponds to a different selection of the 6th water sampled over two snapshots from the TIP3P waterbox. The points are color-coded to show whether the 6th water acts as a donor (pink), acceptor (yellow), or both (blue). Panel (c) shows δ *versus* the shortest $O_{cen}(H_{cen})\cdots H(O)$ distance and panel (d) shows δ *versus* the shortest $O_{1stshell}(H_{1stshell})\cdots H(O)$ distance.

The above analysis shows that waters that are hydrogen-bonded to the first solvation shell have the strongest effect on the IE of the central water. Hence, we can use this criterion of building up the QM system instead of the distance from the central water. Fig. 6 shows the convergence of the core IE of the central water with respect to the QM system size using these two criteria for growing the QM system. The smallest QM comprises 5 water molecules (we pick a water molecule at the center of the

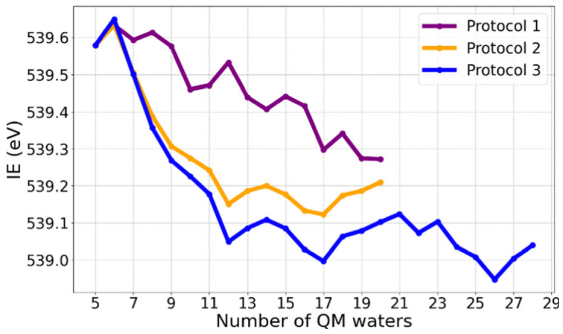

Fig. 6 IE of the central water molecule for a single snapshot computed using different protocols for selecting the QM system in the CVS-EOM-IP-CCSD calculations. The rest of the waters are described by point charges.

simulation box and chose 4 closest waters). We then increase the QM size by adding more waters. In protocol 1, we add the next water based on the $O_{cen}(H_{cen})\cdots H(O)$ distance. In protocols 2 and 3, we add waters based on their $O_{1stshell}(H_{1stshell})\cdots H(O)$ distance. The central water and the first solvation shell are described with the fully uncontracted 6-311+G(3df) basis on oxygen and the 6-311G basis on hydrogens. The rest of the waters in the QM system are treated with a smaller basis – 6-311G* in protocols 1 and 2, and 6-31G in protocol 3. The rest of the waters are described by point charges. We freeze all the oxygen cores in the QM system in the CVS-EOM-IP calculations and request the number of IEs equal to the number of oxygen atoms in the QM system. We then select the IE of the central water.

As Fig. 6 shows, increasing the QM size results in a red shift of the $1s_O$ IE of the central water. The minimal QM is clearly not sufficient – increasing the size from 5 to 20 waters leads to the red shift of 0.3–0.4 eV. The convergence is not monotonous and quite slow – we observe small fluctuations ($\sim$0.1 eV) even beyond 25 waters. Fig. 6 also shows that selecting waters by the distance from the first solvation shell is more effective and results in faster and smoother convergence. For the QM size of 20 water molecules, the difference in the IE between protocol 1 and 2 is 0.06 eV and between protocol 2 and 3 (basis-set effects) – 0.107 eV. We note that the difference due to using a smaller basis increases with the system size (5–12) and then becomes nearly constant (about 0.10 eV for the QM sizes of 12–20). Hence, in our production-level simulations we use protocol 3 and estimate the basis-set correction by taking a difference between protocols 2 and 3 for a small number of snapshots. Fig. 7 shows the convergence of the shift in $1s_O$ IE of the central water for the simulation with the 20 QM waters between protocol 2 and protocol 3. The average value of the shift (δ_{basis}) is 0.105 eV.

To further analyze the convergence with respect to the QM size, Fig. 8 shows the number of hydrogen bonds formed by the water molecules of the first solvation shell as a function of the QM size for the same snapshot as used in Fig. 6. As one can see, beyond the QM size of 10, the protocol based on the distance from the first solvation shell captures more hydrogen bonds than the protocol based on the distance from the central water.

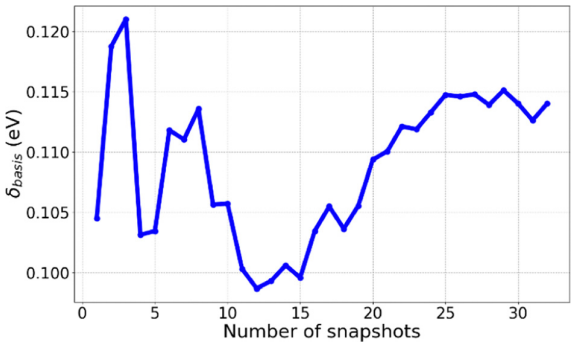

Fig. 7 Convergence of the shift in IE in between protocols 2 and 3 (δ_{basis}) with the 20 QM waters. The estimated shift δ_{basis} = 0.114 eV, with a 0.035 eV standard deviation.

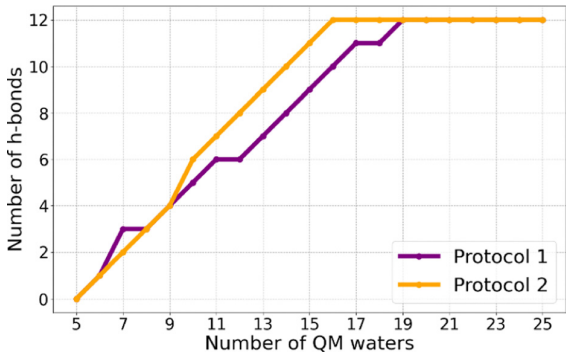

Fig. 8 The number of hydrogen bonds in the first solvation shell as a function of the QM system size for a single snapshot.

We also see that all bonds are captured for a smaller QM size when QM is selected using the former protocol: saturation is reached at 15 waters *versus* 18. On the basis of this analysis, we conclude that the QM size of 20 waters should be sufficient to capture all hydrogen bonds formed by the first solvation shell and we use QM size of 20 waters in our production-level calculations.

Table 3 shows the effect of triple excitation and different types of embedding evaluated for the same snapshot as used in Fig. 6. The u6-311+G(3df) basis was used for 5 waters and 6-31G for the rest of quantum waters. Although the absolute values of IEs vary between the methods, the value of the shift, ΔIE, is rather insensitive to the type of embedding, CVS scheme, or correlation treatment. The effect of MM charges (beyond 20 quantum waters) is about 0.1 eV and the effect of freezing the second solvation shell (15 water molecules) at the Hartree–Fock level of theory is approximately 0.1 eV (decreasing the magnitude of the red shift). The effect of including triple excitations with MLCC3 is almost 0.4 eV (increasing the magnitude of the red shift) for this snapshot.

The multilevel framework makes it possible to use approximate triples for a study such as this. However, the cost remains

Table 3 Effect of triple excitations on IE (eV) from multi-level calculations

Method	Gas-phase IE	Bulk IE	ΔIE
CVS-EOM-IP-CCSD (20)[a]	541.32	540.36	−0.96
CVS-EOM-IP-CCSD-in-HF (5/15)[b]	541.32	540.45	−0.87
CVS-EOM-IP-CCSD (20) in MM[c]	541.32	540.24	−1.08
CVS-EOM-IP-CCSD-in-HF (5/15) in MM[d]	541.32	540.33	−0.99
CVS-EOM-IP-MLCC3 (20)[e]	538.72	537.38	−1.34
CVS-EOM-IP-MLCC3 (20) in MM[f]	538.72	537.25	−1.47
fc-CVS-EOM-IP-CCSD in MM[g]	540.26	539.17	−1.09

[a] 20 water molecules are treated by CCSD, MM water molecules are ignored. [b] 5 water molecules are treated by CCSD, 15 water molecules are treated by HF, MM water molecules are ignored. [c] 20 water molecules are treated by CCSD, MM water molecules are treated as MM charges. [d] 5 water molecules are treated by CCSD, 15 water molecules are treated by HF, MM water molecules are treated as MM charges. [e] 20 water molecules are treated by MLCC3, MM water molecules are ignored. CNTO/PAOs used for active orbital selection. [f] 20 water molecules are treated by MLCC3, MM water molecules are treated as MM charges. CNTO/PAOs used for active orbital selection. [g] 20 water molecules are treated by CCSD, the rest by MM charges.

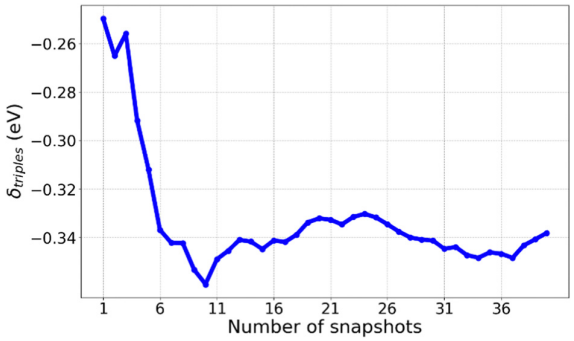

Fig. 9 Convergence of the shift in IE due to approximate triples ($\delta_{triples}$) through the CVS-EOM-IP-MLCC3 approach. The estimated effect of triples on the shift in ΔIE $\delta_{triples}$ = −0.34 eV, with a 0.07 eV standard deviation.

a limitation and performing such calculation for hundreds of snapshots is impractical. We therefore use a subset of 40 snapshots to estimate the effect of approximate triples; the results are shown in Fig. 9. The average shift is −0.34 eV, with a 0.07 eV standard deviation. This correction will be applied to the final averaged fc-CVS-EOM-CCSD result to provide our best estimate for the ΔIE.

D. Extrapolation to bulk

To obtain bulk IE, one needs to extrapolate the results from finite-size simulations, such as our simulation box used in IE calculations, to the infinite system size. The importance of this step has been recently illustrated by Tazhigulov and Bravaya[88] – they have shown that the computed energies converge very slowly with the simulation size. Based on the linear dependence of the energies on the inverse system size, they derived a simple correction using Born solvation model.[88]

For vertical IEs of the solvated neutral species, the Born correction is[88]

$$\mathrm{VIE}^{\infty} = \mathrm{VIE}^{R} - \frac{1}{2}\left(1 - \frac{1}{\varepsilon_{\mathrm{opt}}}\right)\frac{1}{R}, \qquad (6)$$

where VIE^{∞} is intrinsic vertical IE in the bulk, VIE^{R} is VIE computed in a finite cluster of radius R, and $\varepsilon_{\mathrm{opt}}$ is solvent's optical dielectric constant (1.78 for water); all quantities are in atomic units. For our simulation box ($R \approx 11$ Å), the correction results in an additional red shift in IE of 0.29 eV.

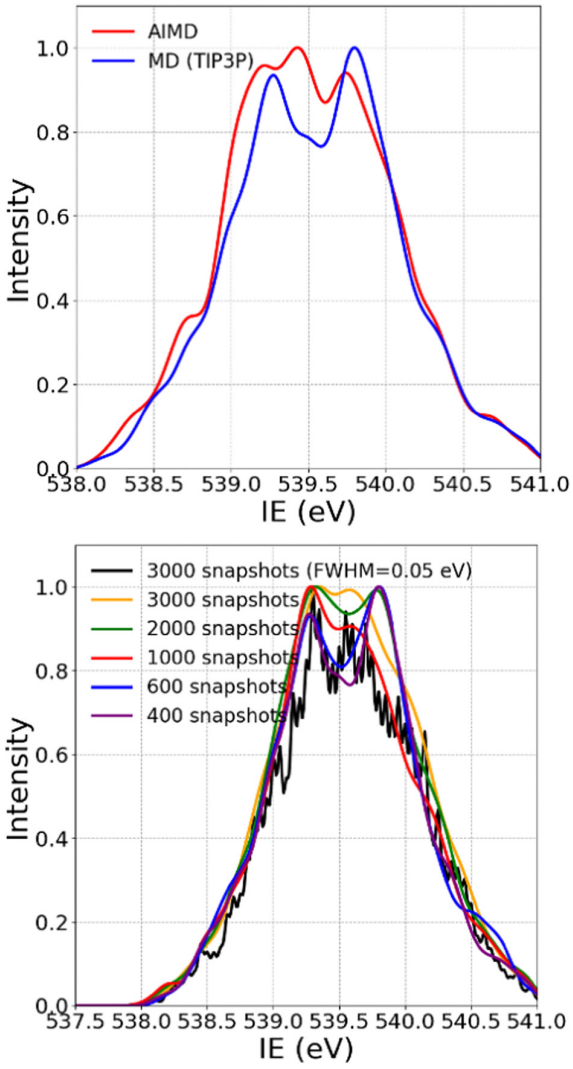

Fig. 10 Theoretical $1s_O$ spectra of liquid water computed using the minimal QM system (5 waters) and assembled using the IE of the central water molecule. Top: Spectra constructed from the MD (TIP3P) and AIMD trajectories (400 snapshots). Bottom: Spectra constructed from the MD trajectory using different number of snapshots. All spectra were obtained using gaussians with FWHM = 0.2 eV, except for the black trace in the right panel which was produced using 0.05 eV FWHM to show the intrinsic roughness of the spectrum produced from the 3000 snapshots.

III. Results and discussion

One of the two goals of this study is to understand the effect of the simulation protocol on the computed spectra. We begin with comparing the results of the MD and AIMD simulations. Fig. 10 (top panel) compares the spectra constructed using the MD and AIMD snapshots. In this calculation, we used the smallest QM size (5 waters only, the rest being treated by point charges) and constructed the spectra using the IE of the central water molecule. The two simulations yield essentially identical spectra, both in terms of the band maximum and in terms of the bandwidth (inhomogeneous broadening). There are small differences in fine structure, but, as we show below, these features become smoothed out when the sampling is increased. Thus, we conclude that for this system the core IE spectra are not sensitive to the differences between TIP3P and AIMD sampling of the equilibrium dynamics. The bottom panel of Fig. 10 shows the convergence of the spectrum with respect to the number of snapshots. As one can see, while the band maximum and its width are captured by the simulation with 400 snapshots, the finer details continue to evolve and the spectrum converges only around 3000 snapshots (although the

remaining roughness remains visible). We note that the convergence can be accelerated when using more than one IEs, as we do in the simulations with larger QM sub-systems.

We also point out that the double-peak structure, which is not visible in the experiments, persists and remains visible even for the simulation with 3000 snapshots. It becomes more pronounced when thinner gaussians are used. This feature is reminiscent of the double-peak structure observed on the X-ray emission spectra in liquid water.[66]

Fig. 11 shows the spectra constructed from either one IE (of the central water) or 5 IEs (of the cluster of 5 waters). As one can see, for the small QM (5 waters), the spectrum changes qualitatively when the IEs of all 5 waters are used, leading to a much smaller red shift relative to the gas-phase peak than the calculation using the IEs of 1 central water (the difference between the two calculations is 0.25 eV). This illustrates the limitations of the electrostatic embedding, which apparently is not able to correctly describe the waters on the QM/MM boundary. Using the QM of 20 waters, results in a larger red shift (by about 0.5 eV), consistent with the benchmark calculations for a single snapshot (Fig. 6). Moreover, the calculations with the large QM yield the same spectra (in terms of the peak maximum and its width) whether using the IEs of one or five waters, suggesting that the boundary is sufficiently far in this case. The spectrum computed with the IEs of 5 waters is smoother, as in such calculation the sampling is effectively improved by a factor of 5. Hence, we construct our best spectrum by using the IEs of the five water molecules embedded in the QM cluster of 20 waters (with the rest of waters described by point charges). The effect of triples is evaluated using the MLCC3 method with 20 CCSD waters and the triples operator restricted to orbitals within the first solvation shell of the central water molecule.

Fig. 12 shows our best spectrum and compares it with the available experimental spectra.[12,13,24] We note that the reported spectrum of Olivieri *et al.*[13] is for zero bias and, therefore, does not include the correction due to the workfunctions difference. The corrected values from this experiment are given in Table 4, which also lists the computed peak positions and widths for different protocols and compares them with the experimental values. To extract the value of the peak maxima and the width from the computed spectra, we convoluted the computed spectra with broad gaussians (FWHM = 0.5 eV) to remove the noise due to finite sampling. The effect of this broadening is analyzed in the ESI:† Fig. S4 in the ESI† shows the sample spectra obtained by convoluting the raw data with Gaussians of different widths. As one can see, using broad gaussians does not introduce noticeable change in the band width.

Our best value of the shift is −1.79 eV – it is computed from the spectrum based on 400 AIMD snapshots with 20 QM waters and using the IEs of 5 waters (AIMD/400, 5w/20QM scheme) treated with fc-CVS-EOM-CCSD, to which we add the triples, basis-set, and Born corrections. This protocol yields the absolute value of the bulk IE of 538.47 eV. Our results agree well with the best experimental values, *i.e.*, by Olivieri *et al.*,[13] who reported the bulk water IE of 538.21 ± 0.07 eV and the shift of 1.61 ± 0.09 eV, and with Thürmer *et al.*,[16] who reported IE of 538.10 ± 0.05 eV.

A. Analysis of structures

The spectra of the bulk species report on the structure of the solvent around the solute and its dynamic fluctuations. In

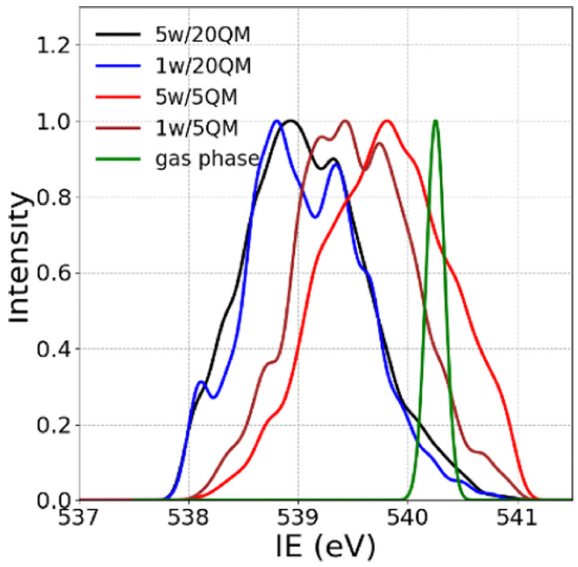

Fig. 11 Theoretical $1s_O$ spectra of liquid water computed using small (5) and large (20) QM systems constructed from the AIMD trajectory (400 snapshots). The spectra were constructed using the $1s_O$ IEs of either one central water or five QM waters.

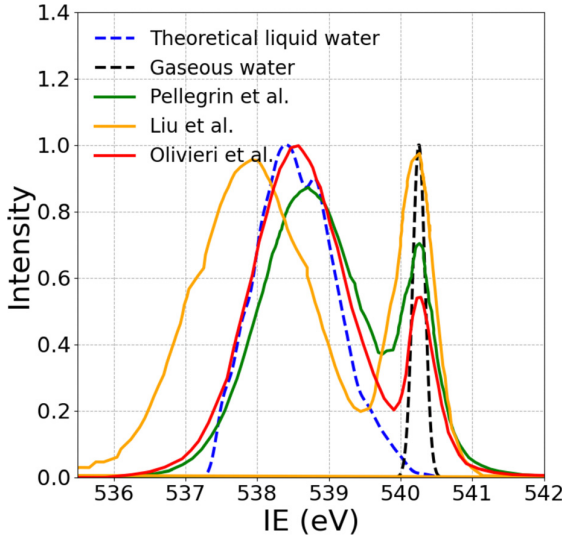

Fig. 12 Experimental[12,13,24] and theoretical $1s_O$ XPS spectra of water. The spectra of Pellegrin *et al.*,[12] Liu *et al.*,[24] and Olivieri *et al.*[13] spectra has been shifted by +0.59 eV, +4.9 eV, and +0.87 eV, respectively, to match the theoretical gas-phase peak. The computationally constructed spectrum is shown by dashed line. The computed spectra of liquid water includes triples, basis-set, and Born corrections.

Table 4 Shift (ΔIE) of the $1s_O$ IE of liquid water relative to the gas-phase water and the width (FWHM) of the band

Setup[a]	ΔIE, eV	FWHM, eV
AIMD/400, 1w/5QM	−0.84	1.40
AIMD/400, 5w/5QM	−0.45	1.55
MD/400, 1w/5QM	−0.74	1.43
MD/400, 5w/5QM	−0.41	1.50
MD/3000, 1w/5QM	−0.75	1.42
MD/3000, 5w/5QM	−0.40	1.51
AIMD/400, 1w/20QM	−1.37	1.41
AIMD/400, 5w/20QM	−1.27	1.44
AIMD/40, 5w/20QM[b]	−1.15	
Best estimate[c]	−1.79	1.44
Exp. (ref. 13), zero bias	−1.77	1.53
Exp. (ref. 13), corrected	−1.61	1.53
Exp. (ref. 16)	−1.72	
Exp. (ref. 24)	−2.2	1.93

[a] Sampling/snapshots, number of IEs/size of QM; fc-CVS-EOM-IP-CCSD/u6311+G(3df). [b] CVS-EOM-IP-CCSD/u6311+G(3df). [c] AIMD/400, 5w/20QM/fc-CVS-EOM-IP-CCSD value with the CC3 ($\delta_{triples}$ = −0.340 eV) and basis-set (δ_{basis} = +0.114 eV) corrections, as well as Born correction (−0.29); see text.

particular, the spectrum of water reflects the contributions from different hydrogen-bonding patterns. Computationally, it is possible to break down the total computed spectra into different contributions, similarly to the analysis in other theoretical studies.[11,66] Fig. 13 shows the breakdown of the total spectrum into the contributions from structures with different hydrogen-bonding patterns. Panel (a) shows contributions from snapshots where the central water molecule forms a single donor–single acceptor (DA), single donor–double acceptor (DAA), double donor–single acceptor (DDA), and double donor–double acceptor (DDAA) hydrogen bonds. Collectively, these types of structures add up to 97.2% of the structures sampled in the simulation. The shape of the band is dominated by the DDAA pattern, which has the largest population. The structures in which the central water acts as hydrogen-bond donor yield red-shifted IEs and the structures in which the central water acts as hydrogen-bond acceptor are blue shifted. The DDAA motif, in which there is an equal number of both kinds of hydrogen bonds, features a high-energy and a low-energy peaks. The DDA and DAA motifs have peaks at lower and higher energies, respectively.

The bottom panel of Fig. 13 shows the analysis of the spectrum in terms of the contributions from different hydrogen-bonding patterns around the first solvation shell. The motifs shown in the figure add up to 68.9% of the structures sampled in by the simulation. The 6D6A motif, which corresponds to the DDAA motif of the central water, has the highest contribution. Due to many different motifs of hydrogen bonding around the first shell, the overall effect is that these structures cannot be mapped into particular spectral features but rather they collectively contribute

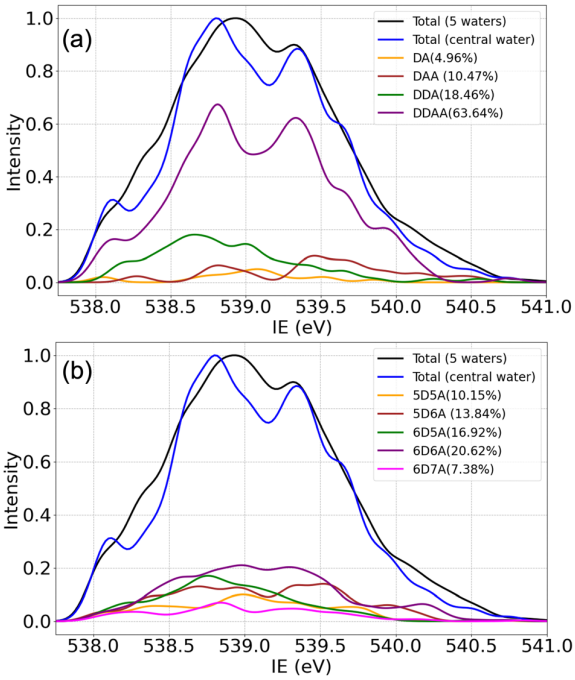

Fig. 13 The $1s_O$ spectra of liquid water constructed from the central water and the first shell (total 5 waters) broken into the contributions from structures with different hydrogen-bonding patterns (raw spectrum without corrections). Panel (a) shows the contributions from structures with different hydrogen-bonding patterns around the central water and panel (b) shows the contributions of different hydrogen-bonding patterns around the first solvation shell. The spectra were computed from the AIMD snapshots trajectory using QM with 20 waters and treating the rest of the waters as point charges.

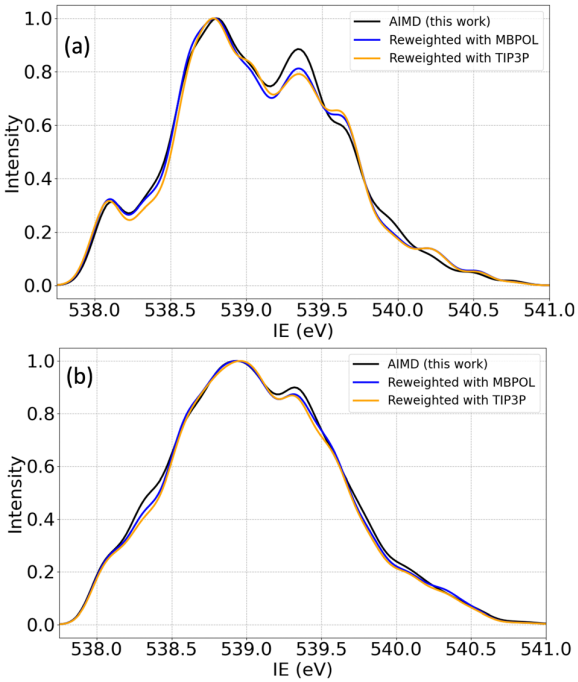

Fig. 14 The $1s_O$ spectra of liquid water (raw spectra without corrections) constructed from the (a) central water molecule and (b) 5 water molecules computed using AIMD snapshots (black) and synthetic spectra obtained by re-weighting the contributions from the dominant hydrogen-bonding patterns to match the distributions from other simulations (see text).

to the inhomogeneous broadening and smoothening of the spectrum.

Finally, we use the contributions of different hydrogen-bonding patterns to estimate the effect of different equilibrium sampling on the computed spectra using the approach from ref. 66. Fig. 14 shows our original spectrum computed using 400 AIMD snapshots and several synthetic spectra obtained by rescaling the relative contributions of the structures with double donor (sum of contributions of the DDA and DDAA structures) and single-donor (sum of the contributions of the DA and DAA structures) motifs to match the results from other simulations (see Table S2 in the ESI†). We see that the effect on the overall shape of the band, its maximum and width, are negligible, especially, when sampling is adequate (such as in the simulations using the IEs of the 5 waters). Hence, imperfections in the equilibrium water structures seem to be washed out by the averaging and are not expected to affect the computed value of the ΔIE.

IV. Conclusion

We presented a state-of-the-art simulation of the $1s_O$ ionization of liquid water. We employed highly accurate EOM-CC methods adapted to core-vacancy states by using the CVS scheme. Equilibrium sampling was carried out using MD and AIMD simulations. We carefully analyzed the effect of the embedding on the computed IEs and show that the convergence of the result with respect to the size of the QM system is slow. Our production-level calculations were carried out using the QM system of 20 water molecules embedded in the MM charges. We also evaluated the effect of triple excitations using the MLCC3 framework. Our calculations yield the value of the intrinsic bulk IE of 538.47 eV and the FWHM of the bulk peak of 1.44 eV; the computed shift relative to the gas-phase IE is −1.79 eV (including MLCC3 and basis-set corrections).

These results agree well with the best experimental values from the most recent liquid-jet experiments of the bulk water IE of 538.21 ± 0.07 eV and the shift of 1.61 ± 0.09 eV (ref. 13) and 538.10 ± 0.05 eV (ref. 16). Given the remaining uncertainties in the experimental determination of the true (intrinsic) bulk IEs (surface potential, presence of solvated species, *etc.*), our results provide an important reference value.

We conclude with listing the aspects of the present theoretical treatment that need improvement. First, improved equilibrium sampling is highly desirable, *i.e.*, using higher-quality *ab initio* treatment, larger QM sizes, and including nuclear quantum effects. Second, one can further increase the size of the QM system in the IE calculations, which can be achieved using multi-level methodologies, to achieve full convergence. Third, improving correlation treatment beyond CC3 is desirable. Fourth, the effect of electrolytes both on the bulk IE and on the surface potential needs to be studied. Fifth, the effect of the finite-depth probed in microjet experiments needs to be investigated by simulations. We hope to address these issues in future studies. The availability of accurate intrinsic bulk IE and reliable experimental value (real IE) can provide an estimate of the surface potential (φ_{in}) of water.

Data availability

The data that support the findings of this study are available within the article and the associated ESI.†

Conflicts of interest

The authors declare the following competing financial interest(s): A. I. K. is the president and a part-owner of Q-Chem, Inc.

Acknowledgements

We thank Professor Stephen Bradforth from USC for stimulating discussions and for his comments on the manuscript. We also thank Professor Stephan Thrümer for his illuminating explanations about subtle details of microjet experiments. S. D. and A. I. K. are grateful to Dr Musahid Ahmed from the Lawrence Berkeley National Laboratory for the initial motivation for this project and for stimulating discussions. This work was supported by the U.S. National Science Foundation (No. CHE-2154482 to A. I. K.) and Research Council of Norway (FRINATEK 275506 to S. D. F., A. P., and H. K.).

References

1 C. P. McKay, Requirements and limits for life in the context of exoplanets, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 12628.
2 E. Brini, C. J. Fennell, M. Fernandez-Serra, B. Hribar-Lee, M. Lukšič and K. A. Dill, How water's properties are encoded in its molecular structure and energies, *Chem. Rev.*, 2017, **117**, 12385.
3 F. Cipcigan, V. Sokhan, G. Martyna and J. Crain, Structure and hydrogen bonding at the limits of liquid water stability, *Sci. Rep.*, 2018, **8**, 1718.
4 A. Nilsson and L. G. M. Pettersson, Perspective on the structure of liquid water, *Chem. Phys.*, 2011, **389**, 1.
5 A. I. Krylov, From orbitals to observables and back, *J. Chem. Phys.*, 2020, **153**, 080901.
6 S. Gozem and A. I. Krylov, The ezSpectra suite: An easy-to-use toolkit for spectroscopy modeling, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2022, **12**, e1546.
7 S. Gozem, R. Seidel, U. Hergenhahn, E. Lugovoy, B. Abel, B. Winter, A. I. Krylov and S. E. Bradforth, Probing the electronic structure of bulk water at the molecular length scale with angle-resolved photoelectron spectroscopy, *J. Phys. Chem. Lett.*, 2020, **11**, 5162.
8 M. Ahmed and W. Lu, Probing complex chemical processes at the molecular level with vibrational spectroscopy and X-ray tools, *J. Phys. Chem. Lett.*, 2023, **14**, 9265.
9 B. Winter and M. Faubel, Photoemission from liquid aqueous solutions, *Chem. Rev.*, 2006, **106**, 1176.

10 D. Ghosh, A. Roy, R. Seidel, B. Winter, S. E. Bradforth and A. I. Krylov, A first-principle protocol for calculating ionization energies and redox potentials of solvated molecules and ions: Theory and application to aqueous phenol and phenolate, *J. Phys. Chem. B*, 2012, **116**, 7269.
11 B. Winter, E. F. Aziz, U. Hergenhahn, M. Faubel and I. V. Hertel, Hydrogen bonds in liquid water studied by photoelectron spectroscopy, *J. Chem. Phys.*, 2007, **12**, 124504.
12 E. Pellegrin, V. Perez-Dieste, C. Escudero, P. Rejmak, N. Gonzalez, A. Fontsere, J. Prat, J. Fraxedas and S. Ferrer, Water/methanol solutions characterized by liquid μ-jet XPS and DFT—the methanol hydration case, *J. Mol. Liq.*, 2020, **300**, 112258.
13 G. Olivieri, A. Goel, A. Kleibert, D. Cvetkoc and M. A. Brown, Quantitative ionization energies and work functions of aqueous solutions, *Phys. Chem. Chem. Phys.*, 2016, **18**, 29506.
14 M. Faubel, B. Steiner and J. P. Toennies, Photoelectron spectroscopy of liquid water, some alcohols, and pure nonane in free micro jets, *J. Chem. Phys.*, 1997, **106**, 9013.
15 K. Nishizawa, N. Kurahashi, K. Sekiguchi, T. Mizuno, Y. Ogi, T. Horio, M. Oura, N. Kosugi and T. Suzuki, High-resolution soft X-ray photoelectron spectroscopy of liquid water, *Phys. Chem. Chem. Phys.*, 2011, **13**, 413.
16 S. Thürmer, S. Malerz, F. Trinter, U. Hergenhahn, C. Lee, D. M. Neumark, B. Winter and I. Wilkinson, Accurate vertical ionization energy and work function determinations of liquid water and aqueous solutions, *Chem. Sci.*, 2021, 10558.
17 J. W. Rabalais, *Principles of UltraViolet Photoelectron Spectroscopy*, John Wiley and Sons, New York, 1977.
18 K. M. Ervin and W. C. Lineberger, Photoelectron spectroscopy of molecular anions, in *Advances in Gas Phase Ion Chemistry*, ed. N. G. Adams and L. M. Babcock, JAP Press, Greenwich, CT, 1992, vol. 1, pp. 121–166.
19 J. Simons, Detachment processes for molecular anions, in *Photoionization and Photodetachment*, ed. C. Y. Ng, Part II of Advanced Series in Physical Chemistry, World Scientific Publishing Co., Singapore, 2000, vol. 10.
20 A. Sanov and R. Mabbs, Photoelectron imaging of negative ions, *Int. Rev. Phys. Chem.*, 2008, **27**, 53.
21 O. Kostko, B. Bandyopadhyay and M. Ahmed, Vacuum ultraviolet photoionization of complex chemical systems, *Annu. Rev. Phys. Chem.*, 2016, **67**, 19.
22 W. F. Egelhoff, Jr, Core-level binding-energy shifts at surfaces and in solids, *Surf. Sci. Rep.*, 1987, **6**, 253.
23 E. Harder and B. Roux, On the origin of the electrostatic potential difference at a liquid-vacuum interface, *J. Chem. Phys.*, 2008, **129**, 234706.
24 J. Liu, Y. Han, C. Liu, B. Yang and Z. Liu, Origin of the liquid/gaseous water binding energy splitting measured *via* X-ray photoelectron spectroscopy, *J. Phys. Chem. Lett.*, 2023, **14**, 863.
25 S. M. Kathmann, F. W. Kuo and C. J. Mundy, Electronic effects on the surface potential at the vapor-liquid interface of water, *J. Am. Chem. Soc.*, 2008, **130**, 16556.
26 S. M. Kathmann, I.-F. W. Kuo, C. J. Mundy and G. K. Schenter, Understanding the surface potential of water, *J. Phys. Chem. B*, 2011, **115**, 4369.
27 C. C. Doyle, Y. Shi and T. L. Beck, The importance of the water molecular quadrupole for estimating interfacial potential shifts acting on ions near the liquid-vapor interface, *J. Phys. Chem. B*, 2019, **123**, 3348.
28 C. Zhang and M. Spirk, Electromechanics of the liquid water vapour interface, *Phys. Chem. Chem. Phys.*, 2020, **22**, 10676.
29 M. N. Pohl, E. Muchová, R. Seidel, H. Ali, Š. Sršeň, I. Wilkinson, B. Winter and P. Slavíček, Do water's electrons care about electrolytes?, *Chem. Sci.*, 2019, **10**, 848.
30 W. L. Jolly, K. D. Bomben and C. J. Eyermann, Core-electron binding energies for gaseous atoms and molecules, *At. Data Nucl. Data Tables*, 1984, **31**, 433.
31 R. Sankari, M. Ehara, H. Nakatsuji, Y. Senba, K. Hosokawa, H. Yoshida, A. De Fanis, Y. Tamenori, S. Aksela and K. Ueda, Vibrationally resolved O 1s photoelectron spectrum of water, *Chem. Phys. Lett.*, 2003, **380**, 647.
32 G. Öhrwall, R. F. Fink, M. Tchaplyguine, L. Ojamäe, M. Lundwall, R. R. T. Marinho, A. Naves de Brito, S. L. Sorensen, M. Gisselbrecht, R. Feifel, T. Rander, A. Lindblad, J. Schulz, L. J. Sæthre, N. Mártensson, S. Svensson and O. Björneholm, The electronic structure of free water clusters probed by Auger electron spectroscopy, *J. Chem. Phys.*, 2005, **123**, 054310.
33 A. Verdaguer, C. Weis, G. Oncins, G. Ketteler, H. Bluhm and M. Salmeron, Growth and structure of water on SiO_2 films on Si investigated by Kelvin probe microscopy and *in situ* X-ray spectroscopies, *Langmuir*, 2007, **23**, 9699.
34 M. A. Brown, B. Winter, M. Faubel and J. C. Hemminger, Spatial distribution of nitrate and nitrite anions at the liquid/vapor interface of aqueous solutions, *J. Am. Chem. Soc.*, 2009, **131**, 8354.
35 C. Weeraratna, C. Amarasinghe, W. Lu and M. Ahmed, A direct probe of the hydrogen bond network in aqueous glycerol aerosols, *J. Phys. Chem. Lett.*, 2021, **12**, 5503.
36 A. Křepelová, J. Newberg, T. Huthwelker, H. Bluhm and M. Ammann, The nature of nitrate at the ice surface studied by XPS and NEXAFS, *Phys. Chem. Chem. Phys.*, 2010, **12**, 8870.
37 M. Lundholm, H. Siegbahn, S. Holmberg and M. Arbman, Core electron spectroscopy of water solutions, *J. Electron Spectrosc. Relat. Phenom.*, 1986, **40**, 163.
38 M. Favaro, C. Valero-Vidal, J. Eichhorn, F. M. Toma, P. N. Ross, J. Yano, Z. Liu and E. J. Crumlin, Elucidating the alkaline oxygen evolution reaction mechanism on platinum, *J. Mater. Chem. A*, 2017, **5**, 11634.
39 K. A. Stoerzinger, M. Favaro, P. N. Ross, Z. Hussain, Z. Liu, J. Yano and E. J. Crumlin, Stabilizing the meniscus for operando characterization of platinum during the electrolyte-consuming alkaline oxygen evolution reaction, *Top. Catal.*, 2018, **61**, 2152.
40 K. A. Stoerzinger, M. Favaro, P. N. Ross, J. Yano, Z. Liu, Z. Hussain and E. J. Crumlin, Probing the surface of platinum during the hydrogen evolution reaction in alkaline electrolyte, *J. Phys. Chem. B*, 2018, **122**, 864.
41 Y. Han, S. Axnanda, E. J. Crumlin, R. Chang, B. Mao, Z. Hussain, P. N. Ross, Y. Li and Z. Liu, Observing the

electrochemical oxidation of Co metal at the solid/liquid interface using ambient pressure X-ray photoelectron spectroscopy, *J. Phys. Chem. B*, 2018, **122**, 666.
42 M. Favaro, J. Yang, S. Nappini, E. Magnano, F. M. Toma, E. J. Crumlin, J. Yano and I. D. Sharp, Understanding the oxygen evolution reaction mechanism on CoOx using operando ambient-pressure X-ray photoelectron spectroscopy, *J. Am. Chem. Soc.*, 2017, **139**, 8960.
43 H. Ali-Löytty, M. W. Louie, M. R. Singh, L. Li, H. G. Sanchez Casalongue, H. Ogasawara, E. J. Crumlin, Z. Liu, A. T. Bell, A. Nilsson and D. Friebel, Ambient-pressure XPS study of a Ni–Fe electrocatalyst for the oxygen evolution reaction, *J. Phys. Chem. C*, 2016, **120**, 2247.
44 M. Favaro, W. S. Drisdell, M. A. Marcus, J. M. Gregoire, E. J. Crumlin, J. A. Haber and J. Yano, An operando investigation of (Ni–Fe–Co–Ce)Ox system as highly efficient electrocatalyst for oxygen evolution reaction, *ACS Catal.*, 2017, **7**, 1248.
45 S. Axnanda, E. J. Crumlin, B. Mao, S. Rani, R. Chang, P. G. Karlsson, M. O. M. Edwards, M. Lundqvist, R. Moberg, P. Ross, Z. Hussain and Z. Liu, Using ''tender'' X-ray ambient pressure X-ray photoelectron spectroscopy as a direct probe of solid-liquid interface, *Sci. Rep.*, 2015, **5**, 9788.
46 L. P. Ramírez, A. Boucly, F. Saudrais, F. Bournel, J.-J. Gallet, E. Maisonhaute, A. R. Milosavljević, C. Nicolas and F. Rochet, The Fermi level as an energy reference in liquid jet X-ray photoelectron spectroscopy studies of aqueous solutions, *Phys. Chem. Chem. Phys.*, 2021, **23**, 16224.
47 M. Pugini, F. Trinter, I. Wilkinson, B. Credidio, I. Walter, S. Malerz, D. Stemer, U. Hergenhahn, G. Meijer, B. Winter and S. Thürmer, How to measure work functions from aqueous solutions, *Chem. Sci.*, 2023, **14**, 9574.
48 J. Liu, H. Zhang, Y. Li and Z. Liu, Disorder in aqueous solutions and peak broadening in X-ray photoelectron spectroscopy, *J. Phys. Chem. B*, 2018, **122**, 10600.
49 A. I. Krylov, Equation-of-motion coupled-cluster methods for open-shell and electronically excited species: The hitchhiker's guide to Fock space, *Annu. Rev. Phys. Chem.*, 2008, **59**, 433.
50 M. L. Vidal, X. Feng, E. Epifanovsky, A. I. Krylov and S. Coriani, A new and efficient equation-of-motion coupled-cluster framework for core-excited and core-ionized states, *J. Chem. Theory Comput.*, 2019, **15**, 3117.
51 M. L. Vidal, A. I. Krylov and S. Coriani, Dyson orbitals within the fc-CVS-EOM-CCSD framework: theory and application to X-ray photoelectron spectroscopy of ground and excited states, *Phys. Chem. Chem. Phys.*, 2020, **22**, 2693.
52 L. S. Cederbaum, W. Domcke and J. Schirmer, Many-body theory of core holes, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1980, **22**, 206.
53 H. Koch, O. Christiansen, P. Jørgensen, A. M. S. de Meras and T. Helgaker, The CC3 model: An iterative coupled cluster approach including connected triples, *J. Chem. Phys.*, 1997, **106**, 1808.
54 R. H. Myhre and H. Koch, The multilevel CC3 coupled-cluster model, *J. Chem. Phys.*, 2016, **145**, 044111.
55 A. C. Paul, R. H. Myhre and H. Koch, New and efficient implementation of CC3, *J. Chem. Theory Comput.*, 2020, **17**, 117.
56 A. C. Paul, S. D. Folkestad, R. H. Myhre and H. Koch, Oscillator strengths in the framework of equation of motion multilevel CC3, *J. Chem. Theory Comput.*, 2022, **18**, 5246.
57 A. Warshel and M. Levitt, Theoretical studies of enzymatic reactions: Dielectric electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme, *J. Mol. Biol.*, 1976, **103**, 227.
58 H. M. Senn and W. Thiel, QM/MM methods for biomolecular systems, *Angew. Chem., Int. Ed.*, 2009, **48**, 1198.
59 A. I. Krylov and P. M. W. Gill, Q-Chem: An engine for innovation, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2013, **3**, 317.
60 E. Epifanovsky, A. T. B. Gilbert, X. Feng, J. Lee, Y. Mao, N. Mardirossian, P. Pokhilko, A. F. White, M. P. Coons, A. L. Dempwolff, Z. Gan, D. Hait, P. R. Horn, L. D. Jacobson, I. Kaliman, J. Kussmann, A. W. Lange, K. U. Lao, D. S. Levine, J. Liu, S. C. McKenzie, A. F. Morrison, K. D. Nanda, F. Plasser, D. R. Rehn, M. L. Vidal, Z.-Q. You, Y. Zhu, B. Alam, B. J. Albrecht, A. Aldossary, E. Alguire, J. H. Andersen, V. Athavale, D. Barton, K. Begam, A. Behn, N. Bellonzi, Y. A. Bernard, E. J. Berquist, H. G. A. Burton, A. Carreras, K. Carter-Fenk, R. Chakraborty, A. D. Chien, K. D. Closser, V. Cofer-Shabica, S. Dasgupta, M. de Wergifosse, J. Deng, M. Diedenhofen, H. Do, S. Ehlert, P.-T. Fang, S. Fatehi, Q. Feng, T. Friedhoff, J. Gayvert, Q. Ge, G. Gidofalvi, M. Goldey, J. Gomes, C. E. González-Espinoza, S. Gulania, A. O. Gunina, M. W. D. Hanson-Heine, P. H. P. Harbach, A. Hauser, M. F. Herbst, M. Hernández Vera, M. Hodecker, Z. C. Holden, S. Houck, X. Huang, K. Hui, B. C. Huynh, M. Ivanov, Á. Jász, H. Ji, H. Jiang, B. Kaduk, S. Kähler, K. Khistyaev, J. Kim, G. Kis, P. Klunzinger, Z. Koczor-Benda, J. H. Koh, D. Kosenkov, L. Koulias, T. Kowalczyk, C. M. Krauter, K. Kue, A. Kunitsa, T. Kus, I. Ladjánszki, A. Landau, K. V. Lawler, D. Lefrancois, S. Lehtola, R. R. Li, Y.-P. Li, J. Liang, M. Liebenthal, H.-H. Lin, Y.-S. Lin, F. Liu, K.-Y. Liu, M. Loipersberger, A. Luenser, A. Manjanath, P. Manohar, E. Mansoor, S. F. Manzer, S.-P. Mao, A. V. Marenich, T. Markovich, S. Mason, S. A. Maurer, P. F. McLaughlin, M. F. S. J. Menger, J.-M. Mewes, S. A. Mewes, P. Morgante, J. W. Mullinax, K. J. Oosterbaan, G. Paran, A. C. Paul, S. K. Paul, F. Pavošević, Z. Pei, S. Prager, E. I. Proynov, Á. Rák, E. Ramos-Cordoba, B. Rana, A. E. Rask, A. Rettig, R. M. Richard, F. Rob, E. Rossomme, T. Scheele, M. Scheurer, M. Schneider, N. Sergueev, S. M. Sharada, W. Skomorowski, D. W. Small, C. J. Stein, Y.-C. Su, E. J. Sundstrom, Z. Tao, J. Thirman, G. J. Tornai, T. Tsuchimochi, N. M. Tubman, S. P. Veccham, O. Vydrov, J. Wenzel, J. Witte, A. Yamada, K. Yao, S. Yeganeh, S. R. Yost, A. Zech, I. Y. Zhang, X. Zhang, Y. Zhang, D. Zuev, A. Aspuru-Guzik, A. T. Bell, N. A. Besley, K. B. Bravaya, B. R. Brooks, D. Casanova, J.-D. Chai, S. Coriani, C. J. Cramer, G. Cserey, A. E. DePrince, R. A. DiStasio, A. Dreuw, B. D. Dunietz, T. R. Furlani, W. A. Goddard, S. Hammes-Schiffer, T. Head-Gordon, W. J. Hehre, C.-P. Hsu, T.-C. Jagau, Y. Jung, A.

Klamt, J. Kong, D. S. Lambrecht, W. Liang, N. J. Mayhall, C. W. McCurdy, J. B. Neaton, C. Ochsenfeld, J. A. Parkhill, R. Peverati, V. A. Rassolov, Y. Shao, L. V. Slipchenko, T. Stauch, R. P. Steele, J. E. Subotnik, A. J. W. Thom, A. Tkatchenko, D. G. Truhlar, T. Van Voorhis, T. A. Wesolowski, K. B. Whaley, H. L. Woodcock, P. M. Zimmerman, S. Faraji, P. M. W. Gill, M. Head-Gordon, J. M. Herbert and A. I. Krylov, Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package, *J. Chem. Phys.*, 2021, **155**, 084801.

61 S. D. Folkestad, E. F. Kjønstad, R. H. Myhre, J. H. Andersen, A. Balbi, S. Coriani, T. Giovannini, L. Goletto, T. S. Haugland, A. Hutcheson, I.-M. Høyvik, T. Moitra, A. C. Paul, M. Scavino, A. S. Skeidsvoll, Å. H. Tveten and H. Koch, e^T 1.0: An open source electronic structure program with emphasis on coupled cluster and multilevel methods, *J. Chem. Phys.*, 2020, **152**, 184103.

62 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX*, 2015, **1–2**, 19.

63 P. Mark and L. Nilsson, Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K, *J. Phys. Chem. A*, 2001, **105**, 9954.

64 L. R. Pestana, N. Mardirossian, M. Head-Gordon and T. Head-Gordon, Ab initio molecular dynamics simulations of liquid water using high quality meta-GGA functionals, *Chem. Sci.*, 2017, **8**, 3554.

65 J. Liu, X. He and J. Z. H. Zhang, Structure of liquid water – a dynamical mixture of tetrahedral and 'ring-and-chain' like structures, *Phys. Chem. Chem. Phys.*, 2017, **19**, 11931.

66 V. W. D. Cruzeiro, A. Wildman, X. Li and F. Paesani, Relationship between hydrogen-bonding motifs and the $1b_1$ splitting in the X-ray emission spectrum of liquid water, *J. Phys. Chem. Lett.*, 2021, **12**, 3996.

67 J.-D. Chai and M. Head-Gordon, Systematic optimization of long-range corrected hybrid density functionals, *J. Chem. Phys.*, 2008, **128**, 084106.

68 A. Luzar and D. Chandler, Effect of environment on hydrogen bond dynamics in liquid water, *Phys. Rev. Lett.*, 1996, **76**, 928.

69 S. Coriani and H. Koch, Communication: X-ray absorption spectra and core-ionization potentials within a core-valence separated coupled cluster framework, *J. Chem. Phys.*, 2015, **143**, 181103.

70 S. Coriani and H. Koch, Erratum: "Communication: X-ray absorption spectra and core-ionization potentials within a core-valence separated coupled cluster framework" [J. Chem. Phys. 143, 181103 (2015)], *J. Chem. Phys.*, 2016, **145**, 149901.

71 R. Sarangi, M. L. Vidal, S. Coriani and A. I. Krylov, On the basis set selection for calculations of core-level states: Different strategies to balance cost and accuracy, *Mol. Phys.*, 2020, **118**, e1769872.

72 J. Liu, D. Matthews, S. Coriani and L. Cheng, Benchmark calculations of K-edge ionization energies for first-row elements using scalar-relativistic core-valence-separated equation-of-motion coupled-cluster methods, *J. Chem. Theory Comput.*, 2019, **15**, 1642.

73 F. Plasser, A. I. Krylov and A. Dreuw, libwfa: Wavefunction analysis tools for excited and open-shell electronic states, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2022, **12**, e1595.

74 R. H. Myhre, T. J. A. Wolf, L. Cheng, S. Nandi, S. Coriani, M. Gühr and H. Koch, A theoretical and experimental benchmark study of core-excited states in nitrogen, *J. Chem. Phys.*, 2018, **148**, 064106.

75 R. H. Myhre, S. Coriani and H. Koch, X-ray and UV spectra of glycine within coupled cluster linear response theory, *J. Phys. Chem. A*, 2019, **123**, 9701.

76 D. A. Matthews, EOM-CC methods with approximate triple excitations applied to core excitation and ionisation energies, *Mol. Phys.*, 2020, **118**, e1771448.

77 A. M. J. Sanchez de Meras, H. Koch, I. G. Cuesta and L. Boman, Cholesky decomposition-based definition of atomic subsystems in electronic structure calculations, *J. Chem. Phys.*, 2010, **132**, 204105.

78 S. J. Bennie, B. F. E. Curchod, F. R. Manby and D. R. Glowacki, Pushing the limits of EOM-CCSD with projector-based embedding for excitation energies, *J. Phys. Chem. Lett.*, 2017, **8**, 5559.

79 S. D. Folkestad and H. Koch, Equation-of-motion MLCCSD and CCSD-in-HF oscillator strengths and their application to core excitations, *J. Chem. Theory Comput.*, 2020, **16**, 6869.

80 I.-M. Høyvik, R. H. Myhre and H. Koch, Correlated natural transition orbitals for core excitation energies in multilevel coupled cluster models, *J. Chem. Phys.*, 2017, **146**, 144109.

81 A. Landau, K. Khistyaev, S. Dolgikh and A. I. Krylov, Frozen natural orbitals for ionized states within equation-of-motion coupled-cluster formalism, *J. Chem. Phys.*, 2010, **132**, 014109.

82 P. Pokhilko, D. Izmodenov and A. I. Krylov, Extension of frozen natural orbital approximation to open-shell references: Theory, implementation, and application to single-molecule magnets, *J. Chem. Phys.*, 2020, **152**, 034105.

83 S. Saebo and P. Pulay, Local treatment of electron correlation, *Annu. Rev. Phys. Chem.*, 1993, **44**, 213.

84 S. D. Folkestad, E. F. Kjønstad, L. Goletto and H. Koch, Multilevel CC2 and CCSD in reduced orbital spaces: electronic excitations in large molecular systems, *J. Chem. Theory Comput.*, 2021, **17**, 714.

85 L. Zheng, M. Chen, H. Y. Ko, B. Santra, P. Pratikkumar and X. Wu, Structural, electronic, and dynamical properties of liquid water by ab initio molecular dynamics based on SCAN functional within the canonical ensemble, *J. Chem. Phys.*, 2018, **148**, 164505.

86 A. K. Soper and C. J. Benmore, Quantum differences between heavy and light water, *Phys. Rev. Lett.*, 2008, **101**, 065502.

87 K. Pathirannahalage, P. Sachini, N. Meftahi, A. Elbourne, A. C. G. Weiss, C. F. McConville, A. Padua, D. A. Winkler, M. C. Gomes, T. L. Greaves, T. C. Le, Q. A. Besford and A. J. Christofferson, Systematic comparison of the structural and dynamic properties of commonly used water models for molecular dynamics simulations, *J. Chem. Inf. Model.*, 2021, **61**, 4521.

88 R. N. Tazhigulov and K. B. Bravaya, Free energies of redox half-reactions from first-principles calculations, *J. Phys. Chem. Lett.*, 2016, **7**, 2490.