



X-ray emission spectroscopy: a genetic algorithm to disentangle core–hole-induced dynamics

Lars G. M. Pettersson¹ · Osamu Takahashi²

Received: 17 July 2021 / Accepted: 19 November 2021 / Published online: 29 November 2021
© The Author(s) 2021

Abstract

A genetic algorithm (GA) is developed and applied to make proper connections of final-state potential-energy surfaces and X-ray emission (XES) cross sections between steps in the time-propagation of H-bonded systems after a core–hole is created. We show that this modification results in significantly improved resolution of spectral features in XES with the semiclassical Kramers–Heisenberg approach which takes into account important interference effects. We demonstrate the effects on a water pentamer model as well as on two 17-molecules water clusters representing, respectively, tetrahedral (D2A2) and asymmetric (D1A1) H-bonding environments. For D2A2, the applied procedure improves significantly the obtained intensities, whereas for D1A1 the effects are smaller due to milder dynamics during the core–hole life-time as only one hydrogen is involved. We reinvestigate XES for liquid ethanol and, by properly disentangling the relevant states in the dense manifold of states using the GA, now resolve the important $3a''$ state as a peak rather than a shoulder. Furthermore, by applying the SpecSwap-RMC procedure, we reweigh the distribution of structures in the sampling of the liquid to fit to experiment and estimate the ratio between the main *anti* and *gauche* conformers in the liquid at room temperature. This combination of techniques will be generally applicable to challenging problems in liquid-phase spectroscopy.

Keywords X-ray emission · Core–hole-induced dynamics · Genetic algorithm · Liquid ethanol · Water structures

1 Introduction

X-ray emission spectroscopy (XES) provides a direct measurement of the valence electronic structure, which is projected onto a selected atom in a molecule or material [1–4]. In XES, one detects the photon that is emitted when a valence electron decays to an empty level in the core region (core–hole), which gives a direct measure of the energy difference between the valence- and core-level. The involvement of the strongly localized core–hole, created by absorption of an X-ray photon, makes XES a local probe of the

valence electronic structure at the atom with the core–hole through the projection onto the empty core-state. The transition is dipole controlled and, with a spherically symmetric core–hole (e.g., a 1 *s*-level), the intensity is determined by the *p*-character of the valence-level and thus corresponds to an experimental measurement of the *p*-contribution in that valence orbital. Since core levels of different elements are very well separated in energy, XES is element-specific and, by exploiting the chemical shift between atoms of the same species in different environments, the core hole can be created on a selected atom in a molecule [5] or the molecule in different environments [3, 6].

The core–hole has a finite life-time before the decay occurs, which, e.g., for oxygen is of the order 4 fs [7, 8]. This is long enough for structural changes through core–hole-induced dynamics [9] in the system if oxygen, as in water and alcohols, is bonded to a light atom such as hydrogen. This effect can be understood in the equivalent core or $Z + 1$ approximation [10, 11] where, through the removal of a screening charge from the inner shell, the valence electrons effectively experience the nuclear charge corresponding to the next element to the right in the periodic table. In the case

Published as part of the specialcollection of articles “20th deMon Developers Workshop”.

✉ Lars G. M. Pettersson
lgm@fysik.su.se

¹ Department of Physics, AlbaNova University Center, Stockholm University, S-106 91 Stockholm, Sweden

² Basic Chemistry Program, Graduate School of Advanced Science and Engineering, Hiroshima University, Higashihiroshima 739-8526, Japan

of a water molecule that is hydrogen-bonded to other molecules in a sample of water, the core-ionization effectively transforms it into an H_2F^+ . This strongly modifies the initial O–H stretch potential energy surface from being uphill in energy when stretching the O–H bond toward the H-bond-accepting neighboring oxygen to having the minimum at the accepting oxygen in the H-bond for the core-ionized system [12]. The significant release of zero-point energy translates into a wave packet that travels toward the accepting oxygen and effectively delocalizes in the modified H-bond potential well [13]. Experimentally this is seen as an isotope dependence in the XES spectral shape [14–21].

The process thus goes from the electronic and vibrational ground state, via an electronically excited core–hole state in a mixed vibrational state (traveling wave packet), to the valence-hole final state with a distribution of vibrational states that is determined by the projection of the wave packet onto the stationary vibrational states of the final state. The initial state is well-defined and the properties of the final valence-hole state can in principle be interrogated. However, the wave packet in the short-lived intermediate state contains a multitude of vibrational states that cannot be resolved. This gives rise to interference between the different channels from initial to final state [12, 22], similar to the famous double-slit experiment.

XES is a photon-in/photon-out spectroscopy that is best viewed as a one-step inelastic scattering process [2, 23, 24]. Here, we will focus on non-resonant excitation by the initial X-ray photon, *i.e.*, the intermediate state is the core-ionized state without the presence of the excited electron which makes the theoretical treatment independent of the initial photon energy. The cross section (σ) for emission at energy ω' is then given by the Kramers–Heisenberg formula

$$\sigma(\omega') \propto \sum_f \left| \sum_n \frac{\langle f | D'_{FN} | n \rangle \langle n | D_{NI} | i \rangle}{\omega' - (E_n - E_f) + i\Gamma} \right|^2 \quad (1)$$

where the electronic transition moments, *e.g.*, D_{NI} given by $D_{NI}(\mathbf{R}) = \langle N | \hat{D} | I \rangle$, depend on the instantaneous geometry and thus need to be projected also onto the vibrational wave functions. Capital and lower case letters indicate, respectively, electronic and vibrational states on the initial (I), intermediate (N) and final (F) electronic potential energy surface (PES). Γ is the natural line-width of the core-level which, due to the short life-time of the core–hole, is significant—for O 1 *s* 0.18 eV full-width-at-half-maximum [7]. The KH formula contains only electronic and vibrational energy eigenstates, and in this picture, there is no dynamics, only different weights for different transitions. The square of the sum over intermediate vibrational states results in interference effects due to the involvement of the different possible vibrational states. The importance of these interference effects depends strongly on the lifetime broadening Γ

that determines to what extent the intermediate vibrational states can be resolved [2]. For heavy atoms, the vibrational level-spacing is small, such that a large fraction of the possible intermediate vibrational states are within the width Γ and thus contribute to the sum at energy ω' , which then approximates the unit operator and there is no influence of the intermediate state. In this case, the transition is directly from the initial to the final state and there is no influence of dynamics, as also expected from a classical picture based on mass. An alternative way to view this is through the concept of scattering duration time [25]. However, for light atoms, the level-spacing can be significant compared to Γ and it becomes necessary to include the intermediate vibrational states in the treatment of XES.

In order to evaluate the transition moments in such cases, the vibrational wave functions of the initial, intermediate and final states are needed. For low-dimensional problems, the relevant potential surfaces can be mapped out and the full wave-packet propagation performed according to the time-dependent Schrödinger Equation [12, 26–29]. However, for the liquid state, the high dimensionality and necessary sampling of a statistically relevant number of bonding situations make a full wave-packet treatment on a precomputed potential energy surface unfeasible. To remedy this situation, we have developed and implemented [30] a semiclassical approximation [12, 13] to the Kramers–Heisenberg expression that builds on averaging over classical dynamical trajectories on the core–hole state. This corresponds to following the time-development of the probability distribution of the light particle after creation of the core hole and requires a proper sampling of the initial quantum probability distributions of both positions and momenta. This is particularly essential for H-bonded liquid systems where the quantum nature of the proton (deuteron) causes significantly increased width of the quantum probability distributions in both positions and momenta compared to for classical initial conditions. On the short time-scale of the core–hole lifetime, such propagation of sampled probabilities has been shown to agree very well with the probability distribution from a wave packet description for, *e.g.*, water dimer [12, 13].

Due to the low mass of the proton, the time-steps in the dynamics need to be small (here 0.25 fs) and, in addition, the trajectories need to be long enough (20–40 fs) to sufficiently well map out the potential energy surface of the intermediate, as well as all final states. At each time-step, a full spectrum is computed where all transitions are ordered according to their energy. Generating the proper potential energy surface for each final state covering the full trajectory thus leads to a combinatorial problem due to changing energy order between states as the dynamics proceeds. Here, we will present a genetic algorithm (GA) to make these connections.

We will discuss calculations of XES for liquid ethanol and water. In the case of ethanol, we have earlier published computed spectra based on the semiclassical Kramers–Heisenberg formalism in quite good agreement with experiment, except for the high-emission-energy $3a''$ state appearing only as a shoulder rather than a resolved peak [31]. Here, we show that, when curve crossings along the trajectory are properly handled, also this missing peak is resolved in the computed spectrum. Water is more challenging, not only due to there being two protons that undergo dynamics, but also because the structure of the liquid is under strong debate [32] and so is the interpretation of the XES spectrum of the liquid [15, 19, 20, 33–39]. Here, we will limit the discussion to a few examples of how the resolution in the simulated spectra can be improved by proper connections between the multitude of final states along the trajectories.

2 Methods

All spectrum and dynamics calculations were performed using the deMon2k [40] code. The trajectories and computed spectra for ethanol were taken from Ref. [31] while for water a tetrahedrally H-bonded pentamer and two clusters of 17 molecules representing a central single-donor/single acceptor (D1A1) and a double-donor/double-acceptor (D2A2) central molecule were selected for study. These structures were extracted from a TIP4P/Ew [41] trajectory at ambient conditions. Here, we have included core–hole-induced dynamics and will focus on the issues that this leads to. The central molecule was described using the all-electron IGLO-III basis for oxygen while for the surrounding molecules the O $1s$ core was replaced by an effective core potential to make the core level of the targeted central molecule unique. The hydrogens were described using the IGLO-II basis set.

The core–hole-induced dynamics trajectories were calculated using Born–Oppenheimer molecular dynamics (BOMD) on clusters sampled from a regular MD trajectory. The starting point is thus a static finite-size cluster where the central molecule is described at the all-electron level with a half-occupied $1s$ (core hole) while for the surrounding molecules an effective core potential (ECP) replaces the $1s$ shell of all equivalent atoms (here oxygen atoms). This makes the location of the core–hole unique and, by at each time-step selecting the orbital with the lowest eigenvalue to be singly occupied, we furthermore avoid collapse of valence electrons into the core-level as all orbitals are fully variationally relaxed at each time-step. For water, the revised Perdew–Burke–Ernzerhof (PBE) functional [42] was used, while for ethanol the trajectories were from ref. [31] which used the same parameters as in earlier simulations of XES on liquid methanol [43] where the original PBE functional [44] was applied.

For each of the structures, 72 trajectories with a core–hole on the central oxygen were run, using different initial conditions sampling the quantum probability distributions in all nine degrees of freedom. The trajectories were run for a total of 20 fs with a step of 0.25 fs as described above. The spectrum contribution from each possible final state was subsequently calculated at each time-step along each trajectory; interference effects were included using the semi-classical Kramers–Heisenberg approach [12, 13]. Spectra were calculated using the ground-state approximation, *i.e.*, neglecting relaxation effects both in the core-ionized and valence-ionized states [45].

2.1 Semiclassical Kramers–Heisenberg (SCKH)

The dynamics in terms of proton-transfer along H-bonds that are initiated by the creation of the core–hole are most correctly described by a wave packet with time-evolution given by the time-dependent Schrödinger equation. However, for a disordered liquid, the dimensionality of the relevant potential energy surface makes a strict quantum mechanical wave packet treatment prohibitive. However, on the ultrashort time-scale of the core–hole life time, the time-evolution of the initial probability distribution mimics very closely the probability distribution obtained from the time-evolving wave packet as long as a proper sampling of the initial quantum probability distributions in positions and momenta is used to initiate the dynamics [13]. We may thus compute the XES from a given structure as a sum of contributions from an ensemble of classical trajectories with initial conditions sampling the quantum distributions. The working equations in this approach are [12, 13]

$$\sigma(\omega') = \sum_{\text{traj}} \sum_F \left| D_F^+(\omega') \right|^2 \quad (2)$$

where $\sigma(\omega')$ is the cross section, the summation is over trajectories and the transition amplitude $D_F^+(\omega')$ given by

$$D_F^+(\omega') = \int_0^\infty dt D_{NI}^+(0) D_{FN}^+(t) e^{-i \int_0^t (E_F(\tau) - E_N(\tau)) d\tau} e^{-\Gamma t} e^{-i\omega' t} \quad (3)$$

The transition amplitude at frequency ω' is thus obtained through the Fourier transform of the autocorrelation between the initial amplitude for excitation into a specific intermediate state (N) and the amplitude to transition to the final state (F), reduced by the life-time decay factor $e^{-\Gamma t}$. The Fourier transform appears when going from the energy representation of Eq. (1) to the time domain by writing

$$\frac{1}{\omega' - (E_n - E_f) + i\Gamma} = -i \int_0^{\infty} dt e^{-i(E_n - E_f)t - \Gamma t} e^{i\omega' t} \quad (4)$$

Due to the presence of the Fourier transform, it becomes essential to properly connect the different final states from one time-step to the next in order to avoid the loss of resolution that results from jumps in the cross sections or potential energy curves at state crossings (Fig. 1c). We design a genetic algorithm to make these connections.

2.2 Genetic algorithm (GA)

A genetic algorithm (GA) was constructed to connect the correct final states from one time-step to the next to produce curves that are maximally smooth in both the potential energy surfaces and their associated cross sections. The genome was encoded as an (n, m) matrix where the number of rows (n) corresponds to the number of states and the columns (m) to the time-steps. The states in the initial file with transition energy, cross section and transition moments are, at each time-step, ordered from 1 to n , and these values are the entries in the genome matrix. Thus, the initial individual is represented by a matrix in which every row has numerical value equal to the row number such that the original energy order is kept and thus encoding zero curve-crossings. For the ethanol trajectories, the matrix was (204,160) while for water the size was (68,80) resulting in gigantic combinatorial problems that required some care. To illustrate the procedure, we will first use the simpler case of a water pentamer for which the dimensionality is reduced to (20,80).

2.3 GA: initialization

The total initial population is built by making the requested number of replicas of the start individual. For half of these, we loop through the states and select a random step along the trajectory and a random partner state (a state being one of the curves in Fig. 1a) at that step that lies within an energy

threshold (< 0.01 Hartree). The energy difference is normalized by the threshold energy and compared to a random number and, if accepted, this is followed by a test on the difference in cross section between time-steps, comparing crossing over (switching positions) to staying on the same curve. This biases changes toward problematic sections. For the second half, the same procedure is followed, but only comparing a random number to the normalized absolute energy difference between the states at the selected time-step. This generates a sufficiently diverse initial population for the following steps using procreation and mutations.

2.4 GA: mutations

Four different kinds of mutations (mutation rate 1–1.5%) were implemented as either directed or random. To ensure that the best individuals would not be destroyed by mutations, a certain number of the highest ranked individuals were exempt. Random mutations were implemented as either point mutations, where a random exchange was made between two individual entries at one random time-step, or as a switch between two randomly selected states from a random time-step as guided by their normalized energy-difference at the switching point. As more directed mutations were required considering the vast search space, the derivative of the cross section for each state at each point along the trajectory was calculated and the product of the derivatives for pairs of states at points along the trajectory computed. If negative, *i.e.*, one losing and one gaining cross section, then the states were interchanged with probability determined by the mutation rate. Finally, with probability set by the mutation rate, each individual was passed through the crossing-over attempts as when initializing.

2.5 GA: fitness

The fitness of each individual was computed based on the sum of five criteria where the first two were the smoothness of its PES and cross sections as measured by the difference

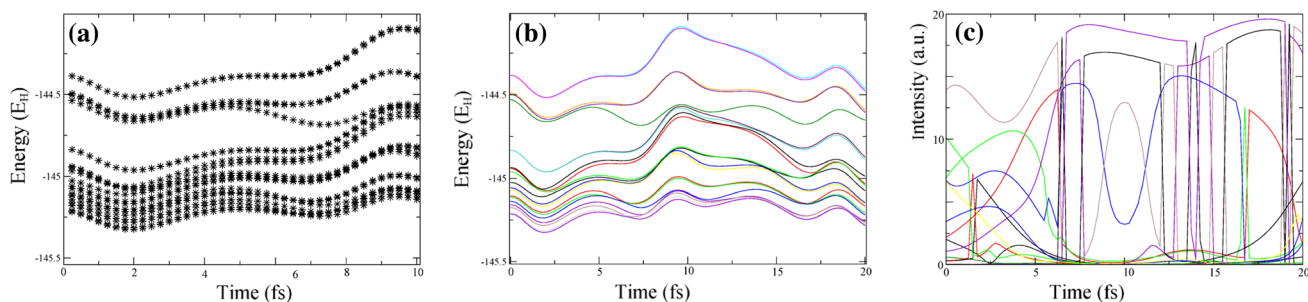


Fig. 1 **A** Final-state energies as function of time-step as calculated along the core-hole-induced trajectory for one initial condition in a water pentamer model. **B** Final-state potential energy surfaces

obtained by ordering states in terms of increasing energy at each time-step. **C** Cross sections along the trajectory with jumps in intensity due to unresolved curve crossings

between the value generated by a five-point Savitzky–Golay filter [46] and the actual energy or cross section at the midpoint of the running filter. As an additional measure, the absolute difference between cross sections at sequential points along the same state was summed to penalize jumps in cross section. Furthermore, a three-point Taylor expansion of the PES was used to predict the energy of the next point and the absolute difference from the real value (according to the individual) was added to the fitness. Finally, in order to eliminate kinks in the PES due to jumps in the curves, the second derivative of the PES was computed at sequential points and the square of the difference added to the fitness. At regular inflection points, this gives a small contribution while jumps up and down between states are associated with large curvature and thus a large contribution to the fitness. Since the range of the values of the cross sections and PES curves are very different, a scaling factor was used on the PES contributions to the fitness to maintain a good balance between the resulting goodness of the two.

2.6 GA: procreation

Once a fitness value was established for each individual, they were sorted in increasing value of the fitness measure. Since the aim was for as low a value as possible, the 50% individuals with the lowest values of the fitness were kept as parents for the next generation. To avoid inbreeding, also the remaining 50% less fit individuals had a chance (set to 2%) of being passed on to contribute to build the next generation.

Procreation was implemented by simply choosing two random individuals among the eligible parents and a random split point (k) so that the generated new individual would have the first section of its genome from parent one and the second half from parent two. In terms of the encoding matrix, this was simply the first k columns from parent one and the remaining from parent two. Apart from participating in procreation, the fittest individual was always transferred directly to the next generation so as not to lose its genetic material. In addition, 25% of the new individuals were subjected to the same crossing-over attempts as when initializing.

2.7 GA: iterations

It was found advantageous to work with two sizes of populations: one small (250–500 individuals) for the iterations that was propagated until no improvements were found for a certain number of generations (typically set to 15). At each such instance, the population was reinitialized based on the best individual using the maximum size population allowed in the code (here 4000 individuals) while retaining untouched the 10% best individuals. These 4000 new individuals were ranked according to fitness, and the specified

number of fittest individuals was retained for the next iterations. The procedure was terminated after a maximum number (200–500) of generations without improvement and a new, reordered trajectory file, as well as individual PES's and their associated cross sections, was written to file.

2.8 GA: handling interacting states

The GA algorithm handles crossings of non-interacting states and states that interact over a few time-steps where the dominant state loses some intensity to the other state, but regains it beyond the reacting region. In the latter case, the effect on the final spectrum is negligible if the interaction is weak and occurring over a limited region. However, for more strongly interacting, temporarily near-degenerate states, a significant loss of cross section can occur that, when the Fourier transform is taken along the individual trajectories, causes loss of resolution even though the cross section has not moved significantly in energy. Thus, in a final separate step, the GA-optimized cross section curves were analyzed for these types of situations and cross sections shared between near-degenerate interacting states were summed into the dominating state while modifying the transition moments to correspond to the summed cross section. This was achieved by setting the transition moments of the lower state to zero and scaling those of the upper state to correspond to the new cross section. Since this modifies the computed transition moments, it has to be done very judiciously and will be reported separately in the following, although experimentally this situation cannot be resolved as long as the energy difference between the involved states is within the experimental resolution (~ 0.3 eV).

3 Results and discussion

The output from the spectrum calculations at each time-step is an energy-ordered list of individual transitions and their cross section (Fig. 1A). Assuming that the energy order is maintained between time-steps results in smooth potential energy surfaces (PES) for each final state, but wildly fluctuating cross sections as the curves in reality should cross as the structure changes during the core–hole-induced dynamics. These fluctuations cause problems in the SCKH formalism due to the Fourier transform in Eq. (3), which, for a cross section that only appears during a limited segment of the trajectory for a given state, results in the contribution being strongly spread out in terms of frequencies/energies. Thus, at each step along the trajectory, it has to be decided which transition each state should be connected to in the following time-step. To simplify the discussion, we will start with a water pentamer model for which limits the number of final states to 20.

3.1 Water pentamer

Figure 1A shows the point-wise computed energies of the 20 valence-ionized final states of a water pentamer model as the geometry changes due to the proton dynamics induced by the created core-hole. The trajectory is 20 fs long with steps of 0.25 fs giving 80 time-steps; the figure is limited to the first 10 fs so that individual time-steps can be resolved. In Fig. 1B, the states have been connected to smooth PES:s by simply keeping the order from one time-step to the next. This procedure does not take into account curve-crossings between non-interacting states with large differences in cross section resulting in the strong jumps along the trajectory as shown in Fig. 1C.

In Fig. 2, we show the results of the GA finding proper connections between time-steps, resulting in continuous curves both for the PES:s (2A) and the cross Sects. (2B). Due to the exponential life-time decay in the integrand in Eq. (3) given by the life-time broadening, Γ (here 0.18 eV⁷) the first 5–10 fs are the most essential for a proper representation of the spectrum. Thus, in this particular case, the exchange of intensity that is evident around 14 fs and ~ 17 fs, with sharp losses in the upper (in terms of cross section) states and corresponding temporary gains in the lower, will not significantly affect the resulting spectrum after life-time vibrational interference is included. However, in order to handle also such instances, the GA is continued in a second round where the focus is on minimizing derivatives in the cross sections between time-steps. This is achieved by searching for near-degenerate interacting partner states that with the cross section of the lower state added into that of the upper state leads to improved fitness according to the same criteria as in the first round of the algorithm. The resulting cross sections are shown in Fig. 3A.

It is clearly seen that the sudden jumps in the cross sections mentioned above are absent after the summation in step 2 of the GA. There are additional state-interactions, most notably around 10 fs where a new state grows up and temporarily takes over most of the intensity of the upper state. The formalism can in principle handle also this situation, *i.e.*, smooth transitions back and forth between states, but the states need to be within the threshold set for near-degeneracy, which here was set to a conservative 0.006 Hartree.

In Fig. 3B, we compare the resulting spectrum contribution from the trajectory with and without properly connecting states along time-steps. It is clear that the ordering improves the resolution of spectral features, in this case most clearly visible in the $1b_2$ region around 524 eV, but a sharpening of features around 528 eV is also evident.

3.2 D2A2 and D1A1 models

We now turn to two more extended water models with 17 molecules representing a central tetrahedrally H-bonded molecule with two donated and two accepted H-bonds (D2A2) and the other, D1A1, with one donated and one accepted. The H-bond criterion used was that of Wernet et al. [47]. Here, the number of valence final states is 68 with 80 time-steps in the trajectory resulting in a combinatorial matrix of (68,80). Since the transition is from an occupied valence orbital to fill the core-hole generated on a specific oxygen (the central molecule in the cluster model), an increasing number of possible final states have negligible intensity as the cluster size is increased. Thus, to simplify the problem, a first scan was made to eliminate final states that showed insignificant intensity along the entire trajectory; these are mainly $2a_1$ (predominantly $2s$) states on the surrounding water molecules. This reduced the combinatorial problem to (53,80). A further reduction to yield two

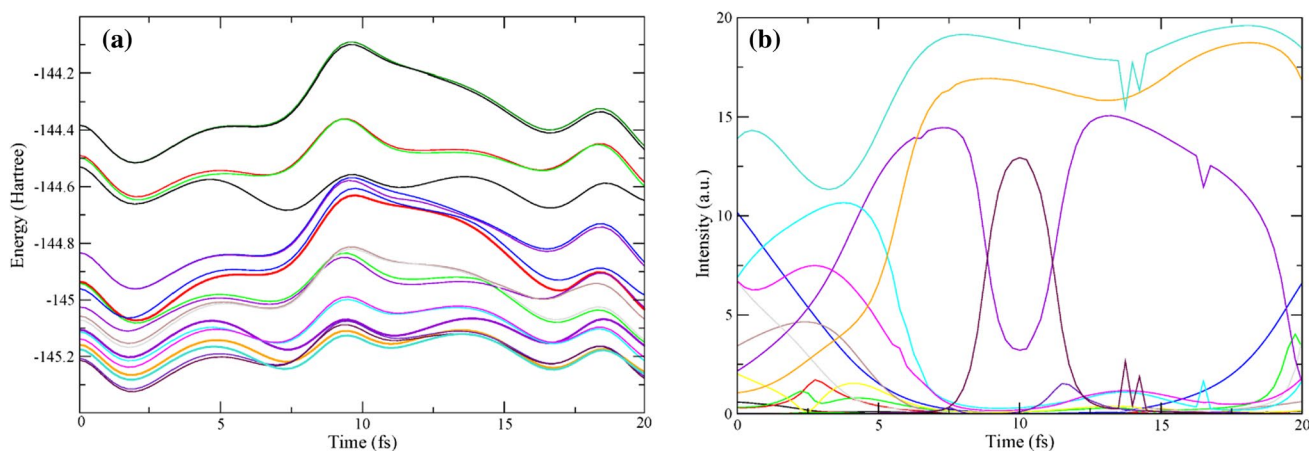


Fig. 2 **A** Pentamer potential-energy surfaces of Fig. 1 after the GA. Some crossing states are highlighted. **B** Cross sections after taking curve crossings into account

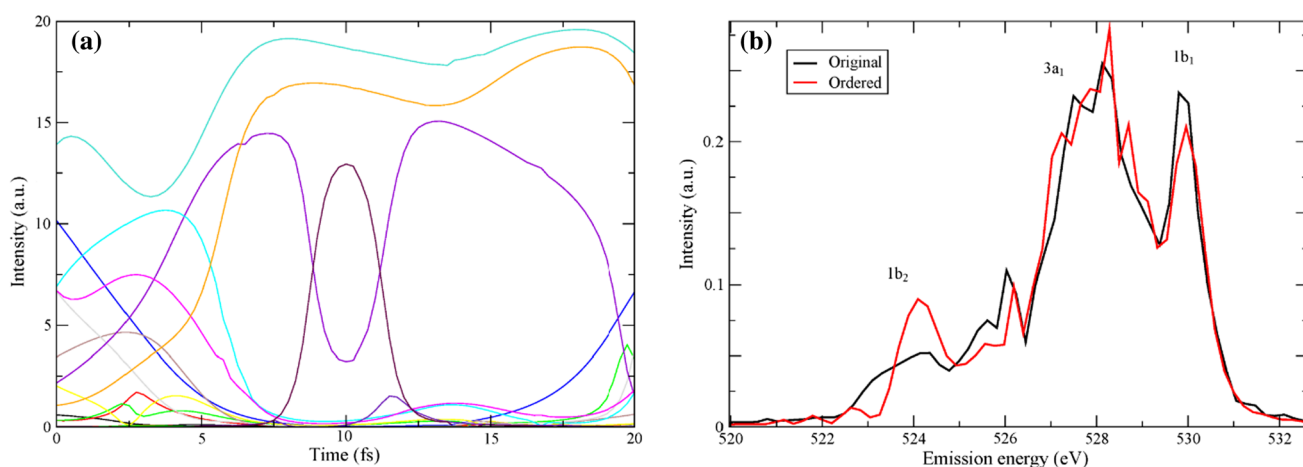


Fig. 3 **A** Pentamer cross sections of Fig. 2 including summation over near-degenerate interacting states. **B** Resulting XES spectrum contribution from the trajectory without (original) and with (ordered)

proper connections between states. The molecular orbital labels indicate the regions to which they mostly contribute

separate problems of (3,80) and (50,80) was possible by identifying a maintained energy gap between remaining $2a_1$ and higher valence states. These reductions were not critical here, but were essential for the simulations of ethanol to be discussed below.

For these clusters, a total of 72 trajectories were run for each, sampling the quantum probability distributions both in positions and momenta based on harmonic modes calculated for the central molecule's internal vibrations, hindered rotations and translations. The cross sections along one such trajectory for the D2A2 model cluster are shown in Fig. 4, where the simply connected PES's yield the distributions in Fig. 4A. Taking into account, non-interacting crossings with the GA gives the time-dependent cross sections in Fig. 4B, while also taking into account near-degenerate interacting states results in Fig. 4C.

The spectra resulting from the three steps are shown in Fig. 5A where it is clear that the resolution is much improved by properly connecting states along the curves;

here, the spectra are built from all 72 trajectories. In particular, the important $1b_1$ lone-pair feature at 529 eV becomes much more well-defined and also the $1b_2$ contribution at 524 eV. In Fig. 5B, we show the corresponding spectra for the D1A1 model, which, in this case, shows significantly smaller effects due to fewer curve-crossings in the initial part of the trajectories. Finally, Fig. 6A compares the D2A2 spectrum contribution to that from the D1A1 model. The shift in energy position between the lone-pair peaks is consistent with the assignment of the experimental XES spectrum for liquid water where the high emission-energy peak is assigned as due to asymmetrically H-bonded molecules (D1A1) while the peak at lower energy is due to tetrahedrally H-bonded [3, 6, 19, 20, 24, 48–50]. The peak shapes are furthermore consistent with this interpretation with more effects of interference for D2A2 resulting in less effects of the dynamics and a more symmetrical line-shape compared to the D1A1 case [12, 19]. This comes about since, in the D2A2 case, both hydrogens undergo dynamics

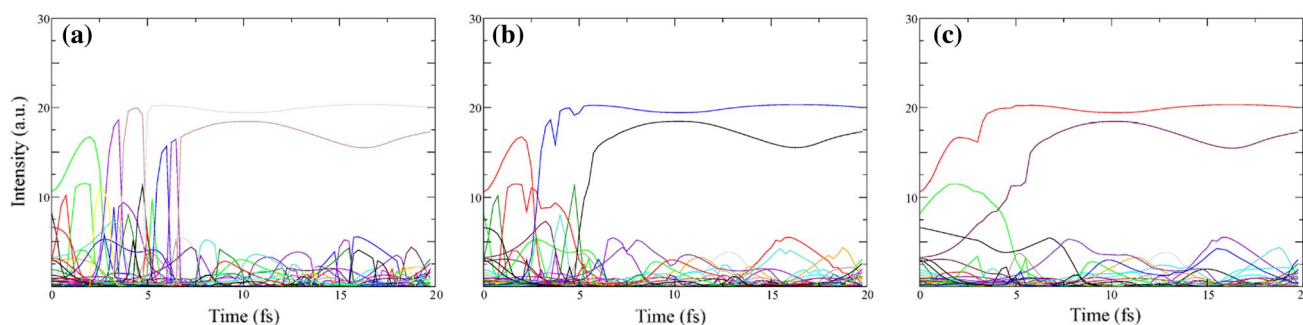


Fig. 4 Cross sections as function of time along one trajectory of the 17 molecule D2A2 water cluster. **A** PESs generated by maintaining the output order for each transition. **B** Cross sections after taking

crossings between non-interacting states into account. **C** Summing cross sections for near-degenerate interacting states

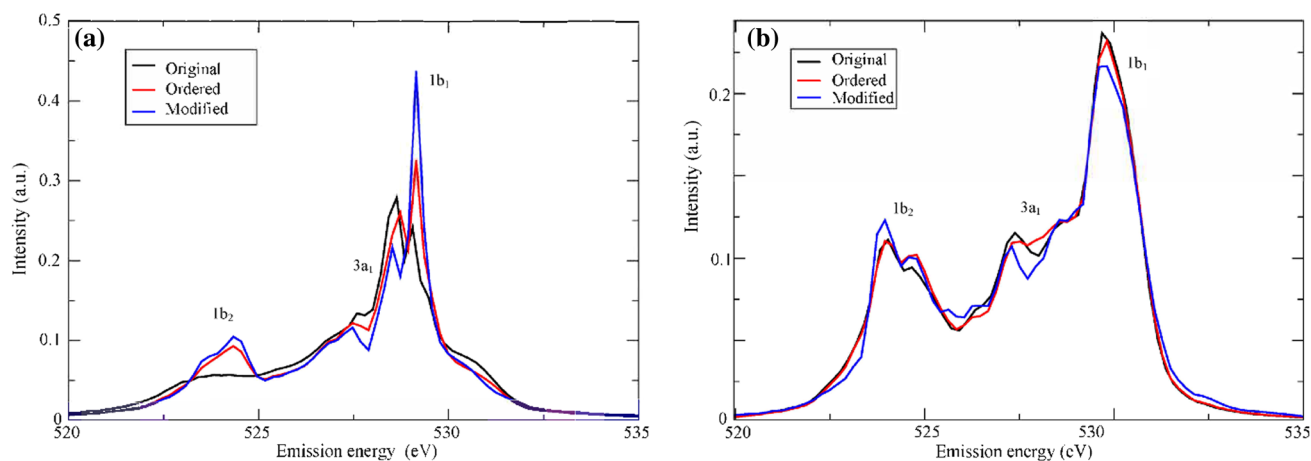


Fig. 5 Computed XES spectra for the two 17-molecule water clusters summed over all 72 initial states. **A** The D2A2 water cluster comparing spectra based on the as computed PESs and cross sections (origi-

nal) with the spectrum after GA (ordered) and after handling near-degeneracies (modified). **B** Same as A, but for the D1A1 cluster. Note the different vertical scales

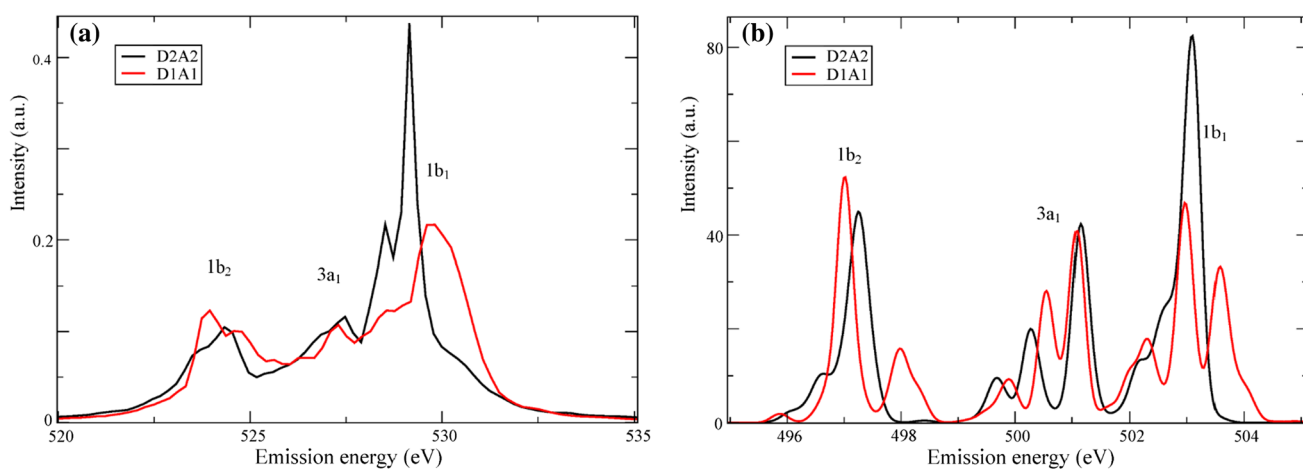


Fig. 6 **A** Comparison of the D2A2 and D1A1 cluster spectra of Fig. 5 including all corrections. **B** D2A2 and D1A1 spectra without including core-hole-induced dynamics, but summed over all 72 initial structures. Note the different scales, where the energy scale in spectrum B is generated as simple orbital energy differences, while

that in A in the SCKH procedure is referenced to the Δ SCF computed total energy difference between the ground and core-ionized states. The core-hole relaxation energy of ~ 27 eV is thus not included in 6B. Relative positions and intensities within each spectrum are significant while the intensity scale between spectra is arbitrary

and contribute intermediate vibrationally excited states that make the summation over these states in Eq. (1) closer to complete and thus closer to a unit operator.

The importance of including the core-hole-induced dynamics when simulating XES spectra of H-bonded systems is illustrated in Fig. 6B. Here, the same 72 trajectories for D2A2 and D1A1 are used, but the spectrum is from only the initial structures, *i.e.*, sampling only the quantum position probability distributions. The main features ($1b_2$, $3a_1$ and $1b_1$) are seen both with and without dynamics (marked in the figure), but with significantly different intensity distributions. The split between the D2A2 and D1A1 lone-pair ($1b_1$) peaks is approximately the same for these two clusters

with and without dynamics. This was the basis for the extensive analysis in ref. [51] where only the initial configurations were included in TDDFT calculations of the deexcitation probabilities and energies from valence states to the core-hole.

3.3 Liquid ethanol

Turning now to ethanol, we apply the GA to the core-hole-induced trajectories used to build the spectra in ref. [31]. In that work, a good agreement with experiment was obtained, which allowed a detailed analysis of the sensitivity to H-bond distributions as well as to the internal geometry in

terms of the distribution of dihedral angles. However, the intensity of the highest-energy feature ($3a''$) was not well represented, appearing as a shoulder rather than a distinct peak. With 17 molecules included in each of the 216 sampled structures, the number of final states becomes 204, which results in a very high density of final states to sort out along the trajectories (Fig. 7). For each structure, 8 initial conditions were used to sample the zero-point vibrational position and momentum distributions in the O–H stretch

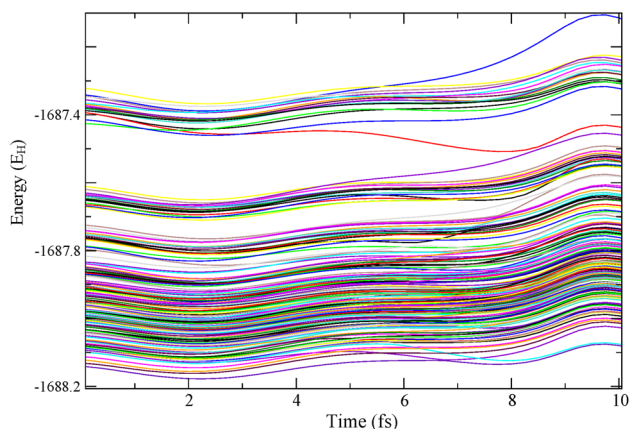


Fig. 7 Final-state potential-energy surfaces for one trajectory of a 17-molecule ethanol cluster. The trajectory length was 40 fs, but was cut at 10 fs for better visibility. Note the density of states in the valence region, i.e., -1687.7 to -1688.2 Hartree

mode giving a total of 1722 trajectories. Of the order, 60 of the final states had cross section below threshold along the full trajectory and could thus be excluded in the GA treatment.

In Fig. 8, we show the results of the GA on cross sections and spectra for one randomly chosen trajectory. Figures 8A and B show the cross sections and spectrum without the GA with large jumps and spikes in the cross sections also within the important first 10 fs, which results in significant loss of resolution and a broadening of the spectrum (8B); only two main peaks are clearly resolved. After applying the GA to resolve non-interacting curve-crossings (8C and D), the cross sections as function of time-step have become significantly smoother, but still with some dips and corresponding spikes as states interact over a limited range of the trajectory. The resolution of spectral features (8D) is significantly improved where, in particular, the $3a''$ peak at 530 eV now appears. In addition, we now resolve a broad feature around 526 eV and the two peaks that are seen in Fig. 8B are better defined. The improved resolution is also reflected in the narrowing of the spectral range as seen from the now sharp drop in intensity above 530 eV.

By also taking into account, near-degeneracies between interacting states the cross sections are further improved (Fig. 8E), which leads to even better defined main peaks and also resolves the broad feature around 523 eV in Fig. 8D as two distinct peaks at 522.5 and 524 eV. By visual inspection of Fig. 8E, it is clear that there are additional interactions

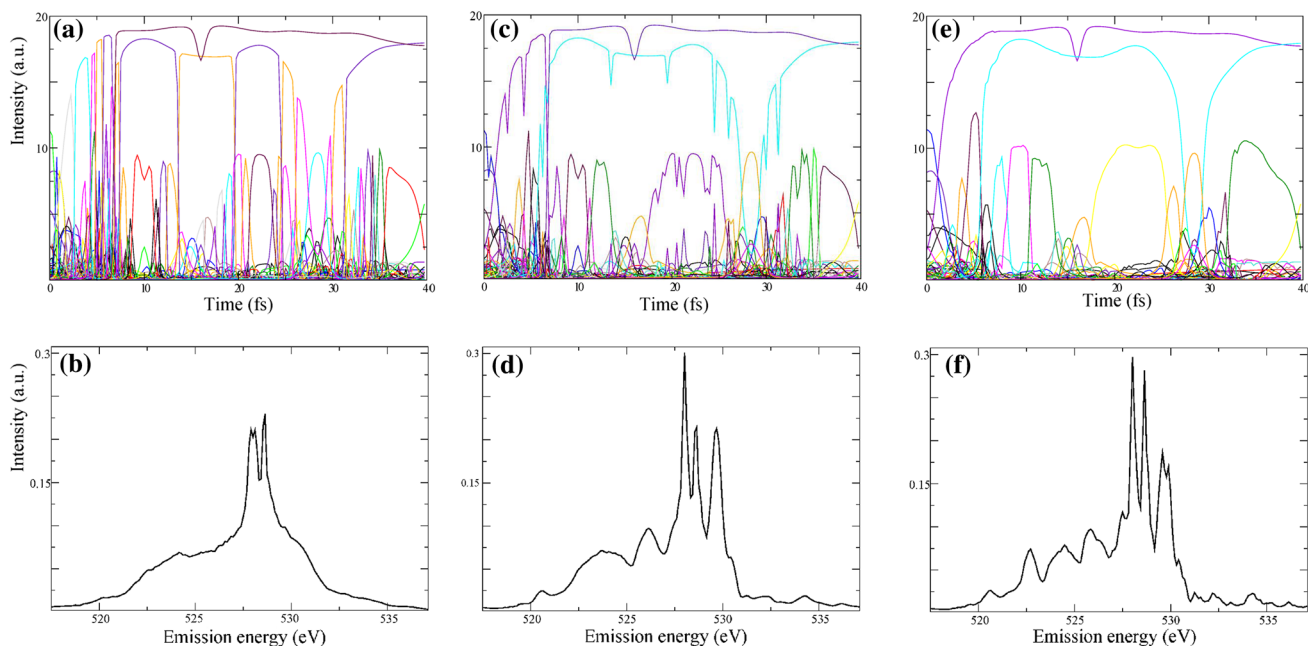


Fig. 8 One trajectory for the 17-molecule ethanol cluster. **A, B** Original cross sections as function of time-step (fs) together with resulting spectrum contribution (eV). **C, D** Cross sections and spectrum after

handling non-interacting curve-crossings. **E, F** Cross sections and spectrum including summation of cross sections for near-degenerate interacting states

present. These are most notable in the rise around 5 fs of the second largest cross section where the two initial spikes likely could be joined as a smooth start of the dominating curve. Similarly, around 10 fs there are four instances where states interact and exchange cross sections. These are, however, separated by more than the conservative energy threshold (0.01 Hartree) applied here. A larger threshold could likely be applied, but a balance between resolution and energy position of the peaks needs to be maintained.

In Fig. 9, we compare the full spectrum over all 1722 trajectories with and without intervention by the GA. As expected from a sampling of instantaneous structures in a disordered liquid, we find that the features of the individual spectrum contributions are in general broadened, but the two highest emission-energy peaks, $10a'$ and $3a''$, are clearly resolved. Most notably, compared to the original spectrum, we now resolve the $3a''$ final state as a peak rather than a shoulder.

In Fig. 10, we compare the final computed spectrum with experiment [21]. It should be noted that the approximation to use orbital energies to estimate the valence binding energies neglects the final-state relaxation, which is larger for deeper valence states than for the more higher-lying. As a consequence, the computed spectrum at lower emission energies than ~ 525 eV can be seen to be compressed compared to experiment. The two marked peaks are obtained in agreement with experiment (after a shift of the computed spectrum by -1.35 eV), albeit the intensity ratio between the $10a'$ and $3a''$ peaks is off.

We have in our earlier study [31] found that the ratio between the two peaks is sensitive to the internal dihedral angle, such that the *anti* configuration (with the OH antiparallel to the terminal methyl group) was seen to give significantly larger contribution to the $10a'$ feature while being reduced in terms of $3a''$ intensity compared to the *gauche*

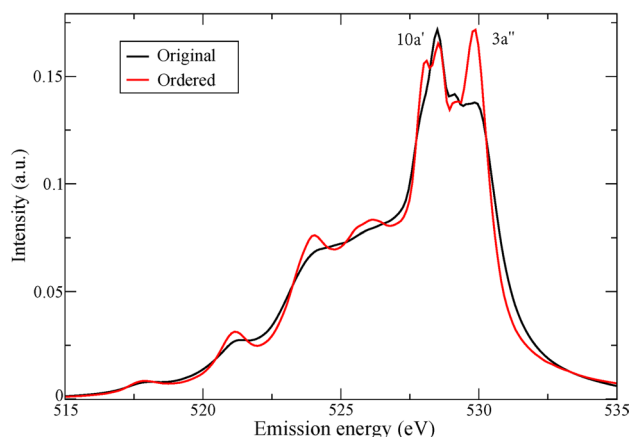


Fig. 9 Computed XES spectra of liquid ethanol sampling 216 random configurations with a total of 1722 trajectories from ref. [31]. Comparison of spectra without (original) and with (ordered) using the GA

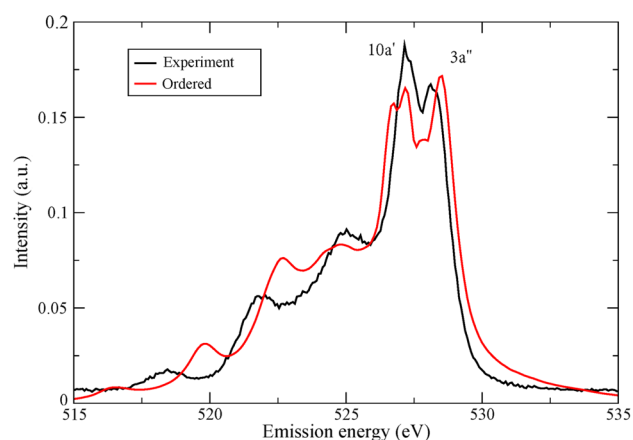


Fig. 10 Comparison of computed and ordered XES spectrum from Fig. 9 with experiment from ref. [21]. The computed spectrum was shifted -1.35 eV

configuration. It is thus likely that a reweighting of the distribution of internal conformations in the simulation of the liquid can give better agreement with the experimental intensity distribution.

In ref. [31], it was suggested that a combination of 60% *gauche* and 40% *anti* best represented the experimental spectrum. Here, we reinvestigate this based on our better-resolved computed spectra by applying the SpecSwap-RMC procedure [52, 53]. This is a Monte Carlo-based method to derive structural models that best fit given experimental data that can be calculated based on a given structure. To allow for also computationally heavy data to be included it uses a library of precomputed data and their associated structures. Thus, instead of making atomistic moves and recomputing the data after each move, it makes a first random selection of a specified number of structures (here 200) and sums their contributions to compare with the experimental data. It then makes random replacements of one structure at a time, compares the new summed contributions with experiment and accepts the move if the agreement is improved. If not, the move is accepted with a probability that depends exponentially on the difference in χ^2 -deviation from experiment. When the procedure has converged, the window of selected structures is interrogated at regular intervals to determine which structures are contained therein and thus build up weights corresponding to how often they are found to contribute. Structures that are important to reproduce the experimental data will be found to contribute more often, while less significant structures obtain smaller weights. These weights can then be applied to the original library of structures to determine what distribution of structures better represents the experimental data. For this to work properly, it is essential that the library of structures is sufficiently diverse and that each associated property is calculated to sufficient accuracy [53]. Since different experiments are

sensitive to different structural properties, it is beneficial to simultaneously fit to several different experimental data sets. Previous applications have combined X-ray diffraction (XRD) and EXAFS measurements on water [54], X-ray absorption spectroscopy (XAS) on hexagonal ice [55], and combining XAS, XES, XRD and NMR data to find possible local structures of liquid water [56]. Here, we apply it to fit the $10a'$ and $3a''$ region (526 to 530 eV) based on the 1722 individual converged XES spectra from the 216 initial structures in our sampling of the MD trajectory for liquid ethanol.

The SpecSwap-RMC works without replacements, i.e., an individual structure can only contribute once to the fit. In order to allow multiple occurrences of particularly fit structures, the library was tripled to contain three copies of each structure giving a total of 5166 basis elements (structures and their associated spectra) after which the procedure was run for 10^8 steps to obtain stable weights. The resulting XES spectrum after applying the weights to sum all spectra in the library is shown in Fig. 11A. The fitted region is between 526 and 530 eV and it is clear that the procedure has resulted in positions and relative intensity of the two main peaks in closer agreement with experiment. To better compare with experiment in terms of the three peaks below 526 eV, the spectral region below 525 eV was linearly stretched in energy with a slope of 0.30 to match the peaks at 522 and 518 eV to account for the neglect of final-state relaxation which becomes successively more important for the deeper valence levels since the ground state orbital energies are used to approximate their energy position.

In Fig. 11B, we compare the distribution of dihedral angles from the library with that obtained after applying the weights that result in agreement with the XES spectrum. The *gauche* structures are enhanced and peaking toward

somewhat higher values while the contribution from *anti* structures in the range 170° – 180° is slightly reduced. We also note an enhanced distinction between the two categories through the reduction between 100° and 140° . The average dihedral angle in the library including all structures is 113.8° , which is slightly reduced to 111.6° with the weights applied. Integrating the areas between 30° and 110° (*gauche*) and between 140° and 180° (*anti*), we now find a 1:1 ratio between the two conformers, which is reasonable considering the small energy difference between the two [57].

4 Conclusions

We have presented a genetic algorithm (GA) to meet the challenge of properly connecting contributions from different final states along core–hole-induced trajectories when computing XES for systems that contain light atoms bound to the probed atom in a molecule. This becomes particularly challenging and important for hydrogen-bonded systems where the creation of the core–hole may dramatically change the potential energy surface and induce significant dynamics. This dynamics takes the form of a wave packet propagating on the intermediate state (i.e., core-ionized state) potential before the decay of a valence electron to fill the core hole occurs, together with emission of an X-ray photon. This leads to several possible channels for the process, i.e., specific intermediate-state vibrational eigenstates that build up the wave packet and lie within the resolution determined by the life-time broadening. These channels interfere, which causes significant redistribution of intensity and affects the shape of the spectrum and thus needs to be taken into account for a quantitative description. This is

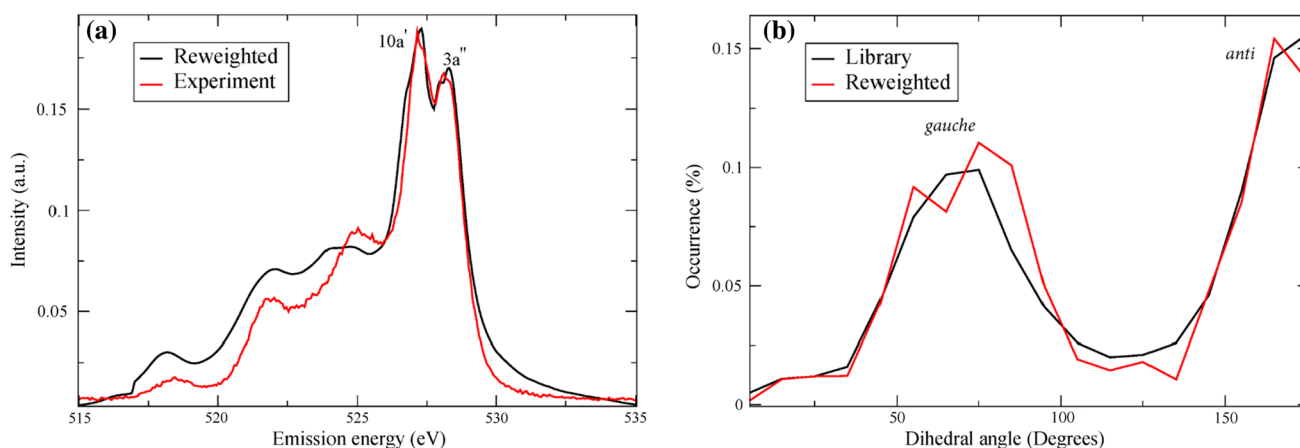


Fig. 11 **A** Comparison of reweighted XES spectrum after SpecSwap-RMC fitting of the energy range 526 to 530 eV. The computed spectrum was treated as described in the text. Note that a linear stretch was applied below 526 eV to compensate for the lack of final-state

relaxation which compresses the spectrum for deeper valence states. **B** Original (library) and reweighted distributions of dihedral angles as result of the SpecSwap-RMC procedure

done based on a semi-classical approximation (SCKH) [12, 13] to the Kramers–Heisenberg description of non-resonant XES. In contrast to a simple summation of spectra along a trajectory with contributions weighted by an exponential decay factor [36, 37], which neglects interference effects, the SCKH is critically dependent on properly connecting different states at consecutive time-steps in order to not lose resolution through the Fourier transform inherent in the SCKH approximation. The here presented GA achieves this objective.

We demonstrate the procedure first on a water pentamer model and then on two randomly selected 17 molecule water clusters where the central molecule is characterized in terms of its H-bonding as either D2A2 or D1A1. The obtained spectra after the GA are consistent with an assignment of the split peaks in the $1b_1$ lone-pair region as due to these types of structures being dominant in the real liquid [3, 6, 19, 20, 49, 50, 58, 59], but do not constitute proof thereof. The procedure was applied to the simulated spectra of liquid ethanol from ref. [31] for which the resulting higher resolution led to the $3a''$ state appearing as a resolved peak rather than a shoulder, giving better agreement with experiment. However, the ratio between the $10a'$ and $3a''$ states as obtained from the statistical sampling of the underlying trajectory [60] was not quite in agreement with the experiment.

We applied SpecSwap-RMC [53] to investigate what type of structures would provide an improved agreement with experiment based on our 1722 individual spectra including interference effects. The result was a close to 1:1 ratio between the two dominating conformers, *gauche* and *anti*, which is consistent with the small energy difference between the two. We suggest that this approach, where various techniques are combined, can be useful in correlating experimental observables with possible underlying structures. Here, the GA is essential to achieve the required resolution in the computed data when dealing with X-ray emission spectroscopy including quantum mechanical interference effects.

Acknowledgements The work was supported by the Swedish Research Council (VR) through grant 2020-05538. The Swedish National Infrastructure for Computing (SNIC) is acknowledged for providing computational resources at the High Performance Computing Center North (HPC2N).

Funding Open access funding provided by Stockholm University.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not

permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Nilsson A, Pettersson LGM (2004) Chemical bonding on surfaces probed by X-ray emission spectroscopy and density functional theory. *Surf Sci Rep* 55:49–167
2. Gel'mukhanov F, Ågren H (1999) Resonant X-ray raman scattering. *Phys Repts* 312:87–330
3. Harada Y et al (2017) Probing the OH stretch in different local environments in liquid water. *J Phys Chem Lett* 8:5487–5491
4. Nilsson A et al (2000) Probing chemical bonding in adsorbates using X-ray emission spectroscopy. *J Electron Spectrosc Relat Phenom* 110/111:15
5. Nilsson A et al (1997) An atom specific look at the surface chemical bond. *Phys Rev Lett* 87:2847
6. Harada Y et al (2013) Selective probing of the OH or OD stretch vibration in liquid water using resonant inelastic soft-X-ray scattering. *Phys Rev Lett* 111:193001
7. Neeb M, Rubensson JE, Biermann M, Eberhardt W (1994) Coherent excitation of vibrational wave-functions observed in core hole decay spectra of O₂, N₂ and CO. *J Electron Spectrosc Relat Phenom* 67:261–274
8. Hennies F et al (2010) Resonant Inelastic scattering spectra of free molecules with vibrational resolution. *Phys Rev Lett* 104:193002
9. Rubensson J-E (2000) RIXS dynamics for beginners. *J Electron Spectrosc Relat Phenom* 110–111:135
10. Mårtensson N, Nilsson A (1995) On the origin of core level binding energy shifts. *J Electron Spectrosc Relat Phenom* 75:209
11. Delesma FA et al (2018) A chemical view on X-ray photoelectron spectroscopy: the ESCA molecule and surface-to-bulk XPS shifts. *ChemPhysChem* 19:169–174
12. Ljungberg MP, Pettersson LGM, Nilsson A (2011) Vibrational interference effects in x-ray emission of a model water dimer: implications for the interpretation of the liquid spectrum. *J Chem Phys* 134:044513
13. Ljungberg MP, Nilsson A, Pettersson LGM (2010) Semi-classical description of nuclear dynamics in X-ray emission of water. *Phys Rev B* 82:245115
14. Benkert A et al (2016) Isotope effects in the resonant inelastic soft X-ray scattering maps of gas-phase methanol. *J Phys Chem A* 120:2260–2267
15. Fuchs O et al (2008) Isotope and temperature effects in liquid water probed by X-ray absorption and resonant X-ray emission spectroscopy. *Phys Rev Lett* 100:027801
16. Odellius M et al (2005) Ultrafast core-hole-induced dynamics in water probed by x-ray emission spectroscopy. *Phys Rev Lett* 94:227401
17. Weinhardt L et al (2012) Nuclear dynamics and spectator effects in resonant inelastic soft x-ray scattering of gas-phase water molecules. *J Chem Phys* 136:144311
18. Weinhardt L et al (2011) Nuclear dynamics in the core-excited state of aqueous ammonia probed by resonant inelastic soft X-ray scattering. *Phys Rev B* 84:104202
19. Tokushima T et al (2010) High resolution X-ray emission spectroscopy of water and its assignment based on two structural motifs. *J El Spec Relat Phenom* 177:192–205
20. Tokushima T et al (2008) High resolution X-ray emission spectroscopy of liquid water: the observation of two structural motifs. *Chem Phys Lett* 460:387–400

21. Schreck S et al (2014) Dynamics of the OH group and the electronic structure of liquid alcohols. *Struct Dyn* 1:054901
22. Gel'mukhanov FK, Mazalov LN, Kondratenko AV (1977) A theory of vibrational structure in the X-ray spectra of molecules. *Chem Phys Lett* 46:133
23. Fransson T et al (2016) X-ray and electron spectroscopy of water. *Chem Rev* 116:7551–7569
24. Nilsson A et al (2013) Resonant inelastic X-ray scattering of water. *J Electron Spectrosc Relat Phenom* 188:84–100
25. Gel'mukhanov F, Salek P, Privalov T, Ågren H (1999) Duration of X-ray Raman scattering. *Phys Rev A* 59:380–389
26. Safek P, Gel'mukhanov F, Ågren H (1999) Wave-packet dynamics of resonant X-ray raman scattering: excitation near the Cl L_{II, III} edge of HCl. *Phys Rev A* 59:1147
27. Vaz da Cruz V et al (2019) Nuclear dynamics in resonant inelastic X-ray scattering and X-ray absorption of methanol. *J Chem Phys* 150:234301
28. Ertan E et al (2018) Ultrafast dissociation features in RIXS spectra of the water molecule. *Phys Chem Chem Phys* 20:14384–14397
29. Da Cruz VV et al (2017) A study of the water molecule using frequency control over nuclear dynamics in resonant X-ray scattering. *Phys Chem Chem Phys* 19:19573–19589
30. Semi-Classical Kramers-Heisenberg code download (GitHub, 2018). <https://github.com/mathiasljungberg/sckh>.
31. Takahashi O, Ljungberg MP, Pettersson LGM (2017) X-ray emission spectrum of liquid ethanol: origin of split peaks. *J Phys Chem B* 121:11163–11168
32. Gallo P et al (2016) Water: a tale of two liquids. *Chem Rev* 116:7463–7500
33. Fuchs O et al (2008) Reply to Comment on “isotope and temperature effects in liquid water probed by X-ray absorption and resonant X-ray emission spectroscopy.” *Phys Rev Lett* 100:249802
34. Pettersson LGM, Harada Y, Nilsson A (2019) Do X-ray spectroscopies provide evidence for continuous distribution models of water at ambient conditions? *Proc Natl Acad Sci (USA)* 116:17156–17157
35. Pettersson LGM et al (2008) Comment on “isotope and temperature effects in liquid water probed by X-ray absorption and resonant x-ray emission spectroscopy.” *Phys Rev Lett* 100:249801
36. Odelius M (2009) Information content in O[1s] K-edge X-ray emission spectroscopy of liquid water. *J Phys Chem A* 113:8176–8181
37. Niskanen J et al (2019) Compatibility of quantitative X-ray spectroscopy with continuous distribution models of water at ambient conditions. *Proc Natl Acad Sci (USA)* 116:4058–4063
38. Niskanen J et al (2019) Reply to Pettersson et al.: why X-ray spectral features are compatible to continuous distribution models in ambient water. *Proc Natl Acad Sci (USA)* 116:17158–17159
39. Yamazoe K, Miyawaki J, Niwa H, Nilsson A, Harada Y (2019) Measurements of ultrafast dissociation in resonant inelastic X-ray scattering of water. *J Chem Phys* 150:204201
40. deMon2k v. 6.1.6 (The deMon developers, Cinvestav, México (2020). See also www.demon-software.com, 2020)
41. Horn HW et al (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J Chem Phys* 120:9665
42. Hammer B, Hansen LB, Nørskov JK (1999) Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals. *Phys Rev B* 59:7413–7421
43. Ljungberg MP, Zhovtobriukh I, Takahashi O, Pettersson LGM (2017) Core-hole-induced dynamical effects in the x-ray emission spectrum of liquid methanol. *J Chem Phys* 146:134506
44. Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. *Phys Rev Lett* 77:3865–3868
45. Föhlisch A et al (2000) The ground state picture of X-ray emission spectroscopy on adsorbates: CO adsorbed on Cu(100). *Phys Rev B* 61:16229
46. Savitzky A, Golay MJE (1964) Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem* 36:1627–1639
47. Wernet P et al (2004) The structure of the first coordination shell in liquid water. *Science* 304:995–999
48. Huang C et al (2010) Reply to Soper et al.: fluctuations in water around a bimodal distribution of local hydrogen bonded structural motifs. *Proc Natl Acad Sci (USA)* 107:45
49. Huang C et al (2009) The inhomogeneous structure of water at ambient conditions. *Proc Natl Acad Sci (USA)* 106:15214–15218
50. Tokushima T et al (2012) Polarization dependent resonant x-ray emission spectroscopy of D₂O and H₂O water; assignment of the local molecular orbital symmetry. *J Chem Phys* 136:044517
51. Zhovtobriukh I, Besley NA, Fransson T, Nilsson A, Pettersson LGM (2018) Relationship between X-ray emission and absorption spectroscopy and the local H-bond environment in water. *J Chem Phys* 148:144507
52. SpecSwap-RMC download (2013). <https://github.com/leetmaa/SpecSwap-RMC>
53. Leetmaa M, Wikfeldt KT, Pettersson LGM (2010) SpecSwap-RMC: a novel reverse Monte Carlo approach using a discrete configuration space and pre-computed properties. *J Phys Condens Mater* 22:135001
54. Wikfeldt KT, Leetmaa M, Mace A, Nilsson A, Pettersson LGM (2010) Oxygen-oxygen correlations in liquid water; addressing the discrepancy between diffraction and EXAFS using a novel multiple-data set fitting technique. *J Chem Phys* 132:104513
55. Zhovtobriukh I, Norman P, Pettersson LGM (2019) X-ray absorption spectrum simulations of hexagonal ice. *J Chem Phys* 150:034501
56. Zhovtobriukh I, Cabral BJC, Corsaro C, Mallamace D, Pettersson LGM (2019) Liquid water structure from X-ray absorption and emission, NMR shielding and X-ray diffraction. *Sci China* 62:107010
57. Sekine M, Nakata M (2011) Reinvestigation of the energy difference between anti and gauche ethanol in low-temperature argon and xenon matrices. *Chem Phys Lett* 508:33–37
58. Nilsson A, Pettersson LGM (2011) Perspective on the structure of liquid water. *Chem Phys* 389:1–34
59. Nilsson A, Pettersson LGM (2015) The structural origin of anomalous properties of liquid water. *Nature Commun* 6:8998
60. Lehtola J, Hakala M, Hämäläinen K (2010) Structure of liquid linear alcohols. *J Phys Chem B* 114:6426–6436

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.