Solvent Effects and pH Dependence of the X-ray Absorption Spectra of Proline from Electrostatic Embedding Quantum Mechanics/Molecular Mechanics and Mixed-Reference Spin-Flip Time-Dependent Density-Functional Theory

Marc Alías Rodríguez, Simone Bonfrate, Woojin Park, Nicolas Ferré, Cheol Ho Choi, Miquel Huix-Rotllant

Solvent Effects and pH Dependence of the X-ray Absorption Spectra of Proline from Electrostatic Embedding Quantum Mechanics/Molecular Mechanics and Mixed-Reference Spin-Flip Time-Dependent Density-Functional Theory

Marc Alías-Rodríguez,*,† Simone Bonfrate,†,¶ Woojin Park,‡,¶ Nicolas Ferré,† Cheol Ho Choi,*,‡ and Miquel Huix-Rotllant*,†

†*Aix-Marseille Univ, CNRS, ICR, Marseille 13013, France.*
‡*Department of Chemistry, Kyungpook National University, Daegu 41566, South Korea*
¶*Contributed equally to this work*

E-mail: marc.alias-rodriguez@univ-amu.fr; cchoi@knu.ac.kr; miquel.huix-rotllant@cnrs.fr

Abstract

The accurate description of solvent effects on X-ray absorption spectra (XAS) is fundamental for comparing simulated spectra with experiments in solution. Currently, few protocols exist that can efficiently reproduce the effects of the solute/solvent interactions on XAS. Here, we develop an efficient and accurate theoretical protocol for simulating the solvent effects on XAS. The protocol combines electrostatic embedding QM/MM based on electrostatic potential fitted (ESPF) operators for describing the solute/solvent interactions and mixed-reference spin-flip time-dependent density functional theory (MRSF-TDDFT) for simulating accurate XAS spectra. To demonstrate the capabilities of our protocol, we compute the X-ray absorption of neutral proline in the gas phase and ionic proline in water in all relevant K-edges, showing an excellent agreement with experiments. We show that states represented by core to π^* transitions are almost unaffected by the interaction with water, whereas the core to σ^* transitions are more impacted by the fluctuation of proline structure and the electrostatic interaction with the solvent. Finally, we reconstruct the pH-dependent X-ray absorption spectra of proline in solution, determining that the N K-edge can be used to distinguish its three protonation states.

1 Introduction

X-ray spectroscopy is gaining renewed interest due to the recent developments of new laser sources with ultrashort pulses like X-ray free electron lasers (XFELs). Such techniques can provide sufficient resolution even at sub-femtosecond timescales, opening the way for developing new experimental techniques based on attosecond pulses. Among the X-ray spectroscopic techniques, X-ray absorption (XAS) is perhaps one of the most widely used and well-established techniques, providing at the same time electronic information from the Near-Edge X-ray Absorption Fine Structure (NEXAFS) and geometric information from the extended X-ray absorption fine-structure (EXAFS).[1] X-ray spectroscopies are an element-specific technique that is particularly useful for following up on the dynamics of chemical reactions.[2] Indeed, time-resolved (TR) XAS spectroscopy provides a wealth of information on (photo)chemical re-

activity in organic and organometallic complexes.[3]

X-ray absorption directly probes core-excited states by promoting core electrons to unoccupied virtual orbitals, following the usual dipole selection rules for absorption. In the K-edge spectra of most common organic molecules, the 1s electron can be used to probe transitions to π^*, σ^*, or Rydberg orbitals. These states can be affected by the electrostatic interaction with the solvent as well as the structural fluctuations due to the temperature. Solvent effects in the XAS spectra are well-documented experimentally for many chemical systems,[3–7] but only a few theoretical protocols exist in the literature capable of efficiently simulating solvent effects on XAS.[4,8–15]

Here we develop a new model for describing efficiently and accurately the solvent effects on the X-ray spectra. Our model is based on a combination of the electrostatic embedding QM/MM method, used to equilibrate a water droplet around the molecule of interest, and mixed-reference spin-flip time-dependent density functional theory (MRSF-TDDFT) coupled to QM/MM to extract the X-ray spectroscopy. MRSF-TDDFT has been recently shown to be able to efficiently and accurately extract XAS, with errors that are within 1 eV of the experimental spectra.[16] In addition, an MRSF-TDDFT/MM coupling was shown to lead to accurate excited-state non-adiabatic dynamics of thymine in water.[17]

As a test case, we compute the XAS spectrum of the aminoacid proline in water. Proline is the only standard amino acid with a secondary amino group. The interaction of zwitterionic proline with water is well documented, indicating that the carboxylate group accepts less than two hydrogen bonds and the amine hydrogen atoms donate less than one hydrogen bond to the surrounding water molecules.[18] Proline has a rich XAS structure for the oxygen, nitrogen, and carbon K-edges,[7,19] and the corresponding spectra depend on the pH.[6,20] Indeed, at neutral pH, proline exists in a zwitterionic form, in which the carboxylic group is deprotonated and the amino group is protonated. At very acidic pH (<2), the carboxylic group is protonated, the proline being in a cationic form, whereas at very basic pH (>11) the amino group is deprotonated and the proline is in anionic form.

This study is organized as follows: first, we describe the MRSF-TDDFT methodology for X-ray spectroscopy and its electrostatic embedding coupling to the solvent via electrostatic potential fitted (ESPF) charges; second, we describe a theoretical protocol to simulate the solvent effects in X-ray spectroscopy via the combination of QM/MM molecular dynamics and MRSF-TDDFT/MM method; finally, we describe the solvent effects on proline by comparing the gas phase spectra, the spectrum generated from molecular dynamics snapshots with and without electrostatics for the three protonation forms of proline.

2 Methods

2.1 Protocol for X-ray spectra simulations in solution

The simulation of solvent effects on X-ray spectra implies the description of the solute-solvent interactions as well as the inhomogeneous broadening due to the fluctuations of the solute in the solvent that acts as a thermal bath. In Fig. 1, we summarize the main steps of our protocol. This protocol makes use of QM/MM partitioning, describing accurately only transition densities that fit well within the spatial extension of the QM fragment and the employed basis set.

Step 0: Model preparation. The preliminary step is the construction of a model system containing the solute and the solvent. To this aim, the gas-phase ground state optimized geometry of the neutralized solute is placed in the center of a box of pure solvent, the temperature and the pressure of which have been equilibrated at MM level by two following dynamics propagated respectively in the NVT and NPT ensembles using the Berendsen thermostat and barostat. In both cases, Periodic Boundary Conditions (PBC) have been employed. A droplet of suitable radius, defined by testing the

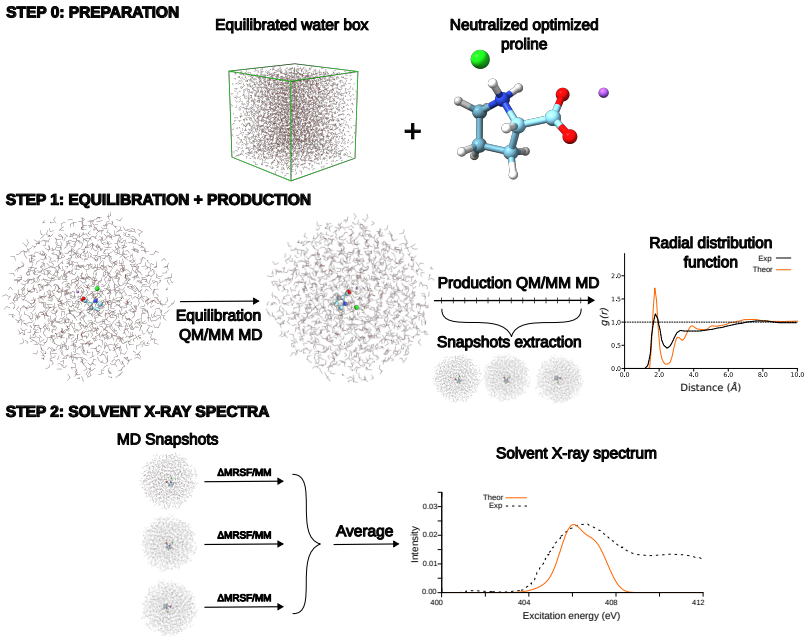


Figure 1: Schematic representation of the theoretical protocol for simulating solvent effects of X-ray spectra.

dependence of the XAS spectra on droplet size, is then extracted from the box.

Step 1: Equilibration and production. The droplet structure is firstly minimized at the QM/MM level, where the solute is described at the QM level while the solvent is described classically. Spherical boundary conditions are imposed on all the simulations on the droplet system to prevent diffusion of the outer shell of the solvent. The QM/MM electrostatic embedding method is based on electrostatic potential fitted (ESPF) operators, as briefly described in Sec. 2.2.

Once the system has been relaxed, a two-step equilibration is done. The droplet is firstly equilibrated in the NVT ensemble with QM/MM molecular dynamics (MD) in which the Berendsen thermostat is used until it reaches the desired temperature. Second, the equilibration is further refined using the Bussi thermostat. An NVT production run using the Bussi thermostat is then performed. If available, a comparison with experimental radial distribution functions is a good test of the convergence of the trajectories. The length of the production run will depend on the desired number of independent snapshots that would be extracted from the trajectory, which is imposed by the autocorrelation function.

Step 2: X-ray spectrum in solution. The lowest states involving core excitations are calculated at ΔMRSF/MM level on the different snapshots extracted from MD. The ΔMRSF/MM method is briefly described in Sec. 2.3. For such calculations, a mixed uncontracted atomic basis set is employed for converging the density at the nuclear region of the core of interest, and a contracted basis set that can describe correctly the valence states for the rest of atoms. The accuracy of ΔMRSF depends on the choice of exchange-correlation functional and basis set, which can be validated by comparing to experimental gas phase spectra. Finally, the XAS spectra are simulated assigning an established Gaussian broadening to each X-ray state.

2.2 Solute-solvent interactions via electrostatic embedding QM/MM

The MRSF-TDDFT has been recently coupled to an MM electrostatic environment via

an electrostatic embedding QM/MM approach based on electrostatic potential fitted (ESPF) charges.[17] The main advantage of ESPF electrostatic embedding is the reduced number of integrals, which only depends on the number of QM atoms and the numerical grid for computing the ESPF atomic charges.[21] Here, we summarize the main features of the model. The total energy of the QM/MM system can be decomposed as a sum of three terms,

$$E = E_{QM} + E_{MM} + E^{int}\,, \tag{1}$$

where E_{QM} is the energy of the QM subsystem in the gas phase, E_{MM} is the energy of the MM subsystem plus the van der Waals interactions between QM and MM subsystems and E^{int} consist in the energy of the electrostatic interactions between the two subsystems. Here, we restrict our model to non-polarizable MM force fields. In the ESPF approach, it is convenient to define the QM/MM interaction energy in terms of Coulombic potential as

$$E^{int} = \sum_{A=1}^{N_{QM}} q_A \Phi_A^{MM}\,. \tag{2}$$

The Coulombic potential felt by a QM atom A at position $\mathbf{r}_A$ is given by,

$$\Phi_A^{MM} = \sum_{j=1}^{N_{MM}} \frac{q_j}{|\mathbf{r}_{Aj}|}\,. \tag{3}$$

in which $\mathbf{r}_{Aj} = \mathbf{r}_A - \mathbf{r}_j$. In the equation, $\mathbf{r}_j$ and q_j are respectively the coordinates and the charge of MM atom j, and N^{QM} and N^{MM} are respectively the number of QM and MM atoms. The ESPF partial charges at the QM centers, $q_A = Z_A - Q_A$, are obtained as the difference between the nuclear charge Z_A, and the atomic charge population $Q_A = \sum_{\mu\nu} P_{\mu\nu} Q_{A,\mu\nu}$ which in turn is derived by tracing the one-electron density matrix $\mathbf{P}$ with the atomic charge operator matrix $\mathbf{Q}_A$. The elements of the charge operator matrix are obtained by solving a linear set of equations to fit the electrostatic integrals computed on the Lebedev atom-centered grid constructed around the QM subsystem,[21,22]

$$\sum_{A}^{N_{QM}} \frac{Q_{A,\mu\nu}}{|\mathbf{r}_{kA}|} = V_{k,\mu\nu}\,. \tag{4}$$

In the equation, the electrostatic integrals on a grid are defined as

$$V_{k,\mu\nu} = \int d^3\mathbf{r} \chi_\mu^*(\mathbf{r}) \frac{1}{|\mathbf{r} - \mathbf{r}_k|} \chi_\nu(\mathbf{r})\,, \tag{5}$$

where $\mathbf{r}_k$ are the coordinates of the grid points and χ are the atomic orbitals. A correction is applied to the charge operator to enforce the conservation of the total charge of the QM subsystem.[22] Deriving the interaction energy (Eq. 2) with respect to the density matrix $\mathbf{P}$, we can obtain the QM/MM interaction in the form of an interaction operator, simply given by

$$h_{\mu\nu}^{int} = -\sum_{A}^{N_{MM}} Q_{A,\mu\nu} \Phi_A^{MM}\,, \tag{6}$$

that is added as a one-electron Hamiltonian contribution to the gas phase Fock operator. For the calculations of ΔMRSF, this accounts for adding $h_{\mu\nu}^{int}$ in the restricted open-shell Kohn-Sham (ROKS) Fock operator, with no extra terms to be added in the response equations.[17]

It is further possible to compute the gradient of the total QM/MM energy (Eq. 1) as $E^x = E_{QM}^x + E_{MM}^x + E^{int,x}$. Hereafter, the superscript $E^x = \partial E/\partial x$ refers to the derivative with respect to an x atom coordinate. The analytical derivative E_{QM}^x depends on the *ab initio* method employed, but it is generally available for the most common electronic structure methods. Likewise, analytical expressions for E_{MM}^x are available. Here we focus on the derivatives of the interaction energy with respect to QM and MM atom coordinates.[22] The derivative with respect to MM atom coordinates is simply given by

$$E^{int,x_i} = \sum_{A=1}^{N_{QM}} q_A \Phi_A^{MM,x_i} \tag{7}$$

where the derivative of the Coulombic potential

with respect to an MM atom is

$$\Phi_A^{MM,x_i} = -q_i \frac{x_{Ai}}{|\mathrm{r}_{Ai}|^3}\,. \tag{8}$$

The derivative of the interaction energy with respect to the QM atom is given by

$$E^{int,x_B} = \sum_{A=1}^{N_{QM}} \left(q_A \Phi_A^{MM,x_B} - Q_A^{(x_B)} \Phi_A^{MM} \right), \tag{9}$$

where Φ_A^{MM} is the derivative of the Coulombic potential (Eq. 3) with respect to the same QM coordinate, which is given by

$$\Phi_A^{MM,x_B} = -\sum_{j=1}^{N_{MM}} q_j \frac{x_{Aj}}{|\mathrm{r}_{Aj}|^3} \delta_{AB}\,. \tag{10}$$

On the other hand, $Q_A^{(x_B)}$ is the derivative of the ESPF charge with respect to the QM coordinate x_B. The parenthesis in the derivatives indicates that the term corresponding to the derivative of the density matrix is excluded (see Ref. 22 for further details). The construction of the ESPF charge operator derivative is straightforwardly obtained by solving a set of linear equations

$$\sum_{A}^{N_{QM}} \frac{Q_{A,\mu\nu}^{x_A}}{|\mathbf{r}_{kA}|} = V_{k,\mu\nu}^{x_A} - \sum_{A}^{N_{QM}} Q_{A,\mu\nu} \left[\frac{1}{|\mathbf{r}_{kA}|} \right]^{x_A}. \tag{11}$$

2.3 X-ray spectrum from mixed-reference spin-flip time-dependent density-functional theory

MRSF-TDDFT[23,24] is a promising new variant of linear response (LR) theory, which overcomes the limitations of LR- as well as Spin-Flip (SF)-TDDFT.[25] The essence of the MRSF-TDDFT is in the utilization of two references, which naturally expands the response space resolving issues such as spin-contamination of SF-TDDFT.[26] For its realization, a new mixed reference state is introduced that has an equiensemble of $M_S = +1$ and $M_S = -1$ components of a triplet state, *e.g.*,

$$\rho_0^{\mathrm{MR}}(x) = \frac{1}{2} \left\{ \rho_0^{M_S=+1}(x) + \rho_0^{M_S=-1}(x) \right\}, \tag{12}$$

where the idempotency of $\rho_0^{\mathrm{MR}}(x)$ was recovered by the concept of *hypothetical* single reference.[23]

The responses from the two components allow exact external spin adaptations, which yields completely decoupled linear-response equations for singlet and triplet excited states as,[23]

$$\begin{aligned} &\sum_{rs} \left(A_{pq,rs}^{(k)(0)} + A_{pq,rs}^{\prime(k)} \right) X_{rs}^{(k)} \\ &= \Omega_{(k)} X_{pq}^{(k)}, \qquad k = S, T \end{aligned} \tag{13}$$

where $k = S, T$ labels singlet and triplet states, $A_{pq,rs}^{(k)(0)}$ is an orbital Hessian matrix derived by the linear response, and $A_{pq,rs}^{\prime(k)}$ is a coupling matrix between configurations originating from different components, $M_S = +1$ and $M_S = -1$, of the mixed reference.[23,24] $X_{pq}^{(k)}$ and $\Omega_{(k)}$ are the amplitude vectors and the excitation energies with respect to the reference state, respectively. As a result, not only the spin-contamination but also the ambiguity with the cumbersome identification of the resulting response states of SF-TDDFT are eliminated.

An important advantage of MRSF-TDDFT over the usual linear-response TDDFT[27–29] is that MRSF-TDDFT enables the proper computation of the S_1/S_0 conical intersections, as both states are treated at the same level of theory in MRSF-TDDFT.[30] Additionally, MRSF-TDDFT not only includes singly but also some important doubly excited configurations that are missing in the LR-TDDFT,[31] significantly improving its prediction accuracy in the study of diradicals[32] and multi-configurational states.[33,34]

The particular formulation of MRSF-TDDFT combined with maximum-overlap method (MOM) also provides a significant advantage in the computation of X-ray absorption.[16,35] Although LR-TDDFT[29] has been the most popular method for valence excited state studies, the accuracy of TDDFT to predict experimental X-ray states is very crude, with errors as large as ∼10 eV for second-period elements like C, N,

O, and F,[36] and even larger errors for heavier elements. These large shifts are attributed to the lack of description of *core-hole orbital relaxation* effects in the approximate exchange-correlation functionals used in TDDFT, which cannot account for the drastic change in the Coulomb interaction of the remaining electrons when a core electron is excited. Properly tuned self-interaction corrected range-separated exchange functionals can remedy the large errors in core-excitation energies by DFT methods.[37] However, a simple remedy to orbital relaxation is not available in the context of the TDDFT approach. These major challenges can be easily overcome by MRSF-TDDFT with the help of its high-spin triplet reference, which allows a core-hole particle (CHP) pair. The particular ΔCHP-MRSF flavor (referred to for simplicity as ΔMRSF in the following) was designed in our previous work,[16] that we briefly summarize here. In the ΔMRSF approach, the excitation energy is calculated as the energy difference between the X-ray state computed with MRSF-TDDFT using a core-hole triplet state as the reference and the singlet ground state computed at the restricted Kohn-Sham level. The orbital relaxation of the CHP pair configuration can be accomplished by replacing the two singly occupied open (O1 and O2) orbitals of restricted open-shell Hartree–Fock (ROHF), or ROKS applying DFT, reference with the particular core (the 1s orbital for XAS) and the virtual orbitals with MOM.[35] In that way, the orbital relaxation due to the excitation of a core electron is embedded in the orbital energies, leading to a larger accuracy of the spectrum. Finally, to simplify the excitation space, a restricted excitation window (REW) is also applied to all calculations. This restricts MRSF-TDDFT excitation space from unwanted occupied orbitals. This simple strategy exhibited near-perfect predictions with RMSE $\sim$ 0.5 eV, featuring a median value of 0.3 and an interquartile range of 0.4 for K-edge excitation energies of 24 second- and 17 third-row molecules.[16]

3 Computational details

The model for proline in the gas phase has been constructed by performing minimum energy geometry optimization of the ground state. The geometries without counterions are optimized in GAMESS-US at BH&HLYP/6-31G* level. For the zwitterionic case, the N-H distance has been kept frozen during the process in order to avoid a proton transfer to the carboxylic group. The geometries including Na^+ and Cl^- counterions are fully relaxed without constraints.

For including the solvent effects, the gas phase proline structures have been soaked in a pre-equilibrated TIP3P[38] water cubic box with an edge size of 50 Å. From such structures, we proceed to extract spherical droplets with a radius of 20 Å centered at C_α of proline. These systems have been first relaxed with a QM/MM minimization at the HF/3-21G//Amber99 level,[39] where the QM sub-region includes proline and counterions, whereas the MM region contains all the water molecules. A spherical solvent boundary has been applied to prevent the diffusion of the outer water molecules during the simulations.[40]

After minimization, the systems were equilibrated at QM/MM MD level using the same conditions as the minimization (HF/3-21G//Amber99 in which the QM sub-region includes proline and counterions). A first equilibration was run for 150 ps in the canonical ensemble using the Berendsen thermostat with a target temperature of 298 K. The QM/MM MD equilibration was further refined at the same temperature for another 50 ps using the Bussi thermostat with spherical solvent boundary (see Fig. S1 in the supporting information.) Finally, we proceeded to the production run for 150 ps using the Bussi thermostat. A set of 50 independent snapshots has been extracted from each production trajectory. The statistical significance of the sampling has been tested by means of Z-test. The analysis of the p-values (reported in Tab. S1 in the supporting information) shows that in all cases but one (i.e. $N - C_1$ distribution in the anion trajectory) the p-values are greater than 0.05, indicating that the snapshot set constitutes a good sam-

ple of the full trajectory in the most relevant proline distances.

For all the MD simulations we used the following parameters. The equations of motion have been integrated using the Beeman integrator with an integration time step of 1 fs. The coupling time (τ_T) of the thermostat was set to 0.1 ps for the Berendsen thermostat and 0.2 ps for the Bussi thermostat. An additional rescaling of the velocities has been applied at each integration step in order to enforce the translational and rotational invariance of the center of mass of the system.

Minimization and molecular dynamics were run with a development version of GAMESS-US, version 2021 (R2 Patch2),[41] coupled to a modified version of Tinker 8.10.1.[42] The image of the droplet model was generated using the UCSF ChimeraX molecular visualization tool.[43]

The K-edge spectra have been simulated using ΔMRSF as implemented in GAMESS-US[41] with the scalar relativistic effects through second-order Douglas-Kroll-Hess Hamiltonian.[44,45] The DFT exchange-correlation functional employed for all calculations is BH&HLYP.[46] The basis set employed has been aug-pc-X2 for the core of interest and aug-pc-1 for the other QM centers.[47,48] For converging the core-hole triplet reference, the standard orbital rotation restriction based on the maximum overlap method has been used. From this reference, the lowest 14 X-ray states were extracted from linear-response MRSF-TDDFT. The overall spectra were obtained as a sum of Gaussian functions for each transition, whose position was determined by the energy at the ΔMRSF level, the height corresponded to the transition dipole moment and the width was fixed at 0.5 eV.

4 Results & Discussion

The X-ray absorption spectra for the carbon, nitrogen, and oxygen K-edge have been computed for a proline model for the ionic forms (see Fig. 2) in the gas phase and in the water solution using the methodology protocol described in the previous sections. The gas phase spectra of the neutral form of proline for all K-edges are found in the supporting information, showing an excellent agreement between the theoretical and experimental spectra. The calculations have been performed with and without counterions. Here, we only show the spectra of proline with counterions, in which the right amount of Na^+ and Cl^- ions have been added so that the total net charge is zero (for the effect of the counterions in the XAS spectrum of proline, see Figures S12, S13 and S14 of the supporting information).

Proline can exist in three protonation forms, namely, cationic, anionic, and zwitterionic forms depending on the pH (see Fig. 2). The neutral form is only stable in the gas phase when there are no counterions present, and will not be considered in the current study. The focus of the following sections is on analyzing the proline-water interactions and describing the X-ray transitions. The energetic shifts due to the solvent for each of the protonation forms of proline are discussed independently for each K-edge, compared to the gas phase, and to the snapshot average deactivating the QM and MM electrostatic interactions (labeled "no electrostatics"), allowing us to determine the energetic shift due to the interaction with the solvent, and due to the fluctuations of proline. Finally, we reconstruct the pH-dependent spectra of proline.

4.1 Proline models in water

In order to understand better the proline-water interactions in the droplet models, we extracted the radial distribution function (RDF) from the production trajectories. In particular, we focused our attention on the interactions between water molecules and the amino and carboxylic groups of proline in the three different protonation states. The RDF plots obtained for the zwitterionic structure (in orange in Fig. 3) are in good agreement with the experimental RDF plots reported by McLain et al. as both the position of the main peaks and their relative intensity match well.[18]

The water-amino group RDF plots (Fig. 3,

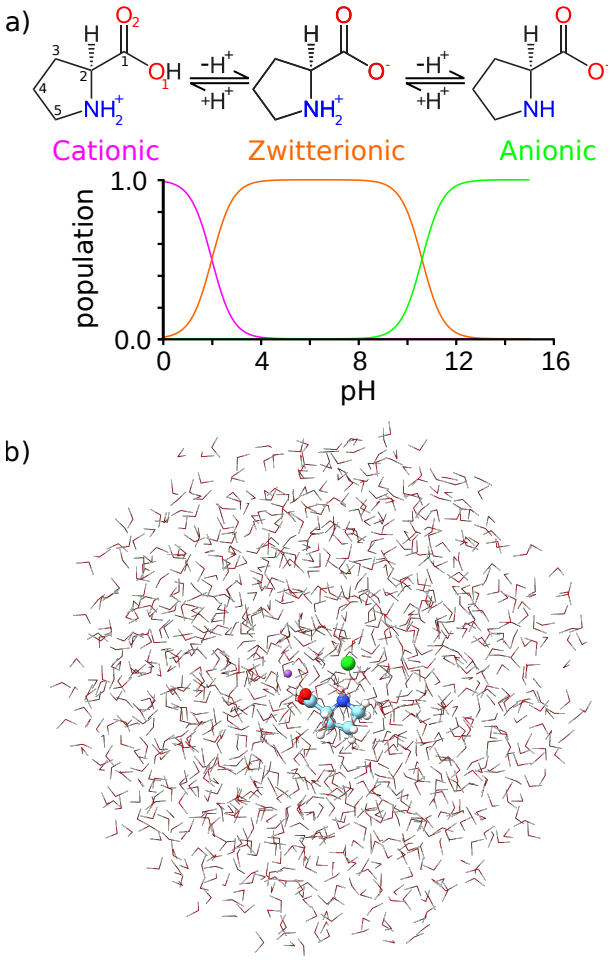


Figure 2: (a) Cationic, zwitterionic, and anionic forms of proline depending on the pH. The atom numbering used throughout the article is indicated in the cationic form. The pH populations (lower panel) have been constructed using the Henderson-Hasselbach equation, with carboxylic pKa 1.99 and amino pKa 10.6. The pKa values in water have been extracted from Ref. 49. (b) Structure of the droplet model for the zwitterionic form of proline including sodium (in purple) and chloride (in green) counterions. Similar structures have been obtained for the anionic and cationic forms, with or without suitable counterions.

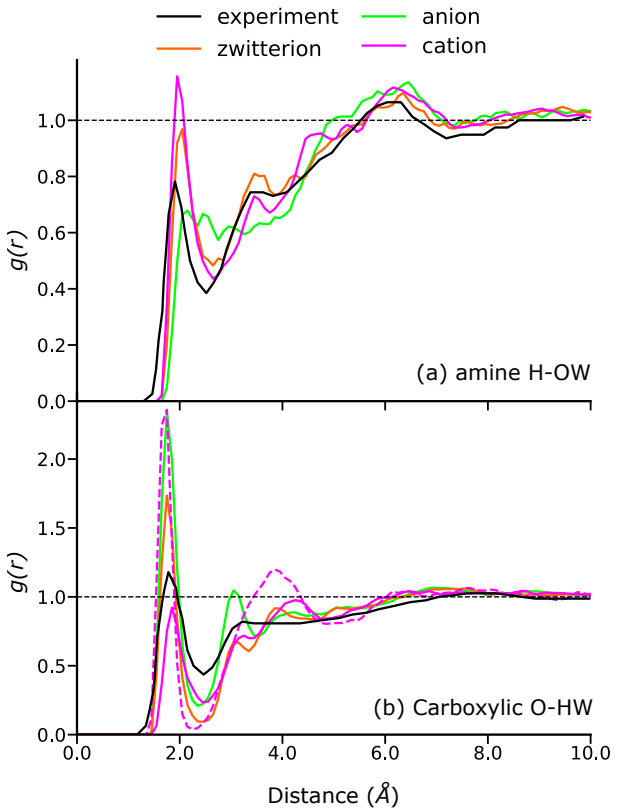


Figure 3: Normalized radial distribution functions for the proline droplet models for the anionic, zwitterionic, and cationic forms during the 150 ps production QM/MM molecular dynamics. The distance shown in (a) corresponds to amine's hydrogen (H) with water oxygen (OW), and in (b) corresponds to the carboxylic oxygen (O) with water hydrogen (HW). The black lines correspond to the experimental spectra of the zwitterionic form, registered in a 1:10 proline/water solution, reported in Ref. 18. The dashed line corresponds to the carboxylic hydrogen and OW in the cationic form.

a), instead, show that in the cationic and zwitterionic form, which present a peak of similar intensity around 2 Å, the hydrogens of the amine group act as hydrogen bond donors toward the oxygens of water. In contrast, the RDF of the anionic structure does not present such a peak suggesting that, in this protonation form, the only hydrogen of the amino group establishes preferentially an intramolecular hydrogen bond with one of the oxygens in the carboxylic group (see Fig. S7 of the supporting information). This hypothesis is further supported by the calculation of the average distance ($\overline{r}_{O_2-HN} = 3.28 \pm 0.25$ Å) between the two oxygens and the hydrogen atom.

In the case of water-carboxylic group RDF plots (Fig. 3, b), the presence of a peak around 1.75 Å for each protonation form, indicating clearly the formation of hydrogen bonds between water and the two oxygen of the carboxylic group (the latter acts as hydrogen bond acceptor.) The distances at which the peaks occur suggest also a correlation in all the protonation forms, although the lower intensity of the peaks for the cation model implies that the carboxylic group is accepting fewer hydrogen bonds than the other forms. The presence of a peak at a comparable distance in the RDF plot of the water oxygen and hydrogen of the carboxylic group in the cationic structure indicates that the latter acts also as a hydrogen bond donor.

As part of the geometrical analysis of the trajectories, we report also the behavior of the counterions along the trajectories that contain them, since it can affect the XAS spectrum, especially for the σ^* peaks and their mixture with Rydberg-type orbitals (see Fig. S2 of the supporting information.) The sodium ion in both anionic and zwitterionic models exhibits interaction with the oxygens of the carboxylic group although, for the anion, the sodium ion interacts preferentially with one of the oxygens ($\overline{r}_{Na-O_2} = 2.17 \pm 0.08$ Å and $\overline{r}_{Na-O_1} = 4.14 \pm 0.12$ Å). In the zwitterion model instead, the interaction is equally shared between the two oxygens ($\overline{r}_{Na-O_2} = 2.40 \pm 0.21$ Å and $\overline{r}_{Na-O_1} = 2.33 \pm 0.16$ Å). The chloride ion does not form a strong interaction with the amine moiety of proline. In the cation model, in fact, the distribution of the distances between the counterion and the nitrogen atom, which is non-gaussian, has a median value of $\tilde{r}_{Cl-N} = 11.51 \pm 3.61$ Å, indicating a weak electrostatic interaction with proline. Similarly, in the zwitterion case, the chloride-amine distance has a value of $\overline{r}_{Cl-N} = 6.70 \pm 0.33$ Å exhibiting thus a larger interaction with the sodium ion.

4.2 N K-edge

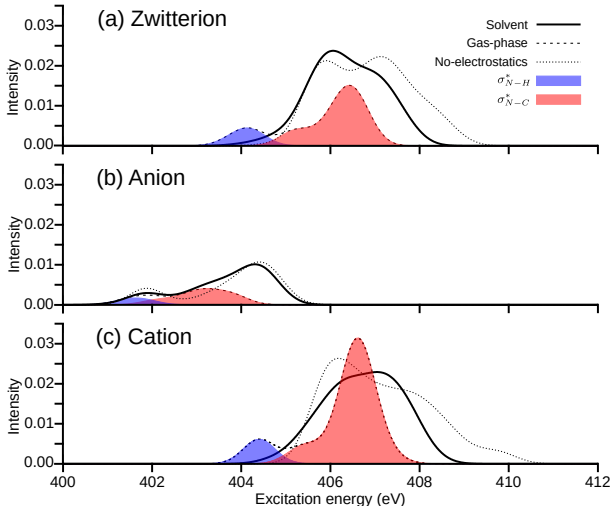


Figure 4: Nitrogen K-edge absorption of (a) zwitterionic, (b) anionic, and (c) cationic in solution (solid line), gas phase (dashed), and no electrostatics (dotted). The gas phase XAS spectrum comes along with its analysis (shaded area) and corresponds to $1s \rightarrow \sigma^*_{N-H}$ (blue) and $1s \rightarrow \sigma^*_{C-N}$ (red). The spectra for the solvent model converged up to 406.96, 404.21 and 407.51 eV for the zwitterionic, anionic, and cationic structures, respectively.

The gas phase, solvent, and no electrostatics N K-edge spectrum for the three protonated structures are shown in Figure 4. Since there is only one nitrogen core, two types of electronic transitions are observed either to σ^*_{N-H} localized along one of the N-H bonds or to σ^*_{N-C} orbitals localized over both N-C_2 and N-C_5 bond (see Fig. S9 in the supporting information). In the gas phase, the spectrum of the zwitterionic structure shows a weakly intense peak at 404.1 eV corresponding to the σ^*_{N-H}

transition and a broader more intense peak at 406.4 eV corresponding to the σ^*_{N-C}. A similar structure is observed for the cationic form, slightly blueshifted with respect to the zwitterionic form. Indeed, the σ^*_{N-H} peak is found at 404.4 eV, whereas the σ^*_{N-C} is located at 406.6 eV. Finally, the anionic spectrum is less intense and redshifted with respect to the zwitterionic or cationic forms. The first peak is located at 401.9 eV, 0.3 eV above the σ^*_{N-H}, whereas the σ^*_{N-C} is found at 403.2 eV.

The N K-edge spectra of proline in water can be largely affected by the changes in the σ^* orbitals, which depend directly on the N-C (see Fig. S4 in the supporting information.) and N-H distances (see Fig. S5 in the supporting information.) The zwitterion and cation distances exhibit similar average distances. The average N-C bond length ($\overline{r}_{\mathrm{N-C}} = 1.54 \pm 0.04$ Å) is 0.05 Å longer than in the gas phase. On the contrary, the average N-H bond distance in solution ($\overline{r}_{\mathrm{N-H}} = 1.03 \pm 0.03$ Å) is 0.07 Å shorter than in the gas phase. For the anionic form, the gas phase and solvent average distances are similar. The average N-C distance in solution ($\overline{r}_{\mathrm{N-C}} = 1.49 \pm 0.03$ Å) is 0.03 Å longer than in the gas phase, while the N-H distance in solution ($\overline{r}_{\mathrm{N-H}} = 1.01 \pm 0.02$ Å) is 0.01 Å longer.

The N K-edge spectra in solution show a strong impact both on the main peak corresponding to the $1s \rightarrow \sigma^*_{N-C}$ transition and the σ^*_{N-H}. Indeed, for the zwitterionic and cationic forms, the σ^*_{N-H} intensity is largely decreased with respect to the gas phase, which only appears as a tail of the σ^*_{N-C} transition. In the anionic form instead, the first peak appears at the same position as in the gas phase. In the zwitterionic and cationic forms, a double peak structure is observed for the σ^*_{N-C} transition in both cases. The full width at half maximum of these peaks has a broader distribution of 2.2 eV for the zwitterion and 2.4 eV for the cation. This distribution is explained by the structural fluctuation of the N-C_2 and N-C_5 distances (see Fig. S4 in the supporting information), which largely affect the energy of the σ^*_{N-C} orbitals. For the zwitterion, the peak maximum is redshifted around 0.4 eV with respect to the gas phase, whereas in the cationic form, the peak maximum is blueshifted 0.4 eV. The anionic form, on the contrary, has a peak structure and width similar to the gas phase. The σ^*_{N-C} transition is blueshifted more than 1.0 eV.

Comparing the spectrum in solution with the spectrum without electrostatic interactions, we can unravel the effect of interactions of proline with water on the σ^* transitions. For the anionic form, the spectra with and without electrostatics are identical, indicating that the effect of the solvent on neutral nitrogen is negligible. For the zwitterionic and cationic forms, the electrostatic interaction modulates slightly the intensity of the first σ^*_{N-C} peak, without affecting its excitation energy. For the second peak, on the contrary, the transition is redshifted by the solvent in both cases.

4.3 O K-edge

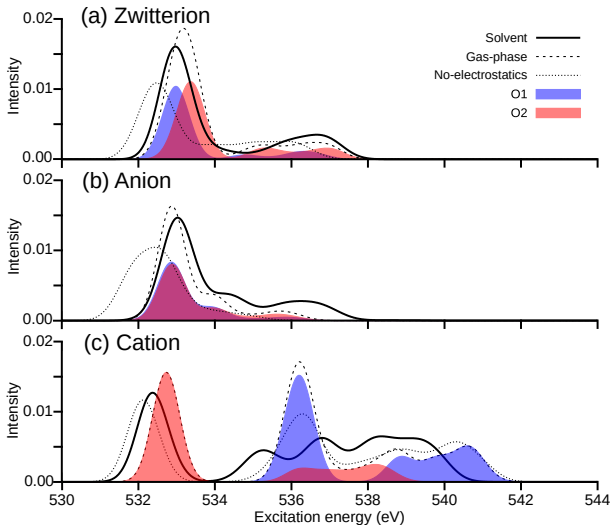


Figure 5: Oxygen K-edge absorption of (a) zwitterionic, (b) anionic, and (c) cationic in solution (solid line), gas phase (dashed), and no electrostatics (dotted). The gas phase XAS spectrum comes along with its analysis (shaded area) and corresponds to oxygen 1 (blue) and 2 (red) according to the nomenclature of Fig. 2. The spectra for the solvent model converged up to 536.51, 536.61, and 537.58 eV for the zwitterionic, anionic and cationic structures, respectively.

The gas phase, solvent, and no electrostatics O K-edge spectrum for the three protonated

structures are shown in Figure 5. The total gas phase spectra are decomposed in the K-edge of O_1 and O_2, according to the nomenclature in Figure 2. For the equivalent decomposition of the solvent spectra, see Fig. S15 of the supporting information. The three forms show a similar spectrum, with a first main peak corresponding to the 1s(O)$\rightarrow \pi^*$ transition between 532 and 534 eVs. In the case, of the anion and zwitterion, the π^* orbital is delocalized over the carboxylate group (see Fig. S10 supporting information). In the anionic forms, the O_1 and O_2 transitions are overlapping due to the symmetry between the two oxygens, leading to a maximum peak at around 532.9 eV. In the case of zwitterionic form, the 1s(O_2)$\rightarrow \pi^*$ transition is slightly blueshifted with respect to O_1 due to the intermolecular hydrogen bond that is formed with the amine group. The absorption maximum is found at 533.2 eV. In the cationic form, the transition from the core-orbital of O_1 is to the σ^*_{O-H}, shifted by more than 3.4 eV due to the protonation of the oxygen. For the lower part of the spectrum, the absorption maximum is found at 532.7 eV. Above 534 eV, the spectrum corresponds to transitions to σ^* or the lone pairs of oxygen, which are mixed with the Rydberg character. These transitions are less intense and structured than the π^* transitions. In the case of the cationic form, they span a larger range of energies due to the shift caused by the protonation of O_1.

The O K-edge spectra of proline in water can be largely affected by the changes both in the π^* and the σ^* orbitals, which depend on the fluctuation of C-O distances for all forms. The cationic form can also be affected by the O-H distance. The torsion of the carboxylic group could also affect the energetics of π^*, but this motion is not observed in our dynamics (see Fig. S8 in the supporting information). From the analysis of the C-O distance along the trajectories, it is possible to observe that there is an almost perfect delocalization of the double bond across the two oxygens in the zwitterion model ($\overline{r}_{\mathrm{C-O}} = 1.26 \pm 0.02$ Å), 0.01 Å larger to the gas phase. The anion model presents a similar behavior although, in such a case, one bond is slightly longer than the other ($\overline{r}_{\mathrm{C-O_1}} = 1.25 \pm 0.02$ Å and $\overline{r}_{\mathrm{C-O_2}} = 1.27 \pm 0.02$ Å), close to the gas phase value of 1.26 Å. Instead, as expected in the cation model, one can distinguish clearly a C=O double bond ($\overline{r}_{\mathrm{C-O}} = 1.21 \pm 0.02$ Å) and a single bond ($\overline{r}_{\mathrm{C-OH}} = 1.34 \pm 0.03$ Å). The double bond is 0.01 Å shorter in the gas phase, whereas the single bond is 0.02 Å shorter.

The O K-edge spectra for the three protonated structures in solution are shown in Fig. 5. In the zwitterionic structure, the main π^* transition is slightly redshifted (around 0.2 eV) and its intensity is slightly reduced in solution compared to the gas phase one. A similar trend is observed for the lowest π^* peak in the cationic form, in which energy is redshifted by around 0.4 eV and a similar decrease of intensity. The σ^*_{O-H} transition from O_1 core of the cationic form appearing at 536 eV is smeared out, overlapping with other σ^* transitions. These transitions fall however at the water window region,[50,51] making it impossible to record the proline spectra at such high energies. Finally, the anionic spectrum appears blueshifted with respect to the gas phase, with a maximum peak shift of 0.1 eV.

From the comparison between the gas phase, the solvent spectra, and the spectrum deactivating electrostatic interactions, we can see a cancellation effect between the structural fluctuations and the solvent electrostatic interactions. On the one hand, the structural fluctuations tend to redshift by as much as 1 eV the π^* peaks, which are broadened and less intense than in the gas phase. This is probably due to the formation of hydrogen bonds with water. On the other hand, the electrostatic interaction blueshifts the π^* due to an unfavorable electrostatic interaction, which brings back the π^* states at almost the same excitation energy as in the gas phase.

4.4 C K-edge

The gas phase, solvent, and no electrostatics C K-edge spectrum for the three protonated structures are shown in Figure 6. The three forms show a similar spectrum, due to the fact that the protonation or deprotonation does not

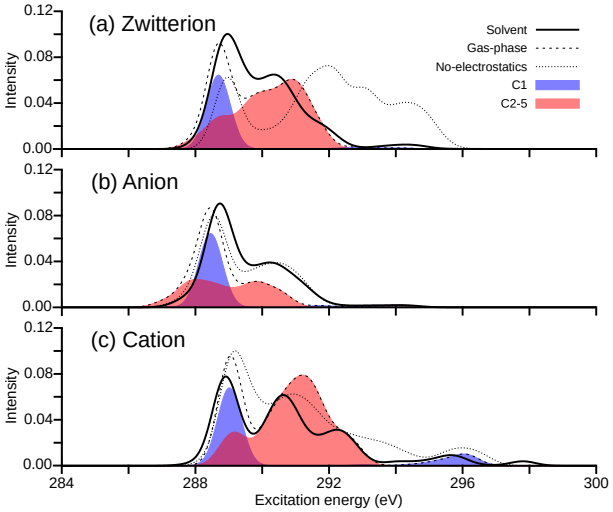


Figure 6: Carbon K-edge absorption of (a) zwitterionic, (b) anionic, and (c) cationic in solution (solid line), gas phase (dashed), and no electrostatics (dotted). The gas phase XAS spectrum comes along with its analysis (shaded area) and corresponds to carbon 1 (blue) and carbon 2 to 5 (red) according to the nomenclature of Fig. 2. The spectra for the solvent model converged up to 290.16, 290.49 and 290.10 eV for the zwitterionic, anionic, and cationic structures, respectively.

affect strongly the electronic structure of the carbons. The spectra have essentially two main contributions (see Fig. S11 for the supporting information for the molecular orbital pictures). A first main peak is found between 288 and 290 eV, and it corresponds to the $1s(C_1) \rightarrow \pi^*$ transition, in which the π^* is centered on the carboxylic group. This peak is slightly blueshifted with respect to the other two forms. The second set of transitions forms a broader band found between 288 and 294 eV, corresponding to transitions of the other carbons to the σ^* orbitals. The σ^* orbitals are mixed with Rydberg-type orbitals, which explains the broad range of energies. For all forms, a second intense peak is found around 291 eV, followed by several shoulders at higher energies.

The C K-edge spectra of proline in water can be largely affected by the changes in the σ^* orbitals, which depend directly on the C-C distances for all the carbons (see S6 in the supporting information), on C-O distances for C_1 (see Fig. S3 in the supporting information) and on N-C distances for C_2 and C_5 (see Fig. S4 in the supporting information.) Unlike the C-N and C-O bonds, the C-C bonds are less affected by the change in the protonation state of proline, with average values that are very close in the three different models ($\overline{r}_{C-C} = 1.54 \pm 0.03$ Å for the zwitterion, $\overline{r}_{C-C} = 1.55 \pm 0.04$ Å for the anion and $\overline{r}_{C-C} = 1.54 \pm 0.04$ Å for the cation). These values are between 0.01 and 0.02 Å larger than in the gas phase.

The solvent effect is small in the case of the C K-edge (for the decomposition of the solvent spectra, see Fig. S16 of the supporting information.) For the zwitterionic and anionic forms, the π^* transition is blue-shifted by 0.1 and 0.3 eV and slightly more intense with respect to the gas phase, whereas the cationic form is redshifted by 0.1 eV and slightly less intense with respect to the gas phase. The effect of the solvent on the σ^* peaks is small. Indeed, only the cationic form has a splitting into two peaks appearing at 290.6 and 292.2 eVs.

Comparing the gas phase, solvent, and no-electrostatics, we observe a similar trend as in the case of the oxygen K-edge in the cationic form. Indeed, the fluctuation of the structure

blueshifts the π^* transition, while the electrostatics redshifts it with respect to the gas phase. In the zwitterionic, this effect is more clearly observed for the σ^* transitions. The anionic form seems to be largely unaffected by the presence of electrostatic interactions.

4.5 pH-dependent spectra

Proline exhibits two pKa values, one corresponding to the deprotonation of its carboxylic acid group (pKa_1=1.95) and one corresponding to the deprotonation of its amine moiety (pKa_2=10.6).[49] Here, we construct the pH-dependent X-Ray spectra. For the sake of simplicity, we will focus the discussion on the N K-edge, for which experimental results are also available.[6] The pH-dependent O and C K-edge spectra can be found in Fig. S17 of the supporting information.

The pH-dependent XAS spectra are constructed by using the computed spectra of the solvated anion, zwitterion, and cation weighted by each population extracted from the Henderson–Hasselbach equation for the zwitterionic (p_Z), anionic (p_A), and cationic (p_C) forms (see Fig. 2). Imposing that their populations add up to one, the populations are given by

$$p_Z = \frac{1}{1 + 10^{pH-pKa_2} + 10^{pKa_1-pH}} \quad (14)$$

$$p_C = p_Z \cdot 10^{pH-pKa_1}$$

$$p_A = p_Z \cdot 10^{pH-pKa_2}. \quad (15)$$

The populations for the different pH values are plotted in Fig. 2.

Fig. 7 (top) shows the computed N K-edge spectra at pHs 6.3 and 12.2 compared to the experimental spectra.[5] Since we simulated only the lower part of the spectrum, the higher-energy region (above 408 eV) is not reproduced in our theoretical model. Converging such a spectrum would require a convergence of the continuum and probably an extension of the QM region to better describe that quasi-ionization region and the electron transfer to water, which is beyond the scope of the present model.

At pH 6.3, the spectrum corresponds essentially to the zwitterionic form, whereas at pH 12.2 the spectrum corresponds to the anionic form. At both pH values, we observe an excellent agreement between the experimental and theoretical spectra, both in terms of intensity and excitation energies. We confirm thus the disappearance of the σ^*_{N-H} peak at pH=6.3, whereas this is still observed for pH=12.2. The broadening of the main σ^*_{C-N} transition at pH=7 is slightly more compact in the theoretical spectrum compared to the experimental spectrum. The double peak structure observed for that peak might is thus not observable experimentally.

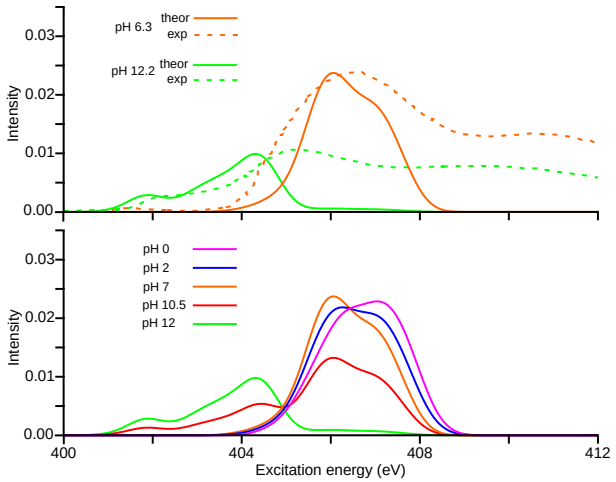


Figure 7: pH-dependent N K-edge spectra of proline. (top) comparison between experimental (dashed) and theoretical (solid) spectra at pH=6.3 (orange), and pH=12.2 (green); (bottom) theoretical spectra at pH=0 (pink), pH=2 (blue), pH=7 (orange), pH=10.5 (red), and pH=12 (green). The spectra converged up to 408.07, 405.19, and 408.55 eV for the zwitterionic, anionic, and cationic structures, respectively. Experimental data are reproduced from Ref. 5.

In Fig. 7, we reconstruct the theoretical XAS spectrum of proline along a range of different pH values (0, 2, 7, 10.5, and 12). At pH=0, the population corresponds to the cationic form. At pH=2, the population is half-half between the cationic and zwitterionic forms. At pH=7, the population is essentially the zwitterionic state. At pH=10.5, a half-half population corresponds to the zwitterionic and anionic forms, and at

pH=12 only the anionic form is present. As we have seen from the experimental spectra, it is possible to clearly distinguish the anionic and zwitterionic forms from the loss of intensity of the σ^*_{C-N} peak, the redshift of the spectrum and the appearance of the σ^*_{H-N} peak. The next question is whether the cationic form (appearing at acidic pH values) can be distinguished from the zwitterionic form through the N K-edge XAS spectrum. In principle, despite the double peak structure being experimentally not observed, the two forms could be distinguished by a change in the absorption maximum of the σ^*_{C-N} peak, which should be slightly blueshifted at acidic pH. For the oxygen K-edge, which could potentially be used for identifying the three protonation forms, most of the spectrum cannot be recorded since it falls outside the water window at which water oxygens absorb above 533 eV. For the carbon K-edge, the absorption spectrum is similar for the main transition to the π^*, and not sufficiently structured for the σ^* transitions to clearly identify the three protonation forms.

4.6 Conclusions

We have developed an efficient theoretical protocol for computing the solvent effects of X-ray spectra using an atomistic description of the solvent. The protocol is based on the combination of mixed-reference time-dependent density-functional theory (MRSF-TDDFT) with electrostatic embedding QM/MM based on electrostatic potential fitting charges.

We have applied this protocol to compute the nitrogen, oxygen, and carbon K-edges of three protonation forms of the amino acid proline. We describe the solvent effects in each of the edges by comparing the gas phase spectra with the spectra generated in water droplets with and without the electrostatic interaction with the solvent. We show that the $1s \rightarrow \pi^*$ peak, the most intense peak in the carbon and oxygen K-edges, is largely unaffected by the presence of the solvent, experiencing only a small shift due to the electrostatic interaction with water. At variance, the $1s \rightarrow \sigma^*$ transitions have an effect on both the electrostatic interaction with the solvent and the fluctuations of the structure. These fluctuations affect the mixed character of σ^* orbitals with Rydberg-type orbitals. In the case of the N K-edge, in which only σ^* transitions are observed, the largest effect of the solvent is observed. Indeed, in this core, the three forms of proline can be distinguished from the X-ray absorption spectra as shown from the reconstruction of pH-dependent spectra.

In summary, the ΔMRSF/MM protocol provides an efficient theoretical framework for describing solvent effects on X-ray spectroscopy. Currently, one limitation exists in this methodology, namely, that the SOMO→SOMO configuration is excluded from the excitation space. This implies that the converged core-hole reference must be in a state with a negligible transition dipole moment for constructing a reasonable X-ray absorption spectrum. This will be solved in the future by extending the excitation space included in the MRSF methodology.

Supporting Information Available

The supporting information contains the comparison of experimental and theoretical proline in the gas phase, statistical analysis of geometrical parameters for the QM/MM MD trajectories, the molecular orbitals representative of the different X-ray transitions, the effect of the counterions on the gas phase X-ray absorption spectra of proline, the decomposition of the solvent spectra in cores, and the pH-dependent oxygen and carbon K-edge absorption spectra.

Acknowledgement M.H.R and N.F. acknowledge the support from "Agence Nationale de la Recherche" through the project MAPPLE (ANR-22-CE29-0014-01). Centre de Calcul Intensif d'Aix-Marseille is acknowledged for granting access to its high-performance computing resources. This work was also supported by the NRF funded by the Ministry of Science and ICT (2020R1A2C2008246 and 2020R1A5A1019141 to C.H.C.).

References

(1) Schnohr, C. S.; Ridgway, M. C. In X-Ray Absorption Spectroscopy of Semiconductors; Schnohr, C. S., Ridgway, M. C., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2015; pp 1–26.

(2) Kim, J.; Kim, K. H.; Oang, K. Y.; Lee, J. H.; Hong, K.; Cho, H.; Huse, N.; Schoenlein, R. W.; Kim, T. K.; Ihee, H. Tracking reaction dynamics in solution by pump–probe X-ray absorption spectroscopy and X-ray liquidography (solution scattering). Chem. Commun. **2016**, 52, 3734–3749.

(3) Schmidt, C. Time-Resolved Soft X-Ray Absorption Spectroscopy of Molecules in the Gas and Liquid Phases; Springer International Publishing, 2021; pp 103–111.

(4) Penfold, T. J.; Curchod, B. F. E.; Tavernelli, I.; Abela, R.; Rothlisberger, U.; Chergui, M. Simulations of X-ray absorption spectra: the effect of the solvent. Phys. Chem. Chem. Phys. **2012**, 14, 9444–9450.

(5) Messer, B. M.; Cappa, C. D.; Smith, J. D.; Drisdell, W. S.; Schwartz, C. P.; Cohen, R. C.; Saykally, R. J. Local Hydration Environments of Amino Acids and Dipeptides Studied by X-ray Spectroscopy of Liquid Microjets. J. Phys. Chem. B **2005**, 109, 21640–21646.

(6) Messer, B. M.; Cappa, C. D.; Smith, J. D.; Wilson, K. R.; Gilles, M. K.; Cohen, R. C.; Saykally, R. J. pH Dependence of the Electronic Structure of Glycine. J. Phys. Chem. B **2005**, 109, 5375–5382.

(7) Zubavichus, Y.; Shaporenko, A.; Grunze, M.; Zharnikov, M. Innershell Absorption Spectroscopy of Amino Acids at All Relevant Absorption Edges. J. Phys. Chem. A **2005**, 109, 6998–7000.

(8) Messer, B. M.; Cappa, C. D.; Smith, J. D.; Wilson, K. R.; Gilles, M. K.; Cohen, R. C.; Saykally, R. J. pH Dependence of the Electronic Structure of Glycine. J. Phys. Chem. B **2005**, 109, 5375–5382.

(9) Cappa, C. D.; Smith, J. D.; Messer, B. M.; Cohen, R. C.; Saykally, R. J. Nature of the Aqueous Hydroxide Ion Probed by X-ray Absorption Spectroscopy. J. Phys. Chem. A **2007**, 111, 4776–4785.

(10) Weinhardt, L. et al. Probing hydrogen bonding orbitals: resonant inelastic soft X-ray scattering of aqueous NH3. Phys. Chem. Chem. Phys. **2015**, 17, 27145–27153.

(11) Folkestad, S. D.; Koch, H. Equation-of-Motion MLCCSD and CCSD-in-HF Oscillator Strengths and Their Application to Core Excitations. J. Chem. Theory Comput. **2020**, 16, 6869–6879.

(12) Loh, Z.-H. et al. Observation of the fastest chemical processes in the radiolysis of water. Science **2020**, 367, 179–182.

(13) Reinholdt, P.; Vidal, M. L.; Kongsted, J.; Iannuzzi, M.; Coriani, S.; Odelius, M. Nitrogen K-Edge X-ray Absorption Spectra of Ammonium and Ammonia in Water Solution: Assessing the Performance of Polarizable Embedding Coupled Cluster Methods. J. Phys. Chem. Lett. **2021**, 12, 8865–8871.

(14) Tsuru, S.; Sharma, B.; Nagasaka, M.; Hättig, C. Solvent Effects in the Ultraviolet and X-ray Absorption Spectra of Pyridazine in Aqueous Solution. J. Phys. Chem. A **2021**, 125, 7198–7206.

(15) De Santis, M.; Vallet, V.; Gomes, A. S. P. Environment Effects on X-Ray Absorption Spectra With Quantum Embedded Real-Time Time-Dependent Density Functional Theory Approaches. Front. Chem. **2022**, 10, 823246.

(16) Park, W.; Alías-Rodríguez, M.; Cho, D.; Lee, S.; Huix-Rotllant, M.; Choi, C. H. Mixed-Reference Spin-Flip Time-Dependent Density Functional Theory for

Accurate X-ray Absorption Spectroscopy. J. Chem. Theory Comput. **2022**, 18, 6240–6250.

(17) Huix-Rotllant, M.; Schwinn, K.; Pomogaev, V.; Farmani, M.; Ferré, N.; Lee, S.; Choi, C. H. Photochemistry of Thymine in Solution and DNA Revealed by an Electrostatic Embedding QM/MM Combined with Mixed-Reference Spin-Flip TDDFT. J. Chem. Theory Comput. **2023**, 19, 147–156.

(18) McLain, S. E.; Soper, A. K.; Terry, A. E.; Watts, A. Structure and Hydration of L-Proline in Aqueous Solutions. J. Phys. Chem. B **2007**, 111, 4568–4580.

(19) Plekan, O.; Feyer, V.; Richter, R.; Coreno, M.; de Simone, M.; Prince, K.; Carravetta, V. An X-ray absorption study of glycine, methionine and proline. J. Electron Spectrosc. Relat. Phenom. **2007**, 155, 47–53.

(20) Meyer, F.; Hauschild, D.; Benkert, A.; Blum, M.; Yang, W.; Reinert, F.; Heske, C.; Zharnikov, M.; Weinhardt, L. Resonant Inelastic Soft X-ray Scattering and X-ray Emission Spectroscopy of Solid Proline and Proline Solutions. J. Phys. Chem. B **2022**, 126, 10185–10193.

(21) Ferré, N.; Ángyán, J. G. Approximate electrostatic interaction operator for QM/MM calculations. Chem. Phys. Lett. **2002**, 356, 331–339.

(22) Huix-Rotllant, M.; Ferré, N. Analytic Energy, Gradient, and Hessian of Electrostatic Embedding QM/MM Based on Electrostatic Potential-Fitted Atomic Charges Scaling Linearly with the MM Subsystem Size. J. Chem. Theory Comput. **2021**, 17, 538–548.

(23) Lee, S.; Filatov, M.; Lee, S.; Choi, C. H. Eliminating spin-contamination of spin-flip time dependent density functional theory within linear response formalism by the use of zeroth-order mixed-reference (MR) reduced density matrix. J. Chem. Phys. **2018**, 149, 104101.

(24) Lee, S.; Kim, E. E.; Nakata, H.; Lee, S.; Choi, C. H. Efficient implementations of analytic energy gradient for mixed-reference spin-flip time-dependent density functional theory (MRSF-TDDFT). J. Chem. Phys. **2019**, 150, 184111.

(25) Shao, Y.; Head-Gordon, M.; Krylov, A. I. The spin-flip approach within time-dependent density functional theory: Theory and applications to diradicals. J. Chem. Phys. **2003**, 118, 4807–4818.

(26) Lee, S.; Park, W.; Nakata, H.; Filatov, M.; Choi, C. H. In Time-Dependent Density Functional Theory; Zhu, C., Ed.; Jenny Stanford Publishing, 2022; Chapter 4, pp 101–140.

(27) Casida, M. E.; Jamorski, C.; Bohr, F.; Guan, J.; Salahub, D. R. ACS Symposium Series; American Chemical Society, 1996; pp 145–163.

(28) Fiolhais, C.; Nogueira, F.; Marques, M. A. A primer in density functional theory; Springer Science & Business Media, 2003; Vol. 620.

(29) Casida, M. E.; Huix-Rotllant, M. Progress in time-dependent density-functional theory. Annu. Rev. Phys. Chem. **2012**, 63, 287–323.

(30) Lee, S.; Shostak, S.; Filatov, M.; Choi, C. H. Conical Intersections in Organic Molecules: Benchmarking Mixed-Reference Spin–Flip Time-Dependent DFT (MRSF-TD-DFT) vs Spin–Flip TD-DFT. J. Phys. Chem. A **2019**, 123, 6455.

(31) Park, W.; Shen, J.; Lee, S.; Piecuch, P.; Filatov, M.; Choi, C. H. Internal Conversion between Bright ($1^1B_u^+$) and Dark ($2^1A_g^-$) States in s-trans-Butadiene and s-trans-Hexatriene. J. Phys. Chem. Lett. **2021**, 12, 9720–9729.

(32) Horbatenko, Y.; Sadiq, S.; Lee, S.; Filatov, M.; Choi, C. H. Mixed-Reference Spin-Flip Time-Dependent Density Functional Theory (MRSF-TDDFT) as a Simple yet Accurate Method for Diradicals and Diradicaloids. J. Chem. Theory Comput. **2021**, 17, 848–859.

(33) Horbatenko, Y.; Lee, S.; Filatov, M.; Choi, C. H. How Beneficial Is the Explicit Account of Doubly-Excited Configurations in Linear Response Theory? J. Chem. Theory Comput. **2021**, 17, 975–984.

(34) Park, W.; Lee, S.; Huix-Rotllant, M.; Filatov, M.; Choi, C. H. Impact of the Dynamic Electron Correlation on the Unusually Long Excited-State Lifetime of Thymine. J. Phys. Chem. Lett. **2021**, 12, 4339–4346.

(35) Gilbert, A. T. B.; Besley, N. A.; Gill, P. M. W. Self-Consistent Field Calculations of Excited States Using the Maximum Overlap Method (MOM). J. Phys. Chem. A **2008**, 112, 13164–13171.

(36) Besley, N. A.; Asmuruf, F. A. Time-dependent density functional theory calculations of the spectroscopy of core electrons. Phys. Chem. Chem. Phys. **2010**, 12, 12024–12039.

(37) Besley, N. A. Density Functional Theory Based Methods for the Calculation of X-ray Spectroscopy. Acc. Chem. Res. **2020**, 53, 1306–1315.

(38) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. **1983**, 79, 926.

(39) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J. Comput. Chem. **2000**, 21, 1049–1074.

(40) Kratky, K.; Schreiner, W. Computational techniques for spherical boundary conditions. J. Comput. Phys. **1982**, 47, 313–320.

(41) Barca, G. M. J. et al. Recent developments in the general atomic and molecular electronic structure system. J. Chem. Phys. **2020**, 152, 154102.

(42) Ponder, J. W. TINKER: Software tools for molecular design, version 8.10.1. 2021; `http://dasher.wustl.edu/tinker`, (accessed 2023-09-18).

(43) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. **2021**, 30, 70–82.

(44) Douglas, M.; Kroll, N. M. Quantum electrodynamical corrections to the fine structure of helium. Ann. Phys. **1974**, 82, 89–155.

(45) Hess, B. A. Relativistic electronic-structure calculations employing a two-component no-pair formalism with external-field projection operators. Phys. Rev. A **1986**, 33, 3742–3748.

(46) Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. J. Chem. Phys. **1993**, 98, 1372–1377.

(47) Jensen, F. Polarization consistent basis sets: Principles. J. Chem. Phys. **2001**, 115, 9113–9125.

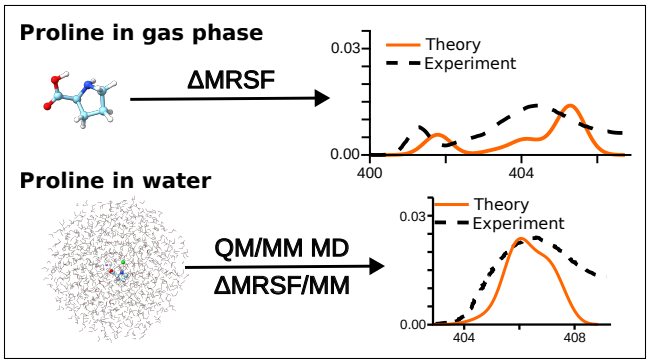
(48) Ambroise, M. A.; Jensen, F. Probing Basis Set Requirements for Calculating Core Ionization and Core Excitation Spectroscopy by the Δ Self-Consistent-Field Approach. J. Chem. Theory Comput. **2019**, 15, 325–337.

(49) Dawson, R. M. C.; Elliott, D. C.; Elliott, W. H.; Jones, K. M. Data for Biochemical Research, 3rd ed.; Clarendon Press: Oxford, England, 1989; p 580.

(50) Spielmann, C.; Burnett, N. H.; Sartania, S.; Koppitsch, R.; Schnürer, M.; Kan, C.; Lenzner, M.; Wobrauschek, P.; Krausz, F. Generation of Coherent X-rays in the Water Window Using 5-Femtosecond Laser Pulses. Science **1997**, 278, 661–664.

(51) De Stasio, G. et al. Feasibility tests of transmission X-ray photoelectron emission microscopy of wet samples. Rev. Sci. Instrum. **2000**, 71, 11–14.

TOC Graphic


Proline in gas phase
ΔMRSF
Theory
Experiment
0.03
0.00
400
404
Proline in water
QM/MM MD
ΔMRSF/MM
Theory
Experiment
0.03
0.00
404
408