Changes in the hydration structure of imidazole upon protonation: Neutron scattering and molecular simulations

Elise Duboué-Dijon, Philip E Mason, Henry Fischer, Pavel Jungwirth

Changes in the hydration structure of imidazole upon protonation: Neutron scattering and molecular simulations

Elise Duboué-Dijon[1,a)], Philip E. Mason[1,b)], Henry E. Fischer[2)], Pavel Jungwirth[1)]

[1]*Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Flemingovo nam. 2, 16610 Prague 6, Czech Republic*

[2]*Institut Laue-Langevin, 71 avenue des Martyrs, CS 20156, 38042 Grenoble cedex 9, France*

Abstract

The imidazole motif is widely encountered in biomolecules, and its biological role, for instance as a proton relay, is often linked to its ability to form hydrogen bonds with water molecules. Detailed characterization of the hydration pattern of imidazole and its changes upon protonation is thus of high interest. Here, we combine neutron scattering experiments with force field simulations to provide an unprecedented characterization of the neutral and protonated imidazole solvation at the atomistic level. We show that neutron diffraction data can be used to assess the quality of the imidazole force field in molecular simulations. Simulations using the CHARMM general force field for imidazole are in excellent agreement with the experimental neutron scattering data and we use them to provide an atomic scale interpretation of the neutron scattering patterns. Upon protonation, we clearly identify the signature of the reorganization in the hydration pattern caused by the change from one H-bond donor and one H-bond acceptor group for imidazole to two H-bond donor groups for imidazolium. We also point the limits of the experiment, which is rather insensitive to details of the H-bond geometry at the deprotonated nitrogen of imidazole and further complement the description of the hydration structure with *ab initio* molecular dynamics simulations.

[a] Author to whom correspondence should be addressed. Electronic mail: elise.duboue-dijon@uochb.cas.cz
[b] Author to whom correspondence should be addressed. Electronic mail: philip.mason@uochb.cas.cz

I. INTRODUCTION

The imidazole ring (FIG. **1**) is a central motif in chemistry and biochemistry – it forms the side chain of the amino acid histidine, the purine DNA bases consist of an imidazole ring fused with a pyrimidine ring, and imidazolium-based salts are one of the most common groups of ionic liquids. The imidazole ring can also form complexes with a number of transition metal cations; histidine residues are thus present in a large number of enzymatic active sites.[1] The singly deprotonated imidazole group behaves as a weak base with a pKa of approximately 7, which makes its protonation state very flexible at metabolic pH. This property is at the origin of the central role played by histidine residues in a number of key biological proton transfer pathways.[2-5] The imidazole motif often acts as a proton relay. This role is tightly linked with its H-bond forming ability with water molecules. It is thus important to characterize in detail the hydration pattern of imidazole and its changes upon protonation. Despite the importance of this motif, only a couple of computational studies have examined this question,[6, 7] while several previous neutron diffraction and simulation studies investigated the hydration and association properties of indole, side chain of the amino-acid tryptophan.[8, 9] The present study represents the first direct experimental characterization, supported by molecular simulations, of the solvation structures of neutral and protonated imidazole.

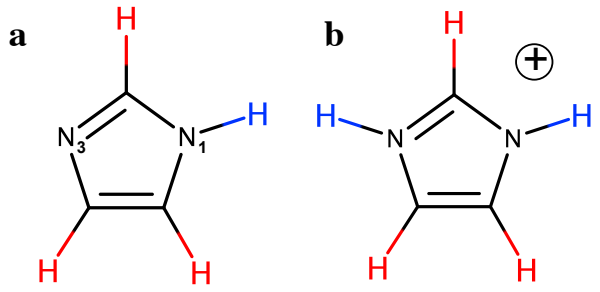

FIG. 1. a) Imidazole and b) imidazolium cation chemical structures. The exchangeable hydrogens "*Hex*" are drawn in blue, while the non-exchangeable ones "*Hsub*" are colored in red.

In particular, we use neutron diffraction with isotopic substitution[10] (NDIS) to probe the structures of imidazole and imidazolium chloride aqueous solutions. The NDIS method has been widely used to study the structure of aqueous solutions (see for instance refs [11-14]), including solutions of aromatic compounds,[9, 15] and of imidazolium containing ionic liquids.[16-20] While the total neutron scattering pattern obtained for the imidazole solution is dominated by water-water interactions, the NDIS method enables to study specifically the hydration structure of the solute by performing measurements on solutions with identical atomic composition but with different isotopic compositions. This method relies on the fact that substitution of an atom by a different isotope does not affect the structure of the solution and only changes the neutron scattering properties of the substituted nuclei. We combine neutron scattering experiments with classical Molecular Dynamics (MD) simulations to obtain a detailed picture of the imidazole hydration at the atomistic level. We show that neutron scattering experiments can be used to validate the quality of the imidazole force field used in classical simulations, but we also examine the limits in the sensitivity of the technique to certain details of the hydration structure. The use of *ab initio* MD simulations finally allows us to complement the neutron scattering data and nuance the picture provided by force field simulations.

The outline of the manuscript is as follows. In section II, we describe the employed experimental and computational methods. Results are presented and discussed in section III, which starts with a description of the experimental results (section III.A). We then compare the experiments with classical simulations, use the simulations to interpret the experimental data and obtain a molecular description of the hydration structure, paying special attention to the H-bonding properties of the studied solute (section III.B). Concluding remarks are given in section IV.

II. METHODS

A. Neutron scattering experiments

Sample preparation. A total of 8 solutions were prepared, 4 containing imidazole, and 4 containing imidazolium chloride. All the imidazole or imidazolium chloride solutions have the same molecular composition of 3 solutes for 55.55 water molecules (further denoted as "3 m") but different isotopic contents (see TABLE **I**). The imidazole and imidazolium molecules contain two different types of hydrogen atoms (FIG. **1**) – three non-exchangeable hydrogens (covalently bonded to carbon atoms, "*Hsub*"), and one or two exchangeable hydrogens "*Hex*", which rapidly exchange with water on the time scale of the experiment. To describe the sample preparation, we adopt a specific nomenclature to distinguish between these two types of hydrogen atoms and their deuterated analogues; for instance, *h-d3*-imidazole refers to an imidazole molecule where the 3 non-exchangeable hydrogen atoms are deuterated "*d3*", while the exchangeable hydrogen atom, "*h*", on imidazole is not.

A total of four imidazole solutions were prepared, with all the reagents (hydrogenated and fully deuterated imidazole) being purchased from Sigma-Aldrich (> 99 % purity and 98 atom % D for the deuterated imidazole). *d-d3*-imidazole (~ 900 mg) was dissolved directly into the correct mass of D_2O (~ 4 ml) to give the *d-d3*-imidazole in D_2O solution. The *h-h3*-imidazole in

H_2O solution was prepared in an identical way. The *h-d3*-imidazole H_2O solution was prepared by dissolving ~ 900 mg of *d-d3*-imidazole in H_2O (~ 10 ml). Water was then fully evaporated, before addition of another 10 ml of H_2O. This process was repeated a total of three times, to ensure that the exchangeable deuterium is replaced by hydrogen. Each washing was determined gravimetrically to have exchanged more than 95 % of the exchangeable hydrogens. The correct amount of H_2O was then added gravimetrically to yield the desired concentration (3 m). The *d-h3*-imidazole D_2O solution was prepared using a similar procedure.

To obtain the imidazolium chloride solutions, two solutions of HCl (~ 3 m) were first prepared – HCl in H_2O and DCl in D_2O. Both solutions were titrated against a common standard in order to accurately assess their concentrations. Aliquots of the above imidazole solutions were taken and an HCl solution was added up to the equivalence point. Water was removed under vacuum from the prepared solution, and the correct amount of water was then added gravimetrically.

TABLE I. Summary of sample compositions.

Sample name	Non-Exchangeable hydrogen atoms, "*Hsub*"	Exchangeable hydrogen atoms, "*Hex*"
h3-imidazole in H_2O	H_3	H
d3-imidazole in H_2O	D_3	H
h3-imidazole in D_2O	H_3	D
d3-imidazole in D_2O	D_3	D
h3-imidazolium chloride in H_2O	H_3	H
d3-imidazolium chloride in H_2O	D_3	H
h3-imidazolium chloride in D_2O	H_3	D
d3-imidazolium chloride in D_2O	D_3	D

Data collection and treatment. Neutron diffraction data[21] were recorded on the D4C diffractometer at the ILL in Grenoble, France, with neutrons of wavelength $\lambda = 0.7$ Å. Data were recorded for ~ 3 h for each D_2O sample and about 6 h for each H_2O sample. The raw scattering data were then corrected for multiple scattering and absorption,[22] before being normalized versus a standard vanadium rod. This provided for each sample the total scattering pattern, *S(Q)*.

The difference was then taken between each pair of isomorphic solutions, where the non-exchangeable hydrogens on imidazole were either deuterium (*d3*), or hydrogen (*h3*), to yield the first order NDIS data. For instance, in the case of imidazole in D_2O, the first order difference was taken as:

$$^{imidazole}\Delta S^{D_2O}_{HsubX}(Q) = {}^{d3\text{–}imidazole}S^{D_2O}(Q) - {}^{h3\text{–}imidazole}S^{D_2O}(Q), \quad (1)$$

where the superscript *d3*-imidazole refers to the isotopic hydrogen constitution of the non-exchangeable hydrogens on the imidazole, and the superscript D_2O refers to the solvent isotopic composition. In total 4 first order differences were obtained, 2 with imidazole ($^{imidazole}\Delta S^{D_2O}_{HsubX}(Q)$ and $^{imidazole}\Delta S^{H_2O}_{HsubX}(Q)$ in D_2O and H_2O) and 2 with imidazolium chloride ($^{imidazolium}\Delta S^{D_2O}_{HsubX}(Q)$ and $^{imidazolium}\Delta S^{H_2O}_{HsubX}(Q)$). Each of the first order differences can be expressed as a sum of pair-wise structure factors, which contains only structural data from the substituted non-exchangeable hydrogen "*Hsub*" to any other atom in solution, "*X*". All the other terms – including water-water contributions, which are dominant in the full total scattering pattern S(Q) – are cancelled out in the difference. For the imidazole solutions, the first-order differences are expressed as:

$$\begin{aligned} ^{imidazole}\Delta S^{D_2O}_{HsubX}(Q) = {} & 38.1\, S_{HsubHex}(Q) + 16.1\, S_{HsubOw}(Q) + 2.81\, S_{HsubN}(Q) \\ & + 2.99\, S_{HsubC}(Q) - 0.66\, S_{HsubHsub}(Q) - 60.6 \end{aligned} \quad (2)$$

$$ {}^{imidazole}\Delta S^{H_2O}_{HsubX}(Q) = -21.3\, S_{HsubHex}(Q) + 16.1\, S_{HsubOw}(Q) + 2.81\, S_{HsubN}(Q) + 2.99\, S_{HsubC}(Q) - 0.66\, S_{HsubHsub}(Q) - 1.24 \quad . \qquad (3) $$

The prefactors (expressed in millibarns) are calculated from the concentrations and coherent scattering length of each nucleus.[23] The corresponding expressions for the imidazolium chloride solutions are provided in the supplementary material.

Direct subtraction of the two first order differences yields the second order difference ${}^{imidazole}\Delta\Delta S(Q)$. It can be expressed as a function of a single pair-wise structure factor ${}^{imidazole}S_{HsubHex}(Q)$ and highlights the structure of exchangeable hydrogens *Hex* around the substituted hydrogens *Hsub*:

$$ {}^{imidazole}\Delta\Delta S\ (Q) = {}^{imidazole}\Delta S^{D_2O}_{HsubX}(Q) - {}^{imidazole}\Delta\Delta S^{H_2O}_{HsubX}(Q) = 59.4\ {}^{imidazole}S_{HsubHex}(Q) - 59.4 \quad . \qquad (4) $$

The second order difference ${}^{imidazole}\Delta\Delta S(Q)$ can be Fourier-transformed to provide structural information, related to a single radial distribution function:

$$ {}^{imidazole}\Delta\Delta G\ (r) = 59.4\ {}^{imidazole}g_{HsubHex}(r) - 59.4 \quad . \qquad (5) $$

A very similar expression is obtained for the imidazolium chloride solution:

$$ {}^{imidazolium}\Delta\Delta G\ (r) = 57.3\ {}^{imidazolium}g_{HsubHex}(r) - 57.3 \quad . \qquad (6) $$

B. Simulation details

Classical molecular dynamics simulations. We performed classical molecular dynamics (MD) simulations of two distinct systems – a 3 m imidazole solution and a 3 m imidazolium chloride solution.

Water molecules were described with the SPCE force field [24] and the chloride anion force field was taken from ref [25]. We also tested the influence of the choice of the water model by replicating some of the simulations with both the original TIP3P model[26] and the CHARMM TIPS3P version with Lennard-Jones parameters added on the water hydrogens, which had been

used in previous simulations of imidazolium/imidazole solutions.[27] The calculated neutron scattering structure factor was found to be fairly insensitive to the water model (see FIG. S5 in supplementary material). Imidazole and imidazolium solutes were described with the CHARMM General Force Field (CGenFF).[28] The behavior of the CGenFF imidazole force field was also compared to that of the OPLS-AA force field.[29]

Gromacs tools were used to prepare the system and all simulations were performed using the Gromacs5.1.1 software.[30] The simulation boxes for imidazole and imidazolium chloride solutions were composed of 1759 water molecules and 95 imidazole molecules (corresponding to a 3 m imidazole solution) or 95 imidazolium cations and 95 chloride anions (3 m imidazolium chloride solution). The solutes were initially randomly distributed in the simulation box. The energy of each system was first minimized for 10000 steps using a steepest descent algorithm. Each system was then equilibrated for 100 ns at 298 K in the NPT ensemble, using a Parinello-Rahman barostat[31] with a coupling time constant of 1 ps, and a velocity rescaling thermostat with a coupling time of 0.5 ps. This equilibration was followed by a 20 ns production run, using the same thermostat and barostat. All simulations used periodic boundary conditions with a Particle Mesh Ewald treatment of long-range electrostatic interactions.[32] All the bonds were constrained using the LINCS algorithm,[33] which helped maintaining the planarity of the imidazole ring.

Non-polarizable force fields are known to often provide a relatively poor description of ion-ion and ion-water interactions,[34-37] which could affect the simulations of concentrated imidazolium chloride solutions. We tested the robustness of our approach by repeating the simulation using a mean-field approach to electronic polarization, the Electronic Continuum Correction (ECC).[38] In this approach, the fast electronic polarization response (which is missing in non-polarizable force

fields) is accounted for in a mean field way by scaling the charges of all ions by the inverse of the square root of the electronic part of water dielectric response $1/\sqrt{\varepsilon_{el}} \approx 0.75$.[38]

Such a description has been already used in several simulations of ionic solutions[34, 35, 39] and has been shown to better capture the amount of ion pairing than standard non-polarizable force fields. Similar "scaled-charges" approaches are widely used to improve ion pairing in simulations of imidazolium ion liquids,[40, 41] where studies suggest that the average charge of each ion should be around +/-0.8 instead of unity. The ECC force field for the chloride anion was already available in the litterature.[39] The ECC force field for the imidazolium cation was simply obtained by scaling the charges, without adjusting the van der Waals radii.

***Ab initio* molecular dynamics simulations.** Born–Oppenheimer *Ab Initio* MD simulations at constant volume and temperature (330 K) were carried out. We employed the Quickstep module of the CP2K3.0 package[42] with hybrid Gaussian and plane waves method (GPW) implemented.[43] The simulation cell consisted of one imidazole molecule and 64 water molecules in a cubic box with an edge length of 12.8554 Å. The volume of the box was determined using a auxiliary classical 5 ns NPT simulation of the same system, with the previously described force fields. AIMD simulations used a 0.5 fs time step, and a velocity rescaling (CSVR) thermostat[44] with a time constant of 50 fs. We used a double-ζ basis set optimized for the condensed phase,[43] DZVP-MOLOPT-SR-GTH, with core electrons described with GTH pseudopotentials.[45] A cutoff of 400 Ry was used for the auxiliary plane wave basis. The system was described with the BLYP density functional.[46, 47] Two simulations were performed using two different schemes to account for dispersion interaction: the older D2 correction scheme[48] and the newer D3(BJ) Grimme

correction scheme[49] with Becke-Johnson damping.[50] The two systems were first equilibrated for 5 ps, before a 70 ps production run.

III. RESULTS AND DISCUSSION

A. Neutron scattering patterns

The raw neutron scattering data were obtained for each of the 8 samples (4 imidazole and 4 imidazolium chloride), their isotopic composition being summarized in TABLE **I** (see Methods). Comparison of the total scattering patterns of the imidazole and imidazolium chloride solutions (FIG. S1 in supplementary material) brings little information about their solvation structures since the total scattering pattern primarily reflects water structure as the most abundant species in the solution. Differences between the two solutions are thus hidden in the total signal. The NDIS method, which consists in performing neutron diffraction experiments on samples differing by their isotopic constitution, enables us to focus on the part of the system we are interested in, *i.e.* the imidazole hydration structure. We obtain the first-order differences in the Q-space, $\Delta S(Q)$, by a direct substraction of the scattering pattern of pairs of samples differing only by the isotope (H or D) of the non-exchangeable hydrogen atoms of the imidazole ring. This first-order difference cancels all the terms which do not involve the substituted hydrogen atoms, such that $\Delta S(Q)$ contains only the structural correlations between the substituted nuclei *Hsub* and any other atom in the system, *X*. The obtained first-order differences $\Delta S(Q)$ are shown in FIG. 2 (blue and red lines) for imidazole and imidazolium solutions, both in H_2O and D_2O. Since the number density – and thus the Placzeck effect – of hydrogen atoms differs between the two solutions entering in each difference, the first-order differences curves exhibit a small positive slope. At this point, differences between imidazole and imidazolium solutions are much more apparent, especially in

the 2-5 Å^{-1} region. However, *ΔS(Q)* still contains a number of intra-molecular correlations (H-C and H-N), that are of little interest here.

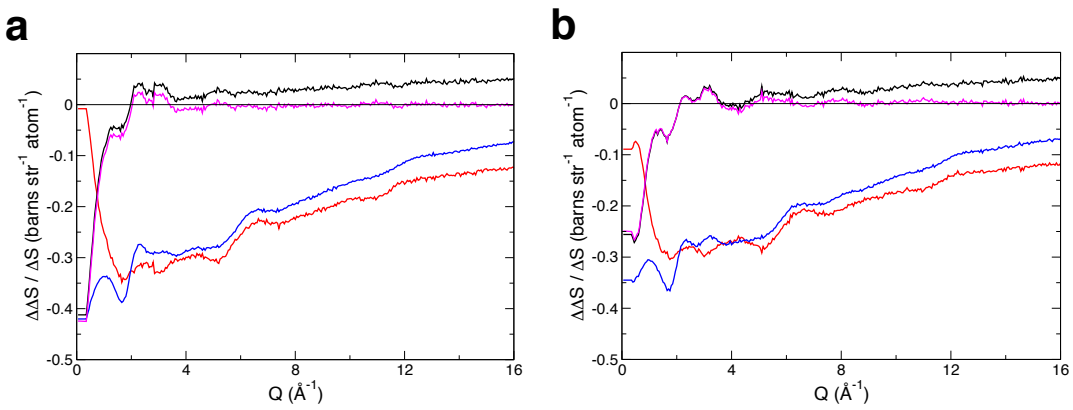

FIG. 2. Experimental first order differences Δ*S(Q)* in H_2O (red) and D_2O (blue) for a) a 3 m imidazole solution and b) a 3 m imidazolium chloride solution, together with the corresponding second order differences ΔΔ*S(Q)*, before (in black) and after (magenta) a correction to remove the residual Placzek.

Our experimental setup allows us to further focus on the imidazole/imidazolium hydration structure: by subtracting pairs of first order differences obtained in H_2O or D_2O, we obtain the second order difference *ΔΔS(Q)* for each solute. The raw *ΔΔS(Q)* are shown in FIG. 2 (black curves); they are obtained by direct subtraction of the two first order differences without any further data treatment. These second order differences still exhibit a small residual positive slope, corresponding to a residual Placzek effect (which was expected to cancel in the second order difference). This Placzek effect was corrected using a linear correction and offset (FIG. 2 magenta line) to make the signal properly terminate to zero and allow later comparison with MD simulations results; we note that this correction has almost no effect on the low-Q part of the signal.

While the H_2O and D_2O solutions used to obtain the second-order difference are not exactly isomorphic due to nuclear quantum effects,[51, 52] the structural difference between heavy and light water is extremely small and hardly detectable by neutron scattering experiments.[51] It is thus

possible to assume the pairs of solutions to be isomorphic. Within this assumption the second-order differences reduce to a single pair-wise structure factor (see Methods section) and thus provide information exclusively on the structure of water around the imidazole or imidazolium rings. All the intramolecular correlations present in the first order difference cancel, so that only remain correlations between the substitution nuclei *Hsub* and the exchangeable hydrogen atoms *Hex* from water and the imidazole NH group. A number of differences clearly appear when comparing these correlations for imidazole and imidazolium chloride solutions (FIG. 2). First, the negative peak at very low Q is more pronounced for imidazole than for imidazolium chloride, which is a sign of a more pronounced long-range structure, *i.e.* aggregation, in the imidazole solution. This finding is not unexpected, since a neutral molecule is clearly more prone to aggregation in water than its positively charged protonated moiety, and will be confirmed below by examination of molecular dynamics simulations of the same system. The shape of the peaks around 2-3 $Å^{-1}$ is also markedly different between the two solutions and so is the peak around 5 $Å^{-1}$, broader for the imidazolium chloride solution. However, given the complexity of the system, further interpretation of these features is not possible without the help of molecular simulations.

B. Molecular interpretation using MD simulations

1. Second-order differences

Classical molecular dynamics simulations of the 3 m imidazole and imidazolium chloride solutions were performed and the radial distribution function between substituted and exchangeable hydrogen nuclei, $g_{HsubHex}(r)$, was calculated for each system. The second order difference in direct space, $\Delta\Delta G(r)$, is directly related to this computed radial distribution function

(see Methods section). A numerical back Fourier transform of $\Delta\Delta G(r)$ provides the Q-space second-order difference $\Delta\Delta S(Q)$, which can be directly compared to the experimental data.

FIG. 3a-b compares the experimental second order difference $^{imidazole}\Delta\Delta S(Q)$ for the imidazole solution with that obtained from simulations performed with two different force fields for imidazole, CGenFF and OPLS. The difference curves between experimental and MD data are provided in the supplementary material (FIG. S2). Clearly, the CGenFF force field compares much better with the neutron data, especially at very low Q: $\Delta\Delta S(Q)$ obtained with the OPLS force field exhibits a much more negative peak at low Q, which can be attributed to a pronounced aggregation of imidazole molecules. Indeed, FIG. **4**a compares the probability for an imidazole molecule to belong to a cluster of a given size in simulations performed with the CGenFF and OPLS force fields and clearly shows that the OPLS imidazole force field leads to an increased formation of large imidazole clusters (containing over 40 imidazole molecules) compared to the CGenFF description. (Two imidazole molecules are considered part of the same cluster if two of their heavy atoms are found at a distance lower than 4.5 Å). This difference in the aggregation behavior of the two imidazole force fields is further illustrated in FIG. **4**b-c, with snapshots of the CGenFF and OPLS imidazole simulations, imidazole molecules belonging to the same cluster being assigned the same color. This increased aggregation of imidazole molecules with the OPLS force field may be due to the slightly smaller amplitude of the imidazole ring charges in the OPLS description. Since the CGenFF imidazole simulation very well describes the shape of the neutron scattering signal at low Q, it must provide a more realistic description of imidazole aggregation in solution; further analysis of the interaction mode between imidazole molecules in our simulations is provided in the supplementary material (FIG. S3). The CGenFF force field also better captures than OPLS the shape of the first two peaks in

the 2-4 Å region. This comparison thus shows that the experiment is sensitive enough to assess the performance of different imidazole force fields and that the simulations performed with the CGenFF imidazole force field compare extremely well with the neutron data, thus allowing us to gain molecular level insight into the structure of the imidazole solutions.

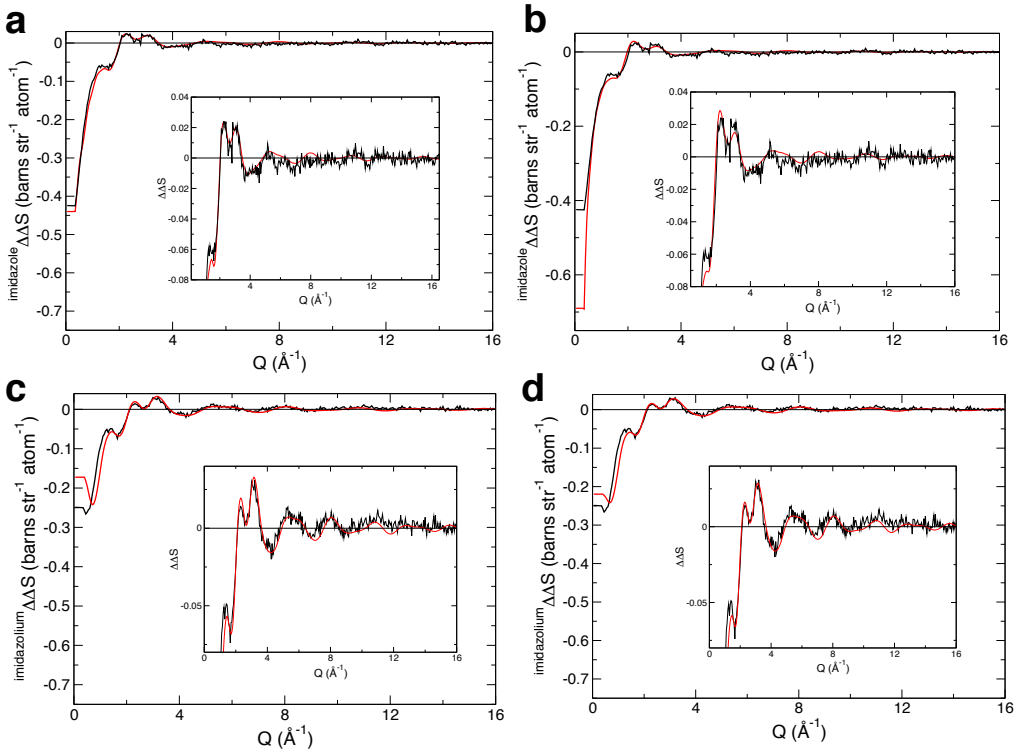

FIG. 3. The second order difference function ΔΔS(Q) as obtained from experiment (black) and from MD simulations (red, the data is cut at low Q at the experimental resolution). The upper plots show results for a 3 m imidazole solution, using in simulations a) the CGenFF force field and b) the OPLS force field for imidazole. The lower plots show results for a 3 m imidazolium chloride solution, using in simulations c) the standard CGenFF full charges and d) the ECC scaled charges force field for imidazolium. Insets highlight the details of the structures at larger Q values.

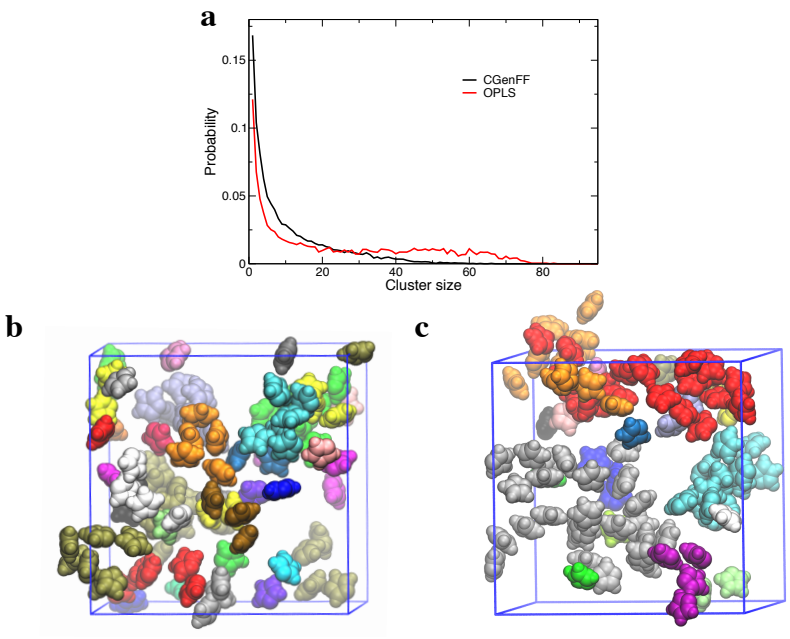

FIG. 4. a) Probability for an imidazole molecule to be part of a cluster of a given size in a simulation of a 3 m imidazole solution, imidazole being described with either the CGenFF or the OPLS force field. b) Representative snapshot of the CGenFF imidazole simulation, the same color being assigned to all imidazole molecules belonging to the same cluster. c) Idem for the OPLS imidazole simulation.

A simulation of the imidazolium chloride solution was performed using the standard CGenFF force field for the imidazolium molecule, and the computed second order difference was compared to the experimental pattern (FIG. 3c). Our simulations compare very well with experiments and correctly capture the main differences between the behavior of imidazole and imidazolium solutions. Namely, the negative peak at low Q is less pronounced for the imidazolium chloride solution, which can be attributed to reduced aggregation in the imidazolium chloride solution (see supplementary material FIG. S4). The difference in the relative height of the first two peaks between the two solutions is also correctly reproduced.

The simulation of the imidazolium chloride solution was repeated with scaled ionic charges. This ECC approach (see Methods section) is a way to include in a mean-field approach the electronic polarization,[38] which has been repeatedly shown[34, 35] to improve the description of ion

pairing in water. In the present case, the use of the ECC description has only a very small effect on the structure of the solution (FIG. 3d), the computed neutron signal with the full and scaled charges force field being almost identical. The ECC, scaled charges, simulation may compare a little better with experiment at low Q than the standard full charge description and it seems to capture slightly better the height of the first peak, but the differences are hardly beyond noise level. In this case, neutron scattering does not allow us to assess which of the two force fields performs better, as the two employed force fields — full and scaled charges — provide a very similar picture of the solution.

2. Hydration structure from MD simulations

Having shown that our force field MD simulations are in a very good agreement with the experimental neutron data, we now use these simulations to provide a molecular interpretation of the neutron patterns in terms of the imidazole / imidazolium hydration. Indeed, MD simulations directly provide us the radial distribution function between substituted and exchangeable hydrogen nuclei $g_{HsubHex}(r)$, which are much easier to interpret than Q-space data.

FIG. 5 shows the calculated second order differences for imidazole and imidazolium chloride solutions and identifies the molecular origin of each characteristic feature. The two sharp peaks at 2.5 Å and 4.2 Å in the r-space signal correspond to intramolecular correlations with the exchangeable hydrogen of the NH bond. These peaks are more intense for the imidazolium chloride solution, because the imidazolium molecule possesses two NH bonds with exchangeable H nuclei. The broader peak around 3 Å in the imidazole radial distribution function is assigned to water H donating a hydrogen bond to N3. As expected, this feature disappears in the imidazolium chloride solution, where both N atoms are protonated. Finally, the broad and

shallow feature around 4.5-5.5 Å as well as the more pronounced peak at 6.5 Å correspond to correlations involving water molecules that receive a H-bond from the imidazole NH group. As expected, these features are twice more intense for imidazolium.

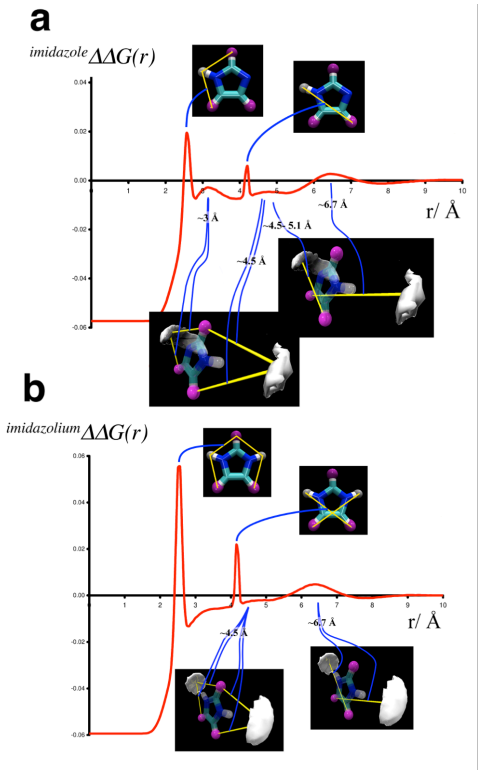

FIG. 5. Second order differences in the r-space $\Delta\Delta G(r)$ as obtained from MD simulations of a) an imidazole solution and b) an imidazolium chloride solution. The origin of each characteristic feature is identified in the corresponding inset, either a molecular correlation (upper insets) or a hydration pattern, as reflected with the density map for H_{ex} around the solute (lower insets).

In principle, one can Fourier-transformed the experimental Q-space signal to convert it into an experimental r-space representation, but numerical artifacts make the interpretation of the experimental r-space data difficult. While instruments like D4C have excellent counting statistics, inevitable counting errors lead to ringing in the Fourier transform of the Q space data. This is especially true here because a total of four independent experiments are needed to obtain a single second-order difference $\Delta\Delta S(Q)$, which results in a higher statistical noise on the final difference function. Some of these problems can be alleviated by splining the Q-space data and

the application of windows functions, but this lowers the resolution of the r-space data (sharp peaks become broader) and can remove signal as well as the statistical noise. A more detailed discussion on the effect of data post-treatment on the Fourier transform of experimental data is provided in the supplementary material (see Fig. S6). FIG. 6 shows the comparison between the experimentally derived and simulated double differences in the r-space. The same treatment was applied to experimental and simulation Q-space data before transformation to r-space (see supplementary material Fig. S7). Compared to the simulated full resolution r-space second order difference (FIG. 5), the processed lower-resolution (reduced Q range) simulation data exhibits much broader peaks and significant ringing artifacts (FIG. 6), which could not be avoided. These ringing artifacts make the comparison in r-space between simulation and experiments difficult, but the main peaks at 2.5 Å, 4.2 Å and 6.5 Å clearly appear at the same distance in simulation and experiments. Furthermore, the change in the second order difference $\Delta\Delta G(r)$ upon imidazole protonation is consistent between simulation and experiment: the intensity of the two sharp peaks (intramolecular correlation with NH) increases, and so does the peak at 6.5 Å (water receiving a H-bond from a solute NH group). We note that the intramolecular peaks at 2.5 Å and 4.2 Å are much sharper in the simulations than in the experimental neutron signal. This behavior is fully expected and due to the fact that in our simulations the imidazole molecules have fixed bond distances and that simulations do not include nuclear quantum effects and the associated partial delocalization of protons, thus resulting in sharper peaks in the neutron signal.[53]

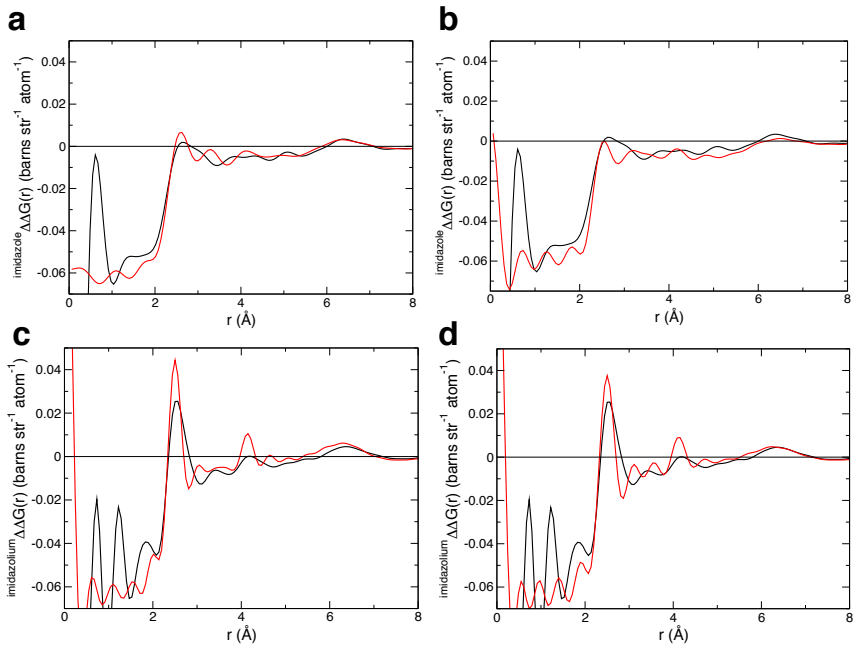

FIG. 6. *ΔΔG(r)* as obtained from neutron experiments (black) and MD simulations (red) for a-b) a 3 m imidazole solution and c-d) a 3 m imidazolium chloride solution. The imidazole is described in simulations with either a) the CGenFF or b) the OPLS force field. The imidazolium cation is described in simulations with c) the standard CGenFF full charges force field or d) with a scaled charges ECC description. The same resolution function (window function and spline) was applied to experimental and MD data.

The hydration structure is further characterized using the density maps of water oxygen and hydrogen atoms around imidazole and imidazolium. The density maps calculated from the MD simulations are shown in FIG. **7**. They very clearly characterize the general hydration patterns around these two species, with the NH groups of either the imidazole or imidazolium donating a linear hydrogen bond (H-bond) to water. This H-bond becomes slightly weaker when using a scaled charges description of the imidazolium cation. The lone pair on the imidazole nitrogen receives a H-bond from hydrating water molecules. While the H-bonds donated by NH groups both from neutral and protonated imidazole exhibit an in plane linear geometry, the H-bonds received from water at the N3 lone pair are much less well localized. These observations are consistent with previous gas phase quantum mechanical calculations on imidazole-water dimers.[7]

Interestingly, the density map shows that H-bonds at N3 are directed outside the imidazole plane, while the expected most favorable geometry would be in plane.[6, 7] The reduced density of water oxygen atoms around the imidazole solute when described by the OPLS force field (FIG. **7**b) reflects the increased imidazole aggregation exhibited by this force field, as already discussed above.

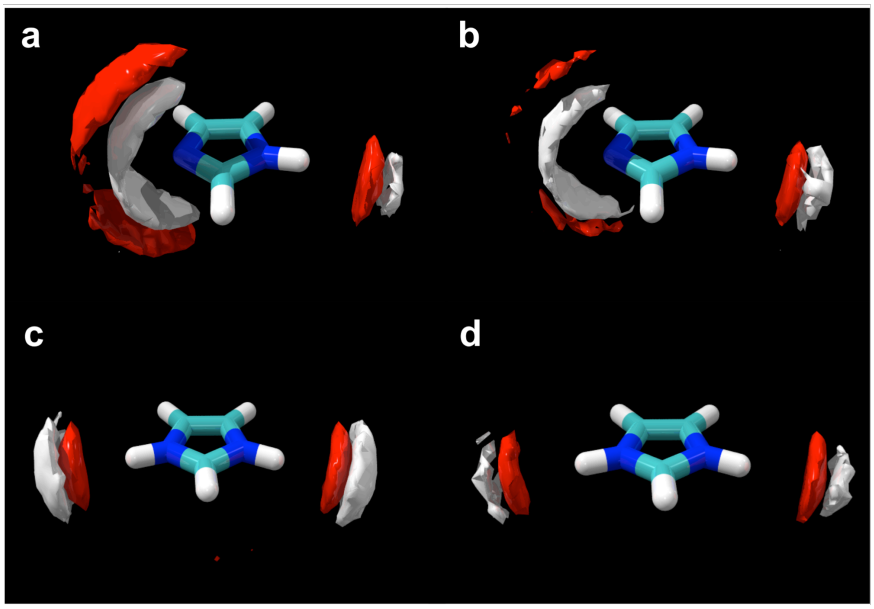

FIG. 7. Density maps of water oxygen and hydrogen atoms around the solute for a 3 m a and b) imidazole solution and c and d) imidazolium chloride solution. The imidazole molecule is described with a) the CGenFF force field or b) the OPLS force field. The imidazolium chloride is described with c) the standard full charges CGenFF force field or d) a scaled charges (ECC) description. The density contour levels are drawn at 0.13 atoms $Å^{-3}$ and 0.16 atoms $Å^{-3}$ for oxygen and hydrogen on water respectively.

3. H-bonds characterization

We now focus more closely on the H-bonds formed between the solute and water molecules. As shown by the density maps (FIG. **7**), the imidazolium molecule possesses two H-bond donor NH groups, while the imidazole molecule possesses one H-bond donor group (N1-H) and one H-bond acceptor (N3). The number of H-bonds donated by the imidazolium cation is sensitive to

the force field used. With the standard full charges description an imidazolium molecule donates on average to water 1.6 H-bonds and this number decreases to 1.3 H-bond when using the ECC description (TABLE II). The imidazole molecule, having only one NH donor group, donates on average only 0.7 H-bonds to water. The imidazole H-bond acceptor nitrogen receives on average slightly less than 2 H-bonds, on average 1.9, in our classical simulations, and the density maps clearly shows that these 2 H-bonds are on average out of the imidazole plane (FIG. **7**).

TABLE II. Average number of hydrogen bonds accepted and donated per solute molecule. The following criteria are used for H-bond definition: $d_{ad} < 3.5$ Å and $\widehat{HDA} < 30$ °, where d_{ad} is the donor-acceptor distance and $\widehat{HDA}$ is the angle between the hydrogen, donor and acceptor atoms involved in the H-bond.

Solute		Number of H-bonds donated to water (per solute)	Number of H-bonds accepted from water (per solute)
Imidazole (CGenFF)		0.7	1.9
Imidazolium	Full charges (CGenFF)	1.6	—
	ECC	1.3	—

This picture of the H-bonded structure for imidazole is quite different from what one would expect based on the N3 *sp2* hybridization, with the nitrogen lone pair receiving a single H-bond in the imidazole plane and on the $\widehat{CNC}$ bisector line. Description of this kind of properties for aromatic compounds is a well-known weak point of classical force field,[54, 55] since no potential term imposes a *sp2* hybridization at the N3 nitrogen. The unexpectedly elevated number of about 2 H-bonds received at N3 and their out-of-plane geometry may thus be an artifact of the classical potential.

Unfortunately, due to the noise in the double-difference neutron data, neutron scattering results cannot be used to further investigate this issue. Moreover the neutron signal is rather insensitive to the local H-bond geometry at N3; the $g_{HsubHex}$ radial distribution function hardly changes when the water H-bonded to N3 is constrained to stay in the imidazole plane (see supplementary material FIG. S8). The excellent agreement between our simulations and the neutron signal thus does not necessarily mean that simulations exhibit the proper H-bond geometry.

To resolve this issue, we performed *ab initio* molecular dynamics (AIMD) simulations of an imidazole molecule in solution (see sec. Methods for details). AIMD simulations should properly capture the aromatic structure of our solute and the local *sp2* hybridization at N3. However, these AIMD simulations suffer from several potential drawbacks, as they are limited to small system sizes and short simulation times. The small system size of a single imidazole and 64 water molecules makes direct comparison with the neutron signal (obtained on a concentrated 3 m imidazole solution) impossible. Moreover, the short 70 ps length of the simulation means that we only partially sample of the configuration space. Calculation of radial distribution functions thus suffers from significant error bars. Finally, DFT GGA functionals suffer from known limitations, especially in the description of weak long-range interactions.[56, 57] Empirical dispersion corrections allow to correct to a large extent these defects. Here we ran simulations with two different flavors of empirical dispersion correction, D2 and D3(BJ) for comparison (see Methods section).

First, we compute the average number of H-bonds received at N3, using the same H-bond definition as previously. The computed H-bond numbers at N3 of 1.6 for D2 or 1.7 for D3(BJ) are very consistent between the two AIMD simulations and are only a little smaller than the value of 1.9 from force field simulations. N3 thus receives on average decidedly more than one

H-bond. The force field description reasonably captures this, even though it slightly overestimates the H-bond number. The intuitive picture of a single H-bond at N3 is thus not supported neither by force field nor AIMD simulations.

We further characterize the H-bond geometry at N3 by plotting the distributions of N-O_w and N-H_w distances for water molecules H-bonded to N3 (FIG. **8**a). This clearly shows that the NO and NH distances are shorter by about 0.1 Å in AIMD simulations compared to the force field ones. Such a shorter N3-H_w distance is expected to lead to a small change in the $\Delta\Delta S(Q)$ signal and the corresponding $\Delta\Delta G(r)$ real space signal (see supplementary material FIG. S9), which unfortunately falls within the experimental noise.

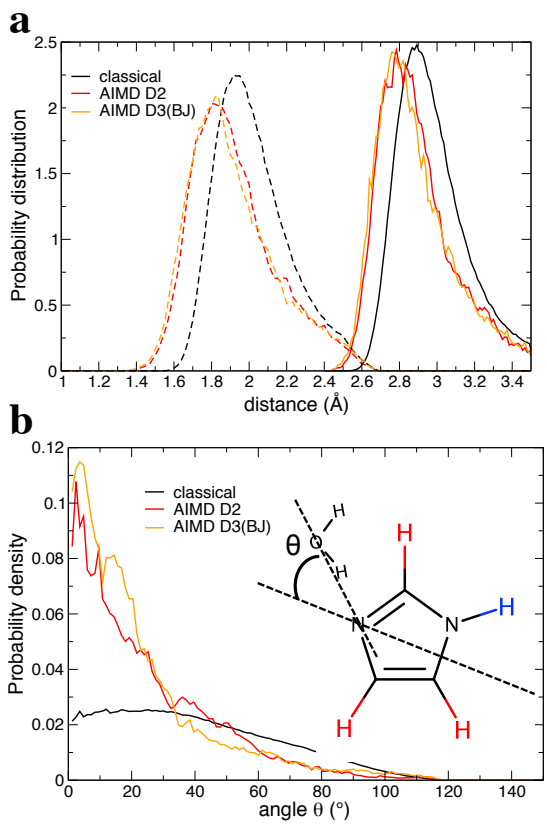

FIG. 8. a) Distance distribution between the H-bond acceptor imidazole N and the H-bond donor water oxygen (solid lines) or the water hydrogen (dashes), in the classical and in the two AIMD simulations. b) θ angle probability density distribution for the classical and the two AIMD simulations. θ is the angle between the N3-O_w vector and the $\widehat{CNC}$ bisector line, where O_w is the oxygen of a water molecule H-bonded to the N3 imidazole atom.

Visual examination of the simulations enabled us to identify one marked difference between AIMD and force field descriptions of the N3 H-bonded geometry. Namely, in the force field simulations, the most common pattern is N3 receiving two H-bonds from two water molecules, both being out of the imidazole plane. In contrast, the most common pattern from the AIMD simulation is one water donating a H-bond in the imidazole plane following the *sp2* hybridization geometry at N3, while another water molecule occasionally donates a second H-bond in an out-of-plane geometry. We can however identify also a fraction of the trajectories that exhibit a similar H-bond pattern as in the force field simulations. This difference in H-bonding patterns between AIMD and force field trajectories is quantitatively depicted in FIG. **8**b, which shows the probability density distribution of the angle θ between the N-O_w vector and the $\widehat{CNC}$ bisector line. The probability of low θ angles is markedly higher in AIMD than in force field simulations.

The AIMD simulations thus bring a useful additional insight to imidazole hydration and nuance the picture emerging from classical simulations. While they confirm that more than one H-bond is donated on average to N3, they also suggest that the force field description leads to a slightly too long N3-water H-bonds, and that these H-bonds are on average too far from the planar trigonal arrangement at N3.

IV. CONCLUSION

This work presents the first combined experimental and simulation characterization of the imidazole hydration structure and its change upon imidazole protonation. The comparison of the experimental neutron signal with results from molecular dynamics simulations allows us to gain an atomistic insight into the hydration structures and to assess the quality of different force fields

for the imidazole molecule. Simulations performed with the CHARMM general force field for both neutral and protonated imidazole are in excellent agreement with the experimental data and thus allow us to interpret the complex neutron signal and provide a molecular level description of the hydration structure.

Upon imidazole protonation the changes in the neutron signal can be assigned to a decreased aggregation of solute molecules and to the reorganization in the hydration pattern caused by the change from one H-bond donor and one H-bond acceptor group for imidazole to two H-bond donor groups for imidazolium. This shows that the choice of the protonation state of the imidazole ring does affect the surrounding water structure. These results showing the significant changes in the hydration structure of imidazole upon protonation are particularly relevant to structural and functional studies of proteins. This problem is exacerbated by the fact that the protonation state of histidine residues is rarely clear from X-ray studies and is usually assigned based on simple examination of the orientation of the imidazole ring toward potential H-bond partners, or based on simple pKa calculations.

In line with previous theoretical studies, the H-bonds donated by the solute NH group(s) are found well localized in the imidazole plane, while the H-bond received at the imidazole N3 atom is less localized. MD simulations show that the N3 atom receives on average more than a single H-bond and this picture of almost two H-bonds is confirmed by *ab initio* MD simulations. Detailed analysis of the neutron experimental data shows that the signal is rather insensitive to the in-plane vs. out-of-plane H-bond geometry at N3. *Ab initio* simulations thus bring us useful complementary data and also show that the classical description slightly underestimates the proportion of the in plane geometry, due to the lack of explicit term to properly enforce a local *sp2* hybridization at N3.

To conclude, we have combined the techniques of neutron diffraction with isotopic substitution with force field and *ab initio* molecular simulations to obtain the first structural determination of neutral and protonated imidazole hydration, yielding important information relevant to imidazole's biological role *e.g.* as a proton relay. This has allowed us to provide an accurate picture of imidazole hydration, to assess the quality of empirical force fields for the biologically relevant imidazole moiety and also to examine the limits in the experimental sensitivity and resolution.

SUPPLEMENTARY MATERIAL

See supplementary material for the raw neutron scattering patterns, difference plots between the experimental and computed neutron signal in Q-space, details on the interaction mode between imidazole molecules and clustering of imidazolium in our simulation, the study of the influence of the water model on the computed second order difference, details on the Q-space processing of the experimental data before Fourier transform and examination of the sensitivity of the neutron scattering signal with respect to the H-bond geometry at N3.

ACKNOWLEDGMENTS

We thank the Institute Laue Langevin for providing the neutron scattering facility. We particularly thank the staff of D4C and of the ILL chemistry laboratory for their help with the neutron scattering experiments. We are grateful to the Czech Science Foundation (grant no. P208/12/G016) for support. EDD acknowledges support from the EMBO and Marie Curie Actions (fellowship ALTF 952-2015).

REFERENCES

1. R. J. Sundberg and R. B. Martin, Chem. Rev. **74**, 471-517 (1974).

2. F. Mamedov, R. T. Sayre and S. Styring, Biochemistry **37**, 14245-14256 (1998).

3. P. Adelroth, M. L. Paddock, A. Tehrani, J. T. Beatty, G. Feher and M. Y. Okamura, Biochemistry **40**, 14538-14546 (2001).

4. Z. Fisher, J. A. H. Prada, C. Tu, D. Duda, C. Yoshioka, H. Q. An, L. Govindasamy, D. N. Silverman and R. McKenna, Biochemistry **44**, 1097-1105 (2005).

5. J. Hu, R. Fu, K. Nishimura, L. Zhang, H. X. Zhou, D. D. Busath, V. Vijayvergiya and T. A. Cross, Proc. Natl. Acad. Sci. USA **103**, 6865-6870 (2006).

6. G. Alagona, C. Ghio, P. Nagy, K. Simon and G. Naray-Szabo, J. Comput. Chem. **11**, 1038-1046 (1990).

7. P. I. Nagy, G. J. Durant and D. A. Smith, J. Am. Chem. Soc. **115**, 2912-2922 (1993).

8. A. Henao, A. J. Johnston, E. Guardia, S. E. McLain and L. C. Pardo, Phys. Chem. Chem. Phys. **18**, 23006-23016 (2016).

9. A. J. Johnston, Y. R. Zhang, S. Busch, L. C. Pardo, S. Imberti and S. E. McLain, J. Phys. Chem. B **119**, 5979-5987 (2015).

10. J. L. Finney and A. K. Soper, Chem. Soc. Rev. **23**, 1 (1994).

11. A. K. Soper, G. W. Neilson, J. E. Enderby and R. A. Howe, J. Phys. C: Solid State Phys. **10**, 1793-1801 (1977).

12. A. Botti, F. Bruni, S. Imberti, M. A. Ricci and A. K. Soper, J. Chem. Phys. **120**, 10154-10162 (2004).

13. P. E. Mason, G. W. Neilson, C. E. Dempsey, A. C. Barnes and J. M. Cruickshank, Proc. Natl. Acad. Sci. U. S. A. **100**, 4557-4561 (2003).

14. P. E. Mason, G. W. Neilson, J. E. Enderby, M. L. Saboungi, C. E. Dempsey, A. D. MacKerell, Jr. and J. W. Brady, J. Am. Chem. Soc. **126**, 11462-11470 (2004).

15. L. Tavagnacco, J. W. Brady, F. Bruni, S. Callear, M. A. Ricci, M. L. Saboungi and A. Cesaro, J. Phys. Chem. B **119**, 13294-13301 (2015).

16. R. Hayes, S. Imberti, G. G. Warr and R. Atkin, Phys. Chem. Chem. Phys. **13**, 3237-3247 (2011).

17. R. Hayes, S. Imberti, G. G. Warr and R. Atkin, Angew. Chem. Int. Ed. Engl. **51**, 7468-7471 (2012).

18. D. T. Bowron, C. D'Agostino, L. F. Gladden, C. Hardacre, J. D. Holbrey, M. C. Lagunas, J. McGregor, M. D. Mantle, C. L. Mullan and T. G. Youngs, J. Phys. Chem. B **114**, 7760-7768 (2010).

19. C. Hardacre, J. D. Holbrey, S. E. J. McMath, D. T. Bowron and A. K. Soper, J. Chem. Phys. **118**, 273-278 (2003).

20. C. Hardacre, J. D. Holbrey, C. L. Mullan, T. G. Youngs and D. T. Bowron, J. Chem. Phys. **133**, 074510 (2010).

21. P. Jungwirth, H. E. Fischer, J. Hladilkova and P. E. Mason (2012) Ion-Ion pairing in aqueous solutions of imidazolium chloride. Institut Laue-Langevin (ILL) doi:10.5291/ILL-DATA.8-03-779 <http://dx.doi.org/10.5291/ILL-DATA.8-03-779> for which the data are available at: G. Neilson, J.W. Brady, H.E. Fischer, J. Hladilkova, P. Jungwirth, P.E. Mason, M-L. Sabougi, L. Tavagnacco (2012) Institut Laue-Langevin (ILL) doi:10.5291/ILL-DATA.6-02-498 <http://dx.doi.org/10.5291/ILL-DATA.6-02-498>.

22. A. C. Barnes, S. B. Lague, P. S. Salmon and H. E. Fischer, J. Phys.: Condens. Matt. **9**, 6159-6173 (1997).

23. G. W. Neilson, P. E. Mason, S. Ramos and D. Sullivan, Phil. Trans. R. Soc. Lond. A **359**, 1575-1591 (2001).

24. H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, J Phys Chem-Us **91**, 6269-6271 (1987).

25. L. X. Dang, G. K. Schenter, V.-A. Glezakou and J. L. Fulton, J. Phys. Chem. B **110**, 23644-23654 (2006).

26. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, J. Chem. Phys. **79**, 926-935 (1983).

27. J. Heyda, P. E. Mason and P. Jungwirth, J. Phys. Chem. B **114**, 8744-8749 (2010).

28. K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. Mackerell, Jr., J. Comput. Chem. **31**, 671-690 (2010).

29. W. L. Jorgensen and N. A. McDonald, Theochem.-J. Mol. Struc. **424**, 145-155 (1998).

30. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, J. Comput. Chem. **26**, 1701-1718 (2005).

31. M. Parrinello and A. Rahman, J. Appl. Phys. **52**, 7182-7190 (1981).

32. T. Darden, D. York and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).

33. B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, J. Comput. Chem. **18**, 1463-1472 (1997).

34. E. Pluharova, P. E. Mason and P. Jungwirth, J. Phys. Chem. A **117**, 11766-11773 (2013).

35. M. Kohagen, P. E. Mason and P. Jungwirth, J. Phys. Chem. B **118**, 7902-7909 (2014).

36. M. Vazdar, P. Jungwirth and P. E. Mason, J. Phys. Chem. B **117**, 1844-1848 (2013).

37. E. Wernersson and P. Jungwirth, J. Chem. Theory Comput. **6**, 3233-3240 (2010).

38. I. Leontyev and A. Stuchebrukhov, Phys. Chem. Chem. Phys. **13**, 2613-2626 (2011).

39. E. Pluhařová, H. E. Fischer, P. E. Mason and P. Jungwirth, Mol. Phys. **112**, 1230-1240 (2014).

40. B. L. Bhargava and S. Balasubramanian, J. Chem. Phys. **127**, 114510 (2007).

41. E. J. Maginn, J. Phys. Condens. Matt. **21**, 373101 (2009).

42. J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing and J. Hutter, Comput. Phys. Commun. **167**, 103-128 (2005).

43. J. VandeVondele and J. Hutter, J. Chem. Phys. **127**, 114105 (2007).

44. G. Bussi, D. Donadio and M. Parrinello, J. Chem. Phys. **126**, 014101 (2007).

45. S. Goedecker, M. Teter and J. Hutter, Phys. Rev. B Condens. Matt. **54**, 1703-1710 (1996).

46. A. D. Becke, Phys. Rev. A **38**, 3098-3100 (1988).

47. C. Lee, W. Yang and R. G. Parr, Phys. Rev. B **37**, 785-789 (1988).

48. S. Grimme, J. Comput. Chem. **27**, 1787-1799 (2006).

49. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).

50. A. D. Becke and E. R. Johnson, J. Chem. Phys. **123**, 154101 (2005).

51. A. K. Soper, ISRN Phys. Chem. **2013**, 279463 (2013).

52. A. K. Soper and C. J. Benmore, Phys. Rev. Lett. **101**, 065502 (2008).

53. P. E. Mason, G. W. Neilson, J. E. Enderby, M. L. Saboungi and J. W. Brady, J. Phys. Chem. B **109**, 13104-13111 (2005).

54. G. Chessari, C. A. Hunter, C. M. Low, M. J. Packer, J. G. Vinter and C. Zonta, Chem. Eur. J. **8**, 2860-2867 (2002).

55. Z. Xu, H. H. Luo and D. P. Tieleman, J. Comput. Chem. **28**, 689-697 (2007).

56. G. A. Dilabio and A. Otero-de-la-Rosa, *Noncovalent Interactions in Density-Functional Theory*. (John Wiley & Sons, Inc, Hoboken, NJ., 2014).

57. J. Ireta, J. Neugebauer and M. Scheffler, J. Phys. Chem. A **108**, 5692-5698 (2004).