Preprint submited to Biochimica et Biophysica Acta (BBA) - General Subjects

Published in final form as https://doi.org/10.1016/j.bbagen.2019.129480

Oxygen-oxygen distances in protein-bound crystallographic water suggest the presence of protonated clusters.

Luigi Leonardo Palese*

*University of Bari "Aldo Moro", Department of Basic Medical Sciences, Neurosciences and Sense Organs (SMBNOS), Bari 70124, Italy.

Contact email: luigileonardo.palese@uniba.it

Abstract

Background. The availability of high-resolution X-ray structures has shown that proteins contain numerous water molecules, but their role is still not fully understood. Protonated and deprotonated water species are often involved in biochemical reactions. However protons are exceedingly difficult to detect directly because they are electron-poor species.

Methods. The oxygen-oxygen distance of the crystallographic water molecules was analyzed in a large high-resolution data set. Moreover, a detailed analysis was carried out on the protein-bound water in the available structures of carbonic anhydrase II and cytochrome *c* oxidase, chosen as protein models in which protonated and deprotonated water species play a significant role.

Results. The analysis shows an excess of water-water distances below the expected value for hydrogen bond. In the cavities and on the surface of the considered model proteins, clusters of water molecules are found, whose structure suggests the presence of chemical species deriving from self-ionization of water.

Conclusions. The presence of a small maximum below the hydrogen bond threshold in the oxygen-oxygen distance distribution of crystallographic water molecules, along with the location of many of these water clusters, suggest the presence of Zundel-like structures in, or near, the proteins. Particularly significant is the presence of such structures in protein regions which have been identified as proton antennae or channels.

General significance. This work shows the possibilities, still unexplored, offered by this type of analysis in detecting in structures obtained by X-ray diffraction the presence of aqueous protons or hydroxide ions, which are chemical species as important as elusive.

Keywords

Water cluster; proton; protein channel; protein surface; hydration shell; X-ray structures.

1. Introduction

The importance of water as a primary component of cellular biochemistry, far from being a mere spectator that acts "simply" as a solvent, is an increasingly pervasive concept in the field of life sciences [1,2]. The availability of high-resolution X-ray structures has shown that proteins contain numerous water molecules, but despite the large number of studies devoted to the protein - water interactions many aspects of this topic are still an intense field of investigation. For example, water molecules participate to proton transfer in biochemical reactions, and how this process, that occurs in random directions in liquid water, becomes directional in biochemically relevant events is an active field of research [1-3]. Also confinement effects on the properties of water in biochemical systems [1-7] are important topics of study.

In the (roughly) exponentially growing number of structures in the Protein Data Bank (PDB) [8], the coordinates of hydrogen atoms are rarely reported, and in general not for solvent water. And even if the presence of protonated water species in some systems is a well established fact [9,10], a systematic search for protonated and deprotonated water species within proteins or near their surface, has not been carried out to date. Over the years, neutron crystallography has become the state-of-the-art technique to visualize hydrogen atoms in biological macromolecules (or deuterium in *ad hoc* prepared samples), because neutron diffraction is due to the presence of nucleons [11]. From the classic studies that showed a doubly protonated histidine in the catalytic triad of trypsin [12], numerous data obtained by this technique have been reported, and in some cases also the presence of protonated water species has been demonstrated [13-15]. But the number of available structures obtained by neutron diffraction in the PDB is only a tiny fraction of those obtained by the (more popular) X-ray crystallography, whose diffraction intensity, at the contrary of neutrons, is related to the electron density of a region. Unfortunately, electron poor species, such as hydrogen atoms, are essentially invisible in the electron density maps obtained by X-ray diffraction [16].

Anyway, the large number of high-resolution structures obtained by X-ray crystallography and the importance of the hydrated proton and hydrated hydroxide in biochemistry (hereafter indicated as H^+_{aq} and OH^-_{aq}) suggest that the search for some kind of trace of these molecular species in the available structures in the PDB could be a fruitful task. The detailed nature of the H^+_{aq} and OH^-_{aq} is an important field of study, to which considerable theoretical and experimental efforts are dedicated [1-3]. The Eigen and Zundel ions [17,18] can be considered adequate descriptions of protonated forms of water in gas phase, but only limiting structures in liquid water. Crystallographic studies on nanoconfined H^+_{aq} have shown that oxygen-oxygen (O-O) distances below 2.70 Å can be associated to the presence of protonated water clusters [19-21]. This suggests immediately that the analysis of the distances between oxygen atoms attributed to water molecules in high-resolution X-ray structures could furnish hints about the presence of protonated water molecules. Also in the case of OH^- ion, the interaction with neighbor water molecules implies an O-O distance below 2.70 Å [22, 23]: the shortest O-O distance in water, of only 2.29 Å, has been observed in the hydrated hydroxide $H_3O_2^-$ in a crystal of a chromium complex [24-26]. Such a short distance implies the presence of a symmetrical single-well hydrogen bond.

In this work we show how the analysis of the O-O distances in the crystallographic water molecules can provide evidences about their protonation state. Our strategy is to search for the presence of O-O distances below the hydrogen bond expected value. This has been done in a large high-resolution data set of X-ray structures obtained from the PDB, showing that short distance water molecules (SDWMs) are present. Even if in forthcoming we will refer more often to "proton" or "protonated water" for short, the reader should bear in mind that these SDWM clusters can represent both positively (H^+_{aq}) or negatively (OH^-_{aq}) charged species of water. Furthermore, a detailed analysis of two systems known to be involved in proton transfer events is presented.

2. Materials and Methods

2.1 Data sets

The X-ray structures were obtained from the PDB [8]. The full data sets considered in this work are reported in the Supplementary Data [27]. The high resolution data set (HR) contains 469 entries; this was obtained by searching for structures with the following constraints: resolution ≤ 1 Å, X-ray only, protein only, monomer only, diffraction temperature 100 K. From this HR set, two subsets were obtained by adding to constraints the absence of the SHELX software [28] in the declared refinement conditions (the no-SHELX HR data set) or the absence of sodium ions in the declared experimental conditions (the sodium free HR data set). A further subset not refined by SHELX and without sodium ions in the declared crystallization conditions was analyzed. The PDB entries for these data sets are reported as Tables 1-4 in Supplementary Data [27].

Two data sets containing a single type of protein were obtained as follows. For the bovine cytochrome *c* oxidase (CcO) data set only structures with a resolution of at least 1.8 Å were considered. The diffraction temperature for these structures was 50 or 100 K (for a more detailed description, see Supplementary Data [27]). The human carbonic anhydrase II (hCA II) data set was obtained by adding to the no-SHELX HR data set the E.C. number 4.2.1.1 as constraint.

2.2 Numerical analyses

The atomic radial pair distribution function (RDF) *g(r)* of the crystallographic water molecules was calculated by a Tcl program using the appropriate built-in functions in VMD [29]. For each pdb file in a data set, this program returned a text file containing *r*, *g(r)* and the number integral over *g(r)*, from 0.0 to 10.0 Å with 0.1 Å resolution, where *r* is the distance from a given reference particle. The *g(r)* values were extracted from these text files, arranged in a matrix form and averaged in a Python (www.python.org) environment described previously [30-34].

The single-type protein data sets were analyzed as follows. First of all, the atomic coordinates of all atoms belonging to the molecule of interest, including water, were used to make a new pdb file; these pdb files were built for the CcO subunit I and for the single monomer in the case of multimeric crystals of the hCA II. Subsequently sequence and structural analyses were performed as described [30-34], including the structural alignment of the protein of interest. The coordinates of each superposed structure were stored in a new pdb file. The atomic RDFs were calculated as detailed above and the mutual Euclidean distances between all the oxygen atoms of the water molecules belonging to each protein were calculated by means of a Tcl program in a VMD environment. Numerical calculations were implemented in Python as described [30-34]. Further details on the data sets and methods are reported in Supplementary Data [27].

3. Results

3.1 The water radial distribution function in a high resolution data set

The crystal structures of the HR data set, refined at resolution of at least 1 Å, contain 174167 water molecules, including 11794 partially occupied positions. The entire data set is composed of structures diffracted at 100 K, in order to reduce the effects of thermal motions on the resolution of solvent molecules and to consider temperature homogeneous data. The RDF in this data set is reported in Figure 1, left panel. The highest peak in this distribution is at 2.75 Å, in excellent agreement with the O-O distance in hexagonal ice Ih at cryogenic temperatures obtained by neutron diffraction [35], in which the reported values were 2.750 Å at 60 K and 2.751 Å at 123 K. Even if the peak is slightly asymmetrical, with an excess of probability at 2.85 Å relative to that observed at 2.65 Å (see also below), we can conclude that water molecules, to which these oxygen atoms belong, are engaged in classic hydrogen bonds, as in ordinary ice. A second lower and broader peak is centered at 4.45 Å, again in agreement with the structure of the second shell in ordinary ice, and a third very weak peak indicates the presence of residual order at even greater distances.

An intriguing evidence is the presence of a small peak centered at 2.25 Å. An enlarged view of the RDF of this data set is shown as Figure 1, left panel, in the Supplementary Data [27]. This maximum was described previously [36] in a survey that included 105 high resolution structures. This peak is completely unexpected if one consider the RDF of pure water, and since it is evident both in the previous [36] as well as in the present analysis, it should be carefully considered.

A simple explanation for this peak could be the presence of sodium ions *incorrectly* assigned as water. This because the magnitude of X-ray diffraction depends on the electrons of the molecule or atom, and the sodium ion is isoelectronic to a water molecule. Protein-bound sodium ions, correctly identified as such, do not represent a problem in this analysis: in the whole HR data set there are 57 sodium ions in 35 different entries, including some obtained without the use of sodium in the crystallization medium. The shape and position of the peak suggest that the misinterpreted sodium

hypothesis is an extremely unlikely explanation, as suggested in [36]. To prove this suggestion, we use a subset of our HR data set, in which all the entries reporting the use of sodium in crystallization methods were excluded. The RDF of this data set is reported as Figure 2 in the Supplementary Data [27]. The peak at 2.25 Å is still present, therefore excluding that incorrect assignment of sodium can explain this finding.

Another possible explanation is a very trivial one, namely an artifact of the refinement programs used in protein crystallography. This eventuality was seriously considered in [36], where it was suggested that this maximum could be the consequence of the default value of the distance cut-off in the SHELX [28] software suite. In [36] this possibility, although strongly suggested, was not analyzed in detail, probably because this would have meant eliminating almost all the components of the data set available at that time. But actually our HR data set is sufficiently large to allow the elimination of the entries refined by this program. The RDF of this subset of structures not refined by SHELX, containing 269 entries, is reported in Figure 1, right panel. This clearly shows that a large part of the signal centered at 2.25 Å in the whole HR data set can be considered an artifact of the refinement process. But a probability excess still remains in the RDF values sampled at 2.30 - 2.50 Å as a small shoulder of the main peak (Figure 1, right panel; see also Figure 1, right panel, in Supplementary Data [27] for an enlarged view of the RDF of this data set).

Although the data presented above seem to exclude that sodium in the crystallization medium may be the cause of some signal in the RDF of protein-bound water, to definitively eliminate this possibility we analyzed the RDF of a subset of structures not refined by SHELX *and* not containing sodium in the reported crystallization protocol (see Figure 3 in Supplementary Data [27]). The RDFs of this last subset and of the no-SHELX HR data set are practically identical, with very small differences outside the region of interest. This is further proved by the scatter plot analysis of the two RDFs (see Figure 4 in Supplementary Data [27]), thus excluding the possibility that the small

shoulder at 2.30 - 2.50 Å in the no-SHELX HR data set may be attributed to the presence of sodium in the crystallization medium erroneously interpreted as water.

A plausible explanation for this shoulder is that it arises from some peculiar structural arrangement of water. Particularly, direct X-ray determination of the locations of hydrogen atoms in water molecules or related species (such as protonated water species) is not possible because of their low electron density. Nevertheless, excess protons in water can be revealed by the O-O distances: if we observe O-O distances shorter than those expected in hexagonal ice (see above) or liquid water (2.85 Å) we are facing potential sites for H^+ bridging [19-21]. Protonated (and deprotonated) forms of water are characterized by an O-O length distance below 2.70 Å, where short separations (≤ 2.42 Å) are associated with symmetrical H-bonding in the $H_5O_2^+$ Zundel ion, which is a limiting structure of the hydrated proton; even shorter distances can be expected in the case of the hydrated hydroxide [24-26]. Our RDF analysis suggests that candidate pairs of water molecules bound by a H^+ bridge are rare, but not absent, in protein crystals.

3.2 The high-resolution structures of human carbonic anhydrase II

The HR data set is extremely heterogeneous, because the only common trait of these structures is that these are obtained at the same temperature and have a resolution of at least 1.0 Å. Obviously the same applies to the subsets of it considered in this work. Consequently, a detailed analysis of all the water molecules characterized by a short O-O distance is practically impossible. However, in the no SHELX HR data set there are some entries that belong to a highly studied protein, even in the context of proton transfer phenomena, namely the hCA II. This is a zinc metalloenzyme that catalyzes the reversible hydration/dehydration of CO_2/HCO_3^- [37]. It is a cytosolic isoform expressed in most cell types and involved in several processes. The primary function of carbonic anhydrases in animals is to maintain the acid-base balance of blood and tissues and to facilitate the

transport of CO_2. The hydration of CO_2 occurs by the nucleophilic attack of a zinc-bound hydroxide. Subsequently the bicarbonate ion is displaced by a water molecule, which loses a proton thus regenerating the zinc-bound hydroxide. This proton transfer process is mediated by ordered water molecules in the active site that connect the zinc-bound H_2O/OH^- to H64 [37-41]. Since this water network participates to proton transfer, it is conceivable that their components can be trapped in protonated forms, at least partially and transiently. To verify this hypothesis, we analyzed the hCA II entries in this data set (4MTY, 4Q78, 4YXI, 5LJT, 5OGO, 5Y2R, 5Y2S and 6B00 [42-47]). The mutual Euclidean distances of all oxygen atoms reported in the pdb file as belonging to water molecules have been calculated. Further, the electron density maps of the atom pairs whose mutual distance was in the range 2.29 - 2.50 Å were analyzed. This interval was chosen considering the shorter O-O distance between water molecules reported in the literature [24-26] and the peak of the shoulder in the RDF (Figure 1 in Supplementary Data [27]). This probably excludes from the analysis some interesting SDWMs: for example, taking into account the uncertainty associated with the temperature factor, pairs separated by distances around 2.20 Å in the crystallographic PDB structures could actually be above the minimum allowed O-O distance. But this restrictive criterion keeps the possibility of false positives to a minimum. The atoms whose position did not correspond to a well defined electron density at 1.0 sigma were excluded. The atoms were subsequently excluded if the electron density map at 4.0 sigma was positive at their location. Using this last criterion, some oxygen atoms belonging to extremely ordered water molecules (and probably also water-related species) were excluded from the analysis, such as the metal-bound ones. However, this avoids to consider as water other chemicals that could be present in the crystallization medium, particularly those of the third period of the periodic table (with the exception of sodium ion, which is isoelectronic to water). The water pairs that meet all these criteria are reported in Supplementary Data [27]; we retained 56 O-O pairs over 560371 considered, as detailed in Supplementary Data [27], Table 5. Inspection of the above mentioned table shows that no meaningful relations exist

between the number of putative SDWM pairs and other parameters, such as resolution, number of water molecules per monomer and so on. In 5LJT the low number of candidates could be related to the refinement process, because no pairs below 2.44 Å can be observed, as opposed to all the other structures of hCA II in the no-SHELX HR data set. In some cases different structures show a raw number of candidate pairs that is similar before the map analyses, but it becomes very different after the quality check. An absolute quantitative analysis of the number of clusters of short-distance water molecules is far beyond the scope of this work. But a noteworthy fact is the high number of SDWMs in hCA II structures acquired at high CO_2 pressure, namely 5Y2R and 5Y2S [46], obtained 2.5 and 7.0 atm CO_2 and containing 12 and 17 SDWM pairs respectively (see Table 5 in Supplementary Data [27]).

The localization on the protein of these pairs of water molecules having an O-O distance compatible with the presence of protonated forms is shown in Figure 2 (note that the Figure reports cumulatively all the SDWM pairs observed in the hCA II PDB entries discussed in this Section). As expected, most of these are in the active site (see Figure 2, left panel). An interesting fact that emerges from this analysis is the presence of numerous pairs (or even larger structures) of SDWMs that are located at the surface of the protein. This can be observed also in the structures of hCA II not included in this data set because of lower resolution or different diffraction temperature (not shown). As can be easily seen by looking at Figure 2, a preferential location of these SDWMs is evident in some regions of the protein surface; similar behavior is observed in the lower resolution hCA II molecules (not shown). The presence of SDWMs on the surface of the protein and their preferential localization in some areas of the surface is a quite general phenomenon, and not limited to the hCAII (not shown).

3.3 Water molecules in cytochrome *c* oxidase

Another interesting system in which proton transfer occurs is CcO. This is a large and complicated membrane complex composed of protein subunits, redox active centers, lipids and water that operates the terminal step of cellular respiration [48-57]. CcO contains two low potential, redox-active metal centers: the bimetallic center Cu_A in subunit II (which titrates as a single electron center) and heme *a* in subunit I. These redox active centers donate electrons to the oxygen reduction site, composed of two high potential centers in subunit I, heme a_3 and Cu_B (the binuclear center). Beside charge movements directly associated to the oxygen reduction chemistry, the enzyme pumps up to four protons during the four-electrons catalytic cycle. Subunit I is the key player both for oxygen reduction and for proton pump; in this protein three channels have been identified in crystallographic structures (labeled as K, D and H channels, based on key conserved amino acids), which can deliver the protons needed for the oxygen redox chemistry and those pumped [48-52,58,59].

In the PDB there are some CcO structures obtained at 50 K and characterized by a resolution below 1.8 Å [50,60,61]: 5B1A (fully oxidized state), 5B1B (fully reduced state), 5B3S (carbon monoxide-bound mixed-valence), 5XDQ (fully oxidized state), 5ZCP and 5ZCQ (azide-bound states obtained by long time exposure to 20 mM or 10 mM azide solutions, respectively). All of these were obtained at pH 6.8, excluding 5XDQ which was crystallized at pH 7.3. These structures were obtained using sodium in the crystallization protocol, but none reports the use of SHELX in the deposited experimental protocol. We analyzed the water molecules contained in the subunit I of these structures (a data set containing CcO structures at lower resolution whose diffraction temperature was 100 K was also considered; see Supplementary Data [27]). The number of crystallographic water molecules attributed to these proteins is reported in Supplementary Data [27] Table 6; in these structures an inverse relationship can be observed between resolution and number of water molecules present in the crystal (Figure 5 in Supplementary Data [27]).

Figure 3 reports the RDF of the CcO high-resolution data set: the general appearance is similar to the HR data set, with a maximum at 2.75 Å. Again, as in the HR data set, the peak is slightly asymmetrical, with an excess of probability at 2.85 Å relative to that observed at 2.65 Å, and partial order in the second shell is also evident. The asymmetry of the main peak is more evident if, instead of the RDF, we use the distribution of the Euclidean distances of the O-O pairs of nearby water molecules. This distribution for the CcO structures obtained at 50 K is reported as Figure 6 in Supplementary Data [27]. This distribution shows a major peak centered at 2.78 Å, close to what is observed in the hexagonal ice Ih [35]. So from the RDF and the distribution of the Euclidean reciprocal distances we can conclude that the water molecules to which these oxygen atoms belong are engaged in classic hydrogen bonds, as in ordinary ice. Moreover, in the Euclidean distance distribution, beside the major peak, a secondary one centered at 2.88 Å is evident. This can be related to the O-O distance in liquid water, which equals 2.85 Å [62]. Other secondary (and noisy) peaks can be observed at greater distances, reflecting a partial order in the local structure of CcO water molecules. Two similar major peaks can be observed also in the O-O distance distribution in a data set obtained considering CcO structures at lower resolution, diffracted at 100 K (Figure 7 in Supplementary Data [27]).

If we hold as threshold criterion for putative bridging H^+ sites O-O distances in the range 2.29 - 2.50 Å as above and considering the electron density maps at 1.0 and 4.0 sigma as described for the hCA II, we obtain 37 candidates in the data set (see Table 6 in Supplementary Data [27] for details) These water clusters are located mainly at the N- and P-surface of the protein, sometimes in groups containing three water molecules (suggesting the presence of hydrated proton as an $H_7O_3^+$ ion). It is interesting to note that at the N-surface of the protein these SDWM clusters are located in the neighbor of the entry mouth of the three proton channels (H, D and K), as shown in Figure 4. This finding supports the hypothesis that these clusters are not mere artifacts, but instead regions of the water network where excess (or defect) protons localize. Another region at the N-side of the protein

surface in which a number of water clusters characterized by low O-O distances are visible is between the residues H503, E506 and E507. Notably, H503 has been identified as one of the residues involved in the binding of a zinc ion at the N-side of CcO, whose depressive effect on the proton pump activity is well documented [52,63]. At the P-side of the protein surface there are some SDWM clusters too. The most notable one is located in a cleft delimited by residues H138, E119, A122, S142 and Q52 (see Figure 4). It should be noted that no SDWMs are detectable near the D51, the heme propionates or the magnesium ion (the reader should keep in mind that this in no way implies that these residues are not involved in pumping).

Inside the protein there are seldom candidate clusters of SDWMs, which are detectable only in the azide inhibited species (see Figure 5). A water pair is visible in the D channel in one of the CcO subunits of the enzyme treated with 10 mM azide, and candidate SDWM clusters are visible in both the azide inhibited species near the conserved H413 in the H channel. These last clusters are at hydrogen bond distance from the imidazole ring of H413 (the cluster toward the P side) or of a water molecule that is at hydrogen bond distance of the same residue (the one located toward the N side). Determining if these structure composed of H413 and water molecules are frozen intermediate of the pump (which would imply that the channel is important for the CcO proton pump [49,50]) or strong proton trap (which would suggest a dielectric well role for the channel [59]) will require further analyses.

4. Discussion

In this report we have shown that it is possible to obtain information about the presence of water clusters containing an excess (or defect) of protons, inside or at the surface of proteins, by the analysis of their O-O distances. In the presence of anomalously low O-O values (the SDWM clusters described above) it is very probable that we are facing such a type of chemical species [19,20]. With the exception of those observed in bacteriorhodopsin and CA II [9,10], to date no *structural* evidences of protonated (or deprotonated) water molecules have been reported in the PDB entries obtained by X-ray diffraction.

Our data confirm the presence of a strong peak at the expected distance for hydrogen-bonded water molecules, whose shape suggest the presence of at least two different more frequent values of the O-O distance. The fact that two types of hydrogen bonds can be observed is in itself interesting. The (old) notion that water can be best described as a two-phase mixture of ice-like and liquid-like configurations, although no longer sustainable in its original formulation, still maintains its influence in the current molecular models of water [2]. However it should be noted that here we are not dealing with bulk water but with protein-bound water; consequently, the presence of ice-like and liquid-like values in the hydrogen bond distances can be justified simply by the microscopic complexity of samples. A small maximum around 2.3 Å was previously reported in the RDF of crystallographic water in a data set containing atomic resolution protein structures. Even if the possibility of protonated water forms was not excluded, this finding was attributed mainly to refinement artifacts [36]. Actually our results clearly show that a large part of this peak in the RDF is due to refinement artifacts. However, after excluding the structures affected by these artifacts, the presence of a residual density in this region of the RDF indicates that protonated water clusters should be taken into account to explain this anomaly in the O-O distance distribution of protein-bound water.

The claim that we can locate the clusters of protonated water within the crystallographic structure of a protein immediately raises some problems. First of all, the H^{+}_{aq} is a transient species, that decays in the picosecond time scale in liquid water: i.e. the hopping phenomenon will occur after (or within) a waiting time of this order of magnitude. Even if we are considering diffraction patterns obtained at 50 K or 100 K, proton diffusion in amorphous ice at cryogenic temperatures can be still significant [64,65]. Rotational motions should be expected also in the case of well ordered water molecules and labile protons of ionizable side-chains can exchange too. Consequently, a hypothetical "free" proton has time to visit a number of water molecules in the crystal during the diffraction time. So how can we expect to see a labile structure in a particular position? Another important point is the expected number of hydronium (and protonated water species in general) and hydroxide ions in the vicinity of a protein. If we compare the number of SDWMs reported in this analysis for each protein with the expected number of H^{+}_{aq} and OH^{-}_{aq} in an equivalent volume of solution at the crystallization pH, the number of SDWMs might seem too large. Again, we must consider that here we are considering not bulk water, but water bound to proteins. Theoretical works have shown that two-dimensional confinement between Lennard-Jones walls enhances the self-dissociation of water [6] (however see also [7]), thus suggesting the possibility that the nanoconfined water inside or at the surface of a protein could exhibit an increased ionic product. Moreover, in the first layers of protein hydration the presence of charges capable of neutralizing those of the protein is expected, therefore it is conceivable that species deriving from the self-dissociation of water can be part of the system of counterions.

A plausible interpretation of reported data is therefore probabilistic: the SDWM clusters are simply crystallographic water molecules that preferentially bind the excess of protons (see scheme in Figure 6). Or equivalently, these are positions in the water molecule network where it is much more probable than average to find an excess of protons in a given instant (or a proton defect). Excess proton localization should be considered as averaged over the X-ray exposure time, which is

typically in the order of seconds. For example, near a negatively charged amino acid is probable on average to find a proton, i.e. a SDWM pair. Amino acid charges are not the only explanation for SDWMs: surface clefts and nearby structures into the crystal can be also considered, as well as the experimental conditions. About this, it is suggestive to note the number of SDWM pairs in the high-resolution hCA II exposed to high CO_2 pressure [46], i.e. obtained in the presence of high chemical activity of the enzyme substrate.

The hypothesis above explains also why in hCA II structures obtained by neutron diffraction no evidences of protonated (or deprotonated) water clusters are reported, where instead their presence is suggested in this work. Although this technique can detect the presence of hydrogen atoms, because it is the presence of nucleons that determines the neutron diffraction, nevertheless this requires the presence of ordered and stable atomic nuclei. Consequently, and similarly to X-ray electron density maps, no direct signals are expected in the case of dynamic or static disorder, as suggested above for the proton excess (or defect) in the crystallographic water network.

It should be noted that, also assuming the average positions reported in the PDB for each water molecule as *absolutely* exact, the Debye-Waller factor associated to each atom implies that static or dynamic mobility of crystallographic water molecules must be taken into account. This source of uncertainty forces us to consider these SDWMs as potential candidates, and not as certain sites where a charged water cluster is located (not shown). To be reasonably sure that we are dealing with protonated forms of water, further considerations have to be made. So we consider significant the presence of such SDWM clusters in protein regions in which proton transfer occurs, i.e. the channels in CcO and the water network that leads protons from the protein surface to the zinc-bound compound in the hCA II (or *vice versa*). Also significant is the presence of such clusters in known antenna regions at the protein surface of hCA II (E69 and D72; not shown) and of CcO (near the residues H503, E506 and E507; see Figure 4) [66-70]. This latter finding is in accord also with a recently proposed mechanism of CcO pump, in which thermal type motions on the P and N surfaces

of CcO subunit I participate in the thermodynamic barrier height fluctuations, which are crucial for proton uptake and release [71]. Finally, the presence of charged water species in the hydration layer at the protein surface is a well documented fact [72,73] and our data show the ubiquitous presence of SDWM clusters at the protein surface, as a sort of electrical double layer of proteins.

Acknowledgements

The author would like to thank the anonymous Reviewers for their valuable comments, which helped to improve significantly the manuscript.

References

[1] P. Ball, Water as an active constituent in cell biology, Chem. Rev. 108 (2008) 74-108. https://doi.org/10.1021/cr068037a.

[2] E. Brini, C.J. Fennell, M. Fernandez-Serra, et al., How water's properties are encoded in its molecular structure and energies, Chem. Rev. 117 (2017) 12385-12414. https://doi.org/10.1021/acs.chemrev.7b00259.

[3] N. Agmon, H.J. Bakker, R.K. Campen, et al., Protons and hydroxide ions in aqueous systems, Chem. Rev. 116 (2016) 7642-7672. https://doi.org/10.1021/acs.chemrev.5b00736.

[4] A. Morozenko, I.V. Leontyev, A.A. Stuchebrukhov, Dipole moment and binding energy of water in proteins from crystallographic analysis, J. Chem. Theory Comput. 10 (2014) 4618-4623. https://doi.org/10.1021/ct500358r.

[5] J.M. Swanson, C.M. Maupin, H. Chen, et al., Proton solvation and transport in aqueous and biomolecular systems: insights from computer simulations, J. Phys. Chem. B 111 (2007) 4300-4314. https://doi.org/10.1021/jp070104x.

[6] D. Muñoz-Santiburcio, D. Marx, Nanoconfinement in slit pores enhances water self-dissociation, Phys. Rev. Lett. 119 (2017) 056002. https://doi.org/10.1103/PhysRevLett.119.056002.

[7] Y.A.P. Sirkin, A. Hassanali, D.A. Scherlis, One-dimensional confinement inhibits water dissociation in carbon nanotubes, J. Phys. Chem. Lett. 9 (2018) 5029-5033. https://doi.org/10.1021/acs.jpclett.8b02183.

[8] H.M. Berman, J. Westbrook, Z. Feng Z, et al., The Protein Data Bank, Nucleic Acids Res. 28 (2000) 235-242. https://doi.org/10.1093/nar/28.1.235.

[9] S. Wolf, E. Freier, M. Potschies, et al., Directional proton transfer in membrane proteins achieved through protonated protein-bound water molecules: a proton diode, Angew. Chem. Int. Ed. Engl. 10 (2010) 6889-6893. https://doi.org/10.1002/anie.201001243.

[10] B. S. Avvaru, C. U. Kim, K. H. Sippel, et al., A short, strong hydrogen bond in the active site of human carbonic anhydrase II, Biochemistry 49 (2010) 249–251. https://doi.org/10.1021/bi902007b.

[11] P. Langan, J. C.-H. Chen, Seeing the chemistry in biology with neutron crystallography, Phys. Chem. Chem. Phys. 15 (2013) 13705-13712. https://doi.org/10.1039/c3cp51760h.

[12] A. A. Kossiakoff, S. A. Spencer, Direct determination of the protonation states of aspartic acid-102 and histidine -57 in the tetrahedral intermediate of the serine proteases: neutron structure of trypsin, Biochemistry 20 (1981) 6462-6474. https://doi.org/10.1021/bi00525a027.

[13] A. Y. Kovalevsky, B. L. Hanson, S. A. Mason et al., identification of the elusive hydronium ion exchanging roles with a proton in an enzyme at lower pH values, Angew. Chem. Int. Ed. 50 (2011) 7520-7523. https://doi.org/10.1002/anie.201101753.

[14] M. G. Cuypers, S. A. Mason, M. P. Blakeley, et al., Near-atomic resolution neutron crystallography on perdeuterated *Pyrococcus furiousus* rubredoxin: implication of hydronium ions

and protonation state equilibria in redox changes, Angew. Chem. Int. Ed. 52 (2013) 1022-1025. https://doi.org/10.1002/anie.201207071.

[15] M. Unno, K. Ishikawa-Suto, K. Kusaka, et al., Insights into the proton tranfer mechanism of a bilin reductase PcyA following neutron crystallography, J. Am. Chem. Soc. 137 (2015) 5452-5460. https://doi.org/10.1021/jacs.5b00645.

[16] A. Wlodawer, W. Minor, Z. Dauter, M. Jaskolski, Protein crystallography for non-crystallographers, or how to get the best (but no more) from published macromolecular structures, FEBS J. 275 (2008) 1-21. https://doi.org/10.1111/j.1742-4658.2007.06178.x.

[17] M. Eigen, Proton transfer, acid-base catalysis, and enzymatic hydrolysis, Angew. Chem. Int. Ed. 3 (1964) 1-19. https://doi.org/10.1002/anie.196400011.

[18] G. Zundel, Hydrogen bonds with large proton polarizability and proton transfer processes in electrochemistry and biology, Adv. Chem. Phys. 111 (2000) 1.217. https://doi.org/10.1002/9780470141700.ch1.

[19] E.S. Stoyanov, I.V. Stoyanova, C.A. Reed, The structure of the hydrogen ion (H(aq)+) in water, J. Am. Chem. Soc. 132 (2010) 1484-1485. https://doi.org/10.1021/ja9101826.

[20] E.S. Stoyanov, I.V. Stoyanova, F.S. Tham, C.A. Reed, H(aq)+ structures in proton wires inside nanotubes, J. Am. Chem. Soc. 131 (2009) 17540-17541. https://doi.org/10.1021/ja907708g.

[21] C.A. Reed, Myths about the proton. The nature of H^+ in condensed media, Acc. Chem. Res. 46 (2013) 2567-2575. https://doi.org/10.1021/ar400064q.

[22] N. Agmon, Mechanism of hydroxide mobility, Chem. Phys. Lett. 319 (2000) 247-252. https://doi.org/10.1016/S0009-2614(00)00136-6-

[23] S. T. Roberts, P. B. Petersen, K. Ramasesha, et al., Observation of a Zundel-like transition state during proton transfer in aqueous hydroxide solutions, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 15154-15159. https://doi.org/10.1073/pnas.0901571106.

[24] K. Abu-Dari, D. P. Freyberg, K. N. Raymond, Coordination chemistry of microbial iron transport compounds. 18. Crystal and molecular structure of disodium triethylmethylammonium tris(thiobenzohydroximato)chromate(III) hemikis(sodium hydroxide hydrate), Na2[(C2H5)3(CH3)N][Cr(PhC(S):N(O))3].1/2NaH3O2.18H2O, Inorg. Chem. 18 (1979) 2427-2433. https://doi.org/10.1021/ic50199a020.

[25] K. Abu-Dari, K. N. Raymond, D. P. Freyberg, The bihydroxide ($H_3O_2^-$) anion. A very short, symmetric hydrogen bond, J. Am. Chem. Soc. 101 (1979) 3688-3689. https://doi.org/10.1021/ja00507a059.

[26] W. W. Cleland, P. A. Frey, J. A. Gerlt, The low barrier hydrogen bond in enzymatic catalysis, J. Biol. Chem. 273 (1998) 25529-25532. https://doi.org/10.1074/jbc.273.40.25529.

[27] Submitted to Data in Brief.

[28] G. M. Sheldrick, A short hystory of SHELX, Acta Cryst. A 64 (2008) 112-122. https://doi.org/10.1107/S0108767307043930.

[29] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33-38. https://doi.org/10.1016/0263-7855(96)00018-5.

[30] F. Bossis, A. De Grassi, L.L. Palese, C.L. Pierri, Prediction of high- and low-affinity quinol-analogue-binding sites in the aa3 and bo3 terminal oxidases from Bacillus subtilis and Escherichia coli, Biochem. J. 461 (2014) 305-314. https://doi.org/10.1042/BJ20140082.

[31] L.L. Palese, Random Matrix Theory in molecular dynamics analysis, Biophys. Chem. 196 (2015) 1-9. https://doi.org/10.1016/j.bpc.2014.08.007.

[32] L.L. Palese, Correlation analysis of Trp-cage dynamics in folded and unfolded states, J. Phys. Chem. B 119 (2015) 15568-15573. https://doi.org/10.1021/acs.jpcb.5b09678.

[33] L.L. Palese, Protein states as symmetry transitions in the correlation matrices, J. Phys. Chem. B 120 (2016) 11428-11435. https://doi.org/10.1021/acs.jpcb.6b09216.

[34] L.L. Palese, Conformations of the HIV-1 protease: A crystal structure data set analysis, Biochim. Biophys. Acta 1865 (2017) 1416-1422. https://doi.org/10.1016/j.bbapap.2017.08.009.

[35] W.F. Kuhs, M.S. Lehmann, The structure of the ice Ih by neutron diffraction, J. Phys. Chem. 87 (1983) 4312-4313. https://doi.org/10.1021/j100244a063.

[36] X. Chen, I. Weber, R.W. Harrison, Hydration water and bulk water in proteins have distinct properties in radial distributions calculated from 105 atomic resolution crystal structures, J. Phys. Chem. B 112 (2008) 12073-12080. https://doi.org/10.1021/jp802795a.

[37] C. L. Lomelino, J. T. Andring, R. McKenna, Crystallography and its impact on carbonic anhydrase research, Int. J. Med. Chem. 2018 (2018) 9419521. https://doi.org/10.1155/2018/9419521.

[38] C. M. Maupin, R. McKenna, D. N. Silverman, G. A. Voth, Elucidation of the proton transport mechanism in human carbonic anhydrase II, J. Am. Chem. Soc. 131 (2009) 7598-7608. https://doi.org/10.1021/ja8091938.

[39] D. Riccardi, P. König, H. Guo, Q. Cui, Proton transfer in carbonic anhydrase is controlled by electrostatics rather than the orientation of the acceptor, Biochemistry 47 (2008) 2369-2378. https://doi.org/10.1021/bi701950j.

[40] A. Roy, S. Taraphder, Identification of proton-transfer pathways in human carbonic anhydrase II, J. Phys. Chem. B 111 (2007) 10563-10576. https://doi.org/10.1021/jp073499t.

[41] S. Z. Fisher, A. Y. Kovalevsky, J. F. Domsic, et al., Neutron structure of human carbonic anhydrase II: implications for proton transfer, Biochemistry 49 (2010) 415-421. https://doi.org/10.1021/bi901995n.

[42] J. Buratto, C. Colombo, M. Stupfel, et al., Structure of a complex formed by a protein and a helical aromatic ologoamide foldamer at 2.1 A resolution, Angew. Chem. Int. Ed. 53 (2014) 883-887. https://doi.org/10.1002/anie.201309160.

[43] P. Mader, A. Pecina, P. Cigler, et al., Carborane-based carbonic anhydrase inhibitors: insight into CAII/CAIX specificity from a high-resolution crystal structure, modeling, and quantum chemical calculations, Biomed. Res. Int. 2014 (2014) 389869. https://doi.org/10.1155/2014/389869.

[44] R. Gaspari, C. Rechlin, A. Heine, et al., Kinetic and structural insights into the mechanism of binding of sulfonamides to human carbonic anhydrase by computational and experimental studies, J. Med. Chem. 59 (2016) 4245-4256. https://doi.org/acs.jmedchem.5b01643.

[45] A. Nocentini, M. Ferraroni, F. Carta, et al., Benzenesulfonamides incorporating flexible triazole moieties are highly effective carbonic anhydrase inhibitors: synthesis, and kinetic, crystallographic, computational, and intraocular pressure lowering investigations, J. Med. Chem. 59 (2016) 10692-10704. https://doi.org/10.1021/acs.jmedchem6b01389.

[46] J. K. Kim, C. L. Lomelino, B. S. Avvaru, et al., Active-site solvent replenishment observed during human carbonic anhydrase catalysis, IUCrJ 5 (2018) 93-102. https://doi.org/10.1107/S2052252517017626.

[47] K. M. Kean, J. J. Porter, R. A. Mehl, P. A. Karplus, Structural insights into a thermostable variant of human carbonic anhydrase II, Protein Sci. 27 (2018) 573-577. https://doi.org/10.1002/pro.3347.

[48] M. Wikström, K. Krab, V. Sharma V. Oxygen activation and energy conservation by cytochrome c oxidase, Chem. Rev. 118 (2018) 2469-2490. https://doi.org/10.1021/acs.chemrev.7b00664.

[49] S. Yoshikawa, K. Muramoto, K. Shinzawa-Itoh, Proton-pumping mechanism of cytochrome c oxidase. Annu. Rev. Biophys. 40 (2011) 205-223. https://doi.org/10.1021/10.1146/annurev-biophys-042910-155341.

[50] A. Shimada, K. Hatano, H. Tadehara, et al., X-ray structural analyses of azide-bound cytochrome c oxidases reveal that the H-pathway is critically important for the proton-pumping activity, J. Biol. Chem. 293 (2018) 14868-14879. doi:10.1074/jbc.RA118.003123.

[51] S. Papa, N. Capitanio, G. Capitanio, L.L. Palese, Protonmotive cooperativity in cytochrome c oxidase, Biochim. Biophys. Acta 1658 (2004) 95-105. https://doi.org/10.1016/j.bbabio.2004.04.014.

[52] N. Capitanio, L.L. Palese, G. Capitanio, et al., Allosteric interactions and proton conducting pathways in proton pumping aa(3) oxidases: heme a as a key coupling element, Biochim. Biophys. Acta 1817 (2012) 558-566. https://doi.org/10.1016/j.bbabio.2011.11.003.

[53] S. Lobasso, L.L. Palese, R. Angelini, A. Corcelli, Relationship between cardiolipin metabolism and oxygen availability in Bacillus subtilis, FEBS Open Bio 3 (2013) Feb 19;3:151-5. https://doi.org/10.1016/j.fob.2013.02.002.

[54] C.Y. Son, A. Yethiraj, Q. Cui, Cavity hydration dynamics in cytochrome c oxidase and functional implications, Proc. Natl. Acad. Sci. USA 114 (2017) E8830-E8836. https://doi.org/10.1073/pnas.1707922114.

[55] M. Wikström, M.I. Verkhovsky, G. Hummer G, Water-gated mechanism of proton translocation by cytochrome c oxidase, Biochim. Biophys. Acta 1604 (2003) 61-65. https://doi.org/10.1016/S0005-2728(03)00041-0.

[56] S. Supekar, A.P. Gamiz-Hernandez, V.R. Kaila, A protonated water cluster as a transient proton-loading site in cytochrome c oxidase, Angew. Chem. Int. Ed. Engl. 55 (2016) 11940-11944. https://doi.org/ 10.1002/anie.201603606.

[57] A. Farahvash, A. Stuchebrukhov, Investigating the many roles of internal water in cytochrome c oxidase, J. Phys. Chem. B. 122 (2018) 7625-7635. https://doi.org/10.1016/10.1021/acs.jpcb.7b11920.

[58] M.M. Pereira, M. Santana, M. Teixeira, A novel scenario for the evolution of haem-copper oxygen reductases, Biochim. Biophys. Acta 1505 (2001) 185-208. https://doi.org/10.1016/S0005-2728(01)00169-4.

[59] P.R. Rich, A. Maréchal, Functions of the hydrophilic channels in protonmotive cytochrome c oxidase, J. R. Soc. Interface 10 (2013) 20130183. https://doi.org/10.1098/rsif.2013.0183.

[60] N. Yano, K. Muramoto, A. Shimada, et al., The Mg^{2+}-containing water cluster of mammalian cytochrome c oxidase collects four pumping proton equivalents in each catalytic cycle, J. Biol. Chem. 291 (2016) 23882-23894. https://doi.org/10.1074/jbc.M115.711770.

[61] F. Luo, K. Shinzawa-Itoh, K. Hagimoto, et al., Structure of bovine cytochrome c oxidase crystallized at a neutral pH using a fluorinated detergent, Acta Cryst. F 73 (2017) 416-422. https://doi.org/ 10.1107/S2053230X17008834.

[62] A.H. Narten, W.E. Thiessen, L. Blum, Atom pair distribution functions of liquid water at 25°C from neutron diffraction, Science 217 (1982) 1033-1034. https://doi.org/10.1126/science.217.4564.1033.

[63] F. Francia, L. Giachini, F. Boscherini, et al., The inhibitory binding site(s) of Zn2+ in cytochrome c oxidase, FEBS Lett. 581 (2007) 611-616. https://doi.org/10.1016/j.febslet.2007.01.017

[64] I. Presiado, J. Lal, E. Mamontov, et al., Fast proton hopping detection in ice Ih by quasi-elastic neutron scattering, J. Phys. Chem. C 115 (2011) 10245-10251. https://doi.org/10.1021/jp2008094.

[65] D.H. Lee, H. Kang, H. Kang, Tunneling diffusion of excess protons in amorphous solid water at 10 and 80 K, J. Phys. Chem. C 123 (2019) 3657-3663. https://doi.org/10.1021/acs.jpcc.8b11829.

[66] V. Sacks, Y. Marantz, A. Aagaard, et al., The dynamic feature of the proton collecting antenna of a protein surface, Biochim. Biophys. Acta 1365 (1998) 232-240. https://doi.org/10.1016/S0005-2728(98)00073-5.

[67] Y. Marantz, E. Nachliel, A. Aagaard, et al., The proton collecting function of the inner surface of cytochrome c oxidase from Rhodobacter sphaeroides, Proc. Natl. Acad. Sci. USA 95 (1998) 8590-8595. https://doi.org/10.1073/pnas.95.15.8590.

[68] Y. Georgievskii, E.S. Medvedev, A.A. Stuchebrukhov, Proton transport via the membrane surface, Biophys. J. 82 (2002) 2833-2846. https://doi.org/10.1016/S0006-3495(02)75626-9.

[69] A.Y. Mulkidjanian, J. Heberle, D.A. Cherepanov, Protons @ interfaces: implications for biological energy conversion, Biochim. Biophys. Acta 1757 (2006) 913-930. https://doi.org/10.1016/j.bbabio.2006.02.015.

[70] S. I. Noor, S. Jamali, S. Ames, et al., A surface proton antenna in carbonic anhydrase II supports lactate transport in cancer cells, eLife 7 (2018) e35176. https://doi.org/10.7554/eLife.35176.

[71] L.L. Palese, Cytochrome c oxidase structures suggest a four-state stochastic pump mechanism, Phys. Chem. Chem. Phys. 21 (2019) 4822-4830. https://doi.org/10.1039/c8cp07365a.

[72] G. Careri, M. Geraci, A. Giansanti, J. A. Rupley, Protonic conductivity of hydrated lysozyme powders at megahertz frequencies, Proc. Natl. Acad. Sci. USA 82 (1985) 5342-5346. https://doi.org/10.1073/pnas.82.16.5342.

[73] G. Careri, A. Giansanti, J. A. Rupley, Proton percolation on hydrated lysozyme powders, Proc. Natl. Acad. Sci. USA 83 (1986) 6810-6814. https://doi.org/10.1073/pnas.83.18.6810.

Figure legends

Figure 1. The RDF in the HR data sets. The Figure shows the oxygen water RDF of the whole HR data set containing 469 entries (left panel) and of the HR subset of structures not refined by SHELX containing 269 entries (right panel), described in the main text. All structures in the data set have a resolution of at least 1 Å and were collected at 100 K.

Figure 2. The hCA II short distance water pairs. The Figure shows the short distance water pairs in the hCA II entries in the HR data set not refined by the SHELX program (see text). These structures, including the water molecules, were aligned to a common references. Protein surface refers to the 5Y2R structure, obtained by a 1.4 Å probe. Red, blue, green and white regions correspond to acidic, basic, polar and non-polar residues, respectively. Left panel is a view from the entry mouth of the active site, whilst the opposite side is reported in the right panel. The positions of water molecules belonging to short distance pairs are marked as orange spheres (all water pairs in the considered structures are cumulatively reported). In the active site are visible also a CO_2 molecule (in licorice), the zinc ion (yellow sphere) and the oxygen of the H_2O/OH^- zinc-bound species (red sphere).

Figure 3. The RDF in the CcO data sets. The Figure shows the oxygen water RDF of the CcO data set containing the high-resolution structures diffracted at 50 K described in the text. All structures in the data set have a resolution < 1.8 Å.

Figure 4. The CcO short distance water pairs. The Figure shows the short distance water pairs in the CcO high-resolution structures obtained at 50 K (see text). These structures, including the water molecules, were aligned to a common references. Protein surface refers to the subunit A of 5B1A,

obtained by a 1.2 Å probe. Red, blue, green and white regions correspond to acidic, basic, polar and non-polar residues, respectively (in cyan a residue classified as unassigned by VMD). The positions of water molecules belonging to short distance pairs are marked as orange spheres (all water pairs in the considered structures are cumulatively reported). Left panel: P-side view of the protein surface; the purple sphere represents the magnesium ion, and immediately below is the heme in licorice. Right panel: N-side view of the protein surface; capital letters indicate the entry mouth of the D, K and H channels.

Figure 5. Stereo view of the azide inhibited CcO. The strucure of the azide inhibited CcO 5ZCP, subunit A, is reported. Heme groups and residues D91, E242, K319 and H413 are in licorice; the magnesium ion and the Cu_B are reported as purple and silver spheres, respectively. The positions of water molecules belonging to short distance pairs are marked as orange spheres (all water pairs in the azide-inhibited structures are cumulatively reported).

Figure 6. O-O distance and excess proton in crystallographic water. The scheme depicts the proposed explanation for the SDWM pairs in high-resolution crystal structures. Hydrated protons live in the water network; depending on the relative residence time of a proton in a particular location of the ordered crystallographic water network, the O-O distance of the water molecules that share it, obtained by X-ray diffraction, varies from values practically indistinguishable from the hydrogen bond length to those typical of the Zundel ion.