Article

Surveying Enzyme Crystal Structures Reveals the Commonality of Active-Site Solvent Accessibility and Enzymatic Water Networks

Caleb M. T. Sindic, Pedro L. Muiño, and Patrik R. Callis*

Cite This: *ACS Omega* 2025, 10, 18419−18427

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Despite the demonstrable dependence of enzyme functionality on solvation, the notion of water being directly chemically required for catalysis inside active sites remains unexplored. Here we report that over 99% of 1013 enzyme crystals obtained by X-ray crystallography with high resolution (<1.5 Å) contain continuous chains of water linking residues within the active site to bulk water. Also reported are the findings which inspired this study—that electric fields experienced by water hydrogen atoms are on average twice as strong in the active sites of both chains of bacterial polynucleotide kinase (PDB 4QM6) structures compared to those in bulk water. These results point to the possibility that water molecules within active sites may be paramount to the immense catalytic power of enzymes, especially for mechanisms requiring hydronium or hydroxide ions.
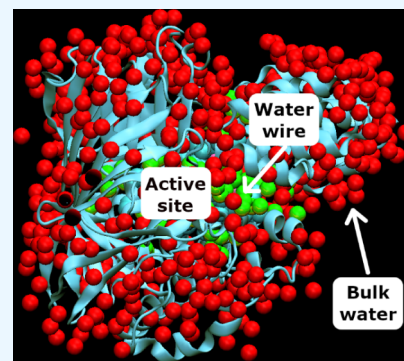
INTRODUCTION

The Grotthuss mechanism is a process wherein rapid proton and electron transfer across water is attributed to the reversible formation of unique structures known as water wires.[1] These water wires consist of chains of water molecules all within proton transfer distance of one another and all oriented in a manner that lowers the activation energy barrier gating proton transfer. Hydronium or hydroxide found on one one end of the wire can be readily transferred to the other end via rapid proton transfer occurring sequentially between each water pair.[2] Water wires have been observed to form in both bulk water and complex molecular environments.[3] In the context of biology, water wires were first observed to play critical roles in transmembrane proteins. They have been especially characterized in the "model membrane protein", the gramicidin A channel.[4,5] In one particularly interesting study, a redox-responsive water wire was calculated to function as a means to prevent the short-circuiting of the cytochrome c oxidase proton pump.[6] Many examples of water wires in non-transmembrane enzymes exist as well. One such example is the heme enzyme ascorbate peroxidase, where the involvement of an intermediary water molecule in a concerted proton transfer pathway from the enzyme's substrate to its heme-bound oxygen has been found to be critical for oxidation.[7] In Green Fluorescent Protein, negatively charged surface residues draw protons from the bulk solvent to be shuttled to the active site via a water wire.[8,9] In Ribonucleotide Reductase, evidence exists that the catalytically critical tyrosine radical is protonated by hydroxamic acid through a proton wire containing four water molecules.[10] In photolyase, QM/MM calculations have

demonstrated the assistance of a water wire to facilitate DNA damage repair through proton-coupled electron transport over >10 Å.[11] Alongside these individual findings, X-ray crystallographic (XRC) studies have demonstrated that 44% of side chain oxygen and nitrogen atoms in enzymes are hydrated.[12]

In addition, computational work by Warshel has provided evidence that enzymes are efficient catalysts because their structural rigidity and protection from bulk water maximizes the electrostatic interaction from active site charged residues, thereby lowering activation energy.[13] Water has been calculated via high-level *ab initio* computations to ionize under fields as low as 0.35 V/Å (6.8×10^{-3} a.u.).[14] The products of ionization, hydronium and hydroxide, are routinely proposed as participants in published mechanisms of enzyme action involving bond breaking and making because of the requirement of strong, rapid acting nucleophiles or electrophiles; there are occasional reports of their observation within active sites of enzymes.[15,16] The production of these species has been associated with a means of transporting charge across large distances within enzymes where reactive components are disparate.[17−20] In GTPase, for example, networks of water recruit charges from distant residues to enhance the nucleophilicity of an active site water molecule that then
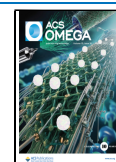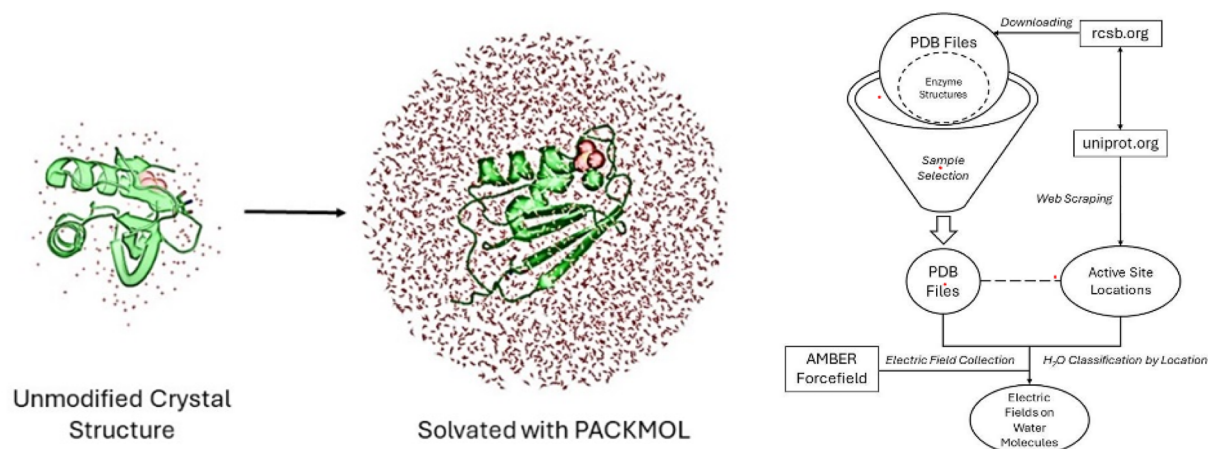
Figure 1. (Left): Graphic depictions of the solvation process of enzymes used in this work. Here the red dots are O atoms labeled as HOH in the raw PDB file. (Right): Schematic diagram depicting the procedure for solvation of enzyme PDB files by PACKMOL.
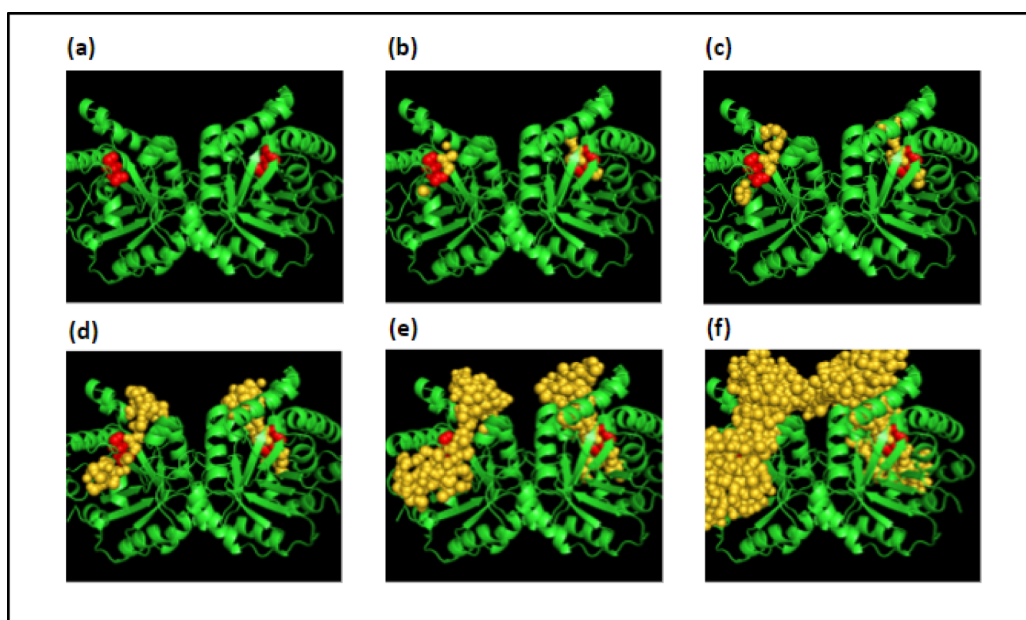
Figure 2. An illustrated example of elucidating a water wire for the active site of triose phosphate isomerase (XRC crystal structure resolution: 1.2 Å, PDB code: 1NEY).[31] The atoms of active site residues Glu 165 and His 95 can be seen as red spheres, the enzyme structure as a green cartoon depiction, and the water atoms belonging to the water wire as yellow spheres. (a) shows the structure before probing for water near the active site residues. (b) shows water molecules within 5 Å of active site atoms. (c), (d), (e), (f) show the recursive expansion of the water wire by adding water molecules whose oxygens are within 5 Å of the current water wire oxygen atoms. In (e) and (f), the water wire has reached the bulk solvent and can be seen extending far from the enzyme. By this point, the wire-generating algorithm would stop as the solvent has been contacted.

hydrolyzes guanosine triphosphate.[21,22] The Boxer group has reported findings connecting strong electric fields within the active site of the enzyme ketosteroid isomerase (KSI) along with evidence that the field is vital to its catalytic power.[23]

Preliminary unpublished work by our group on a bacterial polynucleotide kinase revealed that electric fields promoting proton transfer for active site water and a key ribose hydroxyl are exceptionally strong, ca. twice as strong as we computed for those in bulk water. From a crystal structure of the bacterial polynucleotide kinase (PDB 4QM6)[24] wherein the O2'-H2' bond of the riboguanosine (RG) nucleotide is H-bonded to its 3'phosphate in the C and D chains, we computed the strength of electric fields acting along this O2'-H2' bond and along water hydroxyl bonds in the active site. We therefore established a working hypothesis stating that water and other

hydroxyls will demonstrate an increased proclivity toward proton transfer when located in these enzyme active sites. These results are shown in Figures 3 and 4.

If water is chemically required for catalysis, an obvious expectation is that most enzyme crystal structures will reveal water within the active site and a mechanism, e.g., a water "wire", providing access to bulk water. This manuscript describes a procedure and the results from 1013 published XRC enzyme structures for which active site atoms are identified and for which we identified those that exhibit both water within their active sites and water wires linking the active site to bulk water. These are features that portend the existence of a continuous supply of water to the active site and have been shown to initiate the regeneration of the active site's charge state via the diffusion of charges in multiple enzymes.[25,26] The

number of water molecules in each water wire was then compared to that in bulk solvent to characterize the feasibility of autoionization within them.

■ MATERIALS AND METHODS

Identifying Water Active Sites. A dilemma was encountered when identifying active site residues using the process mentioned above. For many structures, namely those of the oxidoreductase class, the text file corresponding to the structure on uniprot.org did not possess residues denoted by the "active site" keyword but did possess residues denoted by the "binding site" keyword. We visually verified by inspection of the PDB files that these binding site residues were not members of allosteric binding sites but were in the active sites of the structures. In other cases, structures possessed both "active site" and "binding site" residues in their associated uniprot.org text files which were manually found near one another within the structures' active site pockets. For example, the residue GLU168 in glutaminyl cyclase (PDB ID 4MHN), responsible for coordinating substrate binding and denoted as a "binding site" residue on uniprot.org, is located immediately next to the catalytic residue GLU183, a proton acceptor in the active site, which is denoted as an "active site" residue on uniprot.org.[27] Given their proximity, both residues were considered to approximate the active site location within the structure. Accordingly, both residues deemed "binding site" and "active site" on uniprot.org were considered to identify the active site locations within each structure used in this study.

A schematic diagram of the procedure by which we solvated each enzyme is shown in Figure 1 right). In each PDB structure, water molecules whose oxygen was less than 5 Å from any of the active site atoms were identified and counted. The bulk solvent layer is based on the furthest crystal atom (including water) from the geometric center, not just enzyme atoms. These structures have crystal water on their surfaces. Thus, when there is a 5 Å padding for the placement of water via PACKMOL, it is from water on the surface of the enzyme in addition to enzyme atoms themselves. Thus, the padding extended about 5−7 Å from the furthest enzyme atom, if not more (due to the asymmetry of each structure), as shown in Figure 1 (left).

PACKMOL takes PDB files that have been solvated by PACKMOL, extracts active site locations from corresponding index files which contain active site line numbers in the original PDB file, writes the active site lines into their own PDB, then reads this PDB to obtain active site residues in the solvated file. Both the solvated file and the temporary active site PDB file are parsed with Bio.PDB.PDBParser, which then establishes water in the solvated PDB file as SOLVENT or NEAR ENZYME. It then forms a wire of water molecules within 5 Å of the active site atoms and propagates this wire, finding more and more water molecules within 5 Å of each other. It then asks how many solvent water molecules are within 5 Å of any component of the wire. This and more data are written in a file. An illustration of a typical example of the evolution of a water wire leading from an active site is presented for triose phosphate isomerase (PDB 1NEY) in Figure 2.

We did not test any of the structures for missing residues nor did we concern ourselves with irrelevant co-crystallized substances. It was our judgment that because of the rarity of these items compared to the large number of structures examined, their omission would not significantly influence our conclusions.

After characterizing the water wires within each PDB structure, we sought to assess the number of structures which possessed water wires terminating in bulk solvent space outside of the enzyme. Because the crystal structures largely lacked water outside of the enzyme crystal, we employed PACKMOL[28] to simulate bulk water solvation. To prepare PDB files for solvation, a spherical boundary with a padding of 5 Å from the most distant enzyme atom from the geometric center of the structure was constructed around each structure. Using the atomic radii of the atoms in the PDB file, assigned via the ChemLib[29] Python package, the volume occupied by the original structure was ascertained. To calculate the volume to be occupied by the simulated bulk solvent, the volume of the original structure (with no hydrogen added) was subtracted from the volume of the boundary sphere. The resulting volume was multiplied by the density of water at 310 K and 1 atm ($\sim$0.03 molecules/Å^3) to determine the number of water molecules to be added. The files were then solvated by PACKMOL using the TIP3P[30] water structure. In each of the structures, a new partition of water was defined: bulk solvent, defined by water greater than 5 Å from all enzyme atoms, ensuring it was outside of the enzyme. The structures possessing water wires having at least one water molecule with an r(O−O) < 5 Å of any of these bulk solvent molecules (bulk solvent contiguity) were then counted.

Water was not added to the interior of the crystal structure. The water wires reported in this work are composed of O atoms designated HOH in the PDB file.

The use of 5 Å as a cutoff distance for grouping water with respect to other water molecules and enzyme atoms was informed by previous calculations on the autoionization of water. Using a series of LEWIS[32] water recombination MD trajectories, it has been shown that the likelihood of recombination of hydronium and hydroxide is roughly equivalent to the likelihood of their drifting apart at a hydrogen bond distance of $\sim$6 Å.[33] Other QM/MM studies show that below this distance, recombination is highly favorable.[34,35] While the forward process of autoionization in pure water is difficult to simulate over the time scales permitted by QM/MM without biasing parametrization, it is reversible. Thus, the recombination distance is a proper metric for the distance at which water ion pairs can be expected immediately following autoionization. Accordingly, to ensure confidence in the distance between water and other bodies in the structures being sufficient for ionization, and to accommodate for the resolution of the structures (up to 1.5 Å), a 5 Å distance was chosen. More information on this value and its connotations is given in the discussion section below.

As a control, the number of water molecules near target (active site) residues, the number of water molecules in wires emanating from these, and the number of solvent molecules contacted by the wires were obtained for each structure using a random location in the enzyme as a comparative "active site". To this end, two random adjacent residues in each structure were selected to obtain a target point of reasonable size and subjected to the same water topology analysis conducted for the active site. Because the proclivity for proton transfer in a homogeneous liquid is partially dependent on the distance between molecules, the densities of the water wire and bulk solvent were recorded. The chosen metric of density for each partition was the distance between each water oxygen and its closest neighboring water oxygen, calculated using the scikit-

learn NearestNeighbors[36] method for Python. The Python codes and documentation for all methods are available online[1]

Exploratory Computation of Electric Fields on Active Site Hydroxyls.

To calculate the electric fields on active site
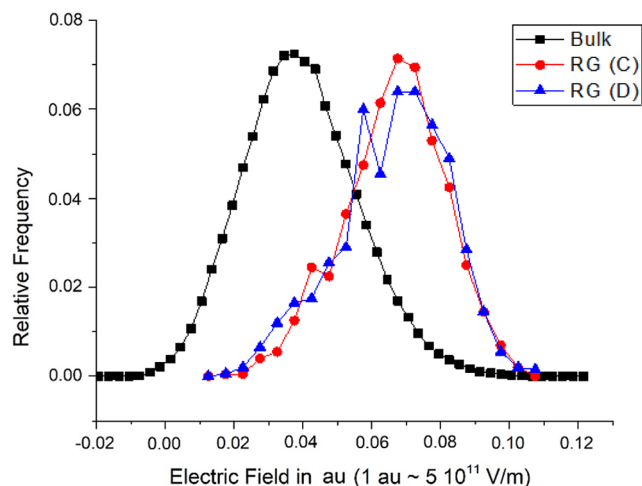
Figure 3. A histogram contrasting the exceptionally high electric fields felt by the O2'−H2' bond of the RG nucleotide in the active sites of the C and D chains of bacterial polynucleotide kinase PDB 4QM6[24] to the fields on liquid water.
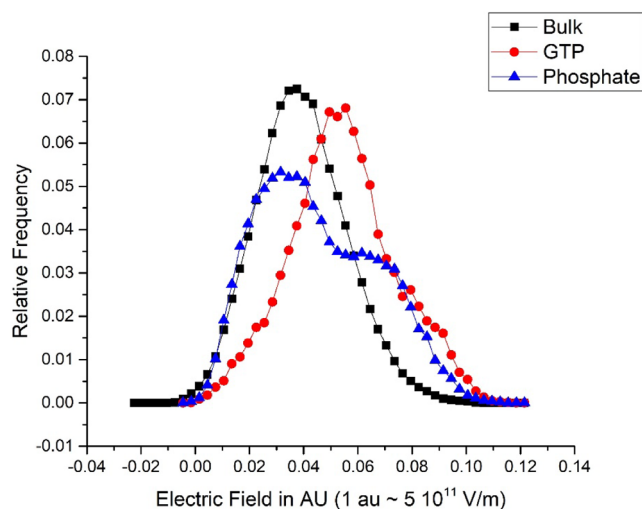
Figure 4. A histogram of the exceptionally high electric fields felt by *water* molecules near the active site bound GTP (red circles) and phosphate (blue triangles) in the bacterial polynucleotide kinase PDB 4QM6,[24] and in bulk water (black squares). The presence of Mg2+ ions likely contributes to the strong fields.

O−H bonds that generated Figures 3 and 4, the following procedure was used:

1 For each enzyme PDB file, including substrate, if used, cofactors, and water of crystallization were used as input for molecular dynamics (MD) using GROMACS (AMBER99SB-ILDN force field) including explicit SPC/E charges and, if necessary, parametrization of the substrate.

2 Missing hydrogen atoms were added to the PDB file by invoking the command *pdb2gmx* within GROMACS, which recognizes the missing hydrogen atoms and allows

the user to specify the protonation state of acidic or basic amino acids. The added hydrogen atoms are then placed according to standard bond lengths, angles, and dihedral angles defined within the force field parameters.

3 Because PDB files do not include atom charges, atom charges were obtained by requesting the MD to include an "itp file" as output, which does list the individual atom charges.

4 When buffer ions are included in the PDB file, they are assigned charges corresponding to the SPC/E parametrization.

5 The resulting structure was energy minimized and equilibrated for 200 ps to reach thermal equilibrium at 300 K.

6 Dynamics were then run for 10 ns, with coordinates extracted every 10 ps.

7 A homemade Fortran code was used to calculate the projection of the electrostatic field onto each of the O−H bonds of bulk water and those in contact with the active site atoms.

Below, we paraphrase the Fortran code written by the authors used to compute the electric field projections along the OH bonds of water that are plotted in Figures 3 and 4.

1 The GROMACS MD output contains the charge of each atom contained in the PDB file. The primary output file from GROMACS contains the Cartesian coordinates of each atom at a given time. From the coordinates and charges, the x, y, z components of the field are computed for each water molecule within 4.0 Å of any active site atom.

2 The output file of the MD trajectory is read line by line, and those lines containing the required information, namely the ith atom indx(i), atom(i), amino(i), chain(i), namino(i), Cartesian coordinates x(i),y(i),z(i), and atom charge, q(i) are stored in their respective arrays.

3 The electric field on the H atoms of any water molecule within 4 Å of any active site atom is computed from the charge and Cartesian coordinates of all protein, water, and counterion atoms in the simulation—except the field coming from the O atom to which the proton is bonded, which is already counted as a primary force involved in establishing the equilibrium O−H bond length.

The Coulomb's Law equation to compute the *x component* of the electric field on atom k due to all other atoms kk is given by $\Sigma_{kk}((x(kk)-x(k))q(kk)/r_1^3)$, where r_1 is the distance between atom kk and atom k.

■ RESULTS AND DISCUSSION

Figures 3 and 4 displayed below are histograms exhibiting average electric fields projected along the OH bonds, i.e., hydroxyl groups, in the active site of the enzyme *Clostridium thermocellum* polynucleotide kinase, an important enzyme in the repair of broken strands of DNA and RNA. In Figure 3 the O−H is attached to the 2' position of the ribose sugar ring of the terminal end of an RNA fragment for the two chains in the crystal structure. In Figure 4, the electric fields are on the H atoms of water found in the active site

Water Wires Connected to Active Sites. Of the 1013 enzyme structures studied, 1005 were found to have water within 5 Å of active site atoms. This finding establishes a structural precedence for solvent-assisted catalysis in enzymes

because the majority contain water within their active site pocket. When water wires were generated from water molecules in the active sites of each structure, most were found to host a large network of water molecules contiguous with their active sites. Across the water wires of the 1013 enzyme structures studied before the addition of solvent molecules, there was an average of 333 water molecules per wire when seeded from active sites and 242 water molecules when seeded from two random adjacent residues (Table 1 and

Table 1. Descriptive Statistics for Water Wire Sizes (Count of Water Molecules/Wire) and the Number of Structures Having a Water Wire That Contacts the Bulk Solvent for Each Class of Enzyme

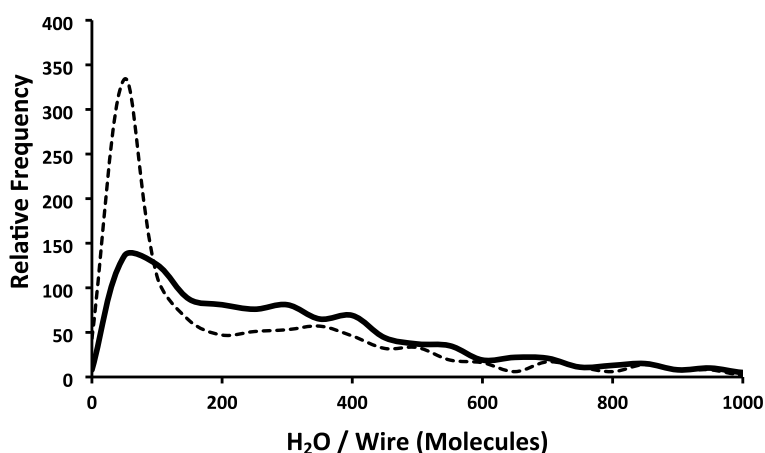|	Water Molecules per Wire (Mean)	Water Molecules per Wire (Std. Dev.)	Number of Structures with Bulk Solvent Contacted	Number of Structures in Class
Hydrolases	330	353	470	482
Isomerases	300	224	45	47
ligases	458	370	20	20
lyases	457	465	57	58
oxidoreductases	444	487	151	162
transferases	248	243	216	219
translocases	87	218	24	25
all structures	333	368	993	1013

Figure 5). Note that the areas under both phosphate curves in Figure 4 are equal because there is only one active site contiguous water cluster, and one random site contiguous water cluster assigned to each enzyme structure.

After the solvation of the structures using PACKMOL, contiguity between the active site and bulk solvent via water wires was observed in 993 of the 1013 structures studied. This feature spatially permits a supply of water to the active site and suggests that a Grotthuss-type mechanism of charge transport between water near the active sites and bulk solvent within these structures could be involved. The mean number of bulk solvent water molecules within the parametrized $r(O-O) < 5$ Å proton transfer distance of any member of each water wire

was 1482, with a standard deviation of 740. This value was drastically increased by large structures whose water wires extended into grooves on the surface of the enzyme, with a large contact area for interaction with the bulk solvent. Irrespective of this subgroup, there are many avenues for water to infiltrate the active sites of enzymes. Likewise, in the instance of an ionizing event on the enzyme surface, there are many water molecules capable of receiving or initiating ionizing events to and from the active site of most enzymes via a Grotthuss-type mechanism.

In 15 structures, water was present within 5 Å of active site atoms but these waters did not form wires that reached the bulk solvent. The average number of water molecules in the water wires found in these structures was 24, significantly lower than the average value of 333 water molecules/wire found for all structures studied. Unless these enzymes were folded with their active sites entombed from the space around the protein, which would hinder substrate access, this finding is likely due to constrictions in the buried tunnels containing the wires that bifurcate them and subsequently prevent contact with bulk solvent. If this is the case, it would be expected that the enlargement of these constrictions during routine molecular motion would elicit a larger water wire size and the contiguity of the wire with bulk solvent.

When two random adjacent residues were selected in each structure as targets in lieu of active site residues, the number of structures that did not exhibit water molecules within 5 Å of target residues was significantly less than when active site residues were selected (48 and 8 structures, respectively). The number of structures that did not have target residue-adjoining water wires that reached the bulk solvent was 94 and 20 for structures after targeting random adjacent residues and true active site residues, respectively. Furthermore, smaller water wires generated from random adjacent residues were much more frequent than those generated from true active site residues (Figure 5). These findings implicate the unique solvent accessibility of the active site of enzymes relative to other locations. This is as expected given the requirements of the active sites to be accessed by substrates through the bulk

Figure 5. Histograms of water molecules per water wire originating from the active site (solid black line) and two random adjacent residues (dashed line) in each of 1013 crystal structures. Each water wire was grouped by the criterion of $r(O-O) < 5$ Å. Relative frequencies were calculated as the number of structures with their water wire in the range of each bin, divided by the total number of structures. These are smoothed histograms with bin size 50.
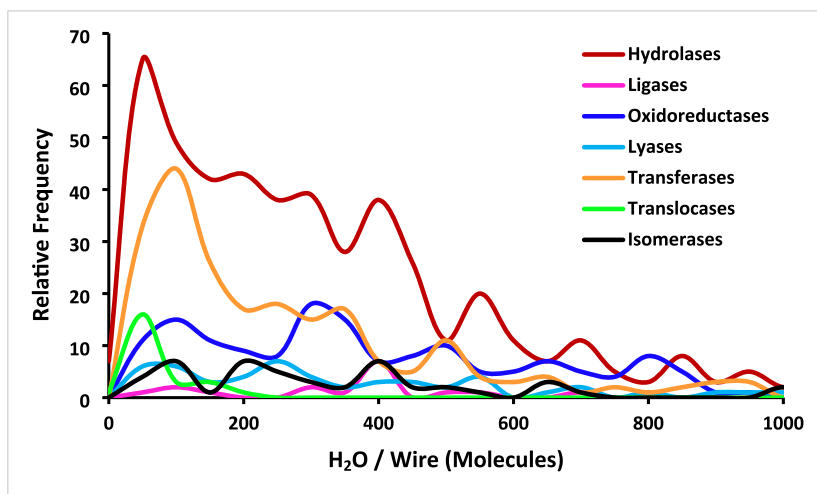
Figure 6. By the basis of enzyme class, histograms of water molecules per wire originating from the active site among 1013 enzyme crystal structures. Each water wire was grouped by the criterion of r(O−O) < 5 Å and required to have one member whose oxygen was this distance from any active site atom. Relative frequencies were calculated as the number of structures with their water wire in the range of each bin, divided by the total number of structures. These are smoothed histograms with bin size 50.

solvent, but it also demonstrates the enhanced ability for active sites to be contacted by the solvent itself.

All enzymes studied belonged to one of seven different classes: hydrolases, oxidoreductases, isomerases, lyases, ligases, transferases, and translocases. The hydrolases composed the largest class in the study (Table 1). The distributions of water wire sizes based on enzyme class show similar features, with a large portion of enzymes falling under a curve at approximately 100−500 water molecules per wire (Figure 6).

Of the eight structures that showed no water near their active site, seven belonged to the hydrolase class and one belonged to the translocase class. These structures were *Klebsiella pneumoniae* carbapenemase 2 beta-lactamase,[37] PI-SceI,[38] fluoroacetate dehalogenase,[39] human "a disintegrin and metalloproteinase with thrombospondin motifs",[40] phage T4 lysozyme,[41] V84D mutant of S. solfataricus acylphosphatase,[42] bovine trypsin,[43] and 2-hydroxy-dATP diphosphatase.[44] A finding of water within 5 Å of an active site is not a completely reliable criterion for identifying water as important for enzyme action. For example, examination of the T4 lysozyme structure[41] reveals that the substrate binds in a cleft located between alpha helices such that the active site may be delocalized among groups that are already normally *in contact with bulk water*, thereby not conforming to the notion of lengthy water wires in the same manner as seen for enzymes with more buried active sites.

Comparing columns 1 and 2 in Table 1 above shows that the mean number of water molecules per wire closely follows the standard deviation in magnitude, except for the translocases, for which the mean molecules (of water) per wire is ∼2.5 times smaller than the standard deviation. The reason for this is almost certainly because the class of translocases uniquely *does not catalyze chemical reactions*. This class uniquely has the purpose only of moving substrates across membranes, the free energy for which comes from coupling with ATP hydrolysis. An obvious explanation for short wires in translocases is that no chemical reactions are required, except for the hydrolysis of ATP, which is bound on the surface in bulk water already.

The average water hydrogen−oxygen bond distance found in the literature is 0.96 Å.[45] Assuming water may rotate to minimize the proximity of the positively charged hydrogen

atoms to the nearest neighboring water oxygen, this distance, r(O−H*), may be within 4.04 Å of the nearest water oxygen at the maximum allowed r(O−O) of 5 Å for waters in these wires. Within the wires, the average r(O−O) between a water and its nearest neighbor, r(O−O)$_{min}$, was 2.85 Å (Figure 7).
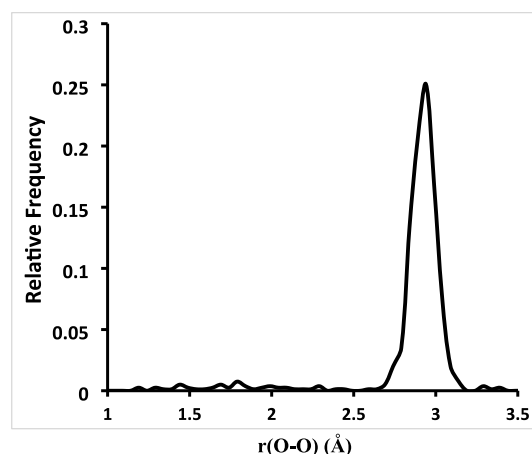
Figure 7. Oxygen−oxygen distances between each water molecule and its nearest neighboring water molecule within the active site-contiguous water wires of 1013 enzyme structures. Relative frequencies were calculated as the number of structures with their water wire in the range of each bin, divided by the total number of structures. These are smoothed histograms with bin size 50.

Provided again that a hydrogen atom may rotate around its central oxygen atom in water, this value corresponds with a possible range of 1.89 < r(O−H*) < 3.81 Å on average. These distances are well within the r(O−H*) values observed in liquid water, where hydrogen bonding is frequent and proton transfers occur on the order of 10^{-5} s^{-1} (about once per 10 h).[46] Moreover, the r(O−O) < 5 Å parameter in the water wires is less than the calculated 6 Å separation distance of hydronium and hydroxide in liquid water immediately after ionization.[46] In nitrogenase, which is responsible for the hydrogenation of N$_2$ to produce NH$_3$, a catalytically critical hydronium-shuttling pathway has been predicted via density

functional theory with water−water distances fluctuating between 2.45 and 3.5 Å and hydrogen bonding characterized by r(O−O) ≈ 2.8 Å.[47] This correlates well with our average r(O−O)$_{min}$ = 2.85 Å found in water wires. Based on distance, water molecules in these networks are sufficient for hydrogen bonding and charge transfer via the Grotthuss mechanism.[46,48]

Finally, we note a closely related work by Pravda et al.[49] who have conducted an extensive investigation of the connection of the active site to the bulk water in 4,823 enzymes whose structures were available in the Catalytic Site Atlas[50] at a resolution better than 2.5 Å. They reported that approximately 64% of those enzymes have at least one channel (and an average of two) of over 15 Å in length leading to the bulk. The scope of that work was to evaluate the accessibility of the substrates to active sites buried deep within the enzyme. They confirmed that the relative abundances of the amino acids lining the walls of the channels deviate from abundances in the bulk of the enzymes, presumably to generate the requisite physicochemical properties (charge and hydrophilicity, for instance) to facilitate rapid progress of the substrates toward the catalytic site. The action of these channels is radically different from the one presented here, substrate accessibility vs water access as a catalytic inducer, but both address the point that communication between the active site and the bulk is a prerequisite for enzyme action.

CONCLUSION

The results presented above point to the possibility that all enzymes require water as a necessary reactant. Water wires emanating from the active sites of the enzymes, clustered by a ≤5 Å oxygen−oxygen distance, were found in 99% of the 1013 X-ray crystal structures surveyed, with an average of 333 water molecules per enzyme residing in water wires. This large value appears from our data to be attributable to water wires that connect water in active sites to water at the surface (bulk water). Because each water molecule in these wires was inherently selected by the condition of it being within a certain distance of its neighbors suitable for autoionization, the wires found here are capable of propagating hydronium and/or hydroxide ions to the active site, in addition to the finding that electric fields within the active sites are computed to be higher on average than in bulk water. Given that each water molecule in a system has a probability of initiating a Grotthuss-type charge transfer reaction, the size of these water wires demonstrates a vast number of locations from which a hydronium or hydroxide ion may be generated and conducted via a Grotthuss-type mechanism to the active site of the enzyme, depending on the local electric fields. Accordingly, most enzymes seem to structurally permit water exchange between the active site and bulk solvent.

Because enzymes are plastic and typically captured in only one conformation by XRC, sampling numerous conformations of each structure in this study would permit a better understanding of water wire dynamics. One computationally intensive but powerful means to this end would be an extension of our analysis using temporal frames extracted from classical molecular dynamics trajectories created from the PDB structures of the enzymes studied. Such an approach would also allow for the visualization of the movement of water within the water wires found in this study if random fluctuations of the protein provide adequate chemical potential differences for water along the wires. Another viable method for confirming our findings is CAVER,[51] a computational tool used to determine whether a given molecule can access a selected point in an enzyme from the bulk solvent.

AUTHOR INFORMATION

Corresponding Author

Patrik R. Callis − *Montana State University, Bozeman, Montana 59717, United States;* orcid.org/0000-0002-6050-4206; Email: pcallis@montana.edu

Authors

Caleb M. T. Sindic − *Montana State University, Bozeman, Montana 59717, United States;* Present Address: USDA Dairy Forage Research Center 1925 Linden Drive, Madison, Wisconsin 53706, United States

Pedro L. Muiño − *Department of Chemistry, Saint Francis University, Loretto, Pennsylvania 15940, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.4c10721

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ADDITIONAL NOTE

[1]All Python codes used for analysis are available at https://github.com/cmsindic/Water_Wire_Connectivity.

REFERENCES

(1) Agmon, N. The grotthuss mechanism. *Chem. Phys. Lett.* **1995**, *244* (5−6), 456−462.

(2) Marx, D. Proton transfer 200 years after von Grotthuss: Insights from ab initio simulations. *ChemPhyschem* **2006**, *7* (9), 1848−1870.

(3) Brewer, M. L.; Schmitt, U. W.; Voth, G. A. The formation and dynamics of proton wires in channel environments. *Biophys. J.* **2001**, *80* (4), 1691−1702.

(4) Kelkar, D. A.; Chattopadhyay, A. The gramicidin ion channel: A model membrane protein. *Biochim. Biophys. Acta, Biomembr.* **2007**, *1768* (9), 2011−2025.

(5) Pomes, R.; Roux, B. Structure and dynamics of a proton wire: A theoretical study of H+ translocation along the single-file water chain in the gramicidin A channel. *Biophys. J.* **1996**, *71* (1), 19−39.

(6) Sharma, V.; Enkavi, G.; Vattulainen, I.; Róg, T.; Wikström, M. Proton-coupled electron transfer and the role of water molecules in proton pumping by cytochrome c oxidase. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (7), 2040−2045.

(7) Efimov, I.; Badyal, S. K.; Metcalfe, C. L.; Macdonald, I.; Gumiero, A.; Raven, E. L.; Moody, P. C. Proton delivery to ferryl heme in a heme peroxidase: Enzymatic use of the Grotthuss mechanism. *J. Am. Chem. Soc.* **2011**, *133* (39), 15376−15383.

(8) Shinobu, A.; Palm, G. J.; Schierbeek, A. J.; Agmon, N. Visualizing proton antenna in a high-resolution green fluorescent protein structure. *J. Am. Chem. Soc.* **2010**, *132* (32), 11093−11102.

(9) Shinobu, A.; Agmon, N. Proton wire dynamics in the green fluorescent protein. *J. Chem. Theory Comput.* **2017**, *13* (1), 353−369.

(10) Offenbacher, A. R.; Barry, B. A. A proton wire mediates proton coupled electron transfer from hydroxyurea and other hydroxamic acids to tyrosyl radical in class Ia ribonucleotide reductase. *J. Phys. Chem. B* **2020**, *124* (2), 345−354.

(11) Wang, H.; Chen, X.; Fang, W. Excited-state proton coupled electron transfer between photolyase and the damaged DNA through water wire: A photo-repair mechanism. *Phys. Chem. Chem. Phys.* **2014**, *16* (46), 25432−25441.

(12) Baker, E. N.; Hubbard, R. E. Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* **1984**, *44* (2), 97−179.

(13) Warshel, A. Energetics of enzyme catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **1978**, *75* (11), 5250−5254.

(14) Saitta, A. M.; Saija, F.; Giaquinta, P. V. Ab initio molecular dynamics study of dissociation of water under an electric field. *Phys. Rev. Lett.* **2012**, *108* (20), 207801.

(15) Kovalevsky, A. Y.; Hanson, B.; Mason, S.; Yoshida, T.; Fisher, S.; Mustyakimov, M.; Forsyth, V.; Blakeley, M.; Keen, D.; Langan, P. Identification of the elusive hydronium ion exchanging roles with a proton in an enzyme at lower pH values. *Angew. Chem., Int. Ed.* **2011**, *123* (33), 7662−7665.

(16) Ikeda, T.; Saito, K.; Hasegawa, R.; Ishikita, H. The existence of an isolated hydronium ion in the interior of proteins. *Angew. Chem., Int. Ed.* **2017**, *56* (31), 9151−9154.

(17) Li, T.-R.; Huck, F.; Piccini, G.; Tiefenbacher, K. Mimicry of the proton wire mechanism of enzymes inside a supramolecular capsule enables β-selective O-glycosylations. *Nat. Chem.* **2022**, *14* (9), 985−994.

(18) Frank, R. A.; Titman, C. M.; Pratap, J. V.; Luisi, B. F.; Perham, R. N. A molecular switch and proton wire synchronize the active sites in thiamine enzymes. *Science* **2004**, *306* (5697), 872−876.

(19) Dai, S.; Funk, L.-M.; von Pappenheim, F. R.; Sautner, V.; Paulikat, M.; Schröder, B.; Uranga, J.; Mata, R. A.; Tittmann, K. Low-barrier hydrogen bonds in enzyme cooperativity. *Nature* **2019**, *573* (7775), 609−613.

(20) Shinobu, A.; Agmon, N. Mapping proton wires in proteins: Carbonic anhydrase and GFP chromophore biosynthesis. *J. Phys. Chem. A* **2009**, *113* (26), 7253−7266.

(21) Tripathi, R.; Noetzel, J.; Marx, D. Exposing catalytic versatility of GTPases: Taking reaction detours in mutants of hGBP1 enzyme without additional energetic cost. *Phys. Chem. Chem. Phys.* **2019**, *21* (2), 859−867.

(22) Tripathi, R.; Glaves, R.; Marx, D. The GTPase hGBP1 converts GTP to GMP in two steps via proton shuttle mechanisms. *Chem. Sci.* **2017**, *8* (1), 371−380.

(23) Fried, S. D.; Boxer, S. G. Electric fields and enzyme catalysis. *Annu. Rev. Biochem.* **2017**, *86*, 387−415.

(24) Das, U.; Wang, L. K.; Smith, P.; Munir, A.; Shuman, S. Structures of Bacterial Polynucleotide Kinase in a Michaelis Complex with Nucleoside Triphosphate (NTP)-Mg^{2+} and 5′-OH RNA and a Mixed Substrate-Product Complex with NTP-Mg^{2+} and a 5′-Phosphorylated Oligonucleotide. *J. Bacteriol.* **2014**, *196* (24), 4285−4292.

(25) Cui, Q.; Karplus, M. Is a "proton wire" concerted or stepwise? A model study of proton transfer in carbonic anhydrase. *J. Phys. Chem. B* **2003**, *107* (4), 1071−1078.

(26) Riccardi, D.; König, P.; Prat-Resina, X.; Yu, H.; Elstner, M.; Frauenheim, T.; Cui, Q. "Proton holes" in long-range proton transfer reactions in solution and enzymes: A theoretical analysis. *J. Am. Chem. Soc.* **2006**, *128* (50), 16302−16311.

(27) Huang, K.-F.; Hsu, H.-L.; Karim, S.; Wang, A.-J. Structural and functional analyses of a glutaminyl cyclase from Ixodes scapularis reveal metal-independent catalysis and inhibitor binding. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2014**, *70* (3), 789−801.

(28) Martínez, L.; Andrade, R.; Birgin, E.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157−2164.

(29) Ambethkar, H.; *Chemlib*. 2020. https://github.com/harirakul/chemlib.

(30) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926−935.

(31) Jogl, G.; Rozovsky, S.; McDermott, A. E.; Tong, L. Optimal alignment for enzymatic proton transfer: Structure of the Michaelis complex of triosephosphate isomerase at 1.2-Å resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (1), 50−55.

(32) Kale, S.; Herzfeld, J.; Dai, S.; Blank, M. Lewis-inspired representation of dissociable water in clusters and Grotthuss chains. *J. Biol. Phys.* **2012**, *38*, 49−59.

(33) Bai, C.; Herzfeld, J. Special pairs are decisive in the autoionization and recombination of water. *J. Phys. Chem. B* **2017**, *121* (16), 4213−4219.

(34) Natzle, W. C.; Moore, C. B. Recombination of hydrogen ion (H+) and hydroxide in pure liquid water. *J. Phys. Chem.* **1985**, *89* (12), 2605−2612.

(35) Hassanali, A.; Prakash, M. K.; Eshet, H.; Parrinello, M. On the recombination of hydronium and hydroxide ions in water. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (51), 20410−20415.

(36) Scikit-Learn.org. *Nearest Neighbors — scikit-learn 0.21.3 documentation*, Scikit-Learn.org.

(37) Papp-Wallace, K. M.; Nguyen, N. Q.; Jacobs, M. R.; Bethel, C. R.; Barnes, M. D.; Kumar, V.; Bajaksouzian, S.; Rudin, S. D.; Rather, P. N.; Bhavsar, S.; et al. Strategic approaches to overcome resistance against Gram-negative pathogens using β-lactamase inhibitors and β-lactam enhancers: Activity of three novel diazabicyclooctanes WCK 5153, zidebactam (WCK 5107), and WCK 4234. *J. Med. Chem.* **2018**, *61* (9), 4067−4086.

(38) Duan, X.; Gimble, F. S.; Quiocho, F. A. Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. *Cell* **1997**, *89* (4), 555−564.

(39) Chan, P. W.; Yakunin, A. F.; Edwards, E. A.; Pai, E. F. Mapping the reaction coordinates of enzymatic defluorination. *J. Am. Chem. Soc.* **2011**, *133* (19), 7461−7468.

(40) Durham, T. B.; Klimkowski, V. J.; Rito, C. J.; Marimuthu, J.; Toth, J. L.; Liu, C.; Durbin, J. D.; Stout, S. L.; Adams, L.; Swearingen, C.; et al. Identification of potent and selective hydantoin inhibitors of aggrecanase-1 and aggrecanase-2 that are efficacious in both chemical and surgical models of osteoarthritis. *J. Med. Chem.* **2014**, *57* (24), 10476−10485.

(41) Liu, L.; Quillin, M. L.; Matthews, B. W. Use of experimental crystallographic phases to examine the hydration of polar and nonpolar cavities in T4 lysozyme. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (38), 14406−14411.

(42) de Rosa, M.; Bemporad, F.; Pellegrino, S.; Chiti, F.; Bolognesi, M.; Ricagno, S. Edge strand engineering prevents native-like aggregation in S ulfolobus solfataricus acylphosphatase. *FEBS J.* **2014**, *281* (18), 4072−4084.

(43) Krzywda, S.; Jaskolski, M.; Rolka, K.; Stawikowski, M. J. Structure of a proteolytically resistant analogue of (NLys) 5SFTI-1 in complex with trypsin: Evidence for the direct participation of the Ser214 carbonyl group in serine protease-mediated proteolysis. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2014**, *70* (3), 668−675.

(44) Kettle, J. G.; Alwan, H.; Bista, M.; Breed, J.; Davies, N. L.; Eckersley, K.; Fillery, S.; Foote, K. M.; Goodwin, L.; Jones, D. R.; et al. Potent and selective inhibitors of MTH1 probe its role in cancer cell survival. *J. Med. Chem.* **2016**, *59* (6), 2346−2361.

(45) Bowen, H. J. M.; *Tables of interatomic distances and configuration in molecules and ions.* Sutton, L. E. Ed.; Chemical Society. 1958, *11*.

(46) Geissler, P. L.; Dellago, C.; Chandler, D.; Hutter, J.; Parrinello, M. Autoionization in liquid water. *Science* **2001**, *291* (5511), 2121−2124.

(47) Dance, I. The pathway for serial proton supply to the active site of nitrogenase: Enhanced density functional modeling of the Grotthuss mechanism. *Dalton Trans.* **2015**, *44* (41), 18167−18186.

(48) Liu, J.; He, X.; Zhang, J. Z.; Qi, L.-W. Hydrogen-bond structure dynamics in bulk water: Insights from ab initio simulations with coupled cluster theory. *Chem. Sci.* **2018**, *9* (8), 2065−2073.

(49) Pravda, L.; Berka, K.; Svobodová Vařeková, R.; Sehnal, D.; Banáš, P.; Laskowski, R. A.; Koča, J.; Otyepka, M. Anatomy of enzyme channels. *BMC Bioinf.* **2014**, *15* (1), 379.

(50) Ribeiro, A. J. M.; Holliday, G. L.; Furnham, N.; Tyzack, J. D.; Ferris, K.; Thornton, J. M. Mechanism and Catalytic Site Atlas (M-CSA): A database of enzyme reaction mechanisms and active sites. *Nucleic Acids Res.* **2018**, *46*, D618−D623.

(51) Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; et al. CAVER 3.0: A tool for the analysis of transport pathways in dynamic protein structures. *PloS Comput. Biol.* **2012**, *8*, No. e1002708.