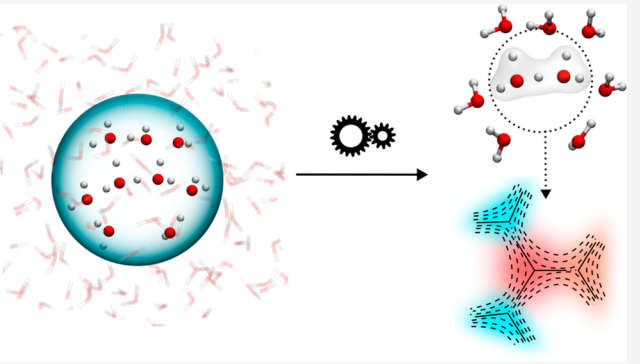
ZundEig: The Structure of the Proton in Liquid Water from Unsupervised Learning

Solana Di Pino, Edward Danquah Donkor, Veronica M. Sánchez, Alex Rodriguez, Giuseppe Cassone, Damian Scherlis, and Ali Hassanali*

ABSTRACT: The structure of the excess proton in liquid water has been the subject of lively debate on both experimental and theoretical fronts for the last century. Fluctuations of the proton are typically interpreted in terms of limiting states referred to as the Eigen and Zundel species. Here, we put these ideas under the microscope, taking advantage of recent advances in unsupervised learning that use local atomic descriptors to characterize environments of acidic water combined with advanced clustering techniques. Our agnostic approach leads to the observation of only one charged cluster and two neutral ones. We demonstrate that the charged cluster involving the excess proton is best seen as an ionic topological defect in water's hydrogen bond network, forming a single local minimum on the global free-energy landscape. This charged defect is a highly fluxional moiety, where the idealized Eigen and Zundel species are neither limiting configurations nor distinct thermodynamic states. Instead, the ionic defect enhances the presence of neutral water defects through strong interactions with the network. We dub the combination of the charged and neutral defect clusters as *ZundEig*, demonstrating that the fluctuations between these local environments provide a general framework for rationalizing more descriptive notions of the proton in the existing literature.



INTRODUCTION

Proton transfer (PT) reactions are at the heart of acid−base chemistry processes relevant for a wide range of phenomena in physical, chemical, and biological systems.[1−13] The thermodynamics associated with PT are central to determining pH and the protonation states of organic molecules in solution. The mechanisms by which protons move through hydrogen-bonded networks (HBNs), such as water, have attracted the attention of experimentalists and theoreticians alike.[10,14] Unlike other ions which diffuse in water *via* hydrodynamic diffusion, protons migrate within the HBN through the interconversion of covalent and hydrogen (H-)bonds, commonly referred to as the Grotthuss mechanism.[15,16]

Textbook chemistry teaches us that the proton in liquid water does not exist as an isolated entity as it does in the gas phase but instead electronically interacts with water.[17] Since the mid 20th century, the structure of the proton in liquid water has been discussed in terms of two apparently limiting structures, namely, Eigen and Zundel.[18,19] Eigen proposed that the excess proton is localized on a single water molecule, forming the $(H_3O)^+$ moiety. This species is bound *via* tight H-bonds to three surrounding water molecules, resulting in the $(H_5O_9)^+$ complex.[20] On the other hand, Zundel presented a more delocalized form of the proton, suggesting instead that it is more evenly shared between two flanking water molecules forming the $(H_5O_2)^+$ structure.[19] This feature was invoked to rationalize the broad absorption in the IR region between 2000 and 3000 cm^{-1}, commonly referred to as the Zundel continuum.[21,22] Since the mid-20th century, these two assignments of the proton structure have fueled intense and lively debates regarding their relative importance in the Grotthuss mechanism.[23−27]

In a seminal paper in 1995, Agmon showed, with very elegant back-of-the-envelope type calculations, that proton diffusion in water involves a rather low activation barrier of approximately 2−3 kcal/mol and that this is essentially coupled to changes in the local solvation of the excess

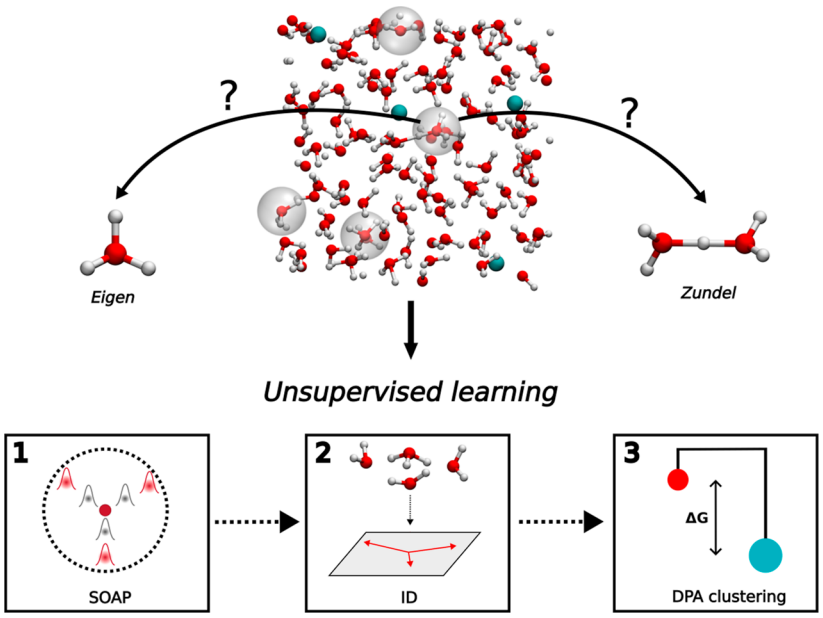


Figure 1. Schematic protocol for the unsupervised learning that we perform on protonated water with the aim of determining whether the dominant species is an Eigen or Zundel. Trajectories obtained from AIMD simulations of protonated water are used to build local atomic descriptors (SOAP) (1) from which we extract the intrinsic dimension (ID) of the local solvation environment (2). This then serves as input for performing high-dimensional clustering and the extraction of free energy (3).

proton,[16] challenging previous ideas of a more collective phenomenon.[28] A lot of our modern understanding of the proton structure in water has emerged from several decades of atomistic molecular dynamics simulations using both first-principles density functional theory (DFT)[29−35] as well as reactive potentials such as empirical valence bond.[36−41] Quantum chemistry-type calculations in the gas phase and in cluster environments, which typically ignore thermal and entropic effects, have also played an important role in understanding the structure of the proton and how it is affected by hydration water.[42−49]

Different conclusions have been reached regarding the dominant structure of the proton, including, besides the canonical Eigen and Zundel complexes, descriptions such as the *distorted Eigen cation*[50] and the *distorted Zundel structure*.[51−53] These so-called limiting states also form part of the inherent language used to rationalize the Grotthuss mechanism, where proton diffusion is described in terms of the interconversion between Eigen and Zundel.[54,55] Since the identity of the proton changes as it migrates through the HBN, various types of chemically inspired criteria need to be used to identify where the proton is instantaneously localized.[33,34,38] Within this context, chemical bias requires some form of dimensionality reduction, which can lead to uncontrolled information loss when examining fluctuations of the system in a reduced space of coordinates. In the past decade, there has been a tremendous spurt in the use of data-science techniques to study complex molecular systems.[56] These approaches offer a framework in which patterns in the data emerging from the atomistic models are used to derive structural hierarchies and subsequently inform the underlying physics and chemistry.[57]

We have recently employed a battery of state-of-the-art data-driven techniques to study the nature of the fluctuations in liquid water at room temperature.[58] Therein, using the smooth overlap of atomic descriptors (SOAP) to characterize local environments in liquid water and subsequently dimensionality reduction and high-dimensional clustering, we showed that liquid water at room temperature is characterized by a single free-energy minimum that is effectively flat and broad, with no evidence supporting the 2-state models of liquid water. In the current work, we extend this protocol to the study of the free-energy landscape of acidic water using previous *ab initio* molecular dynamics (AIMD) simulations of HCl at various concentrations reported by Markland and co-workers.[59] The unsupervised clustering of the high-dimensional SOAP space of HCl yields three statistically dominant clusters, of which only one corresponds to the excess proton. This charged cluster contains both Eigen and Zundel moieties, and these idealized structures do not emerge as limiting thermodynamic states. We demonstrate that the excess proton is best seen as a charged topological defect that enhances the concentration of neutral defects close to it, which together can be chemically interpreted as a mix of the Eigen and Zundel species, which we refer to as *ZundEig*.

METHODS

Our protocol for performing unsupervised learning of the molecular environments of protonated water follows closely our recent work on studying the free-energy landscape of liquid water at room temperature.[58] Steps involved in our general framework for performing the unsupervised learning approach are summarized in Figure 1. First, we build up the SOAP descriptors to characterize local environments.[60] Subsequently, the SOAP descriptors serve as input for computing the intrinsic dimension (ID) using the Two-NN method.[61] Finally, this ID is used to perform high-dimensional free-energy construction.[62,63] A complementary dimensionality reduction technique using UMAP[64] provides a useful visualization of the results obtained with this framework.

***Ab Initio* Molecular Dynamics.** In this work, we analyzed AIMD trajectories of HCl solutions that were previously generated by Markland and co-workers.[59] These simulations were performed using the revPBE exchange−correlation DFT

functional[65,66] with D3 dispersion corrections.[67] This set of simulations includes a single excess proton in bulk water as well as two other concentrations, namely, 2 and 4 M HCl. The box size for the 2 and 4 M solutions was 14.91 Å, containing 106 H_2O molecules and either 4 or 8 excess protons and Cl^- anions, respectively. The trajectory with only a single excess proton was performed in a simulation box with a side length equal to 12.42 Å and 64 water molecules. Unless otherwise stated, we focus most of our analyses throughout the paper on the trajectories stemming from the 2 M HCl aqueous solution, as these are the ones where we had access to significantly more statistics.

Smooth Overlap Atomic Positions (SOAP). Over the past decade, SOAP has emerged as a powerful technique for describing the local environments around atoms and molecules, including organic and inorganic materials.[68−71] In addition, it has also recently been applied to investigate the properties of liquid water.[58,72,73] The SOAP starting point is that of treating the local density of an atomic environment χ as a sum of Gaussian functions with variance σ^2, centered on all species that are neighbors of the central atom

$$\rho_\chi(\mathbf{r}) = \sum_{i\in\chi} \exp\left(\frac{-|\mathbf{r_i} - \mathbf{r}|^2}{2\sigma^2}\right) \tag{1}$$

The atomic density in eq 1 is then expanded in terms of spherical harmonics and radial basis functions such that

$$\rho_\chi(\mathbf{r}) = \sum_{n=0}^{n_{max}} \sum_{l=0}^{l_{max}} \sum_{m=-l}^{l} c_{nlm} g_n(r) Y_{lm}(\theta, \phi) \tag{2}$$

Upon accumulation of the expansion coefficients, a rotationally invariant power spectrum is constructed

$$p_{nn'l}(\chi) = \pi \sqrt{\frac{8}{2l+1}} \sum_m (c_{nlm})^\dagger c_{n'lm} \tag{3}$$

Equation 3 defines the components of the SOAP features that encode the information on the atomic environments that arise in the simulations of the excess proton.

In this work, to build the SOAP descriptors, we used the DScribe package.[74] A total of 265 000 SOAP descriptors were constructed from the 2 M HCl simulations. In order to employ the SOAP descriptors, several hyperparameters must be set: (1) the species whose density we would like to describe, (2) the cutoff radius (r_{cut}) that defines the region that is considered as environment χ of an atom, (3) the σ value entering into eq 1, and (4) the number of radial and angular basis functions employed in the description of the densities (n_{max}, l_{max}).

Thus, to build the SOAP environments, we used oxygen atoms as centers, including both oxygen and hydrogen species within a radial cutoff of 3.7 Å. The σ value chosen for the Gaussian function was 0.25 Å, consistent with a recent study by our group.[75] This ensures that the SOAP descriptor contains important information about the HBN structure. Finally, $n_{max} = 12$ and $l_{max} = 8$ were set to build the descriptors.

Intrinsic Dimension (ID). High-dimensional data generated from molecular systems is often characterized by correlations, implying that the system of interest resides in a lower-dimensional manifold.[56] In the context of our analysis here, the power spectra associated with the SOAP descriptors are made up of high-dimensional vectors consisting of over 1000s of elements. In previous studies, we have shown that this corresponds to the number of independent directions needed to characterize the fluctuations of the local atomic environments of water[58] as well as in proteins.[76]

In this work, we determined the ID using the Two-NN estimator[61] which uses information on the first and second neighbors of each data point. This method is based on the assumption of local uniformity in the data set. It can be demonstrated that, if this condition is satisfied, the ratio of the first and second neighbor distances (r_1 and r_2, respectively) is characterized by the following distribution

$$P(\mu) = \frac{d}{\mu^{d+1}} \tag{4}$$

with $\mu = r_1/r_2$ and d being the ID. If the sampled μ_is are independent, the ID can be calculated by eq 5.

$$d = \frac{N}{\sum_i \log(\mu_i)} \tag{5}$$

DPA Clustering. The calculation of the free-energy surface (FES) of molecular systems involves knowing the relevant coordinates capable of representing the different (meta)stable states of the system. Usually, FES is estimated using a reduced number of coordinates, assuming that these latter will suffice to capture the relevant fluctuations of the system one wants to represent. However, this dimensionality reduction unavoidably leads to information loss. In this work, we use the point adaptive k-nearest (PAk) estimator that has been shown to correctly recover the high-dimensional free-energy surface (FES) of different molecular systems.[76−78] This method is an extension of the standard k-nearest estimator, in which the local density of every point i in the data set is calculated as

$$\rho_i = \frac{\hat{k}_i}{r_{i,k}^d} \tag{6}$$

where d is the ID of the data set extracted from the Two-NN estimator, and $\hat{k}_i$'s is chosen to be maximal under the constraint that the density within $r_{i,k}$ can be considered constant. As shown in ref 77, this selection provides a good bias-variance trade-off in the density estimation. The free energies are then calculated as $F_i = -\log(\rho_i)$. Once the point-dependent free energies F_i have been determined, we then employ the Density Peaks Advanced (DPA) algorithm to find the minima on the FES, often referred to as *clusters*. This method automatically determines the clusters and identifies only those that are within a statistical confidence interval, Z, which is an adjustable parameter. This parameter determines when a density maximum can be considered a consequence of a fluctuation due to limited sampling or a real cluster. In practice, we set the Z parameter as the minimum value that results in consistency between the results on two independently generated data sets.

Finally, in order to visualize the features of the data set as well as the results of the PAk-DPA algorithms, we used the Uniform Manifold Approximation and Projection (UMAP).[64,79] The UMAP method gives a low-dimensional projection of a high-dimensional data set (similar to t-SNE[80]), preserving its topography as much as possible and therefore allowing the visualization of the main features of the high-dimensional FES.

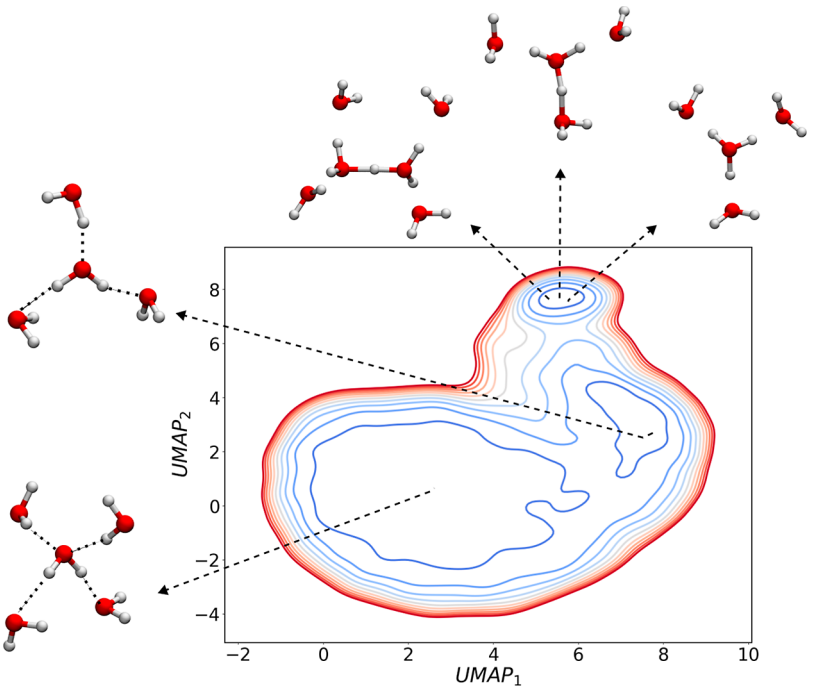


Figure 2. Two-dimensional contour plot obtained from the projection of the SOAP space onto two UMAP variables. Three distinct minima are observed, and some of the representative molecular structures associated with them are reported.

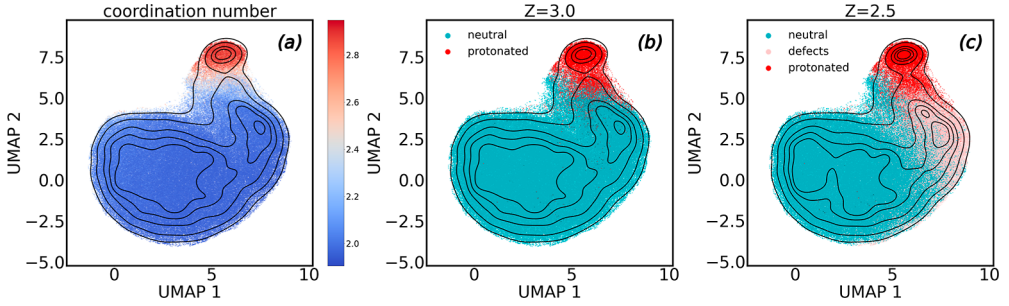


Figure 3. Two-dimensional UMAP projection coloring data by coordination number (a) and by cluster labels obtained with DPA clustering analysis (b,c). Two confidence levels are displayed: Z = 3.0 (b) and Z = 2.5 (c).

RESULTS

Free-Energy Landscape of 2 M HCl. We determined the SOAP descriptors centered on all of the oxygen atoms with a cutoff of 3.7 Å, including both oxygen and hydrogen in the environment. Such a choice ensures that the first solvation shell of each putative molecule or ion in the system is included in the atomic environment. We subsequently computed the ID of the SOAP descriptors, obtaining a value of ∼33. With these two inputs (*i.e.*, the SOAP descriptors and their ID), we then performed the Density Peaks Advanced (DPA) clustering with a confidence interval of Z = 3.0, obtaining two major clusters, one enclosing 91% of all data points and the other 8% of all the data points. The remaining 1% of the data does not belong to any of the two clusters. The presence of two clusters is consistent with our intuitive expectation that in water with an excess proton, there are likely to be at least two, broadly speaking, chemical topologies: one corresponding to neutral water molecules and another to the excess proton.

In order to build our intuitions and visualize the underlying topography and high-dimensional free-energy landscape, we projected the SOAP descriptors using UMAP. Figure 2 shows a contour plot of the resulting 2D-projection, where we observe the presence of three distinct minima. Each of these minima essentially corresponds to different molecular patterns present in the system. Upon visual inspection of the atomic environments associated with each minimum, one of these patterns corresponds to neutral, on average tetrahedrally coordinated, water molecules (bottom left), another one to charged water environments including the excess proton (top), and finally, the third minimum involves neutral defect-looking water molecules (middle left).

To better quantify the structural origins of the different minima, we constructed a standard coordination number using a switching function defined as $n_{\mathrm{H}} = \sum_{i=1}^{N_{\mathrm{H}}} \frac{1}{\exp[\kappa(r - r_c)] + 1}$ (see, *e.g.*, ref 81), which counts the number of hydrogen atoms around every oxygen atom within a given radial cutoff. This allows for quick identification of the oxygen atoms having either two or three covalently bonded hydrogens, allowing for an initial assignment of the clusters. Figure 3a shows the same UMAP projection, colored as a function of n_{H}. We observe a gradient in the coordination number starting from ∼3 and approaching ∼2 as one moves from the top to the bottom. This analysis confirms the results obtained from visual

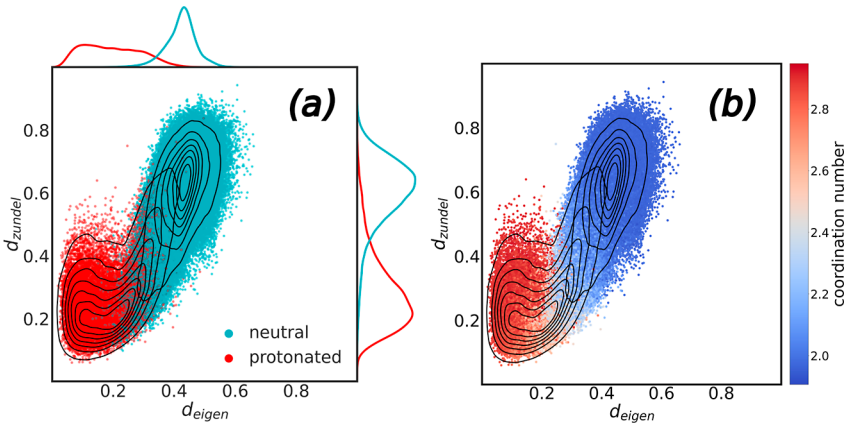


Figure 4. 2D density plot showing the SOAP distance to the idealized Eigen and Zundel references for the protonated and neutral clusters. The protonated cluster is characterized for smaller distances (structures are more similar to the references) than the neutral cluster, although a big overlap is observed between both of them. Coloring the distribution by coordination number shows that the protonated configurations have a coordination number ∼ 3.0 whereas the neutral ones ∼ 2.0.

inspection of the trajectories, which are shown in the snapshots of Figure 2. The minimum observed associated with higher coordination numbers (∼3) corresponds to the typical molecular environments that would be associated with the excess proton, while the other two minima—which exhibit a lower coordination number (∼2)—are those associated with neutral water molecules.

Although the UMAP projection seems to suggest the presence of three distinct minima, our DPA analysis shows the existence of only two clusters using Z = 3.0. The granularity of the underlying free-energy landscape—and therefore the number of clusters—depends on the choice of the confidence parameter Z. Although lowering the value of Z leads to the emergence of more clusters, the confidence in their statistical significance is reduced[77] and, thus, has to be combined with a chemical and physical interpretation. To examine how our results depend on the Z parameter, we compared the outcomes of the clustering obtained for Z = 3.0 and Z = 2.5. We observe that with Z = 2.5, we obtain three distinct clusters, which include 67, 19, and 8% of all data points. Subsequently, we have used the same reduced representation of UMAP, coloring the data points with the labels obtained from the clustering, as shown in Figure 3b,c, which compare the results for Z = 3.0 (middle panel) to those for Z = 2.5 (rightmost panel), respectively.

Interestingly, for Z = 3.0, we observe that one of the clusters obtained by means of DPA partially overlaps with the minimum involving a higher coordination number (∼3), while the other cluster is dominated by a coordination number of 2. Based on these features, the two clusters can be assigned as the protonated and neutral water clusters, respectively. Figure 3c, instead, shows that the effect of reducing the Z parameter and recoloring the UMAP landscape with the labels of the three clusters is that of splitting the large neutral water cluster into two regions, which essentially cover the two minima associated with neutral water. This third cluster, located in the middle of the UMAP projection, is characterized by a coordination number of 2, indicating that it has the local topology of a neutral water molecule.

With the aim of characterizing the origins of this cluster, we studied the H-bond patterns associated with these environments. Interestingly, water molecules in the new third cluster exhibit a deficiency in the number of donating and accepting H-bonds, implying that they are neutral water coordination defects.[82,83] Specifically, relative to neutral water, these defects have a reduced tendency to accept H-bonds (see Supporting Information Figure S1). In order to understand whether these defect water molecules have a tendency to form in close proximity to the excess proton, we determined the pair-correlation function between the oxygen atom centers of the protonated and the two neutral water clusters. We find that these neutral defective water molecules prefer to lie closer to the excess proton compared to waters coming from the other neutral basin (see Supporting Information Figure S2).

In a nutshell, these results demonstrate that the unsupervised approach going from the SOAP descriptors to free-energy clustering allows for an automatic identification of local environments in protonated liquid water, a feature that is consistently reproduced in path-integral simulations, different concentrations of the HCl, and also more accurate treatment of the electronic structure (see Supporting Information Figure S3). The existence of a defect neutral water pattern as a cluster forming close to the excess proton is fully consistent with the "special-pair dance" proposed by Voth and co-workers, where a strong interaction is formed between the excess proton and its nearest neighbor water molecule.[84] Within our unsupervised framework, it can be rationalized as the highly fluxional protonated cluster interacting and creating the neutral defect. However, we do not find any evidence of the existence of two different clusters that correspond to Eigen and Zundel species. Although reducing the confidence parameter to Z = 2.0 leads to the formation of more clusters inside the protonated basin, these inner clusters do not segregate into any statistically significant minima resembling either the Eigen or the Zundel moiety (see Supporting Information Figure S4).

Eigen or Zundel? The extent to which the proton delocalizes in the H-bond network is sensitive to factors such as temperature, zero-point energy fluctuations, and hydration. At low temperatures and in isolated (gas-phase) clusters, the idealized Eigen and Zundel geometries have been proven to be more easily distinguishable.[47−49,85,86] Agmon and co-workers have shown that in the gas phase, a proton added to the water tetramer yields $(H_3O)^+$, while for the water dimer, it yields $(H_5O_2)^+$.[10] Similarly, Johnson and co-workers have shown that the excess proton with 20 water molecules, often referred to as magic number protonated clusters, involves the Eigen cation

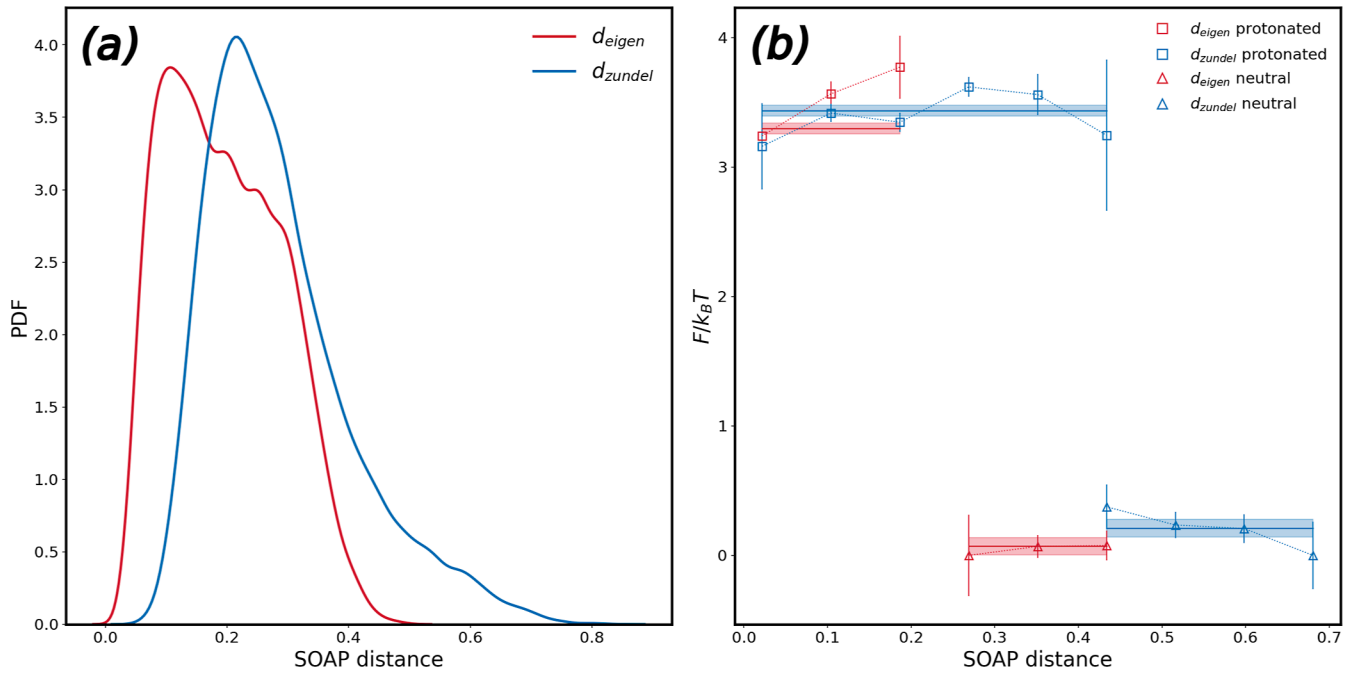


Figure 5. (a) Density distributions for d_{eigen} and d_{zundel} calculated from all structures in the protonated cluster. (b) Free energies calculated in the SOAP space from the DPA clustering algorithm. Only the energies of the core points of the protonated (squares) and neutral (triangles) clusters are considered. The coordinates d_{eigen} and d_{zundel} were divided into bins, and each point in the plot represents an average of free-energy values in each bin. Bars indicate the standard error of the sample in each bin. Horizontal lines follow the average free energy of the points in the protonated and neutral clusters as a function of d_{eigen} and d_{zundel}. Shaded areas display the standard error.

$(H_3O)^+$ on the surface.[87] Pipelining these low-temperature configurations through the same protocol (Figure 1) leads to the identification of two statistically significant clusters corresponding to Eigen and Zundel (see Supporting Information Figure S5).

To provide a more agnostic characterization of the similarity of the condensed phase proton structures to Eigen or Zundel geometries, we identified relevant milestone structures sampled from simulations in the gas phase, where the structures can be identified as Eigen or Zundel without any ambiguity similar to the geometries illustrated in Figure 1. For more details on how the reference structures are chosen and built, the reader is referred to Supporting Information. In summary, this is a postprocessing procedure where we recompute the SOAP descriptors using the Eigen and Zundel configurations obtained from 20 randomly chosen configurations in the magic number protonated simulations. Using these new SOAP vectors, we have computed the similarity measure $d(\chi, \chi')$ between the protonated and neutral clusters obtained with Z = 3.0 (χ) at 2 M HCl with respect to these reference configurations (χ'). $d(\chi, \chi')$ is chosen to be the minimum distance obtained from the 20 configurations. This yields two distances, namely, d_{eigen} and d_{zundel}.

Figure 4a shows the distribution of distances to the Eigen (d_{eigen}) and Zundel (d_{zundel}) reference structures for both the protonated and neutral clusters. A substantial overlap in the proximity of the protonated cluster to Eigen- and Zundel-looking geometries is observed. Figure 4b illustrates the distribution of d_{eigen} and d_{zundel} colored now as a function of coordination number n_H. In this manner, we see clearly that neutral water molecules are dominated by a coordination number close to 2, while that of the protonated cluster involves an enhanced coordination number of 3. A continuous transition between the neutral and protonated environments following the d_{eigen} and the d_{zundel} coordinates is observed, resulting in a large overlap between the neutral and protonated clusters. This arises from the fact that since the excess proton is a highly fluxional topological ionic defect, it strongly distorts the HBN so that water molecules in close vicinity to the excess proton can be classified as either a charged or neutral cluster.

The scatter plots and contour lines of Figure 4a are slightly skewed in favor of Eigen-like configurations, as observed by the fact that the density peak in the protonated cluster of Figure 4a is closer to zero along the d_{eigen} coordinate rather than along the d_{zundel}. This is illustrated more clearly in Figure 5a, which shows the one-dimensional distributions of the d_{eigen} and d_{zundel} distances. The d_{eigen} and d_{zundel} distributions of Figure 5a are one-dimensional projections from a high-dimensional distribution and, therefore, cannot be used to infer the dominance of Eigen *vs* Zundel. In order to understand the impact of this projection, we have determined the high-dimensional free energies, as described earlier from $F_i = -\log(\rho_i)$, of the protonated and neutral clusters as a function of d_{eigen} and d_{zundel}. These free energies are illustrated in Figure 5b. In order to construct these curves, we computed the average of the point free energies for the core points of each cluster along the d_{eigen} and d_{zundel} coordinates. The neutral environments lie in a free-energy basin located about 2−3 k_BT lower than the protonated ones. Interestingly, this free-energy difference is fully consistent with the small activation energy required for proton migration.[16] Examining the free-energy changes for the protonated cluster along the d_{eigen} and d_{zundel} coordinates essentially confirms the findings obtained from the clustering analysis presented earlier. Within the statistical errors of our simulations, the free energy along these two variables is essentially flat. Furthermore, by measuring the proximity of the excess proton using d_{eigen} and d_{zundel}, it is clear that the free energies along these variables are also within statistical error of

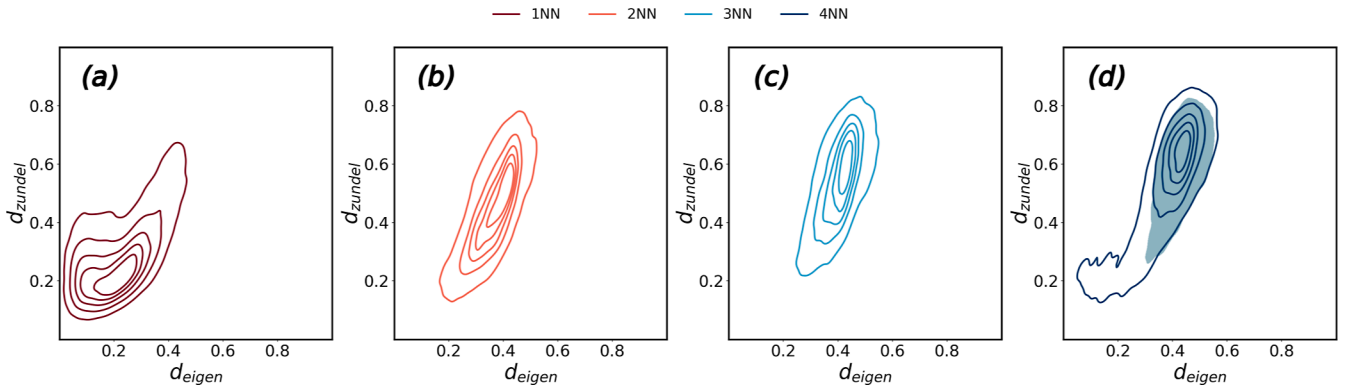


Figure 6. d_{eigen} (x-axes) and d_{zundel} (y-axes) density maps refer to the first (a), second (b), third (c), and fourth (d) nearest neighbors of the molecules in the protonated cluster.

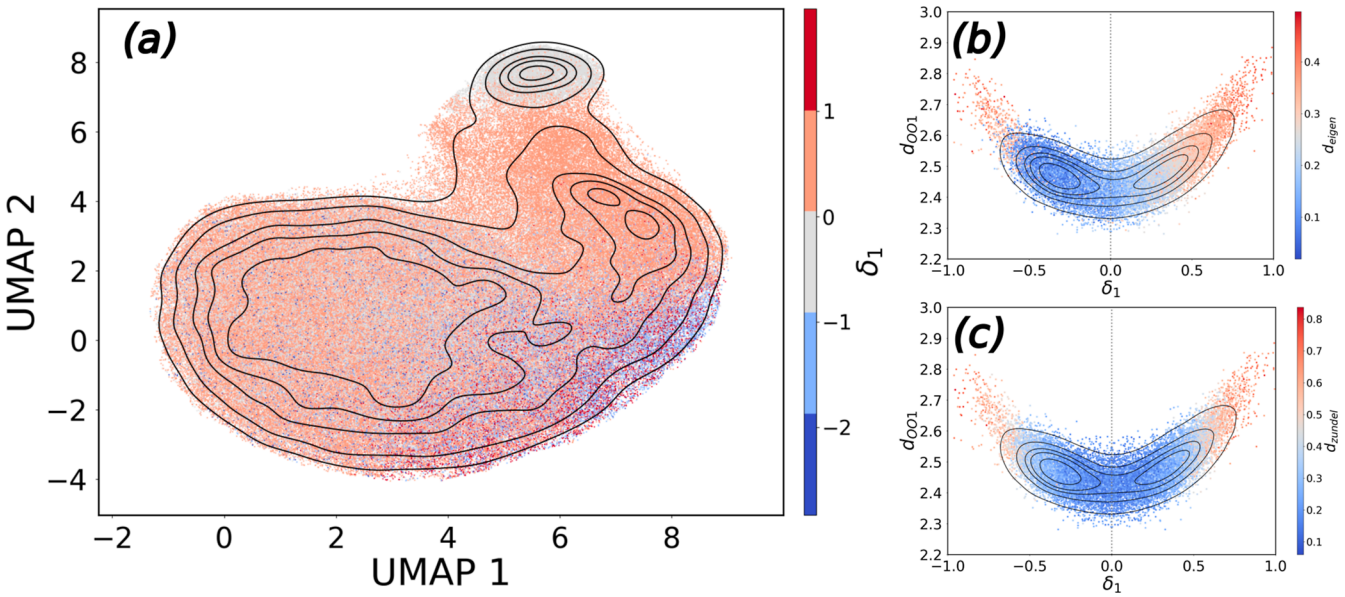


Figure 7. UMAP projection of the SOAP vectors colored as a function of proton sharing coordinate (PTC) δ_1 (a). Distances of the first-neighbor oxygen atom from the protonated oxygen atom d_{OO1} as a function of the PTC δ_1 and colored by the SOAP distances d_{eigen} (b) and d_{zundel} (c).

each other. This evidence implies that fluctuations of the excess proton in water occur on a *flat* free-energy landscape, where transformations between Eigen- and Zundel-like configurations occur. Eigen- and Zundel-like geometries are, however, neither limiting nor stable thermodynamic states on the free-energy landscape.

Up to this point, we have mainly focused on examining the structural similarities between the charged protonated cluster with respect to the Eigen and Zundel species. On the other hand, this ionic and highly fluxional topological defect strongly perturbs the nearby environment. It is hence interesting to investigate the extent of the similarity to the Eigen and Zundel species of the atomic environments in the proximity of the charged protonated cluster. To do so, we construct the d_{eigen} and d_{zundel} coordinates for the 4 nearest neighbors (NNs) of the charged cluster. For every snapshot of our trajectory, we identify the 4 closest oxygen atoms to the charged protonated cluster, extracting the d_{eigen} and d_{zundel} coordinates for those selected oxygen centers. Figure 6 shows the density plots for d_{eigen} (x-axes) and d_{zundel} (y-axes) for the first, second, third, and fourth NNs of the protonated cluster. The first NN (1NN), which essentially corresponds to the special pair discussed in the literature,[84] exhibits a sizable overlap in the d_{eigen} and d_{zundel} distributions. This further corroborates the idea that within the special-pair, a sharp distinction between Eigen and Zundel molecular structures does not adequately describe the fluctuations of the proton. Furthermore, the fact that the 1NN displays substantial similarity to the Zundel is also fully consistent with vibrational spectroscopy measurements of HCl.[51−53]

Going beyond the 1NN, we find that for both 2NN and 3NN (Figure 6b,c), the d_{eigen} and d_{zundel} distributions also exhibit a significant overlap. Thus, in addition to the special-pair atomic environment that resides close to the charged cluster (1NN), proton delocalization to other neighbors involves a substantial overlap between the Eigen and the Zundel configurations. The picture that emerges from the current analysis is in good agreement with previous characterizations of the PT mechanisms associated with the special-pair dance, where the excess proton dynamically fluctuates between forming H-bonds between three different atomic environments.[38] Finally, Figure 6d displays the distributions obtained for the 4NN, where the d_{eigen} and d_{zundel} coordinates essentially converge to those obtained for the neutral cluster (shown as a shaded area).

Combining SOAP and Chemical Descriptors. Due to the fact that the Grotthuss mechanism involves the interconversion of hydrogen and covalent bonds in a highly cooperative process,[38] identifying the relevant degrees of freedom that describe the proton hopping phenomenon is challenging. The extent of PT along the H-bond is typically quantified through the proton transfer coordinate (PTC), which is measured by the difference in distance between the proton and the oxygen atoms along which the transfer occurs. In order to enhance the chemical interpretability of the SOAP projections in UMAP, we begin by showing in the left panel of Figure 7 the original UMAP projection colored as a function of the PTC. To construct the PTC, we sit on the oxygen atom of each of the three clusters identified with the confidence parameter $Z = 2.5$, identify the 1NN, and then compute the PTC as $d_{O_CH} - d_{HO_{1NN}}$, where O_C is the oxygen atom associated with the charged protonated cluster. In the bulk neutral water cluster, the PTC coordinate is randomly distributed, meaning that it takes on both positive and negative values that correspond to water molecules involved in accepting and donating H-bonds, respectively.

Honing in on the charged protonated and neutral defect minima, the PTC is on average negative (−0.4) and positive (0.5), respectively. The interconversion between Eigen and Zundel species has typically been rationalized through the PTC: values close to zero are often interpreted as the creation of a Zundel cation, while deviations from this are interpreted as the formation of the Eigen. The neutral water defect structure forming the minima in the middle of the UMAP projection appears to be dominated by PTC values that are on average greater than zero, consistent with the notion of a molecular structure that is most likely to accept a proton through PT.

Besides the special-pair dance which emerges from our analysis, PT has also been shown to be strongly coupled to the lower-frequency vibrational modes of the heavy atoms that sandwich the proton.[30,33,88] Figure 7b,c show the distributions of the first-neighbor oxygen atom distances from the protonated oxygen d_{OO1} obtained along the PTC, which reproduce the banana-like profiles observed in many previous studies (see, *e.g.*, ref 33). On each banana-shaped profile, we color the plot with the d_{eigen} (Figure 7b) and d_{zundel} (Figure 7c) distances. Interestingly, the asymmetric nature of the Eigen cation with respect to the location of the proton is consistent with its position on the banana plot, *i.e.*, asymmetrically populated basins at short (blue dots) and long (orange dots) d_{eigen} distances (Figure 7b). At the same time, however, PTC values close to zero can also resemble Eigen-like species. The banana-plot of the centrosymmetric Zundel complex, displayed in Figure 7c, shows that there exists a very diffuse basin accessible to the excess proton across the full landscape. In fact, the proton rattling phenomenon along the H-bonds is always characterized by short d_{zundel} coordinates. In other words, the Zundel cation remains a Zundel complex even when the PTC departs from zero, confirming once again that the discretization of the excess proton into two species does not adequately account for the underlying relevant fluctuations. These observations are fully consistent also with the delocalized nature of the excess proton as a distorted Zundel cation presented by Tokmakoff and co-workers.[51−53]

CONCLUSIONS

In the current work, we exploited the smooth overlap of atomic descriptors (SOAP) along with advanced clustering techniques to characterize local environments in acidic water. Using previous AIMD simulations of HCl aqueous solutions at various concentrations reported by Markland and co-workers,[59] we show that the unsupervised clustering of the high-dimensional SOAP space yields three statistically dominant clusters, of which only one corresponds to the excess proton. This charged cluster contains both Eigen and Zundel configurations, although the idealized structures do not emerge as limiting thermodynamic states. By exploiting the definition of metrics allowing for the evaluation of the distance d from well-defined Eigen (d_{eigen}) and Zundel (d_{zundel}) gas-phase milestone structures, we also have determined the free-energy changes for the protonated cluster along the d_{eigen} and d_{zundel} coordinates. Interestingly, the free energy along these two variables is flat while the free-energy difference between the Eigen and the Zundel moieties is nonexistent, implying that fluctuations of the excess proton in water take place on a flat free-energy landscape, where transformations between Eigen- and Zundel-like configurations easily occur.

Another relevant feature emerging from our unsupervised analyses of the global free-energy surface of the system is that the protonated and neutral clusters are separated by a peculiar basin associated with nontetrahedrally coordinated water molecules. In fact, the charged protonated cluster enhances the concentration of a specific type of neutral defect that tend to lie in the proximity of the proton. In particular, this neutral defect is characterized by a H-bond topology that is defective, a feature that is enhanced in HCl solutions compared to that of bulk water with excess protons. Conversion between the charged and neutral clusters essentially describes the fluctuations of the proton involved in the Grotthuss mechanism, where structural interconversions of a wide spectrum of molecular structures form part of a broad spectrum of fluctuations of a charged topological defect. Eigen and Zundel, as originally conceived, form just one of many different structures, and perhaps it is time to let go of very specific assignments based on two-state type descriptions.

Finally, when our clustering analysis is combined with the SOAP distance metrics and previously used chemical-based coordinates, a clear scenario emerges. There exists a very diffuse basin accessible to the excess proton associated with a highly fluxionally charged defect whose chemical identity is fuzzy. In other words, the proton rattling phenomenon along the H-bonds is always characterized by moieties that are simultaneously Eigen- and Zundel-like. Instead, fluctuations between protonic and neutral defects—that we dub as *ZundEig*—is a more appropriate representation for studying PT mechanisms. We believe our results provide a general framework for examining the evolution of the structure of the proton in more complex environments, such as at interfaces,[89,90] under confinement,[91,92] and also in the presence of external electric fields where the extent of proton delocalization can be modulated.[35,93,94]

AUTHOR INFORMATION

Corresponding Author

Ali Hassanali − *International Centre for Theoretical Physics, 34151 Trieste, Italy;* orcid.org/0000-0002-3208-1488; Email: ahassana@ictp.it

Authors

Solana Di Pino − *Departamento de Química Inorgánica, Analítica y Química Física/INQUIMAE, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, C1428EHA Buenos Aires, Argentina*

Edward Danquah Donkor − *International Centre for Theoretical Physics, 34151 Trieste, Italy; Scuola Internazionale Superiore di Studi Avanzati (SISSA), 34136 Trieste, Italy;* orcid.org/0000-0001-7565-7603

Veronica M. Sánchez − *Departamento de Química Inorgánica, Analítica y Química Física/INQUIMAE, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, C1428EHA Buenos Aires, Argentina;* orcid.org/0000-0003-1949-9748

Alex Rodriguez − *International Centre for Theoretical Physics, 34151 Trieste, Italy; Dipartimento di Matematica e Geoscienze, Universitá degli Studi di Trieste, 34127 Trieste, Italy;* orcid.org/0000-0002-0213-6695

Giuseppe Cassone − *Institute for Chemical-Physical Processes, National Research Council (CNR-IPCF), 98158 Messina, Italy;* orcid.org/0000-0003-1895-2950

Damian Scherlis − *Departamento de Química Inorgánica, Analítica y Química Física/INQUIMAE, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, C1428EHA Buenos Aires, Argentina;* orcid.org/0000-0002-0588-287X

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

A.H. acknowledges funding by the European Union (ERC, HyBOP, grant number: 101043272). The views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. G.C. acknowledges CINECA for an award under the ISCRA initiative (ELFI-HB—HP10BZLOOF) and for the availability of high-performance computing resources.

REFERENCES

(1) Kreuer, K.-D. Proton conductivity: Materials and applications. *Chem. Mater.* **1996**, *8*, 610−641.

(2) Cukierman, S. Et tu, Grotthuss! and other unfinished stories. *Biochim. Biophys. Acta, Biomembr.* **2006**, *1757*, 876−885.

(3) Wraight, C. A. Chance and design - Proton transfer in water, channels and bioenergetic proteins. *Biochim. Biophys. Acta, Biomembr.* **2006**, *1757*, 886−912.

(4) Raymand, D.; van Duin, A. C.; Goddard, W. A. I.; Hermansson, K.; Spångberg, D. Hydroxylation Structure and Proton Transfer Reactivity at the Zinc Oxide-Water Interface. *J. Phys. Chem. C* **2011**, *115*, 8573−8579.

(5) Yamaguchi, A.; Inuzuka, R.; Takashima, T.; Hayashi, T.; Hashimoto, K.; Nakamura, R. Regulating proton-coupled electron transfer for efficient water splitting by manganese oxides at neutral pH. *Nat. Commun.* **2014**, *5*, 4256.

(6) Ma, X.; Shi, Y.; Liu, J.; Li, X.; Cui, X.; Tan, S.; Zhao, J.; Wang, B. Hydrogen-Bond Network Promotes Water Splitting on the TiO2 Surface. *J. Am. Chem. Soc.* **2022**, *144*, 13565−13573.

(7) Fajín, J. L. C.; D S Cordeiro, M. N.; Gomes, J. R. B. Density Functional Theory Study of the Water Dissociation on Platinum Surfaces: General Trends. *J. Phys. Chem. A* **2014**, *118*, 5832−5840.

(8) Warburton, R. E.; Mayer, J. M.; Hammes-Schiffer, S. Proton-Coupled Defects Impact O−H Bond Dissociation Free Energies on Metal Oxide Surfaces. *J. Phys. Chem. Lett.* **2021**, *12*, 9761−9767.

(9) Creazzo, F.; Luber, S. *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*; Elsevier, 2023.

(10) Agmon, N.; Bakker, H. J.; Campen, R. K.; Henchman, R. H.; Pohl, P.; Roke, S.; Thämer, M.; Hassanali, A. Protons and Hydroxide Ions in Aqueous Systems. *Chem. Rev.* **2016**, *116*, 7642−7672.

(11) Stuchebrukhov, A. A. Mechanisms of proton transfer in proteins: Localized charge transfer versus delocalized soliton transfer. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2009**, *79*, 031927.

(12) Persson, F.; Halle, B. How amide hydrogens exchange in native proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 10383−10388.

(13) Li, C.; Voth, G. A. A quantitative paradigm for water-assisted proton transport through proteins and other confined spaces. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2113141118.

(14) Hynes, J. T. Physical chemistry: The protean proton in water. *Nature* **1999**, *397*, 565−567.

(15) de Grotthuss, C. J. T. Theory of decomposition of liquids by electrical currents. *Ann. Chim.* **1806**, *58*, 54−74.

(16) Agmon, N. The Grotthuss mechanism. *Chem. Phys. Lett.* **1995**, *244*, 456−462.

(17) Atkins, P.; de Paula, J. *Physical Chemistry*, 8th ed.; Oxford University Press, 2006.

(18) Eigen, M.; De Maeyer, L. Self-dissociation and protonic charge transport in water and ice. *Proc. R. Soc. London, Ser. A* **1958**, *247*, 505−533.

(19) Zundel, G. Hydration structure and intermolecular interaction in polyelectrolytes. *Angew Chem. Int. Ed. Engl.* **1969**, *8*, 499−509.

(20) Eigen, M. Proton transfer, acid-base catalysis, and enzymatic hydrolysis. I. Elementary processes. *Angew Chem. Int. Ed. Engl.* **1964**, *3*, 1−19.

(21) Danninger, W.; Zundel, G. Intense Depolarized Rayleigh-Scattering in Raman-Spectra of Acids Caused by Large Proton Polarizabilities of Hydrogen-Bonds. *J. Chem. Phys.* **1981**, *74*, 2769−2777.

(22) Zundel, G. Hydrogen Bonds with Large Proton Polarizability and Proton Transfer Processes in Electrochemistry and Biology. *Adv. Chem. Phys.* **1999**, *111*, 1−217.

(23) Botti, A.; Bruni, F.; Ricci, M. A.; Soper, A. K. Eigen Versus Zundel Complexes in Hcl-Water Mixtures. *J. Chem. Phys.* **2006**, *125*, 014508.

(24) Tielrooij, K. J.; Timmer, R. L. A.; Bakker, H. J.; Bonn, M. Structure Dynamics of the Proton in Liquid Water Probed with Terahertz Time-Domain Spectroscopy. *Phys. Rev. Lett.* **2009**, *102*, 198303.

(25) Thamer, M.; De Marco, L.; Ramasesha, K.; Mandal, A.; Tokmakoff, A. Ultrafast 2d Ir Spectroscopy of the Excess Proton in Liquid Water. *Science* **2015**, *350*, 78−82.

(26) Dahms, F.; Costard, R.; Pines, E.; Fingerhut, B. P.; Nibbering, E. T. J.; Elsaesser, T. The Hydrated Excess Proton in the Zundel Cation H5O2+: The Role of Ultrafast Solvent Fluctuations. *Angew. Chem., Int. Ed.* **2016**, *55*, 10600−10605.
(27) Kozari, E.; Sigalov, M.; Pines, D.; Fingerhut, B. P.; Pines, E. Infrared and NMR Spectroscopic Fingerprints of the Asymmetric H7+O3 Complex in Solution. *ChemPhysChem* **2021**, *22*, 716−725.
(28) Hückel, Z. Theorie der Beweglichkeiten des Wasserstoff- und Hydroxylions in wässeriger Lösung. *Z. Elektrochem.* **1928**, *34*, 546−562.
(29) Tuckerman, M. E.; Marx, D.; Klein, M. L.; Parrinello, M. On the quantum nature of the shared proton in hydrogen bonds. *Science* **1997**, *275*, 817−820.
(30) Marx, D.; Tuckerman, M. E.; Hutter, J.; Parrinello, M. The nature of the hydrated excess proton in water. *Nature* **1999**, *397*, 601−604.
(31) Marx, D.; Chandra, A.; Tuckerman, M. E. Aqueous basic solutions: Hydroxide solvation, structural diffusion, and comparison to the hydrated proton. *Chem. Rev.* **2010**, *110*, 2174−2216.
(32) Tse, Y.-L. S.; Knight, C.; Voth, G. A. An analysis of hydrated proton diffusion in ab initio molecular dynamics. *J. Chem. Phys.* **2015**, *142*, 014104.
(33) Hassanali, A.; Giberti, F.; Cuny, J.; Kühne, T. D.; Parrinello, M. Proton transfer through the water gossamer. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 13723−13728.
(34) Hassanali, A. A.; Giberti, F.; Sosso, G. C.; Parrinello, M. The role of the umbrella inversion mode in proton diffusion. *Chem. Phys. Lett.* **2014**, *599*, 133−138.
(35) Cassone, G. Nuclear Quantum Effects Largely Influence Molecular Dissociation and Proton Transfer in Liquid Water under an Electric Field. *J. Phys. Chem. Lett.* **2020**, *11*, 8983−8988.
(36) Lobaugh, J.; Voth, G. A. The quantum dynamics of an excess proton in water. *J. Chem. Phys.* **1996**, *104*, 2056−2069.
(37) Pavese, M.; Voth, G. A. Quantum and classical simulations of an excess proton in water. *Ber. Bunsenges. Phys. Chem.* **1998**, *102*, 527−532.
(38) Lapid, H.; Agmon, N.; Petersen, M. K.; Voth, G. A. A bond-order analysis of the mechanism for hydrated proton mobility in liquid water. *J. Chem. Phys.* **2005**, *122*, 014506.
(39) Wu, Y.; Chen, H.; Wang, F.; Paesani, F.; Voth, G. A. An Improved Multistate Empirical Valence Bond Model for Aqueous Proton Solvation and Transport. *J. Phys. Chem. B* **2008**, *112*, 467−482.
(40) Knight, C.; Voth, G. A. The Curious Case of the Hydrated Proton. *Acc. Chem. Res.* **2012**, *45*, 101−109.
(41) Tse, Y.-L. S.; Chen, C.; Lindberg, G. E.; Kumar, R.; Voth, G. A. Propensity of Hydrated Excess Protons and Hydroxide Anions for the Air−Water Interface. *J. Am. Chem. Soc.* **2015**, *137*, 12610−12616.
(42) Headrick, J. M.; Diken, E. G.; Walters, R. S.; Hammer, N. I.; Christie, R. A.; Cui, J.; Myshakin, E. M.; Duncan, M. A.; Johnson, M. A.; Jordan, K. D. Spectral signatures of hydrated proton vibrations in water clusters. *Science* **2005**, *308*, 1765−1769.
(43) Yu, H.; Cui, Q. The vibrational spectra of protonated water clusters: A benchmark for self-consistent-charge density-functional tight binding. *J. Chem. Phys.* **2007**, *127*, 234504.
(44) Park, M.; Shin, I.-H.; Singh, N. J.; Kim, K. S. Eigen and Zundel forms of small protonated water clusters: Structures and infrared spectra. *J. Phys. Chem. A* **2007**, *111*, 10692−10702.
(45) Dong, F.; Nesbitt, D. J. Jet cooled spectroscopy of H_2DO^+: Barrier heights and isotope-dependent tunneling dynamics from H_3O^+ to D_3O^+. *J. Chem. Phys.* **2006**, *125*, 144311.
(46) Sobolewski, A. L.; Domcke, W. Ab Initio Investigation of the Structure and Spectroscopy of Hydronium-Water Clusters. *J. Phys. Chem. A* **2002**, *106*, 4158−4167.
(47) Fournier, J. A.; Wolke, C. T.; Johnson, M. A.; Odbadrakh, T. T.; Jordan, K. D.; Kathmann, S. M.; Xantheas, S. S. Snapshots of proton accommodation at a microscopic water surface: Understanding the vibrational spectral signatures of the charge defect in cryogenically cooled H+(H2O)n = 2−28 clusters. *J. Phys. Chem. A* **2015**, *119*, 9425−9440.
(48) Fournier, J. A.; Wolke, C. T.; Johnson, C. J.; Johnson, M. A.; Heine, N.; Gewinner, S.; Schöllkopf, W.; Esser, T. K.; Fagiani, M. R.; Knorke, H.; et al. Site-specific vibrational spectral signatures of water molecules in the magic $H^+(H_2O)_{20}$and $Cs^+(H_2O)_{20}$clusters. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 18132−18137.
(49) Shin, J.-W.; Hammer, N. I.; Diken, E. G.; Johnson, M. A.; Walters, R. S.; Jaeger, T. D.; Duncan, M. A.; Christie, R. A.; Jordan, K. D. Infrared Signature of Structures Associated with the $H^+(H_2O)_n$(n = 6 to 27) Clusters. *Science* **2004**, *304*, 1137−1140.
(50) Calio, P. B.; Li, C.; Voth, G. A. Resolving the Structural Debate for the Hydrated Excess Proton in Water. *J. Am. Chem. Soc.* **2021**, *143*, 18672−18683.
(51) Fournier, J. A.; Carpenter, W. B.; Lewis, N. H. C.; Tokmakoff, A. Broadband 2D IR spectroscopy reveals dominant asymmetric H5O2+ proton hydration structures in acid solutions. *Nat. Chem.* **2018**, *10*, 932−937.
(52) Carpenter, W. B.; Fournier, J. A.; Lewis, N. H. C.; Tokmakoff, A. Picosecond Proton Transfer Kinetics in Water Revealed with Ultrafast IR Spectroscopy. *J. Phys. Chem. B* **2018**, *122*, 2792−2802.
(53) Carpenter, W. B.; Lewis, N. H. C.; Fournier, J. A.; Tokmakoff, A. Entropic barriers in the kinetics of aqueous proton transfer. *J. Chem. Phys.* **2019**, *151*, 034501.
(54) Vuilleumier, R.; Borgis, D. Quantum Dynamics of an Excess Proton in Water Using an Extended Empirical Valence-Bond Hamiltonian. *J. Phys. Chem. B* **1998**, *102*, 4261−4264.
(55) Vuilleumier, R.; Borgis, D. An Extended Empirical Valence Bond Model for Describing Proton Mobility in Water. *Isr. J. Chem.* **1999**, *39*, 457−467.
(56) Glielmo, A.; Husic, B. E.; Rodriguez, A.; Clementi, C.; Noé, F.; Laio, A. Unsupervised Learning Methods for Molecular Simulation Data. *Chem. Rev.* **2021**, *121*, 9722−9758.
(57) Glielmo, A.; Zeni, C.; Cheng, B.; Csányi, G.; Laio, A. Ranking the information content of distance measures. *PNAS Nexus* **2022**, *1*, pgac039.
(58) Offei-Danso, A.; Hassanali, A.; Rodriguez, A. High-Dimensional Fluctuations in Liquid Water: Combining Chemical Intuition with Unsupervised Learning. *J. Chem. Theory Comput.* **2022**, *18*, 3136−3150.
(59) Napoli, J. A.; Marsalek, O.; Markland, T. E. Decoding the spectroscopic features and time scales of aqueous proton defects. *J. Chem. Phys.* **2018**, *148*, 222833.
(60) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B* **2013**, *87*, 184115.
(61) Facco, E.; d'Errico, M.; Rodriguez, A.; Laio, A. Estimating the intrinsic dimension of datasets by a minimal neighborhood information. *Sci. Rep.* **2017**, *7*, 12140.
(62) Rodriguez, A.; Laio, A. Clustering by fast search and find of density peaks. *Science* **2014**, *344*, 1492−1496.
(63) Rodriguez, A.; d'Errico, M.; Facco, E.; Laio, A. Computing the free energy without collective variables. *J. Chem. Theory Comput.* **2018**, *14*, 1206−1215.
(64) McInnes, L.; Healy, J.; Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. **2020**, arXiv:1802.03426. arXiv preprint.
(65) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(66) Zhang, Y.; Yang, W. Comment on "Generalized Gradient Approximation Made Simple". *Phys. Rev. Lett.* **1998**, *80*, 890.
(67) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(68) Maksimov, D.; Baldauf, C.; Rossi, M. The conformational space of a flexible amino acid at metallic surfaces. *Int. J. Quantum Chem.* **2021**, *121*, No. e26369.

(69) Grant, W. Network modularity and local environment similarity as descriptors of protein structure. Ph.D. Thesis, University of Cambridge, 2020.
(70) De, S.; Bartók, A. P.; Csányi, G.; Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754−13769.
(71) Appignanesi, G. A.; Rodriguez Fris, J. A.; Sciortino, F. Evidence of a two-state picture for supercooled water and its connections with glassy dynamics. *Eur. Phys. J. E* **2009**, *29*, 305−310.
(72) Capelli, R.; Muniz-Miranda, F.; Pavan, G. M. Ephemeral ice-like local environments in classical rigid models of liquid water. *J. Chem. Phys.* **2022**, *156*, 214503.
(73) Monserrat, B.; Brandenburg, J. G.; Engel, E. A.; Cheng, B. Liquid water contains the building blocks of diverse ice phases. *Nat. Commun.* **2020**, *11*, 5757.
(74) Himanen, L.; Jäger, M. O.; Morooka, E. V.; Federici Canova, F.; Ranawat, Y. S.; Gao, D. Z.; Rinke, P.; Foster, A. S. DScribe: Library of descriptors for machine learning in materials science. *Comput. Phys. Commun.* **2020**, *247*, 106949.
(75) Donkor, E. D.; Laio, A.; Hassanali, A. Do Machine-Learning Atomic Descriptors and Order Parameters Tell the Same Story? The Case of Liquid Water. *J. Chem. Theory Comput.* **2023**, *19*, 4596−4605.
(76) Sormani, G.; Rodriguez, A.; Laio, A. Explicit Characterization of the Free-Energy Landscape of a Protein in the Space of All Its Cα Carbons. *J. Chem. Theory Comput.* **2020**, *16*, 80−87.
(77) d'Errico, M.; Facco, E.; Laio, A.; Rodriguez, A. Automatic topography of high-dimensional data sets by non-parametric density peak clustering. *Inf. Sci.* **2021**, *560*, 476−492.
(78) Ansari, N.; Laio, A.; Hassanali, A. Spontaneously forming dendritic voids in liquid water can host small polymers. *J. Phys. Chem. Lett.* **2019**, *10*, 5585−5591.
(79) Diaz-Papkovich, A.; Anderson-Trocmé, L.; Ben-Eghan, C.; Gravel, S. UMAP reveals cryptic population structure and phenotype heterogeneity in large genomic cohorts. *PLoS Genet.* **2019**, *15*, No. e1008432.
(80) van der Maaten, L. Accelerating t-SNE using Tree-Based Algorithms. *J. Mach. Learn. Res.* **2014**, *15*, 3221−3245.
(81) Sprik, M. Computation of the pK of liquid water using coordination constraints. *Chem. Phys.* **2000**, *258*, 139−150.
(82) Gasparotto, P.; Hassanali, A. A.; Ceriotti, M. Probing Defects and Correlations in the Hydrogen-Bond Network of ab Initio Water. *J. Chem. Theory Comput.* **2016**, *12*, 1953−1964.
(83) Henchman, R. H.; Irudayam, S. J. Topological hydrogen-bond definition to characterize the structure and dynamics of liquid water. *J. Phys. Chem. B* **2010**, *114*, 16792−16810.
(84) Markovitch, O.; Chen, H.; Izvekov, S.; Paesani, F.; Voth, G. A.; Agmon, N. Special pair dance and partner selection: Elementary steps in proton transport in liquid water. *J. Phys. Chem. B* **2008**, *112*, 9456−9466.
(85) Kulig, W.; Agmon, N. Both Zundel and Eigen Isomers Contribute to the IR Spectrum of the Gas-Phase H9O4+ Cluster. *J. Phys. Chem. B* **2014**, *118*, 278−286.
(86) Mouhat, F.; Peria, M.; Morresi, T.; Vuilleumier, R.; Saitta, A. M.; Casula, M. Thermal dependence of the hydrated proton and optimal proton transfer in the protonated water hexamer. *Nature* **2023**, DOI: 10.1038/s41467-023-42366-4.
(87) Fournier, J. A.; Johnson, C. J.; Wolke, C. T.; Weddle, G. H.; Wolk, A. B.; Johnson, M. A. Vibrational spectral signature of the proton defect in the three-dimensional $H^+(H_2O)_{21}$ cluster. *Science* **2014**, *344*, 1009−1012.
(88) 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*, 20410−20415.
(89) Björneholm, O.; Hansen, M. H.; Hodgson, A.; Liu, L.; Limmer, D. T.; Michaelides, A.; Pedevilla, P.; Rossmeisl, J.; Shen, H.; Tocci, G.; et al. Water at Interfaces. *Chem. Rev.* **2016**, *116*, 7698−7726.
(90) Creazzo, F.; Pezzotti, S.; Bougueroua, S.; Serva, A.; Sponer, J.; Saija, F.; Cassone, G.; Gaigeot, M.-P. Enhanced conductivity of water at the electrified air−water interface: a DFT-MD characterization. *Phys. Chem. Chem. Phys.* **2020**, *22*, 10438−10446.
(91) Di Pino, S.; Perez Sirkin, Y. A.; Morzan, U. N.; Sánchez, V. M.; Hassanali, A.; Scherlis, D. A. Water Self-Dissociation is Insensitive to Nanoscale Environments. *Angew. Chem., Int. Ed.* **2023**, *62*, No. e202306526.
(92) Tielrooij, K. J.; Cox, M. J.; Bakker, H. J. Effect of Confinement on Proton-Transfer Reactions in Water Nanopools. *ChemPhysChem* **2009**, *10*, 245−251.
(93) 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*, 207801.
(94) Cassone, G.; Giaquinta, P. V.; Saija, F.; Saitta, A. M. Proton Conduction in Water Ices under an Electric Field. *J. Phys. Chem. B* **2014**, *118*, 4419−4424.