PNAS RESEARCH ARTICLE | BIOCHEMISTRY CHEMISTRY OPEN ACCESS

Direct evidence of acid-driven protein desolvation

Farzad Hamdi[a,b,1], Ioannis Skalidis[a,c,1], Inken Kaja Schwerin[d,1], Jaydeep Belapure[a,e], Dmitry A. Semchonok[a,f], Fotis L. Kyrilis[a,g], Christian Tüting[a,b], Johannes Müller[h], Georg Künze[d,i,j], and Panagiotis L. Kastritis[a,b,g,k,2]

Affiliations are included on p. 10.

Edited by Nikolaus Grigorieff, University of Massachusetts Medical School, Worcester, MA; received September 15, 2025; accepted February 4, 2026

Water and its ability to modulate the protonation states of biomolecules govern the physical chemistry of life, dictating their metabolic functions. However, how amino acid protonation alters protein hydration and solubility is an open question since Kuntz and Kauzmann proposed p*H*-driven protein desolvation in 1974. Here, in a series of high-resolution cryoelectron microscopy structures of a protein complex at different p*H* values (from p*H* 9.0 to 3.5), we examined thousands of observable hydration sites. Cryoelectron microscopy data, in agreement with constant-p*H* molecular dynamics simulations, show that nearly half of protein-bound waters exchanged with the bulk solvent upon acidification, with ~100 waters lost per p*H* unit per molecule. The loss of waters was most significant around the side chains of glutamate and aspartate residues while specific polar residues, mostly asparagine, anchored persistent waters. A positionally conserved hydration layer was observed across all p*H* conditions, accounting for 40% of resolved waters. Those waters displayed denser packing than less persistent waters, forming a p*H*-independent solvation shell. Acid-induced water exchange also displaced bound iron, providing a mechanistic link between solvation and metal release. Our findings demonstrate the core principles of acid-driven protein desolvation, resolving a 50-y-old biochemical hypothesis.

protein hydration | cryo-EM | water molecules | molecular dynamics

Significance

Life depends on proteins, and proteins depend on water. Yet for 50 y, the open question around what happens to the water around proteins in acidic conditions has not been resolved. Here, we visualized biomolecular hydration at the atomic level as a function of increasing acidity. We saw hundreds of water molecules leave the structure, while a persistent shell of water remained, organized by ~40% of resolved waters. We also found that acidity shifts how specific metals, i.e., iron, are held in their binding sites. Our results resolve a long-standing question in biochemistry and reveal simple rules for how acidity affects protein solvation. Our findings may also aid the design of more stable or p*H*-tolerant proteins, critical for biotechnological applications.

Water molecules produce complex solvation layers around macromolecules that may extend beyond 20 Å (1). Their dynamics correlate with protein structure, function, interaction, and dynamics (2–4). Atomic modeling of ordered water molecules in high-resolution three-dimensional (3D) electron density maps (5), revealed their roles in critical biological processes, such as RNA function (6), peptide bond cleavage (7), signal transduction by G-protein-coupled receptors (GPCRs) (8), e.g., in rhodopsin (8), and proton-translocation pathways in the respiratory pathway or in photosynthesis (9). Such water molecules, typically resolved in high-resolution structures, are classified as "structural" or "ordered waters", i.e., integral parts of the macromolecular architecture (2) and together with faster, low-occupancy waters define solvation dynamics (10). X-ray crystallography (11) and nuclear magnetic resonance spectroscopy (12) (NMR) have been traditionally applied to experimentally probe the structure and dynamics of water molecules, respectively, while AI-based approaches do not, yet, model water positions in biomolecular interactions at near-atomic resolution (13) or, when they do, response to changes in ionic strength is not considered (14). The dynamics of the water shell that surrounds proteins have been studied in great detail by methods such as NMR, terahertz time-domain spectroscopy, and neutron scattering, among others, to reveal that water molecules around proteins are much more ordered in comparison to bulk water (10, 15–23). Interestingly, NMR studies employing, e.g., nuclear Overhauser effect spectroscopy (NOESY) have shown that there is no difference in the residence time of surface water molecules which are observed in X-ray structures vs. those that are not detected (24).

Single-particle cryoelectron microscopy (cryo-EM) is an alternative method for determining water molecule structures interacting with biomolecules (25)—but its application has been limited, being the most recent of those in terms of high-resolution analysis. Cryo-EM structures where water molecules can be resolved have recently become available due to the ongoing resolution revolution (26), following advances in instrumentation, image processing and analysis software, and integration with standard protein modeling software (27). An intrinsic advantage of cryo-EM is its ability to probe macromolecular structures of proteins that are rapidly frozen in vitreous ice preserving near-native conformations without requiring crystallization (28). Therefore, specimens can be compared under different physical-chemical conditions to probe their structural response, e.g., the temperature dependence of enzyme catalysis (29). This is particularly important when studying the solvation layer and its response to the concentration of hydronium ions in

Author contributions: F.H., I.S., J.B., F.L.K., G.K., and P.L.K. designed research; F.H., I.S., I.K.S., F.L.K., C.T., G.K., and P.L.K. performed research; F.H., I.S., F.L.K., J.M., and P.L.K. contributed new reagents/analytic tools; F.H., I.S., I.K.S., J.B., D.A.S., C.T., and P.L.K. analyzed data; P.L.K. conceived the project, supervised the project; and F.H., I.S., G.K., and P.L.K. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]F.H., I.S., and I.K.S. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: panagiotis.kastritis@bct.uni-halle.de.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2525949123/-/DCSupplemental.

Published March 5, 2026.

aqueous macromolecular solutions. Indeed, p*H* dependency of proteins has been well known to have critical functional implications (30), driving conformational changes (31), dynamics (32, 33), and, ultimately, cellular function (34). At the molecular level, p*H* can affect the hydration layers of macromolecules (35), and importantly, local pKa of crucial residues due to localized protonation changes. Minuscule changes might render an atom from a nucleophile to an electrophile, and thereby allowing certain biochemical reactions to occur. Kunz and Kauzmann (35) observed via ultracentrifugation, slight density increases in protein sedimentation experiments as a function of decreasing p*H* and interpreted them as changes in the amount of bound water resulting from changes in the ionization state of protein carboxyl groups.

Despite decades of research on the effect of p*H* on bound water with an array of biophysical (36–38) and biocomputing (39) methods, experimental observation of individual water molecules and their dependence on p*H* has yet to be probed. Here, we have undertaken this task and analyzed waters in the structure of human apoferritin, a biomolecule that can withstand such drastic changes in p*H* (40), at varying hydronium ion concentrations. Apoferritin was selected as an example because of its highly stable nature (41), making it an ideal specimen for cryo-EM (42, 43), previous experience of the laboratory in its high-resolution structure determination (44), and excellent tolerance to a wide range of p*H* (40).

Results

Recovery of High-Resolution Structures at p*H* 3.5 to 9.0. We overexpressed the human apoferritin gene construct and purified the protein complex (*Materials and Methods* and *SI Appendix*, Fig. S1 *A* and *B*), recovering a monodisperse peak of the molecule in size exclusion chromatography at an expected molecular weight of 450 kDa (*SI Appendix*, Fig. S1*C*). Before vitrification, we optimized the p*H* of the retrieved aqueous solution using a broad range buffer series (45). The selected range ensured that the structural effects to be observed are not affected by any distinct composition of the aqueous solution studied and are only effects of the p*H* of the aqueous solution. Cryo-EM screening showed that the molecule was stable in the p*H* range of 3.5 to 9.0 (*SI Appendix*, Fig. S2*A*), as, at p*H* ~3, the molecule dissociates (46) (*SI Appendix*, Fig. S2 *B–D*). After data collection (*SI Appendix*, Table S1), image processing (*SI Appendix*, Fig. S3), and subsequent refinement procedures (*SI Appendix*, Table S1 and *Materials and Methods*), we were able to reconstruct cryo-EM maps of human apoferritin at p*H* 3.5, 4.0, 5.0, 7.0, and 9.0 at resolutions around 2 Å (*SI Appendix*, Fig. S3 and Table S1). Water molecule densities in all modeled structures were highly visible while accurate modeling of water molecules, cofactors, and side chains (*SI Appendix*, Fig. S4) was feasible at this resolution.

Cryo-EM maps did not show meaningful p*H*-dependent size deviations as previously reported by atomic force microscopy (47), in agreement with low-resolution small angle X-ray scattering data (40). Instead, at varying p*H* conditions, apoferritin suspended in vitreous ice retains its expected molecular structure and higher-order structure. Specifically, the protein backbone remained stable, with Cα RMS-deviation (RMSD) values relative to p*H* 7 being small, i.e., 0.75 Å at p*H* 3.5, 0.49 Å at p*H* 4, 0.62 Å at p*H* 5, and 0.32 Å at p*H* 9, corroborating that changes in nonpolypeptide components are not due to large-scale protein rearrangements. To investigate the structural stability and dynamics of apoferritin at different p*H* values at the atomic level, we conducted all-atom molecular dynamics (MD) simulations. We simulated the varying p*H* conditions by adjusting the charge state of protonatable residues (Glu, Asp, Lys, His) according to the pKa values of their side chain functional groups, which we determined using p*H* titration calculations (*Materials and Methods* and *SI Appendix*, Fig. S5). In agreement with the cryo-EM data, the apoferritin complex remained close to the starting structure in the MD simulations conducted at p*H* values of 5.0 and above. However, with the side chain protonations adjusted to mimic p*H* values of 3.5 and 4, the structure deviated more strongly from the starting coordinates (backbone RMSD of ~3 Å, *SI Appendix*, Fig. S6), indicative of an increased instability of the apoferritin complex at low p*H* values. This structural deviation at lower p*H* likely reflects residues with pKa values slightly above 4. In our simulations, these residues were modeled as fully protonated, leading to changed protein solvation and intermolecular interactions, whereas under the cryo-EM conditions, these residues may remain partially deprotonated or constitute a minority of the apoferritin molecules that did not end up in the final reconstruction during processing.

Distinct *pH* Dependency of Bound Water Molecules and Metals. The modeled structures within the high-resolution cryo-EM maps provided us with a series of hypotheses. First, we wanted to investigate if modeled water molecules exhibited conformational variation higher than that of amino acid residues or bound metals, as it is known from NMR systematic investigations that water reorientation and translation at the protein surface is 3 to 5 times slower than in bulk (2). To address this, we estimated the per-atom resolvability in all five cryo-EM maps. Results show that the resolvability of atoms, calculated by their Q-scores and translated into B-factors (48), is comparable for all distinct chemical species, irrespective of the p*H* in which apoferritin was buffered (Q-scores > 0.8, Fig. 1*A*). This exercise corroborates the observation that resolvability and confidence for the positions of water molecules are not compromised as a function of p*H* or slight differences (~0.1 Å) around 2 Å resolution.

Next, we studied the iron ion coordination in the resolved structures for the different p*H* conditions. Apoferritin is a ferroxidase that converts Fe(II) to Fe(III), and internalizes the iron in its core; however, the transfer of iron at the atomic level is not well understood (49). Previously resolved structures show a network of histidines organizing the iron (50). In our structures, we have unambiguously determined an iron ion at the iron entry channels, corresponding to 8 bound iron ions in total (Fig. 1*B*). Because we resolved the molecule at different p*H* values, we sought to directly observe the localization of the iron and its coordination as a function of decreasing p*H*. This may reflect the general principle that metal solubility increases at lower p*H*, as protons compete with metal ions for coordination sites. It is also critically connected to cellular iron homeostasis (51), and overall, important for iron homeostasis, e.g., in leukemias (52) or neurodegeneration (53).

We observe that with decreasing p*H*, the iron is displaced toward the geometric center of the entire protein complex (Fig. 1 *C–F*). This behavior is highly specific for the iron ion and not for the structural magnesium (*SI Appendix*, Fig. S7 *A–C*). This discrete behavior of bound metals as a function of p*H* underlines the fine-tuning of metal local coordination and further exchange with the bulk solvent. We then specifically asked if these metals were also further coordinated by stable water molecules in their proximity. Stable water molecules are often identified in metal-binding sites of proteins and are directly related to function (54). Structural water is present in proximity to the iron (Fig. 1*D*) but does not persist with increasing p*H*. Whereas at p*H* = 5.0 or lower, a water molecule appears in the proximity of the iron ion, coordinated by four histidine residues, this interaction disappears at a higher p*H* (Fig. 1*D*). Therefore, we identified at least one example in the

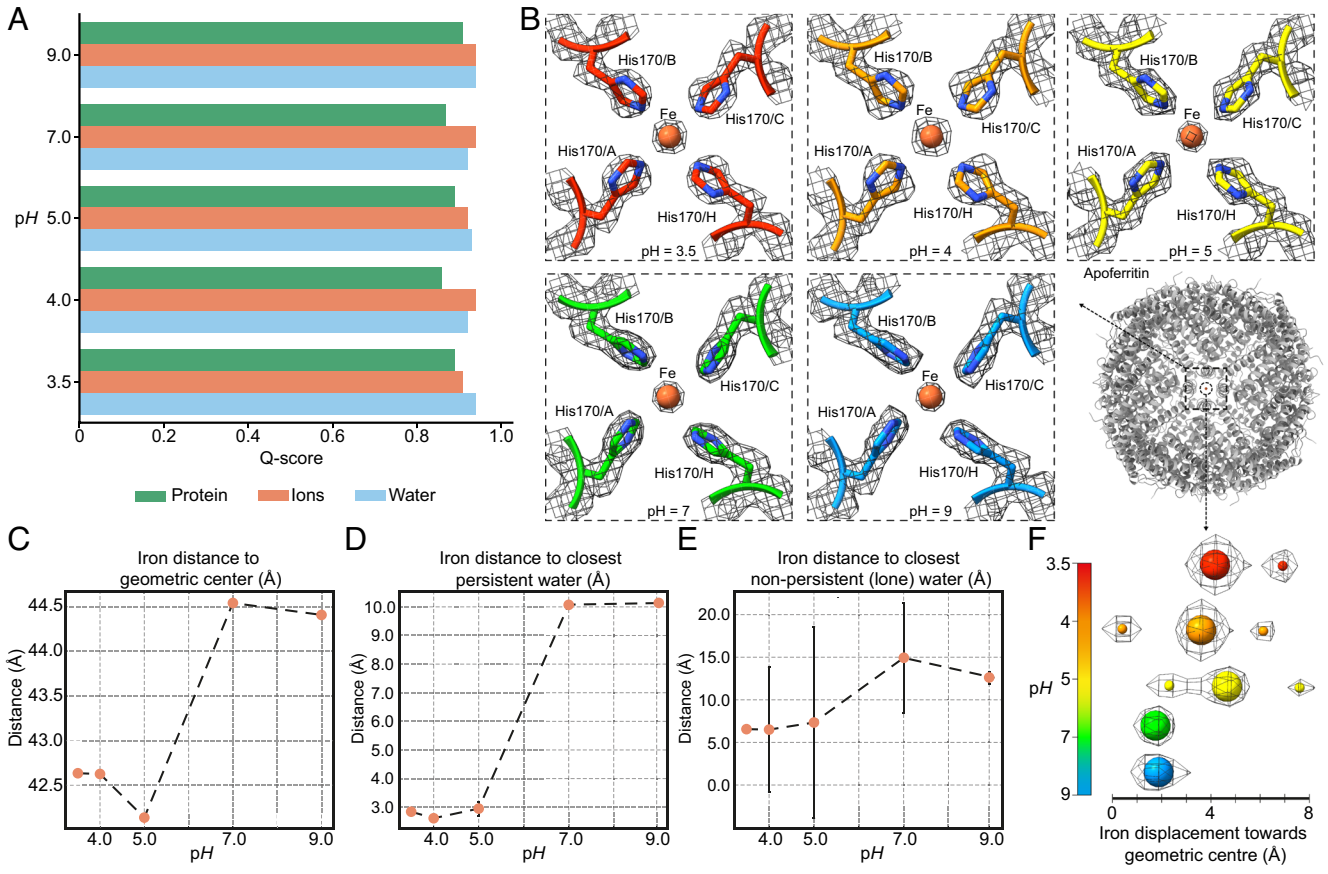

Fig. 1. Analysis of atom resolvability and iron positioning in apoferritin cryo-EM maps determined at varying p*H*. (*A*) High Q-scores are calculated for the resolved Coulomb potential maps corresponding to residues (green), bound ions (brown), and water molecules, (blue) at all p*H* values. (*B*) The iron is resolved in all cryo-EM maps, coordinated by histidines which show minimal side-chain conformational changes. (*C*) Calculation of distances from the apoferritin geometric center to individual iron ions in cryo-EM reconstructions. (*D*) Distances of persistent and (*E*) nonpersistent (lone) water molecules in proximity to the iron ion. Persistent correspond to water molecules present in two or more p*H* values, nonpersistent are not. Only persistent water molecules are identified in proximity to the iron, with their distances getting shorter with decreasing p*H* (at p*H* 5 and below, distances correspond to those of a hydrogen bond). (*F*) Relative positional change of the bound iron in the channel sites. Proximal water molecules appear at p*H* values of 5. The scale bar represents colors corresponding to the different p*H*, similar to panel (*B*) where residues organizing the iron are also shown.

high-resolution apoferritin structure where p*H* effects can directly drive the solvation of its iron channel site combined with a correlated inward movement of the iron (Fig. 1 *C* and *F*) mediated by acid-forming water-metal hydrogen bonds (Fig. 1*D* and *SI Appendix*, Fig. S8).

p*H*-Dependent Protein Desolvation. Kuntz and Kauzmann predicted, based on sedimentation data at different p*H*, that biomolecules are desolvating with decreasing p*H* (35). Due to the high-resolution structures that we determined at different p*H*, we can validate or reject this prediction. After modeling bound water molecules in all five cryo-EM maps (*SI Appendix*, Fig. S4 and Table S1) and unambiguously showing comparable resolvability with amino acid residues and metals (Fig. 1*A*), we visualized the distribution of the water molecules around the resolved structures (Fig. 2*A*). We then counted the bound water molecules in the complete 24-mer structures (*SI Appendix*, Table S1). An apparent decrease in the number of bound water molecules upon acidification can be observed (Fig. 2*B*). A trend is revealed where ~100 bound water molecules are released per *pH* unit (R ≈ 0.97). This trend can be well reproduced when analyzing the number of water binding sites in the corresponding MD simulations (Fig. 2*B*). When the p*H* is lowered from 9 to 3.5, the number of resolved water molecules within a distance of 5 Å from the protein surface decreased by almost 50%.

This striking observation can effectively rationalize the origins of acidic p*H*-induced dissociation at p*H* < 3.5 (Fig. 2*C*). Effectively, the exchange of bound water to bulk is gradually increased up to a threshold; above this threshold, at p*H* < 3.5, solvation of the 24-mer can no longer be stabilized once the water molecules nearest to its surface become too disordered to maintain effective solvation, eventually dissociating the higher-order structure to its more stable monomers (Fig. 2*B*). Next, we asked if the resolved bound water molecules exhibit similar distance deviations as those determined for the iron ion (Fig. 1*B*). Results show that distances of water molecules follow a complex behavior: At low p*H*, they have a less canonical distance to their nearby atoms as shown by the flattening of their distance distribution (*SI Appendix*, Fig. S9). In contrast, in higher p*H*, equivalent distribution is tighter (*SI Appendix*, Fig. S9) consistent with faster water motions as observed before (55). This observation also aligns with Kuntz and Kautzmann (35) who hypothesized that protein hydration reflects an ensemble of bound water states with varying mobilities rather than a single homogeneous population. It is also important to take into account changes in protein flexibility as a function of p*H*, which could affect or even orchestrate protein hydration. Even if we do not observe major conformational changes of apoferritin across p*H*, minor populations might still exist, as also shown by the MD simulations at acidic p*H*. Furthermore, studies on human serum albumin that populates at least five states as a function of

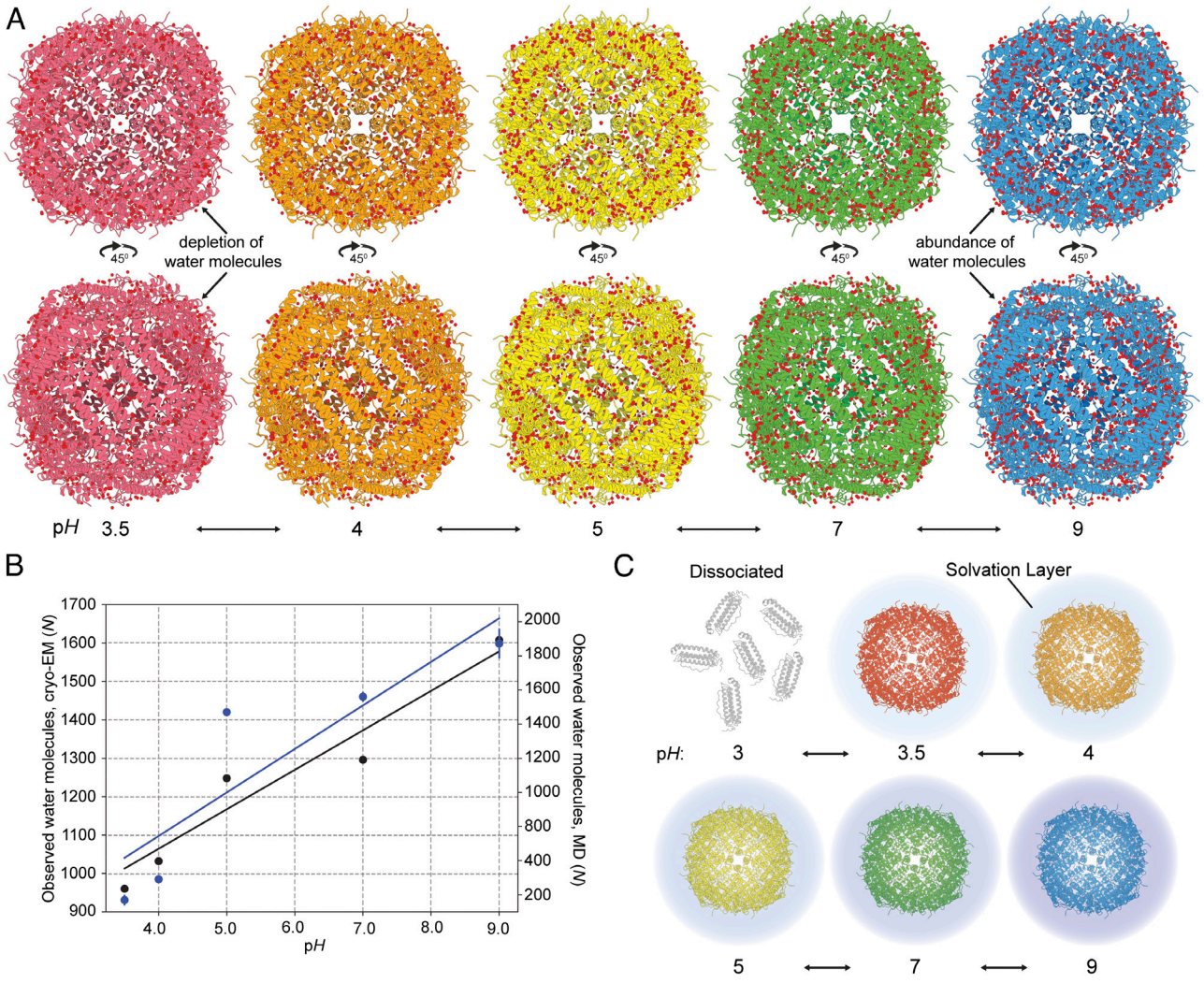

Fig. 2. Solvation layer density and localization across p*H*. (*A*) Distribution of water molecules around the resolved apoferritin structures. The embedded ions have been hidden from the representation to show water density. The water molecule at the iron binding site (*Middle* of structures, *Upper* panel part) appears at pH 5, 4, and 3.5 (see also Fig. 1*D*). (*B*) Total number of surface-bound water molecules determined from the cryo-EM maps (black points) or by running MD simulation (blue points), respectively, as a function of p*H*. A clear upward trend is visible with increasing p*H*. (*C*) Graphical representation of solvation layer density. At p*H* 3, there is no complex formation, whereas with increasing p*H*, the density of the visible water solvation layer increases.

p*H* (ranging from an extended state at p*H* 2.7 to an "aged" state at p*H* 10) showed discrete hydration dynamics for each state (55), in agreement with our observation (*SI Appendix*, Fig. S9). Our study establishes that acidic p*H*-dependent protein dissociation is coupled to progressive water layer flexibility and redistribution of bound water molecules, hydration being the key determinant of structural stability.

Side-Chain Specific Contributions to Bound Water Molecule Retention. Protein solvation is influenced by global p*H* and residue-specific protonation states (56). We therefore asked whether the observed changes in hydration arise from p*H*-dependent shifts in residue charge. To investigate if the p*H*-dependent hydration differences in the structure are linked to specific residue types, we calculated per-residue bound water molecule counts as a function of p*H* and compared these values with the per-residue hydration site numbers obtained from the MD simulations. This analysis showed that certain residue types get considerably desolvated with decreasing p*H* while others show only a small or no significant decrease of their number of bound water molecules (Fig. 3*A*). Specifically, glutamate and aspartate side chains have the largest number of bound waters of all amino acids at p*H* 9 and show the most pronounced desolvation with decreasing p*H*. Interestingly, this effect is also observed for basic residues (lysine and arginine) and to a lesser extent for histidine and polar residues (glutamine, asparagine, threonine, serine, tyrosine) (Fig. 3*A*). As expected, residues with apolar side chains (alanine, valine, leucine, isoleucine, phenylalanine, tryptophan, methionine, proline) or glycine show the smallest bound water numbers and least p*H*-dependent desolvation. Although low-resolution methods such as X-ray photoelectron spectroscopy have previously described p*H*-dependent desolvation effects and related them to changes in the residues' charge state and hydrogen bonding capability with water (57), we here directly visualize and quantify this effect at an atomic level.

To evaluate if p*H*-dependent desolvation effects observed are generalizable, we correlated identified bound waters with total accessible surface area (Fig. 3 *B* and *C* and *SI Appendix*, Fig. S10*A*) and hydrophilicity (*SI Appendix*, Fig. S10*B*) for each amino acid, deriving strong correlations at neutral and alkaline p*H* (R^2 up to 0.75) (Fig. 3*B*). At these conditions, overall hydration is largely dictated by surface area and polarity (*SI Appendix*, Fig. S10*B*), as expected (58, 59). However, at lower p*H*, correlations weakened markedly (R^2 as low as 0.24), reflecting a gradual breakdown of

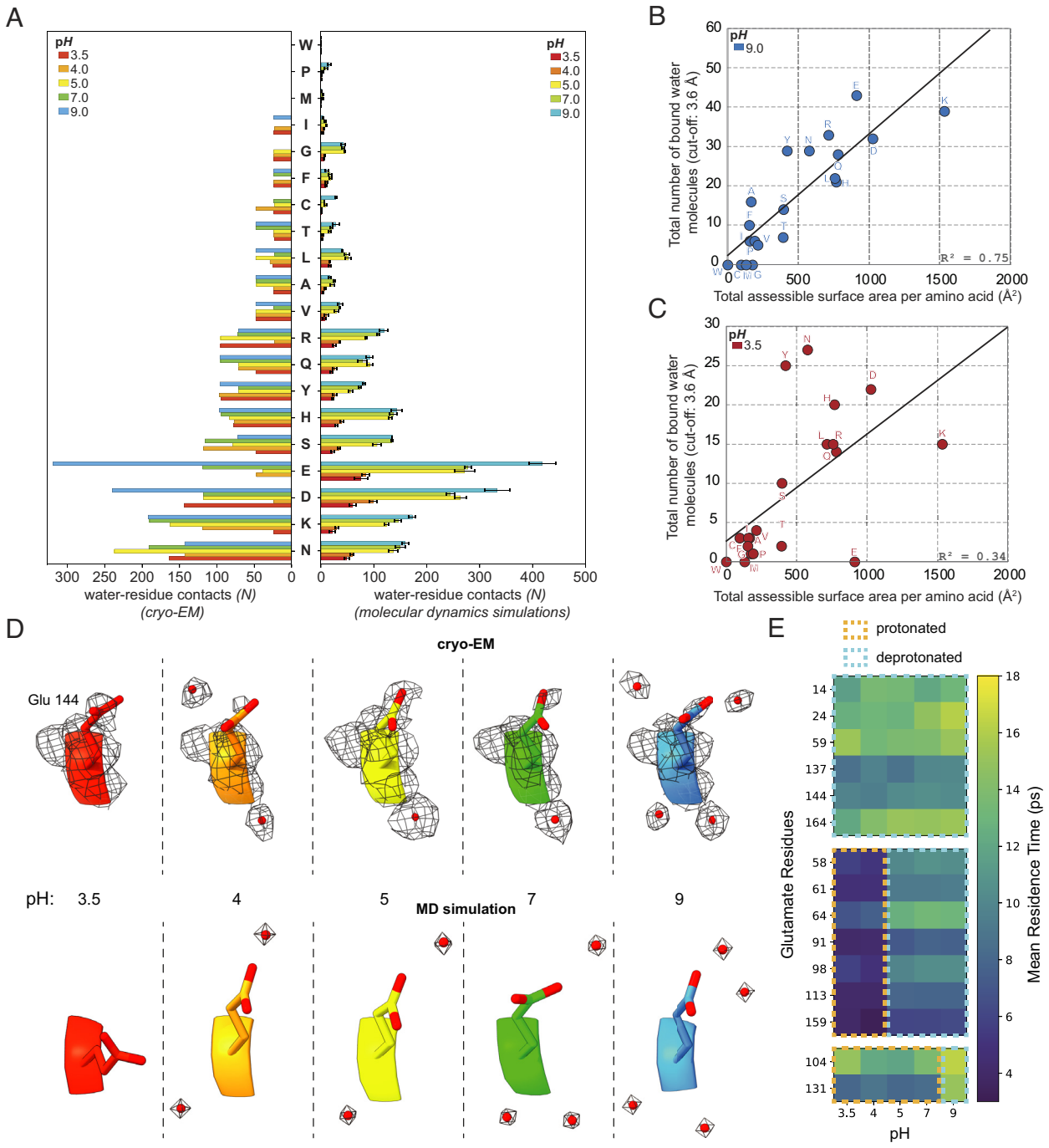

Fig. 3. Per-residue bound water molecule counts at different p*H* values. (*A*) Total modeled water count per p*H* in proximity to each amino acid type in the cryo-EM maps (*Left*) and in the MD simulations (*Right*). (*B* and *C*) p*H*-dependent relationship between amino acid surface area and bound water molecules in apoferritin. Amino acid-specific counts of bound water molecules within 3.6 Å are plotted against their total accessible surface area, for each residue. Each point corresponds to all residues of a given amino acid type within the apoferritin structure. Correlations are strong, but become increasingly weaker at increasingly lower p*H* (*SI Appendix*, Fig. S10). Acidic and charged residues display a decreased water count in their proximity compared to other residue types. (*D* and *E*) Residence times and positions of waters surrounding glutamates. (*D*) Visualization of Glu144 and its surrounding water molecules across p*H* in the cryo-EM maps and MD simulations. (*E*) Residence times of waters surrounding the glutamate residues at different simulated p*H* values.

residue-specific hydration (Fig. 3*C* and *SI Appendix*, Fig. S10*A*). Acidic side chains lost most bound waters upon protonation, whereas polar residues retained waters. These results show that protein hydration is organized and residue type-dependent at high or neutral p*H*, but collapses under increasingly acidic conditions, where many waters are displaced to bulk solvent. To investigate how the persistency of water molecules in the macromolecular structure is affected by p*H* changes, we calculated the occupancy and residence time for the water binding sites that interacted with the amino acid residues in the MD simulations. The most considerable changes in the water molecule persistence were observed for glutamate residues, which showed a progressive reduction in their bound water molecule number with decreasing p*H*. For example, Fig. 3*D* illustrates the water sites interacting with residue E144, with the water positions closely matching between cryo-EM structure and the MD simulation. While at p*H* 9, E144 is interacting with 4 water molecules, it is bound to only 2 at p*H* 4.0 and is entirely desolvated at p*H* 3.5. The water residence times

of glutamates also vary substantially from 4 ps up to 18 ps (Fig. 3*E*). We noted three groups of glutamate residues, demonstrating distinct water stabilities. The first group of glutamates has low pKa values of 2 to 4, as determined by constant p*H* MD simulations (*Materials and Methods* and *SI Appendix*, Fig. S5), and are deprotonated at all p*H* values. These residues show high residence times of 10 ps or higher, which vary minimally between different p*H* levels, reflecting stable water binding. The second group of glutamates has pKa values between 4 and 5 (*SI Appendix*, Fig. S5). Consequently, these residues are protonated below p*H* 5 and deprotonated when the p*H* is 5 or higher. The change in the side-chain protonation state of these glutamates is accompanied by a jump of their bound water residence times from 4 ps at p*H* ≤ 4 to around 10 ps at p*H* ≥ 5, possibly due to stronger hydrogen bonding of their associated water molecules when the side chains get deprotonated. The third group includes two glutamate residues (E104, E131), which have exceptionally high pKa values (pKa_{E104} = 8.9 and pKa_{E131} = 7.9) and exist in a protonated state at p*H* ≤ 7. Their side-chain deprotonation at p*H* 9 leads to increased water residence times, reflecting again more stable water binding. In addition to glutamates, similar trends in the dynamics of bound water molecules, due to changes in the residues' pKa values and protonation states, can be observed also for aspartates (*SI Appendix*, Fig. S11).

Finally, to understand global retention propensities of bound water across p*H* values, we clustered water oxygens based on their distance (1.0 Å cutoff) across aligned modeled structures along with a nearest-neighbor residue assignment using a 3.6 Å cutoff. This yielded a definition of persistent waters as sites present at the same location in ≥2 p*H* values, numbering 1,052 ± 197 sites per p*H* (Fig. 4*A*). In contrast, nonpersistent (lone) waters, i.e., sites found only in one p*H* value showed an upward trend and amounted to 75 at p*H* 3.5, 177 at p*H* 4.0, 173 at p*H* 5.0, 139 at p*H* 7.0, and 333 at p*H* 9.0 (Fig. 4*B*). The most persistent waters (N = 499) were those observed across all five p*H* values (Fig. 4*C*), corresponding to 40.6% of all water molecules observed across all maps. Such a high percentage indicates a localized, well-defined hydration shell that acts as a molecular buffer against p*H*-dependent desolvation and contributes to structural stability, where mostly polar and positively charged residues retain those structural waters (Fig. 4*D*). Lone waters, on the other hand, predominantly appear at higher p*H* (Fig. 4*A*), consistent with total water counts. A visualization of water molecules across sites of differing persistence is shown in Fig. 4*E*, highlighting a comparison between persistent and nonpersistent hydration locations. Analysis of distances between water molecules further revealed that those that persist, cluster substantially closer together than less persistent or lone waters (Fig. 5 *A* and *B* and *SI Appendix*, Table S2). This distribution of distances

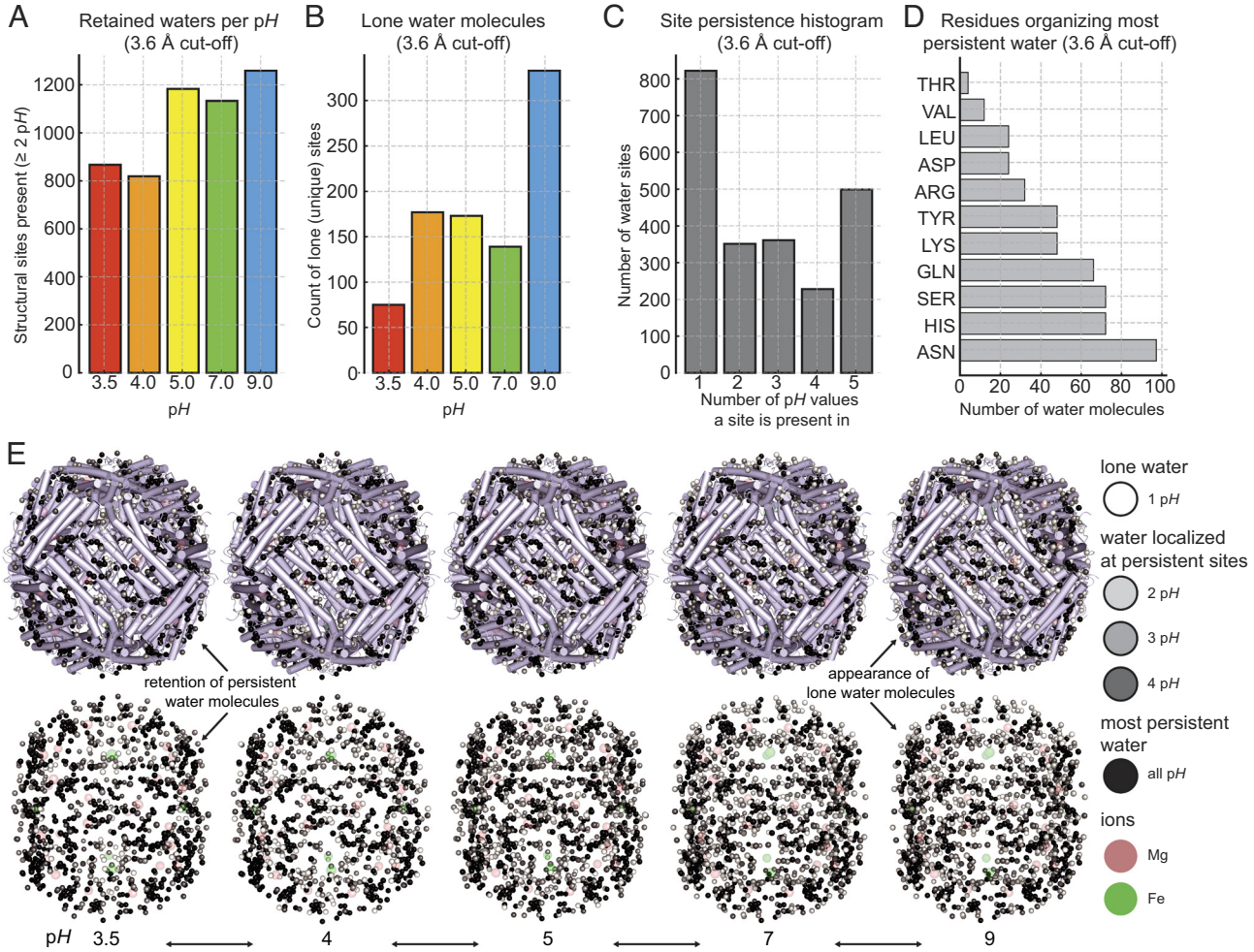

Fig. 4. Persistence of water molecules at their sites. (*A*) Retained water molecules per p*H* bin, found in at least one other p*H* map. (*B*) Water molecules per p*H* bin identified in only one map. (*C*) Number of water sites that are identified within a single p*H* map, two, three, or four maps, and across all five p*H* maps. (*D*) Water molecules organized across all maps (most persistent) amino acid preferences. (*E*) Visualization of the hydration layer across p*H*. Apoferritin is shown in cartoon representation, surrounded by water molecules classified by persistency, from white "lone waters" to black waters persistent across all p*H*. In the bottom, the hydration layer is shown alone for clarity.

was maintained across all p*H* conditions (*SI Appendix*, Fig. S12). Our observations imply that persistent water molecules are stabilized by forming water–water hydrogen bonds, in contrast to more isolated, weakly bound waters that are readily exchanged with the bulk solvent. Mapping of these sites on the apoferritin structure solved at p*H* 7 shows a trend where localized negative charge is not likely to organize persistent water molecules (Fig. 5*C* and *SI Appendix*, Fig. S13).

Discussion

Bound water molecules, their nature and adaptability have inspired biochemists, such as Linus Pauling, since the 1950s (60). No direct method was available to visualize their physical-chemical adaptations, such as those we report here. Cryo-EM can directly visualize samples of different physical-chemical conditions and interpret stable water molecules that interact with a given biomolecule due to the derived high-resolution. We have quantitatively expanded the 50-y-old hypothesis on protein dehydration as a function of decreasing p*H* (35) and directly visualized and quantified this effect on a biomolecular complex. During this effect, which we term *Kuntz-Kauzmann acid-driven protein desolvation*, the loss of waters leads to a gradual protein destabilization due to missing compensation of protein surface charges by waters (Fig. 6).

This effect is observed across the complete protein surface, and distances of water molecules to the protein surface show broader distributions upon acidification, indicating increased localized instability. Kuntz-Kauzmann acid-driven protein desolvation is more pronounced in proximity to specific side chains, especially negatively charged ones, but also in proximity to lysine but less to arginine and histidine; Lysine is known to change its solvation layer as a function of decreasing p*H* (57). However, these residues do not organize stable water molecules; in contrast, the asparagine residues, which are highly solvated across all p*H* values, show the highest degree of proximal structural water molecules that are also retained at their positions. A hypothesis is that their side-chain adaptability, having both partial positive and negative charges, may substantially enhance the localized attraction of stable water molecules.

It remains to be investigated how protonation states could potentially be determined using cryo-EM. As a Coulomb potential map is being calculated using cryo-EM, the charge information

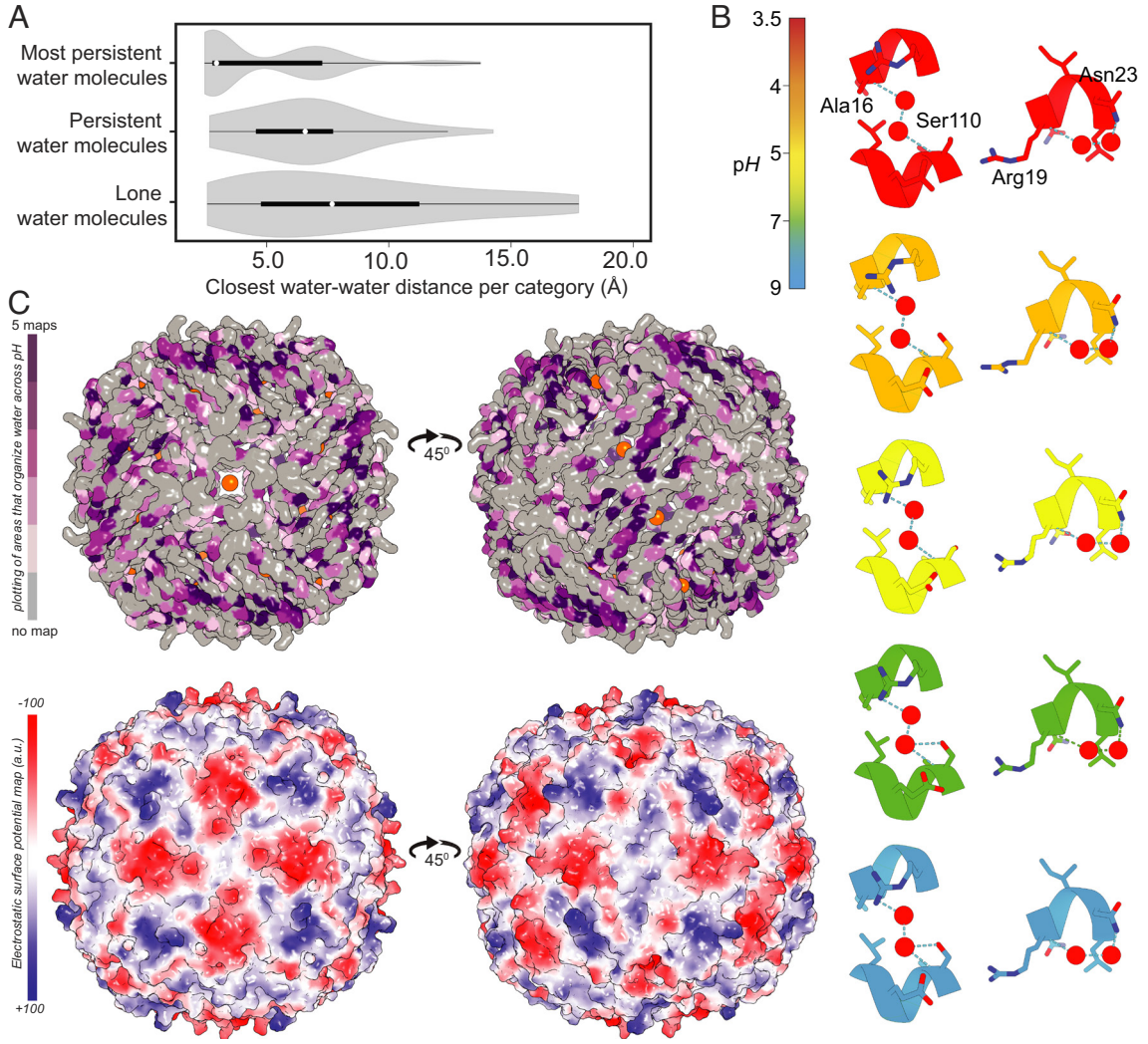

Fig. 5. Water–water distances across water molecule categories and relation of water positions to the electrostatic surface potential. (*A*) Calculated closest water–water distances across the three categories of water molecules, i.e., "lone" water found only in 1 p*H* map; persistent water, found at least in 2 different p*H* maps; and "most persistent," found across all resolved maps. Their Kolmogorov–Smirnov statistics can be found in *SI Appendix*, Table S2. (*B*) Two examples of water–water mediated hydrogen bonds localized around residues Arg19, Asn23, and Ser110 with Ala16. (*C*) On top: mapping of persistent sites for the water molecules on top of the biomacromolecule. On the bottom: electrostatic potential map calculated, showing the anticorrelation between persistent water and high electrostatic surface potential.

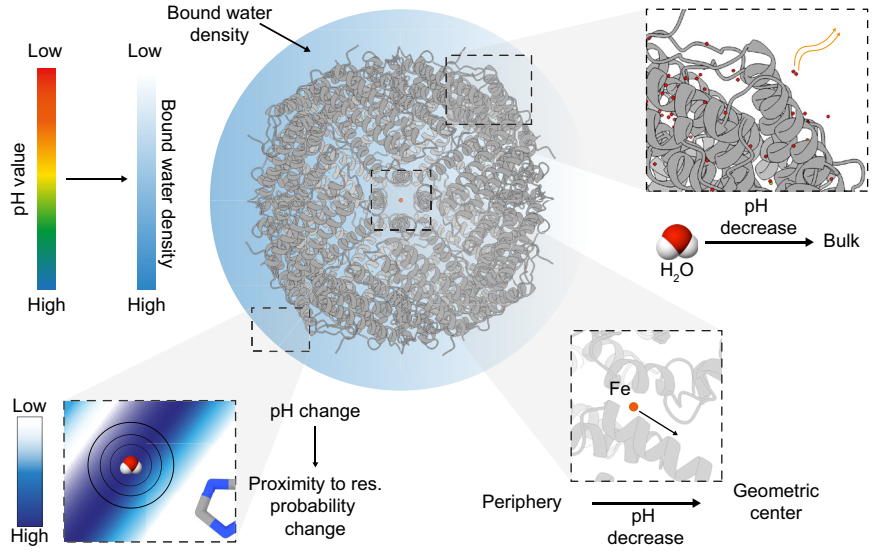

Fig. 6. Solvation layer effects as a function of p*H*. The modeled waters belonging to the solvation layer numerically differ with decreasing p*H*. Density of bound waters shows a decrease in accordance with p*H* decrease, moving toward the bulk and not contributing to protein complex stabilization, whereas the probability distribution of water molecule localization in proximity to the nearest amino acid side chain becomes narrower. Fe atoms are affected by water density, being pushed to the geometric center of the protein complex.

is residing in the data, even at relatively low resolutions (61, 62). For unambiguous visualization of residue protonation states, atomic resolution of approximately 1.2 Å or better is generally needed to resolve hydrogen atoms and accurately assess protonation states, which has indeed been achieved (42, 43), but is an exception rather than the rule. Most importantly, radiation damage prohibits accurate visualization of side chains, manifesting as site-specific damage at high resolution, even with exposures less than 1 $e^- Å^{-2}$, which are well below those applied using low dose acquisition protocols (63). Considering low exposure of initial frames, radiation damage could be further avoided by improvements in beam-induced motion correction, possibly by considering particle rotations and height changes (64), revealing potential protonation states.

Overall, we find that the p*H* dependence of resolved structural waters is described by two p*H*-responsive site classes that titrate around acidic and basic conditions, and a third p*H*-independent "persistent shell." The three classes constitute a coarse-grained envelope of proton-linked binding in Wyman's framework (65) where negative linkage explains loss of waters upon acidification. Additionally, our MD shows that deprotonation at acidic sites lengthens water residence times from a few picoseconds to ~10 ps, corresponding to ~0.5 kcal mol^{-1} per site. When considered across total hydration sites, values are compatible with local charge-regulation energetics (66).

The experimental analysis of hydration presented captures vitrified rather than bulk liquid water. An issue, then, remains to be addressed, on the nature of this water, especially close to protein surfaces. Plunge freezing of thin aqueous films yields vitreous ice within ≤200 µs, which is sufficient to preserve the near-native aqueous structure (67), because of their high cooling rates (~10^5-10^6 K s^{-1}) (68). Although the sample cools on a relevant timescale for protein motions (69), differential quenching could, in principle, bias the very last stages of shell rearrangements as water translates at least an order of magnitude faster than larger proteins. In our data, three observations argue that p*H*-dependent dehydration is relevant beyond an analysis of "frozen water": First, we independently visualize an effect postulated by sedimentation experiments (35). Second, we recover the same trend in both cryo-EM and constant-p*H* MD, despite the latter being performed in liquid water. Third, a positionally conserved hydration shell persists across p*H* conditions and displays dense packing; such properties are difficult to reconcile with vitrification-induced rearrangement alone. Vitrification is, thus, likely a fast snapshot that approximates the solution ensemble close to the protein but we also acknowledge that short-lived, highly mobile subpopulations are underrepresented by our methodology. Capturing such transient water molecules constitutes an opportunity for future work, e.g., by combining time-resolved vitrification and analysis pipelines to retain conformational and solvent heterogeneity of the system under investigation.

Here we report high-resolution analyses for a single hydrated protein complex, an approach that limits breadth but promotes depth. We visualize global effects of acidification, namely acid-driven protein desolvation together with the retention of a positionally resolved hydration layer. Coupled with the derived correlations between acidification and physical-chemical properties (e.g., residue types, solvent-accessible surface areas, protonation states) these findings are expected to extend to other systems. However, their quantitative magnitudes will vary with surface topology, oligomeric state, intrinsic conformational dynamics, posttranslational modifications, and environment (e.g., membrane-embedded vs. soluble proteins). Assessing universality and system-specific deviations will require extending our analysis to additional proteins and environments by benchmarking the water layer behavior as function of p*H*. One other concrete future direction can follow our observation in apoferritin function, that of acid-driven inward displacement of the bound iron accompanied by local water remodeling. This result suggests that related, but not necessarily identical, responses in other metalloproteins exist, where acid-induced metal release may depend on specific interplay of local effects that include hydration, coordination geometry, amino acid residue dynamics and redox states.

Conclusively, our observations and experimental set-up utilizing broad range buffer series for the study of biological macromolecules in variable p*H*, combined with MD simulations to get insight into residence times, accelerates understanding of solvation layer effects on biological macromolecules. Our results can drive

applications in lead compound development for active sites (e.g., the observation of iron displacement with decreasing p*H*) or allosteric sites where water molecules play an active structural/functional role. Finally, there is an increasing demand for biotechnological protein applications at low p*H* [e.g., extraction of metals or rare earth elements (70)]. Our statistical observations based on the high-resolution structures of a protein complex can also act as a lead for protein design of highly soluble and stable proteins, informing the current era of machine-learning-based protein design (71).

Materials and Methods

Protein Expression and Purification. The genomic sequence for human apoferritin was present in a pGEX2T vector, containing a N-terminal GST-fusion and modified to contain a TEV cleavage site, instead of a thrombin cleavage site, resulting in Plasmid LF2422. This plasmid was kindly provided by Louise Fairall and Christos Savva (Protex facility at the University of Leicester). Protein expression and purification was achieved, following a modified protocol from ref. 72. Briefly overexpression of human apoferritin was achieved using *Escherichia coli* BL21 Rosetta™ 2(DE3) (Novagen) transformed with plasmid LF2422. Cells were grown at 37 °C until A600 = 1.0 before adding 1 mM 1-thio-b-d-galactopyranoside (IPTG) and shifting the temperature to 20 °C. After growth overnight, cells were harvested by centrifugation (2,967 × g) at 4 °C, for 30 min. The pellet was resuspended in 100 mL/L (bacterial culture) of phosphate-buffered saline (PBS), containing 0.5% Triton X-100 and 0.5 mM tris(2-carboxyethyl) phosphine (TCEP) and Roche Complete protease-inhibitors. Cells were lysed using a continuous cell disruptor (Constant Systems), and the lysate was centrifuged at >100,000 × g, 4 °C for 1 h. Cleared lysate was initially purified using a 5 mL GSTrap HP column (Cytiva). After binding and washing the column with 5 CV of washing buffer (1 × PBS, 0.5% Triton X-100, 0.5% TCEP) a buffer exchange to the SEC buffer (5 CV, 50 mM Tris, p*H* 7.5, 100 mM NaCl, 0.5 mM TCEP) and on-column digestion of the bound protein was performed with 100 μg self-prepared TEV-protease. After flushing the cleaved protein from the column, the bound GST-tag and any noncleaved protein were eluted from the column using the elution buffer (50 mM Tris, p*H* 7.5, 100 mM NaCl, 0.5 mM TCEP, 10 mM reduced Glutathione). Cleaved apoferritin was then further purified using a Phenomenex BioSep SEC-s4000. After size-exclusion chromatography, the peak with fractions containing fully assembled apoferritin (MW: 450 kDa) was pooled.

p*H* Change, Cryopreparation, Vitrification, and Data Acquisition. The pool was concentrated to about 500 μL using a Millipore 100 kDa cutoff Amicon tube. The buffer-exchange and further concentration was performed using Millipore 100 kDa cutoff tube filters. Buffers for p*H*-change were prepared, following the procedure described elsewhere (45). In brief, buffers were prepared by mixing two stock solutions in different volumes. Stock solution A contained 200 mM boric acid and 50 mM citric acid. Stock solution B contained 100 mM Na_3PO_4. The relative volumes of the two stock solutions were adjusted to obtain the desired p*H* values. This mixed weak-acid buffer system provides continuous buffering capacity over a broad p*H* range due to overlapping pK values of the constituent acids. Also, the chemical composition of the buffer remains constant, with p*H* being the only variable between samples. Therefore, confounding effects arising from buffer-specific interactions with the protein are minimized and structural differences observed among apoferritin reconstructions can be solely attributed to p*H*-dependent effects and not changes in buffer chemistry. Surplus protein sample was frozen in liquid nitrogen and stored at −80 °C. All different p*H* samples (3.5, 4, 5, 7, 9) were diluted to 2 mg/mL prior to vitrification. For cryo-EM sample preparation, R2.1 200-mesh holey carbon grids (Quantifoil Micro Tools) were glow-discharged using a PELCO easiGlow system (15 mA, 25 s, 0.4 mbar residual air pressure). Approximately 3.5 μL of each sample was applied to the grids in a Vitrobot Mark IV (Thermo Fisher Scientific) operated at 100% relative humidity and 4 °C. Grids were blotted for 4 s using 595-grade ashless filter paper (Thermo Fisher Scientific) with blot force 0 and immediately plunge-frozen into liquid ethane. Vitrified grids were transferred under cryogenic conditions to our 200 keV Thermo Fisher Scientific Glacios transmission electron microscope equipped with a Falcon IIIEC direct electron detector. For each of the five p*H* conditions, ~2,000 to 4,000 movies were recorded (*SI Appendix*, Table S1) at a calibrated pixel size of 0.59 Å/pixel, with a total accumulated electron dose of 30 $e^-/Å^2$ fractionated across 30 frames.

Image Processing of Cryo-EM Data, Model Building, and Refinement. All image processing steps were performed within the framework of SCIPION v3.1.0 (73). In detail, each of the collected datasets (p*H* 3.5: 3,725 images, p*H* 4: 2,308 images, p*H* 5: 2,541 images, p*H* 7: 1,891 images, p*H* 9: 1,867 images) was imported, corrected for beam-induced motion with RELION motioncorr (74), and CTF parameters were calculated with gctf (75). Manual particle picking of ~1,000 particles followed, and this particle set was used for TOPAZ training and picking (76). In parallel, picking was also performed with xmipp3 (77), as well as crYOLO (78) and a consensus pick was derived from all different picking methods. After combining the multiple picking strategies, each dataset yielded an initial particle set of (p*H* 3.5: 710,073 particles, p*H* 4: 280,270 particles, p*H* 5: 521,922 particles, p*H* 7: 529,107 particles, p*H* 9: 438,640 particles) which was subsequently refined by iterative 2D classification with 300 classes in cryoSPARC v3.2, leading to final particle sets of (pH 3.5: 217,187 particles, p*H* 4: 267,457 particles, p*H* 5: 421,029 particles, p*H* 7: 432,247 particles, p*H* 9: 343,262 particles). The refined particle sets were employed in RELION-4 (79) 3D initial modeling, followed by cryoSPARC 3D nonuniform refinement of the ab-initio models, taking into consideration per-particle motion and CTF parameters, and then the final particle sets and volumes were imported again to RELION-4 (79) for Bayesian polishing, leading to final reconstructions of 1.99 Å, 2.06 Å, 1.96 Å, 2.08 Å, and 1.89 Å (Gold-standard FSC = 0.143) for p*H* 3.5, 4, 5, 7, and 9 respectively.

The amino acid sequence of the *H. sapiens* ApoF heavy chain (Uniprot id: P02794) monomer was fit into each of the different p*H* EM reconstructions using ChimeraX (80), and then real space refined in PHENIX (27). After automatic refinement, each ApoF monomer model was manually inspected and side-chain fits were refined in COOT v0.9.2 (81). Modeling of waters was performed in two steps. First, each p*H* map and its corresponding model were inspected in COOT and waters were manually modeled, taking into consideration physical-chemical restraints. Simultaneously, each different p*H* model-map combination was used as input in the DOUSE module of PHENIX, for automated water placement in bulk, with cutoffs from 2 to 4 Å. Subsequently, the monomer density with the added waters was once again manually inspected in COOT, with a uniform sigma level (0.5), and water placement was refined manually as needed, comparing automated and manual water placement in the models. For the generation of the complete ApoF complex, the find_symmetry and apply_NCS_symmetry modules in PHENIX were used, and therefore, octahedral (O) symmetry was applied. Finally, all waters were assigned a chain by the closest Euclidean distance to all protein atoms. Water molecules, as well as iron ions, with multiple identical distances and identical atom indices–indicating location within a symmetry axis–were assigned an appropriate occupancy value (0.25 for fourfold axis, 0.33 for threefold axis, 0.5 for twofold axis). These ligands, as well as molecules with multiple identical distances, but different closest atom indices, were assigned the alphabetically first chain in the set of interactors. Further explanations are provided in *SI Appendix*, Fig. S14. Eventually, in this fully annotated biological set, only chain A was isolated, representing the asymmetric unit of the protein complex.

Constant p*H* (Cp*H*) MD Simulations. The cryo-EM structures of apoferritin, obtained at p*H* 4, 7, and 9, were used as starting models. For each structure, seven neighboring protomers were extracted to build a minimal system for running CpHMD simulations (*SI Appendix*, Fig. S15). The phbuilder tool (82) was used to prepare the CpHMD simulations for the GROMACS software. Titration was allowed only for the glutamate, aspartate, histidine, and lysine residues of the middle monomer as well as for residues of neighboring monomers that can directly interact with the middle monomer. Residues coordinating the iron(II) and magnesium(II) ions were kept at their deprotonated state (for E24, E59) or neutral state (for HID62, HIE170) in order to retain the binding site of these ions. Further information about the titrated residues for each monomer can be found in *SI Appendix*, Table S3. The system was solvated in a box with a border of a minimum distance of 1 nm to the protein and neutralized with NaCl and buffer molecules. The topology was generated using the CHARMM36-mar2019-cphmd force field (83) in combination with the CHARMMff additive force field for iron(II) (84) and TIP3P water model. After minimization using the steepest descent algorithm, the system was further equilibrated in an NVT ensemble and subsequently in an NPT ensemble, while gradually releasing restraints on the heavy atoms. Both equilibration steps were run

for 10 ps using a time step of 1 fs with the temperature regulated by the V-rescale thermostat at 310 K and the pressure kept at 1 bar using the C-rescale barostat, respectively. The PME method was used to perform electrostatics, and the LINCS-algorithm was used to constrain bond lengths. Titration MD runs were performed for 40 ns with a timestep of 2 fs using the GROMACS CpHMD beta software, obtained from https://gitlab.com/gromacs-constantph/constantph. A position restraint of 50 kJ mol^{-1} nm^{-2} was applied to all backbone heavy atoms of the six outer monomers and the iron(II) ion in all simulations. The starting titration states were chosen according to the standard pKa values of the residues and the titration was performed in steps of 0.5 from p*H* 1.0 to p*H* 10.5. Titration curves were fitted using the Scipy module (85) after calculating the protonation fractions according to Eq. **1** with n_{prot} including all lambda values smaller than 0.2 and n_{deprot} higher than 0.8 using the accompanying script from the GROMACS constant-p*H* Gitlab repository (https://gitlab.com/gromacs-constantph). A summary of this MD simulation setup can be found in *SI Appendix*, Table S4.

$$S_{deprot} = \frac{n_{prot}}{(n_{prot} + n_{deprot})}. \quad [1]$$

Whole-Assembly Apoferritin MD Simulations. For all determined cryo-EM structures (p*H* 3.5, p*H* 4, p*H* 5, p*H* 7, p*H* 9), MD simulations of the whole 24-mer assembly were run in triplicate with the protonation states of titratable residues adjusted according to the results of the CpHMD simulation. This time, titration states were not allowed to change within one simulation. The protein structures were placed in a cubic box with 1 nm distance between protein and box and subsequently solvated with TIP3P water. NaCl was used to neutralize the system and to ensure an ionic strength of 150 mM. The minimization and equilibration steps were performed as described above, for 100 ps in both NVT and NPT ensembles. Production runs were conducted for 100 ns in triplicates each using GROMACS version 2021 (https://doi.org/10.5281/zenodo.4457626). A summary of this MD simulation setup can be found in *SI Appendix*, Table S4.

Hydration Site and Residence Time Analysis. For the analysis, 500 consecutive MD frames (corresponding to 1-ns sampling window) were extracted at regular 10-ns intervals throughout the 100-ns production trajectory. Specifically, frames were sampled at 10 ns, 20 ns, 30 ns, 40 ns, 50 ns, and so on up to 100 ns and the obtained residence times were afterward averaged over all 10 samples. Rotational and translational movements of the protein were removed by aligning the whole trajectory to the apoferritin structure in the first frame. A water density map of the regions up to 5 Å around the protein with a grid spacing of 0.5 Å was calculated using the MDAnalysis package (86). This map was smoothed by calculating the mean water density of each grid cell and its adjacent grid cells that were sharing an edge. Close density peaks with a distance smaller than 1 Å between them were averaged using the DBscan algorithm (87) implemented in the Sklearn package (88). Hydration sites were counted if they had a smoothed density of 6 kg/L or more. Closest residues were defined as the residues that most often had the closest heavy atom distance to the hydration site. The water residence times for all glutamates were calculated by integrating the survival probability function of water 3.6 Å around each respective glutamate residue using MDAnalysis (86).

Data, Materials, and Software Availability. Cryo-EM maps and models were deposited in EMDB and PDB, respectively, and correspond to the following entries: p*H* 3.5: EMD-54951 (89), 9SJR (90); p*H* 4.0: EMD-54952 (91), 9SJS (92); p*H* 5.0: EMD-54953 (93), 9SJT (94); p*H* 7.0: EMD-54954 (95), 9SJU (96); p*H* 9.0: EMD-54955 (97), 9SJV (98).

ACKNOWLEDGMENTS. We would like to deeply thank Dr. Christos Savva (Diamond Light Source) who provided the plasmid and Prof. Milton T. Stubbs (University of Halle) for initiating the scientific discussions on the correlation of p*H* dependence and charge distribution of cryo-EM maps. P.L.K. was supported by the European Union through funding of the Horizon Europe ERA Chair "hot4cryo" project no. 101086665, the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, Zentrum für Innovationskompetenz program) (grant nos. 03Z22HN23, 03Z22HI2, and 03COV04), the European Regional Development Funds (Europäischer Fonds für regionale Entwicklung) for Saxony-Anhalt (grant nos. ZS/2016/04/78115 and ZS/2024/05/187255), the Deutsche Forschungsgemeinschaft (project nos. 391498659, RTG 2467 and 514901783, SFB 1664 [A04, C04, and D01]), and the Martin Luther University Halle-Wittenberg. G.K. acknowledges funding from Deutsche Forschungsgemeinschaft (project no. 514901783, SFB 1664, subproject D01). We further thank the high-performance computing centers at Leipzig University and at the Nationales Hochleistungsrechnen center of TU Dresden for providing the computational resources. D.A.S. was supported by FCT–Fundação para a Ciência e a Tecnologia, I.P., through MOSTMICRO-ITQB. R&D Unit (DOI: 10.54499/UIDB/04612/2020; DOI: 10.54499/UIDP/04612/2020) and LS4FUTURE Associated Laboratory (DOI: 10.54499/LA/P/0087/2020). I.S. acknowledges funding from the European Union HORIZON-MSCA-2024-PF-01 under grant agreement no. 101207412 (Protein ER-ways).

Author affiliations: [a]Department of Integrative Structural Biochemistry, Institute of Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg, Halle/Saale 06120, Germany; [b]Interdisciplinary Research Center HALOmem, Charles Tanford Protein Center, Martin Luther University Halle-Wittenberg, Halle/Saale 06120, Germany; [c]Structural Biochemistry, Bijvoet Centre for Biomolecular Research, Utrecht University, Utrecht 3584 CG, the Netherlands; [d]Institute for Drug Discovery, Leipzig University, Leipzig 04103, Germany; [e]Navigo Proteins GmbH, Halle/Saale 06120, Germany; [f]Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras 2780-157, Portugal; [g]Institute of Chemical Biology, National Hellenic Research Foundation, Athens 11635, Greece; [h]Institute of Pharmacy, Martin-Luther-University Halle-Wittenberg, Halle (Saale) D-06120, Germany; [i]Interdisciplinary Center for Bioinformatics, Leipzig University, Leipzig 04107, Germany; [j]Center for Scalable Data Analytics and Artificial Intelligence, Leipzig University, Leipzig 04105, Germany; and [k]Biozentrum, Martin Luther University Halle-Wittenberg, Halle/Saale 06120, Germany

1. P. Ball, Water is an active matrix of life for cell and molecular biology. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 13327–13335 (2017).
2. M. C. Bellissent-Funel *et al.*, Water determines the structure and dynamics of proteins. *Chem. Rev.* **116**, 7673–7697 (2016).
3. J. C. Smith, F. Merzel, A. N. Bondar, A. Tournier, S. Fischer, Structure, dynamics and reactions of protein hydration water. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359**, 1181–1189; discussion 1189–1190 (2004).
4. P. Ball, Water as an active constituent in cell biology. *Chem. Rev.* **108**, 74–108 (2008).
5. M. Levitt, B. H. Park, Water: Now you see it, now you don't. *Structure* **1**, 223–226 (1993).
6. R. C. Kretsch *et al.*, Complex water networks visualized by cryogenic electron microscopy of RNA. *Nature* **642**, 250–259 (2025).
7. J. J. Perona, C. S. Craik, R. J. Fletterick, Locating the catalytic water molecule in serine proteases. *Science* **261**, 620–622 (1993).
8. A. J. Venkatakrishnan *et al.*, Diverse GPCRs exhibit conserved water networks for stabilization and activation. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 3288–3293 (2019).
9. R. Hussein *et al.*, Cryo-electron microscopy reveals hydrogen positions and water networks in photosystem II. *Science* **384**, 1349–1355 (2024).
10. D. Laage, T. Elsaesser, J. T. Hynes, Water dynamics in the hydration shells of biomolecules. *Chem. Rev.* **117**, 10694–10725 (2017).
11. M. Nakasako, Water–protein interactions from high-resolution protein crystallography. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359**, 1191–1204; discussion 1204–1196 (2004).
12. G. Otting, NMR studies of water bound to biological molecules. *Prog. Nucl. Magn. Reson. Spectrosc.* **31**, 259–285 (1997).
13. J. Abramson *et al.*, Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
14. A. Zamanos, G. Ioannakis, I. Z. Emiris, Hydraprot: A new deep learning tool for fast and accurate prediction of water molecule positions for protein structures. *J. Chem. Inf. Model.* **64**, 2594–2611 (2024).
15. B. Bagchi, Water dynamics in the hydration layer around proteins and micelles. *Chem. Rev.* **105**, 3197–3219 (2005).
16. B. Born, S. J. Kim, S. Ebbinghaus, M. Gruebele, M. Havenith, The terahertz dance of water with the proteins: The effect of protein flexibility on the dynamical hydration shell of ubiquitin. *Faraday Discuss.* **141**, 161–173 (2009).
17. C. J. Crilly, J. E. Eicher, O. Warmuth, J. M. Atkin, G. J. Pielak, Water's variable role in protein stability uncovered by liquid-observed vapor exchange NMR. *Biochemistry* **60**, 3041–3045 (2021).
18. S. Ebbinghaus *et al.*, An extended dynamical hydration shell around proteins. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20749–20752 (2007).
19. A. C. Fogarty, E. Duboué-Dijon, F. Sterpone, J. T. Hynes, D. Laage, Biomolecular hydration dynamics: A jump model perspective. *Chem. Soc. Rev.* **42**, 5672–5683 (2013).
20. T. Li, A. A. Hassanali, Y.-T. Kao, D. Zhong, S. J. Singer, Hydration dynamics and time scales of coupled water−protein fluctuations. *J. Am. Chem. Soc.* **129**, 3376–3382 (2007).
21. S. Mondal, S. Mukherjee, B. Bagchi, *Protein Hydration Dynamics: Much Ado About Nothing?* (ACS Publications, 2017), pp. 4878–4882.
22. O. Sushko, R. Dubrovka, R. S. Donnan, Sub-terahertz spectroscopy reveals that proteins influence the properties of water at greater distances than previously detected. *J. Chem. Phys.* **142**, 055101 (2015).
23. K. Wüthrich *et al.*, NMR studies of the hydration of biological macromolecules. *Faraday Discuss.* **103**, 245–253 (1996).
24. G. Otting, E. Liepinsh, K. Wüthrich, Protein hydration in aqueous solution. *Science* **254**, 974–980 (1991).

25. M. Beckers, D. Mann, C. Sachse, Structural interpretation of cryo-em image reconstructions. *Prog. Biophys. Mol. Biol.* **160**, 26–36 (2021).
26. W. Kuhlbrandt, Biochemistry. The resolution revolution. *Science* **343**, 1443–1444 (2014).
27. D. Liebschner *et al.*, Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. D. Struct. Biol.* **75**, 861–877 (2019).
28. J. Dubochet, Cryo-EM–the first thirty years. *J. Microsc.* **245**, 221–224 (2012).
29. C. Y. Chen, Y. C. Chang, B. L. Lin, C. H. Huang, M. D. Tsai, Temperature-resolved cryo-EM uncovers structural bases of temperature-dependent enzyme functions. *J. Am. Chem. Soc.* **141**, 19983–19987 (2019).
30. K. Talley, E. Alexov, On the pH-optimum of activity and stability of proteins. *Proteins* **78**, 2699–2706 (2010).
31. T. Fleckenstein *et al.*, The chaperone activity of the developmental small heat shock protein Sip1 is regulated by pH-dependent conformational changes. *Mol. Cell* **58**, 1067–1078 (2015).
32. T. Otosu, K. Ishii, T. Tahara, Microsecond protein dynamics observed at the single-molecule level. *Nat. Commun.* **6**, 7685 (2015).
33. F. Damberger *et al.*, NMR characterization of a pH-dependent equilibrium between two folded solution conformations of the pheromone-binding protein from *Bombyx mori*. *Protein Sci.* **9**, 1038–1041 (2000).
34. J. Srivastava, D. L. Barber, M. P. Jacobson, Intracellular pH sensors: Design principles and functional significance. *Physiology (Bethesda)* **22**, 30–39 (2007).
35. I. D. Kuntz Jr., W. Kauzmann, Hydration of proteins and polypeptides. *Adv. Protein Chem.* **28**, 239–345 (1974).
36. S. Ebbinghaus *et al.*, Protein sequence- and pH-dependent hydration probed by terahertz spectroscopy. *J. Am. Chem. Soc.* **130**, 2374–2375 (2008).
37. B. Halle, Protein hydration dynamics in solution: A critical survey. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359**, 1207–1223; discussion 1223–1204, 1323–1208 (2004).
38. D. S. Grebenkov, Y. A. Goddard, G. Diakova, J. P. Korb, R. G. Bryant, Dimensionality of diffusive exploration at the protein interface in solution. *J. Phys. Chem. B* **113**, 13347–13356 (2009).
39. M. Marchi, F. Sterpone, M. Ceccarelli, Water rotational relaxation and diffusion in hydrated lysozyme. *J. Am. Chem. Soc.* **124**, 6787–6791 (2002).
40. M. Kim *et al.*, pH-dependent structures of ferritin and apoferritin in solution: Disassembly and reassembly. *Biomacromolecules* **12**, 1629–1640 (2011).
41. S. Dostalova *et al.*, Apoferritin as an ubiquitous nanocarrier with excellent shelf life. *Int. J. Nanomed.* **12**, 2265–2278 (2017).
42. T. Nakane *et al.*, Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152–156 (2020).
43. K. M. Yip, N. Fischer, E. Paknia, A. Chari, H. Stark, Atomic-resolution protein structure determination by cryo-EM. *Nature* **587**, 157–161 (2020).
44. F. Hamdi *et al.*, 2.7 A cryo-EM structure of vitrified M. musculus H-chain apoferritin from a compact 200 keV cryo-microscope. *PLoS One* **15**, e0232540 (2020).
45. W. R. Carmody, Easily prepared wide range buffer series. *J. Chem. Educ.* **38**, 559 (1961).
46. R. R. Crichton, C. F. Bryce, Subunit interactions in horse spleen apoferritin. Dissociation by extremes of pH. *Biochem. J.* **133**, 289–299 (1973).
47. L. Stuhn, J. Auernhammer, C. Dietz, pH-depended protein shell dis- and reassembly of ferritin nanoparticles revealed by atomic force microscopy. *Sci. Rep.* **9**, 17755 (2019).
48. G. Pintilie *et al.*, Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nat. Methods* **17**, 328–334 (2020).
49. G. A. Clegg, J. E. Fitton, P. M. Harrison, A. Treffry, Ferritin: Molecular structure and iron-storage mechanisms. *Prog. Biophys. Mol. Biol.* **36**, 56–86 (1980).
50. I. Bertini *et al.*, Structural insights into the ferroxidase site of ferritins from higher eukaryotes. *J. Am. Chem. Soc.* **134**, 6169–6176 (2012).
51. T. Asano *et al.*, Distinct mechanisms of ferritin delivery to lysosomes in iron-depleted and iron-replete cells. *Mol. Cell. Biol.* **31**, 2040–2052 (2011).
52. W. Yi *et al.*, Ferritin-mediated mitochondrial iron homeostasis is essential for the survival of hematopoietic stem cells and leukemic stem cells. *Leukemia* **38**, 1003–1018 (2024).
53. A. Ndayisaba, C. Kaindlstorfer, G. K. Wenning, Iron in neurodegeneration–Cause or consequence? *Front. Neurosci.* **13**, 180 (2019).
54. S. Gupta, C. Merriman, C. J. Petzold, C. Y. Ralston, D. Fu, Water molecules mediate zinc mobility in the bacterial zinc diffusion channel ZIPB. *J. Biol. Chem.* **294**, 13327–13335 (2019).
55. W. Qiu *et al.*, Ultrafast solvation dynamics of human serum albumin: Correlations with conformational transitions and site-selected recognition. *J. Phys. Chem. B* **110**, 10540–10549 (2006).
56. C. J. Cramer, D. G. Truhlar, Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chem. Rev.* **99**, 2161–2200 (1999).
57. D. Nolting *et al.*, pH-induced protonation of lysine in aqueous solution causes chemical shifts in X-ray photoelectron spectroscopy. *J. Am. Chem. Soc.* **129**, 14068–14073 (2007).
58. C. Chothia, The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol.* **105**, 1–12 (1976).
59. D. Eisenberg, A. D. McLachlan, Solvation energy in protein folding and binding. *Nature* **319**, 199–203 (1986).
60. L. Pauling, The adsorption of water by proteins. *J. Am. Chem. Soc.* **67**, 555–557 (1945).
61. M. A. Marques, M. D. Purdy, M. Yeager, CryoEM maps are full of potential. *Curr. Opin. Struct. Biol.* **58**, 214–223 (2019).
62. T. Bick, P. M. Dominiak, P. Wendler, Exploiting the full potential of cryo-EM maps. *BBA Adv.* **5**, 100113 (2024).
63. J. Hattne *et al.*, Analysis of global and site-specific radiation damage in cryo-EM. *Structure* **26**, 759–766.e754 (2018).
64. D. Tegunov, L. Xue, C. Dienemann, P. Cramer, J. Mahamid, Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 A in cells. *Nat. Methods* **18**, 186–193 (2021).
65. J. Wyman, The binding potential, a neglected linkage concept. *J. Mol. Biol.* **11**, 631–644 (1965).
66. B. Honig, A. Nicholls, Classical electrostatics in biology and chemistry. *Science* **268**, 1144–1149 (1995).
67. L. V. Bock, H. Grubmuller, Effects of cryo-EM cooling on structural ensembles. *Nat. Commun.* **13**, 1709 (2022).
68. N. J. Mowry, C. R. Krüger, M. Drabbels, U. J. Lorenz, Direct measurement of the critical cooling rate for the vitrification of water. *Phys. Rev. Res.* **7**, 013095 (2025).
69. D. P. Klebl, L. Aspinall, S. P. Muench, Time resolved applications for Cryo-EM; Approaches, challenges and future directions. *Curr. Opin. Struct. Biol.* **83**, 102696 (2023).
70. A. M. Schmitz *et al.*, High efficiency rare earth element bioleaching with systems biology guided engineering of Gluconobacter oxydans. *Commun. Biol.* **8**, 815 (2025).
71. P. Notin, N. Rollins, Y. Gal, C. Sander, D. Marks, Machine learning for functional protein design. *Nat. Biotechnol.* **42**, 216–228 (2024).
72. W. Zou *et al.*, Expression, purification, and characterization of recombinant human H-chain ferritin. *Prep. Biochem. Biotechnol.* **46**, 833–837 (2016).
73. J. M. de la Rosa-Trevin *et al.*, Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J. Struct. Biol.* **195**, 93–99 (2016).
74. S. Q. Zheng *et al.*, MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
75. K. Zhang, Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
76. T. Bepler *et al.*, Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).
77. J. M. de la Rosa-Trevin *et al.*, Xmipp 3.0: An improved software suite for image processing in electron microscopy. *J. Struct. Biol.* **184**, 321–328 (2013).
78. T. Wagner *et al.*, SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019).
79. J. Zivanov *et al.*, A Bayesian approach to single-particle electron cryo-tomography in RELION-4.0. *eLife* **11**, e83724 (2022).
80. E. F. Pettersen *et al.*, UCSF chimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
81. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
82. A. Jansen, N. Aho, G. Groenhof, P. Buslaev, B. Hess, phbuilder: A tool for efficiently setting up constant pH molecular dynamics simulations in GROMACS. *J. Chem. Inf. Model.* **64**, 567–574 (2024).
83. J. Huang, A. D. MacKerell Jr., CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
84. Y. Won, Force field for monovalent, divalent, and trivalent cations developed under the solvent boundary potential. *J. Phys. Chem. A* **116**, 11763–11767 (2012).
85. P. Virtanen *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
86. N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, O. Beckstein, MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).
87. M. Ester, H.-P. Kriegel, J. Sander, X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise" in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, E. Simoudis, J. Han, U. Fayyad, Eds. (AAAI Press, Portland, OR, 1996), pp. 226–231.
88. F. Pedregosa *et al.*, Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
89. I. Skalidis, D. A. Semchonok, C. Tueting, F. Hamdi, P. L. Kastritis, Cryo-EM structure of Human Apoferritin at pH 3.5. EMDB. https://www.ebi.ac.uk/emdb/EMD-54951. Deposited 23 September 2025.
90. I. Skalidis, D. A. Semchonok, C. Tueting, F. Hamdi, P. L. Kastritis, Cryo-EM structure of Human Apoferritin at pH 3.5. PDB. https://www.rcsb.org/structure/9SJR. Deposited 23 September 2025.
91. I. Skalidis, D. A. Semchonok, C. Tueting, F. Hamdi, P. L. Kastritis, Cryo-EM structure of Human Apoferritin at pH 4. EMDB. https://www.ebi.ac.uk/emdb/EMD-54952. Deposited 23 September 2025.
92. I. Skalidis, D. A. Semchonok, C. Tueting, F. Hamdi, P. L. Kastritis, Cryo-EM structure of Human Apoferritin at pH 4. PDB. https://www.rcsb.org/structure/9SJS. Deposited 23 September 2025.
93. I. Skalidis, D. A. Semchonok, C. Tueting, F. Hamdi, P. L. Kastritis, Cryo-EM structure of Human Apoferritin at pH 5. EMDB. https://www.ebi.ac.uk/emdb/EMD-54953. Deposited 23 September 2025.
94. I. Skalidis, D. A. Semchonok, C. Tueting, F. Hamdi, P. L. Kastritis, Cryo-EM structure of Human Apoferritin at pH 5. PDB. https://www.rcsb.org/structure/9SJT. Deposited 23 September 2025.
95. I. Skalidis, D. A. Semchonok, C. Tueting, F. Hamdi, P. L. Kastritis, Cryo-EM structure of Human Apoferritin at pH 7. EMDB. https://www.ebi.ac.uk/emdb/EMD-54954. Deposited 23 September 2025.
96. I. Skalidis, D. A. Semchonok, C. Tueting, F. Hamdi, P. L. Kastritis, Cryo-EM structure of Human Apoferritin at pH 7. PDB. https://www.rcsb.org/structure/9SJU. Deposited 23 September 2025.
97. I. Skalidis, D. A. Semchonok, C. Tueting, F. Hamdi, P. L. Kastritis, Cryo-EM structure of Human Apoferritin at pH 9. EMDB. https://www.ebi.ac.uk/emdb/EMD-54955. Deposited 23 September 2025.
98. I. Skalidis, D. A. Semchonok, C. Tueting, F. Hamdi, P. L. Kastritis, Cryo-EM structure of Human Apoferritin at pH 9. PDB. https://www.rcsb.org/structure/9SJV. Deposited 23 September 2025.