Proteins maintain hydration at high [KCl] concentration regardless of content in acidic amino acids

Hosein Geraili Daronkola[1] and Ana Vila Verde[1,*]

[1]Department of Theory & Bio-Systems, Max Planck Institute of Colloids and Interfaces, Potsdam, Germany

ABSTRACT Proteins of halophilic organisms, which accumulate molar concentrations of KCl in their cytoplasm, have a much higher content in acidic amino acids than proteins of mesophilic organisms. It has been proposed that this excess is necessary to maintain proteins hydrated in an environment with low water activity, either via direct interactions between water and the carboxylate groups of acidic amino acids or via cooperative interactions between acidic amino acids and hydrated cations. Our simulation study of five halophilic proteins and five mesophilic counterparts does not support either possibility. The simulations use the AMBER ff14SB force field with newly optimized Lennard-Jones parameters for the interactions between carboxylate groups and potassium ions. We find that proteins with a larger fraction of acidic amino acids indeed have higher hydration levels, as measured by the concentration of water in their hydration shell and the number of water/protein hydrogen bonds. However, the hydration level of each protein is identical at low (b_{KCl} = 0.15 mol/kg) and high (b_{KCl} = 2 mol/kg) KCl concentrations; excess acidic amino acids are clearly not necessary to maintain proteins hydrated at high salt concentration. It has also been proposed that cooperative interactions between acidic amino acids in halophilic proteins and hydrated cations stabilize the folded protein structure and would lead to slower dynamics of the solvation shell. We find that the translational dynamics of the solvation shell is barely distinguishable between halophilic and mesophilic proteins; if such a cooperative effect exists, it does not have that entropic signature.

SIGNIFICANCE Acidic amino acids are known to bind the most water of all amino acids, so it has been proposed that their excess is necessary to retain protein hydration in environments with low water activity. Our results do not support this scenario for the case of concentrated KCl solutions, typical of the cytoplasm of some halophilic organisms; the 10 proteins we investigate keep their hydration level constant under widely different external KCl concentrations despite having very different content in acidic amino acids. The hydrated electrolyte ions integrate the protein solvation shell rather than compete with the protein for available water. Maintaining hydration in the folded state is not the evolutionary driving force behind the excess acidic amino acids observed in halophilic proteins.

INTRODUCTION

Halophilic organisms, unlike most life on Earth, have the uncanny ability to survive at molar external NaCl concentrations. Despite their relative scarcity, halophilic organisms have been found in multiple kingdoms of life: archaea (1), bacteria (2), protozoa, fungi, algae, and multicellular eukaryotes (3,4). To counterbalance the high osmotic stress

induced by high external NaCl concentrations, some halophiles accumulate equivalent concentrations of KCl in their cytoplasm (5). At such high salt concentrations, the interactions that dictate the structure and structural stability of proteins differ markedly from those at the much lower salt concentrations found in most organisms.

As a result, many proteins of mesophilic organisms (here, termed mesophilic proteins) are poorly soluble at KCl concentrations typical of the cytoplasm of halophilic organisms (6–8). In contrast, proteins of halophilic organisms (here, termed halophilic proteins) are structurally stable and functioning at high salt concentrations, but often show lower (or no) stability and activity at KCl concentrations typical of the cytoplasm of mesophilic organisms (5). Their resilience is

related to their amino acid composition (3,5,9–11). Cytoplasmic proteins of halophilic organisms are, on average, longer than their mesophilic counterparts; they also have a higher fraction of random coil structure at the expense of α-helix structure (12); they are richer in small+polar and small+apolar amino acids and poorer in large+hydrophobic amino acids; they are also poorer in cationic amino acids, which contain longer alkyl sections in their side chains than the anionic ones. The reduction in the fraction of large-+hydrophobic and cationic amino acids and the increase in small+polar and small+apolar is thought to compensate the increased attraction between nonpolar groups as salt concentration increases (5).

Halophilic proteins are also much richer in acidic amino acids (3,13)—mostly located on their surface (9)—than their mesophilic counterparts, leading to substantially negative net protein charges.

The role played by the excess acidic amino acids, however, is not yet well understood. One of the most intuitive explanations is that large surface charge prevents protein aggregation at high salt concentrations. The excess surface charge in halophilic proteins would compensate for charge screening at high salt concentrations and would prevent aggregation (3,5,14–16). Several results, however, suggest that it is unlikely that charge repulsion is the only–or even main–function of excess acidic amino acids. Experiments show that the stability of halophilic proteins at 0.05 mol/dm^3 NaCl and $\sim 10^{-6}$ mol/dm^3 NADH$^+$ is as high as that reached with molar concentrations of NaCl (17). Charge screening alone does not explain the strong dependence of the stability of the folded state of halophilic proteins, and/or their function, on salt concentrations up to several molar, because measurements (5) and calculations (14) indicate that screening of electrostatic interactions by salt is largely complete at $c_{salt} = 0.5$ mol/dm^3. At the high salt concentrations seen inside halophiles, electrostatic repulsion between the carboxyl groups at the surface of proteins should have only a small effect on their structural stability, as one NMR study shows (16), and likewise on their ability to aggregate. At present, two other explanations have been proposed for the large fraction of acidic amino acids in halophilic proteins:

Ion-solvent stabilization model

Experiments have shown that some folded halophilic proteins bind large amounts of water, NaCl, and KCl (18,19); in contrast, salt binding was not detected in mesophilic proteins in the native state (20). Moreover, quasielastic neutron spectroscopy measurements of water translational dynamics in the cytoplasm of halophilic and nonhalophilic bacteria suggest that halophiles have a water fraction with extremely slow dynamics that is absent from nonhalophiles (21). Based on these results, it has been proposed that the abundance of acidic amino acids enables cooperative, stabilizing interactions with cations. The acidic amino acids contribute to a net stabilization of the folded protein structure (despite intramolecular electrostatic repulsion) by forming cooperative hydrated ion networks (20,22). These networks keep the folded protein hydrated despite the high salt concentration and should lead to highly ordered protein hydration shells (3). According to this view, excess acidic amino acids indeed prevents aggregation of proteins at high salt concentration not by charge repulsion but because the cooperative ion-water-protein networks stabilize the protein solvation layer.

Solvent-only stabilization model

The ion-solvent stabilization model has been challenged, however, based on multiple observations. ^{17}O magnetic relaxation experiments did not find any differences in the rotational dynamics in hydration shells of halophilic and nonhalophilic versions of protein L (23), in contrast to the predictions of the ion-solvent stabilization model. Additionally, recent experiments have shown that some extremely halophilic organisms can thrive at low external NaCl concentrations, i.e., also with cytoplasmic KCl concentrations equal to those found in nonhalophiles, despite having a markedly acidic proteome (1); this result suggests that stabilizing interactions between cations and the acidic residues are not critical. To explain these observations, it was proposed that an abundance of acidic amino acids at the surface of halophilic proteins is necessary to compete–rather than cooperate–with the ions in solution for available water and, thus, ensure that the protein surface remains sufficiently hydrated–and the protein remains soluble, conformationally stable, and thus functional–at high salt concentrations (1,24).

Here, we report a molecular dynamics study of five pairs of halophilic and mesophilic proteins in which we comparatively examine how the structure and dynamics of protein hydration shells is affected by the concentration of KCl and by the fraction of acidic amino acids in the protein. We present also parameters for the interaction between K$^+$ and carboxylate, critical for this study, optimized to reproduce the experimental solution activity derivative of potassium acetate solutions up to 2 mol/kg and crystallographic information of potassium ions in the vicinity of acidic amino acids.

MATERIALS AND METHODS

Force fields

To gain insight into the role played by acidic amino acids in halophilic proteins using simulations, we require force fields for proteins, water, and ions with the correct balance of interactions between all the species, both at low and high (up to several molals) KCl concentrations. The force fields we use—the TIP3P water model (25), the AMBER ff14SB (26,27) force field for proteins, the general AMBER force field (GAFF) for acetate (28), and the potassium and chloride parameters of Joung and Cheatham (29) for TIP3P water meet this requirement for most interactions, so only a few interactions were modified as described below. The AMBER ff14SB force field used with TIP3P water reproduces the hydration free energies of

small-molecule analogs of the side chains of neutral amino acids to within 1 kcal/mol root mean-square error of the experimental reference values (30), and the partition coefficients of the same analogs between water and several organic solvents to within 0.5 log units (31). This combination of solvent and protein force field reproduces the secondary structure content of short peptides (27) and the backbone dynamics in the native state of globular proteins (27). The experimental hydration free energies of K^+ and Cl^- were one of the target properties used in the parameterization done by Joung and Cheatham (29). Their parameters reproduce the solution activity of aqueous solutions of this salt up to $b = 2$ mol/kg (32), so they are appropriate for both dilute and concentrated solutions. Metallic ligands present in the proteins are simulated using AMBER-compatible parameters reported in the literature. The $[2Fe-2S]^{2+}$ ligand present in both ferredoxin proteins studied here is simulated based on the parameters developed by Carvalho et al. (33) for the [2Fe-2S] ferredoxin from *Mastigocladus laminosus*, with slight modifications: 1) all equilibrium angles are set to the values found in the crystal structure of our ferredoxin proteins. And 2) the force constant for the Fe-S-C angle, missing in the parameter set of Carvalho et al. for the $[2Fe-2S]^{2+}$ from *M. laminosus*, is set to the value reported by the same authors for the desulforedoxin protein from *Desulfovibrio gigas*. The desulforedoxin protein contains Fe(III) coordinated by four cysteines in the same configuration as the ferredoxins studied here. Both carbonic anhydrase proteins studied here have a four-coordinated zinc metal center where the zinc ion is connected to three histidine residues and one water molecule. The metal center is simulated using the zinc AMBER force field (34,35) and the Lennard-Jones (LJ) parameters for the zinc ion are from Li et al. (35).

Modifications to these force fields are indispensable for some of the interactions involving carboxylate groups or K^+. We modify the self-interaction LJ parameters for all carboxylate oxygens to the values proposed by Kashefolgheta et al. (36), which reproduce the hydration free energy of acetate in TIP3P water better than the original AMBER parameters; note that these parameters are used in all interactions derived using combination rules. The LJ parameters for the interaction between carboxylate and the side chain of the positively charged amino acid lysine are also modified to those in (36) to reduce excessively strong salt bridges. The LJ parameters for the interaction between K^+ and carboxylate oxygens are modified as described in the following section. Data S1 includes Amber ff14SB, dat, and lib files modified to include the optimized parameters mentioned here. These files can be used to seamlessly generate the final prmtop and inpcrd input files necessary to run simulations with Amber. Example input files for one of the proteins simulated here (protein L) are also given in Data S1.

Results from simulations without any modifications to the default LJ parameters, provided for comparison, are marked as "AMBER and GAFF only."

Optimizing parameters for $K^+ \cdots$ carboxylate interactions

Our tests (described in the Results) show that the default parameters for $K^+ \cdots$ carboxylate interactions cause excessive contact between these ions. Because this interaction is critical for our study, our first step was to find optimal parameters to describe it.

Intermolecular interactions between any atoms i and j have the following form in the AMBER force field:

$$V(r_{ij}) = \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^6\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad (1)$$

In this expression, q_i and q_j are the charges of the atoms, r_{ij} is the distance between them, ε_0 is the dielectric permittivity in the vacuum. $R_{\min,ij}$ and ε_{ij} are parameters determining the LJ potential, which mimics the van der Waals interatomic interaction. LJ parameters for the interaction between

different atom types are typically obtained from their self-interaction parameters using combination rules:

$$R_{\min,ij} = \frac{(R_{\min,ii} + R_{\min,jj})}{2}, \varepsilon_{ij} = \sqrt{(\varepsilon_{ii}\,\varepsilon_{jj})}, \quad (2)$$

where the indices ii and jj denote the self-interaction parameters.

In the AMBER force field, ions have their nominal charge, and the partial charges of polyatomic ions are determined using a well-defined charge fitting procedure. We opted to tune the $K^+ \cdots$ carboxylate interaction by modifying the LJ potential between K^+ and the carboxylate oxygen (O) only. The partial charges on the carboxylate atoms remain unchanged, which ensures compatibility within the AMBER force field. In principle, both $R_{\min,ij}$ and ε_{ij} obtained using Eq. 2 could be overridden by optimized values for the $i = K^+$, $j = O$ pair. This dual optimization is, however, not feasible because of its high computational cost. We optimized only the $R_{\min,K^+O}$ parameter and left ε_{K^+O} at the value obtained with the combination rules. Our prior work confirms that substantial improvements in the description of intermolecular interactions are achieved even with this limited parameterization freedom (36). In Results, we show that the original AMBER parameters lead to substantial deviations between various calculated properties and the experimental reference values. Because of this substantial deviation, the $R_{\min}$ parameter, with its power of 6 and 12 in the LJ potential, is a better choice for parameterization than ε.

The $R_{\min,K^+O}$ parameter is optimized to reproduce 1) the experimental solution activity derivative of aqueous solutions of potassium acetate with molality $b_{KCH3COO} \in \{0.5, 1, 2\}$ mol/kg, and 2) the distances between K^+ and carboxylate groups found in the crystal structure of the halophilic 2Fe-2S ferredoxin protein with Protein Data Bank, PDB: 1DOI. This protein is one of the few for which the potassium ions present during crystallization were resolved in the crystal structure.

The second target property was added because we found that the solution activity derivative varies nonmonotonically with $R_{\min,K^+O}$ to the extent that $R_{\min,K^+O}$ values leading to dramatically different solution structure yielded identical solution activity derivatives, as described below. Our final choice of $R_{\min,K^+O}$, shown in Results, is that which best reproduces selected distances between K^+ and carboxylate oxygens in the protein crystal structure while yielding activity derivatives deviating not more than 7% from the reference value for all three concentrations. Our optimized LJ potential for the interactions between K^+ and carboxylate groups are thus appropriate for simulations of biomolecular systems in a wide range of KCl concentrations and are also appropriate for simulations of systems with acetate ions modeled with the AMBER force field or GAFF.

Calculating the electrolyte activity derivative in simulations

The molar electrolyte activity derivative can be calculated from simulation using Kirkwood-Buff theory (37,38), which links particle fluctuations to thermodynamic properties. Below, we summarize the main expressions used in this work. Examples of its application to other systems can be found in the literature (39–42). The central quantity of this theory is the Kirkwood-Buff integral G. In this work, we use expressions that enable its use with simulations of closed systems (43). For any two species i and j, the integral is as follows:

$$G_{ij}(R) = \int_0^{2R} \left[f_{ij}g_{ij}^{NVT}(r) - 1\right]4\pi r^2\left(1 - 3x/2 + x^3/2\right)dr,$$

$$(3)$$

Where R is 1/4 of the simulation box size, $x = r/(2R)$ and $g_{ij}^{NVT}(r)$ is the radial distribution function (RDF) in the canonical ensemble. The factor

f_{ij} corrects for the fact that the tail of RDFs in closed systems does not converge to 1. Even though there is no formal relation between the RDFs in open and closed systems, the ratio of their tails is expected to be of the order of $1 \pm 1/N$, where N is the number of particles in the simulation box (44–46). For this reason, a multiplicative correction factor is appropriate.

Because of electroneutrality, applying Kirkwood-Buff theory to an electrolyte solution requires the solution to be treated as a binary system of indistinguishable ions (called the cosolvent) and water (38,40,41). The molar cosolvent activity derivative, a_c', for a solution of a 1:1 electrolyte is thus defined as

$$
\begin{aligned}
a_c' &= \left. \frac{\partial \ln a_c}{\partial \ln \rho_c} \right|_{p,T} \\
&= 1 + \left. \frac{\partial \ln \gamma_c}{\partial \ln \rho_c} \right|_{p,T}
\end{aligned}
\tag{4}
$$

In this expression, γ_c is the molar activity coefficient of the cosolvent, ρ_c is its number density, defined as $\rho_c = n_c/V$, where $n_c = n_+ + n_-$ is the total number of positive and negative ions in the solution volume V, and a_c is its molar activity ($a_c = \gamma_c \times \rho_c$). Kirkwood-Buff theory allows the calculation of the electrolyte activity derivative from simulation, as

$$
a_c' = \frac{1}{1 + \rho_c(G_{cc} - G_{cw})}
\tag{5}
$$

where the subscript w refers to the water. The terms in the above expression are sums of Kirkwood-Buff integrals between the ions and water:

$$
G_{cc} = \frac{1}{4}(2G_{+-} + G_{++} + G_{--})
\tag{6}
$$

and

$$
G_{cw} = G_{+w} + G_{-w}
\tag{7}
$$

Experimental electrolyte activity derivative

Experimental activities for electrolytes are typically reported considering the neutral salt unit—as opposed to considering a cosolvent of indistinguishable ions—and in the molal rather than the molar scale. In what follows, the subscript s indicates quantities defined in terms of neutral salt units and the superscript (b) indicates quantities in the molal scale. The mean molal activity coefficient, $\gamma_{s,\pm}^{(b)}$, of potassium acetate in aqueous solution has been experimentally determined for molalities up to $b_s = 3$ mol/kg (47). In Fig. 1, we show the experimental values of $\gamma_{s,\pm}^{(b)}$ as a function of the molality of KCH₃COO together with the result of the Pitzer equation applicable to this system (48). The Pitzer equation reproduces the experimental data excellently, so we use it to calculate the molal salt activity derivative

$$
a_s'^{(b)} = \left. \frac{\partial \ln a_s^{(b)}}{\partial \ln b_s} \right|_{p,T} = \left. \frac{\partial \ln\left(\gamma_{s,\pm}^{(b)} \, b_s\right)}{\partial \ln b_s} \right|_{p,T},
\tag{8}
$$

using Mathematica.

To enable direct comparisons between experiment and simulation, the experimental molal activity derivative of the salt must first be converted to the corresponding molar quantity:
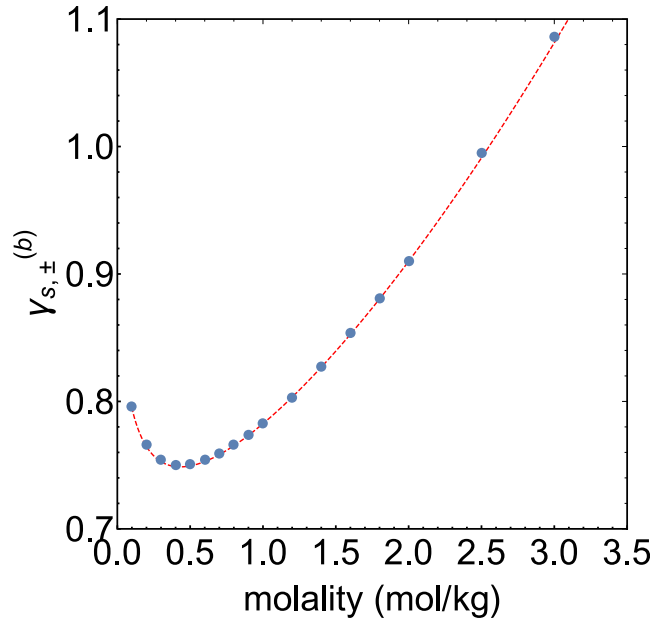

FIGURE 1 Mean molal activity coefficient $(\gamma_{s,\pm}^{(b)})$ of potassium acetate as a function of the molality of KCH₃COO. The blue circles are the experimental data (47), and the red dashed line is calculated using the relevant Pitzer equation (48). To see this figure in color, go online.

$$
a_s' = \left. \frac{\partial \ln a_s}{\partial \ln \rho_s} \right|_{p,T} = 1 + \left. \frac{\partial \ln \gamma_{s,\pm}}{\partial \ln \rho_s} \right|_{p,T},
\tag{9}
$$

where ρ_s is the number density of the salt and $\gamma_{s,\pm}$ its mean molar activity coefficient. The mean molal and molar activity coefficients are related as (49)

$$
\gamma_{s,\pm} = \frac{b_s \, \rho_w}{\rho_s} \gamma_{s,\pm}^{(b)},
\tag{10}
$$

where ρ_w is the density (mass per unit volume) of water and ρ_s is the number density of the salt, for a 1:1 electrolyte would be $\rho_s = \rho_c/2$. Substituting Eq. 10 into Eq. 9 shows that the salt activity derivatives in the two scales are related as

$$
a_s' = a_s'^{(b)} \frac{\partial \ln b_s}{\partial \ln \rho_s}
\tag{11}
$$

We use Eq. 11 with the solution density reported in (50) to obtain a_s' from $a_s'^{(b)}$. The molar solution activities expressed in terms of the salt and the cosolvent are related as $a_s = 0.5 a_c$ for 1:1 electrolytes; their derivatives and their activity coefficients are identical ($a_s' = a_c'$ and $\gamma_{s,\pm} = \gamma_c$); the experimental a_s' can thus be directly compared with a_c' calculated from the simulation.

Simulation details

Potassium acetate solutions

We perform molecular dynamics simulations using the GROMACS simulation package in its 2018 version (51,52). Simulations are performed for three different concentrations, $b_{\text{KCH₃COO}} \in \{0.5, 1, 2\}$ mol/kg. Each

simulation–minimization, equilibration, and production run–is performed for a given test value of R_{min,K^+O}, and multiple values are tested at each concentration. These test values are related to the default, $R_{min,K^+O,LB}$ (obtained using the Lorentz-Berthelot combination rule; Eq. 2), via a scaling factor:

$$R_{min,K^+O} = f_{R_{min,K^+O}} \cdot R_{min,K^+O,LB} \qquad (12)$$

For convenience, we refer to the tested values of R_{min,K^+O} via the scaling factor $f_{R_{min,K^+O}}$.

The simulation boxes are prepared by putting 72, 144, or 288 neutral salt units, KCH_3COO–corresponding to the low, intermediate, and high salt concentrations–in a cubic box with edge length $L \approx 6$ nm and solvating the systems by adding $\sim$8000 TIP3P water molecules using the editconf and solvate tools from Gromacs 2018 (51,52). The LJ interactions are cut-off at 1.2 nm. Electrostatic interactions are calculated using direct summation up to 1.2 nm. Beyond this cutoff distance, the electrostatic interactions are calculated with the particle mesh Ewald scheme with a grid spacing of 0.12 nm and sixth order interpolation (53). LJ interactions are smoothly shifted to 0 between 1.0 and 1.2 nm using the switch function available in GROMACS. Long-range dispersion corrections are applied to both the energy and pressure. A leap-frog stochastic (SD) integrator (54) is used to integrate the equations of motion in all simulations, with the temperature fixed at 298 K. All bonds with H-atoms are restrained using the LINCS algorithm (55) in all simulation steps (except the minimization step with l-bfgs method), which enables integration using a 2-fs time step. The systems are equilibrated following a protocol commonly used in biological studies: 1) two initial minimization steps with the steepest-descent and l-bfgs algorithms. The latter is a quasi-Newtonian algorithm for energy minimization which converges faster than the conjugate gradient algorithm. 2) A 500-ps heating equilibration simulation in the canonical ensemble (*NVT*) using Langevin thermostat with a coupling constant of 1.0 ps to 298 K. 3) Another 10-ns simulation is performed in the isothermal-isobaric ensemble (*NpT*) to equilibrate the system density at the pressure of 1 bar, using the Berendsen barostate (56) with a relaxation time of 1.0 ps. We select three distinct configurations from the *NpT* equilibration simulation, with a box volume similar to the average box volume obtained during that simulation. Each of these configurations is then used as the initial state of a production simulation in the *NVT* ensemble and lasting 50 ns, making up 150 ns for each concentration. This large simulation time is necessary to calculate activity derivatives with high precision. The trajectories of production simulations are saved every 2 ps for analysis.

Restrained ferredoxin in $c_{KCl} = 1$ mol/dm^3

Simulations are performed with Gromacs 2018; unless otherwise noted, all simulation details are identical to those used to simulate potassium acetate solutions. Each simulation–minimization, equilibration, and production run–is performed for a given test value of R_{min,K^+O}. The simulation box is prepared by solvating the protein with a cubic box of TIP3P water, with the distance between the protein surface and the box face being at least 15 Å. During all simulation steps, the backbone atoms are restrained to their position in the crystal structure at all times, using a harmonic potential with a force constant of 1000 kJ·mol^{-1}·nm^{-2}. Applying this restraint is necessary to compare the distance of the K^+ ions to carboxylate groups between the simulated system and the crystal structure for each value of R_{min,K^+O} tested. A single *NpT* production simulation with duration 200 ns is performed for each R_{min,K^+O}.

Unrestrained proteins at high and low KCl concentrations

Most simulation steps are performed using the GPU version of the pmemd engine in AMBER 2018 (57). The only exception is the l-bfgs minimization step, which is performed on the Sander engine of the AMBER simulation package because it is not available in the pmemd engine.

The starting configurations for the simulations are prepared using the tLeap module of the AMBER software (58). Each protein is simulated at two different salt concentrations: $b_{KCl} = 0.15$ mol/kg, corresponding to mesophilic conditions, and $b_{KCl} = 2$ mol/kg, corresponding to halophilic conditions. The simulation boxes are cubic, with edge length $\approx$100 Å. Each initial configuration is prepared by putting potassium and chloride ions in numbers corresponding to the desired concentration, and adding extra potassium or chloride for neutralization using the addIons tool of AMBER; the systems are then solvated with the appropriate number of TIP3P water molecules (25) using the tLeap tool from AMBER. The distance between the protein surface and the box face is at least 20 Å in all cases. Periodic boundary conditions are applied in the *XYZ* directions. The nonbonded potential cutoff distance is 12 Å for both LJ and electrostatic interactions. Beyond this cutoff distance, electrostatic interactions are calculated with the particle mesh Ewald scheme with a grid spacing of 1.0 Å, and fourth order interpolation (53). Long-range dispersion corrections are applied to both the energy and pressure. All bonds with H-atoms are constrained using the SHAKE algorithm (59) during the *NpT* equilibration and the production simulations. We use a 2-fs time step for all simulations except for those of the carbonic anhydrase proteins, which use a 1-fs time step. The carbonic anhydrase proteins have a zinc metal center that coordinates with one water molecule; for technical reasons, the SHAKE algorithm cannot be applied to this molecule, which limits the maximal time step that can be used.

Each system is subject initially to four minimization cycles, each comprising 2500 minimization steps using the steepest-descent algorithm and 7500 steps using the conjugate gradient algorithm. The protein atoms are restrained to their positions in the pdb file, with the value of the restraint bond constants (500, 300, 100, and 50 kcal·mol^{-1}·Å^{-2}) decreasing for each cycle. This procedure removes bad contacts that may be created in the process of adding ions and water to the system. Subsequently, each system is minimized for 10,000 steps using the l-bfgs algorithm without any restraints. The system temperature is progressively increased from 0 to 298 K during a 500-ps simulation in the canonical ensemble (*NVT*), with the protein atoms restrained to their positions with a constant of 10 kcal·mol^{-1}·Å^{-2}, using the Langevin thermostat with a collision frequency of 1.0 ps^{-1}. Each system is equilibrated for 10 ns in 10 steps of 1 ns in the isothermal-isobaric ensemble (*NpT*) using the Berendsen barostat (56) with a relaxation time of 2.0 ps while increasing the skinnb parameter to 5 Å to prevent unnecessary rebuilds of the pair list when there is no possibility of missing atoms in the pairwise calculation within the cutoff. Increasing the value of this parameter is indispensable because the GPU version of the pmemd engine is very sensitive to changes in the box size. Fig. 2 shows an equilibrated simulation box.

The starting configuration for each production simulation is selected from the last 1 ns of the *NpT* equilibration, under the condition that it has the box density closest to the average for that simulation. The production simulations are performed in the *NVT* ensemble, for 5×10^8 steps, corresponding to 0.5 μs for carbonic anhydrase and 1 μs for all other proteins. The simulations for the carbonic anhydrases were not extended to 1 μs because the proteins remain conformationally stable within the simulated time. For the production simulations, we use the Langevin thermostat with a collision frequency of 0.01 ps^{-1}. This low collision frequency is necessary to reduce the impact of the thermostat on the dynamics of the system. Prior work (60) has shown that this low collision frequency leads to self-diffusivity, rotational correlation time, and shear viscosity values within 2% of those obtained from simulations in the *NVE* ensemble at the same temperature. The average temperature in our simulations remains at its target value despite the low collision frequency of the thermostat. Production simulations are saved every 100 ps for analysis.

For the calculation of the mean-square displacement (MSD) for time-scales <100 ps, we perform three production simulations for each protein and each KCl concentration, each lasting 1 ns and, with configurations, saved every 100 fs. The starting configuration for each simulation is taken from the longer production runs, at $t \in \{100, 500, 900\}$ ns.
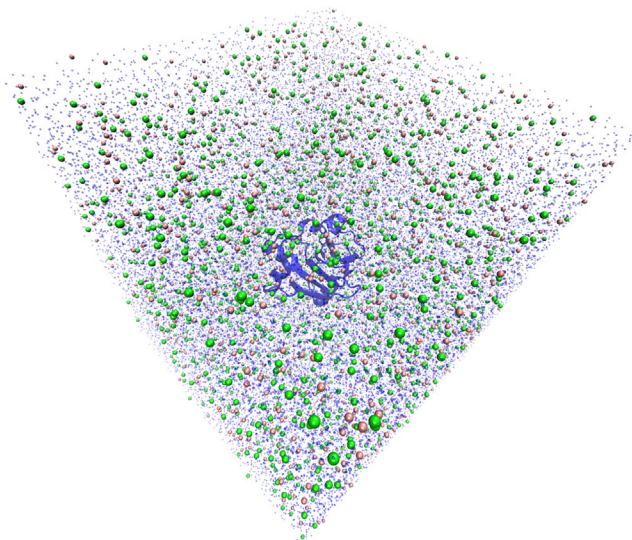
FIGURE 2   Example simulation box, of the halophilic ferredoxin protein (PDB: 1DOI). Dark blue shape represents the New Cartoon representation of protein, pink spheres represent K^+, green spheres represent Cl^-, and transparent small blue dots represent water molecules. To see this figure in color, go online.

RESULTS

Optimizing parameters for $K^+\cdots$ carboxylate interactions

The activity derivative of potassium acetate solutions with molality $b_{KCH_3COO} \epsilon\{0.5, 1, 2\}$ mol/kg calculated in simulation is shown in Fig. 3.

For each concentration, the activity derivative is calculated for different values of R_{min,K^+O}, which govern the LJ interaction between K^+ and the carboxylate oxygens; these values are expressed in the figures in terms of a multiplicative scaling factor (defined in Eq. 12). The unoptimized anion-cation parameters, corresponding to the scaling factor of 1 in Fig. 3, yield activity derivatives far from the reference value from experiment (*green lines*). This deviation demonstrates that the unoptimized parameters fail dramatically to adequately describe $K^+\cdots CH_3COO^-$ interactions in the concentration range we tested; optimizing this interaction is indispensable. With this in mind, we first examine the solution activity derivative for $b_{KCH_3COO} = 0.5$ mol/kg (Fig. 3 *a*). It varies nonmonotonically with an increasing scaling factor: initially it rapidly increases, then decreases slowly before again increasing. Perfect agreement with the reference experimental value at this concentration occurs for $f_{R_{min,K^+O}} = 2.0$. Despite this agreement, $f_{R_{min,K^+O}} = 2.0$ does not result in the correct solution structure at $b_{KCH_3COO} = 0.5$ mol/kg: the corresponding anion-cation RDF, shown in Fig. 3 *d*, would suggest that only solvent-separated ion pairs exist in potassium acetate solutions and that neither contact-shared (CIP) nor solvent-shared (SIP) occur (CIP, the two ions are in direct contact; SIP, the ions share one hydration layer; and solvent-separated,

each ion retains its first hydration layer). The absence of CIPs disagrees with potentiometric measurements, which indicate that K^+ and CH_3COO^- associate–albeit weakly–to form neutral complexes (61,62) best understood as pairs of ions in direct contact (63).

The anion-cation RDFs in Fig. 3 *d* indicate that scaling factors in the range $1.02 < f_{R_{min,K^+O}} < 1.1$ allow the presence of CIPs in solution, i.e., yield a solution structure qualitatively in line with experiment. This range of scaling factors also yields solution activity derivatives within 7% of the target experimental values for all three concentrations (Fig. 3, *a–c*). However, despite the fact that solution activity derivatives remain constant for this range of $f_{R_{min,K^+O}}$-values, the solution structure changes substantially: e.g., Fig. 3 *d* shows that for $f_{R_{min,K^+O}} = 1.02$, CIPs are still abundant, although much less so than SIPs or 2SIPS; for $f_{R_{min,K^+O}} = 1.1$, CIPs are residual. Clearly, randomly choosing a value of $f_{R_{min,K^+O}}$ within the $1.02 < f_{R_{min,K^+O}} < 1.1$ range is insufficient, and another experimental reference property is desirable to pinpoint the optimal parameter value.

The pdb structure with pdb code 1DOI, of a halophilic ferredoxin, contains five cocrystalized K^+ near acidic amino acids at the protein surface. This crystal structure was obtained from both room temperature and 100 K diffraction data and contains 237 water molecules, corresponding to 55% of the crystal volume. Flash freezing, typically used in crystallography to reach cryogenic temperatures, leaves the solvent in an amorphous state but often contracts the unit cell by 2–7% with most of the contraction arising from the solvent (64). These conditions mean that the protein surface and the potassium ions in this crystal structure are heavily solvated and that the distances between K^+ and the nearby carboxylates in the crystal structure are a reasonable estimate of the distance between these ions when forming CIPs for the solvated protein at room temperature. We used this information to select the optimal parameter within the $1.02 < f_{R_{min,K^+O}} < 1.1$ range. To the best of our knowledge, other experimental observables that directly report on the structure of $K^+\cdots$ carboxylate ion pairs in solvated conditions do not exist. We simulated the ferredoxin protein in $c_{KCl} = 1$ mol/dm^3 at $T = 298$ K. This high concentration of K^+ is within the range used in the previous parameterization step. The concentration of ions does not affect the position of the peaks of the RDF, as shown in Fig. 3 *d*, so other concentrations could also have been used; a high concentration is advantageous because it yields RDFs with less noise, given the same simulation time. The backbone atoms were constrained to the coordinates of the 1DOI crystal structure so that any differences in the $K^+\cdots$ carboxylate interaction between simulations with different $f_{R_{min,K^+O}}$-values reflect the impact of the simulation parameters and not steric changes arising from changes in protein conformation.

For each of the five acidic amino acids functioning as the contact site with K^+ in the crystal structure (Fig. 4 *a*), we calculated the carboxylate(O)-K^+ RDF.
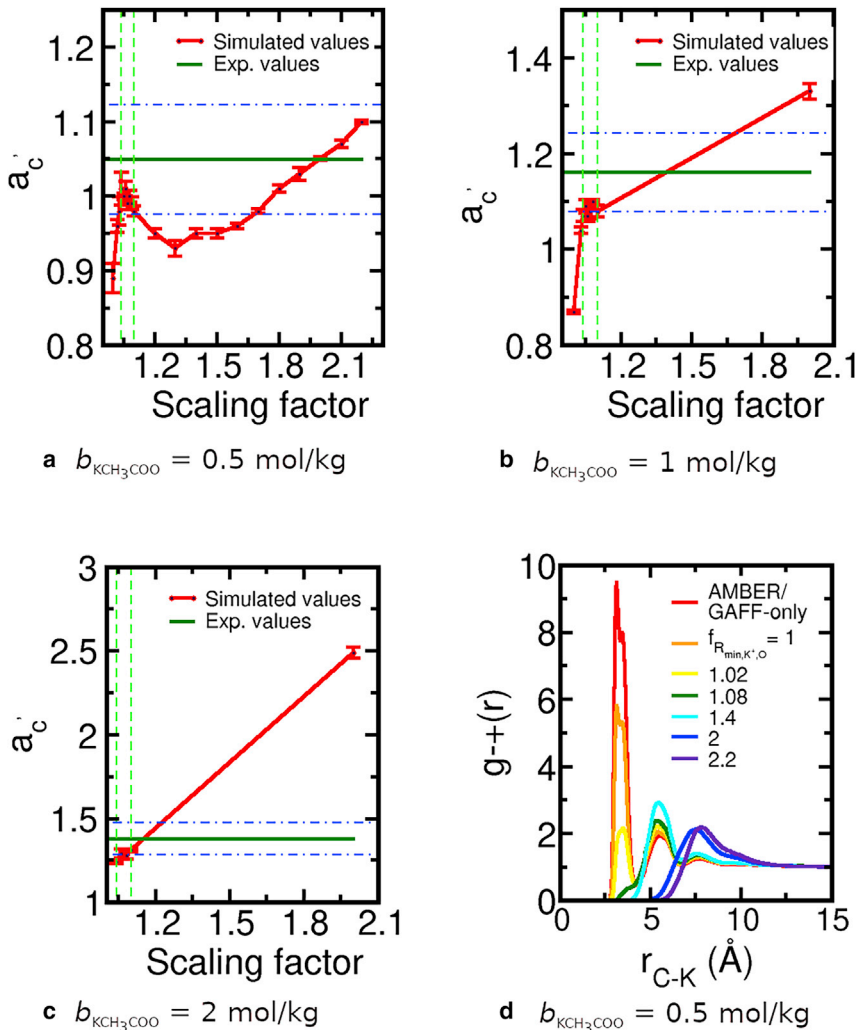
FIGURE 3 (*a–c*) Molar solution activity derivative (*red points*) of potassium acetate solutions as a function of the multiplicative scaling factor ($f_{R_{\min,K^+O}}$; see Eq. 12) applied to the LJ $R_{\min,K^+O,LB}$ parameter governing the interactions between K^+ and carboxylates. The error bars are the standard error of the mean calculated from three independent production simulations. The red lines are a guide to the eye. The green line shows the experimental reference value; see also Table 2. The area between the two horizontal dashed lines regions show the $\pm 7\%$ deviation from the experimental value. The two vertical dashed lines delimit the range of scaling factors acceptable for the three concentrations. (*d*) RDF of potassium and the carbon bonded to the oxygens of acetate (the same simulations as in *a*) for the indicated values of the scaling factor $f_{R_{\min,K^+O}}$. To see this figure in color, go online.

Fig. 4 *b* shows example RDFs for one of the protein sites and for different values of $f_{R_{\min,K^+O}}$. The position, r_{sim}, of the maximum of the first peak of each RDF was obtained as the mean of a Gaussian curve fitted to each peak. This fit allows a better identification of the position of the peak than taking the data point corresponding to the global maximum of each curve. The r_{sim} distances and the corresponding reference values (r_{cryst}) from the crystal are tabulated in Table S1 for each of the protein sites; r_{cryst} is the distance between K^+ and the carboxylate oxygen identified in Fig. 4 *c*. The r_{sim} are always smaller than the reference values for all parameter values tested. The optimal value of $f_{R_{\min,K^+O}}$ is selected as that which leads to the smallest deviation between r_{sim} and r_{cryst}. In Fig. 4 *c*, we show the unsigned relative deviation between r_{sim} and the reference value from the crystal structure r_{cryst} for the five protein sites for multiple values of $f_{R_{\min,K^+O}}$. A scaling factor $f_{R_{\min,K^+O}} = 1.08$ (corresponding to $R^\dagger_{\min,K^+O}$; Table 1) yields the lowest deviation over all protein sites. Table 2 shows the activity derivative

of potassium acetate solutions at the three concentrations investigated, for $f_{R_{\min,K^+O}} = 1.08$. These values deviate 4.7, 5.3, and 7.2% from the corresponding experimental values for $b_{KCH^3COO} = (0.5, 1, \text{and } 2)$ mol/kg, respectively. The position of the contact ion pair peak between the carboxylate oxygens and K^+ obtained with the optimal parameter is $r_{\text{sim}} \approx 2.8$ Å, consistent with ab initio calculations using a polarizable continuum model for the solvent (65,66) but larger than the optimal distance of 2.5 Å observed in gas phase calculations (67). Despite the difference in optimal distance observed between those ab initio calculations, both predict that the contact ion pair between carboxylate and K^+ is less stable than that with Na^+, i.e., these cations follow the normal Hofmeister series when interacting with the carboxylate group. Those results agree with x-ray absorption measurements of acetate solutions (66), with prior molecular simulations using parameters optimized for lower concentrations (68) and with our simulations using our optimized parameters (Fig. S7).
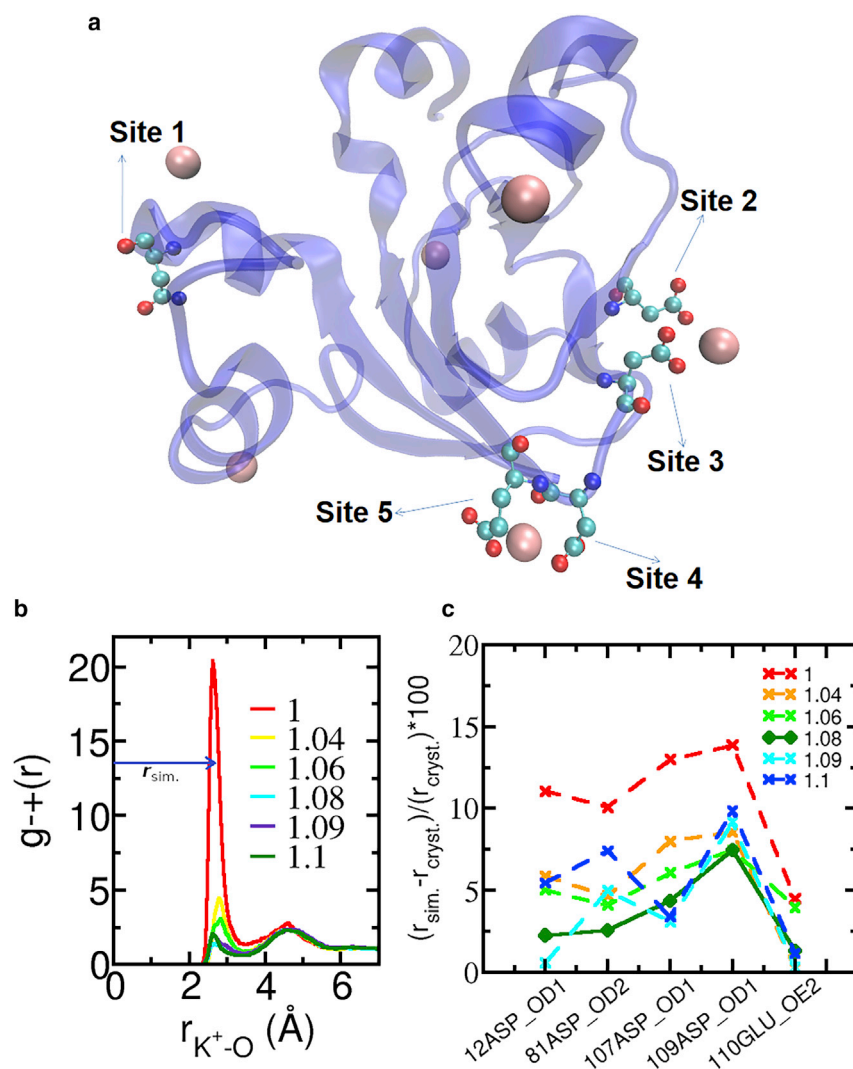
FIGURE 4 (*a*) Crystal structure (PDB: 1DOI) of the halophilic 2Fe-2S ferredoxin from *Haloarcula marismortui* (24); the K^+ ions are shown in pink. The five acidic amino acid sites that have nearby K^+, used to parameterize $R_{\min,K^+O}$, are indicated. (*b*) Example RDF of K^+ and the carboxylate oxygens of site 2 at $T = 298$ K and $c_{KCl} = 1$ mol/dm^3. The distance to the first maximum is identified as $r_{sim.}$. The legend shows the values of the scaling factor $f_{R_{\min,K^+O}}$. (*c*) Unsigned relative deviation between r_{sim} and r_{cryst} for the five indicated sites of the halophilic ferredoxin. The protein site is identified by the residue number, residue name, and oxygen name. The legend shows the values of the scaling factor $f_{R_{\min,K^+O}}$. Numerical data are shown in Table S1. To see this figure in color, go online.

Structure and dynamics of the hydration shell of mesophilic versus halophilic proteins

The ion-solvent stabilization model and the solvent-only stabilization model, proposed to explain the function of excess acidic amino acids in halophilic proteins, imply that the structure and dynamics of the hydration shell of halophilic proteins differs from that of mesophilic ones: according to both models, excess acidic amino acids are indispensable in halophilic proteins, so they remain hydrated at high KCl concentrations. The solvent-only stabilization model proposes that acidic amino acids enhance hydration by direct interactions with the water and that ion-protein interactions are not relevant for hydration or for protein stability. In contrast, the ion-solvent stabilization model proposes that protein hydration is maintained, and the folded protein structure is stabilized, by cooperative interactions between the acidic amino acids, the cations in solution (K^+) and water. According to this model, these cooperative interactions

are possible only in the folded conformation of halophilic protein because only specific tertiary and quaternary protein structures enable them. These cooperative interactions are responsible for the large amounts of cations bound to halophilic proteins and should result in excessively slow dynamics of water of hydration around those proteins. In

TABLE 1   Recommended value of R_{min} for the LJ potential between K^+ and any carboxylate oxygen when using TIP3P water, the Joung and Cheatham (29) parameters for K^+, and the GAFF (28) or AMBER (27) force field with the optimized self-interaction parameters for carboxylates from (36)

Parameter	Source	Value
$R_{\min,K^+O}$ (Å)[a]	this work	3.6355
$f_{R_{\min,K^+O}}$ [b]	this work	1.08
$R_{\min,K^+O,LB}$ (Å)[c]	AMBER or GAFF	3.3662

[a]Given is the optimized parameter value.
[b]Given is the optimized scaling factor, defined in Eq. 12.
[c]Given from combination rules (see Eq. 2), provided for comparison.

TABLE 2 Molal and molar solution activity derivative of aqueous solutions of KCH$_3$COO with the indicated molality

b_{KCH3COO} (mol/kg)	$a'^{(b)\ \exp}_s$[a]	a'_s[b]	a'_c[c]	a'_c[d]
0.5	1.0143	1.0488	1.00 ± 0.01	0.83 ± 0.01
1	1.1229	1.1611	1.100 ± 0.004	NC
2	1.3345	1.3798	1.28 ± 0.02	NC

NC, not calculated.

[a]Given is the experimental molal activity derivative from Pitzer equations (48).

[b]Given is the experimental molar activity derivative from Pitzer equations and Eq. 11.

[c]Given is the simulation (this work) of the molar activity derivative using the optimized LJ parameters listed in Table 1 and the self-interaction parameters for carboxylate from (36). The uncertainty is the standard error of the mean, calculated from four independent simulations.

[d]Given is the simulation (this work) of the molar activity derivative using the AMBER force field or GAFF.

what follows, we compare the structure and dynamics of the hydration shells of five pairs of halophilic-mesophilic proteins and discuss the results in the context of the predictions and assumptions of both models.

Selection of halophilic-mesophilic protein pairs

The five protein pairs investigated here were selected considering the following: 1) availability of experimentally determined structures, 2) existence of experimental studies on their conformational changes and/or activity as a function of salt concentration, 3) availability of parameters for simulation, 4) diversity of size and surface charge density, and 5) similarity of the amino acid sequence and structure between each halophilic protein and its mesophilic pair. In Table S2, we list the pdb IDs of the proteins selected for the study and show the structural and sequence similarity between each of the five halophilic-mesophilic pairs. The pair sequence identity varies between 19 and 90%. Despite this broad range in sequence identity, the proteins in each pair share a common structure, visible in the narrow range of the root mean-square deviation (RMSD) of the C$^\alpha$ atoms of the backbone of the amino acids between protein pairs, after structural alignment: $0.95 < \text{RMSD/Å} < 2.41$.

A comparison of four sets of orthologous proteins–homologous protein sequences that share the same ancestral sequence separated by a speciation event–shows that halophilic proteins have typically 17–20% acidic amino acids and only 8–10% basic amino acids, whereas their orthologous mesophilic proteins are weakly negatively charged, having 12–14% acidic and 10–12% basic amino acids. Table S3 shows the length of each sequence, the number of acidic and basic amino acids, and the total charge of the proteins investigated here; these data show that the sets of halophilic and mesophilic proteins are representative of their respective category. The halophilic proteins range from the very highly charged ferredoxin ($-29\ e$), in which 27% of amino acids are acidic and only 5% of amino acids are basic, to dihydrofolate reduc-

tase ($-15\ e$), with only 19 and 9% of acidic and basic amino acids, respectively. The mesophilic proteins have an acidic amino acid content ranging from 8 (β-lactamase) to 18% (ferredoxin) and are on average only weakly negative. Both ferredoxin proteins are outliers: the halophilic ferredoxin is the most highly charged protein known, and even the mesophilic ferredoxin has an unusually high negative charge.

Structural stability

We simulate each protein at both low ($b_{\text{KCl}} = 0.15$ mol/kg) and high ($b_{\text{KCl}} = 2$ mol/kg) concentration of KCl. The mesophilic proteins at high salt concentration and the halophilic ones at low salt concentration are thus under nonnatural conditions. Experiments have shown that the structure of proteins is typically more stable, and proteins have higher activity when in media with their natural KCl concentration (5). Mesophilic proteins are expected to have low solubility at high KCl or NaCl concentrations; under those conditions, many mesophilic proteins have a negative osmotic second virial coefficient (6,7), known to correlate strongly with low protein solubility (8).

Within the duration of the production simulations—0.5 μs for the carbonic anhydrase proteins and 1 μs for the others—all proteins retain their structure irrespective of the concentration of KCl. Comparison with experimental data suggests that under non-natural electrolyte conditions, the native fold of some of the proteins studied here indeed continues to be the stable state. This seems to be the case for the halophilic and mesophilic carbonic anhydrases (12) and the halophilic β-lactamase (69) studied here: experimental measurements of protein activity indicate that these proteins remain active–and thus presumably folded and in soluble form–under non-natural KCl concentrations. For other proteins, however, the conformational stability seen in the simulation may reflect a metastable state with lifetime longer than the simulation time. This may be the case for the halophilic protein L: experimental studies show that it is largely unfolded at low salt concentrations (23), even though it remains folded during the simulation.

Solvation layer structure

We first assess how hydrogen bonds donated by water molecules to the protein (termed water-protein hydrogen bonds) are affected by salt concentration. We focus our attention on strong hydrogen bonds (70), present if the distance between the water oxygen and the hydrogen bond acceptor (N, O) is below 3 Å, and the angle formed by the water hydroxy, and the acceptor is larger than 135°. In Fig. 5, we show the surface density, σ_{HB}, of water-protein hydrogen bonds, calculated as

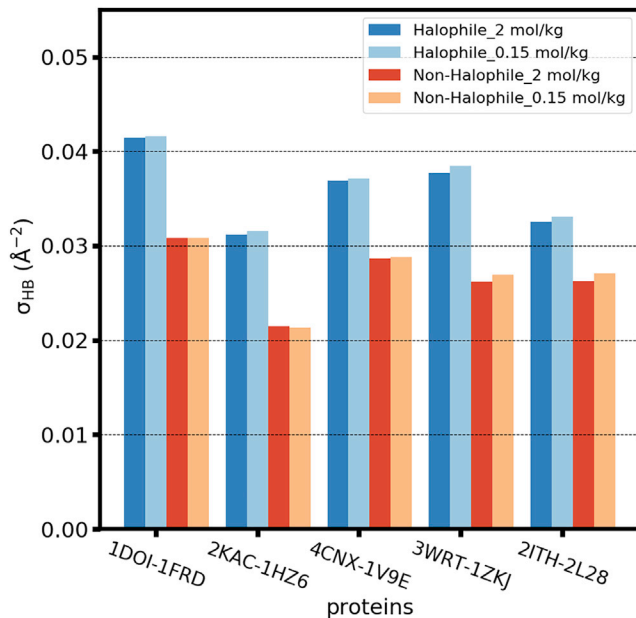$$\sigma_{\text{HB}} = \frac{\overline{n_{\text{HB}}}}{SASA}, \tag{13}$$

FIGURE 5 Surface density of water-protein hydrogen bonds for the indicated halophilic-mesophilic proteins, identified by their pdb ID, for different KCl concentrations. To see this figure in color, go online.

where $\overline{n_{HB}}$ is the time-averaged total number of water-protein hydrogen bonds and $\overline{SASA}$ is the time-averaged solvent-accessible surface area of each protein at the position of the first maximum of the hydration layer of each protein. Further details about the SASA calculation are provided below in the text accompanying Eq. 14.

Surface density of water-protein hydrogen bonds varies substantially with protein identity but is insensitive to KCl concentration

Fig. 5 shows that all halophilic proteins in this study accept significantly more hydrogen bonds from water, per unit area, than their nonhalophilic counterparts. The ferredoxin proteins, in particular, have the highest surface density of water-protein hydrogen bonds than any other protein in their halophilic or mesophilic cohort. In Fig. S1, we show that the number of water-protein hydrogen bonds for each amino acid type (acidic, basic, polar, and apolar) is only weakly sensitive to the identity of the protein and is highest for acidic amino acids. The simulation results are consistent with NMR experiments that demonstrate that the number of water molecules that remain unfrozen at $T \ll 0\,^{\circ}C$ in homopolypeptide solutions—i.e., water molecules perturbed by the homopolypeptides—is substantially higher for the acidic amino acids [71]. The variability of the surface density of water-protein hydrogen bonds with protein identity thus largely originates from the different content in acidic amino acids of each protein.

This fact has been used to propose that excess acidic amino acids in halophilic proteins is indispensable to compete with the electrolyte in solution for the available wa-

ter, thus ensuring that the proteins remain hydrated at high salt concentrations [1,22,24]. That possibility, however, is not supported by the results in Fig. 5 and Fig. S1: the surface density of water-protein hydrogen bonds and the number of water-protein hydrogen bonds formed by acidic amino acids are almost identical at $b_{KCl} = 0.15$ mol/kg and at $b_{KCl} = 2$ mol/kg.

But why does K^+ in solution not compete with the protein for available water at high KCl concentration? To answer this question, we calculate the proximal (also called the perpendicular [72]) number density, $\rho(r)$, of solvent species at distance r to the protein surface:

$$\rho_X(r) = \frac{\overline{n_X(r)}}{dr\,\overline{SASA(r)}} \qquad (14)$$

$\overline{n_X(r)}$ is the time-averaged number of solvent species X within a shell of thickness $dr = 0.05$ Å and at distance r of the protein surface. This quantity is calculated by determining the distance between each X and the nearest nonhydrogen protein atom, as implemented in the rdf function in Gromacs 2018; the position of the oxygen atom is used in the calculation of the number density of water. $\overline{SASA(r)}$ is the time-averaged protein solvent-accessible surface area, calculated using tcl scripts implemented in VMD; each point of the surface is at distance r from the nearest nonhydrogen protein atom.

Concentration of water near proteins is highest around halophilic proteins but is insensitive to KCl concentration

In Fig. 6, we show the proximal water number density around the halophilic (Fig. 6 *a*) and mesophilic (Fig. 6 *b*) proteins. The colored curves are obtained at $b_{KCl} = 2$ mol/kg; for $r < 0.45$ nm, overlapping each colored curve is a black, dashed one, obtained from simulations of the same protein simulated at $b_{KCl} = 0.15$ mol/kg. The height of the peaks of the proximal water number density is sensitive to protein identity, and the peaks are on average higher around the halophilic proteins. Irrespective of the halophilic or mesophilic character of each protein, however, the concentration of water in its hydration layer is insensitive to the KCl concentration in the bulk. Moreover, for all proteins and at both KCl concentrations, the water content for $r < 0.9$ nm exceeds the concentration of water in the bulk. These results indicate that proteins do not compete with ions in solution for available water and that mesophilic proteins, despite their lower content in acidic amino acids, are able to retain their hydration layer even at high KCl concentrations.

Halophilic and mesophilic proteins perturb water structure to the same length scale

The curves in Fig. 6 retain the same qualitative features despite the different protein sizes and charges. This similarity does not support the hypothesis proposed by several authors [17,19] that the hydration of halophilic proteins is qualitatively different from that of mesophilic ones and that
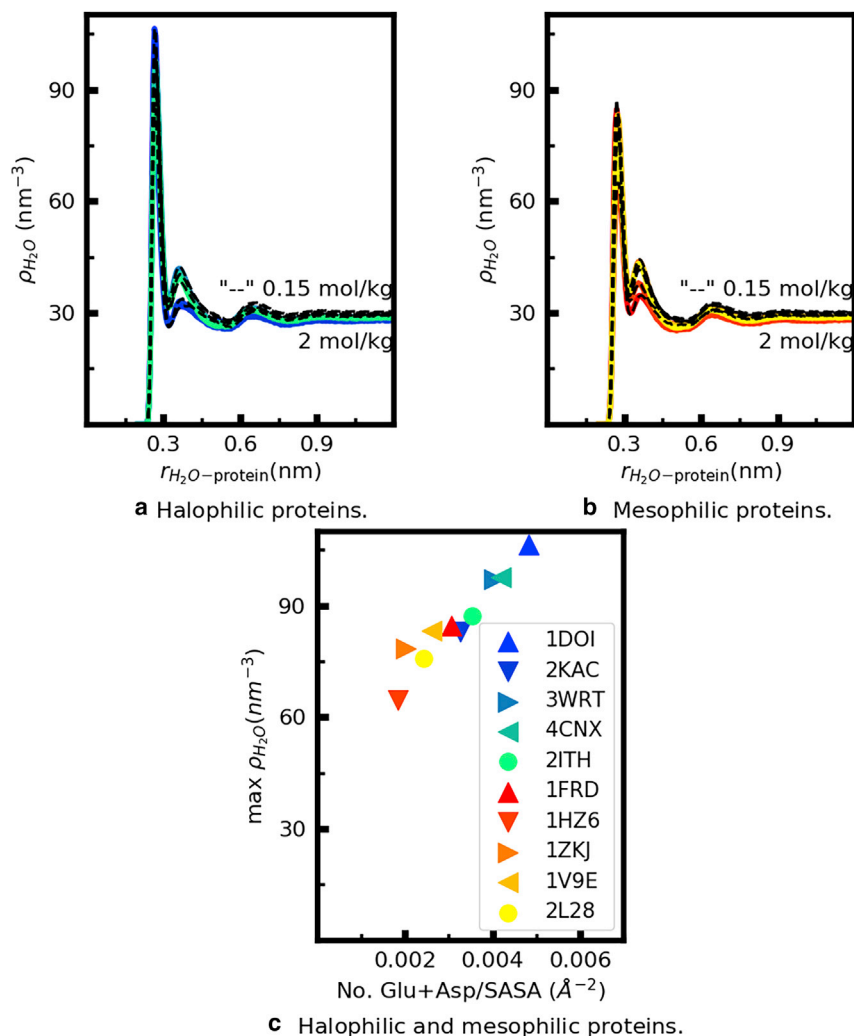
a Halophilic proteins.

b Mesophilic proteins.

c Halophilic and mesophilic proteins.

FIGURE 6 (*a* and *b*) Proximal number density of water molecules as a function of the distance to the surface of the indicated proteins, simulated at $b_{KCl} = 2$ mol/kg (*color*) and $b_{KCl} = 0.15$ mol/kg (*dashed black lines*). (*c*) Height of the first peak of the number density curves for $b_{KCl} = 2$ mol/kg, shown in the other panels, as a function of the surface density of acidic amino acids. Each color corresponds to a protein, identified by its pdb ID in the legend of the bottom panel. To see this figure in color, go online.

halophilic proteins perturb water structure to particularly large length scales (17). Instead, it is in line with reported calculations of proximal RDFs of water around different proteins, from molecular dynamics simulations and from pdb structures, that indicate that this function has a universal character (72,73). According to this view, differences between the proximal rdfs of different proteins reflect different protein shapes and amino acid composition (72,73). Fig. 6 *c* shows that for our set of proteins, there is a strong correlation between the first maximum of $\rho_{H_2O}(r)$ and the surface density of acidic amino acids, i.e., between protein hydration and its content in acidic amino acids. Our results suggest that proximal rdfs may be confidently used to aid x-ray refinements of halophilic proteins, using models such as that in (72).

Potassium integrates the first hydration shell of both halophilic and mesophilic proteins in a charge-density-dependent manner

Fig. 7, *a* and *b* shows the proximal number density of K$^+$, $\rho_{K^+}(r)$, around each protein. The distributions clarify that

K$^+$ accumulates in the vicinity ($r < 0.9$ nm) of each protein to concentrations between two and six times higher than in the bulk, depending on the protein identity and the bulk concentration of KCl. The position of the first peak of $\rho_{K^+}(r)$ coincides with that of $\rho_{H_2O}(r)$ at both low and high c_{KCl}. Nevertheless, K$^+$ cannot measurably displace water from the first hydration shell of the protein because it does not accumulate near the protein in sufficiently large amounts: the height of the peaks of $\rho_{K^+}(r)$ is always below 3 ions/nm^3; in contrast, the maximal value of $\rho_{H_2O}(r)$ is ~105 molecules/nm^3. Rather than displacing water, K$^+$ integrates its own first hydration shell with that of the protein.

Fig. 7, *a* and *b* also shows that the density of K$^+$ in the vicinity of the protein depends strongly on protein identity and is highest for the halophilic proteins. This origin of dependence is clarified in Fig. 7 *c*, which shows that the height of the first peak of $\rho_{K^+}(r)$ strongly correlates with the charge density of the protein. In other words, the total content of K$^+$ in the solvation shell of proteins correlates both with protein net charge and surface area. Halophilic
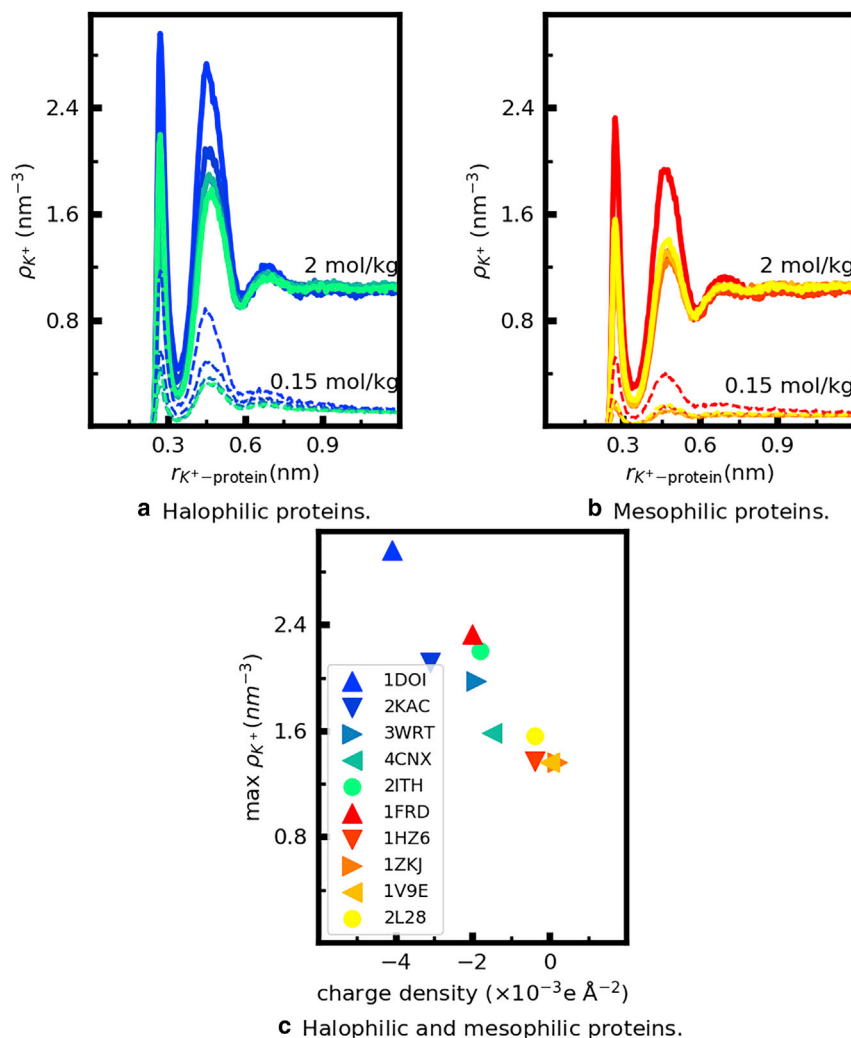
a Halophilic proteins.

b Mesophilic proteins.

c Halophilic and mesophilic proteins.

FIGURE 7 (*a* and *b*) Proximal number density of K^+ as a function of the distance to the surface of the indicated proteins in solutions with the indicated molality of KCl. (*c*) Height of the first peak of the number density curves for $b_{KCl} = 2$ mol/kg shown in the other panels as a function of the protein charge density. Each color corresponds to a protein, identified by its pdb ID in the legend of the bottom panel. To see this figure in color, go online.

proteins are, on average, larger and have a higher net negative charge than mesophilic ones and, therefore, a larger amount of K^+ in their solvation layers, as quantified by the cumulative number of potassium ions in the vicinity of each protein shown in Fig. S2.

Dynamics of the protein solvation shell

We evaluate the impact of salt concentration and protein identity on solvent translational dynamics by calculating the $MSD(\tau)$ of water or of K^+ according to the following:

$$MSD(\tau) = \left\langle \left(\vec{x}(t+\tau) - \vec{x}(t) \right)^2 \right\rangle \quad (15)$$

$\vec{x}(t)$ is the position of each species at time t and the bracket indicates both a time and an ensemble average over the relevant species population. We calculate the MSD for water or K^+ belonging to the first solvation layer at $t = 0$ because if halophilic proteins indeed substantially slow down water dynamics, this effect should be strongest for these subpopula-

tions. We are interested in the short-time dynamics of each species because this dynamics should reflect the strongest impact of their initial position in the hydration shell and thus differ the most from the dynamics of the same species averaged over all its elements in the simulation box. For this reason, it is sufficient to calculate the MSD using coordinates wrapped back into the main simulation cell.

In Figs. S3 and S4, we show the MSD of water and of K^+ in the first hydration layer around the halophilic and mesophilic ferredoxin proteins; the results for the other proteins (data not shown) are qualitatively similar. As expected under these conditions, the curves saturate at long times, which is evidence of confinement to the main simulation cell. To facilitate comparisons between different proteins and simulation conditions, we calculate the diffusion coefficient, D, as

$$MSD(\tau) = 6 D \tau + C, \quad (16)$$

by fitting each $MSD(\tau)$ curve in the interval $\tau = [20, 30]$ ps. The constant C is added to account for nondiffusive

movement at short times. This time interval is chosen because the dynamics of all species is already diffusive beyond 20 ps, but the impact of confinement by the finite simulation box remains marginal much beyond 30 ps (see example in Fig. S3). To facilitate comparisons, the diffusion coefficient averaged over all K^+ or water molecules in the simulation box (calculated using unwrapped coordinates) is estimated from fits using the same time interval.

Halophilic and mesophilic proteins have solvation layers with identical translational dynamics at both low and high KCl concentrations

In Fig. 8, we compare the diffusion coefficients of water molecules that initially belong to the first hydration layer of each protein with the diffusion coefficient of all water molecules in each simulation, at $b_{KCl} = 2$ mol/kg; results for $b_{KCl} = 0.15$ mol/kg are shown in Fig. S5. The TIP3P water model is known to predict translational and rotational dynamics, which are approximately twice faster than the reference experimental values (74). Accordingly, the absolute values of D_{H_2O} reported in Fig. 8 are likely overestimated; only their relative magnitude is meaningful. The translational dynamics in the first hydration layer is indeed lower than that of water in the bulk but only by a factor of 1/2 at most. Moreover, the translational dynamics of water in the first hydration layer of the mesophilic proteins is barely faster that of the nonhalophilic proteins. Our results are not consistent with the possibility that water near halophilic proteins has dramatically slower dynamics than around nonhalophilic ones.

Fig. 9 compares the self-diffusion coefficients of potassium ions initially in the first solvation layer of the protein with those averaged over all K^+ in the simulation box, again at the highest salt concentration (see Fig. S6 for the results at $b_{KCl} = 0.15$ mol/kg). Similar to the trends observed for water dynamics, the translational dynamics of K^+ in the first solvation layer is lower than dynamics in the bulk but only by a factor of 1/3. Moreover, the diffusion coefficients

of this subpopulation are very similar for the halophilic and mesophilic proteins. Even though halophilic and mesophilic proteins differ substantially in surface charge density and these differences lead to considerable differences in the number of K^+ in the solvation layer of each protein (Fig. S2), they do not impact the translational dynamics of the ions near the protein.

DISCUSSION

Both the solvent-only and the ion-solvent stabilization models propose that excess acidic amino acids are necessary in halophilic proteins to maintain their hydration at high salt concentrations. Our results do not support this scenario. Halophilic proteins do accept more hydrogen bonds from water (Fig. 5) and have larger concentration of water in their hydration shells (Fig. 6) because of their higher content in acidic amino acids (Fig. 6 c; Fig. S1; Table S3). These results are consistent with sedimentation measurements (18,19), which indicate that halophilic proteins bind larger amounts of water than mesophilic ones. That fact does not imply that excess acidic amino acids are necessary to maintain protein hydration at high KCl concentration, however. Our study demonstrates that all mesophilic and halophilic proteins investigated here remain equally well hydrated at both low ($b_{KCl} = 0.15$ mol/kg) and high ($b_{KCl} = 2$ mol/kg) KCl concentration, as measured by the number of hydrogen bonds that they accept from water (Fig. 5) and the water concentration at a distance below 4.5 Å from the protein surface (Fig. 6). These results imply that the proteins simulated here do not compete with ions in solution for available water even at concentrations as high as $b_{KCl} = 2$ mol/kg. The force fields used here for carboxylates and K^+ ((36) and this work) were developed to reproduce experimentally determined hydration free energies, lending confidence to our observations. The proteins selected for this study have diverse sizes, net charges, and surface charge density and include both typical examples of halophilic
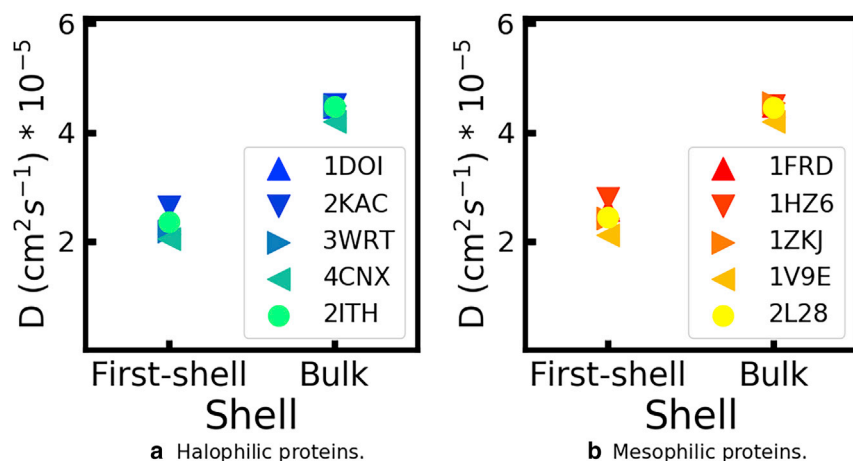

FIGURE 8 Diffusion coefficients of water around the indicated proteins, simulated in $b_{KCl} = 2$ mol/kg. The first shell consists of water molecules that, at $t = 0$, belong to the first hydration shell of the proteins; bulk consists of all water molecules in the same simulation. To see this figure in color, go online.
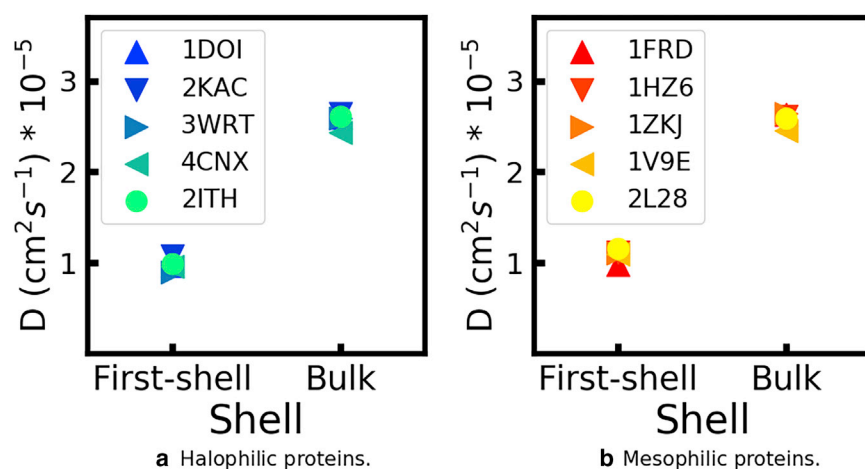
a Halophilic proteins.

b Mesophilic proteins.

FIGURE 9 Diffusion coefficients of potassium ions around the indicated proteins, simulated in $b_{KCl} = 2$ mol/kg. The first shell consists of potassium ions that, at $t = 0$, belong to the first solvation shell of the proteins; bulk consists of all potassium ions in the same simulation. To see this figure in color, go online.

and mesophilic proteins as well as extreme examples (the ferredoxin proteins) of each category (Table S3). The conclusion that protein hydration is concentration independent for $b_{KCl} \leq 2$ mol/kg and that proteins do not compete with KCl for available water thus may apply to all proteins.

Moreover, we suggest that this conclusion may hold also for aqueous NaCl solutions because the water activity of NaCl and KCl solutions is very similar (Fig. S8, based on (75)). The limited experimental data comparing protein hydration in NaCl vs. KCl solutions is consistent with this possibility: neutron scattering and ultracentrifugation studies show that halophilic malate dehydrogenase binds equally large amounts of water when immersed in KCl or in NaCl solutions (18). We note, however, that Na^+ is expected to form more contact ion pairs with carboxylates than K^+ (compare the acetate..K^+ and acetate..Na^+ RDFs shown in Fig. S7, and also prior work by others (65,66,68)). Differences in the stability and dynamics of halophilic proteins in NaCl and KCl solutions (76) likely reflect the different interactions between the cations in solution and the acidic amino acids in the protein. It is of interest to comparatively characterize protein solvation shells in KCl vs. NaCl solutions using nonperturbative methods such as solvation-shell spectroscopy (77) and to understand how their differences influence the conformational stability and the functionality of enzymes. This understanding is critical to rationally develop proteins for use in biotechnological devices (e.g., to produce molecular hydrogen) that function in brackish or sea water, thus reducing pressure on our planet's freshwater resources.

The water activity decreases from 1 to 0.8 as the concentration of KCl increases to $b_{KCl} \leq 4$ mol/kg (Fig. S8). Low water activity also occurs in solvent mixtures composed of water and organic liquids. At first sight, halophilic proteins offer clues to understand which sequence-structure features enable the ability to function at low water activity arising from the presence of organic solvents; in fact, some halophilic proteins remain functional under these conditions (78,79). A closer look into the limited available data, how-

ever, suggests that protein hydration may respond very differently to the two environments and that further studies are necessary before drawing analogies between the proteins at high salt concentration and proteins in mixtures of water with organic solvents. Our results indicate that protein hydration is insensitive to KCl concentrations up to $b_{KCl} = 2$ mol/kg, i.e., protein hydration levels are unaffected by changes in water activity between 1 and 0.9. In contrast, the fraction of water bound to proteins decreases from 60% to 40% (weight percentage of bound water in the protein+bound water system isolated by centrifugation) in solutions of water with immiscible organic solvents (80,81) as the water activity decreases from 1 to 0.9. It is at present unclear whether the different dependence between protein hydration and water activity is real or whether it reflects a bias imposed by the different observables being compared. There is a clear need for nonperturbative experiments to comparatively characterize the hydration shells of proteins in aqueous electrolyte media and in aqueous media containing organic solvents to assert the extent to which halophilic proteins may be useful starting points to design enzymes and catalysts that function in low water activity environments that occur, e.g., when producing ethanol in large quantities or when catalyzing reactions in the presence of organic solvents.

The ion-solvent stabilization hypothesis proposes that cooperative ion-solvent-protein networks arise in halophilic proteins because of their tertiary and quaternary structure (18,19); cooperative interactions would not exist in the unfolded or nonoligomerized state of the protein because in that state, the acidic amino acids are further apart or in unfavorable orientation. These networks have been proposed to explain experimental observations (based on analysis of centrifugates), showing that halophilic proteins bind more cations than mesophilic ones (18,19). Our simulation results agree with the experiment: halophilic proteins indeed have larger amounts of K^+ in their solvation shell than mesophilic proteins (Fig. 7). The fact that the maximum of the

number density of K^+ around the proteins correlates linearly with the charge density of the protein (Fig. 7 c) suggests that cation binding has a predominantly electrostatic origin. We are currently investigating whether protein-ion-solvent interactions indeed have a cooperative component.

Neutron scattering experiments have suggested that the cytoplasm of halophilic organisms has a fraction of water with translational diffusion ~250 times slower than that of water in the bulk; in contrast, that slow component was not observed in mesophilic organisms (21). A fraction of water with extremely slow dynamics has been proposed to arise in the context of the ion-solvent stabilization model: the cooperative ion-water-protein networks would strongly reduce the mobility of water in the first hydration layer of the protein (17). Our results do not support this view: the translational dynamics of water and of K^+ near halophilic and mesophilic proteins is indistinguishable, and only two to three times slower than the dynamics of the same species in the bulk (Figs. 8 and 9). Our simulations are consistent with ^{17}O magnetic relaxation measurements of halophilic and mesophilic versions of protein L (one of the protein pairs simulated here), which show that they have similar hydration dynamics (23). This result does not imply that cooperative ion-solvent-protein interactions are necessarily absent at high KCl concentrations. Cooperative interactions may well exist, but our results indicate that they should not result on unusually slow solvent dynamics, i.e., they do not come at the expense of a high entropic price.

CONCLUSIONS

Many halophilic organisms contain molar concentrations of KCl in their cytoplasm, necessary to balance the large osmotic pressure induced by equally high external concentrations of NaCl (5). The cytoplasmic proteins of these organisms are substantially richer in acidic amino acids than mesophilic ones (1). Because acidic amino acids can bind substantially larger amounts of water than any other natural amino acid (71), it has been proposed that halophilic proteins require acidic amino acids to remain hydrated in their low water activity environment (1,24). This work does not support this possibility. Our simulations of five halophilic proteins and five mesophilic counterparts indicate that halophilic proteins indeed contain larger amounts of water in their hydration shells than mesophilic ones and that their larger hydration level correlates with their high content in acidic amino acids. However, all proteins remained equally hydrated at low ($b_{KCl} \leq 0.15$ mol/kg) and high ($b_{KCl} \leq 2$ mol/kg) KCl concentrations, demonstrating that a higher content in acidic amino acids is not necessary to remain hydrated at high KCl concentrations. We note, however, that to understand the connection between acidic amino acids, solvation, and protein stability, it is also necessary to investigate the unfolded state of proteins; it is the difference in protein hydration and protein-salt interactions between the folded state and the unfolded ensemble that matters (82,83). We will focus on this point in future studies.

Cooperative interactions between acidic amino acids, water and cations have been proposed to exist in the folded structure of halophilic proteins because their high content in acidic amino acids would enable specific, favorable ion-water-carboxylate configurations to form (18,19,22). According to this view, these interactions would stabilize the folded protein configuration, they would be necessary to maintain the protein hydrated at high KCl concentration (18,19,22), and would manifest themselves in a fraction of water with very slow translational dynamics (21). Our simulations show that halophilic proteins have a higher concentration of K^+ in their solvation shells than mesophilic ones, consistent with experiment (18,19). The concentration of K^+ in the protein solvation shell linearly correlates with the protein net charge/SASA ratio. Halophilic proteins are more negative than mesophilic ones, hence the higher concentration of K^+ in their hydration shells. Despite this high concentration, the solvation shell of halophilic proteins remains as labile as that of mesophilic ones, and we find no evidence of water or ions with unusually slow translational dynamics around halophilic proteins. The absence of slow dynamics, we point out, does not imply that cooperativity is impossible; it suggests, however, that if present, it does not unusually slow down water dynamics. If a stabilizing interaction between acidic amino acids (mediated by K^+) indeed exists and is more prevalent in folded rather than unfolded protein conformations, it might be one of the driving forces behind the abundance of acidic amino acids in halophilic proteins. Conversely, if the interactions between acidic amino acids are predominantly destabilizing, they might be necessary to retain protein flexibility (17) or simply to prevent aggregation in a media that greatly enhances the hydrophobic effect (15). We are currently investigating whether water-cation-carboxylate interactions in halophilic proteins have a cooperative nature and the impact of the content in acidic amino acids on protein conformational dynamics and protein-protein interaction.

AUTHOR CONTRIBUTIONS

A.V.V. designed and guided the research. H.G.D. performed the research. Both authors wrote the manuscript.

ACKNOWLEDGMENTS

REFERENCES

1. Deole, R., J. Challacombe, …, W. D. Hoff. 2013. An extremely halophilic proteobacterium combines a highly acidic proteome with a low cytoplasmic potassium content. *J. Biol. Chem.* 288:581–588.

2. van der Wielen, P. W. J. J., H. Bolhuis, …, K. N. Timmis; BioDeep Scientific Party. 2005. The enigma of prokaryotic life in deep hypersaline anoxic basins. *Science.* 307:121–123.

3. Graziano, G., and A. Merlino. 2014. Molecular bases of protein halotolerance. *Biochim. Biophys. Acta.* 1844:850–858.

4. Gunde-Cimermana, N., P. Zalarb, …, A. Plemenitasd. 2000. Hypersaline waters in salterns - natural ecological niches for halophilic black yeasts. *FEMS Microbiol. Ecol.* 32:235–240.

5. Lanyi, J. K. 1974. Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriol. Rev.* 38:272–290.

6. Tessier, P. M., and A. M. Lenhoff. 2003. Measurements of protein self-association as a guide to crystallization. *Curr. Opin. Biotechnol.* 14:512–516.

7. Dumetz, A. C., A. M. Snellinger-O'brien, …, A. M. Lenhoff. 2007. Patterns of protein protein interactions in salt solutions and implications for protein crystallization. *Protein Sci.* 16:1867–1877.

8. Guo, B., S. Kao, …, W. William Wilson. 1999. Correlation of second virial coefficients and solubilities useful in protein crystal growth. *J. Cryst. Growth.* 196:424–433.

9. Kennedy, S. P., W. V. Ng, …, S. DasSarma. 2001. Understanding the adaptation of Halobacterium species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Res.* 11:1641–1650.

10. Paul, S., S. K. Bag, …, C. Dutta. 2008. Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. *Genome Biol.* 9:R70.

11. Siglioccolo, A., A. Paiardini, …, S. Pascarella. 2011. Structural adaptation of extreme halophilic proteins through decrease of conserved hydrophobic contact surface. *BMC Struct. Biol.* 11:50.

12. Warden, A. C., M. Williams, …, V. S. Haritos. 2015. Rational engineering of a mesohalophilic carbonic anhydrase to an extreme halotolerant biocatalyst. *Nat. Commun.* 6:10278.

13. Madern, D., and G. Zaccai. 2004. Molecular adaptation: the malate dehydrogenase from the extreme halophilic bacterium Salinibacter ruber behaves like a non-halophilic protein. *Biochimie.* 86:295–303.

14. Date, M. S., and B. N. Dominy. 2013. Modeling the influence of salt on the hydrophobic effect and protein fold stability. *Commun. Comput. Phys.* 13:90–106.

15. Elcock, A. H., and J. A. McCammon. 1998. Electrostatic contributions to the stability of halophilic proteins. *J. Mol. Biol.* 280:731–748.

16. Tadeo, X., B. López-Méndez, …, O. Millet. 2009. Structural basis for the aminoacid composition of proteins from halophilic archaea. *PLoS Biol.* 7:e1000257.

17. Mevarech, M., F. Frolow, and L. M. Gloss. 2000. Halophilic enzymes: proteins with a grain of salt. *Biophys. Chem.* 86:155–164.

18. Calmettes, P., H. Eisenberg, and G. Zaccai. 1987. Structure of halophilic malate dehydrogenase in multimolar KCl solutions from neutron scattering and ultracentrifugation. *Biophys. Chem.* 26:279–290.

19. Pundak, S., and H. Eisenberg. 1981. Structure and activity of malate dehydrogenase from the extreme halophilic bacteria of the Dead Sea. 1. Conformation and interaction with water and salt between 5 M and 1 M NaCl concentration. *Eur. J. Biochem.* 118:463–470.

20. Madern, D., C. Ebel, and G. Zaccai. 2000. Halophilic adaptation of enzymes. *Extremophiles.* 4:91–98.

21. Jasnin, M., A. Stadler, …, G. Zaccai. 2010. Specific cellular water dynamics observed in vivo by neutron scattering and NMR. *Phys. Chem. Chem. Phys.* 12:10154–10160.

22. Zaccai, G., F. Cendrin, …, H. Eisenberg. 1989. Stabilization of halophilic malate dehydrogenase. *J. Mol. Biol.* 208:491–500.

23. Qvist, J., G. Ortega, …, B. Halle. 2012. Hydration dynamics of a halophilic protein in folded and unfolded states. *J. Phys. Chem. B.* 116:3436–3444.

24. Frolow, F., M. Harel, …, M. Shoham. 1996. Insights into protein adaptation to a saturated salt environment from the crystal structure of a halophilic 2Fe-2S ferredoxin. *Nat. Struct. Biol.* 3:452–458.

25. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

26. Cornell, W. D., P. Cieplak, …, P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197.

27. Maier, J. A., C. Martinez, …, C. Simmerling. 2015. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11:3696–3713.

28. Wang, J., R. M. Wolf, …, D. A. Case. 2004. Development and testing of a general amber force field. *J. Comput. Chem.* 25:1157–1174.

29. Joung, I. S., and T. E. Cheatham, III. 2008. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B.* 112:9020–9041.

30. Zhang, H., C. Yin, …, D. van der Spoel. 2018. Force field benchmark of amino acids: I. hydration and diffusion in different water models. *J. Chem. Inf. Model.* 58:1037–1052.

31. Zhang, H., Y. Jiang, …, C. Yin. 2018. Force field benchmark of amino acids. 2. partition coefficients between water and organic solvents. *J. Chem. Inf. Model.* 58:1669–1681.

32. Joung, I. S., and T. E. Cheatham, III. 2009. Molecular dynamics simulations of the dynamic and energetic properties of alkali and halide ions using water-model-specific ion parameters. *J. Phys. Chem. B.* 113:13279–13290.

33. Carvalho, A. T. P., A. F. S. Teixeira, and M. J. Ramos. 2013. Parameters for molecular dynamics simulations of iron-sulfur proteins. *J. Comput. Chem.* 34:1540–1548.

34. Peters, M. B., Y. Yang, …, K. M. Merz, Jr. 2010. Structural survey of zinc containing proteins and the development of the zinc AMBER force field (ZAFF). *J. Chem. Theory Comput.* 6:2935–2947.

35. Li, P., B. P. Roberts, …, K. M. Merz, Jr. 2013. Rational design of particle mesh ewald compatible Lennard-Jones parameters for +2 metal cations in explicit solvent. *J. Chem. Theory Comput.* 9:2733–2748.

36. Kashefolgheta, S., and A. Vila Verde. 2017. Developing force fields when experimental data is sparse: AMBER/GAFF-compatible parameters for inorganic and alkyl oxoanions. *Phys. Chem. Chem. Phys.* 19:20593–20607.

37. Kirkwood, J. G., and F. P. Buff. 1951. The statistical mechanical theory of solutions. I. *J. Chem. Phys.* 19:774–777.

38. Kusalik, P. G., and G. N. Patey. 1987. The thermodynamic properties of electrolyte solutions: some formal results. *J. Chem. Phys.* 86:5110–5116.

39. Klasczyk, B., and V. Knecht. 2010. Kirkwood-Buff derived force field for alkali chlorides in simple point charge water. *J. Chem. Phys.* 132:024109.

40. Gee, M. B., N. R. Cox, …, P. E. Smith. 2011. A Kirkwood-Buff derived force field for aqueous alkali halides. *J. Chem. Theory Comput.* 7:1369–1380.

41. Fyta, M., and R. R. Netz. 2012. Ionic force field optimization based on single-ion and ion-pair solvation properties: going beyond standard mixing rules. *J. Chem. Phys.* 136:124103–124111.

42. Vila Verde, A., M. Santer, and R. Lipowsky. 2016. Solvent-shared pairs of densely charged ions induce intense but short-range supra-additive slowdown of water rotation. *Phys. Chem. Chem. Phys.* 18:1918–1930.

43. Krüger, P., S. K. Schnell, …, J.-M. Simon. 2013. Kirkwood-Buff integrals for finite volumes. *J. Phys. Chem. Lett.* 4:235–238.

44. Ramshaw, J. D. 1980. Functional derivative relations for a finite non-uniform molecular fluid in the canonical ensemble. *Mol. Phys.* 41:219–227.

45. Hill, T. L. 1987. Statistical Mechanics: Principles and Selected Applications. Dover, New York.

46. Lebowitz, J. L., and J. K. Percus. 1961. Long-range correlations in a closed system with applications to nonuniform fluids. *Phys. Rev.* 122:1675–1691.

47. Robinson, R. A., and R. H. Stokes. 2002. Electrolyte Solutions, Second Edition. Dover Publications, Inc., New York.

48. Pitzer, K. S. 1991. Activity Coefficients in Electrolyte Solutions, Second Edition. CRC Press, London, UK.

49. Bart, H.-J. 2001. Reactive Extraction. Springer-Verlag, Berlin, Germany.

50. Bury, C. R., and G. A. Parry. 1935. The densities of aqueous solutions of potassium acetate and laurate. *J. Chem. Soc.* 0:626–628.

51. Berendsen, H. J. C., D. van der Spoel, and R. van Drunen. 1995. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91:43–56.

52. Van Der Spoel, D., E. Lindahl, …, H. J. C. Berendsen. 2005. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26:1701–1718.

53. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an NaNElog(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.

54. van Gunsteren, W. F., and H. J. C. Berendesen. 1988. A leap-frog algorithm for stochastic dynamics. *Mol. Simul.* 1:173–185.

55. Hess, B., H. Bekker, …, G. E. M. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.

56. Berendsen, H. J. C., J. P. M. Postma, …, J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.

57. Case, D., I Ben-Shalom, …, P. A. Kollman. 2018. Amber 2018 Reference Manual. University of California, San Francisco, CA.

58. Case, D. A., T. E. Cheatham, III, …, R. J. Woods. 2005. The Amber biomolecular simulation programs. *J. Comput. Chem.* 26:1668–1688.

59. Ryckaert, J. P., G. Ciccotti, and H. J. Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23:327–341.

60. Basconi, J. E., and M. R. Shirts. 2013. Effects of temperature control algorithms on transport properties and kinetics in molecular dynamics simulations. *J. Chem. Theory Comput.* 9:2887–2899.

61. Fournier, P., R. Gout, and E. H. Oelkers. 2005. A Raman spectrographic and potentiometric study of aqueous lithium and potassium acetate complexation at temperatures from 20 to 200°C. *J. Solution Chem.* 34:881–898.

62. Daniele, P. G., A. De Robertis, …, C. Rigano. 1985. On the possibility of determining the thermodynamic parameters for the formation of weak complexes using a simple model for the dependence on ionic strength of activity coefficients: Na^+, K+, and Ca^{2+} complexes of low molecular weight ligands in aqueous solution. *J. Chem. Soc. Dalton Trans.* 11:2353–2361.

63. Fuoss, R. M. 1958. Ionic association. III. The equilibrium between ion pairs and free ions. *J. Am. Chem. Soc.* 80:5059–5061.

64. Juers, D. H., and B. W. Matthews. 2001. Reversible lattice repacking illustrates the temperature dependence of macromolecular interactions. *J. Mol. Biol.* 311:851–862.

65. Vrbka, L., J. Vondrásek, …, P. Jungwirth. 2006. Quantification and rationalization of the higher affinity of sodium over potassium to protein surfaces. *Proc. Natl. Acad. Sci. USA.* 103:15440–15444.

66. Aziz, E. F., N. Ottosson, …, B. Winter. 2008. Cation-specific interactions with carboxylate in amino acid and acetate aqueous solutions: X-ray absorption and ab initio calculations. *J. Phys. Chem. B.* 112:12567–12570.

67. Stevens, M. J., and S. L. B. Rempe. 2016. Ion-specific effects in carboxylate binding sites. *J. Phys. Chem. B.* 120:12519–12530.

68. Hess, B., and N. F. A. van der Vegt. 2009. Cation specific binding with protein surface charges. *Proc. Natl. Acad. Sci. USA.* 106:13296–13300.

69. Arai, S., Y. Yonezawa, …, R. Kuroki. 2015. Structure of a highly acidic β-lactamase from the moderate halophile Chromohalobacter sp. 560 and the discovery of a Cs(+)-selective binding site. *Acta Crystallogr. D Biol. Crystallogr.* 71:541–554.

70. Tang, F., T. Ohto, …, Y. Nagata. 2018. Definition of free O-H groups of water at the air-water interface. *J. Chem. Theory Comput.* 14:357–364.

71. Kuntz, I. D. 1971. Hydration of macromolecules. III. Hydration of polypeptides. *J. Am. Chem. Soc.* 93:514–516.

72. Makarov, V. A., B. K. Andrews, and B. M. Pettitt. 1998. Reconstructing the protein-water interface. *Biopolymers.* 45:469–478.

73. Lin, B., and B. M. Pettitt. 2011. Note: on the universality of proximal radial distribution functions of proteins. *J. Chem. Phys.* 134:106101.

74. Mark, P., and L. Nilsson. 2001. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A.* 105:9954–9960.

75. Robinson, R. A., and R. H. Stokes. 1949. Tables of osmotic and activity coefficients of electrolytes in aqueous solution at 25°C. *Trans. Faraday Soc.* 45:612–624.

76. Tehei, M., D. Madern, …, G. Zaccai. 2001. Fast dynamics of halophilic malate dehydrogenase and BSA measured by neutron scattering under various solvent conditions influencing protein stability. *Proc. Natl. Acad. Sci. USA.* 98:14356–14361.

77. Sun, Y., and P. B. Petersen. 2017. Solvation shell structure of small molecules and proteins by IR-MCR spectroscopy. *J. Phys. Chem. Lett.* 8:611–614.

78. Karan, R., M. D. Capes, and S. Dassarma. 2012. Function and biotechnology of extremophilic enzymes in low water activity. *Aquat. Biosyst.* 8:4.

79. Fukushima, T., T. Mizuki, …, R. Usami. 2005. Organic solvent tolerance of halophilic α-amylase from a Haloarchaeon, Haloarcula sp. strain S-1. *Extremophiles.* 9:85–89.

80. Lee, S. B., and K.-J. Kim. 1995. Effect of water activity on enzyme hydration and enzyme reaction rate in organic solvents. *J. Ferment. Bioeng.* 79:473–478.

81. Zaks, A., and A. M. Klibanov. 1988. The effect of water on enzyme action in organic media. *J. Biol. Chem.* 263:8017–8021.

82. Smiatek, J. 2017. Aqueous ionic liquids and their effects on protein structures: an overview on recent theoretical and experimental results. *J. Phys. Condens. Matter.* 29:233001.

83. Diddens, D., V. Lesch, …, J. Smiatek. 2017. Aqueous ionic liquids and their influence on peptide conformations: denaturation and dehydration mechanisms. *Phys. Chem. Chem. Phys.* 19:20430–20440.