

Prevalence and mechanism of synergistic carboxylate-cation-water interactions in halophilic proteins

Hosein Geraili Daronkola¹ and Ana Vila Verde^{1,*}

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

ABSTRACT The cytoplasmic proteins of some halophilic organisms remain stable and functional at multimolar concentrations of KCl, i.e., under conditions that most mesophilic proteins cannot withstand. Their stability arises from their unusual amino acid composition. The most dramatic difference between halophilic and mesophilic proteins is that the former are rich in acidic amino acids. It has been proposed that one of the evolutionary driving forces for this difference is the occurrence of synergistic interactions between multiple acidic amino acids at the surface of the protein, the potassium cations in solution, and water. We investigate this possibility with molecular dynamics simulations, using high-quality force fields for the protein-water, protein-ion, and ion-ion interactions. We create a rigorous thermodynamic definition of interactions between acidic amino acids on proteins that can be used to distinguish between synergistic, noninteracting and interfering interactions. Our results demonstrate that synergistic interactions between neighboring acidic amino acids in halophilic proteins are frequent at multimolar KCl concentration. Synergistic interactions have an electrostatic origin, and are associated with stronger water-to-carboxylate hydrogen bonds than for acidic amino acids without synergistic interactions. Synergistic interactions are not observed in minimal systems of carboxylates, indicating that the protein environment is critical for their emergence. Our results demonstrate that synergistic interactions are neither associated with rigid amino acid orientations nor with highly structured and slow moving water networks, as had been originally proposed. Moreover, synergistic interactions can also be found in unfolded protein conformations. However, because these conformations are only a small subset of the unfolded state ensemble, synergistic interactions should contribute to the net stabilization of the folded state.

SIGNIFICANCE X-ray crystallography and NMR studies have suggested that acidic amino acids in folded halophilic proteins interact synergistically with water and cations at molar KCl concentrations, stabilizing the folded protein. We provide, for the first time, an operational definition of synergistic interactions between acidic amino acids that enables their quantification in molecular simulations. This definition can easily be adapted to investigate synergistic interactions between other amino acids. The results confirm that synergistic interactions exist at high KCl concentration and should stabilize the folded protein. They have an electrostatic origin and are associated with strong water-to-carboxylate hydrogen bonds. The results suggest nevertheless that other mechanisms may contribute to the evolutionary driving force behind the abundance of acidic amino acids in halophilic proteins.

INTRODUCTION

A few environments on Earth have physical-chemical properties far from the typical earthly environment; nevertheless, life thrives in these extreme environments as well.

Submitted February 23, 2023, and accepted for publication May 9, 2023.

*Correspondence: ana.araujo-vila-verde@uni-due.de

Hosein Geraili Daronkola and Ana Vila Verde's present address is University of Duisburg-Essen, Faculty of Physics, Duisburg, Germany.

Editor: Lucie Delemotte.

<https://doi.org/10.1016/j.bpj.2023.05.011>

© 2023 Biophysical Society.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



One striking example of such places is the Dead Sea, with its multi molar concentration of bromide and chloride salts, i.e., much higher than the typical salt concentration ($\approx 0.7 \text{ mol}\cdot\text{dm}^{-3}$) in oceans. Microorganisms living at high salt concentrations are a specific type of extremophile; they are called halophiles (from the Greek word for “salt-loving”) (1). Halophilic microorganisms are mainly archaea (2) and bacteria (3), but halophilic fungi, algae, protozoa, and multicellular eukaryotes also exist (4,5). To prevent rupture of their membrane due to the high concentration of salt in the environment, and therefore a high osmotic

pressure, halophiles accumulate osmolytes inside their cytoplasm up to the same concentration as the external environment (1). For some halophiles, the main osmolyte is KCl; these halophiles boast cytoplasmic KCl concentrations in the multimolar range, far higher than the $0.15 \text{ mol} \cdot \text{dm}^{-3}$ that is typical of nonhalophilic organisms (termed mesophiles). Halophilic proteins (of halophiles where KCl is the main osmolyte) thus remain soluble, structurally stable, and functional (1,4,6–8) under conditions that proteins of mesophiles (i.e., mesophilic proteins) cannot typically tolerate (1,9–12).

Halophilic and mesophilic proteins differ in their amino acid composition (7). Revealing and quantifying how amino acid composition impacts protein-solvent, protein-ion, and protein-protein interactions so proteins remain stable and functional at high KCl concentration is important from a fundamental perspective, because halophilic environments are ancient and are candidates for proteogenesis (the origin of polypeptides) and abiogenesis (the origin of living systems) (13–15). The amino acids most abundant in halophilic proteins are part of the prebiotic set of amino acids; polypeptides formed from this set have the ability to fold when under halophilic conditions (13–15). Clarifying the physical-chemical mechanisms that result in structurally stable and functional proteins under halophilic conditions is thus critical for a broader understanding of evolution and of protein folding. It is also important from a biotechnological perspective, to develop methods to rationally determine the optimal and minimal modifications that must be introduced in mesophilic enzymes (e.g., for production of chiral drug molecules (16), for production of biofuels (17)) so they remain functional at high NaCl concentration. Enzymes with this ability are advantageous in a world where fresh water resources are increasingly scarce.

Mesophilic and halophilic proteins differ substantially in composition and structure. The most noticeable difference between them is that halophilic proteins are highly depleted of positively charged residues, and are highly enriched in negatively charged amino acids, especially aspartic acid, located on the surface of the protein (4,6,7,18,19). Halophilic proteins are thus on average substantially negatively charged. The physical mechanisms making up the evolutionary driving force for this characteristic are not yet well-understood. Multiple explanations have been proposed, as we described elsewhere (20); here we provide only a short summary. At high salt concentration, electrostatic interactions are strongly screened. A seemingly obvious explanation for the excess acidic amino acids in halophilic proteins is that a high net protein charge is necessary for sufficient electrostatic repulsion to prevent protein aggregation, i.e., to increase protein solubility (4,21). However, the fact that enzyme activity, structural stability, and association of subunits strongly depend on salt concentration beyond $0.5 \text{ mol} \cdot \text{dm}^{-3}$, even though charge screening is largely complete by this concentration, indicates that,

while maintenance of solubility may well be one of their roles, acidic amino acids should have other roles as well (1). The *solvent-only stabilization* model claims that acidic amino acids, which have the highest water-binding ability of all amino acids (22), are necessary to keep the protein hydrated by competing with the ions in solution for available water (2,23). However, our previous simulation work (20) does not support that scenario. Our simulations indicate that the amount of hydration water for any given protein is barely affected by changing the KCl concentration from 0.15 to $2 \text{ mol} \cdot \text{dm}^{-3}$ (20).

In this work, we investigate another proposed explanation for the abundance of acidic amino acids in halophilic proteins: the *ion-solvent stabilization model*. This hypothesis claims that hydrated ion networks form around the surface of the folded protein, enabled by specific arrangements of carboxyl groups that attract hydrated potassium ions and which thus induce synergistic interactions between the acidic amino acids (4,24–26). This hypothesis has been described in qualitative terms only. Because analogous arrangements of acidic amino acids cannot form (or at least are not dominant) in the ensemble of unfolded protein structures, the folded structure of halophilic proteins is stabilized relative to the unfolded state (4,24–26). This hypothesis is supported by the presence of highly ordered protein hydration shells in some of the few halophilic proteins that have been crystallized (27). Synergistic hydrated ion networks would explain why some halophilic proteins have higher hydration levels than predicted based on their content in acidic amino acids alone, and why halophilic proteins bind larger amounts of salt than mesophilic proteins (4,24–26,28). Such an ordered solvation layer around the protein is proposed to function as a barrier that prevents aggregation of halophilic proteins at high salt concentration (4,24–26). This hypothesis was further supported by measurements of water translational dynamics in the cytoplasm of halophilic and nonhalophilic bacteria using quasielastic neutron spectroscopy, which were interpreted as indicating that water near halophilic proteins had extremely slow translational dynamics (29). However, this interpretation of the quasielastic neutron spectroscopy measurements was subsequently challenged by another experimental study (30), and by our previous simulation work, which demonstrated that the translational dynamics of water of hydration of mesophilic and halophilic proteins are quite similar (20).

We use simulations to determine whether synergistic water + potassium + carboxylate networks indeed arise in halophilic proteins and contribute toward the stability of the folded protein structure relative to the unfolded state. We use force fields for the carboxylate-potassium, carboxylate-amine, and carboxylate-water interactions previously optimized by us to correctly capture these interactions, a critical aspect of this work (20,31). With the optimized parameters, the models reproduce the KCl activity derivative of aqueous potassium acetate solutions up to $b_{\text{KCH}_3\text{COO}} = 2$

$\text{mol} \cdot \text{kg}^{-1}$, the carboxylate-potassium distances visible in the crystal structure of a protein, the difference between the experimental hydration free energies of the carboxylate and the chloride ions, and the osmotic pressure of concentrated glycine solutions. Although multiple references to synergistic or cooperative effects involving acidic amino acids in halophilic proteins can be found in the literature, a quantifiable definition has not been offered. We thus start by defining synergistic and nonsynergistic (i.e., interfering) interactions between acidic amino acids in a precise and quantifiable manner. The minimum system in which synergistic interactions between acidic amino acids and the ions in solution can exist is that involving two neighboring amino acids. We investigate whether synergistic interactions exist between pairs of neighboring amino acids on halophilic proteins, and also in model systems composed of dimers or trimers of mimics of amino acid side chains. Our results make clear that synergistic interactions between neighboring acidic amino acids are indeed possible in proteins at high KCl concentration, and give insight into the physical mechanisms behind them.

METHODS

All simulations used the TIP3P water model (32), and modified versions of the AMBER ff14SB (33,34) force field for proteins and of the potassium and chloride parameters of Joung and Cheatham (35) for TIP3P water. The original force fields significantly overestimate the interaction between carboxylates and amines and between carboxylates and the potassium ion. To overcome these shortcomings, we modified the Lennard-Jones (LJ) interactions between carboxylate and water, carboxylate and amines, and carboxylate and potassium ions as described in our previous work (20,31).

Free energy calculations

Free energies associated with the mutation of acidic residues—Asp (D) or Glu (E)—to their neutral counterparts—Asn (N) or Gln (Q), respectively—were calculated using thermodynamic integration as implemented in the pmemd GPU engine of the AMBER 18 molecular dynamics package (36), using a dual-topology approach and soft core potentials. Calculations were performed for halophilic proteins with Protein Data Bank identifying codes (PDB) PDB: 1DOI (2Fe-2S ferredoxin from *H. marismortui* (23)), PDB: 2KAC (protein L mutant Kx6E (37)), and PDB: 2ITH (dihydrofolate reductase from *H. volcanii* (38)). Folded proteins PDB: 1DOI and PDB: 2KAC were simulated at two different salt concentrations: $b_{\text{KCl}} = 0.15 \text{ mol} \cdot \text{kg}^{-1}$, corresponding to mesophilic conditions (the control situation), and $b_{\text{KCl}} = 2 \text{ mol} \cdot \text{kg}^{-1}$, corresponding to halophilic conditions. The folded dihydrofolate reductase and the unfolded protein L were simulated at the higher salt concentration to increase the number and diversity of systems under halophilic conditions, for which synergistic effects have been proposed to occur. Given the very high computational cost of these calculations, these two systems were not simulated at low salt concentration. The starting structure of each folded protein for the free energy calculations was the last saved configuration of 1 μs production runs at the desired salt concentration, reported in our previous publication (20). The starting structure for the simulations with the unfolded protein L was obtained as described in supporting material, section 6. The simulation boxes (Fig. 1) are cubic, with edge length $\approx 100 \text{ \AA}$.

To choose appropriate pairs of acidic amino acids for the mutation study, we first selected residues with a minimum solvent-accessible surface area of

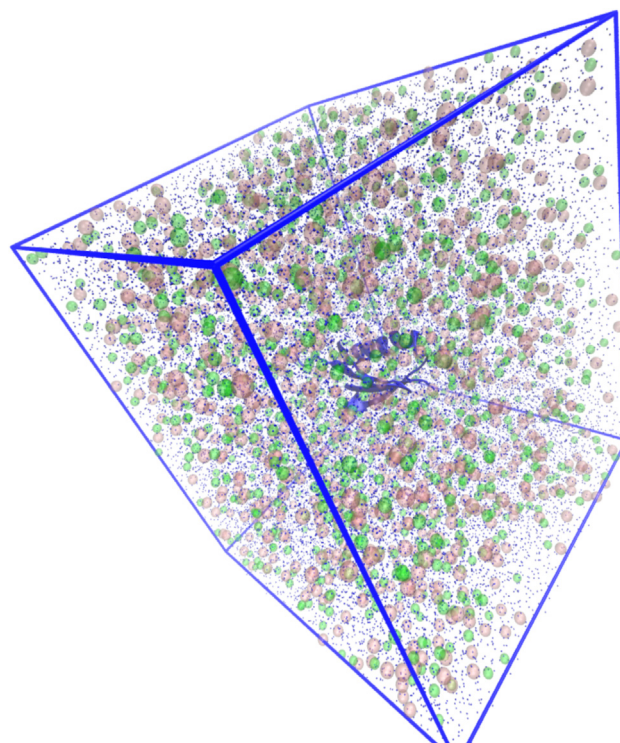
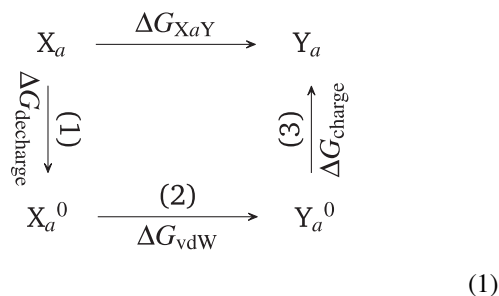


FIGURE 1 A snapshot of the simulation box of protein L in an aqueous solution of KCl at $b_{\text{KCl}} = 2 \text{ mol} \cdot \text{kg}^{-1}$. Protein L, dark blue; K^+ , transparent pink spheres; Cl^- , transparent green spheres; oxygen atoms of water molecules, blue dots. To see this figure in color, go online.

50 \AA^2 using the package *chimera* (39), thus ensuring they were solvent exposed. Of this set, we selected amino acid pairs for which the distance between their carboxylate carbons was below 7 \AA at $t = 0$.

Simulations were performed in three steps, described in Eq. 1 for the mutation of a generic amino acid X into amino acid Y. In the equation, a denotes the position of the residue in the amino acid sequence. Y_a^0 and X_a^0 indicate nonphysical forms of the amino acid where all atomic charges are set to zero. The free energy of mutation, ΔG_{XaY} , is equal to the sum of the three partial terms shown in Eq. 1. This free energy reflects only the different nonbonded interactions of the initial- and final-state amino acids with the solvent (water and ions) and the surrounding protein.



Complete details of the free energy simulations are given in supporting material, section 2.1; here we summarize only the most relevant aspects. The protein backbone atoms were restrained to their initial position using a harmonic restraint with a force constant of 35 $\text{kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$. Doing so is indispensable to obtain reproducible values of ΔG_{XaY} for each mutation. Our tests, discussed in supporting material, section 3, indicate that, without backbone restraints, small conformational changes away from the mutation site affect the value of ΔG_{XaY} by a few $\text{kcal} \cdot \text{mol}^{-1}$. Changes of

this magnitude are similar to those that may arise from the synergistic effect, and would thus make results impossible to interpret. The introduction of backbone restraints eliminates this difficulty. We emphasize that the energy of the restraints does not contribute to $\partial V/\partial\lambda$, and therefore does not contribute to the mutation free energy.

The Python tool *alchemical-analysis* (40) was used to calculate the free energy difference for each of the steps in Eq. 1. This tool calculates multiple free energy estimates, using different methods, from the simulation data. In all cases, the first 1 ns of the production simulations was ignored and considered as equilibration time. The values reported here were estimated integrating $\partial V/\partial\lambda$ using a natural cubic spline (“TI-3”). A detailed analysis of computational accuracy and choice of the best estimator for our system is given in [supporting material](#), section 2.1.1.

Potential of mean force calculations

To generate minimal models of the side chains of aspartate and of asparagine, we replaced the amino ($-\text{NH}_3^+$) and carboxylate ($-\text{CO}_2^-$) functional groups attached to the C_α of the zwitterionic form of the amino acids by two hydrogen atoms, to form a $-\text{C}_\alpha\text{H}_3$ group. The three hydrogen atoms attached to the C_α were given identical charges so that the net charge of the side chain model is zero (for asparagine) or -1 (for aspartate). Finally, the van der Waals (vdW), bond, angle, and dihedral parameters for interactions involving the newly added hydrogen atoms and the rest of the side chain atoms were given the same values as those of the standard hydrogen bonded to C_α in the respective amino acid. The side chain models are shown in [Fig. 2](#).

All umbrella sampling simulations from which the potential of mean force (PMF) curves were obtained were performed using the GROMACS 2020 simulation package (41,42). Complete simulation details are given in [supporting material](#), section 2.1; here we summarize only the most relevant simulation choices. The two-body PMFs were calculated as a function of the distance, ξ , between the C_α of the side chains. For all cases, the side chains were kept approximately coplanar by restraining the dihedral ϕ defined by atoms $\text{C}_{\beta 1}-\text{C}_{\alpha 1}-\text{C}_{\alpha 2}-\text{C}_{\beta 2}$ of the two amino acids (denoted by the subscripts 1 and 2) to $\phi = 0^\circ$. Calculations were performed for different relative orientations of the side chains, enforced by restraining the angles θ_1 and θ_2 to different values; θ_1 is defined by atoms $\text{C}_{\beta 1}-\text{C}_{\alpha 1}-\text{C}_{\alpha 2}$; θ_2 is defined by atoms $\text{C}_{\alpha 1}-\text{C}_{\alpha 2}-\text{C}_{\beta 2}$. The collective variables used in the two-body PMF calculations are illustrated in [Fig. 2 A](#). The two-body PMF was calculated for three different systems—D-D, N-N, and D-N—in explicit water and KCl at $b_{\text{KCl}} = 0.15$ and $2 \text{ mol} \cdot \text{kg}^{-1}$.

The three-body PMF was calculated as a function of the distances, ξ_1 and ξ_2 , of one amino acid to the other two, as illustrated in [Fig. 2 B](#). The C_α s of two of the amino acids are restrained to fixed positions in space so they are

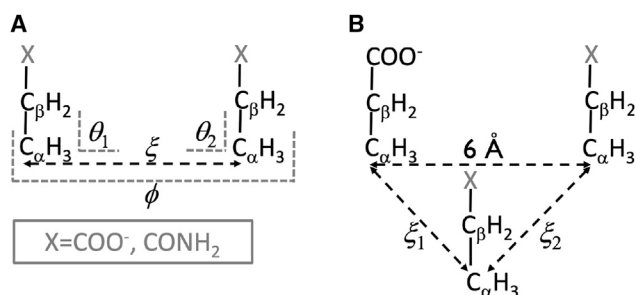


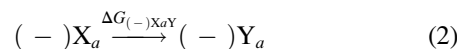
FIGURE 2 Systems and collective variables used in the potential of mean force calculations; $X = \text{COO}^-$ for the model aspartate side chain and $X = \text{CONH}_2$ for the model asparagine side chain. (A) Two-body systems, showing the reaction coordinate ξ and the restrained angles (θ_1 and θ_2) and dihedral (ϕ). (B) Three-body systems, showing the reaction coordinates ξ_1 and ξ_2 ; angle and dihedral restraints (not shown) analogous to those applied in the two-body system were used to keep the side chains parallel to each other and perpendicular to the plane defined by the three C_α s.

on average 6 \AA apart. The position of the C_α of the third amino acid is governed by the restraints applied to ξ_1 and ξ_2 . Six angular and three dihedral restraints (not shown in [Fig. 2 B](#)) analogous to those employed in the two-body system were used to keep the side chains approximately perpendicular to the plane defined by the three C_α s, by setting the angle equilibrium values to 90° and the dihedral ones to 0° . PMFs were calculated for three different systems—D-D-D, D-D-N, and D-N-N—in explicit water and KCl at $b_{\text{KCl}} = 2 \text{ mol} \cdot \text{kg}^{-1}$. The WHAM (43) analysis program implemented in GROMACS was used to compute the PMF curves from the output of the umbrella simulations.

RESULTS

Defining and quantifying synergistic interactions between acidic amino acids

As summarized above, the ion-solvent stabilization model claims that synergistic interactions between spatially close acidic amino acids, mediated by hydrated potassium cations, stabilize folded halophilic proteins at high KCl concentrations (24,26,44). To investigate this effect in simulations, a precise and quantifiable definition of synergistic interactions between acidic amino acids is necessary. The smallest system in which synergistic interactions between acidic amino acids can exist consists of two acidic residues. With this in mind, we designed a simulation protocol to assess the existence and magnitude of these synergistic interactions by a difference of mutation free energies. A pair of solvent-exposed acidic amino acids in a protein, in positions a and b of the amino acid sequence, is selected. We first calculate the free energy of mutating of one of them to the neutral natural amino acid that most closely resembles it: aspartate (D) is mutated to asparagine (N) and glutamate (E) to glutamine (Q). This mutation is schematically represented as:



Equation 2 indicates that amino acid X in position a is mutated to amino acid Y; the minus sign in parentheses $(-)$ emphasizes that this mutation is done in the presence of a particular negative neighbor in position b . The free energy $\Delta G_{(-)XaY}$ contains only the contribution of the interactions between the environment (i.e., the water, ions in solution, and the rest of the protein) and the mutated residues to the mutation free energy, as detailed in the [methods](#). The structure and size of the neutral amino acid Y (Q or N) are very similar to that of X (E or D) as can be seen in [Fig. S1](#), so the initial and final states are expected to have similar vdW interactions with their environment, i.e., these differences will contribute minimally to $\Delta G_{(-)XaY}$. Our results below confirm that this assumption holds. $\Delta G_{(-)XaY}$ is thus dominated by electrostatic interactions with the solvent (i.e., water + ions). These interactions are much more favorable for charged amino acids than for uncharged ones (45), so this free energy is positive (i.e., the mutation is unfavorable). If solvent-mediated synergistic electrostatic

interactions exist between the two acidic amino acids, they will be present in the initial state but not the final state of the system, and will thus affect the value of $\Delta G_{(-)XaY}$.

In a second step, we calculate the free energy associated with the same mutation, but now for a modified version of the protein where D or E in position b was replaced by N or Q, as appropriate, before the calculation is performed. This second mutation is schematically represented as



where (0) emphasizes that the neighboring amino acid is electrically neutral.

The difference

$$\Delta\Delta G = \Delta G_{(0)XaY} - \Delta G_{(-)XaY} \quad (4)$$

gives direct insight into the presence or absence of synergistic interactions between amino acids, and quantifies the magnitude of this effect. If amino acids a and b interact very weakly (e.g., if they are spatially distant), $\Delta\Delta G = 0$. If they are nearby, conventional understanding would suggest that two charged amino acids repel each other. Moreover, this repulsion should be more intense than any interactions between two neutral amino acids or between one neutral and one charged amino acid (in the end state). To first approximation, the electrostatic interaction between two neutral amino acids is a dipole-dipole interaction, which decays as $1/r^3$, where r is the distance between the two dipoles; likewise, the electrostatic interaction between a neutral and a charged amino acid is a charge-dipole interaction, which decays as $1/r^2$. In contrast, charge-charge interactions decay as $1/r$. In this conventional picture, the $(-) XaY$ mutation should be less unfavorable than the $(0) XaY$ one, i.e., $\Delta G_{(-)XaY} < \Delta G_{(0)XaY}$, so $\Delta\Delta G > 0$; we call these interactions repulsive or interfering.

A negative value of $\Delta\Delta G$ indicates a synergistic interaction: it is harder to mutate an acidic amino acid to a neutral one in the presence of a nearby acidic residue than when the neighbor is not acidic. Our results below will show that this synergistic interaction indeed stems from the interactions between the two acidic amino acids.

Practically, to account for the uncertainty in the calculated values of $\Delta\Delta G$, we consider that negligible interactions exist if $|\Delta\Delta G| < 0.5 \text{ kcal}\cdot\text{mol}^{-1}$. Significant interfering interactions occur if $\Delta\Delta G > 0.5 \text{ kcal}\cdot\text{mol}^{-1}$ and significant synergistic interactions exist if $\Delta\Delta G < -0.5 \text{ kcal}\cdot\text{mol}^{-1}$.

We note that other approaches to assess whether neighboring acidic amino acids have synergistic interactions are in principle possible. For example, mutating charged amino acids to their protonated versions should lead to results comparable with ours, because protonating the acid minimally perturbs its LJ interactions with its neighbors. The definition proposed here (Eqs. 2, 3, and 4) has the advantage of being more intuitive because it relies on protonation states realistically accessible at the same pH. Quantitative definitions

relying on mutations of charged amino acids to purely hydrophobic ones, however, will likely introduce larger changes also in the LJ interactions between them, making it harder to isolate the electrostatic effect that is thought to be at the source of synergistic interactions.

Free energies of D \rightarrow N and E \rightarrow Q mutations in folded proteins

We calculated $\Delta\Delta G$ for selected pairs of acidic residues close to each other (the distance between the a and b carboxylate carbons is $< 7 \text{ \AA}$ at $t = 0$) on the surface of three halophilic proteins: halophilic ferredoxin (PDB: 1DOI), halophilic protein L (PDB: 2KAC), and halophilic dihydrofolate reductase (PDB: 2ITH). These proteins were selected because they differ in net electric charge and size. Protein L has a net charge of $-15e$ but is substantially smaller (only 64 amino acids) than either the halophilic dihydrofolate reductase (162 amino acids) or the halophilic ferredoxin (128 amino acids). The halophilic dihydrofolate reductase is substantially less charged ($-15e$) than the halophilic ferredoxin ($-29e$). Their diverse characteristics thus enable us to draw general conclusions regarding the importance of synergistic interactions between acidic amino acids in halophilic proteins. Given the high computational cost of these calculations, we investigated a smaller set of amino acid pairs in two of the proteins at a low salt concentration ($b_{\text{KCl}} = 0.15 \text{ mol}\cdot\text{kg}^{-1}$) and a larger set of pairs in the three proteins at the highest salt concentration ($b_{\text{KCl}} = 2 \text{ mol}\cdot\text{kg}^{-1}$), for which the force fields yield reliable results.

Fig. 3 shows the calculated values of $\Delta\Delta G$ in (i). The position of the residues of each pair is shown in (ii). The amino acid pairs and the mutations performed are indicated using condensed notation: e.g., (D,N36) ... E41Q indicates that the glutamate in position 41 of the amino acid sequence is mutated to glutamine; the vicinal amino acid is in position 36 and is either an aspartate (when calculating $\Delta G_{(-)E41Q}$) or an asparagine (when calculating $\Delta G_{(0)E41Q}$). Fig. 3 A and B shows values of $\Delta\Delta G$ for the two proteins investigated at low KCl concentration in light blue. Most values of $\Delta\Delta G$ are positive, but vary substantially (between 0.5 and 2 $\text{kcal}\cdot\text{mol}^{-1}$) depending on the pair of amino acids being investigated. These results are qualitatively in line with the conventional expectation that, at low salt concentration, neighboring acidic amino acids should repel. Synergistic interactions between acidic amino acids at low KCl concentration are only observed for one of the sites ((D,N38) ... E41Q of protein L) and are very weak ($\Delta\Delta G = -0.5 \text{ kcal}\cdot\text{mol}^{-1}$).

At high KCl concentration, however, the picture that emerges is very different: significant synergistic interactions are observed in 5 out of the 18 pairs of amino acids investigated (Fig. 3 A–C), i.e., for those with $\Delta\Delta G < -0.5 \text{ kcal}\cdot\text{mol}^{-1}$, as described in the previous subsection. Moreover, synergistic interactions have similar

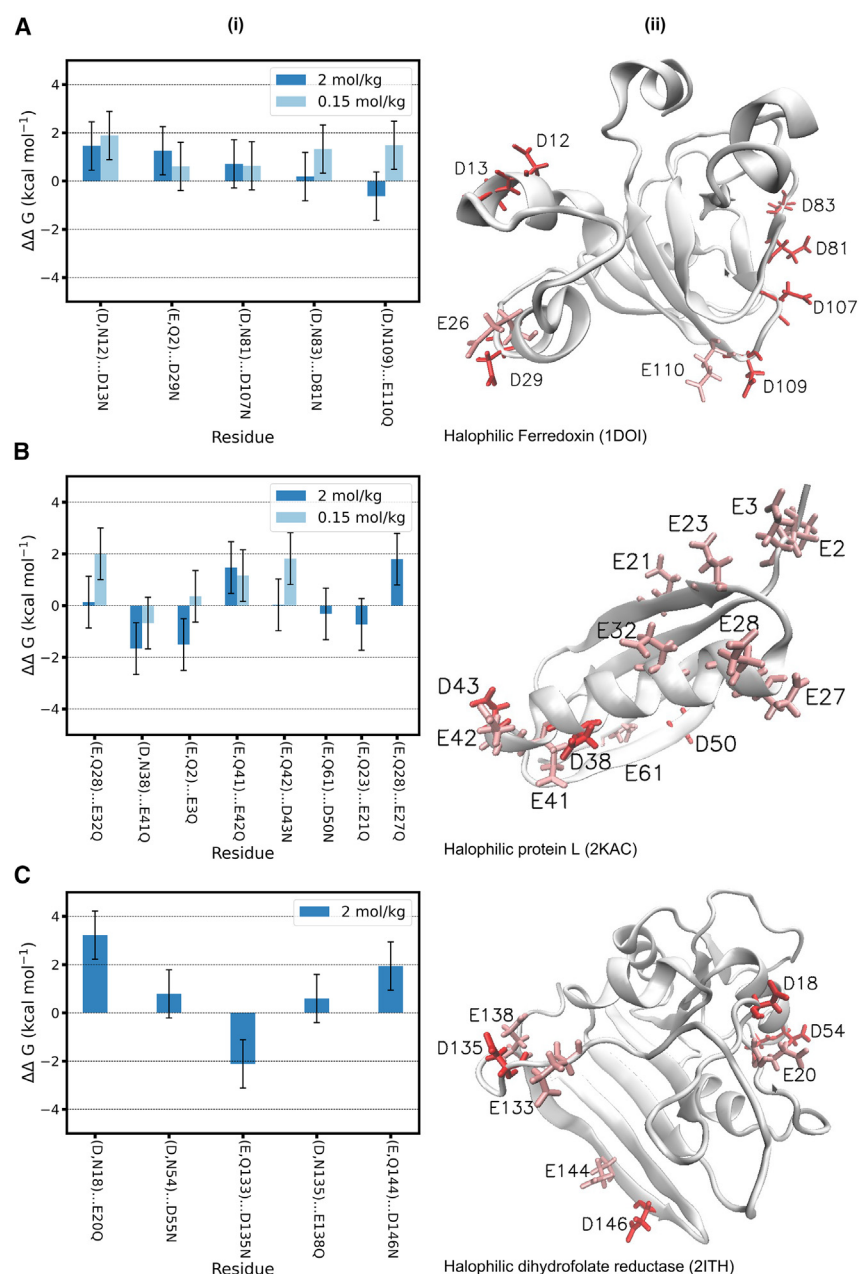


FIGURE 3 Interactions between pairs of acidic amino acids in folded halophilic proteins. (i) Change in free energy of mutation ($\Delta\Delta G \pm 1.0 \text{ kcal} \cdot \text{mol}^{-1}$; Eq. 4) for selected pairs of amino acids of the indicated proteins. The $\Delta\Delta G$ values are compiled in Tables S4–S6. The error bars are calculated using error propagation, using as input the standard error of the mean of each ΔG value estimated from five independent calculations for one of the protein sites as described in supporting material, section 3. (ii) Structure of each protein during the free energy calculation simulations; the acidic amino acids in (i) are displayed in red (D, aspartic acid) and in pink (E, glutamic acid). To see this figure in color, go online.

magnitude (-0.5 to $-2 \text{ kcal} \cdot \text{mol}^{-1}$) as the observed repulsive interactions (positive $\Delta\Delta G$ cases) in other amino acid pairs at the same concentration; in other words, the synergistic effect, when present, is of considerable magnitude. In half of the pairs investigated at high salt concentration, we observe interfering (i.e., repulsive) interactions. The values of positive $\Delta\Delta G$ are clearly lower than those at low KCl concentration for the same pairs, reflecting the expected higher electrostatic shielding brought by the more concentrated electrolyte solution. This result illustrates the limitations of mean field descriptions of electrostatic interactions for short distances: mean field calculations predict that the electrostatic interaction between acidic

amino acids at $b_{\text{KCl}} = 2 \text{ mol} \cdot \text{kg}^{-1}$ is fully shielded for distances larger than 2 \AA , i.e., $\Delta\Delta G$ should be very close to zero for the pairs we selected. In our calculations, full shielding is only observed for 3 of the 18 pairs tested.

The results in Fig. 3 demonstrate that a synergistic interaction between neighboring acidic amino acids is possible, and in fact occurs frequently at high salt concentration: in our data set, we observe it in $\approx 1/3$ of the amino acid pairs tested. The solvent-mediated (i.e., water- and ion-mediated) interaction between the amino acids in these pairs thus contributes to the stabilization of the folded protein structure. Moreover, for $1/6$ of the amino acid pairs tested at high salt concentration, the interaction between the neighboring

amino acids was reduced to zero: those pairs neither stabilize nor destabilize the folded protein structure. For half of the pairs investigated at high salt concentration, however, we observe positive values of $\Delta\Delta G$. Overall, our results suggest that the ion-solvent stabilization effect exists. The fact that a synergistic interaction is more frequent at high salt concentration confirms that it is mediated by the ions and water. Nevertheless, this effect contributes only to a limited extent to the stabilization of the folded structure of halophilic proteins, at least at $b_{\text{KCl}} = 2 \text{ mol} \cdot \text{kg}^{-1}$. It is possible that the ion-solvent stabilization effect becomes more intense and more likely at even higher salt concentrations, but at present this possibility cannot be investigated using molecular simulations because force fields that remain accurate up to the solubility limit of the salt do not exist.

Free energies of D → N and E → Q mutations in the unfolded halophilic protein L

We next tested whether synergistic interactions between acidic amino acids in unfolded proteins are possible: we calculated $\Delta\Delta G$ for multiple pairs of spatially close acidic amino acids belonging to an unfolded structure of halophilic protein L, and compared the frequency of occurrence and magnitude of synergistic effects relative with those observed for the same protein in the folded state. Protein L was selected because, of the three proteins investigated in the folded state, it had the most frequent synergistic interactions. The unfolded configuration was obtained from a replica exchange molecular dynamics (REMD) simulation, as described in [supporting material](#), section 6. The REMD simulation could not sample the full ensemble of unfolded configurations; rather, it sampled a subset of configurations where the protein is denatured but collapsed. For our purpose this behavior is not an impediment because synergistic interactions, if present in the unfolded state, should be more likely when pairs of acidic amino acids are sufficiently close. Theseus (46) was used to find the configuration with the structure most similar to the average structure in the REMD trajectory at 298 K. This configuration is shown in [Fig. 4](#). The backbone of the structure was restrained during the free energy calculations, similarly to the calculations done for the folded proteins.

[Fig. 4 A](#) shows the $\Delta\Delta G$ values for the denatured structure of protein L at $b_{\text{KCl}} = 2 \text{ mol} \cdot \text{kg}^{-1}$. Surprisingly, almost all of the pairs of amino acids show synergistic interactions, and the value of $\Delta\Delta G$ is more negative in some cases than the synergistic cases in the folded structure of the same protein. These results do not support the scenario that synergistic effects between pairs of amino acids are enabled by specific amino acid configurations exclusive to the folded protein structure (44). On the contrary, it seems that pairs of acidic residues assume relative distances and conformations leading to synergistic interactions very easily in denatured conformations of halophilic protein L.

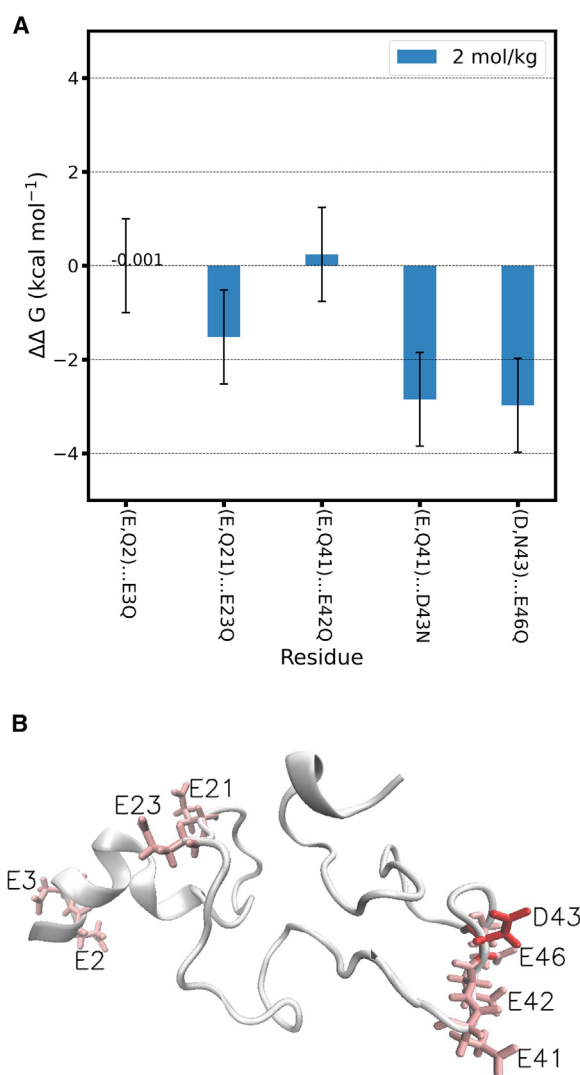


FIGURE 4 Interactions between pairs of acidic amino acids in the unfolded halophilic protein L at $b_{\text{KCl}} = 2 \text{ mol} \cdot \text{kg}^{-1}$. (A) Change in free energy of mutation ($\Delta\Delta G \pm 1.0 \text{ kcal} \cdot \text{mol}^{-1}$; Eq. 4) for selected pairs of amino acids. The $\Delta\Delta G$ values are compiled in [Table S7](#). The error bars are calculated using error propagation, using as input the standard error of the mean of each ΔG value estimated from five independent calculations for one of the protein sites as described in [supporting material](#), section 3. (B) Unfolded conformation of protein L used in the free energy calculations; the acidic amino acids in (A) are displayed in red (D, aspartic acid) and in pink (E, glutamic acid). To see this figure in color, go online.

Mechanism behind the synergistic effect

Negative values of $\Delta\Delta G$ result from synergistic electrostatic interactions between acidic amino acids

To further understand the mechanism leading to the synergistic effect, we examined the three contributions to $\Delta\Delta G$ along the mutation path. The free energy associated with a generic mutation XaY is calculated in three steps

$$\Delta G_{XaY} = \Delta G_{\text{decharge}} + \Delta G_{\text{vdW}} + \Delta G_{\text{charge}} \quad (5)$$

following the thermodynamic cycle shown in Eq. 1. The term $\Delta G_{\text{decharge}}$ is the free energy change associated with decharging amino acid X, i.e., setting all atomic charges of this residue to zero. The term ΔG_{vdW} corresponds to mutating the decharged residue X^0 into the decharged residue Y^0 ; this term includes only the contributions of the changes in the LJ potentials of the residue to the interactions with its environment and is often termed the vdW contribution. The final term, ΔG_{charge} , is the free energy change associated with reinstating the atomic charges of amino acid Y. The values of $\Delta\Delta G$ can thus be decomposed into decharging, vdW, and charging contributions as:

$$\Delta\Delta G_{\text{decharge}} = \Delta G_{(0)\text{decharge}} - \Delta G_{(-)\text{decharge}} \quad (6)$$

$$\Delta\Delta G_{\text{vdW}} = \Delta G_{(0)\text{vdW}} - \Delta G_{(-)\text{vdW}} \quad (7)$$

$$\Delta\Delta G_{\text{charge}} = \Delta G_{(0)\text{charge}} - \Delta G_{(-)\text{charge}} \quad (8)$$

These components are shown in Fig. 5 for the folded halophilic protein L at $b_{\text{KCl}} = 2 \text{ mol} \cdot \text{kg}^{-1}$. The $\Delta\Delta G_{\text{vdW}}$ component takes both positive and negative values, but its absolute value is always below $\approx 0.5 \text{ kcal} \cdot \text{mol}^{-1}$. The vdW component of the mutation free energy is thus essentially independent of the identity of the vicinal amino acid. The sign of

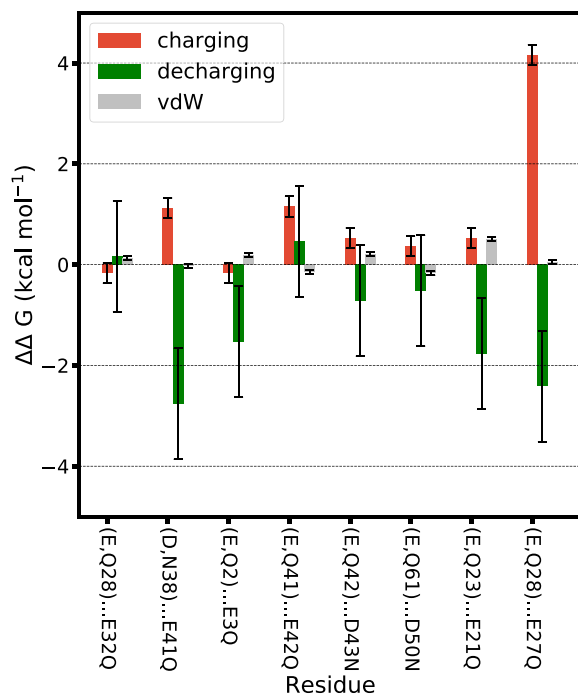


FIGURE 5 Components of $\Delta\Delta G$ for the indicated pairs of amino acids of the halophilic protein L at $b_{\text{KCl}} = 2 \text{ mol} \cdot \text{kg}^{-1}$. $\Delta\Delta G_{\text{charge}} \pm 0.2 \text{ kcal} \cdot \text{mol}^{-1}$, $\Delta\Delta G_{\text{decharge}} \pm 1.1 \text{ kcal} \cdot \text{mol}^{-1}$, and $\Delta\Delta G_{\text{vdW}} \pm 0.04 \text{ kcal} \cdot \text{mol}^{-1}$ (Eqs. 6, 7, and 8). The error bars are standard errors of the mean, estimated from 5 independent calculations for one of the protein sites as described in [supporting material](#), section 3. The sum of the components for each amino acid pair equals the values shown in Fig. 3 Bi). To see this figure in color, go online.

$\Delta\Delta G$ is thus determined by the balance between the charging and the decharging steps. The charging step, which introduces the atomic charges in the final (neutral) amino acid, predominantly has a positive contribution to $\Delta\Delta G$. In contrast, the decharging step, which removes the atomic charges in the initial (negative) amino acid, negatively contributes to $\Delta\Delta G$.

This result confirms that a negative $\Delta\Delta G$ is associated with the decharging step. A negative value of $\Delta\Delta G_{\text{decharge}}$ could arise either from 1) unexpected *stabilizing* electrostatic interactions between the two vicinal acidic amino acids, or between 2) *destabilizing* electrostatic interactions between vicinal acidic and neutral amino acids. As discussed in detail in [supporting material](#), section 5, the first possibility is the one at play here.

Synergistic interactions are not observed in minimal systems of acidic amino acids

Do synergistic interactions between acidic amino acids occur in minimal systems, i.e., outside of a protein environment? Does the synergistic effect occur for particular distances or relative orientations? To answer these questions, we characterized the thermodynamics of the interaction of minimal systems mimicking the side chains of aspartate and of asparagine by calculating the PMF as a function of their distance. In the minimal systems, the carbon corresponding to the C_α in the full amino acid now forms a CH_3 group; that carbon and the hydrogen atoms directly bonded to it have modified charges, so the nominal charge of the side chain is the same as that of the original amino acid. For conciseness, we refer to the amino acid models as side chains, and by the names (aspartate or asparagine) and one-letter code (D or N) of the corresponding amino acids.

Fig. 6 A shows the PMF as a function of $C_\alpha \cdots C_\alpha$ distance (ξ) for the side chain pairs D-D, N-N, and D-N, at $b_{\text{KCl}} = 0.15 \text{ mol} \cdot \text{kg}^{-1}$ and at $b_{\text{KCl}} = 2 \text{ mol} \cdot \text{kg}^{-1}$, for the case where the side chains are parallel to each other and perpendicular to ξ . The curves are shifted along the y axis for ease of viewing, so the absolute values of $\Delta G(\xi)$ have no physical meaning. Differences between the three curves at each salt concentration, or within each curve, are physically meaningful. The D-N and N-N interactions at large distances can be assumed to be zero; for this reason, the curves for the D-N and N-N pairs at each salt concentration were shifted along the y axis to coincide at $\xi = 15 \text{ \AA}$. The interaction energy between the D-D pair at $\xi = 15 \text{ \AA}$ at each salt concentration was estimated analytically as the Coulomb interaction between two point charges. Each D-D curve was shifted so that the difference $\Delta G_{\text{D-D}}(15 \text{ \AA}) - \Delta G_{\text{N-N}}(15 \text{ \AA})$ reproduces the analytical estimate for the respective salt concentration. At both concentrations, the maximum repulsion experienced by two negative side chains occurs at $\xi \approx 6 \text{ \AA}$, and is weak ($< 1 \text{ kcal} \cdot \text{mol}^{-1}$). The D-N and N-N pairs have negligible interaction energy down to $\xi \approx 5 \text{ \AA}$, as expected because

charge-dipole and dipole-dipole interactions are short range and are screened at the salt concentrations considered. Below $\xi \approx 4.5 \text{ \AA}$, the CH_3 and CH_2 groups of the side chains are very close, whereas the terminal atoms point away from each other, as the representative configuration shown in Fig. S10 illustrates. As a result of this preferential configuration, attractive LJ interactions dominate over electrostatics, giving rise to a minimum in the PMF in all three cases (calculations shown in supporting material, section 7). Results (not shown) for other relative orientations of the side chains have the same qualitative features; the main differences are observed in the position and height of the repulsive maximum of the D-D pair. PMF curves calculated based on the distance between the terminal (i.e., the γ) carbons of the side chain show similar characteristics, as illustrated in Fig. S11 for the D-D pair.

To assess whether a synergistic effect can arise from side-chain pairs, we computed the quantity $\Delta\Delta G(\xi)$ from the PMF curves:

$$\begin{aligned} \Delta\Delta G(\xi) &= [\Delta G(\xi)_{\text{N-N}} - \Delta G(\xi)_{\text{N-D}}] - [\Delta G(\xi)_{\text{D-N}} - \Delta G(\xi)_{\text{D-D}}] \\ &= \Delta G(\xi)_{\text{D-D}} + \Delta G(\xi)_{\text{N-N}} - 2 \times \Delta G(\xi)_{\text{D-N}} \end{aligned} \quad (9)$$

The subscripts indicate the pair of side chains for each PMF. The quantity $\Delta\Delta G(\xi)$ is analogous to the $\Delta\Delta G$ calculated for pairs of acidic amino acids on proteins (Figs. 3 and 4) but gives insight into cooperativity as a function of the distance between the side chains. Fig. 6 B shows $\Delta\Delta G(\xi)$ for both salt concentrations, for side chains at different relative orientations. The value and position of the maxima of each curve depend strongly on relative orientation. Nevertheless, this quantity never assumes negative values, indicating that a synergistic effect does not occur in this two-body system. Could it be that a synergistic effect arises in minimal systems with more side chains? To answer this question we performed an analogous PMF study for a system of three side chains. The results, presented in supporting material, section 8, indicate that a synergistic effect does not occur in that system either.

The absence of the synergistic effect in the two-body and three-body systems suggests that synergistic interactions between acidic amino acids do not result solely from water- and salt-mediated interactions between the acidic amino acids, but are also enabled by particular local protein environments.

The synergistic effect does not require particular amino acid orientations

The large variation observed in the values of $\Delta\Delta G$ ($-3 < \Delta\Delta G \text{ (kcal} \cdot \text{mol}^{-1}) < +3$) shown in Figs. 3 and 4 indicates that solvent-mediated interactions (synergistic or repulsive) between charged amino acids at high salt concentration are

strongly affected by local protein composition and local structure. Obvious correlations between the sign or magnitude of $\Delta\Delta G$ and the secondary structural motifs in which the amino acids of the pair are located or the identity of other neighboring amino acids are not present. Moreover, in our data set the frequency of synergistic interactions differs strongly between proteins: they occur more frequently in protein L than for the other two proteins. Because the data set is small, it is at present unclear whether these differences between the proteins are significant. Clarifying these aspects will require their systematic study through a substantially larger data set of mutation free energy calculations, which is beyond what is currently possible.

We have observed a substantial synergistic effect in a denatured configuration of the halophilic protein L. This observation contradicts the claim of the lack of necessary preorientation for cation-acidic residue interaction in the unfolded structure. To gain further insight into the connection between the relative orientation of amino acids and

their synergistic or repulsive interactions, we simulated the folded halophilic protein L, at $b_{\text{KCl}} = 2 \text{ mol} \cdot \text{kg}^{-1}$, to obtain a long, continuous trajectory (see supporting material, section 2.2 for simulation details). We analyzed this trajectory to determine whether characteristic configurations of acidic amino acids differ between pairs of amino acids showing synergistic versus interfering effects. Fig. 7 compares the histograms of distances between the carboxylate carbons of the amino acid pairs for which $\Delta\Delta G$ was calculated (Fig. 3 Bi). The distribution of distances differs substantially between pairs, and a correlation between the characteristics of the distribution and the magnitude or sign of $\Delta\Delta G$ cannot be discerned. The distribution of distances for pairs of amino acids showing synergistic interactions (in blue) can be narrow and associated with short distances (5–6 \AA) corresponding to contact ion pair configurations, or can be quite broad and include distances that correspond to solvent-shared ion pair configurations. The distributions of distances for pairs of amino acids showing interfering interactions (in red) are equally broad and span the same distances. These results do not support the claim that the synergistic effect, when it exists, arises from amino acids with particular and well-defined relative orientations.

Enthalpic/entropic origins of the synergistic effect

To assess whether the synergistic effect observed for some of the protein sites could originate from stronger hydrogen bonds between water and the carboxylate groups of the

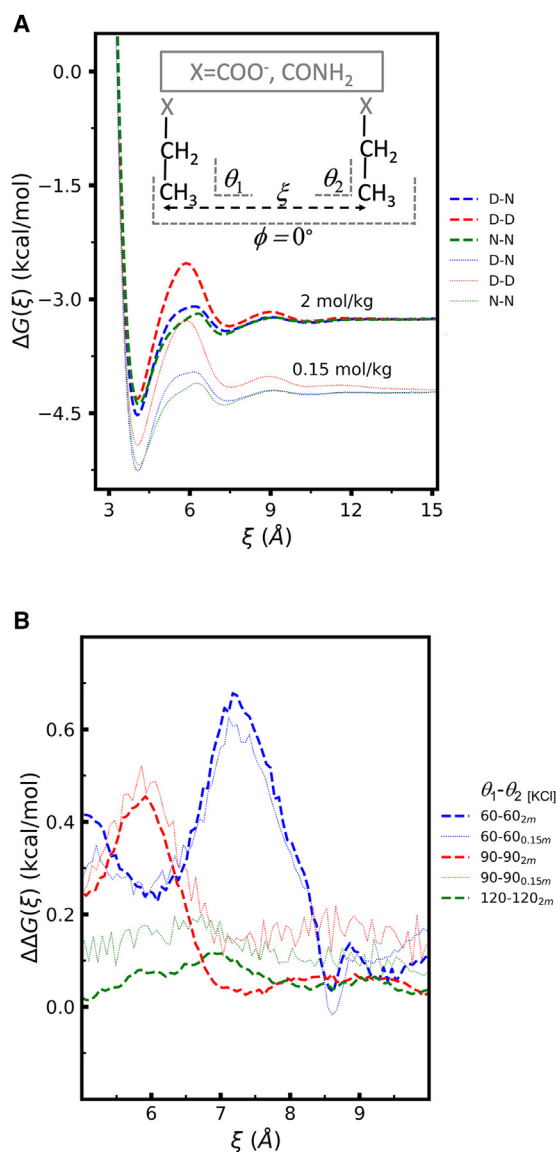


FIGURE 6 Potential of mean force as a function of C α ...C α distance, for different orientations of pairs of side chain mimics, at $b_{\text{KCl}} = 0.15 \text{ mol} \cdot \text{kg}^{-1}$ (0.15m) and $b_{\text{KCl}} = 2 \text{ mol} \cdot \text{kg}^{-1}$ (2m). (A) PMF for the indicated pairs of amino acid side chains, for parallel side chains (restraint angles $\theta_1 = \theta_2 = 90^\circ$; $\phi = 0^\circ$). The ideal gas entropic term ($-2k_B T \ln(\xi)$) is not included in the curves. (B) $\Delta\Delta G(\xi)$ (Eq. 9) for different relative orientations (θ_1 and θ_2 ; $\phi = 0^\circ$) of the side chains, at $b_{\text{KCl}} = 0.15 \text{ mol} \cdot \text{kg}^{-1}$ (0.15m) or $b_{\text{KCl}} = 2 \text{ mol} \cdot \text{kg}^{-1}$ (2m). To see this figure in color, go online.

acidic amino acids in those cases, we quantified the geometry of those hydrogen bonds separately for the synergistic and the interfering pairs of amino acids of protein L from the same simulation on which Fig. 7 is based. Fig. 8 A shows the normalized histogram ($p_{\text{syn}}(d, \cos \theta)$) of the distance (d) between the water and the carboxylate oxygens versus the cosine of the angle θ for water molecules near carboxylate groups of the amino acid pairs showing a synergistic effect. The hydrogen bond strength increases for shorter distances and for cosine values close to -1 , i.e., for θ near 180° . Water

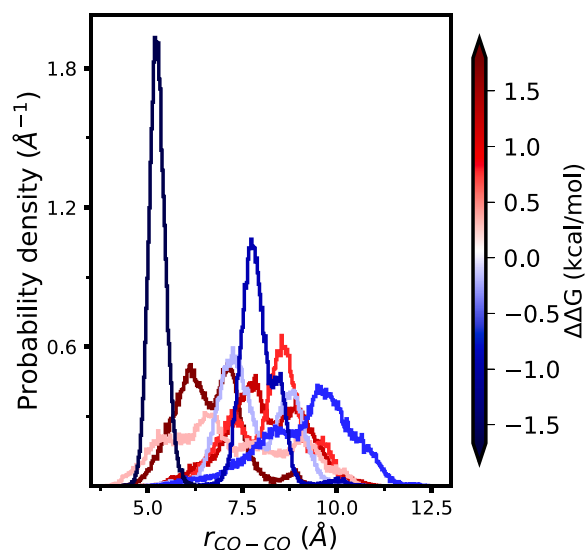


FIGURE 7 Histogram of the distances between carboxylate carbons of the folded protein L from an MD simulation at $b_{\text{KCl}} = 2 \text{ mol} \cdot \text{kg}^{-1}$, for the amino acid pairs for which $\Delta\Delta G$ was calculated (Figure 3 Bi). The color scale shows the value of $\Delta\Delta G$, with blue used for pairs showing a synergistic interaction and red used for pairs with an interfering interaction. To see this figure in color, go online.

preferentially donates strong hydrogen bonds to the carboxylate groups, as indicated by the short values of d and by cosine values near -1 . These results are consistent with previous work, which indicates that carboxylates accept strong hydrogen bonds from water (47). In Fig. 8 B we show the difference between the histograms obtained for the interfering and for the synergistic residues ($p_{\text{syn}}(d, \cos \theta) - p_{\text{int}}(d, \cos \theta)$). This difference shows that the synergistic sites accept stronger hydrogen bonds and accept fewer weaker hydrogen bonds than the interfering sites. This result confirms that the synergistic effect has an enthalpic contribution arising from solute-water interactions.

We found no mechanistic correlation between the radial distribution function of the cation and the presence or absence of synergistic interactions. Further assessing the contribution of the cation toward the difference in hydrogen bond strength between water and the carboxylate groups, and assessing whether entropy also contributes to the synergistic effect, will require the quantification of local entropy and enthalpy at the interfering and synergistic protein sites. At present, algorithms to quantify local enthalpic and entropic contributions to solvation free energies (48,49) can only be applied to solutes in pure water, so these aspects cannot be investigated in our system.

CONCLUSIONS

We used molecular simulations to investigate whether synergistic—rather than the expected interfering—interactions between neighboring acidic amino acids exist and contribute to the stabilization of folded halophilic proteins, as originally

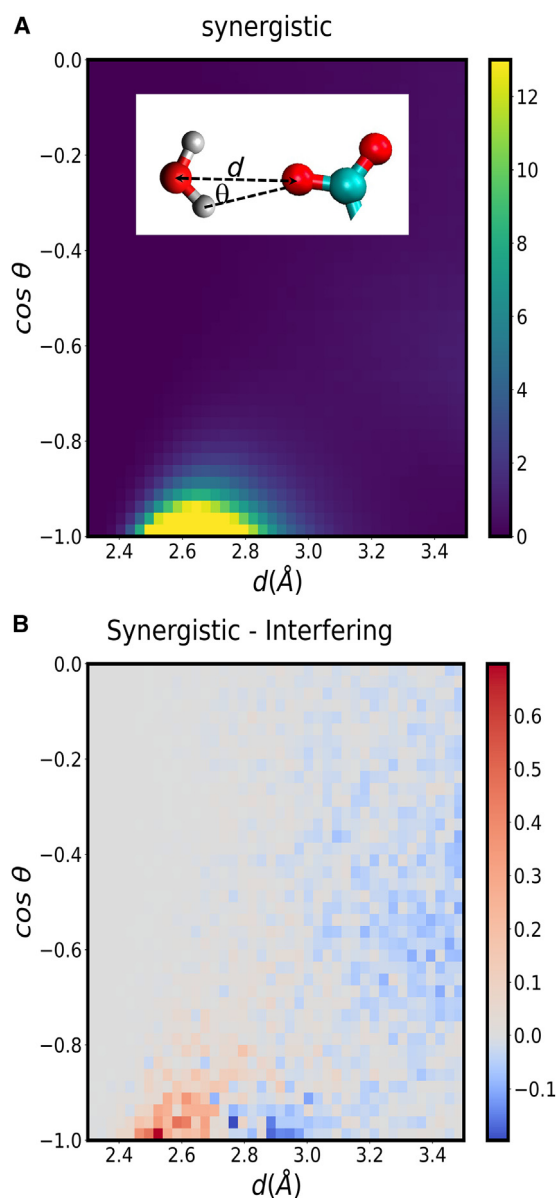


FIGURE 8 Geometric features of hydrogen bonds donated by water molecules to the carboxylate groups of acidic amino acids in protein L. (A) Histogram, $p_{\text{syn}}(d, \cos \theta)$, of the distance and angle characterizing hydrogen bonds accepted by the residues of synergistic pairs. The inset shows a water molecule near a carboxylate group, and illustrates the distance and angle used for the histograms. (B) Difference, $p_{\text{syn}}(d, \cos \theta) - p_{\text{int}}(d, \cos \theta)$, between the histograms characterizing hydrogen bonds accepted by residues of synergistic and interfering pairs. To see this figure in color, go online.

proposed by Zaccai et al. (24) and supported by NMR studies of halophilic proteins (44). Our results partially support this hypothesis, clarify its mechanism, and point to further studies necessary to fully understand the origin of synergistic interactions.

Quantitative investigations of this hypothesis have been hampered by the lack of an operational definition of synergistic interactions. Our first step was thus to address this difficulty: we quantify the interactions between neighboring

amino acids as a difference, $\Delta\Delta G$, of two mutation free energies, which can be calculated with standard molecular dynamics simulation packages.

Our $\Delta\Delta G$ calculations show that pairs of neighboring acidic amino acids at the surface of proteins in multimolar KCl solutions may indeed interact synergistically. Synergistic interactions are frequent at multimolar KCl concentration but very infrequent at low KCl concentration, suggesting that they have an electrostatic origin. Decomposing $\Delta\Delta G$ into the decharging, vdW, and charging components confirms that synergistic interactions are associated with the decharging step, i.e., with solvent-mediated electrostatic interactions between two acidic amino acids. Hydrogen bond analysis confirms that acidic amino acids with synergistic interactions accept stronger hydrogen bonds from water more often than those with interfering interactions. Further work is necessary to investigate whether synergistic interactions have also an entropic contribution. We find that synergistic interactions do not occur in minimal systems mimicking amino acid side chains, indicating that the protein environment influences the solvent- and ion-mediated interactions between acidic amino acids and contributes to the occurrence of synergistic interactions between them. However, obvious correlations between the secondary structural motif to which the amino acids belong or their spatially close neighboring amino acids and their synergistic/interfering interactions do not exist. Further work—e.g., investigating whether the protein environment creates electric fields of favorable strength and direction at synergistic but not interfering sites—is necessary to fully clarify the mechanism by which the protein environment allows synergistic interactions to occur in some sites.

Our results support the existence of synergistic interactions between acidic amino acids proposed by Zaccai et al. (24), but do not support their proposed mechanism that synergistic interactions are enabled by specific amino acid configurations exclusive to folded halophilic proteins. Instead, the simulations show that synergistic interactions are not associated with specific or rigid amino acid configurations, and exist also in unfolded conformations. According to Zaccai et al., specific amino acid configurations would lead to unusually slow dynamics of the water of hydration because the anchoring of the acidic residues in the protein would limit the number of available configurations and would slow down the hydrogen bond dynamics. Our previous simulations do not support that scenario: we found that the translational dynamics of water of hydration of halophilic proteins is very similar to that of mesophilic ones (20).

Synergistic interactions should nevertheless stabilize folded relative to unfolded configurations, as initially proposed, because the ensemble of unfolded configurations has many more configurations than the folded one. The configurations for which synergistic interactions occur (where the acidic amino acids are spatially close) are only a small

subset of the total unfolded ensemble, whereas they dominate the ensemble of folded configurations. Therefore, the stabilizing contribution of synergistic interaction to the unfolded ensemble should be smaller than for the folded one. We emphasize that it remains unclear whether synergistic interactions are the main mechanism by which acidic amino acids stabilize the folded structure of halophilic proteins at high salt concentration. It is also possible that nonsynergistic interactions involving acidic amino acids, solvent, and ions lead to different solvation of folded versus unfolded configurations, and that these differences stabilize the folded state (50). Future studies in our group will investigate this possibility.

SUPPORTING MATERIAL

Supporting material can be found online at <https://doi.org/10.1016/j.bpj.2023.05.011>.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENTS

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under grant no. 639429 and under Germany's Excellence Strategy – EXC 2033 – 390677874 – RESOLV.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

- Lanyi, J. K. 1974. Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriol. Rev.* 38:272–290.
- 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.
- van der Wielen, P. W. J. J., H. Bolhuis, ..., BioDeep Scientific Party. 2005. The enigma of prokaryotic life in deep hypersaline anoxic basins. *Science.* 307:121–123.
- Graziano, G., and A. Merlino. 2014. Molecular bases of protein halotolerance. *Biochim. Biophys. Acta.* 1844:850–858.
- Gunde-Cimermana, N., P. Zalarb, ..., A. Plemenitaš. 2000. Hypersaline waters in salterns - natural ecological niches for halophilic black yeasts. *FEMS Microbiol. Ecol.* 32:235–240.
- 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.
- 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.
- 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.
- Tessier, P. M., and A. M. Lenhoff. 2003. Measurements of protein self-association as a guide to crystallization. *Curr. Opin. Biotechnol.* 14:512–516.
- 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.
- 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.
- Reed, C. J., S. Bushnell, and C. Evilia. 2014. Circular dichroism and fluorescence spectroscopy of cysteinyl-tRNA synthetase from Halobacterium salinarum ssp. NRC-1 demonstrates that group I cations are particularly effective in providing structure and stability to this halophilic protein. *PLoS One.* 9, e89452.
- Rode, B. M. 1999. Peptides and the origin of life. *Peptides.* 20:773–786.
- Longo, L. M., J. Lee, and M. Blaber. 2013. Simplified protein design biased for prebiotic amino acids yields a foldable, halophilic protein. *Proc. Natl. Acad. Sci. USA.* 110:2135–2139.
- Longo, L. M., and M. Blaber. 2014. Prebiotic protein design supports a halophile origin of foldable proteins. *Front. Microbiol.* 4:418.
- Rossino, G., M. S. Robescu, ..., S. Collina. 2022. Biocatalysis: a smart and green tool for the preparation of chiral drugs. *Chirality.* 34:1403–1418.
- Bangaru, A., K. A. Sree, ..., C. N. Reddy. 2022. Role of Enzymes in Biofuel Production: Recent Developments and Challenges. Springer Nature Singapore, pp. 81–112.
- 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.
- Nath, A. 2016. Insights into the sequence parameters for halophilic adaptation. *Amino Acids.* 48:751–762.
- Geraili Daronkola, H., and A. Vila Verde. 2021. Proteins maintain hydration at high [KCl] concentration regardless of content in acidic amino acids. *Biophys. J.* 120:2746–2762.
- Elcock, A. H., and J. A. McCammon. 1998. Electrostatic contributions to the stability of halophilic proteins. *J. Mol. Biol.* 280:731–748.
- Kuntz, I. D. 1971. Hydration of macromolecules. III. Hydration of polypeptides. *J. Am. Chem. Soc.* 93:514–516.
- 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.
- Zaccai, G., F. Cendrin, ..., H. Eisenberg. 1989. Stabilization of halophilic malate dehydrogenase. *J. Mol. Biol.* 208:491–500.
- 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.
- Madern, D., C. Ebel, and G. Zaccai. 2000. Halophilic adaptation of enzymes. *Extremophiles.* 4:91–98.
- Irimia, A., C. Ebel, ..., F. M. D. Vellieux. 2003. The oligomeric states of Haloarcula marismortui malate dehydrogenase are modulated by solvent components as shown by crystallographic and biochemical studies. *J. Mol. Biol.* 326:859–873.
- 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.
- 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.

30. 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.
31. 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.
32. Jorgensen, W. L., J. Chandrasekhar, ..., M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
33. 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.
34. 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.
35. Joung, I. S., and T. E. Cheatham. 2008. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B.* 112:9020–9041.
36. Case, D., I. Ben-Shalom, ..., P. A. Kollman. 2018. Amber 18. University of California.
37. Tadeo, X., B. López-Méndez, ..., O. Millet. 2009. Structural basis for the aminoacid composition of proteins from halophilic archaea. *PLoS Biol.* 7, e1000257.
38. Binbuga, B., A. F. B. Boroujerdi, and J. K. Young. 2007. Structure in an extreme environment: NMR at high salt. *Protein Sci.* 16:1783–1787.
39. Pettersen, E. F., T. D. Goddard, ..., T. E. Ferrin. 2004. UCSF Chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25:1605–1612.
40. Klimovich, P. V., M. R. Shirts, and D. L. Mobley. 2015. Guidelines for the analysis of free energy calculations. *J. Comput. Aided Mol. Des.* 29:397–411, HHS Public Access.
41. Berendsen, H. J. C., D. van der Spoel, and R. van Drunen. 1995. GRO-MACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91:43–56.
42. Van Der Spoel, D., E. Lindahl, ..., H. J. C. Berendsen. 2005. GRO-MACS: fast, flexible, and free. *J. Comput. Chem.* 26:1701–1718.
43. Kumar, S., J. M. Rosenberg, ..., P. A. Kollman. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13:1011–1021.
44. Ortega, G., T. Diercks, and O. Millet. 2015. Halophilic protein adaptation results from synergistic residue-ion interactions in the folded and unfolded states. *Chem. Biol.* 22:1597–1607.
45. Deng, Y., and B. Roux. 2004. Hydration of amino acid side chains: nonpolar and electrostatic contributions calculated from staged molecular dynamics free energy simulations with explicit water molecules. *J. Phys. Chem. B.* 108:16567–16576.
46. Theobald, D. L., and D. S. Wuttke. 2008. Accurate structural correlations from maximum likelihood superpositions. *PLoS Comput. Biol.* 4:43.
47. Vinogradov, S. N., and R. H. Linnell. 1971. Hydrogen Bonding. Van Nostrand Reinhold.
48. Lazaridis, T. 1998. Inhomogeneous fluid approach to solvation thermodynamics. 1. Theory. *J. Phys. Chem. B.* 102:3531–3541.
49. Heyden, M. 2019. Disassembling solvation free energies into local contributions—toward a microscopic understanding of solvation processes. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 9:e1390.
50. 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.