

Tesis Doctoral

STRUCTURAL STUDIES OF THE HIPERSALINE ADAPTATION OF PROTEINS BELONGING TO HALOPHILIC ARCHAEA

Memoria presentada por

D . Gabriel Ortega Quintanilla

dirigida por

Dr. Óscar Millet Aguilar-Galindo
CIC bioGUNE

para optar al título de Doctor por la

eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

Facultad de Ciencia y Tecnología

Departamento de Biomedicina y
Biología Molecular

Programa de Doctorado en Biología
Molecular y Biomedicina

Leioa, Octubre de 2015

capítulo cero: agradecimientos

esta tesis ha sido posible gracias al apoyo económico de la Fundación “Obra Social” de la Caixa, del Ministerio de Economía y Competitividad y del tristemente extinto Ministerio de Ciencia e Innovación

quisiera dar las gracias a todos los que han contribuido en mayor o menor medida a la consecución de este trabajo. Me refiero principalmente a los compañeros del CIC bioGUNE, pero también al grupo del Dr. Bertrand García-Moreno de la JHU

mención aparte merecen los compañeros y amigos que forman y han formado parte del equipo de Resonancia Magnética Nuclear del CIC bioGUNE. Sin ellos nada hubiera sido igual. Especialmente quiero agradecer a Óscar su excelente desempeño como guía y maestro durante estos cinco años: no se me ocurre un Virgilio más adecuado

gracias también a ti, estimado lector, por halagarme con tu interés. Esta tesis es una parte importante de mi, pues puse en ella todo mi empeño. Espero que la encuentres interesante

por último, y por ello lo más importante, quiero dar las gracias a mi familia por enseñarme el valor del esfuerzo y la fuerza del cariño, y a Myriam por su apoyo pasado, presente y futuro

1. INTRODUCTION	1
1.1. Live at extreme environments: halophilic organisms	1
1.2. Adaptive strategies	3
1.2.1. Compatible solutes	3
1.2.2. Intracellular KCl accumulation	4
1.3. Molecular basis of haloadaptation	7
1.3.1. Halophilic signature: aminoacid composition	7
1.3.2. Salt dependence of protein thermodynamic stability	10
1.3.3. Salt modulation of enzymatic activity	11
1.3.4. Salt modulation of protein solubility	12
1.4. Models for haloadaptation	13
1.4.1. The halophilic model: molecular basis and limitations	13
1.4.2. The solvent exposed area and the Hofmeister effect	15
1.5. Biological interest and potential application of the study of halophilic proteins	17
2. OBJECTIVES AND WORKING HYPOTHESIS	19
3. MATERIALS AND METHODS	20
3.1. Proteins under study	20
3.1.1. Naturally mesophilic protein L <i>ProtL</i>	21
3.1.2. Halotolerant beta-lactamase from <i>Oceanobacillus iheyensis</i> β LAC	22
3.1.3. Unstructured N-terminal domain from human SRC kinase <i>uSRC</i>	23
3.1.4. Design of the halophilic unstructured homologue <i>haloSRC</i>	23
3.1.5. NAD ⁺ -dependent DNA Ligase from <i>Haloferax volcanii</i> <i>LigN</i>	25
3.1.5.1. <i>Adenylation and deadenylation assays</i>	26
3.1.5.2. <i>Activity assays: DNA Ligation</i>	27

3.2. Sample preparation	28
3.2.1. Culture growth and isotopic labelling	28
3.2.2. Protein purification	29
3.2.2.1. <i>ProtL</i>	30
3.2.2.2. β LAC	31
3.2.2.3. <i>Hv LigN</i>	32
3.2.2.4. <i>uSRC & haloSRC</i>	32
3.3. Thermodynamic stability measurements	33
3.3.1. Chemical denaturation experiments	34
3.3.1.1. <i>Linear extrapolation method: two–states fitting</i>	35
3.3.1.2. <i>Three–states fitting</i>	36
3.3.2. Thermal denaturation experiments	37
3.4. Nuclear Magnetic Resonance	40
3.4.1. Introduction	40
3.4.2. Backbone and side chain assignment experiments	41
3.4.3. pH titrations	45
3.4.4. Relaxation experiments	47
3.4.4.1. <i>Amide proton ^{15}N relaxation experiments</i>	50
3.4.4.2. <i>Relaxation models</i>	50
3.4.4.3. <i>^{17}O Magnetic Relaxation Dispersion experiments</i>	53
3.4.4.4. <i>^{17}O relaxation experiments</i>	53
3.4.5. Exchange experiments	55
3.4.5.1. H/D exchange experiments	55
3.4.5.2. H/H exchange: CLEANEX	57
3.4.5.3. Chemical exchange: EXSY	58
3.4.6. Chemical Shift Perturbation	59

4. MECHANISTIC INSIGHTS IN PROTEIN HALOADAPTATION	60
4.1. Setting the problem. Global stabilisation upon KCl addition	61
4.2. The role of charged groups	65
4.2.1. Electrostatic contribution of carboxylic groups	66
4.2.2. Electrostatic contribution of the unfolded state	70
4.2.3. Electrostatic contribution of Lysines	71
4.2.4. Global thermodynamic contribution of electrostatic interactions	72
4.3. Study of the hydration shell of halophilic proteins	74
4.3.1. Protein and internal water dynamics	75
4.3.2. Dynamics of the hydration layer of halophilic proteins	78
4.4. Role of the unfolded state in salt adaptation	81
4.4.1. The unfolded state becomes destabilized upon salt addition	83
4.4.2. Structural and dynamic analysis of the unfolded conformational ensemble	86
4.4.3. Analysis of the interactions between KCl and the unfolded state	91
4.4.4. Thermodynamic contribution of the unfolded state m^{unf}	94
4.5. Excluded volume contribution	99
5. A MODEL FOR PROTEIN HALOADAPTATION	101
5.1. Disentangling the energetic contributions to haloadaptation	102
5.1.1. Non specific interactions in the folded state	102
5.1.2. Preferential ion exclusion in the unfolded state	104
5.1.3. Towards a semi-quantitative model for protein haloadaptation	106
5.2. Contribution of the hydration of solvent exposed hydrophobic area	109
5.3. Why this aminoacid selection? Strict halophilicity	112
5.4. Halophilic adaptation and folding landscape	113
5.5. Comparison with other ions and the Hofmeister effect	117
5.6. Mechanistic comparison between haloadaptation and osmoprotection	121

6. ENZYME ACTIVATION INDUCED BY SALTS AS A SUBROGATE OF PROTEIN HALOADAPTATION	130
6.1. Complex salt modulation of activity	130
6.2. NAD ⁺ -dependent DNA Ligase from <i>Haloferax volcanii</i> LigN	131
6.3. Modulation of activity by KCl and NaCl	135
6.4. Salt dependence of stability and coupling with activity	136
6.4.1. Stabilizing cosolutes select conformations with lower solvent exposed hydrophobic area	138
6.4.2. Design of mutations reactivate the enzyme in the presence of NaCl	138
6.5. Effect of other Hofmeister ions	140
6.5.1. Stabilizing ions	140
6.5.2. Destabilizing ions	141
7. CONCLUDING REMARKS	143
8. SUMMARY	144
PUBLICATIONS	161
REFERENCES	162
Appendix	
Abbreviations	
List of Figures, Tables & Equations	

1. INTRODUCTION

*"La ciencia se construye de aproximaciones
que gradualmente se acercan a la verdad"*

Isaac Asimov

1. INTRODUCTION

1.1. LIFE AT EXTREME ENVIRONMENTS: HALOPHILIC ORGANISMS

Along evolution, life has successfully adapted to conquer almost every possible environment on Earth. Biological organisms have evolved to survive in extreme conditions such as high / low temperatures, high pressure, acidic or basic media, absence of oxygen, high salt concentrations... These organisms are widely known as **extremophiles**, as opposed to **mesophiles** that only thrive at mild regular conditions. Particularly, those adapted to high salt environments are named **halophiles**. *Halo-* derives from *alos* (*αλος*), the greek word for salt, whereas *-philic* derives from the greek *filos* (*φιλος*), meaning “friendly” or “love for”. Altogether, the term *halophilic* is referred to salt-loving or salt-friendly. In the context of this thesis, the word *halophilic* is applied to a series of organisms that have adapted to live in extreme conditions of high salinity.

Halophilic organisms thrive in hypersaline environments in which salt concentrations exceed 150-200 g/L (> 3M). They inhabit salt lakes such as the Great Salt Lake in Utah, the Dead Sea, other hypersaline water bodies such as the crystallizer ponds of solar salterns in which sea water is evaporated for the salt production, natural salines like Salar de Uyuni in Bolivia or Salinas de Añana in the Basque region of Alava (Figure I1), salted fish and hides preserved by treatment with salt... (Oren 2006)



Figure I1. Salinas de Añana. Located at the natural water springs of Muera River, which travels across underground salt deposits in the Basque region of Álava. First reports of exploitation date from 822 A.D.

Microbial deterioration of salted hides and salted fish by members of the halobacteriales is clearly visible by the pink-red color of the developing colonies. Many of the early studies in the 20s on the red halophilic *archaea* were initiated as an attempt to understand the cause of the damage to fish treated with salt for its preservation (Harrison and Kennedy 1922) (Klebahn 1919). Ironically, many of these new forms of life were first discovered at the Death

Sea, named after the belief that no life forms could survive due to its high salt content. Around 1940, prokaryotic and unicellular eukaryotic organisms were reported to live in these harsh waters, long time thought to be depleted of any life form. Benjamin Elazari-Volcani published the first report of microbial life in the Dead Sea (Elazari-Volcani 1936). *Haloferax volcanii*, which was isolated from Dead Sea mud in 1977, was named after him to commemorate his work (Mullankhanbhai and Larsen 1975). The most abundant organisms found were extreme halophilic halobacteria from the *archaea* domain, but also some bacterial halophiles, anaerobic halophilic *bacteria* and some *eukarya*, such as the algae *Dunaliella salina* (Oren 2006).

Quite often, extreme halophilic organisms require a minimum salt concentration in the media to survive. In that case, they are named **strict halophiles**. For example, *Haloferax volcanii* and *Haloarcula marismortui* need at least a minimum NaCl concentration of 1 and 1.7 M, respectively. On the other hand, **halotolerant** organisms can endure low salt concentrations, such as *Oceanobacillus iheyensis* that can thrive in total absence of salt (Figure I2).

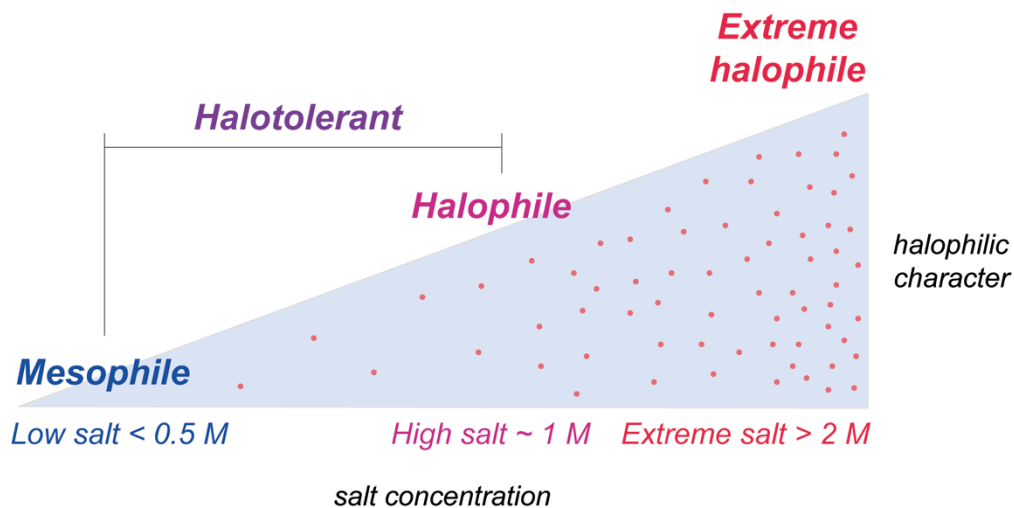


Figure I2. Halophiles and mesophiles. Mesophile organisms thrive in media near regular physiological conditions. Halophiles, on the other hand, populate media with molar salt concentrations. Strict halophiles need a minimum salt concentration in order to survive, whereas halotolerant organisms can endure both high and low salt concentrations. Most halophilic *archaea* can endure extremely high salinities up to 4-5 M.

1.2. ADAPTIVE STRATEGIES

High salt environments pose a strong bias to life development. The large extracellular salt concentration exerts a negative osmotic pressure on the cell, driving water out of the cytoplasm, causing cell shrinkage and, ultimately, cell death by desiccation. This process, commonly known as **osmotic shock**, is the ultimate reason for using high solute concentrations as food preservatives, e.g. salted foods and syrups.

Halophilic organisms have adopted a common protective strategy to avoid osmotic shock: they **accumulate molar concentrations of cosolutes in the cytoplasm** in order to match osmolarities at both sides of the cell membrane. Two principal classes of cosolutes are accumulated: organic compatible solutes and inorganic ionic salts.

1.2.1. Compatible solutes

Some *eukarya* – unicellular algae, vascular plants, invertebrates and vertebrates – and *bacteria* (Yancey, Clark et al. 1982) synthesize and accumulate large amounts of neutral organic molecules in the cytoplasm (between 1 and 1.5 M). The fact that such phylogenetically diverse organisms share a small set of molecules, suggests a strong selective pressure associated with this striking example of convergent evolution.

These neutral molecules are denominated **compatible solutes** due to their neutral effect on the native structure of proteins and on macromolecule-solvent interactions. Likely, this results from the absence of interactions with biological substrates and cofactors. These osmoprotectants are chemically classified in different groups: *saccharides* such as sucrose, trehalose or mannose; *polyols* like glycerol, sorbitol or mannitol; *amino acids* and derivatives such as glycine, taurine or β -alanine; *methylamines* like trimethyl ammonium oxide (TMAO), betaine, ectoine, sarcosine, and others, usually accompanied by urea (Yancey, Clark et al. 1982). Many of these solutes also show cytoprotective properties that go beyond simple compatibility and vary from solute to solute (Table I1), such as antioxidation or stabilization of proteins. That's why they are also known as **protecting osmolytes**. As a result, high quantities of these molecules are also present in other organisms under stress conditions, especially in hyperthermophilic bacteria. This is not surprising, since a variety of other stresses (temperature, high pressure, oxidative, etc.) can co-occur with water stress in the environments of extreme salinity where halophilic organisms survive. Usually, environmental stresses are denaturing. Nature's incorporation of protecting osmolytes within the cellular milieu stabilizes intracellular proteins while protecting the cell against water stress. This vital strategy establishes a direct link between cellular function and folding energetics, underscoring the central role of solvent-protein interactions in living systems.

Accumulation of organic cosolutes acts as a driving force for water flow towards the interior of the cell, which in turn compensates the osmotic pressure exerted by the large extracellular salt concentration. This is called the *salt-out* strategy. Unicellular organisms generally complement this strategy with an ion-pumping mechanism to maintain a low (physiological) intracellular concentration of ions (Yancey, Clark et al. 1982), thus protecting macromolecules and essential processes from the adverse effects of salt ions.

Cytoprotective / Stabilizing Property	Compatible Solute in Nature
Increase thermostability	Trehalose, anionic polyols (e.g. vent archaeon)
Counteract hydrostatic pressure	TMAO (e.g. deep-sea fish)
Protect membranes against freezing	Trehalose, proline (e.g. frozen insect)
Counteract urea inhibition	Methylamines, especially TMAO (e.g. shark)
Preserve in dry state	Carbohydrates, especially trehalose (e.g. dried yeast)
Counteract inorganic ion inhibition	Methylamines (e.g. salt marsh plant)
Antioxidation	Hypotaurine; DMSP; polyols (e.g. water-stressed plant)
Redox balancing, hypoxia	Proline, β -alanine betaine, glycerol (e.g. <i>Dunaliella</i> alga)
Sulphide detoxification/storage	Hypotaurine (e.g. hydrothermal vent tubeworm, <i>Riftia</i>)
Sulphate detoxification	Choline-O-sulphate (e.g. mangrove plant)
Energy reserve	Glucose, trehalose, etc. (e.g. frozen wood worm)
Predator repellent	DMSP, <i>trans</i> -hydroxyprolinebetaine (e.g. diatom)
Calcium modulation	Taurin (e.g. mammalian neuron)

Table 11. Compatible solutes. Summary of protective properties of osmolytes through metabolic reactions and stabilization of macromolecules and membranes (Yancey 2005).

By accumulating compatible or stabilizing solutes, osmotically enriched organisms may achieve simultaneously a genetically “simple” solution and a temporarily far more flexible adaptive mechanism to survive in hypersaline environments. For these organisms the need for adaptive changes and mutations in proteins is greatly minimized as compared to organisms that accumulate inorganic ions. Rather than modifying hundreds to thousands of enzymes to cope with high or widely fluctuating intracellular salt concentrations, only a small fraction of the intramolecular repertoire is ultimately altered. Systems needing modification presumably include the enzymes controlling the synthesis and degradation of the organic osmolytes themselves.

This adaptive mechanism is successful in environments with salt concentrations up to about 1.5 M; e. g. ocean water with osmotic concentration of 1000 milliosmoles per liter (~1000 mOsm), well above the ~300-400 mOsm created by the common solutes found in most cells (K^+ , metabolites, proteins, etc.). However, this strategy is no longer effective at higher concentrations, where the synthesis of the large quantities required of those organic molecules would be energetically too expensive for the cell.

1.2.2. Intracellular KCl accumulation

Alternatively, a small group of halophilic prokaryotes is capable of exclusive exploitation of hypersaline environments with salt concentration between 2 and 5 M. Most **extreme halophiles** of this group belong to the order Halobacteriales (Archaea), as detailed in Table 12.

These halophilic organisms, typically **halophilic archaea**, have developed a much simpler strategy to deal with the osmotic shock: they **allow salt inside the cytoplasm**, accumulating molar quantities of ions to equilibrate the ionic forces inside and outside the cell membrane. It is called *salt-in* strategy, as opposed to the *salt-out* strategy, previously described.

Genus	Species	NaCl opt. (M)	NaCl range (M)	T opt. (°C)	pH opt.
Halobacterium	<i>Halobacterium salinarum</i>	4-5	3.0-5.2	35-50	Neutral
Halobaculum	<i>Halobaculum gomorrense</i>	1.5-2.5	1.0-2.5	40	6-7
Halorubrum	<i>Halorubrum saccharovorum</i>	3.5-4.5	1.5-5.2	50	Neutral
	<i>Halorubrum sodomense</i>	1.7-2.5	0.5-4.3	40	Neutral
	<i>Halorubrum lacusprofundi</i>	2.5-3.5	1.5-5.2	31-37	Neutral
	<i>Halorubrum coriense</i>	2.2-2.7	3.2	-	Neutral
	<i>Halorubrum distributum</i>	-	-	37-45	9.5
	<i>Halorubrum vacuolatum</i>	3.5	2.6-5.1	35-40	Neutral
	<i>Halorubrum trapanicum</i>	-	-	-	Neutral
	Haloarcula	<i>Haloarcula vallismortis</i>	3.5-4.3	-	40
<i>Haloarcula marismortui</i>		3.4-3.9	1.7-5.1	40-50	Neutral
<i>Haloarcula hispanica</i>		2.6	2.0-5.2	35-40	Neutral
<i>Haloarcula japonica</i>		3.4	2.5-5.1	42	7.0-7.5
<i>Haloarcula argentinensis</i>		2.5-3.0	2.0-4.5	40	Neutral
<i>Haloarcula mukohataei</i>		3.0-3.5	2.5-4.5	40	Neutral
<i>Haloarcula quadrata</i>		3.4-4.3	2.7-4.3	53	6.5-7.0
Natromonas	<i>Natromonas pharaonis</i>	3.5	2.0-5.2	45	9.5-10.0
Halococcus	<i>Halococcus morrhuae</i>	3.5-4.5	2.5-5.2	30-45	Neutral
	<i>Halococcus saccharolyticus</i>	4.3	2.5-5.2	37-40	6-8
	<i>Halococcus salifodinae</i>	3.4-4.3	2.6-5.2	40	6.8-9.5
Natrialba	<i>Natrialba asiatica</i>	3.5-4.0	2.0-5.2	35-40	6.6-7.8
	<i>Natrialba magadii</i>	3.5	2.0-5.2	37-40	9.5
Natronobacterium	<i>Natronobacterium gregory</i>	3.0	2.0-5.2	37	9.5
Halogeometricum	<i>Halogeometricum borinquense</i>	3.4-4.3	1.4-5.2	40	7
Natronococcus	<i>Natronococcus occultus</i>	3.4-3.8	1.4-5.2	35-40	9.5
	<i>Natronococcus amylolyticus</i>	2.5-3.4	1.4-5.2	40-45	9.0
Haloferax	<i>Haloferax volcanii</i>	1.7-2.5	1.0-4.5	40	Neutral
	<i>Haloferax gibbonsii</i>	2.5-4.3	1.5-5.2	35-40	6.5-7.0
	<i>Haloferax denitrificans</i>	2-3	1.5-4.5	50	6.7
	<i>Haloferax mediterranei</i>	2.9	-	40	6.5
Natrinema	<i>Natrinema pellirubrum</i>	3.4-4.3	-	-	7.2-7.8
	<i>Natrinema pallidum</i>	3.4-4.3	-	37-40	7.2-7.6
Haloterrigena	<i>Haloterrigena turkmenica</i>	-	-	45	Neutral
Natronorubrum	<i>Natronorubrum bangese</i>	3.8	-	45	9.5
	<i>Natronorubrum tibetense</i>	3.4	-	45	9.0

Table I2. Halophilic organisms. Family *Halobacteriaceae* in the order *Halobacteriales*: genera, species and properties (Oren 2006).

	Non-halophilic	Moderately halophilic	Extremely halophilic	
Organism	<i>Staphylococcus aureus</i>	<i>Micrococcus halodenitrificans</i>	<i>Halobacterium salinarum</i>	<i>Sarcina morrhuae</i>
Na ⁺ (*)	98 ± 20	311 ± 16	1370 ± 210	3170 ± 275
K ⁺ (*)	680 ± 52	474 ± 38	4570 ± 121	2030 ± 359
Cl ⁻ (*)	8 ± 3	55 ± 6	3610 ± 70	3660 ± 250
Amino acids (*)	438 ± 25	244 ± 12	209 ± 26	88 ± 2
NaCl external (M)	0.15	1.0	4.0	
K ⁺ external (M)	0.025	0.004	0.032	

(*) : Intracellular concentrations expressed as *mmoles/kg cell water*

Table I3. Intracellular salt concentrations. Apparent concentrations of solutes within cells of halophilic and non-halophilic organisms (Christian and Waltho 1962).

The accumulation is ion selective, being K^+ by far the most abundant cation in the cellular media, and Cl^- the major counterion (Christian and Waltho 1962). Regardless of Na^+ being the most abundant extracellular cation, intracellular KCl concentrations always exceed that of NaCl by several folds, and they can reach values near saturation and as high as 5 M, like in *Halobacterium salinarum* (Table I3). Such a selective KCl accumulation is achieved through active ion transporting across the membrane. ATP dependent TrkA potassium uptake system and Na^+/H^+ antiporter have been identified as key players in *Haloferax volcanii* haloadaptation, whereas in other organisms such as *Halobacterium halobium* bacteriorhodopsin coupled systems generate a H^+ electrochemical potential. (Meury and Kohiyama 1989).

This protective strategy is based in the same physical principle that drives the accumulation of compatible solutes. However, since the extensive synthesis of organic molecules is no longer required, cells avoid major changes in their metabolic pathways. As a result, a **simpler adaptive strategy** from a metabolic point of view, with a more efficient use of nutrients and energy, is achieved.

KCl accumulation is typical from halophilic archaea, but it has also been observed, to a lesser extent, in the fermentative obligatory anaerobic halophilic bacteria from the order Halanaerobiales (Oren 2006), in the Halobacteria *Salinibacter ruber* (Antón, Oren et al. 2002) and in the Proteobacteria *Halorhodospira halophila* (Deole, Challacombe et al. 2013). Thus, it constitutes an adaptive strategy that, although limited, is quite phylogenetically spread over unrelated families of organisms ranging from *archaea* to *bacteria* kingdoms.

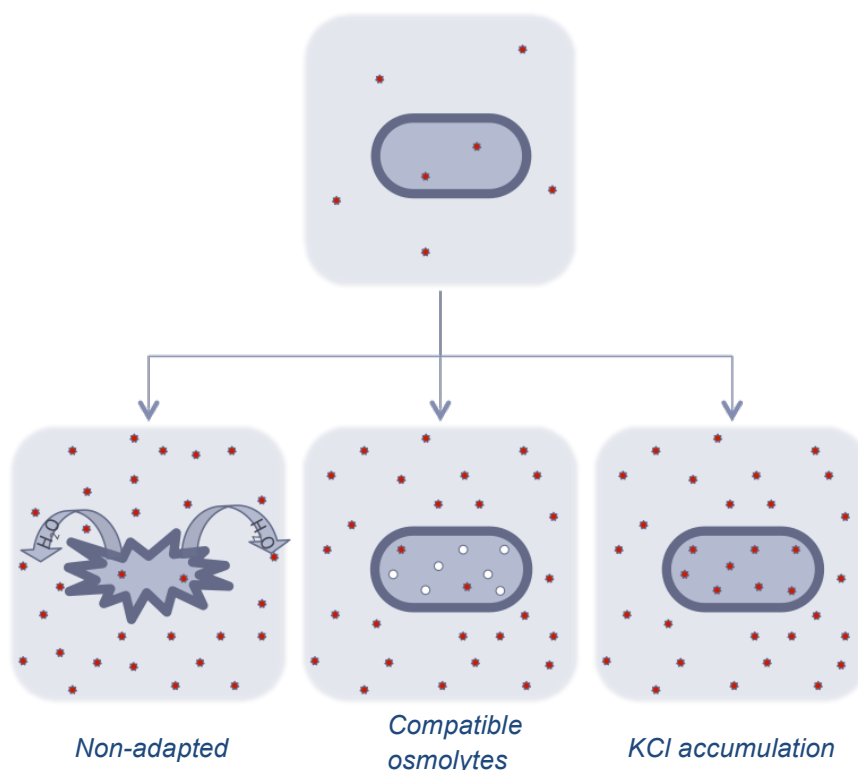


Figure I3. Adaptive strategy to avoid osmotic shock. Cells maintain the osmotic balance with the extracellular milieu (up). Non-adapted cells die of dissection in high salt concentrations (left). Halophilic organisms resist the osmotic shock by equilibrating the intracellular osmotic pressure with the one outside. They do so by accumulating compatible organic solutes (middle) or, for most extreme conditions, KCl (right).

Extreme halophilic organisms selectively accumulate molar quantities of KCl in their cytoplasm

1.3. MOLECULAR BASIS OF HALOADAPTATION

Despite the apparent simplicity of the adaptive mechanism, the presence of a high intracellular concentration of solute ions is generally devastating for the functional conformation of proteins and other macromolecules (Dennis and Shimmin 1997):

- i. it causes **aggregation** or structural collapse of proteins due to the enhancement of hydrophobic interactions,
- ii. it interferes with essential **electrostatic interactions** within or among macromolecules weakening their salt bridge interactions and
- iii. because of competition with salt ions **hydration**, the availability of free water is reduced below the threshold required to sustain essential biological processes.

In order to remain functional and active at such high intracellular salt concentrations, the cellular machinery has undergone a major reengineering along evolution. The adaptive features of halophiles are likely related to the physical principles governing DNA and protein stability in response to the extreme environmental conditions under which they thrive.

It has been reported a higher genomic GC-content (well above 60%) in some halophilic organisms; presumably to avoid UV induced thymidine dimer formation and the possible accumulation of mutations in their specialized habitat: shallow coastal lagoons characterized by high levels of UV irradiation (Paul, Bag et al. 2008). However, high salt concentrations affect mostly the conformational stability of proteins, and it is widely agreed that **the main halophilic adaptation has occurred at the protein level** (Madern, Ebel et al. 2000).

1.3.1. Halophilic signature: amino acid composition

Halophilic proteins exhibit a strongly biased amino acid composition as compared to their mesophilic counterparts (Figure I4). **Aspartic acid, to a lesser extent, glutamic acid, are found at higher frequencies** compared to mesophilic proteins. Residues with **short polar side chains, particularly threonine, are also enriched**. On the other hand, **lysines are greatly penalized**, accompanied by an **overall reduction of residues with large hydrophobic side chains**: cysteine, methionine, tryptophan, phenylalanine, leucine, isoleucine, proline and tyrosine (Paul, Bag et al. 2008). Figure I4 shows the relative abundance of each amino acid type in halophilic proteins compared to the average protein amino acid composition calculated following equation I1, where \bar{n}^{aa} is the average number of a given amino acid *aa* in halophilic or mesophilic sequences:

$$\text{Relative abundance (\%)} = \Delta aa_{h-m} = \frac{\bar{n}_h^{aa} - \bar{n}_m^{aa}}{\bar{n}_m^{aa}} \cdot 100 ;$$

Equation I1. Average relative abundance in halophilic proteins of a given amino acid as compared to mesophilic proteins.

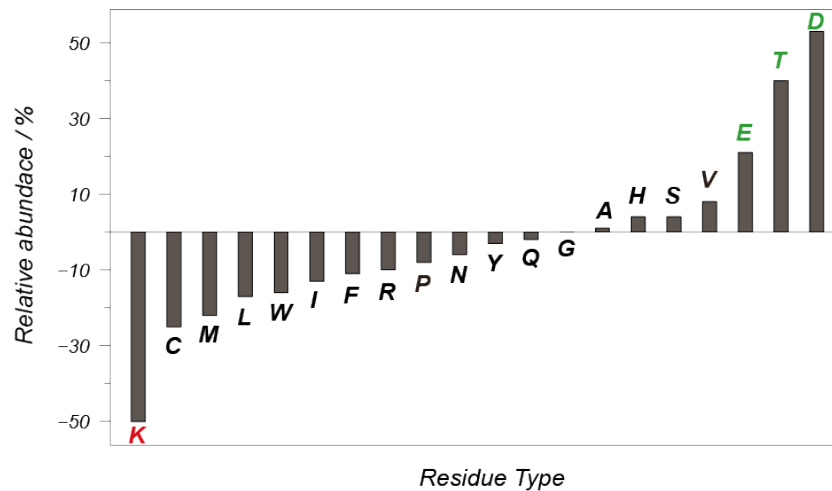


Figure 14. Halophilic amino acid composition. Residue abundance in halophilic proteins compared to mesophilic proteins expressed as the percentage of relative variation in the average amino acid composition. (Paul, Bag et al. 2008)

Such a biased amino acid composition is known as **halophilic signature** and it is exclusive from proteins belonging to halophilic organisms that accumulate KCl inside their cytoplasm. Genomic analyses of taxonomically distant species (Fukuchi, Yoshimune et al. 2003) and statistical comparison of several halophilic and mesophilic protein primary sequences (Madern, Pfister et al. 1995) reflect similar distributions in the amino acid composition of halophilic proteins, indicating that this halophilic signature is a **universal evolutionary trait**.

These deviations from mesophilic amino acid compositions are found to occur mainly, but not exclusively, **at the surface of the protein**, therefore suggesting that haloadaptation mechanism relies on protein-solvent interaction (Fukuchi, Yoshimune et al. 2003).

As a result, halophilic proteins exhibit a low hydrophobic content and a surface with a large **negative net charge** (Figure 15), falling amongst the most acidic proteins in the proteome. Despite all these changes, **fold is preserved** with respect to their mesophilic counterparts (Figure 16 & Table I4). This is not surprising since the folded conformation entails most of the biological function (Paul, Bag et al. 2008).

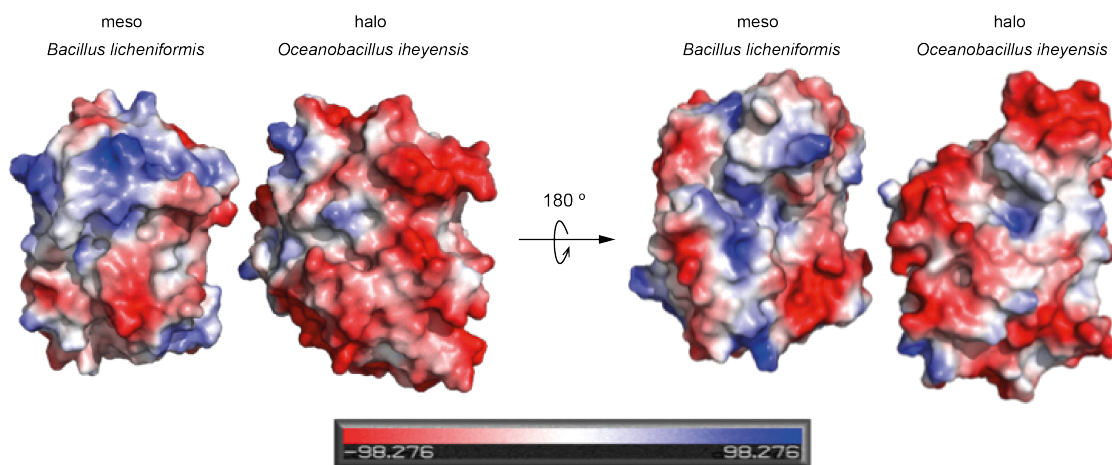


Figure 15. Surface negative charge. Comparison of the surface potential obtained for halophilic (*Oceanobacillus iheyensis*, PDB: 3LEZ) and mesophilic (*Bacillus licheniformis*, PDB: 1NRF) β -lactamases. Poisson Boltzmann calculations were run on PyMol with APBS plug-in (Baker, Sept et al. 2001). Halophilic proteins show a more negative potential in the surface (in red).



Figure 16. The structure is conserved. Structure of dihydrofolate reductase from mesophilic *Escherichia coli* (left, PDB: 5DFR) and halophilic *Haloferax volcanii* (right, PDB: 2ITH). The overall fold is conserved.

Protein	Halophilic organism	PDB, resol. (Å) oligomeric st., N _{res}	Meso/Thermophilic organism	PDB, resol. (Å) oligomeric st., N _{res}	Sequence Identity (%)	C _α RMSD (Å)
Ferredoxin	<i>Haloarcula marismortui</i>	1doi, 1.9, monomer, 128	<i>Anabaena sp.</i>	1fxa, 2.5 monomer, 98	51	1.96
Catalase- peroxidase	<i>Haloarcula marismortui</i>	1itk, 2.0 dimer, 731	Burkholderia pseudomallei	2fxg, 1.8 dimer, 748	60	1.54
Photoreceptor	<i>Halorhodospira halophila</i>	1nwz, 0.8, monomer, 125	Rhodospirillum centenum	1mzu, 2.0 monomer, 129	46	1.42
Cupin 2–domain containing protein	<i>Halorhodospira halophila</i>	3ibm, 2.0 dimer, 167	Rhodopseudomonas palustris	3kgz, 1.8 dimer, 156	44	1.08
DNA–protecting protein	<i>Halobacterium salinarum</i>	1tjo, 1.6 dodecam., 182	Thermosinecho- coccus elongates	2vxx, 2.4 dodecamer, 192	36	0.94
Nucleoside di- phosphate kinase	<i>Halobacterium salinarum</i>	2az3, 2.2 hexamer, 164	Saccharomyces cerevisiae	3b54, 3.1 hexamer, 161	54	0.73
Dodecin	<i>Halobacterium salinarum</i>	2cc6, 1.3 dodecamer, 68	Thermus thermophilus	2v18, 2.6 dodecamer, 68	42	1.94
Periplasmic ectoin binding protein	<i>Halomonas elongata</i>	2vpn, 1.5 monomer, 316	Ruegeria pomeroyi	3fxb, 2.9 monomer, 326	62	0.92
Proliferating cell nuclear antigen	<i>Haloferax volcanii</i>	3ifv, 2.0 trimer, 247	Archaeoglobus fulgidus	1rwz, 1.8 trimer, 245	36	1.72
GGDEF domain	<i>Marinobacter aquaei vt8</i>	3ign, 1.8 monomer, 177	Pseudomonas aeruginosa pao1	3i1c, 1.9 monomer, 206	40	2.91
D–mannonate dehydratase	<i>Chromohalobacter saalexigens</i>	3bsm, 2.2 octamer, 413	Novosphingobium aromaticivorans	2qij, 1.8 octamer, 402	66	0.59
Glucose dehydrogenase	<i>Haloferax mediterranei</i>	2b5w, 1.6 dimer, 356	Sulfolobus solfataricus	2cd9, 1.8 tetramer, 366	30	1.47
Malate dehydrogenase	<i>Haloarcula marismortui</i>	2j5k, 1.9 monomer, 304	Clostridium thermocellum	1y6j, 3.0 monomer, 318	33	1.47
Cytochrome c552	<i>Marinobacter hydrocarbonoclasticus</i>	1cno, 2.2 monomer, 87	Pseudomonas stutzeri	1etp, 2.2 monomer, 190	47	1.00
Cytochrome c peroxidase	<i>Marinobacter hydrocarbonoclasticus</i>	1nml, 2.2 dodec., 326	Geobacter sulfurreducens	3hq6, 2.2 dimer, 345	64	1.73

Table 14. Comparison between the structures of proteins from halophilic and non–halophilic organisms. Despite the sometimes-low sequence homology, the structures are highly conserved, as extracted from the low C_α RMSD (Root-Mean-Square Deviation) values (Pica, Krauss et al. 2013, Graziano and Merlino 2014).

1.3.2. Salt dependence of protein thermodynamic stability

The thermodynamic stability of folded proteins corresponds to the Gibbs free energy difference between the folded and the unfolded state ΔG^{U-F} . Both enthalpic and entropic contributions modulate this magnitude. Enthalpic intermolecular interactions between the protein and the solvent favor the unfolded state, where more interaction sites are available. Intramolecular interactions, on the other hand, occur mainly in the folded conformation, lowering its chemical potential and resulting in folded state stabilization. Regarding entropic contributions, the favorable gain of protein entropy upon unfolding is greatly compensated by the huge decrease in solvent entropy: the translational and rotational diffusion of polar water molecules solvating the exposed hydrophobic side chains are greatly restrained, in turn, no interactions are formed due to the distinct character of water and newly exposed hydrophobic area. The latter contribution is the dominant one in folded proteins, and it is commonly known as the **hydrophobic effect**.

Cosolutes such as ionic species or denaturants alter the balance between entropic and enthalpic contributions, resulting in stabilization or destabilization of the protein. As opposed to the already mentioned negative effects on mesophilic proteins, *biased amino acid composition makes halophilic proteins to become **monotonically stabilized upon KCl addition*** (and, to a lesser extent, NaCl) (Polosina, Zamyatkin et al. 2002). Other ionic solutes have also been studied (Ebel, Faou et al. 1999) and, similarly to non halophilic proteins, the ion induced effects on protein's stability have been found to correlate well with the ranking established in the **Hofmeister series** more than 100 years ago (Hofmeister 1888). Moreover, the stabilizing or destabilizing effects are greatly amplified on halophilic proteins (Tadeo, López-Méndez et al. 2009a). Remarkably, this stabilization/destabilization, understood as the increase/decrease in the unfolding free energy upon salt addition, shows a strong **linear correlation** with the molar cosolute concentration (Tadeo, Pons et al. 2007).

Interestingly, many halophilic proteins (obligate/strict halophiles) became greatly destabilized in the absence of salt, and extensively unfold at KCl / NaCl concentrations below 1 – 2 M. This is thought to be due to the large electrostatic repulsion between the abundant negatively charged residues in the surface of the protein.

The rigorous reporter for the conformational stability of a protein is given by the free energy difference between the folded and unfolded states. However, it is commonly assayed as residual activity measurements, especially for complex systems showing non-ideal behavior; e.g. irreversible unfolding, aggregation, oligomerization, etc. Such an assumption can be easily misleading, since the mechanisms modulating salt dependence of activity might be unrelated to those determining the energetics of the unfolded – folded equilibrium.

Halophilic signature:

- *increase of acidic (Asp, Glu), short (Val) and polar (Thr) residues*
- *decrease of Lys and bulky hydrophobic side chains.*

Adaptation occurs mainly in the surface: the structure is conserved.

Halophilic proteins' stability increases linearly with KCl.

1.3.3. Salt modulation of enzymatic activity

The monotonic stabilization of halophilic proteins by KCl and NaCl contrasts with its **complex modulation** on the functionality of halophilic enzymes. To name a few representative examples, β -galactosidase from *Haloferax alicantei* (Holmes, Scopes et al. 1997) and dihydrofolate reductase (Wright, Banks et al. 2002) from *Haloferax volcanii* are activated by salt, as opposed to other proteins such as glucose dehydrogenase from *Haloferax mediterranei* (Esclapez, Pire et al. 2007) and catabolic ornithine transcarbamylase of *Halobium salinarum* (Madern, Ebel et al. 2000) that show a decrease in their catalytic activity at increasing salt concentrations. More complex phenomena are the different dependency on NaCl addition seen for the two activities of catalase-peroxidase from *Halobacterium halobium* (Brown-Peterson and Salin 1993).

More strikingly, some enzymes show an ion dependent modulation of activity: e.g. 3-hydroxy-3-methylglutaryl-coenzyme A reductase from *Haloferax volcanii* increases activity with KCl concentration while decreases it with NaCl concentration (Madern, Ebel et al. 2000). Also DNA ligase N from *Haloferax volcanii* (Hv LigN) shows an equivalent complex behavior: addition of KCl activates the enzyme, whereas it becomes inactive at all NaCl concentrations (Poidevin and MacNeill 2006). Moreover, the addition of other inorganic salts causes various effects on protein activity.

Thus, the variable salt effects on the activity of halophilic enzymes have been demonstrated for many case studies and tested over different ionic species, but still a unified molecular mechanism is missing.

The complex salt modulation of halophilic enzymes contrasts with the monotonic dependence of stability

1.3.4. Salt modulation of protein solubility

Protein solubility is mainly determined by solvent-protein interface properties. Attractive weak protein – protein interactions promote aggregation, whereas repulsion forces favor dissolution. High salt concentrations alter the equilibrium between the soluble and the aggregated form. *Salting-out* ions promote aggregation, whereas *salting-in* ions favor dissolution. On the one hand, if ions are excluded from protein's surface (*salting-out*), local concentration of the solvation shell is lower than that of the bulk solvent, the hydrophobic effect is enhanced and aggregation processes are favored. Stabilizing salts include sodium or potassium chloride, sulphate and phosphate, among others. On the other hand, when ions interact with the protein surface (*salting-in*) the bulk and the solvation shell concentrations are similar and the protein is dissolved. Denaturants, like guanidinium chloride, and destabilizing salts, like nitrates or perchlorates, act mainly through this mechanism.

Mesophilic proteins tend to aggregate in highly concentrated NaCl and KCl solutions, whereas halophilic enzymes remain soluble (Madern, Ebel et al. 2000). Small angle neutron scattering (SANS), small angle X-ray scattering (SAXS) and analytical ultracentrifugation (AUC) experiments were performed over a series of halophilic and non-halophilic systems (Coquelle, Talon et al. 2010). Analysis of the sedimentation coefficients obtained from AUC and the hydrodynamic information obtained from SAXS and SANS showed that, at high salt concentrations, mesophilic proteins were in the attractive regime. Behavior of halophilic proteins, on the other hand, corresponded to the repulsive regime. Therefore, halophilic proteins are less prone than non-halophilic proteins to intermolecular attractive interactions, ultimately leading to aggregation.

Attractive forces are dominated by weak hydrophobic protein – protein interactions. In halophilic proteins, a decrease in the surface exposed hydrophobic area, mainly achieved through a reduction of surface lysines, diminishes these attractive intermolecular interactions. On the other hand, **repulsive electrostatic forces** are enhanced by the higher abundance of acidic residues in the surface of the protein. Therefore, the electrostatic potential and the exposed hydrophobic area of the surface ultimately determine the solubility of halophilic proteins. This observation has been experimentally corroborated. For instance, correlation between surface charge and solubility was found by Tadeo and coworkers (Tadeo, López-Méndez et al. 2009b) and mechanistic studies on the effect of lysine glycation in aggregation also underscore the role of surface net charge and area exposed to solvent in aggregation processes (Adrover, Mariño et al. 2014).

1.4. MODELS FOR HALOADAPTATION

Considerable efforts have been dedicated to understand the mechanistic basis of halophilic adaptation. Experiments particularly focused on how the evolutionary selected set of amino acids modulates the properties of halophilic proteins. Deep characterization of halophilic systems from a thermodynamic, functional and structural point of view, complemented with studies of the effects of different ionic species on proteins, lead to the establishment of the first mechanistic **halophilic model**. However, the model is incomplete and some unexplained evidence challenged this traditionally accepted theoretical frame. Based on those limitations, novel aspects and contributions to haloadaptation have been unveiled.

1.4.1. The halophilic model: molecular basis and limitations

G. Zaccai, M. Mevarech, H. Eisenberg and coworkers established the traditional halophilic model during the late '80s – '90s. Their pioneering work on halophilic malate dehydrogenase from *Haloarcula marismortui* (*HmMalDH*) constitutes the first biochemical, thermodynamic and structural characterization of a halophilic enzyme (Mevarech, Eisenberg et al. 1977) (Pundak, Aloni et al. 1981). Extensive data on *HmMalDH* and other halophilic systems lead to the solvation-stabilization model, based on three key observations (Madern, Ebel et al. 2000):

- i. enthalpic mechanisms dominate the kinetic deactivation in molar KCl and NaCl,
- ii. folded proteins bind relatively large amounts of salt and water in KCl, NaCl or MgCl₂ solutions and
- iii. the excess of acidic amino acids in the protein composition could provide favored sites for specific water and ion binding to the tertiary and quaternary structure.

According to the model, the protein forms **specific structural arrangements** in the tertiary or quaternary structure: **organized patches of carboxyl groups stabilized by the interaction with cooperative hydrated cation networks, particularly potassium**. As a result, the folded conformation is preferentially stabilized as compared to the unfolded conformation by the presence of a high concentration of hydrated ions (Zaccai, Cendrin et al. 1989, Mevarech, Frolov et al. 2000, Ebel, Costenaro et al. 2002).

The specific arrangement of acidic side chains, and not only the excess of negative charges in the surface, would therefore be the main adaptive feature in halophilic proteins. This is consistent with the fact that acidic residues are highly hydrated, more than any other amino acid (Christian and Waltho 1962), and can coordinate the organization of a hydrated salt ion network at the surface of the protein. Accordingly, a model based in interactions of electrostatic nature is expected to be **enthalpy driven**, and is also in agreement with large amounts of salt and water bound to the protein surface. It is important to note, though, that since *HmMalDH* does not show a reversible cooperative unfolding, the stability on the thermodynamic studies was measured as residual activity (Zaccai, Cendrin et al. 1989). Provided that the salt modulation of the enzyme activity is far more complex than protein stability and the unfolding events are controlled by kinetics rather than by a reversible thermodynamic equilibrium, these data should be regarded with caution.

On the other hand, evidences for **water and ion binding** came mainly from SANS, SAXS and AUC (Ebel, Costenaro et al. 2002). As the authors admit, those approaches “cannot determine ion distribution or mechanisms of binding”, so the interpretation is heavily model – dependent. For instance, negligible effects of anion concentration on the preferential binding parameters are reported and attributed to the presence of a limited number of strong anion binding sites. Posterior thorough thermodynamic studies on the role on anion stabilization provide convincing evidence that discriminate anion stabilizing effects between high and low salt concentrations. Anion binding sites exist, indeed, and the interactions are not purely electrostatic, but those sites are already saturated at low concentrations (< 0.1 M) (Ramos and Baldwin 2002). Surprisingly, moderate anion dependent linear stabilization still occurs at much higher concentrations (> 1 M), in agreement with Hofmeister effects (Ramos and Baldwin 2002, Tadeo, Pons et al. 2007). Strong specific anion binding sites are not consistent with such experimental observations, since they are expected to be saturated at low millimolar concentrations. They are therefore unable to account for a linear stabilization at high molar values. The same applies to cation binding. A model based in the interactions between carboxylic groups and hydrated cations has a limited number of binding sites, expected to be fulfilled at salt concentrations around 0.1 M.

Furthermore, crystal structures of halophilic proteins have **failed to identify hydrated cation networks**. Particularly 2Fe-2S ferredoxin (Frolow, Harel et al. 1996) and *HmMalDH* (Dym, Mevarech et al. 1995) fail to show constructive interactions between negatively charged clusters in the surface and hydrated cations. Some ions are reported, but mainly at positions involved in oligomerization interfaces or salt-bridges between side chains. Support is given, nevertheless, to the notion of exceptional hydration of halophilic proteins by an “immobile”, “quasi-crystalline” first layer of “tightly bound” water molecules inducing four additional organized water layers. Latter structures of halophilic proteins at higher resolution also show that counterions from the solvent do not appear to play a dominant role in neutralization of the negative charge of the surface (Britton, Baker et al. 2006), i.e., for the five K^+ ions bound to glucose dehydrogenase from *Haloferax mediterranei*, only one interaction with a carboxyl oxygen is found, whereas most of them interact with carbonyl oxygens from the backbone chain. Interestingly, again the number of total water molecules reported is higher than that for non-halophilic proteins, revealing a “highly ordered, multilayered solvation shell”.

In summary, the traditionally accepted solvation-stabilization halophilic model interprets salt adaptation of halophilic proteins as the process of remodeling protein surface to create specific ion binding sites. The electrostatic interactions between ions and particular structural arrangements – groups of carboxylates – provide the energetic contribution favorable for the folded conformation. Subsequently, this model can successfully explain the high abundance of acidic residues in halophilic sequences and the stabilization at high salt concentrations. Unfortunately, such interactions have not yet been experimentally demonstrated at high resolution, neither accurately quantified in energetic terms. Observations of moderate linear stabilization upon salt addition at high molar concentrations cannot be attributed to such interactions since they should be saturated at much lower concentrations and are not expected to promote a linear increase in stability. Alternatively, these observations point to Hofmeister effects. In addition, the model fails to apply its mechanistic principles to successfully rationalize the complex salt modulation of the activity of halophilic enzymes.

Halophilic model explains haloadaptation in terms of electrostatic interactions between hydrated K^+ and specific binding sites formed by arrangements of carboxylates at the surface of halophilic proteins

1.4.2. The solvent exposed area and the Hofmeister effect

A linear increase in the protein's unfolding free energy at high molar salt concentrations has already been observed in the context of the Hofmeister series. **Franz Hofmeister** studied in the late 1800s the effectiveness of different cosolutes in precipitating the proteins of egg white (Hofmeister 1888), ordering the different anions and cations accordingly in a series named after him (Figure I7). The original ion ranking in the series has been proved applicable to plenty of other biological phenomena, and extensive research has been dedicated to the topic for more than a hundred years (Cacace, Landau et al. 1997). Originally, ions were divided between *kosmotropic* for water structure enhancing ions, and *chaotropic* for water structure breakers. The mechanistic theory underlying the Hofmeister series has changed considerably throughout the years, and today the classification of **stabilizer versus destabilizer** ions is preferred.

A complete mechanistic understanding of the effect of cosolutes on macromolecules still remains elusive, but the role of solvation, interactions and preferential exclusion seem to be central to the current models.

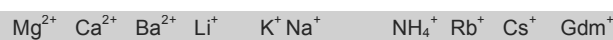
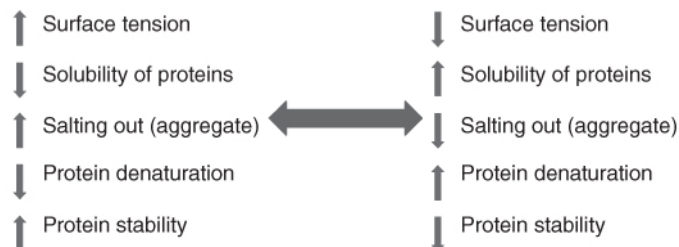


Figure I7. Hofmeister series. A commemorative plaque at the Medical Faculty of the Charles University in Prague reading (in Czech and German): “Professor Franz Hofmeister (1850-1922), who carried out research in this building, predicted that amino acids in proteins are connected by a peptide bond and, in 1888, derived the lyotropic (Hofmeister) series of ions.” Photo by Pavel Jungwirth.

Recent studies (Ramos and Baldwin 2002, Fayos, Pons et al. 2005) show that protein stability is linearly dependent on ion concentration, even at high molar values, and the extent of this stabilizing effect is heavily modulated by the nature of the ionic specie following Hofmeister series, specially anions (Tadeo, Pons et al. 2007). More detailed experiments outlined the correlation between the area exposed to solvent and the ultimate Hofmeister modulation of protein properties. The stability of mutants of a model protein was measured at increasing concentrations of various salt species. A nice empirical correlation between the salt dependence of protein stability and the **change in solvent accessible area upon unfolding** (Δ ASA) introduced upon mutation was found, in particular for the non-polar Δ ASA of the side chains (Tadeo, López-Méndez et al. 2009a). Further efforts have recently focused in understanding the role of accessible solvent area in halophilic proteins. Extensive mutagenesis and a thorough thermodynamic and structural characterization allowed an unprecedented understanding of the contributions of sequence halophilicity to salt adaptation (Tadeo, López-Méndez et al. 2009b). Several cumulative mutation lines were designed in order to “tune” the halophilic character of model proteins. Selected features, such as net surface charge or solvent exposed hydrophobic area, were progressively altered mimicking the evolutionary process of haloadaptation. Strikingly, mutations increasing the negative charge in the surface of the protein had negligible effects on salt stabilization, whereas mutations decreasing the solvent-accessible surface enhanced halophilicity. Again, a strong empirical correlation between the halophilic character of the protein and the changes in the area accessible to solvent introduced upon mutation was found.

The traditional halophilic model underestimates a distinct feature of halophilic sequences: their significant **reduction in the overall hydrophobic character**. A model based in ionic interactions between charged particles or hydrated ions does not require a reduction of the hydrophobic exposed area, but the fact that this is a universal evolutionary feature suggests that it plays an important role in protein adaptation to high salt environments. In fact, all the studies about halophilic proteins have focused mainly on its acidic character, whereas their reduced hydrophobicity has been largely ignored for long time. Reduction of the nonpolar content occurs mainly in the surface, but also in the core of the protein. On the one hand, reduction of solvent exposed hydrophobic area is mainly achieved through a strong restriction of the number of surface **lysines** (Britton, Baker et al. 2006). Solvation of the long and disordered aliphatic side chain of lysine introduces a large energetic cost in terms of water loss of entropy. As a result, it becomes highly penalized in environments with restricted water activity. Nevertheless, key lysines for catalysis and function are preserved. On the other hand, a minor prevalence of residues with **bulky hydrophobic side** chains occurs throughout the whole sequence (Fukuchi, Yoshimune et al. 2003). Obviously, hydrophobic contacts essential for a proper folded conformation are maintained (Siglioccolo, Paiardini et al. 2011). Accordingly, isoleucine and leucine substitutions to valine or other smaller hydrophobic residues (e.g. threonine) rank amongst the ten most common mutations between non-halophilic and their halophilic orthologues (Paul, Bag et al. 2008). Such conservative mutations maintain the hydrophobic nature of the residue while decreasing the size of the side chain and the overall hydrophobic content.

The determinant role of exposed hydrophobic surface in the extent of salt stabilization of halophilic proteins suggests that solvation plays a relevant role in haloadaptation, at least comparable to that of electrostatic interactions with negative charges, if not larger. Consistent with this idea is the slightly higher number of water molecules reported in X-Ray crystal structures of halophilic proteins as compared to those of mesophilic proteins (Britton, Baker et al. 2006).

1.5. BIOLOGICAL INTEREST AND POTENTIAL APPLICATION OF THE STUDY OF HALOPHILIC PROTEINS

Approximately 97% of all water in Earth is present in saline oceans, saline lakes, inland seas and saline groundwater; and salt deposits underlie roughly one-quarter of the land on Earth (Deole, Challacombe et al. 2013). Thus, saline and **hypersaline environments** are highly abundant and of great ecologic value. Moreover, salinity is a major determinant for microbial community composition. Although the adaptive value and molecular basis for protein halophilicity have attracted significant attention, certain biological aspects of haloadaptation still lack a proper mechanistic model and the molecular basis involved remains unclear.

For example, electrostatic interactions play a key role in **macromolecule intercommunication**, especially in DNA – protein interactions, where a high salt concentration can prevent binding (Wojtas, Moggi et al. 2012). However, adaptation to high salt concentrations of DNA-binding proteins has not yet been studied in detail. Electrostatics also underlie the triggering mechanism for **viral infection** and, although halophilic viruses have already been described (Pietila, Laurinmaki et al. 2013), it is still unclear how they adapt to such novel conditions and the dependence upon salinity of their infective activity. Studies of the adaptation to high salt concentrations of such biomolecular processes may reveal novel biologically relevant mechanisms.

Understanding the mechanisms of haloadaptation, and particularly the role of evolutionary selected amino acid residues, remains a controversial and interesting issue in modern molecular biology. Elucidation of the underlying mechanisms would also shed some new light in fundamental aspects of **protein thermodynamics, stability** and the nature of the influence of **solvent composition** in the properties of biological macromolecules. Some particular issues that could also benefit from these studies would be **Hofmeister series**, the role of **hydrophobic solvation, electrostatic interactions**, etc.

From an **evolutionary** perspective, there are at least two aspects that could potentially benefit from the study of the mechanisms underlying halophilic adaptation. First, the strong amino acid selection in extremely halophilic organisms largely coincides with the minimal amino acid set for foldable prebiotic polypeptides (Longo, Lee et al. 2013), suggesting that a halophilic environment may have been required in early abiogenic events. The fact that life may have been originated in a halophilic-like environment is a suggestive and intriguing concept of crucial conceptual significance. Second, the high degree of convergent evolution is surprising among such unrelated species, thus suggesting that this particularly biased amino acid composition is based on fundamental biophysical grounds.

The fields of **biotechnology and industrial enzymatic processes** could also benefit from the use of halophilic enzymes and halophilic organisms. Halophilic enzymes are particularly attractive because they are active in environments with reduced water activity. Salt, as organic solvents, lowers water activity. For instance, the activity of an aqueous saturated

NaCl solution ($a_w = 0.75$) is equivalent to a 60% solution of N,N-dimethyl formamide (DasSarma, Coker et al. 2010). Organic solvents are often used in industrial applications to enhance the solubility of products or reactives, or to promote enantiomer specificity. Unfortunately, this comes at the cost of a severe reduction in the activity or the stability of most enzymes. However, lowered water activity is ideal for halophilic enzymes, which show optimal performance under similar conditions. Reports have shown that halophilic enzymes display substantial levels of activity in organic media, potentially making them an ideal choice for **biocatalysts** in the presence of ethylene glycol, glycerol and dimethyl sulfoxide (Lanyi and J. 1969). Understanding the molecular basis of salt adaptation would allow tuning the halophilic character of a particular enzyme according to the specific requirements of the process.

Of particular interest for detergent and food industry would be halophilic proteases, lipases, and amylases. In addition, other **industrial processes** requiring pure water could be adapted to the use of salted seawater, with is a major advantage from an environmental point of view. Glycosyl hydrolases, for instance, break down cellulose, which is important for biomass processing and biofuel production. Genetic engineering into crop plants of salt resistance or mechanisms developed in halophilic organisms, enabling the use of seawater for irrigation, could truly revolutionize **agriculture** and dramatically increase the area of arable land (DasSarma, Coker et al. 2010). Those are just a few examples of hypothetical technologies that could benefit from a deep understanding of the mechanisms underlying salt adaptation, both at the protein and at the cellular level.

2. OBJECTIVES

*"¡Atreveos!
El progreso solamente se logra así"*

Víctor Hugo

2. OBJECTIVES AND WORKING HYPOTHESIS

We have approached halophilic adaptation largely from a thermodynamic perspective, focusing on protein stability in environments with high salt concentration. Our main goals are:

- ✓ Unveiling the determinants for the salt stabilization of halophilic proteins.
- ✓ Understanding the general mechanisms behind salt stabilization of proteins.
- ✓ Successfully explaining the biased amino acid selection found in halophilic proteins.
- ✓ Defining the functional role of acidic residues and their electrostatic interactions with cations in the stabilization of halophilic proteins.
- ✓ Establishing a model compatible with the linear dependence of stability on salt concentration.
- ✓ Explaining the selective accumulation of K^+ inside the cellular cytoplasm.

Following previous research in our lab, our approach is grounded on the evidence that the reduction of solvent-exposed area, rather than the increase in negative charge, is the key determinant for halophilic protein stabilization and, hence, haloadaptation. Based on that hypothesis we expect to:

- ✓ Establish a mechanistic link between the reduction in solvent-exposed hydrophobic area and the increased stabilization (halophilicity) of proteins.
- ✓ Find natural halophilic systems and design new ones from mesophilic homologues, replicating *in vitro* the process of haloadaptation.
- ✓ Obtain biophysical and structural information by means of high resolution Nuclear Magnetic Resonance Spectroscopy.
- ✓ Establish a mechanistic model, based on fundamental protein–cosolute interaction principles, able to explain a range of experimental observations wider than previous models.
- ✓ Elucidate the relationship between Hofmeister series and protein haloadaptation.
- ✓ Provide a mechanistic unified theory to explain haloadaptation as the effects of salt on protein stability and enzymatic activity.
- ✓ Investigate the forces driving the unfolding of obligate halophilic proteins in the absence of salt.
- ✓ Underscore the implications of our findings in the evolutionary process leading to haloadaptation.
- ✓ Extend the range of validity of current haloadaptation theories to other fields: such as folding or organic osmolytes.

3. MATERIALS & METHODS

*"The blunders are all there in the board,
waiting to be made"*

Ksawery Tartakower

3. MATERIALS AND METHODS

3.1. PROTEINS UNDER STUDY

A number of natural and designed proteins were employed as halophilic and non-halophilic case studies. Wild-type Protein L (*WT ProtL*) was employed as a natural mesophilic reference system for a folded protein. A number of lysine to glutamate mutations were introduced in order to obtain designed folded halophilic systems (*KxnE ProtL*). *Oceanobacillus iheyensis* β -lactamase (β LAC), on the other hand, was chosen as a naturally occurring halophilic protein and it serves as a model for a halotolerant protein. Intrinsically disordered N-terminal domain of the human c-SRC kinase, uSRC, was used as a model system to investigate the mesophilic unfolded state. Rationally designed mutations were introduced in order to make its amino acid sequence similar to the average halophilic amino acid composition. The resulting peptide haloSRC constitutes our model for unfolded halophilic conformation. Finally, a naturally occurring obligate halophile was chosen for the study of the complex salt-modulation of activity: NAD⁺-dependent DNA ligase from *Haloferax volcanii* (*Hv LigN*). The amino acid compositions can be found in table M1, and the corresponding sequences in the Appendix.

	WT ProtL	KxnE ProtL	βLAC	uSRC	haloSRC	Hv LigN
<i>Ala</i>	9	9	21	17	17	89
<i>Arg</i>	0	0	13	5	4 (-1)	62
<i>Asn</i>	4	4	9	3	3	18
<i>Asp</i>	4	4	28	3	6 (+3)	79
<i>Gln</i>	1	1	5	4	4	10
<i>Glu</i>	6	6 + <i>n</i>	34	4	6 (+2)	70
<i>Gly</i>	5	5	20	11	11	58
<i>His</i>	0	0	1	3	3	16
<i>Ile</i>	3	3	20	0	0	20
<i>Leu</i>	3	3	23	3	2 (-1)	60
<i>Lys</i>	7	7 - <i>n</i>	10	6	2 (-4)	11
<i>Met</i>	1	1	7	1	1	6
<i>Phe</i>	4	4	3	5	4 (-1)	26
<i>Pro</i>	0	0	7	11	10 (-1)	38
<i>Ser</i>	2	2	10	13	14 (+1)	38
<i>Thr</i>	8	8	24	3	5 (+2)	34
<i>Tyr</i>	3	3	10	0	0	15
<i>Val</i>	3	3	14	2	2	77
<i>Trp</i>	1	1	2	1	1	4
<i>Cys</i>	0	0	0	0	0	6
Total	64	64	261	95	95	737

Table M1. Amino acid compositions for all the proteins under consideration.

3.1.1. Naturally occurring mesophilic protein L *ProtL*

The immunoglobulin G binding domain B1 of Protein L from the anaerobic bacteria *Peptoespectrococcus magnus* is a cellular wall protein that binds immunoglobulin light chains. Altogether with four other homologous repeats (B2-B5), this domain in the N-terminal part of the protein conforms its immunoglobulin G binding activity. The 64 residues peptide is composed of a α -helix packed against a four-stranded β -sheet (Figure M1) (O'Neill, Kim et al. 2001). *ProtL* shows a moderate stability ($7 \text{ kcal}\cdot\text{mol}^{-1}$ at 298 K), cooperative unfolding and a two-states reversible thermal and chemical unfolding equilibrium, without any detectable intermediates (Yi and Baker 1996). As a result, it has extensively been used as a model for folding and stability studies, as well as an appropriate system for NMR since it is highly soluble, stable in the long term and has a well defined and disperse spectrum (Yi, Scalley et al. 1997). The variant traditionally used differs slightly from the original sequence: tyrosine 47 has been substituted by a tryptophan in order to increase the fluorescent signal for an easier monitoring of the purification and unfolding processes (O'Neill, Kim et al. 2001). It is used as a model for a **natural mesophilic system**. It has also been used as the starting frame for the introduction of lysine to glutamate substitutions (K_xnE , being n the total number of substitutions) in order to design artificial **mutants with progressively increasing halophilic character** (Tadeo, López-Méndez et al. 2009b). For instance, K_x5E to K_x7E *ProtL* mutants have nearly the same density of carboxylate groups as ferredoxin from *Haloarcula marismortui* (128 residues).

The gene coding for *ProtL* was previously cloned into the ampicillin resistant pET16b vector between the restriction sites *NcoI* and *BamHI*. Detailed description of the procedures and the primers employed for site-directed mutagenesis can be found in the work from Tadeo *et al.* (Tadeo, López-Méndez et al. 2009b).

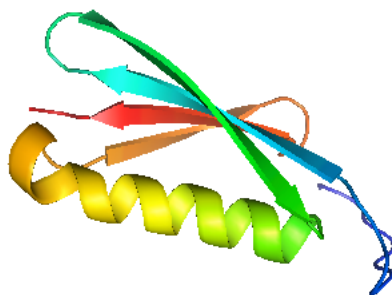


Figure M1. ProtL: Immunoglobulin G binding domain of Protein L from *Peptostreptococcus magnus*.

ProtL	Substitutions
Kx1E	42
Kx2E	42, 54
Kx3E	28, 42, 54
Kx4E	28, 42, 54, 61
Kx5E	23, 28, 42, 54, 61
Kx6E	23, 28, 41, 42, 54, 61
Kx7E	7, 23, 28, 41, 42, 54, 61

Table M2. Lysine to glutamate substitutions introduced to WT ProtL (amino acid sequence: MEEVTIKANLIFANGSTQTAEFKGTFEKATSEAYAYADTLKKDNGEYTVDVADKGYTLNIKFAG). K_xnE involves n substitutions from K to E.

3.1.2. Halotolerant beta-lactamase from *Oceanobacillus iheyensis* β LAC

An extracellular β -lactamase from *Oceanobacillus iheyensis*, OIH-1, was chosen as a model system for a naturally occurring, folded and stable halophilic protein. *Oceanobacillus iheyensis*, whose habitat is the sediment at 1050 m depth in the Pacific Ocean off the coast of Japan, is a halotolerant alkaliphilic bacterium that can thrive in mesophilic habitats but prefers to grow in salt concentrations up to 3.6 M NaCl. It secretes an extracellular β -Lactamase OIH-1, ultimately responsible for antibiotic resistance, which constitutes a major evolutionary asset. OIH-1 is an **extremely halotolerant enzyme** capable of hydrolyzing β -lactames even at saturating NaCl concentration. The amino acid composition is typical of halophilic proteins with an acidic content of almost 25% (Table M1), whereas the structure is highly conserved when compared to other mesophilic homologues (Figure M2) (Toth, Smith et al. 2010).

A plasmid containing the clone for β -lactamase was kindly provided by Professor Sergei Vakulenko, from the University of Notre Dame (Indiana, US). Quality tests on the construct suggested that the 14 C-terminal amino acids reduced the protein's long-term stability. The protein contains leader sequences, short polypeptides that target the protein for transport outside the cytoplasm. These signaling sequences are usually small and disordered so they can trigger unfortunate processes such as unfolding, aggregation and proteolysis. Thus, we amplified and cloned a shorter version of the gene without these residues into the ampiciline resistant pET16b overexpression plasmid between the restriction sites NcoI (primer: 5' – ACACCATGGGCAGTGAAGATCTGAAA – 3') and BamHI (primer: 5' – GGTGGATCCTTAATTACGCAGTTCCTGTAACAC – 3'). The resulting protein β LAC shows long-term stability, a well dispersed ^1H , ^{15}N -HSQC spectrum and reversible thermal (but not chemical) unfolding equilibrium.

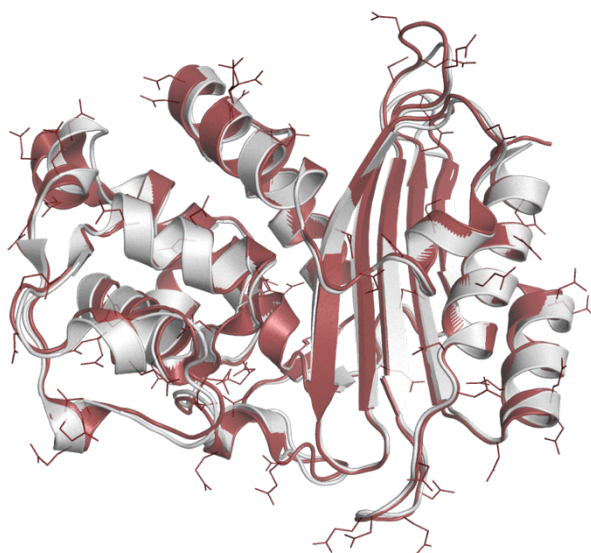


Figure M2. Structural overlay of β -lactamases from halotolerant *Oceanobacillus iheyensis* OIH-1 (PDB: 3LEZ, in red) and mesophilic *Bacillus licheniformis* (PDB: 1NRF, in white). Side chains of aspartic and glutamic acids in the surface of the halophilic homologue are shown as red lines.

3.1.3. Unstructured N-terminal domain from human Src kinase *uSRC*

uSRC corresponds to the SH4 and *unique* domains of human c-Src kinase, which belongs to the Src family of non-receptor tyrosine kinases (SFK). SFKs are key players in the initiation of different signal transduction pathways, ultimately involved in cell cycle regulation. They all comprise six functional domains: SH4, unique domain, SH3, SH2, the catalytic kinase domain and the C-terminal regulatory region. The N-terminal domain, composed of SH4 and unique domains, has been related to subcellular localization and trafficking, membrane attachment and protein-protein interactions contributing to substrate specificity. It also exhibits sites for phosphorylation and other covalent modifications. They are thought to modulate the enzyme activity but their biological role remains unclear (Perez, Gairi et al. 2009).

The 84 residues long *uSRC* (Figure M3), composed of the SH4 and unique domain, is largely disordered but shows transient secondary structural elements. It has been chosen here as a model of the **unfolded state of a non-halophilic protein**. The clone was kindly provided by Dr. Miquel Pons from the University of Barcelona (Perez, Gairi et al. 2009).

3.1.4. Design of the halophilic unstructured homologue *haloSRC*

In order to obtain a model for a **halophilic unfolded state**, *uSRC* was used as a scaffold and its amino acid composition modified in order to match that of a hypothetical halophilic counterpart. Figure I2 shows the percentages of the relative abundance (in percentage) for halophilic sequences as compared to mesophilic sequences for each amino acid type, Δaa_{h-m} (Paul, Bag et al. 2008). The total number of residues in the halophilic sequence is calculated as follows:

$$n_h^{aa} = n_m^{aa} \cdot \left(1 + \frac{\Delta aa_{h-m}}{100} \right) ;$$

Equation M1. Design of a halophilic sequence composition. Formula used for the calculation of the number of residues of a hypothetical halophilic sequence homologous to a given mesophilic sequence.

where n_h^{aa} and n_m^{aa} are the number of residues of a given amino acid type *aa* in a halophilic or mesophilic sequence respectively. When the relative abundance coefficients are applied to *uSRC* sequence, only 8 substitutions are required to obtain a halophilic homologue (Table M3). Mutations were introduced at random positions evenly distributed along the polypeptidic chain and designed to maximally preserve the hydrophobic/electrostatic nature of the mutated residues (Figure M3). The resulting polypeptide *haloSRC* (BioMagResBank access code 25196) is also disordered and exhibits residual secondary structure (see the assigned ^1H , ^{15}N -HSQC spectrum in Figure R14).

The gene for *haloSrc* was synthesized by *euofins MWG Operon*® and cloned into a pET16b expression vector between the restriction sites NcoI and BamHI. A streptavidin affinity tag was also added in C-terminal position for purification purposes.

aa	Δaa_{halo} (%)	<i>uSRC</i>	Δaa	<i>haloSRC</i>
Ala	+ 1	17		17
Arg	- 10	5	- 1	4
Asn	- 6	3		3
Asp	+ 53	3	+ 3	6
Gln	- 2	4		4
Glu	+ 21	4	+ 2	6
Gly	0	11		11
His	+ 4	3		3
Ile	- 13	0		0
Leu	- 17	3	- 1	2
Lys	- 50	6	- 4	2
Met	- 22	1		1
Phe	- 11	5	- 1	4
Pro	- 8	11	- 1	10
Ser	+ 4	13	+ 1	14
Thr	+ 40	3	+ 2	5
Tyr	- 3	0		0
Val	+ 8	2		2
Trp	- 16	1		1
Cys	- 25	0		0

Table M3. Amino acidic content in *uSRC* and *haloSRC*. Relative abundance coefficients are applied to the mesophilic sequence *uSRC* in order to obtain the amino acid composition of a putative homologous halophilic sequence *haloSRC*.

uSRC

MASN**K**SKP**K**DASQRRRSLE**P**AENVH
 1 25
 GAGGG**A**F**P**ASQTP**S**K**P**ASADGH**R**G**P**
 26 50
 SAAFAPAAA**E**P**K**LFGGFNSSDTVTS
 51 75
 PQRAG**P**L**A**GG SAWSH**P**Q**F**E**K**
 76 95

Strep-tag

haloSRC

MAS**N**DSK**P**S**D**ASQRRRSLE**T**AENVH
 1 25
 GAGGG**A**T**P**ASQTP**S**E**P**ASADGH**D**G**P**
 26 50
 SAAFAPAAA**E**P**D**LFGGFNSSDTVTS
 51 75
 PQRAG**P**E**A**GG SAWSH**P**Q**F**E**K**
 76 95

Strep-tag

Figure M3. *uSRC* versus *haloSRC* sequence. Primary sequence of the mesophilic (*uSRC*) and halophilic (*haloSRC*) intrinsically disordered proteins. Substitutions introduced to convert *uSRC* into *haloSRC* are highlighted in red and green respectively. Lipid binding regions are shaded.

3.1.5. NAD⁺-dependent DNA Ligase from *Haloferox volcanii* LigN

DNA ligases play key roles in all forms of cellular life, e.g. joining of DNA molecules or Okazaki fragments during DNA repair processes and DNA replication respectively. These enzymes interact with a nucleotide cofactor to form a covalent enzyme-nucleotide monophosphate moiety. Particularly, DNA ligases utilize either NAD⁺ (nicotinamide adenine dinucleotide) or ATP (adenosine triphosphate) as the nucleotide cofactor. The mechanism of DNA ligation is similar for both types of enzymes. In a first step, nucleophilic attack by the enzyme on the α -phosphorous of either NAD⁺ or ATP results in the formation of an enzyme-adenylate intermediate (a lysine residue covalently bound to adenosine monophosphate AMP) and the release of NMN (nicotinamide mononucleotide) or pyrophosphate respectively. In a second step the AMP moiety is transferred to the 5' end of the nicked DNA strand to form a DNA-adenylate complex. Finally, attack by the 3'-OH group of the nicked DNA strand on the DNA-adenylate, catalyzed by the enzyme, results in DNA strand joining and simultaneous AMP release.

There are two distinct DNA ligases in the extreme halophilic archaea *Haloferox volcanii*: NAD⁺-dependent ligase N (*Hv* LigN) and ATP-dependent ligase A. *Hv* LigN is a 699 amino acid protein (75.7 kDa). It shares structure and sequence homology with bacterial NAD⁺ dependent ligases, typically displaying a toroidal form and a highly modular architecture. Four different domains form *Hv* LigN (Figure M4): Oligomer Binding domain (OB, gray), Zinc finger (yellow), helix-hairpin-helix (HhH, green), BRCT (purple) and the adenylation domain. The latter is composed of two subdomains, 1A (red) and 1B (blue), linked by a long flexible loop.

Professor Stuart MacNeill from the University of Copenhagen kindly provided us with the *Hv* LigN clone. The gene is hosted into a pET47b(+) kanamycin resistant expression vector between the restriction sites *Sal*I and *Bam*HI. Resulting plasmid encodes a polyhisitidine tag for purification purposes, which can be removed with HRV 3C protease.

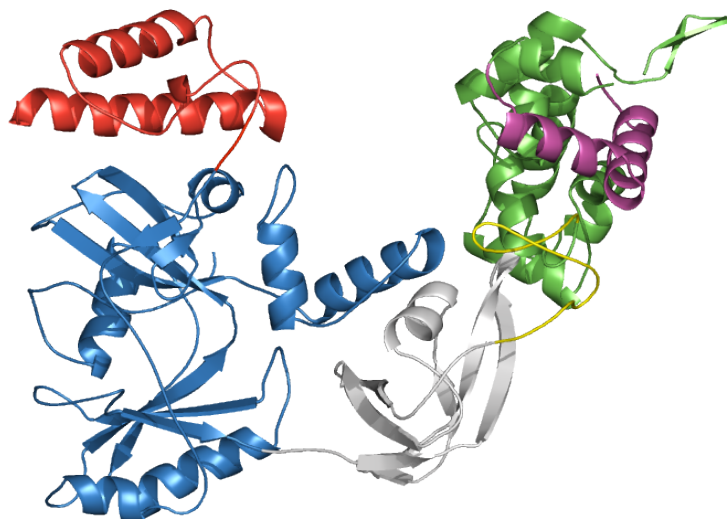


Figure M4. Structural homology model for *Hv* LigN enzyme. The enzyme shows a highly modular architecture: OB (gray), zinc finger (yellow), HhH (green) and part of the disordered BRCT (purple) domains are shown. Catalytic adenylation domain is divided in two subdomains: 1A (red) and 1B (blue).

3.1.5.1. Adenylation and deadenylation assays

Adenylation domain is essential for cofactor binding and therefore for catalytic activity. Subdomain rearrangement and closure of subdomain 1A over domain 1B is required in order to conform the NAD^+ binding site and the adenylation of the enzyme. Later opening of subdomain 1A allows the release of NMN, the entrance of a new molecule of NAD^+ and the restart of the catalytic cycle. Therefore, open – closure dynamic of the adenylation domain is involved in alternatively hosting the NMN, AMP and NAD^+ molecules along the ligation reaction (Figure M5).

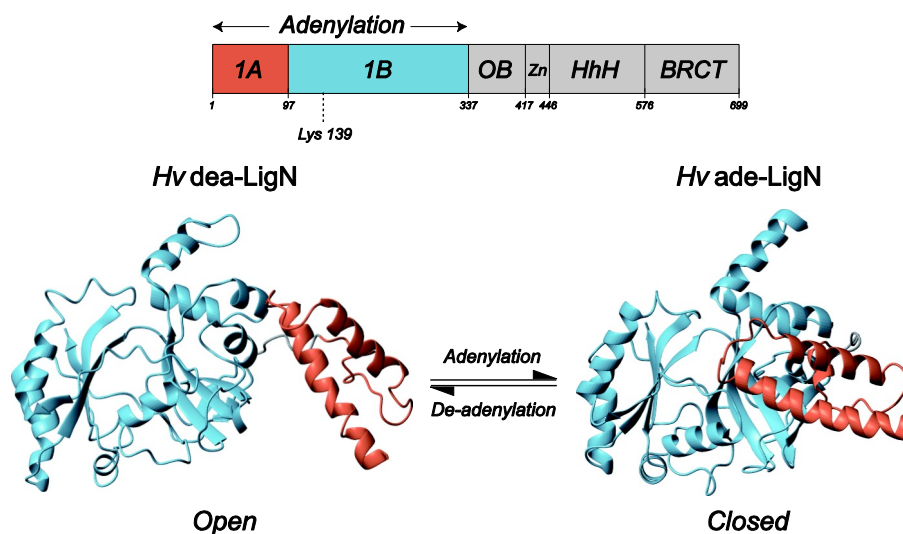


Figure M5. Adenylation dynamics. Adenylation domain of *Enterococcus faecalis* ligase represents the open and closed forms along the reaction coordinates. Domain 1B is shown in blue and domain 1A in red. Both structures share the same orientation of domain 1B. NAD^+ binding causes the domain 1A to close, represented by the protein structure on the right. Deadenylation of the adenylation domain with excess NMN results in the open NMN-bound conformation shown here on the left, leaving the adenylation lysine exposed to solvent and the nicked strand of DNA.

Active close-adenylated (*Hv ade-LigN*) and inactive open-deadenylated (*Hv dea-LigN*) conformations of *Hv LigN* can be isolated. Adenylation and deadenylation assays were adapted from previously reported protocols for NAD^+ dependent proteins from other organisms (Timson and Wigley 1999, Poidevin and MacNeill 2006):

Adenylation was achieved by addition of 2.5 mM NAD^+ to a solution of approximately 30 μM of protein (80-fold excess) and subsequent incubation of the mixture for 20 h at 25 °C.

For **deadenylation** assay, again a 80-fold excess of NMN and MgCl_2 was added to the protein and the mixture was incubated at 25 °C during at least 75 h.

In both cases, assays were followed by gel filtration chromatography in a Superdex75 column (Amersham), the elution buffer being 2 M KCl, 20 mM sodium phosphate (pH = 8).

3.1.5.2. Activity assays: DNA Ligation

Hv LigN selectively joins fragments I (8454 bp) and IV (5686 bp) of *Bst*EII-digested λ DNA to produce a new oligomer that runs in an agarose gel at a molecular weight equal to the sum of the two fragments (Figure M6). This reaction has been used to evaluate the relative enzymatic activity upon addition of different co-solute concentrations with respect to that at the reported *in vivo* concentration 3.2 M KCl. DNA ligase assays were performed as previously described, but with minor modifications (Poidevin and MacNeill 2006). Reactions (20 μ L) contained 1.25 μ g of *Hv*LigN in 20 mM Tris-HCl (pH 7.6 at 25 °C), 25 mM potassium acetate, 10 mM magnesium acetate, 10 mM DTT, 1 mM NAD⁺, 0.01 % Triton X100 and 3 μ l of 500 μ g/ml of *Bst*EII (restriction endonuclease)-digested λ DNA (New England Biolabs) and the tested concentration of cosolute. Ligation reactions were run at 45 °C for 10 minutes and were stopped by addition of 4 μ L of 50 mM EDTA.

Samples were then desalted using the MinElute reaction clean-up kit (Qiagen), eluted in 20 μ l of EB buffer (10 mM Tris pH 8.0), heated to 65 °C for 5 min to melt the super-coiled DNA and chilled on ice for another 1 min prior to loading 10 μ L of the reaction product + 10 μ L of loading buffer onto a 1% agarose gel prepared in TAE buffer (0.4 M Tris pH 8.0 – 0.2 M Acetate – 10 mM EDTA). The gel contained Sybr® Safe DNA gel stain (Invitrogen # S33102) and was run submerged in 1x TAE at 120 V for 120 minutes. Gels were visualized in Chemidoc™ XRS System (BioRad) and band intensities were determined by using ImageJ 1.41o analysis software (Wayne Rasband, National Institutes of Health, USA). Activity values were calculated by measuring the net intensity of the product band (fragment I+IV) relative to non-substrate fragments (bands II + III or bands V + VI) (Figure M6).

The assay was tested in the presence of sodium and potassium chloride, fluoride, bromide, sulphate, acetate, phosphate, carbonate, nitrate and perchlorate. In the case of phosphate salts, buffer was incubated at 75 °C for 10 min prior to running the ligase activity test, due to the low solubility of magnesium phosphate.

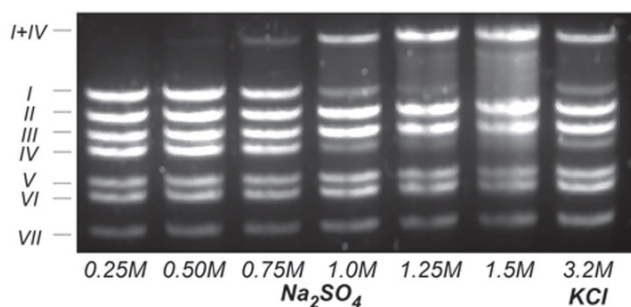


Figure M6. Example of an assay of ligation activity at increasing concentrations of Na₂SO₄. Increase in ligase activity is clearly observed by the progressive disappearance of substrate bands I and IV and appearance of a band of higher molecular weight equal to the sum of both. 3.2 M KCl control is also shown.

3.2. SAMPLE PREPARATION

Protein samples were prepared by transformation of the plasmid with the expression vector into *Escherichia coli* BL-21 DE(3), induction of protein overexpression with IPTG (*isopropyl- β -D-thiogalactopyranoside*) and posterior purification by affinity, size-exclusion and/or anionic exchange chromatography, depending on the system. Proteins for biophysical and enzymatic activity tests were grown in regular LB (Lysogeny Broth, Pronadisa #1231.00) media, whereas protein samples required for NMR experiments included an specific isotopic labelling scheme.

3.2.1. Culture growth and isotopic labelling

Plasmids carrying the gene coding for the protein of interest were transformed into *E. coli* BL-21 DE(3) expression strain cells following the transformation protocol adapted from Stratagene QuickChange Mutagenesis Kit[®]. Cells were plated in petri plates with 50 mg/mL of the corresponding antibiotic (Table M4) and incubated over night at 37 °C. Cells were then grown at 37 °C for 16 h in small preinocules of 100mL of LB media for unlabelled proteins or M9 minimal media for isotopic labelled proteins. Cells were then separated by centrifugation and resuspended into 1–1.5L of culture to grow at 37 °C. Once cells reached optical density of 0.45–0.60 for M9 cultures or 0.60–0.80 for LB cultures, protein overexpression was induced with 1 mM IPTG. After several expression tests, optimal induction conditions were found to be 4 h at 37 °C for ProtL, uSRC and haloSRC and 16 to 20 h at 20–25 °C for *Hv* LigN and β LAC (Table M4). Cells were then separated by centrifugation and resuspended in 30 mL of resuspension buffer per litre of culture (Table M4). Resuspended cells can be stored at -20 °C prior to purification process.

Protein	Plasmid	Induction conditions	Resuspension Buffer	Yield (mg/L of culture)
ProtL	pET16b (Amp) NcoI – BamHI	4 h @ 37 °C	20mM PO _x (pH 6.0)	70
β Lac	pET16b (Amp) NcoI – BamHI	20 h @ 25 °C	20mM PO _x (pH 7.5) 0.5 M NaCl	30
<i>Hv</i> LigN	pET47b (Kan) SanDI – BamHI	20 h @ 25 °C	20mM PO _x (pH 8.0) 2M KCl	120
uSRC	pET14b (Kan) NcoI – BamHI	4 h @ 37 °C	50mM PO _x (pH 7.0) 0.2mM PMSF & EDTA	10
haloSRC	pET16b (Amp) NcoI – BamHI	4 h @ 37 °C	100mM tris/HCl (pH 8.0) 150mM NaCl 1mM PMSF & EDTA	20

Table M4. Summary of protein expression. Plasmid, antibiotic resistance, restriction sites, resuspension buffer and purification yield in mg / L of cell culture for each one of the proteins used.

Protein samples prepared for NMR experiments require specific isotopic labelling: typically ^{15}N , ^{13}C and sometimes ^2H . M9 minimal media is a restrictive media in which the sources of nutrients are tightly controlled (Table M5). Ammonium chloride and glucose were used as the sole sources of nitrogen and carbon for protein synthesis. As a result, the addition of $^{15}\text{NH}_4\text{Cl}$ or ^{13}C -glucose ensures that all nitrogen or carbon atoms in the protein are isotopically labelled.

Reactive	Quantity	Company
Na_2HPO_4	6 g/L	Sigma # S3264
K_2HPO_4	3 g/L	Sigma # P9791
NaCl	0.5 g/L	Sigma # S3014
$^{15}\text{NH}_4\text{Cl}$	1 g/L	Cambridge Isotopes Labs.
^{13}C -glucose	2-3 g/L	Cambridge Isotopes Labs.
MgSO_4	1 mM	Sigma # M7506
CaCl_2	0.1 mM	Sigma # C1016
Biotin	1 mg	Sigma # T4639
Thiamine hydrochloride	1 mg	Sigma # T4625
Antibiotic	0.5 mM	Sigma # A9393
H_2O (D_2O)	up to 1L	(Cambridge Isotopes Labs.)

Table M5. Composition of M9 minimal media. ^{15}N -labelled ammonium chloride, ^{13}C -labelled glucose or deuterated water can be added depending on the isotopic labeling required

3.2.2. Protein purification

Extraction of pure protein from the cells was usually done by cellular lysis and different chromatographic techniques, which depended on the biophysical properties of the protein and the affinity tags incorporated in the gene construct. Purity and quantity of the purified proteins was assessed by UV absorption spectroscopy, circular dichroism and SDS-PAGE electrophoresis.

In a first step, 1:100 diluted Protease Inhibitor Cocktail was added (*PIC* Sigma # P2714) after pellet resuspension. It is a mixture of chemicals whose principal component is phenylmethylsulfonyl fluoride (PMSF). The sample must be kept at cold temperature on an ice bucket at all times to prevent proteolysis. Pressure induced cell lysis was performed in a cell disruptor Constant Systems TS 1.1KW at 27 kpsi, followed by ultracentrifugation in a Beckman Ultracentrifuge at > 30-45k rpm for 30-45 min at 4 °C. Pellet was immediately discarded, whereas supernatant contained the protein. Further steps are protein specific.

3.2.2.1. ProtL

Sample was heated to 75 °C for 5 min and agitated at room temperature for 30 min. wt and ProtL mutants exhibit a thermal reversible unfolding, so the thermal shock precipitates all non-reversible proteins while keeping ProtL soluble. Next, centrifugation at 30k rpm for 30 min separated insoluble aggregates from the soluble protein supernatant. Supernatant was loaded to a Superdex75[®] (GE Healthcare) size exclusion column and eluted under isocratic conditions (Figure M7). Loading and elution buffer was 20 mM sodium phosphate, pH 6.0. An additional purification step was usually required, especially for halophilic KxnE mutants, which tend to bind DNA and oligonucleotides. Anion exchange chromatography separates molecules based on their net charge. Samples from size-exclusion chromatography were loaded onto a cationic resin (Q Sepharose Fast Flow[®] from GE Healthcare[®]; quaternary ammonium strong ion-exchange groups attached to a polymer bed of crosslinked 6 % agarose beds) and eluted using a BioLogic LP[®] low-pressure chromatography system (Bio-Rad[®]). ProtL was bound to the resin and eluted using a gradient from 0 to 1 M NaCl. WT ProtL was displaced around 100 mM NaCl (Figure M8), whereas the other mutants eluted at higher salt concentrations (Table M6). Kx5E ProtL, for instance, eluted at > 0.5 M NaCl.

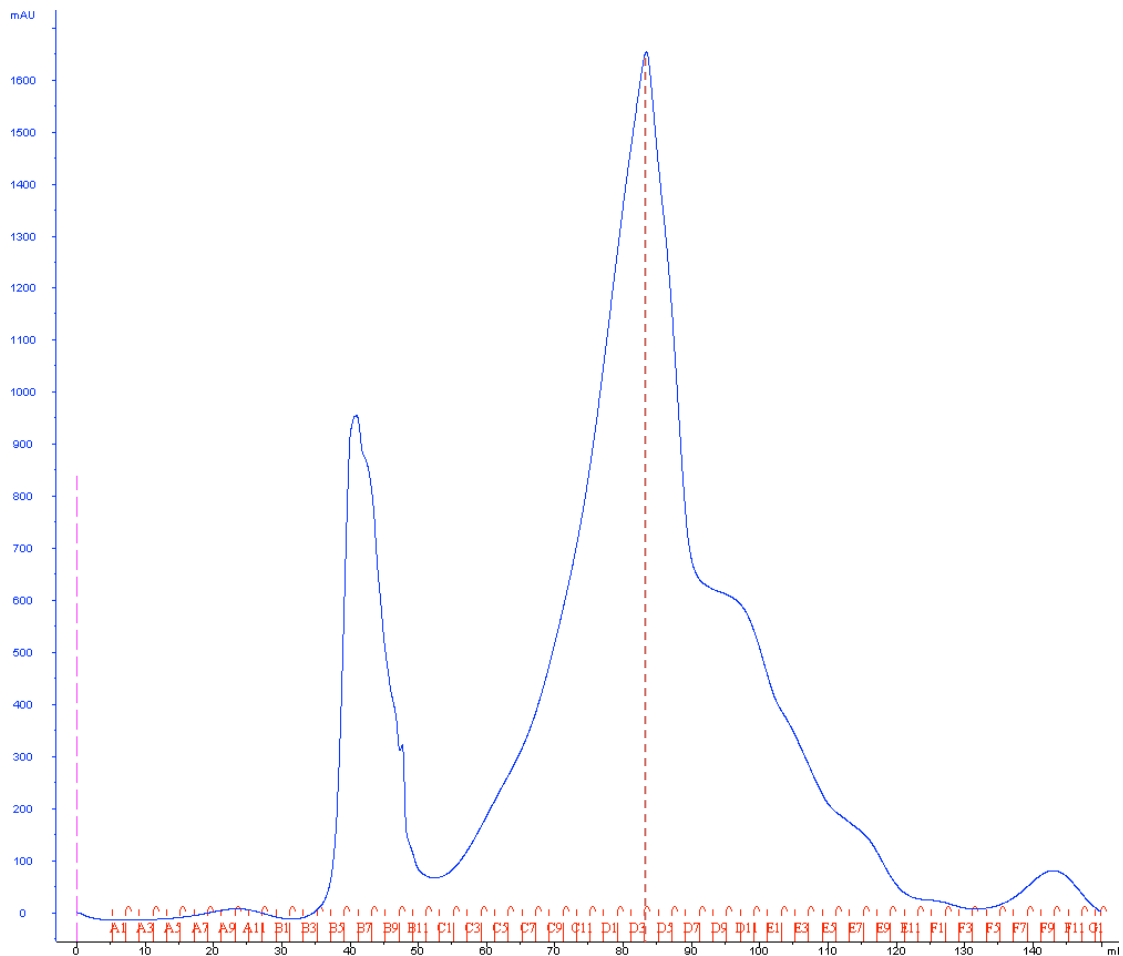


Figure M7. Size-exclusion chromatography. Elution profile of wt ProtL in Superdex75[®] after thermal shock at 75 °C. Blue line corresponds to absorbance in arbitrary units (vertical axis). Horizontal axis shows the elution volume and the elution fraction, the red dashed line marks the protein elution peak.

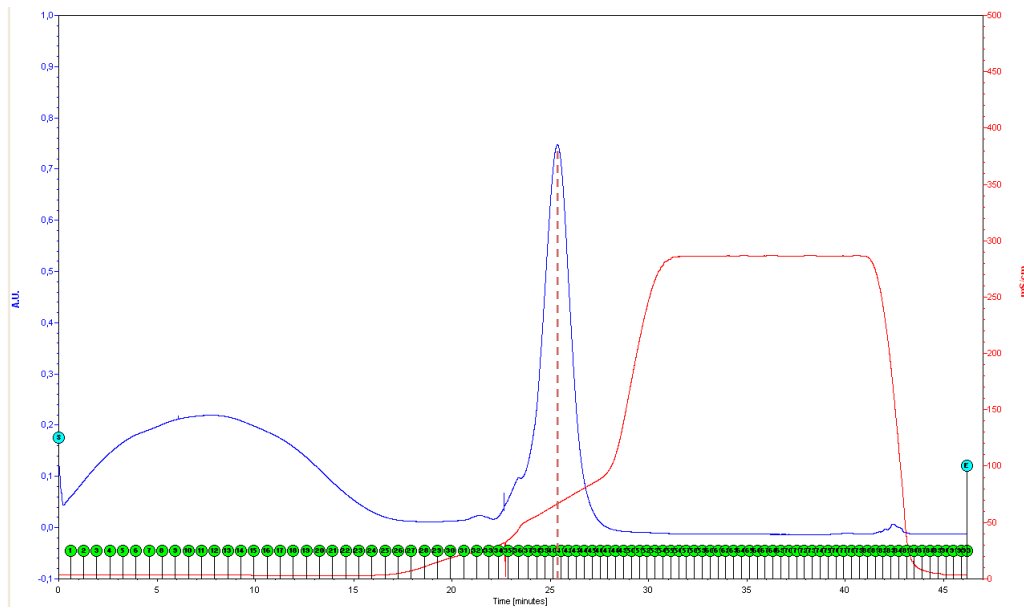


Figure M8. Anion-exchange chromatography. Elution profile of wt ProtL in Q Sepharose Fast Flow[®] anion exchange chromatography after thermal shock at 75 °C and size exclusion. The protein elutes as a sharp peak between 50 and 100 mM NaCl marked by the red dashed line.

ProtL	pI	Charge @ pH 7.0	~ [NaCl] eluted (mM)
wild type	4,77	-3,1	< 100
Kx5Q	3,84	-8,1	350
Kx5E	3,65	-13,1	500
Kx6E	3,37	-15,1	550

Table M6. Anion-exchange versus negative charge. Salt concentration of elution at which the different variants of ProtL elute in anion exchange chromatography correlates with total negative charge.

3.2.2.2. β LAC

The β LAC gene construct does not include an affinity tag. After resuspension in 20 mM sodium phosphate (pH 7.5) 0.5 M NaCl, cellular lysis and centrifugation, supernatant was diluted in phosphate buffer (pH 7.5) to a final concentration of 250 mM NaCl. The solution was then loaded onto a Q Sepharose Fast Flow[®] resin in a BioLogic LP[®] chromatography system for anionic exchange. Elution with a gradient between 0.1 and 1 M KCl released protein at approximately 500 mM KCl, as expected for a halophilic protein. Finally, fractions were immediately loaded onto a Superdex 75[®] column for gel filtration chromatography under isocratic regime with 20 mM sodium phosphate (pH 7.5) and 0.5 M KCl.

3.2.2.3. *Hv LigN*

Hv LigN incorporates a hexa-histidine tag. Immediately after cell lysis, supernatant was incubated with nickel-nitrilotriacetic acid 6% agarose cross-linked beds (Ni-NTA[®] Qiagen[®] # 30230) on a disposable column for 90-120 min at 4 °C with gentle shaking. The histidine tagged protein selectively binds to the nickel ions of the resin. Flow trough was discarded and the unspecific bound substances were washed out with 20 mM sodium phosphate (pH 8.0), 2 M KCl buffer with 5 mM imidazole. Protein was eluted with buffer with 500 mM imidazole and immediately loaded onto a Superdex 75[®] column and eluted under isocratic conditions with 20 mM sodium phosphate (pH 8.0), 2 M KCl.

3.2.2.4. *uSRC* & *haloSRC*

Both *uSRC* and *haloSRC* have a Strep-tag for streptavidin affinity. Purification protocol is therefore the same for the two proteins. Expression yield, however, differs due to the presence of human codons in the sequence of *uSRC*, which were removed in *haloSRC*. Cell lysate was loaded onto a Strep-Tactin[®] Superflow[®] H-PR cartridge (iba[®]). Flow-through was discarded and a wash step with 100 mM tris/HCl (pH 8.0), 150 mM NaCl, 1 mM EDTA, 1 mM PMSF buffer was applied to remove unspecific bound substances. The protein was eluted with 2.5 mM desthiobiotin buffer solution and immediately loaded on a Superdex 75[®] for gel filtration chromatography under isocratic regime with 50 mM sodium phosphate (pH 7.0), 0.2 mM PMSF, 0.2 mM EDTA.

3.3. THERMODYNAMIC STABILITY MEASUREMENTS

Assuming a two state equilibrium for folding, the reporter for the stability of folded proteins is the Gibbs free energy difference between the folded and the unfolded state chemical potentials $\Delta\mu^{U-F}$, which at constant pressure correspond to ΔG^{U-F} . Addition of denaturants and increase of temperature alter the balance between entropic and enthalpic contributions both in the protein and in the protein–water interactions, ultimately leading to the unfolding of the protein. For a reversible unfolding equilibrium, it is possible to calculate the stability of the protein ΔG^{U-F} from the relative populations of folded and unfolded state along the denaturant or temperature coordinate.

Circular dichroism (CD) along with fluorescence are the techniques most commonly used to monitor the chemical and thermal unfolding experiments and to report on the populations of the folded and unfolded state. CD originates from the differential absorption of left and right-handed circularly polarized light upon its interaction with optically active molecules. Unlike linearly polarized light, the electric field vector has a constant length, but rotates along the propagation axis forming a helix in space. If the helix is left-handed, the radiation is referred to as left-polarized and viceversa. Chiral structures interact asymmetrically with these two components of plane-polarized light, and the transmitted radiation possesses elliptical polarization that can be measured and is usually reported as *ellipticity* θ (Figure M9). Obviously, since light absorption is wavelength dependent, so is the difference in absorption, yielding the CD spectrum of the irradiated matter.

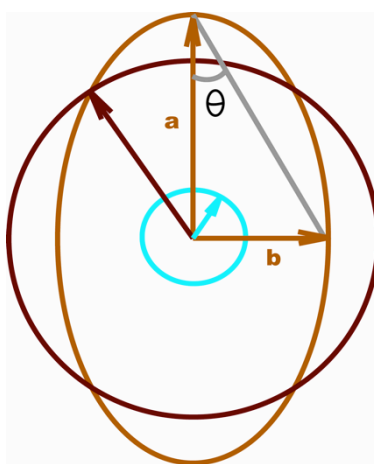


Figure M9. Elliptical polarization after the interaction with a chiral element. The major and minor axes of the resulting ellipse (orange) are a -the sum of the left (red) and right (blue) polarization- and b - the difference between them-. Circular dichroism signal is reported as ellipticity (θ), equal to $\arctan(b/a)$.

Not only chiral molecules differentially interact with plane-polarized light; also elements placed in a 3D chiral environment give rise to a CD signal. Secondary and tertiary structures of proteins induce asymmetry in the residues and bonds of the proteins. As a result, typical secondary structures such as α -helices and β -sheets, as well as random coil, exhibit characteristic CD spectra under 240 nm, which is the wavelength of absorption of the peptide bonds (Figure M10). Therefore, CD is a reference technique for obtaining qualitative structural information and to monitoring conformational changes induced by denaturing agents. Particularly, we use it to monitor the unfolded and folded state populations at increasing temperature or denaturant concentration. From those stability curves, global protein stability can be estimated.

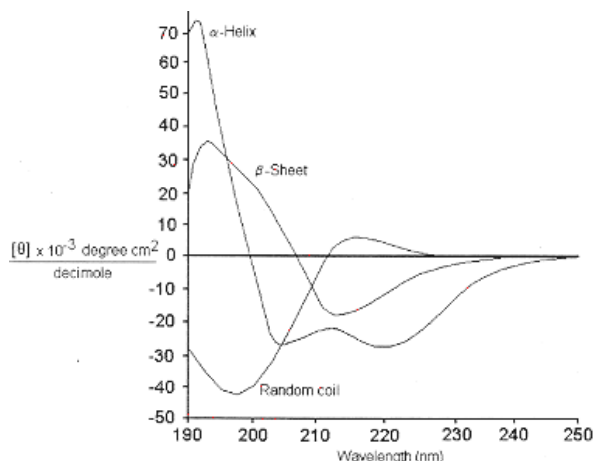


Figure M10. Far UV CD spectra associated with various types of secondary structure. α -helix and β -sheet spectra are shown together with random coil spectrum.

Calculation of global stability from chemical denaturation experiments is quite straightforward, however it requires a more complicated experimental set-up and large amounts of protein. Thermal denaturations, on the other hand, use much less protein amounts but, to extract stability values, complementary measurements are required to obtain the changes upon unfolding in heat capacity ΔC_p and enthalpy ΔH^0 . Previous works in the lab already reported those magnitudes for ProtL, so thermal denaturation is used in a regular basis (Tadeo, Pons et al. 2007). In the case of β LAC, chemical unfolding is not reversible, so thermal melts were also employed and the thermodynamic quantities were estimated from the literature. For *Hv* LigN, chemical denaturations with guanidinium chloride (GuHCl) are performed for both adenylated and deadenylated isoforms.

3.3.1. Chemical denaturation experiments

The effect of inorganic salts on the stability of *Hv* LigN was determined by measuring equilibrium denaturation curves with guanidinium chloride (GuHCl) of both enzyme isoforms *Hv* dea-LigN and *Hv* ade-LigN at increasing concentrations of co-solute. The following salts were considered: potassium chloride, sodium chloride, sodium sulfate and sodium phosphate.

Denaturing agents show a preferential binding to protein surface and/or backbone. As the unfolded state is more exposed to solvent than the folded state, the former is stabilized by denaturants (conversely, folded protein is destabilized). Typical denaturing agents are urea and GuHCl, but also other salts such as thiocyanates or perchlorates destabilize folded conformation of proteins. GuHCl was chosen for this experimental setup because it allows the recording of highly cooperative curves exhibiting a three states-model shape (see below).

A JASCO J-810 spectropolarimeter with a quartz cuvette of 1 cm path length was used to monitor the denaturation experiments by means of circular dichroism. CD experiments were monitored at 222 nm (wavelength at which maximum signal for α -helix is recorded) with a bandwidth of 4 nm and collection of data every 0.2 degrees. GuHCl denaturation experiments employed an initial volume of 1.7 mL of a protein solution about 0.7 μ M in 20 mM sodium phosphate buffer at pH 8.0 and at the appropriate concentration of the inorganic salt.

Protein denaturation was achieved by the addition of aliquots (performed with an automatic titrator) of a solution with the same protein concentration and buffer conditions but also containing the denaturant agent. The concentration of the denaturant agent was adjusted by measuring the refractive index in an Atago RX-7000DX refractometer. The range of co-solute concentrations tested was limited by the reduced solubility of the inorganic salts in the presence of denaturant. Thus, denaturant concentration was lowered to expand the range solubility without compromising the proper determination for the unfolded state baseline (see below), with final denaturant concentrations falling between 4.5 M and 7 M.

Between two and four independent measurements were obtained for each experimental condition and duplicate points were used to obtain an estimation of the error. Data analysis was completed with in-house built scripts programmed in MatLab[®] (Simulink). Guanidinium chloride denaturation data monitored by CD were processed assuming the linear extrapolation method adapted to a three states model.

3.3.1.1. Linear extrapolation method: two-states fitting

Linear extrapolation method assumes a linear preferential stabilization of the unfolded state versus the folded state as the denaturant concentration increases (Greene and Pace 1974). For a simple reversible two-states model, $F \rightleftharpoons U$, monitoring of the native (F) and unfolded (U) populations allows the determination of the midpoint of chemical denaturation C_m and the slope at the unfolding region (Figure M11). The former corresponds to the denaturant concentration at which the stability of the protein is zero; therefore the populations of native and unfolded state are equal. The latter, in turn, is related to the cooperative unit m_{U-F} that reflects the increment upon unfolding of exposed binding sites and is proportional to $\delta(\Delta G_{U-F}) / \delta[\text{denaturant}]$.

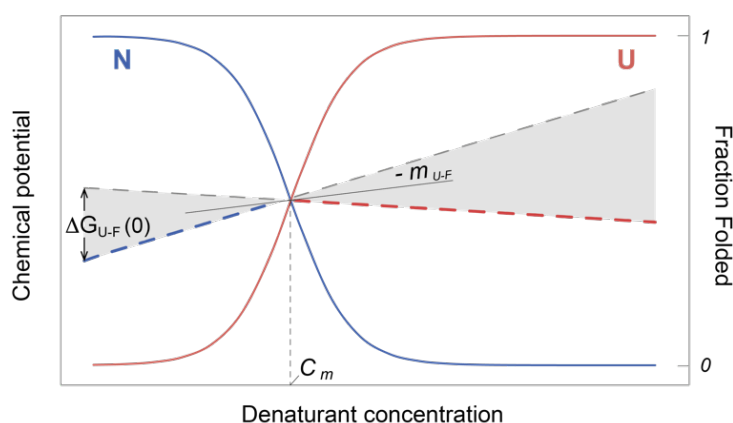


Figure M11. Linear Extrapolation Method. Addition of denaturant preferentially lowers the chemical potential of the unfolded state (blue dashed line) over that of the folded state (red dashed line). As a result, the stability of the protein is linearly decreased (grey area) and the population of the folded state (blue solid line) is displaced to the unfolded state (red solid line). Denaturant concentration at which the ΔG_{U-F} is zero and both populations are equal is known as the midpoint denaturation concentration C_m , and the ratio of populations around that region allows the calculation of the cooperative unit of denaturation m_{U-F} .

From the C_m and the m_{U-F} a linear regression can be carried out to get the $\Delta G^0(0)$ value in the absence of denaturant. The equation that relates the CD signal detected to the thermodynamic quantities involved is:

$$K_{eq} = \frac{U}{F} = 1 - \frac{F}{F} = e^{-\Delta G_{U-F}^0/RT} \rightarrow F = \frac{1}{1 + e^{-\Delta G_{U-F}^0/RT}} ;$$

then

$$\theta_{obs} = \theta_F \cdot F + \theta_U \cdot U = \frac{\theta_F + \theta_U \cdot e^{\frac{-\Delta G_{U-F}^0 + m_{U-N}[\text{denat}]}{RT}}}{1 + e^{\frac{-\Delta G_{U-F}^0 + m_{U-N}[\text{denat}]}{RT}}} ;$$

Equation M2. Fitting of CD data to two-states model. Observed ellipticity θ_{obs} as the sum of the ellipticity of folded (θ_N) and unfolded (θ_U) states modulated by their relative stabilities and the denaturant concentration.

where θ_F , θ_U and θ_{obs} correspond to the native state, unfolded state and total observed ellipticity respectively, and ΔG^0_{U-F} to the unfolding free energy in the absence of denaturant.

3.3.1.2. Three-states fitting

In the case of *Hv* LigN, a simple two-state model is not enough to explain to stability curves. An intermediate is clearly visible, so a three state model $F \rightleftharpoons I \rightleftharpoons U$ is more appropriate (Figure M12).

The two-state equation previously described is no longer valid, the new equation used for the raw data fitting is (Georlette, Blaise et al. 2004):

$$\theta_{obs} = \theta_F \cdot F + \theta_I \cdot I + \theta_U \cdot U = \frac{\theta_F + p \cdot [denat] + \{\theta_I + (\theta_U + q \cdot [denat] \cdot e^b)\} \cdot e^a}{1 + (1 + e^b) \cdot e^a} ;$$

where

$$a = \frac{-\Delta G_{I-F}^0(0) + m_{I-F} \cdot [denat]}{RT} \quad \text{and} \quad b = \frac{-\Delta G_{U-I}^0(0) + m_{U-I} \cdot [denat]}{RT} ;$$

Equation M3. Equation used for fitting of chemical denaturation data to a three-states model.

and where $\Delta G_{I-F}^0(0)$ and $\Delta G_{U-I}^0(0)$ are, respectively, the free energy differences between the intermediate and the native state and between and the unfolded and the intermediate states in the absence of denaturant; θ_{obs} , θ_F , θ_I , and θ_U are the ellipticity observed and the ellipticity for the native, intermediate and unfolded state respectively; m_{I-F} and m_{U-I} are the slopes of the transitions and p and q are the empirical factors that account for the linear dependence of the CD signal of the folded and unfolded state. The experimental total free energy of unfolding in the absence of denaturant for both protein isoforms and for each salt concentration corresponds to:

$$\Delta G_{U-F}^0(0) = \Delta G_{I-F}^0(0) + \Delta G_{U-I}^0(0)$$

Equation M4. Total free energy difference upon unfolding of *Hv* LigN.

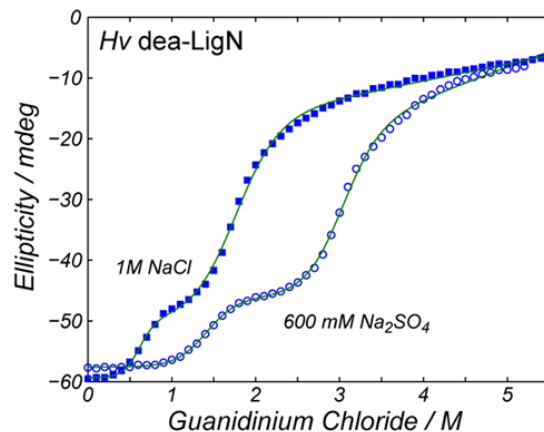


Figure M12. Chemical denaturation with guanidinium chloride of deadenylated *Hv* LigN at two different salt concentrations. Blue squares represent experimental CD measurements and green lines the best fittings to equation M3.

3.3.2. Thermal denaturation experiments

In contrast with the unfolding free energy ΔG_{U-N}^0 linear dependence on denaturant concentration, the temperature modulation becomes non-linear. As shown in Gibbs–Helmholtz equation (Equation M5 and figure M13), it depends largely in the changes in heat capacity and the enthalpic contribution upon unfolding (Becktel and Schellman 1987):

$$\Delta G_{U-F}^0 = \Delta H_m^0 \cdot \left(1 - \frac{T}{T_m}\right) + \Delta C_p^0 \cdot \left[(T - T_m) - T \cdot \ln\left(\frac{T}{T_m}\right)\right] \quad ;$$

Equation M5. Gibbs–Helmholtz equation for the dependence of the unfolding free energy versus temperature.

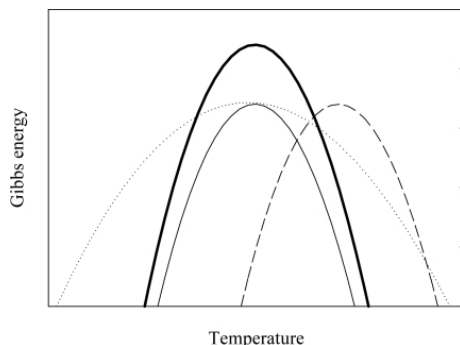


Figure M13. Protein stability curves as a function of temperature. The higher the ΔC_p the more sharply the ΔG depends on T.

Both ProtL and β LAC exhibit thermally reversible unfolding and a two-states equilibrium. Thermal denaturations at different salt concentrations were carried out in order to characterize the stabilization profile upon salt addition. A JASCO J-810 spectropolarimeter with a quartz cuvette of 2 mm path length was used to monitor the denaturation experiments by means of CD and fluorescence. CD experiments were monitored at 222 nm (wavelength at which maximum signal for α -helix is recorded) with a bandwidth of 4 nm and collection of data every 0.2 degrees. Thermal denaturation experiments employed an initial volume of 350 μ L of ProtL (β LAC) solution about 15 μ M (1 μ M) in 20 mM sodium phosphate buffer at pH 6.0 (pH 7.5) and at the appropriate concentration of the tested cosolute.

At each cosolute concentration, both the melting temperature T_m and the enthalpy contribution at that same temperature ΔH_m were evaluated by fitting the experimental data assuming a two-state process (Santoro and Bolen 1988):

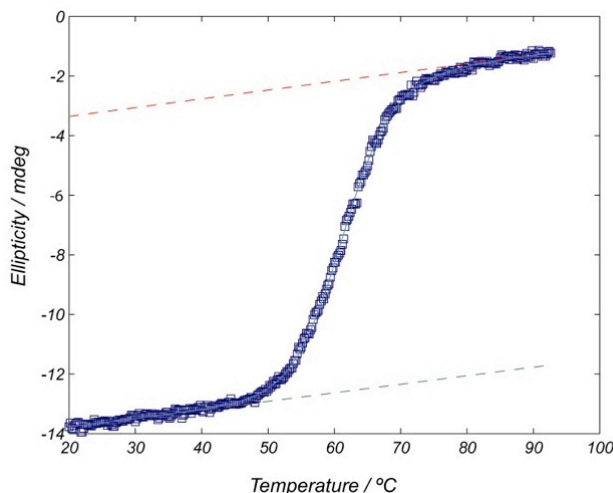


Figure M14. Thermal denaturation. wt ProtL thermal melt measured by CD. Blue squares correspond to experimental points, blue line corresponds to the best linear fitting to equation 5, red and blue dashed lines correspond to T dependence of ellipticity baseline for unfolded and native conformations.

The total observed signal can be expressed as the linear combination of the native and unfolded species:

$$\theta_{obs} = (\theta_F + p \cdot [denat]) \cdot F + (\theta_U + q \cdot [denat]) \cdot U ;$$

the populations of which determine the equilibrium

$$\frac{U}{F} = \frac{1-F}{F} = e^{-\Delta G_{U-F}/RT} = e^{-(\Delta H - T \cdot \Delta S)/RT}$$

where,

$$\Delta H = \Delta H_m^0 - \Delta C_p \cdot (T_m - T) \quad \text{and} \quad \Delta S = \frac{\Delta H_m^0}{T_m} - \Delta C_p \cdot \ln \left(\frac{T_m}{T} \right) ;$$

Equation M6. Thermal denaturation analysis. Equation used for fitting of thermal denaturation data to a two-states model. Adjustable parameters are T_m and ΔH^0 .

where p and q represent empiric factors for the baseline temperature dependence of ellipticity for the native and unfolded state, ΔG_{U-F} , ΔH and ΔS are the free energy, enthalpy and entropy at a given temperature T , ΔC_p the increment in heat capacity upon unfolding and ΔH_m^0 the melting enthalpy at standard conditions. Experimental curves are numerically fitted to equation M6 using in-house built MatLab[®] scripts with two adjustable parameters: T_m and ΔH_m^0 . ΔC_p can be extracted from the ΔG versus temperature curves (Figure M13) or from differential scanning calorimetric *DSC* experiments. Alternatively, since ΔC_p is directly proportional to the area exposed upon unfolding, it can also be estimated from the change in accessible solvent area upon unfolding ΔASA , which in turn shows an empirical correlation with protein size (Myers, Pace et al. 1995):

$$\Delta C_p = 0.19 \cdot (\Delta ASA) - 251 \quad \text{and} \quad \Delta ASA = 93 \cdot (n^\circ aa) - 907 \quad ;$$

Equation M7. ΔC_p calculation. Empirical expressions for the estimation of ΔC_p ($\text{cal} \cdot \text{mol}^{-1} \cdot \text{M}^{-1}$) from ΔASA (\AA^2), which can also be estimated from the total number of residues in the protein, as developed by (Myers, Pace et al. 1995)

Since the model is only accurate for ΔG calculation at $T \sim T_m$ and the lineal extrapolation method is not valid for the temperature dependence of stability (Figure M13), thermal denaturations do not allow the straight determination of absolute ΔG_{U-F}^0 . Alternatively, increments in T_m can be interpreted as relative changes in global free energy $\Delta \Delta G_{U-F}^C$ (Santoro and Bolen 1988):

$$\Delta \Delta G_{U-F}^C = \Delta H_m^R \cdot \left(1 - \frac{T_m^R}{T_m^C} \right) - \Delta \Delta C_p^C \cdot \left[(T_m^C - T_m^R) + T_m^0 \cdot \ln \left(\frac{T_m^R}{T_m^C} \right) \right] \quad ;$$

Equation M8. Stabilization estimates from thermal melts. Calculation of changes in free energy upon cosolute addition with respect to a reference as a function of melting temperature variations.

where the subscript R and the superscript C stand for the value of the property at the reference (usually zero) and tested salt concentrations. $\Delta\Delta C_p^C$ corresponds to the change of heat capacity value upon salt addition. It can be extracted from the slope of T_m versus ΔH_m plot, and it has already been reported to be zero for cosolute addition (Tadeo, Pons et al. 2007). Therefore, the expression is simplified to:

$$\Delta\Delta G_{U-F}^C = \Delta H_m^R \cdot \left(1 - \frac{T_m^R}{T_m^C}\right) ;$$

Equation M9. Simplified expression for the calculation of salt stabilization measured as increments in melting temperature.

ΔH_m^C and T_m^C are obtained from the fitting of the stability curve to equation M6, so the changes in ΔG_{U-F}^0 at different cosolute concentrations with respect to a reference value can be calculated.

3.4. NUCLEAR MAGNETIC RESONANCE

Nuclear Magnetic Resonance (NMR) is a powerful technique for the study of DNA or proteins. It provides high-resolution structural information and is the only technique that allows exploring the dynamic behaviour of biological systems in solution. In recognition of its early development and applications several Nobel Prizes have been awarded to Otto Stern (1943), Isidor Rabi (1944), Felix Bloch & Edward M. Purcell (1952), Richard Ernst (1991), Kurt Wüthrich (2002) and Paul C. Lauterbur & Peter Mansfield (2003). Our group makes extensive use of NMR as a tool to obtain information about structural and biophysical properties at atomic resolution, about dynamics, disordered conformations, global properties and solvation of biological systems, mostly proteins. A brief introduction to NMR basic principles and applications follows. More detailed description is beyond the scope of this thesis and can be found in specialized books (Keeler 2002, Rule and Hitchens 2006).

3.4.1. Introduction

Nuclear Magnetic Resonance (NMR) Spectroscopy exploits the properties that certain magnetic nuclei have in the presence of a magnetic field and applied electromagnetic pulse or pulses. Nuclei absorb energy from the electromagnetic pulse and radiate this energy back out at a specific resonance frequency, which depends on the strength of the magnetic field and the chemical environment, allowing the observation of specific magnetic properties of an atomic nuclei.

NMR is related to the nuclear *spin*, an intrinsic quantum property of atomic nuclei. Rotating charged particles, like atomic nuclei, acquire magnetic properties. For a single nucleus, its spin determines the magnitude and orientation of magnetic angular momentum. Atomic nuclei behave like small rotating magnets, therefore interacting with external magnetic fields. In the presence of a magnetic field B_0 , nuclei with a spin quantum number I different from zero adopt one of the possible angular momentums determined by the magnetic quantum number m (which goes from $+I$ to $-I$ in integer steps). Dipolar nuclei (nuclei with spin $\frac{1}{2}$) are the most studied by NMR. Under the influence of a magnetic field B_0 , they split in two energetic levels (Zeeman effect), one with the magnetic momentum aligned to the magnetic field ($m = +1/2$; lower energy state) and the other with an opposite orientation ($m = -1/2$; higher energy state) (Figure M15). As a result only one transition is possible, giving rise to sharp spectral lines in contrast to larger spin quantum numbers (quadrupolar nuclei) that give rise to more transitions and much more complex spectra; e.g., nuclei with spin $I = 3/2$ split in four energetic states ($m = +3/2, +1/2, -1/2$ and $-3/2$). Hydrogen, carbon, nitrogen and oxygen, which are the most common elements in proteins, all have isotopes with spin $I = \frac{1}{2}$: ^1H , ^{13}C , ^{15}N and ^{17}O respectively. However, only for hydrogen it corresponds to the most naturally abundant, whereas isotopic enrichment is required for the other elements.

Each possible orientation of a magnetic moment is associated to a determined energetic state (Figure M15). The energy difference between the two states is proportional to the magnetic field B_0 and the constant gyromagnetic ratio γ , characteristic of each nucleus type, where h is the Planck's Constant ($6.626 \cdot 10^{-34} \text{ J} \cdot \text{s}$) and $\hbar = h/2\pi$:

$$\Delta E = h \cdot \nu = \hbar \cdot \gamma \cdot B_0;$$

Equation M10. Nuclear energetic transitions. The energetic transition between the quantum states of a proton with spin $\frac{1}{2}$ is proportional to the gyromagnetic ratio γ and the magnetic field B . The oscillating frequency of the equivalent radiofrequency pulse is the characteristic *resonance frequency*, ν .

Resonance frequency ν depends on the external applied magnetic field B_0 , and the local magnetic field induced by the atoms surrounding the observed nuclei, B_{loc} . As the chemical environment is unique for each atom, so are the observed magnetic field $B = B_0 + B_{loc}$ and the characteristic resonating frequency.

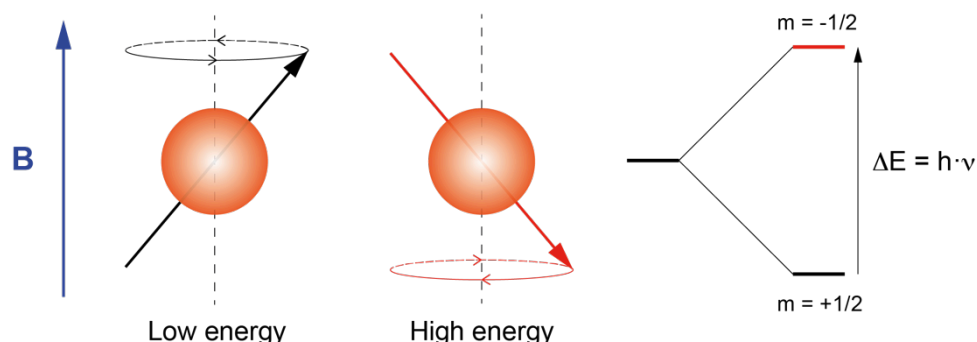


Figure M15. Spin precession and energetic diagram. For a proton with spin $I = \frac{1}{2}$ under the influence of a magnetic field B , the z component of the magnetic moment of the nucleus is aligned along the direction of the magnetic field. Two orientations of z magnetization are possible: the more stable state (black) shares the orientation with the magnetic field, whereas the highest energetic state (red) corresponds to opposite orientation. The energetic difference corresponds to a photon of an electromagnetic radiation at Larmor frequency. According to quantum physics, the x and y components of the magnetization cannot be simultaneously determined and, in this representation, they oscillate around the z -axis (spin precession) at the Larmor frequency (ν) with no net component.

With the currently available magnets, the energetic gaps between those states are relatively small when compared to those between the electronic states detected in fluorescence spectroscopy or the vibrational transitions detected by infrared spectroscopy. Low energetic gaps produce small population differences, thus NMR is a low-sensitivity technique that requires a high level of accumulation to increase the signal-to-noise ratio. Nuclear spin transitions in NMR fall in the range of 10-1000 MHz, corresponding to the radiofrequency part of the electromagnetic spectrum. Therefore, radiofrequency pulses are used to interact with the energetic states. Radiofrequency pulses are electromagnetic radiation associated to an oscillating magnetic field. The frequency of the pulse or *Larmor frequency* corresponds to the carrier frequency of the radiofrequency pulse required for the transit between the two energetic states of a given nuclei, and depends heavily on the external magnetic field B_0 .

According to quantum physics, if spin rotation for the z component of the magnetic moment is quantified, x and y components cannot be simultaneously determined. They oscillate around the z -axis at a characteristic resonance frequency. As a result, the magnetic moments precess around the z -axis in what is called *spin precession* (Figure M15 and M16). Properties detected in NMR experiments result from the time average of the ensemble of spins in the sample. Net magnetization along z -axis (Figure M16, left) arises from the addition of all the small magnetic moments in the sample. Radiofrequency pulses alter the populations and the orientation of the net magnetic moment of the sample. If the oscillation frequency (Larmor frequency) matches that of spin precession, the magnetic moment of the nuclei is pushed out from its equilibrium position along the z -axis. Net longitudinal magnetization along z -axis is transformed into transversal magnetization on the xy -plane (Figure M16, right), where the magnetization oscillates at the resonating frequency of the nucleus with an *offset* with respect to the Larmor frequency. This oscillation at a unique frequency for each atom is what is measured in NMR experiments. *Fourier transformation* of these oscillating signals into frequencies creates the NMR spectrum.

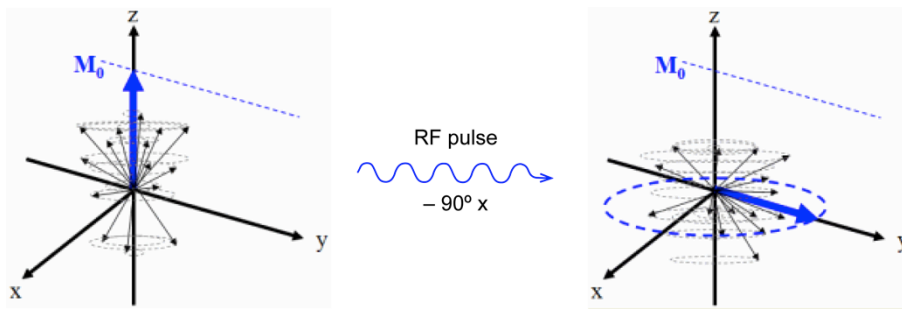


Figure M16. Magnetic moment oscillation. Individual magnetic moments (black vectors) oscillate around z-axis at Larmor frequency. At equilibrium, net magnetization (blue vector) is aligned with the magnetic field. A radiofrequency pulse at Larmor frequency along x-axis shifts the z and y components of individual magnetic moments. As a result, net magnetization is transferred to y-axis, where it precesses around z-axis at Larmor frequency.

Since frequencies depend on the magnetic field (Equation M10), they are usually normalized to *chemical shift* scale to allow for homogeneous comparison (Equation M11). The larger the magnetic field the more expanded is the frequency range and the higher is the resolution. That is why newer and more advanced spectrometers always seek to increase the power of the magnetic field. Most powerful spectrometer so far is a 23.5 Tesla magnet at 'Centre de RMN à Très Hauts Champs' in Lyon, France.

$$\text{chemical shift, } \delta \text{ (ppm)} = \frac{\nu_0 - \nu_{TMS}}{\text{Spectrometer frequency (MHz)}} \cdot 10^6 \quad ;$$

Equation M11. Chemical shift scale in parts per million (ppm). Obtained from the subtraction of the resonance frequency from that of a reference compound (TMS, try-methyl silyl or similar) and normalized by the frequency of the spectrometer.

Protein NMR frequently utilizes multidimensional experiments because, although each distinct nucleus experiences a distinct chemical environment and thus has a distinct chemical shift by which it can be recognized, in large molecules such as proteins the number of resonances can typically be several thousand and a single frequency dimension spectrum (1D) has too much overlapping:

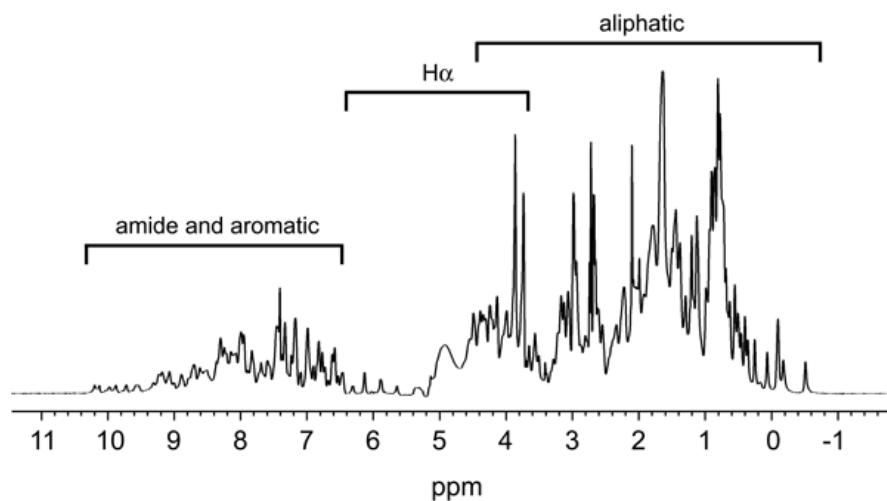


Figure M17. Example of a ^1H monodimensional spectrum of a protein. Thousands of ^1H resonances appear throughout the proton frequency range of the spectrum distributed by their bonding pattern and their chemical environment. The high degree of overlapping is evident, and makes individual assignment of the resonances an unbearable task.

Multidimensional experiments correlate the frequencies of distinct nuclei, decreasing the chance of overlaps and providing more information since they correlate signals from nuclei within a specific part of the molecule. Magnetization is transferred into the sample and between nuclei using pulses of electromagnetic energy and delays. These pulse sequences select specific types of connections between nuclei (coherences). There are different mechanisms for magnetization transfer, three of which will be outlined:

- i. Spin-spin scalar coupling through the electron cloud (J coupling) allows transferring the magnetization to nuclei typically within distance of one to three bonds. It provides important connectivity information with neighbouring magnetically active nuclei – either from the same element (homonuclear coupling) or from different elements (heteronuclear coupling) – essential for the assignment of resonance frequencies to the nuclei in the sample.
- ii. Through-space nuclear spin coupling is the basis for the Nuclear Overhauser Effect (NOE), which allows to get short-range structural restraints ultimately used in the structure calculation by NMR.
- iii. Dipolar coupling is the through-space interaction between two dipoles that allows getting long-range structural restraints. It is the most efficient route in solid-state NMR.

Transfer of magnetization sets the foundations for multidimensional and heteronuclear NMR. Heteronuclear Single Quantum Correlation (HSQC) is the most used 2D protein NMR experiment in biomolecular NMR spectroscopy. HSQC spectra have one peak for each H bound to a heteronucleus. Thus, in the ^{15}N -HSQC one signal is expected for each amino acid residue (Figure M18), with the exception of proline, which has no amide-hydrogen due to the cyclic nature of its backbone. Moreover, tryptophan and certain other residues with N-containing sidechains will give rise to additional signals. The ^{15}N -HSQC is the fingerprint of a protein because each protein has a unique pattern of signal positions, it allows a quick check on sample quality and is already very informative on the protein backbone. ^{15}N -HSQC also constitutes the framework for most of further multidimensional experiments.

Multidimensional and heteronuclear NMR experiments provide useful structural information ultimately leading to the high-resolution structure of proteins and DNA. Alternatively to X-Ray crystallography, solution NMR of biological macromolecules makes it possible to study their structure and properties in conditions close to physiological environments. Moreover, dynamic properties from the *ns-ps* to the *s-ms* (and even *hours-days*) timescales can be measured. Limitations of NMR usually involve requirements of highly soluble, stable and relatively small proteins. However, those limits are being constantly pushed further by technical advances increasing sensitivity and by new methods, such as selective labelling and methyl-TROSY, which allow tackling complexes as large as 1 MDa.

In the following sections, the experiments used in this thesis are briefly described.

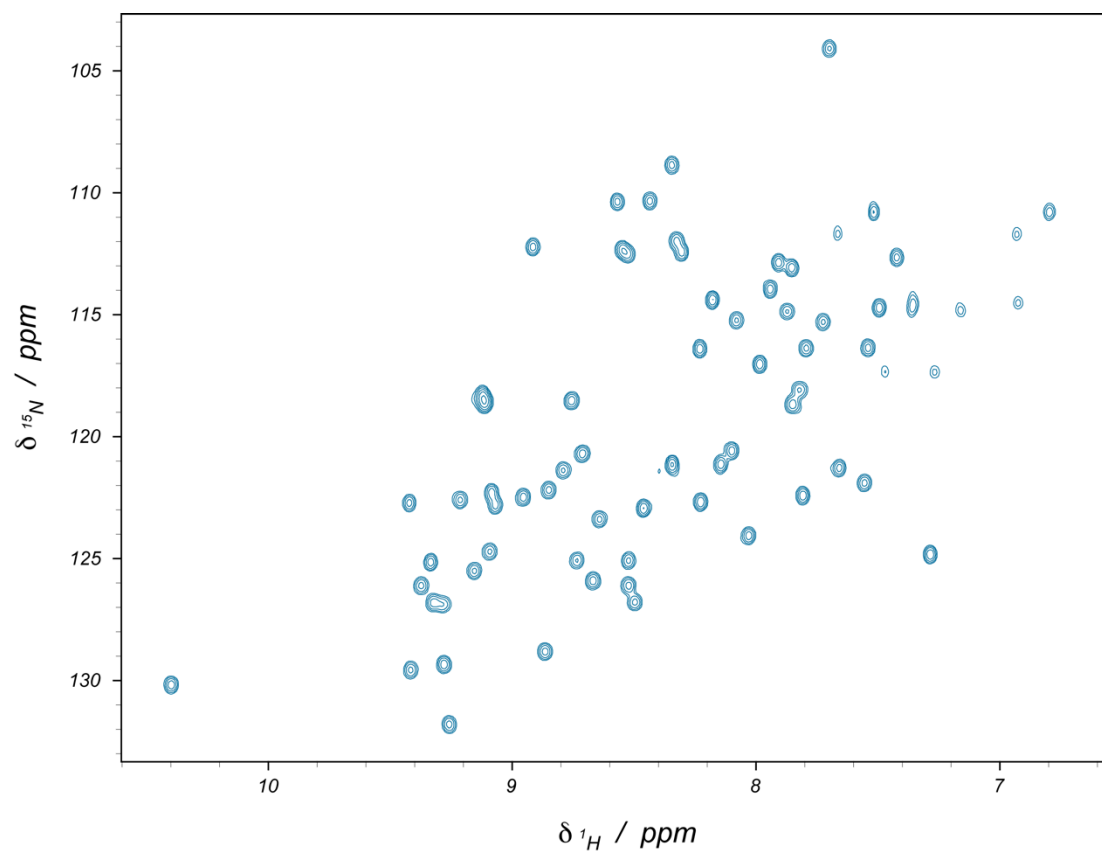


Figure M18. $^1\text{H}, ^{15}\text{N}$ -HSQC spectrum of ProtL. Each signal corresponds to a N–H bond, resonating at the ^1H and ^{15}N frequency on the horizontal and vertical axis respectively.

3.4.2. Backbone and side-chain assignments

Each one of the signals detected in a NMR spectrum must be assigned to its corresponding nucleus in order to give a structural context to the information from the experiments. Multidimensional heteronuclear NMR experiments combine complex radiofrequency pulses and delays that transfer the magnetization between different elements, typically from protons to other heteronuclei such as carbon and nitrogen. This is particularly useful for protein characterization by NMR. If the protein is labelled with magnetically active isotopes (^1H , ^{15}N and ^{13}C), the magnetization can be transferred along the backbone and the side-chain atoms. Alpha (CA), beta (CB) and further carbons of the side-chains have chemical shifts characteristic for each amino acid type. Chemical shifts are essential for identification purposes; unfortunately, these alone are certainly not enough for unambiguous assignment.

The information about the sequential connectivity of the signals complements the information of the carbon chemical shift. Typically, 3D multidimensional heteronuclear experiments recorded are the HNCA / HN(CO)CA and HN(CA)CB / HN(COCA)CB pairs. Magnetization generated at the amide proton of residue i is first transferred to the amide nitrogen. Next, in HN(CO)CA and HN(COCA)CB pulse sequences, magnetization is transferred through the backbone carbonyl to the CA and CB of the residue $i-1$ (Figure M19). We get, then, the inter-residue connectivity between the amide bond and the CA and CB of the previous residue. The complementary experiments HNCA and HN(CA)CB transfer the magnetization from the amide backbone of residue i to the CA and CB of both the same i residue and the previous $i-1$ residue (Figure M19). They establish both the intra and the inter-residue connectivity. The combination of both pairs of experiments helps following the magnetization back and forth along the polypeptide chain and allows assigning each resonance signal in the NMR spectrum to its corresponding atom in the protein.

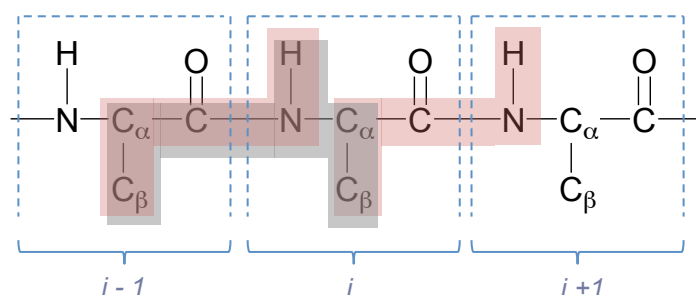


Figure M19. Schematic diagram of the connectivity established by the 3D-experiments for protein assignments. HNCA and HN(CA)CB (grey shadow) gives the inter ($i-1$) and intra (i) residue correlation of the CA and CB with amide NH from residue i . Complementary HN(CO)CA and HN(COCA)CB (red shadow) show the intra-residue correlation between the NH bond from residues i and the CA and CB from residue $i-1$.

Freshly prepared samples of βLAC (900 μM [^{15}N , ^{13}C , ^2H] βLAC in 20 mM phosphate buffer at pH 7.5, 500 mM KCl, 1mM NaN_3 and 10% D_2O) and haloSRC (1 mM [^{15}N , ^{13}C] haloSRC in 50 mM phosphate buffer at pH 7.0, 0.2 mM EDTA, 0.2 mM PMSF, 1 mM NaN_3 and 10% D_2O) were used. The βLAC sample was labeled in ^2H (D) instead of ^1H to reduce sources of relaxation such as dipolar interactions and chemical shift anisotropy, which decrease the lifetime of the detected magnetization. The sample was afterwards incubated for two days at 298 K to maximize back-exchange of labile ND deuterons with the solvent. For the disordered haloSRC the experiment temperature was decreased to 280 K to minimize fast HN/ H_2O

exchange that significantly broadens the NMR signals. All 3D-TROSY type experiments for backbone assignment (HNCO, HNCA, HN(CO)CA, constant-time HN(CA)CB and HN(COCA)CB) were acquired on an 800 MHz Bruker Avance III spectrometer equipped with a TCI cryoprobe. Use of a constant CB evolution time thereby provides valuable additional information for assigning the CB signals as their sign depends on the side chain topology (Castaño and Millet 2010). Still, since haloSRC is disordered, the high spectral overlap hampers complete assignment. Another set of complementary experiments, HNCO and HN(CA)CO are recorded in order to add a third independent layer of constraints to the sequential connectivity. β LAC spectra were processed with NMRPipe and analyzed with NMRView assisted by a suite of in-house scripts, whereas haloSRC spectra were processed with Topspin 3 and analyzed with Sparky. For β LAC, 230 (91 %) non-proline residues were assigned, and the chemical shift data was deposited in the BioMagResBank (access code: 18373). For haloSRC, even more complete backbone signal assignment (> 95 %) was achieved (BioMagResBank access code: 25196).

3.4.3. pH titrations

The charge of an ionizable group depends greatly on the chemical environment experienced by acidic and basic residues. The absence or presence of charges, in turn, alter the local magnetic field experienced by the observed atom. As a result, chemical shifts of atoms from ionizable groups are highly sensitive to their protonation state. Measurements of chemical shift along the pH titration coordinate allows monitoring the charged and neutral populations as a function of pH at atomic resolution, ultimately leading to the determination of the acidity constants for each residue.

A 2D H(C)CO experiment was used to resolve the carboxylate ^{13}C signals of aspartate and glutamate residues while a ct- ^1H , ^{13}C -HSQC was used to monitor the NMR signals of CD of lysines. The protonation – deprotonation equilibrium is fast compared to the timescale of NMR detection. As a result, observed signal frequency δ_{obs} shifts continuously between the protonated and deprotonated chemical shifts as a function of pH. The titration curve can be fitted to the general Henderson-Hasselbach expression to extract individual pK_a values (Bosshard, Marti et al. 2004):

$$\delta_{obs} = \delta_d \cdot q(i) + \delta_p \cdot (1 - q(i)) \quad ; \quad \text{where} \quad q(i) = \frac{K_a(i)}{K_a(i) + [H^+]} \quad ;$$

$$\text{then} \quad \delta_{obs} = \frac{\delta_d \cdot K_a(i) + \delta_p \cdot [H^+]}{K_a(i) + [H^+]} \quad ;$$

Equation M12. Chemical shift as a probe of protonation states. The observed chemical shift δ_{obs} reflects the relative populations of protonated and deprotonated species along the pH titration, which is a function of the acidity constant K_a .

where δ_d and δ_p are the chemical shifts of the deprotonated and protonated species, respectively, $q(i)$ is the pH dependent charge and $K_a(i)$ the acidic constant of a given ionizable group i .

pH titration experiments were recorded on 200 μM [$U\text{-}^{15}\text{N}$, ^{13}C] labeled protein samples at 298 K for ProtL and βLAC or 280 K for haloSRC and uSRC. Proteins were first lyophilized at low buffering conditions (5 mM) and then resuspended in 450 mL D_2O (> 95 %) containing 200 mM buffer at the desired pH and salt concentration. Carbonate/bicarbonate buffer at high pH [9.14 to 13], sodium phosphate buffer at intermediate pH [6.0 to 8.0] and deuterated acetic/acetate buffer was used at low pH [3.0 to 5.5]. Deuterated instead of regular acetate was used to avoid strong buffer H-CO correlation signals interfering with the ones from the protein. The pH was adjusted (at about 0.5 unit intervals) using 1 M DCl or NaOD solutions.

3.4.4. Relaxation experiments

Magnetization is eventually lost after some time in a process known as relaxation. Longitudinal relaxation (R_1) brings the magnetization back to the equilibrium orientation along the external magnetic field, whereas transversal relaxation (R_2) destroys the coherence of the transversal magnetization, i.e. it destroys the synchronous oscillation at Larmor frequency of the x and y components of individual magnetic moments.

Relaxation processes eventually bring the magnetization to its equilibrium value and are intimately ligated to dynamics of the molecule (Figure M20). Their contribution to “relax” the spins in the sample by helping them to release the excess energy constitutes the link between energetics and internal and external dynamics. Local and global motions cause fluctuating magnetic fields, which interact with the individual magnetic moments and thus rotate them to new positions, in just the same way as a radiofrequency pulse. The effects can be explained as two different contributions:

- i. **Non-secular contributions:** local fields oscillating at Larmor frequency act as radiofrequency pulses that change the orientation of the magnetic moments of individual spins. It accounts for longitudinal relaxation (bringing z-magnetization to equilibrium) and part of transverse relaxation (pushing transverse magnetization out of xy-plane back to z-axis).
- ii. **Secular contributions:** oscillating local fields provoke a variable inhomogeneity in the magnetic field observed by each equivalent atom in the sample. As a result, the Larmor frequency at which each spin precesses around the z-axis is slightly different. The global transversal magnetization detected arises from the addition of synchronous oscillating magnetic moments, therefore when some of them fall out of step, the total magnetization decreases. This process contributes to transversal relaxation.

A particular source of a local magnetic field is called a relaxation mechanism. While there are quite a lot of these, two tend to be dominant for spin-half nuclei.

First, in the **dipolar mechanism** the local field is due to the presence of the magnetic moment of another spin. There are thus two spins involved: one generating the field and the other one experiencing it. The local field due to the neighbouring spin depends on the distance between the two spins (falls off at $1/r^3$), their gyromagnetic ratio and the orientation of the vector joining the two spins with respect to the applied magnetic field along the z-axis. In two-spins systems coupled through the dipolar mechanism, **cross-relaxation**, in which magnetization is mutually transferred between two spins, strongly contributes to relaxation.

As opposed to the scalar coupling, in which magnetization transfer occurs via electron clouds and chemical bonds, magnetization in dipolar coupling is transferred through space. This gives rise to a very relevant phenomenon, the **Nuclear Overhauser Effect (NOE)**, which is the cross-relaxation between the z-magnetization of two spins. Not only is NOE useful to probe the dynamics of the sample, but it is also a very important tool for obtaining structural information since the magnitude of cross-relaxation and magnetization transfer heavily depends on the distance. As a matter of fact, NOE is the central experiment in the obtention of high-resolution structures by NMR. The equivalent phenomenon to NOE for transversal xy-magnetization is the ROE, with important applications in CLEANEX experiments, as explained in section 3.4.5.2.

The second mechanism considered is the **chemical shift anisotropy**. Electrons in the molecule give rise to a small local field at the nucleus. For most molecules, this induced local field is anisotropic and is characterized by a chemical shielding tensor. As the molecule tumbles in solution or local motions occur within the molecule, the orientation of the tensor changes with respect to the external strong magnetic field. As a result, the induced magnetic field changes.

Other important relaxation mechanisms involve **scalar coupling (J-coupling)**, the presence of paramagnetic species with strong induced anisotropic magnetic field (e.g. dissolved oxygen or lanthanides) or chemical exchange between two different conformations. Since we haven't considered those for our present work, they fall beyond the scope of this thesis.

For our purposes, it is enough to consider the *link between relaxation and dynamics* (Figure M20). The key about relaxation mechanisms is their *stochastic nature*, which makes different parts of the sample to be randomly affected to a different extent. Due to molecular tumbling, rotational diffusion, local motions and random orientations of the molecules in the sample, these local fields are different for each spin and, more importantly, variable over time; same as the local magnetic field, Larmor frequency and chemical shift experienced for each individual spin. That causes coherence to be lost and magnetization to go back to equilibrium. As a consequence, signal intensity decreases exponentially. Specific pulse sequences selectively measure the decrease of longitudinal or transversal magnetization over time. Each experiment and relaxation process informs about dynamics at different timescales (Figure M20). Particularly, relaxation rates R_1 , R_2 and NOE calculated from the experimental data can be interpreted in terms of the local and global dynamics of the protein.

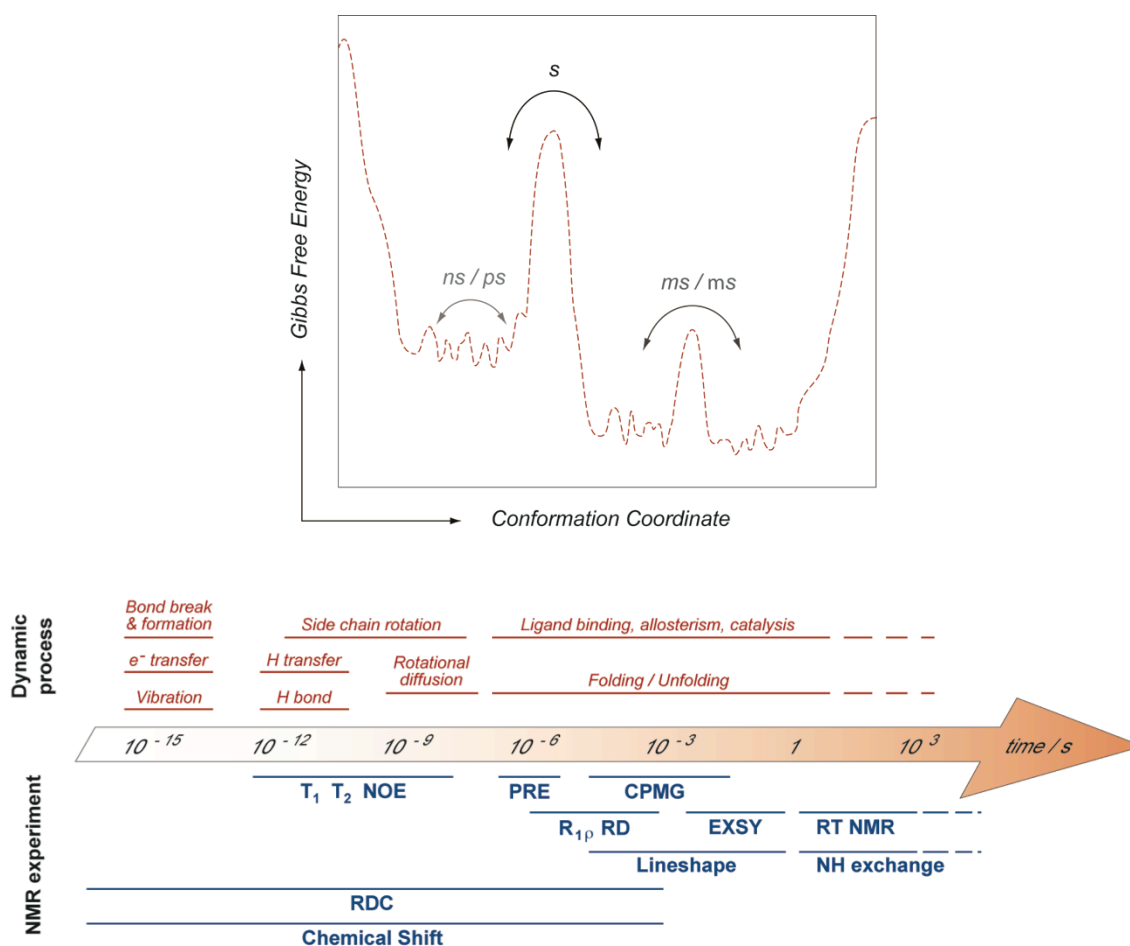


Figure M20. Molecular dynamics and NMR experiments. Upper graphic shows a schematic representation of the energetic landscape of the dynamic events within a molecule and their corresponding timescales. The item below shows a detailed listing of main dynamic processes and NMR experiments, together with the corresponding timescale. Abbreviations for NMR experiments: PRE, *paramagnetic relaxation enhancement*; CPMG, *Carl, Purcell, Meiboom & Gill sequence*; RD, *relaxation–dispersion*; EXSY, *exchange spectroscopy*; RT NMR, *room temperature NMR*; RDC, *residual dipolar couplings*.

3.4.4.1. Amide proton ^{15}N relaxation experiments

Backbone amide ^{15}N relaxation data (R_1 , $R_{1\rho}$ and heteronuclear $^{15}\text{N}[^1\text{H}]$ NOE) was measured for haloSRC and uSRC at 600 MHz using described methods (Farrow, Muhandiram et al. 1994). For the $R_{1\rho}$ experiment the spin-lock field was 3 KHz. Typically, 10 to 12 time points were recorded for each relaxation rate (R_1 or $R_{1\rho}$) in interleaved mode and including at least two duplicate time points for error estimation. The sampled time points ranged from 0.1 to 1.2 s for R_1 , and 50 to 250 ms for $R_{1\rho}$. For the heteronuclear $^{15}\text{N}[^1\text{H}]$ NOE experiment, steady-state HN saturation was achieved with a 4 s train of 120° pulses, and a 12 s interscan delay was used in the reference (non saturated) spectrum to avoid bias from residual HN saturation.

3.4.4.2. Relaxation models

NMR relaxation data on macromolecules in solution contain information concerning the nature of internal motions and the overall motion. There are different approaches to extract such information. The use of particular dynamic models for motion establishes a correlation function specific for each type of motion and helps obtaining parameters with direct physical interpretation. Whilst such analysis can be useful, experimental data do not allow choosing the appropriate model. There is the danger of overinterpretation of limited data and the possibility that the resulting picture is not unique.

Alternatively, Giovanni Lipari and Attila Szabo developed in 1982 a model-free approach (Lipari and Szabo 1982a, Lipari and Szabo 1982b). They state that the simplest possible description of internal dynamics involves specifying its rate (time scale) and spatial restriction. As a result, the information on fast internal motions contained in NMR relaxation experiments can be completely described by two model-independent quantities: a generalized order parameter, S^2 , which is the measure of the spatial restriction of the motion (from 1, totally rigid, to 0, totally flexible), and the effective correlation time, τ_e (in seconds), which is a measure of the rate of the motion. Assuming an isotropic macromolecular overall rotation characterized by a single correlation time, τ_c ; and fast intramolecular motions ($\tau_e < 100$ ps) independent of molecular overall rotation, they propose a certain type of spectral density function characterized by a limited number of parameters, but without any specific information about the kind of motion. This approach is referred to as “model-free” because the expression for the spectral density is derived without invoking a specific model for internal motions:

$$J(\omega_0) = \frac{2}{5} \cdot [S^2 \cdot \tau_c + (1 - S^2) \cdot \tau] \quad \text{and} \quad J(\omega) = \frac{2}{5} \cdot \left[\frac{S^2 \cdot \tau_c}{1 + (\omega \cdot \tau_c)^2} + \frac{(1 - S^2) \cdot \tau}{1 + (\omega \cdot \tau)^2} \right]$$

$$\text{where} \quad \frac{1}{\tau} = \frac{1}{\tau_c} + \frac{1}{\tau_e}$$

Equation M13. Spectral density function proposed by Lipari and Szabo for the model-free analysis. The three model-independent parameters are the order parameter S , effective correlation time τ_e and overall correlation time τ_c .

The spectral density function characterizes the dynamic behavior of a system. It quantifies the amount of motion at each determined frequency: ω_0 corresponds to frequency zero, ω_H and ω_N to the proton and nitrogen resonance frequencies, and ω_+ and ω_- to the sum and the subtraction of ω_H and ω_N respectively. Relaxation mechanisms are therefore intimately ligated to spectral density function. Relaxation rates for amide ^{15}N - ^1H derived from Solomon equations can be expressed as:

$$R_1 = \frac{d_{HN}}{4} \cdot [J(\omega_-) + 3 \cdot J(\omega_N) + 6 \cdot J(\omega_+)] + \text{CSA} \cdot J(\omega_N)$$

$$R_{1\rho} = \frac{d_{HN}}{4} \cdot [4 \cdot J(\omega_0) + J(\omega_-) + 3 \cdot J(\omega_+) + 6 \cdot J(\omega_H)] + \frac{\text{CSA}}{6} \cdot [3 \cdot J(\omega_N) + 4 \cdot J(\omega_0)]$$

$$\text{NOE} = 1 + \frac{\frac{d_{HN}}{4} \cdot \frac{\gamma_H}{\gamma_N} \cdot [6 \cdot J(\omega_+) - J(\omega_-)]}{\frac{d_{HN}}{4} \cdot [J(\omega_-) + 3 \cdot J(\omega_N) + 6 \cdot J(\omega_+)] + \text{CSA} \cdot J(\omega_N)}$$

where,

$$\text{CSA} = \frac{(\Delta \cdot \omega_N)^2}{3} = \frac{(2 \pi \cdot 170 \cdot \omega_H \cdot \gamma_N / \gamma_H)^2}{3} \quad \text{and} \quad d_{HN} = \left(\frac{\gamma_N \cdot \gamma_H \cdot h / 2\pi}{r_{NH}^3} \right)^2$$

Equation M14. General expressions derived from theory and Solomon equations for the longitudinal (R_1) and transversal ($R_{1\rho}$) relaxation rates, the dipolar NOE cross-relaxation and the chemical shift anisotropy contributions.

R_1 and $R_{1\rho}$ account for the longitudinal and transversal relaxation rates respectively, NOE for the dipolar cross-relaxation term and CSA for the chemical shift anisotropy. h is the Planck's constant, r_{NH} the length of the NH bond, γ_N and γ_H the nitrogen and proton gyromagnetic ratios.

Experimental rates measured for each residue at various magnetic fields can be least squares fitted to these expressions. The floating variables in the LS-2 model (Lipari-Szabo model-free approach with two adjustable parameters) are two: order parameter S^2 and correlation time τ_e . However, if the molecule overall tumbling cannot be explained by a single correlation time, such as in disordered peptides, LS-3 model also considers macromolecular correlation time τ_c as an additional adjustable variable. Further LS models employ more complex spectral density functions incorporating new variables in the fitting; e.g. motion in two regimes at different timescales or anisotropic molecular tumbling. Statistical analysis on our data suggests LS-3 as an appropriate model for our intrinsically disordered uSRC and haloSRC.

The backbone ^{15}N relaxation data was analysed in terms of the Lipari-Szabo formalism using own MatLab[®] scripts. For a given residue i , all relaxation data was simultaneously fit to a model-free function with three adjustable parameters (LS-3): a generalised order parameter $S^2(i)$ for fast local motions with a local correlation time $\tau_e(i)$ on the picosecond timescale, and an effective *per residue* correlation time $\tau_c(i)$ for superimposed slower local (nanosecond) and molecular tumbling motions (Bruschweiler, Liao et al. 1995), which cannot be disentangled in intrinsically disordered proteins.

Additionally, $R_{1\rho}$ transversal relaxation rates were also directly fitted to a model probing for correlated local motion along an unfolded polypeptide. The unstructured chain dynamics are modelled as a correlated motion (first summatory of the expression) altered by the presence of rigid clusters modelled as Gaussian distributions (second summatory of the expression) (Klein-Seetharaman, Oikawa et al. 2002):

$$R_{1\rho}(i) = R_{1\rho,\text{int}} \cdot \sum_{j=1}^N e^{-\frac{|i-j|}{\lambda_0}} + \sum_{\chi=1}^M R_{1\rho,\chi} \cdot e^{-\frac{|i-j_\chi|^2}{\lambda_\chi}}$$

Equation M15. Expression for model of correlated local motion along a polypeptide chain.

Determined by the length of persistence λ_0 and the presence of M rigid clusters χ .

Here, $R_{1\rho,\text{int}}$ is the intrinsic relaxation rate of the polypeptide chain, index j runs over all its N residues and λ_0 is the persistence length of the chain (in number of residues) as a measure of its stiffness. More compact clusters (index χ) with restricted motion lead to locally increased relaxation rates with a Gaussian distribution around their central residue j_χ , a persistence length λ_χ and an intrinsic cluster relaxation rate $R_{1\rho,\chi}$.

3.4.4.3. ^{17}O Magnetic Relaxation Dispersion experiments

In contrast to ^{16}O , which has a nuclear spin null value, ^{17}O is magnetically active because it has nuclear spin $-5/2$. Oxygen-17 spin relaxation monitors water molecules selectively (without complications from labile hydrogens) and reports on the single-molecule rotational correlation time (without complications from cross-correlations and intermolecular couplings) (Halle, Denisov et al. 2002). In collaboration with Professor Bertil Halle and Dr. Johan Qvist from Lund University (Lund, Sweden), relaxation of ^{17}O -labelled water molecules was measured in order to probe the hydration shell of halophilic versus non-halophilic proteins.

By recording the longitudinal spin relaxation rate R_1 as a function of resonance frequency – the magnetic relaxation dispersion (MRD) profile – information about long-lived internal water molecules as well as about the more mobile water molecules interacting with the external protein surface can be obtained (Halle, Denisov et al. 2002). Here, the ^{17}O MRD profile was measured for WT ProtL and for the obligate halophile Kx6E with and without salt.

The proteins were dissolved in water enriched to 21 % in ^{17}O (Isotec). Kx6E solutions were prepared without added salt and with 750 mM NaCl (> 99.5 %, Merck). The protein solutions (about 10 mM unlabeled protein) also contained phosphate buffer (19 mM for ProtL and 7 mM for Kx6E), 2 mM NaN_3 and pH was adjusted to 7.8 ± 0.1 with NaOH and HCl. Three solvent reference samples were prepared with the same solvent composition as in the corresponding protein solutions. The final protein concentrations were determined from the absorbance at 280 nm in the presence of 4 – 5 M guanidinium chloride, using an extinction coefficient of $9952 \text{ mol}^{-1} \cdot \text{cm}^{-1}$ based on a complete amino acid analysis of the ProtL sample. To rule out protein aggregation, water ^1H MRD profiles were recorded at 300 K in the frequency range 10 kHz – 40 MHz using a Stellar Spinmaster 1 T fast field cycling instrument. This technique can detect even very small populations of protein aggregates (Gottschalk and Halle 2003), but none were seen in the samples used for ^{17}O relaxation measurements.

The dispersion of the water ^{17}O longitudinal relaxation rate R_1 was measured at 27.1, 67.8 and 81.3 MHz with the aid of fixed-field superconducting magnets (Bruker Avance 200, Varian Unity Plus 500 and 600) and in the frequency range 2.1 – 11.5 MHz with a field variable resistive magnet (Drusch) interfaced to a Tecmag Discovery console. The sample temperature was maintained at 300.2 ± 0.1 K with a regulated flow of dried air and was measured before and after each relaxation measurement with a copper-constantan thermocouple referenced to an ice-water bath. At each field, R_1 was measured for a protein sample and then for the corresponding reference sample, using the inversion recovery method with 30 – 40 relaxation delays and a sufficient number of transients to obtain a signal-to-noise ratio larger than 100. Based on repeated measurements, the accuracy in R_1 was estimated to be 1.0 % at the two lowest frequencies and 0.5 % at all other frequencies.

3.4.4.4. ^{17}O relaxation experiments

To characterize the surface hydration dynamics, the temperature dependence of ^{17}O -oxygen R_1 is measured at high frequency (where internal water molecules do not contribute) over a 60 K temperature range extending into the supercooled regime (Mattea, Qvist et al. 2008) for WT ProtL and for the obligate halophile Kx6E with and without 0.75 M NaCl.

For relaxation measurements at subzero temperatures, freezing of the solvent was prevented by dispersing the protein and reference solutions in μm -sized emulsion droplets (Mattea, Qvist et al. 2008) that are large enough so the oil-water interface has negligible effect on water dynamics, but small enough to allow supercooling to near the homogeneous nucleation temperature (Rasmussen and MacKenzie 1973). To detect any effects of the emulsion droplet interface, ^{17}O relaxation measurements were carried out at 300 K on all the three protein solutions before and after incorporation into emulsion droplets. No significant difference in R_1 was detected between the original solutions and the corresponding emulsified samples.

For the emulsified samples, the water ^{17}O longitudinal relaxation rate R_1 was measured at 81.3 MHz on a Varian Unity Plus 600 spectrometer as described above, except that the pre-cooled air used for temperature regulation was subjected to an extra drying step. During thermal equilibration (at least 20 min, but up to 40 min at the lowest temperatures), R_1 was measured repeatedly to ensure that temperature oscillations had been damped out and that stable R_1 values were being obtained. At sufficiently low temperatures, where a significant fraction of the emulsion droplets freeze by homogenous nucleation (as seen by a reduction of the integrated ^{17}O peak area), the released latent heat makes accurate temperature control difficult. Rather than attempting to correct for this effect, data obtained at temperatures where significant freezing effects were detected were simply discarded. At low temperatures, where the sample temperature is the principal source of error, the experimental uncertainty in R_1 was estimated by propagating a 0.1 K temperature error using the known temperature dependence of R_1 for bulk water. At 238 K, for example, this yields an R_1 uncertainty of 1.3 %, in good agreement with reproducibility tests at this temperature.

3.4.5. Exchange experiments

NMR can monitor chemical exchange processes such as exchange of labile protons with solvent, conformational exchange between different protein conformations, equilibria between folded and unfolded states of the protein, etc. Each process has a unique timescale, but NMR is versatile enough to offer different experiments at the appropriate dynamic range.

3.4.5.1. H/D exchange experiments

H/D exchange was monitored via the signal decay in a series of $^1\text{H},^{15}\text{N}$ -HSQC experiments (Figure M21). For this, samples of βLAC (1 mM $[\text{U-}^{15}\text{N}]$ βLAC in 20 mM phosphate buffer, pH 7.5) or ProtL (1 mM $[\text{U-}^{15}\text{N}]$ wt or Kx5E ProtL in 20 mM phosphate buffer, pH 6.0) were lyophilized and resuspended in D_2O containing the desired KCl concentration. The sample was immediately inserted into the NMR spectrometer and the HSQC series started (approximate lag time for setup: 30 minutes). Measurements were performed at 0.5, 0.75, 1.0 M KCl for wt ProtL and βLAC , and 0.5, 1.0, 1.5, 2.0 M KCl for Kx5E ProtL. $^1\text{H},^{15}\text{N}$ -HSQC experiments were recorded at 298K every 1.5 h up to 48-96 h on a 600 MHz (ProtL) or 800 MHz Bruker Avance III spectrometer (βLAC). For βLAC , a final $^1\text{H},^{15}\text{N}$ -HSQC was acquired after 1 month of sample storage at 298K to better cover the core residues with slower H/D exchange.

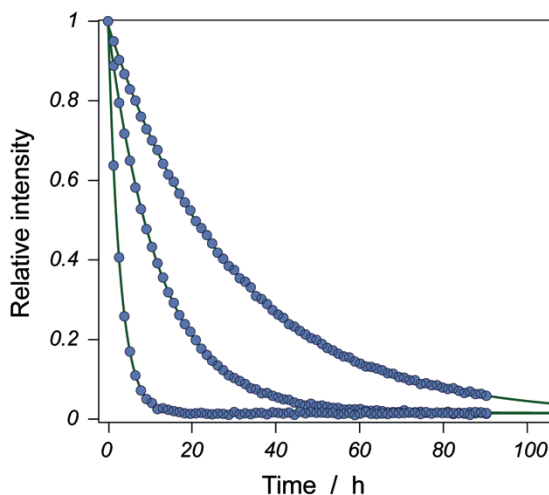


Figure M21. H/D exchange. Three examples of signal intensity decay during a H/D exchange experiment. Circles correspond to experimental points recorded and green lines to the monoexponential fitting.

Each series of spectra was processed with NMRpipe and integrated using a single peak list and in-house built scripts. For a given residue i , intensities were fitted to a single exponential decay with exchange rate $k_{\text{ex}}^{\text{H/D}}(i)$. H/D exchange in proteins involves two events: i) transition from a *closed*, solvent-protected conformation to an *open*, solvent-accessible conformation; followed by ii) exchange with the proton/deuteron of the solvent:

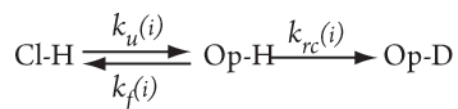


Figure M22. Scheme for the H/D exchange process. Open – closed equilibrium is coupled with the exchange of a protein proton with a solvent deuteron. Under EX2 regime for stable proteins,

$k_u(i)$ and $k_f(i)$ are the kinetic rate constants for unfolding and folding at residue i , respectively; $k_{rc}(i)$ is the intrinsic H/D exchange rate constant in the solvent-accessible conformation computed using the SPHERE web server (Bai, Milne et al. 1993) (<http://www.fccc.edu/research/labs/roder/sphere/>).

At pH > 5, H/D exchange in stable proteins typically occurs via the bimolecular EX2 mechanism (Englander and Hiller 2001), and it was previously shown that ProtL exchanges in the EX2 regime at pH 6.0 (Yi, Scalley et al. 1997). Under EX2 conditions, $k_{rc}(i) \ll k_f(i)$ and the overall $k_{ex}^{H/D}(i)$ rate adopts the form:

$$k_{ex}^{H/D}(i) = \frac{k_u(i)}{k_f(i)} \cdot k_{rc}(i) = K_{U-F}(i) \cdot k_{rc}(i)$$

Equation M16. Thermodynamic analysis of H/D exchange. Under EX2 conditions, the exchange rate is directly proportional to the equilibrium constant $K_{U-F}(i)$ and the exchange in the open (random coil-like) conformation $k_{rc}(i)$.

$K_{U-F}(i)$ is the thermodynamic equilibrium constant for local unfolding at residue i that can be directly converted into the free energy of local unfolding, $\Delta G_{U-F}^0(i)$:

$$\Delta G_{U-F}^0(i) = -R \cdot T \cdot \log \frac{k_{ex}^{H/D}(i)}{k_{rc}(i)}$$

Equation M17. Local stability. Formula for the calculation of local stability from experimentally measured exchange rates and intrinsic exchange in the open conformation.

Experiments with the dipeptide Ala-Ala at different concentrations of KCl (0 and 1M) revealed a negligible effect of KCl on $k_{rc}(i)$ rates (data not shown), which were therefore assumed to remain constant over the whole range of KCl concentrations. Contrarily, exchange rates measured for our test proteins show a strong dependence on the cosolute salt, and $\Delta G_{U-F}^0(i)$ varies linearly with the molar concentration of KCl yielding a *salt coefficient* $m_{KCl}^{H/D}(i)$ as proportionality factor:

$$m_{KCl}^{H/D}(i) = \frac{\partial \Delta G_{U-F}^0(i)}{\partial [KCl]}$$

Equation M18. Local salt stabilization. Local stability derived from H/D exchange depends linearly on KCl concentration.

The salt coefficient, $m_{KCl}^{H/D}(i)$, is a reporter for haloadaptation: a positive (negative) $m_{KCl}^{H/D}(i)$ value reflects protein stabilization (destabilization) in the presence of KCl. Residues decaying too fast ($k_{ex}^{H/D}(i) < 0.5$ h) or too slow ($k_{ex}^{H/D}(i) > 1$ year) could not be accurately determined with our experimental setup. Therefore, they were excluded from the analysis. A total of 104, 39 and 38 residues for bLAC, wt ProtL and Kx5E ProtL, respectively, were used to derive $m_{KCl}^{H/D}(i)$ values.

3.4.5.2. H/H exchange: CLEANEX

Solvent exposed NH and labile protons usually exchange at rates < 1 s, therefore falling out of the experimental range of H/D experiments. Fast HN/H₂O exchange rates of solvent exposed amide protons in the disordered haloSRC and uSRC polypeptides were measured on a Bruker Avance III 600 MHz NMR spectrometer using the Phase-Modulated CLEAN chemical EXchange (CLEANEX-PM) spin-locking experiment (Hwang, van Zijl et al. 1998) with a ¹⁵N,¹H-HSQC detection scheme and fully interleaved acquisition. Artifacts like intramolecular NOEs, TOCSY transfer from C α protons coincident with water frequency, or exchange-relayed NOE from fast exchanging hydroxyl or amine protons, are removed by CLEANEX sequence by using a NOESY-ROESY mixing scheme. In the slow motion regime, the ratio of cross-relaxation rates for ROE versus NOE is 2 to -1. Proper manipulation of magnetization trajectories to let the spins spend twice as much time along the z-axis than in the xy-plane can, therefore, cancel NOE and ROE contributions, leaving only pure exchange contributions. The effective exchange rates are quantitated by using initial slope analysis (Hwang, van Zijl et al. 1998):

$$\frac{V}{V_0} = k_{ex} \cdot \frac{e^{-R_{1B} \cdot \tau_{ex}} - e^{-(R_{1A} + k_{ex}) \cdot \tau_{ex}}}{R_{1B} + R_{1A} + k_{ex}} ;$$

Equation M19. Fast water-exchange rate constant. Initial slope analysis allows extracting water-NH exchange rates k_{ex} from the volumes of the exchange signals (V) with respect to their reference values (V_0). At low mixing times τ_{ex} , signal decay due to relaxation of water (R_{1B}) and NH (R_{1A}) is almost negligible.

where V and V_0 are the experimentally measured volumes under a peak after a exchange time τ_{ex} and in the reference HSQC respectively. Peak intensity is modulated by the exchange rate with the water k_{ex} and the relaxation decay rates for water (R_{1B}) and NH (R_{1A}).

Four CLEANEX-PM spectra at HN/H₂O exchange times (τ_{ex}) of 0, 15, 30 and 45 ms were acquired. For $\tau_{ex} > 50$ ms the HN/H₂O exchange signals are increasingly biased by longitudinal T_1 relaxation and polarization transfer from residual protein protons, compromising the accurate determination of exchange constants. A ¹⁵N[¹H] fHSQC was acquired as a reference spectrum. All spectra were recorded at 280K, the same temperature as used for the polypeptide assignments. A ¹⁵N labeled protein sample (1 mM concentration) was dissolved in 50 mM phosphate buffer at pH 7.0, 0.2 mM EDTA and 0.2 mM PMSF, and salt was added for measurements at 0.5, 1 and 1.5 M KCl.

3.4.5.3. Chemical exchange: EXSY

At a low range of KCl concentrations, both unfolded and folded states are detectable by NMR for the obligate halophilic *Kx7E ProtL*. The slow exchange between the folded and the unfolded conformation falls in the range of seconds – milliseconds. Being the folded and unfolded state peaks quite well resolved and separated, a ^1H – ^1H NOESY or EXSY experiment allows the calculation of the unfolding/folding kinetic rate constants (Jeener, Meier et al. 1979). This experiment yields a 2D spectrum with up to four signals arising per exchanging structural probe: two diagonal peaks accounting for the equilibrium populations of folded and unfolded state, and two cross peaks indicating exchange from folded to unfolded and viceversa.

For slow exchange and within the initial rate approximation, the equilibrium constant (K_{eq}), the exchange rate (k_{ex}), and the kinetic rate constants of folding (k_f) and unfolding (k_u) can be calculated from (Rule and Hitchens 2006):

$$K_{eq} = \frac{I_{FF}}{I_{UU}} = \frac{\chi_F}{\chi_U} = \frac{k_f}{k_u} \quad ;$$

$$k_{ex} = \frac{1}{\tau_m} \cdot \text{Ln} \left[\frac{1 + \frac{I_{UF}}{I_{UU}}}{1 - \frac{(1 - \chi_F) \cdot I_{UF}}{\chi_F \cdot I_{UU}}} \right] = \frac{1}{\tau_m} \cdot \text{Ln} \left[\frac{1 + \frac{I_{FU}}{I_{FF}}}{1 - \frac{\chi_F \cdot I_{FU}}{(1 - \chi_F) \cdot I_{FF}}} \right] \quad ;$$

$$k_{ex} = k_f + k_u \quad ; \quad k_f = \frac{k_{ex} \cdot K_{eq}}{1 + K_{eq}} \quad \text{and} \quad k_u = \frac{k_{ex}}{1 + K_{eq}}$$

Equation M20. Kinetic and thermodynamic analysis of EXSY experiments. Equations for the calculation of kinetic rate constants of folding (k_f) and unfolding (k_u) from the measured intensities of diagonal (I_{UU} and I_{FF}) and cross peaks (I_{UF} and I_{FU}) in an EXSY experiment.

where χ_F is the molar fraction of the folded species; I_{UU} and I_{FF} are the diagonal signal intensities for unfolded and folded states; I_{FU} and I_{UF} are the exchange cross signal intensities of the unfolding and folding transitions, respectively. For data analysis we used in-house MatLab[®] scripts.

NMR Exchange Spectroscopy (EXSY) experiments were measured at 600MHz and 298 K on 1mM *Kx7E ProtL* in 20 mM phosphate buffer (pH = 6) at different KCl concentrations. The ϵ -HN signal of tryptophan 47 in *ProtL* was used to monitor protein folding. The EXSY mixing time τ_m was set to 400 ms. Folding and unfolding kinetic rate constants were determined over a range of KCl concentrations (0 to 600 mM).

3.4.6. Chemical Shift Perturbation

Temperature coefficients (dependence of chemical shift with temperature) are mostly determined by the hydrogen–bonding network of the amide bond (Rothmund, Weisshoff et al. 1996), being considered reliable reporters on the thermal fluctuations and rigidity of the structure. On the other hand, chemical shift perturbation upon KCl addition informs about the changes in the chemical environment of the amide bonds in the presence of salt. These two magnitudes monitor changes in the native structure at a *per* residue level as a consequence of the presence of ions in solution.

To derive temperature dependence and KCl coefficients (Tadeo, Pons et al. 2007) for ProtL (*wt* and *Kx5E*), amide signal positions in the $^1\text{H},^{15}\text{N}$ -HSQC were measured at several temperatures (283, 288, 293, 298, 303, 308 and 313 K) at increasing KCl concentration (0.5, 1, 1.5 and 2 M of KCl). Similarly, KCl–induced chemical shift perturbation coefficients for uSRC and haloSRC were extracted from chemical shift measurements at 0, 0.5, 1 and 2 M KCl. Calibration was implemented via internal referencing to *trimethylsilylpropionate* (TMSP). A matrix of total 28 points *per* signal was obtained.

Statistical F–test analysis showed that a simple model considering independently the effect of both variables is enough to explain the experimental data. In other words, temperature dependence of chemical shifts is independent of salt dependence. Since chemical shifts varied linearly with temperature or KCl concentration, the slope *CSP* of the linear fitting can be used as a reporter for the chemical shift perturbation coefficients.

4. RESULTS

"El perseverante se hace inteligente"

Napoleón Bonaparte

4. MECHANISTIC INSIGHTS IN PROTEIN HALOADAPTATION

Adaptation to extreme halophilic environments occurred through modifications on the constituting proteome of salt adapted organisms. Opposite to mesophilic proteins, halophilic proteins remain stable and folded at high molar intracellular KCl concentrations. This is achieved by evolutionary selection of a biased set of amino acids that favours certain features: a lowered hydrophobic content and a high prevalence of short polar and negatively charged side chains. Thus, lysine content is lowered while halophilic proteins are enriched in Asp, Thr and Glu. Phenomenological and descriptive information about haloadaptation is readily available, but a complete mechanistic model is still missing. Our research has addressed the mechanism by which characteristic halophilic composition modulates the sensitivity to the cosolute. More specifically, we have explored the monotonic stabilisation of halophilic proteins upon salt addition, the obligate halophilicity and the salt modulation of enzymatic activity. To that end, we have used biophysical techniques in combination with NMR spectroscopy to investigate with atomic resolution a set of proteins, including natural and rationally designed halophilic proteins as well as folded and intrinsically disordered sequences, under *in vitro* conditions mimicking the cytosol of extreme halophiles.

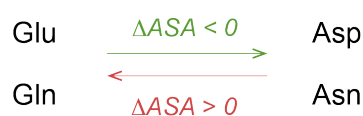
4.1. SETTING THE PROBLEM. PROTEIN STABILISATION UPON KCl ADDITION

KCl constitutes, by far, the most abundant ionic specie in the cytosol of salt accumulating halophiles (despite being NaCl the major extracellular specie). We have performed a thorough thermodynamic characterization of different proteins at a wide range of KCl concentrations in order to obtain the salt dependence of the stability of halophilic systems as compared to non-halophilic. Assuming a two state folding–unfolding equilibrium, we define protein stability as the difference between the chemical potentials of folded and unfolded states, reported as the change in free energy upon unfolding ΔG_{U-F}^0 . Such thermodynamic parameter can be readily obtained through thermal or chemical denaturations monitored by some spectroscopic technique, e.g. circular dichroism, fluorescence, NMR...

A number of systems were employed for the mechanistic studies on stability (see Materials and Methods 3.1 for detailed description). **WT ProtL** was chosen as a model for a non-halophilic folded protein. WT ProtL (64 aa) follows a canonical two-state folding equilibrium, which is chemically and thermally reversible and shows no detectable intermediates (Yi and Baker 1996, Scalley, Yi et al. 1997). The energetic profile of ProtL and the effect of the different cosolutes have already been thoroughly characterized (Tadeo, Pons et al. 2007): global stability of the WT was reported to be 7 kcal·mol⁻¹ at 298 K, ΔC_p (needed for the analysis of thermal denaturation curves) was reported to be 770 cal·mol⁻¹·K⁻¹ and insensitive to salt concentration.

Complementarily, rationally designed variants with lysine to glutamate substitutions (**KxnE ProtL**) were devised as counterparts with progressively increasing halophilic character. A thorough mutation study carried out in our lab already demonstrated that lysine removing mutations confer proteins a halophilic character (Tadeo, López-Méndez et al. 2009b). The effect on salt stabilization of systematic substitutions selectively changing protein properties was measured. Mutations decreasing hydrophobic area (such as ExD, ExN, KxE, KxD...) were found to increase KCl and NaCl stabilization, whereas those changing surface charge (QxE, ExQ, DxN, NxN...) had negligible influence (Scheme R1). Moreover, the effect of the mutations is cumulative and independent of the positions mutated. Thus, Kx5E ProtL is a good example for a halotolerant protein able to fold even in the absence of KCl. Further mutations such as Kx6E and Kx7E ProtL, turn ProtL into an obligate halophile unfolded at salt concentrations below ~ 0.5 and 1 M, respectively.

Changes in exposed surface:



Changes in surface charge:



Scheme R1. Changes in accessible surface area modulate the halophilic character of proteins. Mutations decreasing the ASA favour salt stabilization, as opposed to proteins increasing the exposed surface. Mutations altering the surface charge, however, show negligible effects in salt stabilization.

The β -lactamase from halotolerant *Oceanobacillus iheyensis*, **β LAC**, was chosen as a case study for natural halotolerant proteins. β LAC exhibits a typical halophilic composition, with nearly 25 % acidic content and a low number of lysines and bulky hydrophobic side-chains. β LAC is a large protein (261 aa) with a complex folding energy landscape: in our hands, thermal equilibrium is reversible at NaCl or KCl concentrations above 0.5 M, whereas chemical unfolding is non-reversible in any of the conditions experimentally tested. Subsequently, chemical denaturations can't report on the absolute free energy of unfolding of β LAC (ΔG^0_{U-F}). For this reason, thermal melts were used to probe the thermodynamic stability of β LAC. Its ΔC_p was estimated to be $4188.5 \text{ cal}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$ following empirical correlations with ΔA_{SA} and protein size (Myers, Pace et al. 1995).

Thermal denaturation curves monitored by circular dichroism were performed at different KCl concentrations and denaturation curves were analysed with *in-house* built MatLab[®] scripts (see Materials and Methods 3.3.2). Direct ΔG^0_{U-F} values cannot be extracted from thermal denaturation experiments, but relative changes with respect to a reference system, $\Delta\Delta G^0_{U-F}$, can be calculated if a reasonable estimation of the enthalpy ΔH_m is available (Equation M9 on Materials and Methods 3.3.2):

$$\Delta\Delta G^C_{U-F} = \Delta H_m^R \cdot \left(1 - \frac{T_m^R}{T_m^C}\right) ;$$

Equation R1. Stabilization calculated from changes in T_m . Simplified expression for the calculation of protein changes of unfolding free energy at salt concentration C measured as the increment in melting temperature to reference conditions.

For all the proteins studied, the salt dependence of protein stability was found to be linear and the slope serves as a measure of protein halophilicity (Fayos, Pons et al. 2005):

$$m_C = \partial \Delta G^0_{U-F} / \partial [C] ;$$

Equation R2. Salt stabilization of proteins. Slope for the monotonic linear salt stabilization upon salt addition, where C is the ionic specie (KCl). *m* value is used as a reporter for haloadaptation, since it contains all the contributions to the effect of the salt in the protein.

As expected, proteins show a **linear stabilization** upon KCl addition and the extent of salt stabilization is larger for halophilic proteins β LAC and Kx5E ProtL than for non-halophilic WT ProtL (Figure R1 and Table R1).

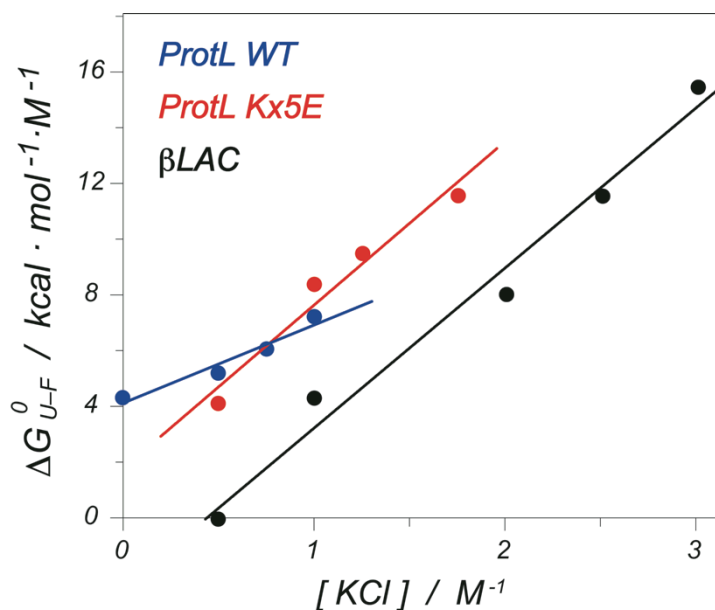


Figure R1. Global stabilization upon KCl addition for βLAC, WT and Kx5E ProtL calculated from thermal melts. Circles mark the relative stabilities calculated from experimental data, whereas lines represent the best linear fitting to KCl concentration. ΔG_{U-F}^0 increases linearly with cosolute addition, being the magnitude of the stabilization larger for halophilic sequences βLAC and Kx5E ProtL. βLAC values are $\Delta\Delta G_{U-F}^0$ referenced to that at 0.5 M KCl.

How can we define the halophilicity of a protein? We have devised a method to quantify the halophilic character of a given sequence based on the statistical deviation of its amino acid composition from that of a hypothetical halophilic proteome, reported as the sum of the square differences (χ^2) for each amino acid type:


$$\chi^2 = \sum^{aa} \left(\Delta aa_{h-m} \cdot \frac{n^{aa} - n_h^{aa}}{N} \right)^2 \quad \text{where} \quad n_h^{aa} = N \cdot \left(1 + \frac{\Delta aa_{h-m}}{100} \right) \cdot \frac{\overline{\%_{aa}^m}}{100} ;$$

Equation R3. Halophilic character of a sequence. Deviation from the consensus halophilic sequence calculated from the relative abundance of each residue type Δaa_{h-m} in halophilic proteome as compared to mesophilic proteome abundance ($\overline{\%_{aa}^m}$).

where, for a given amino acid type aa , n^{aa} corresponds to the number of residues among a total number of residues N . That number is compared to n_h^{aa} , the number of residues of a hypothetical homologue with an amino acid composition matching the average of halophilic sequences. The latter is calculated from the relative abundance coefficients Δaa_{h-m} (expressed in % on figure I2) and the average mesophilic protein composition ($\overline{\%_{aa}^m}$) following the expression on the right. The sum of the square differences is weighted by the relative abundance for each amino acid type Δaa_{h-m} in order to emphasize the residues relevant for haloadaptation. Thus, for instance, changes in alanine composition are irrelevant when evaluating the halophilic character of a protein, whereas changes in the number of lysines or aspartates are the major determinants. The smaller the χ^2 value, the more halophilic is a protein and viceversa. Ideally, a protein with the exact average amino acid composition would have a χ^2 value equal to zero.

For all the proteins tested, the slope of the KCl stabilization agrees well with the similarity to the consensus halophilic signature (Table R1). It confirms that **the amino acid composition is ultimately responsible for the salt induced stabilisation** of the folded conformation, and is consistent with other studies involving different cosolutes and previously reported in the literature (Ebel, Faou et al. 1999, Polosina, Zamyatkin et al. 2002, Tadeo, López-Méndez et al. 2009a).

Protein (MW / KDa)	χ^2	m_{KCl} kcal·mol ⁻¹ ·M ⁻¹
β LAC (29)	3.9	5.7 ± 0.5
HvCopG (5)	4.9	6.5 ± 1.5
HvLigN ade (77)	5.5	4.1 ± 0.5
HvLigN dea (77)	5.5	1.3 ± 0.5
Kx5E ProtL (7)	8.7	5.8 ± 0.9
EcGGBP (34)	17.4	0 ± 0
WT ProtL (7)	20.7	1.7 ± 0.7
haloSRC (10)	4.3	n.a.
uSRC (10)	15.9	n.a.



 χ^2

] disordered

Table R1. Stability and halophilicity. Deviation from a consensus average halophilic amino acid composition, χ^2 , and global stabilization upon KCl addition, m_{KCl} , obtained from thermodynamic data of global protein unfolding for a series of protein systems (Molecular Weight in kDa). HvCopG: translational repressor from halophilic archaea *Haloflex volcanii*. EcGGBP: Glucose/Galactose Binding Protein from *Escherichia coli*.

KCl induces a monotonic linear protein stabilization of halophilic proteins.

Halophilic amino acid sequence is directly responsible for KCl stabilization.

4.2. THE ROLE OF CHARGED GROUPS

The surface of halophilic proteins is enriched in acidic groups (Asp and Glu) that remain negatively charged at physiological pH (Paul, Bag et al. 2008). Focusing on this particular evolutionary feature, the halophilic model explains the stabilizing effect of salt on halophilic proteins through specific non-covalent interactions of carboxylate clusters with hydrated potassium cations (Dym, Mevarech et al. 1995, Mevarech, Frolow et al. 2000). Due to the electrostatic nature of these intermolecular interactions, neutral and charged forms of carboxylic acids are expected to have distinct K^+ affinities, which therefore depend on their individual acidity constants (pK_a). Reciprocally, both K^+ affinity and local environment (especially in folded proteins) modulate pK_a values. To disentangle this intricate interplay we used NMR to measure, with atomic resolution, the **changes in the individual pK_a values of ionizable groups as a function of KCl concentration**.

Chemical shift changes upon pH titration ultimately report on the populations of charged and neutral species for each ionizable group along the protonation equilibrium. Individual pK_a values can be extracted from the analysis of the chemical shift versus pH curves on the basis of the Henderson–Hasselbach classic formulation (Equation M11). Measurements of the changes in pK_a at increasing KCl concentration can be interpreted as increments in the free energy of the protonation / deprotonation reaction. The thermodynamic link between the ΔpK_a and the $\Delta G^{0, elec}(i)$ is established by (Tanford 1970):

$$\Delta\Delta G^{0, elec}(i) = -2.303 \cdot R \cdot T \cdot \Delta pK_a(i) \quad ;$$

Equation R4. Changes in the stability of the neutral and charged species in an acid – base equilibrium can be calculated from the experimentally detected shifts in the acidity constant pK_a .

The ion–residue interaction is proposed to be mainly electrostatic. Accordingly, changes in free energy associated to the acid–base equilibrium could be mostly attributed to changes in the stability of the charged specie. Moreover, since ionizable groups in the surface are expected to be charged at physiological conditions (Isom, Castañeda et al. 2011), the charged specie is the dominant at the experimental conditions tested and its changes in stability would be directly reflected in the global protein stability ΔG_{U-F}^0 .

In order to obtain the salt contribution to halophilic electrostatic stabilization, pH titrations and pK_a calculations were performed in the presence of increasing KCl concentrations on halophilic and mesophilic systems. The analysis has been divided into carboxylic and amine groups contributions. Furthermore, effects on the folded and unfolded conformations are investigated in order to extract the net contribution to global unfolding stability.

4.2.1. Electrostatic contribution of carboxylic groups

We have measured the chemical shift of Asp-C γ and Glu-C δ to track changes in individual carboxylate pK_a values as a function of KCl concentration (Figure R2, see Materials and Methods 3.4.3 for details) for three folded proteins: mesophilic WT ProtL (Figure S1), designed halotolerant Kx5E ProtL (Figure S2) and natural halotolerant β LAC (Figure S3). Complete assignment of WT and Kx5E ProtL carboxylate carbons was achieved, providing structural information at a *per* residue level for all the acidic groups. Unfortunately, due to the complexity and overlap of the spectrum of β LAC, α and β carbons were assigned but not carboxylates. Moreover, titration curves could be resolved only for 11 out of 65 carboxylates. For all residues measured, characteristic sigmoidal shapes are found (Figure R3). A single inflexion point is observed, which accounts for a unique transition determined by its pK_a value (Table S1). Under the conditions tested, KCl induced small but measurable shifts (inset in figure R3). Shift of a titration curve to lower pH values indicates a decrease in the pK_a value equivalent to an electrostatic stabilization of the charged specie, and *vice versa*.

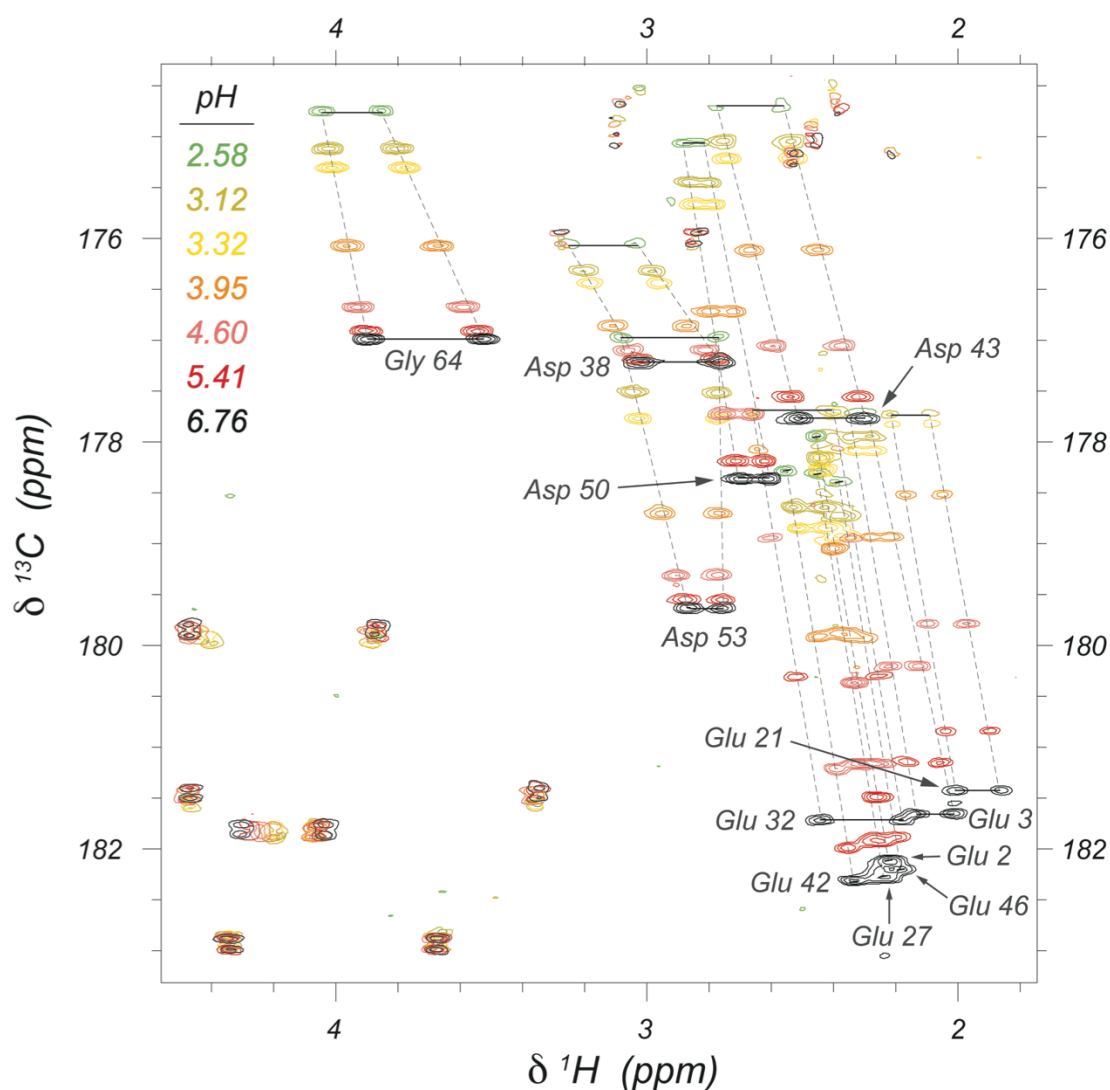


Figure R2. pH titration for WT ProtL. The above image shows an overlay of several HA(CA)CO spectra acquired at different pH values for WT ProtL in the presence of 1M KCl. Signals corresponding to carbonyl carbons of aspartic and glutamic acids shift as a function of pH.

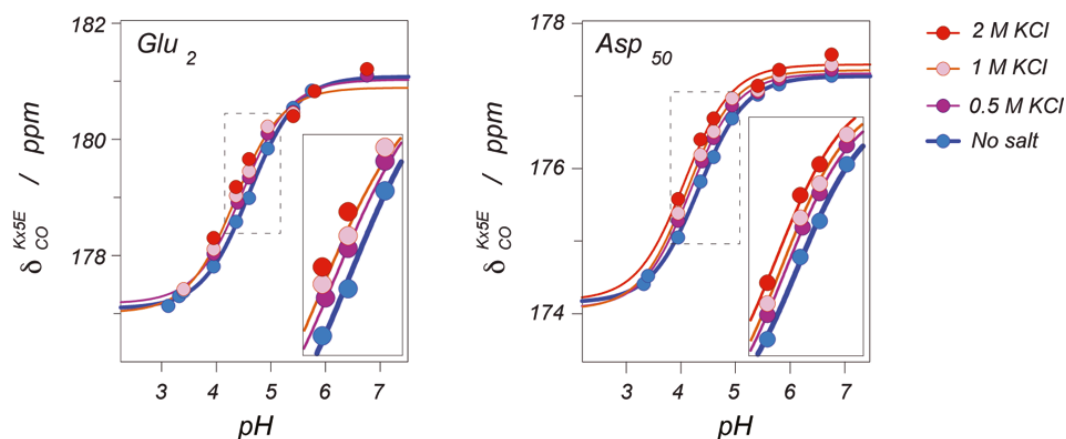


Figure R3. Close-up of the pH titration for two residues of Kx5E ProtL. The inflexion point of the titration curve corresponds to a pK_a value. Here, the pK_a is shifted towards lower pH values by KCl, which corresponds to a stabilisation of the deprotonated specie.

For all residues, only a **weak linear dependence** was observed between pK_a (as a measure for the free energy $\Delta G^{0, elec}$) and KCl salt concentration (Figure R4, left). This opposes the idea of a strong specific binding event, since no binding isotherm could be observed. In the best-case scenario, a binding model just saturated at our furthest experimental condition can be assumed (Figure R4, dashed lines). Fitting this dataset would yield a lower limit for the individual cation–protein **dissociation constant (K_d) in the order of $\sim 10^3$ M**, equivalent to an interaction energy of $\sim 4 \text{ kcal}\cdot\text{mol}^{-1}$ (disfavourable, notice the positive sign of the ΔG) or $\sim 3\cdot 10^{-20}$ J per each binding event. On the basis of a purely electrostatic interaction, the distance between both charges can be estimated as,

$$E = \frac{q^+ \cdot q^-}{4\pi \epsilon_0 \epsilon_r \cdot r^2} \quad ;$$

Equation R5. Electrostatic potential of interactions. Formula for the calculation of the electrostatic potential of the interaction between two opposite charges.

where q is the electron charge, ϵ_0 the vacuum permittivity $8.85\cdot 10^{-12} \text{ F}\cdot\text{m}^{-1}$, ϵ_r the relative permittivity (80 for a 1 M KCl solution) and r the distance between the two charges. The distance corresponding to the energy described above is in the order of $\sim 10 \text{ }\mu\text{m}$, clearly too far away from the protein scale ($\text{\AA} - \text{nm}$). This is a very simple estimation based on a rough approximation, but at the same time constitutes plain evidence that the assumption for a binding model is unsatisfactory. Individual carboxylate–cation interactions cannot be explained through specific binding, thus providing insight into the lack of K^+ ions bound to carboxylic groups in the crystal structures of halophilic proteins (Britton, Baker et al. 2006).

In an analogous way to the global m_{KCl} , the slope reports on the linear changes in stability induced by salt for a given ionizable residue. The proportionality factor $m_{\text{KCl}}^{elec}(i)$ (Table S1) corresponding to the linear electrostatic stabilization was derived from equation R4 (Tanford 1970) as:

$$m_{\text{KCl}}^{elec}(i) = \frac{\partial \Delta G^{0, elec}(i)}{\partial [\text{KCl}]} = -2.303 \cdot R \cdot T \cdot \frac{\partial \Delta pK_a(i)}{\partial [\text{KCl}]} \quad ;$$

Equation R6. Electrostatic KCl stabilization. Calculation of the m value from the electrostatic contribution to overall protein stability.

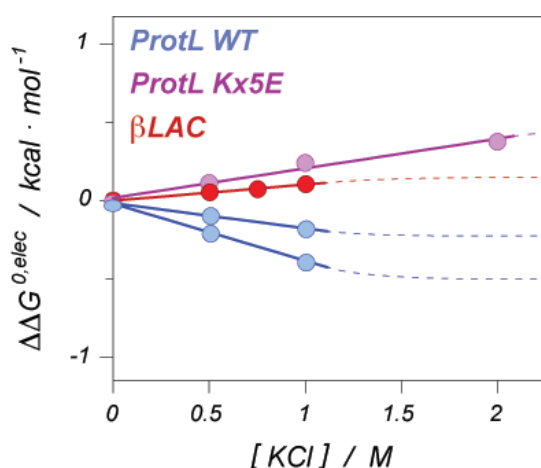


Figure R4. Stability of individual charged groups calculated from the pK_a values. Solid lines show the linear stabilizing (positive slope) or destabilising (negative slope) electrostatic interaction between the KCl and the negatively charged carboxylate. Dashed lines show the extrapolation to a putative binding model.

For the three folded proteins, the $m_{KCl}^{elec}(i)$ values were virtually identical for aspartates and glutamates, proving that the KCl stabilization is not affected by the nature of the side chain; instead, the interaction is inherent to the carboxyl moiety. In the context of halophilic adaptation, these m -values reflect the electrostatic interaction between the carboxylates and the K^+ cations. To our knowledge, this is the first quantitative report of such contributions for halophilic proteins. *The measured individual values are small: around $1 \text{ kcal} \cdot \text{mol}^{-1} \cdot M^{-1}$ at its best, consistent with the low affinity above reported.* Furthermore, negative local values < 0 illustrate the complex and ambiguous impact of the protein's charge network. The experimental $m_{KCl}^{elec}(i)$ values for ProtL can be placed on the structure (Figure R5) because the assignments of NMR signals are available (Tadeo, López-Méndez et al. 2009b). There is *no evidence from our data pointing to a specific arrangement of clusters of carboxylic groups forming a binding site for K^+ ions.* Instead, new acidic groups provoke small increases in salt stabilization of all surrounding acidic groups. Moreover, glutamates side chain orientation in Kx5E seems to be conserved as compared to the former lysines in WT ProtL (Tadeo, López-Méndez et al. 2009b).

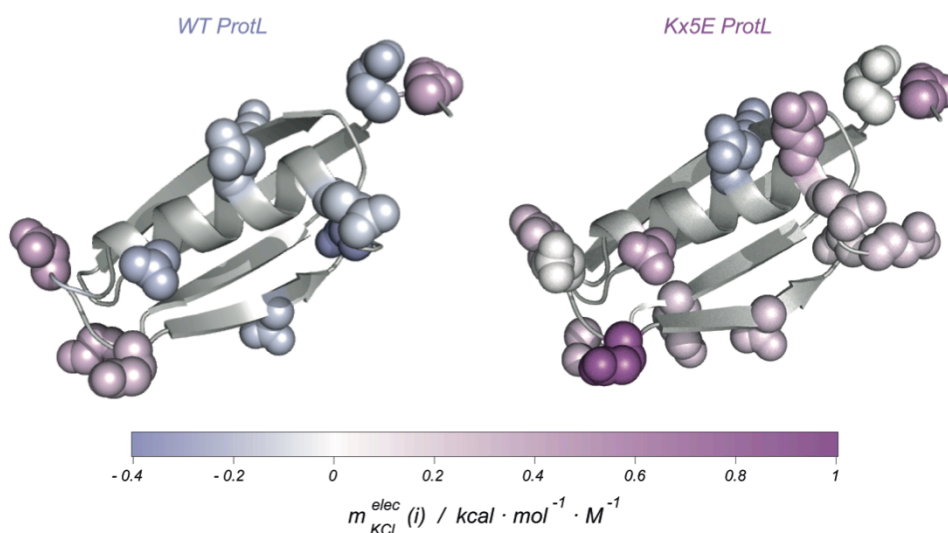


Figure R5. Structural analysis of the electrostatic stabilization of ProtL. Acidic residues on WT and Kx5E ProtL shown as spheres colour-coded according to the $m_{KCl}^{elec}(i)$ value: blue for destabilization and purple for stabilization. Side chain orientations of WT ProtL (PDB id: 1hz6) are maintained in Kx5E ProtL structure, resulting from a SWISS-PROT homology model on the NMR-structure of Kx6E ProtL (PDB id: 2kac) determined at our lab (Tadeo, López-Méndez et al. 2009b).

Results are summarized in the histograms on figure R6. An individual carboxyl group can thereby show either positive (stabilizing) or negative (destabilizing) $m_{KCl}^{elec}(i)$ values (Figures R4 – R6) for its interaction with solute cations. Overall, the mesophilic WT ProtL shows a bimodal distribution of values around zero, consistent with the observed minimal changes in global stability induced by KCl (Figure & Table R1). The designed halotolerant protein Kx5E ProtL has more carboxylic groups with positive $m_{KCl}^{elec}(i)$, resulting in a small net positive contribution $\overline{m_{KCl}^{elec}}$ to the protein's stabilization by KCl (Figure R6). KCl effect on natural halotolerant β LAC is similar to Kx5E ProtL.

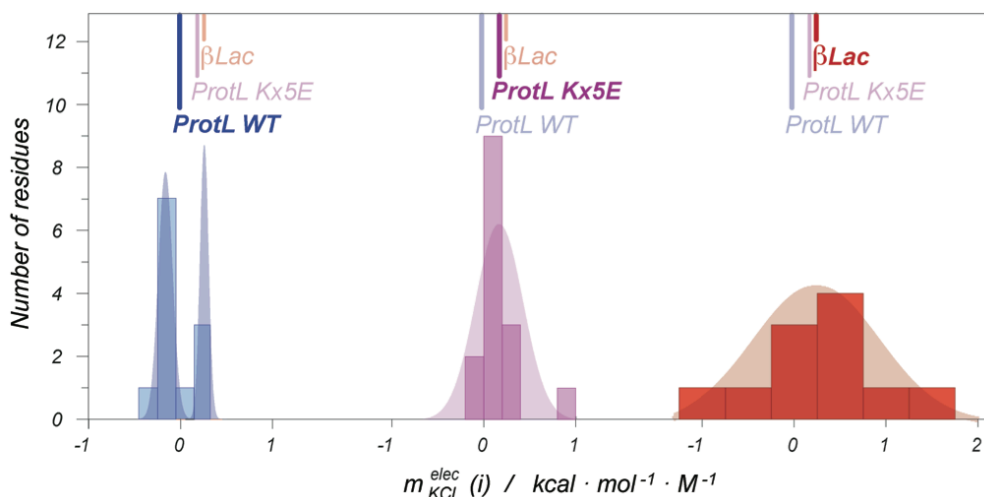


Figure R6. Electrostatic contribution of carboxylic groups. Coloured bars represent the histograms of the experimentally measured $m_{KCl}^{elec}(i)$ values for the three folded proteins studied. Shaded areas correspond to the calculated normal distribution representative of the experimental data for each protein. Solid lines in the upper part of the plot mark the average of the values.

Our data show that the increase of acidic residues slightly enhances KCl stabilization of negative charges due to electrostatic interactions. Again, the enhancement is small and unspecific, not consistent with a model of specific cation binding to arrangements of carboxylates. Alternatively, a model in which the overall increase in negative potential of protein surface modulates the **weak, unspecific binding** of K^+ cations seems more plausible.

4.2.2. Electrostatic contribution of the unfolded state

Regarding halophilic adaptation, salt induced increment of overall protein stability results from the difference between the folded and the unfolded states. Therefore the net electrostatic stabilization is obtained upon subtraction of the unfolded state contribution:

$$m_{KCl}^{elec} = \sum_i [m_{KCl}^{elec, F}(i) - m_{KCl}^{elec, U}(i)] \quad ;$$

Equation R7. Electrostatic contribution to native stabilization. Total electrostatic stabilization comes from the subtraction of the individual electrostatic contributions of the folded state and unfolded states for each of the ionizable groups of the protein.

The unfolded state of ProtL and β LAC is accessible usually through denaturants or high temperature. Such extreme conditions would difficult the comparison with experimental data for folded proteins. Specifically, the addition of denaturants could interfere with the effects of other cosolutes and would certainly complicate the experimental setup due to cosolubility issues. That's why we decided to use intrinsically disordered uSRC and haloSRC as models for the unfolded state of mesophilic and halophilic proteins respectively.

The pH titrations and the pK_a measurements were performed on disordered proteins following the same analysis reported above (Table S2). Both uSRC and haloSRC $m_{KCl}^{elec, U}(i)$ values for local electrostatic stabilization of negative charges were very similar and close to zero (Figure R7). Disordered chains are more extended and have more freedom to populate conformations with lower negative charge density and repulsion. As a result, they are expected to be less sensitive to K^+ electrostatic screening, as confirmed by our data.

Average $\overline{m_{KCl}^{elec, U}}$ was used as an approximation for the KCl-induced electrostatic contribution for the stability of individual residues from mesophilic and halophilic unfolded states (-0.05 ± 0.1 for uSRC and $+0.02 \pm 0.2$ kcal·mol⁻¹·M⁻¹ for haloSRC). Substitution in equation R7 yields the overall electrostatic contribution of carboxylic groups to KCl stabilization:

$$m_{KCl}^{elec} = \sum_i [m_{KCl}^{elec, F}(i) - \overline{m_{KCl}^{elec, U}}] \quad ;$$

Equation R8. Electrostatic contribution to global stability. The average of the $m_{KCl}^{elec, U}(i)$ values calculated from experimental data for uSRC and haloSRC is used as an approximation of the individual unfolded state contributions in equation R7.

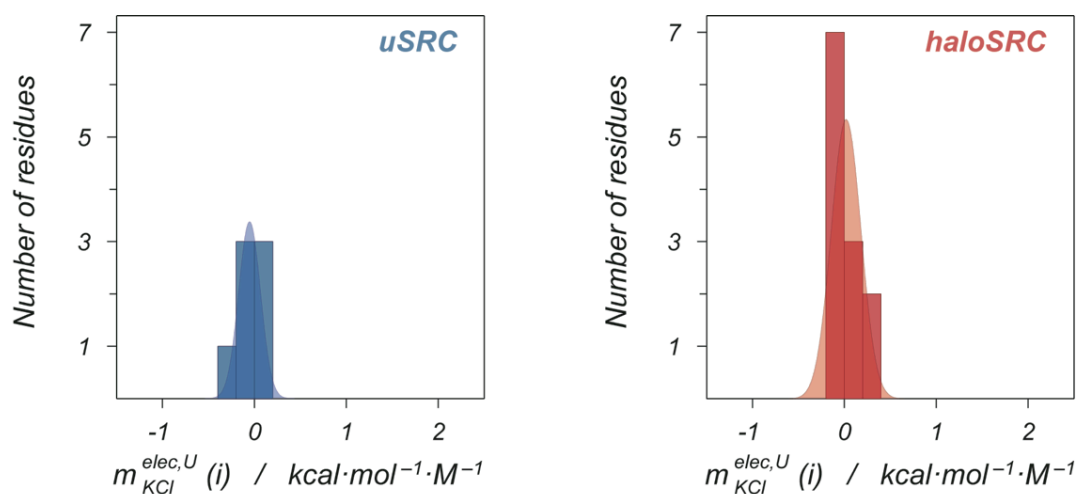


Figure R7. Electrostatic contribution of individual carboxylic groups in the unfolded state. Histograms of the $m_{KCl}^{elec, U}(i)$ values for the acidic groups present in disordered uSRC (left) and haloSRC (right). Values are very similar between both unfolded proteins and close to zero.

4.2.3. Electrostatic contribution of Lysines

A reduction in the number of lysines is a ubiquitous evolutionary feature for halophilic proteins. Unlike acidic groups, the running hypothesis is that they destabilize proteins in salted environments due to its long hydrophobic side chain (Britton, Baker et al. 2006, Tadeo, López-Méndez et al. 2009b) and not that much because of electrostatic interactions with their positive charge. This interaction between Cl^- anions and lysine amino groups is studied by determining its pK_a values as a function of KCl concentration (Table S3).

Two effects can be differentiated (Figure R8, left): a larger stabilization of positively charged specie at low salt concentrations and a much smaller dependence at high molar salt concentrations. The earlier stabilization can be explained through specific anion binding sites, which have already been reported and suggested to saturate at relatively low salt concentrations (Ramos and Baldwin 2002), whereas the later stabilization occurs through weak unspecific anion–amino interactions equivalent to the ones reported before for acidic groups. Halophilic organisms accumulate high molar salt concentrations in the cytoplasm. Therefore, we would only consider the changes at conditions close to the halophilic physiological conditions of organisms ($> 1 \text{ M}$) for the study of haloadaptation, discarding the specific contributions at lower salt concentrations.

WT ProtL lysine amine pK_a values show a very small (almost negligible) dependence on KCl (figure R8), as so does Kx5E ProtL. That stabilization is still very low even when considering also the data at low salt concentrations (dashed lines). Histogram on the right plot of Figure R8 shows the narrow distribution around zero for the electrostatic m -values of WT ProtL. In contrast to acidic residues, which spread up to $1 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$, the largest value measured for lysines is below $0.05 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$. The practical **absence of interactions between positively charged amines and Cl^- anions** at molar salt concentrations could provide an explanation for the low lysine abundance in halophilic proteins.

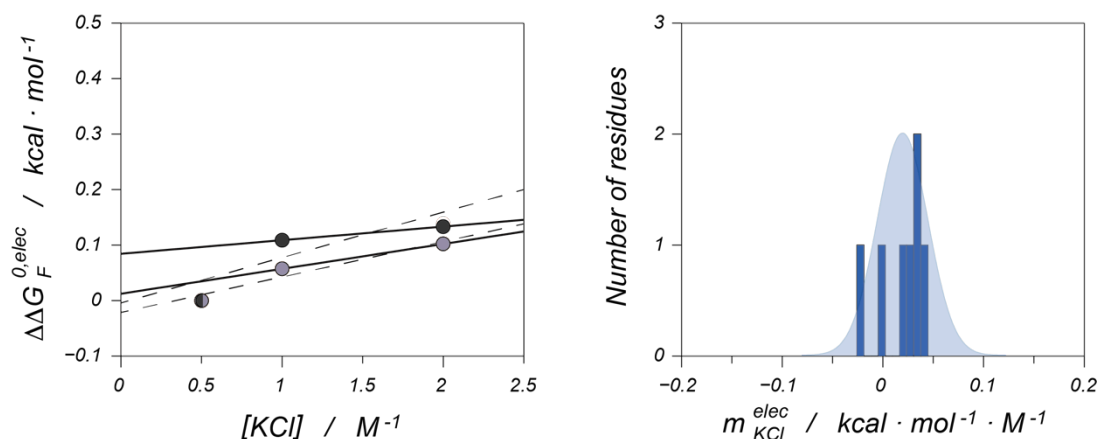


Figure R8. Lysine electrostatic contribution. Summary of the $m_{\text{KCl}}^{elec}(i)$ values for the Lysine amino groups present in WT ProtL. Left plot represents the calculated $\Delta\Delta G_F^{0,elec}(i)$ values and the least square difference fitting to a linear equation. Solid lines only consider the two points at higher KCl concentrations, whereas the dashed lines take into account the extra stabilization at low salt concentration, which is presumably caused by anion–specific binding sites. Slopes are shown in the histogram on the right. Values are narrowly clustered around zero.

4.2.4. Global thermodynamic contribution of electrostatic interactions

The overall contribution of the negatively charged surface to the stability of halophilic proteins in the presence of KCl can be computed as the sum of the differences between stabilizing and destabilizing effects of each charged group in the folded and unfolded state (Equation R8). Figure R9 shows, for WT and Kx5E ProtL, the KCl induced electrostatic contribution $m_{\text{KCl}}^{\text{elec}}$ for carboxyl groups (purple) and lysines (blue) relative to the total protein stabilization by KCl (white), where m_{KCl} values were derived from chemical and thermal denaturation experiments (Figure & Table R1). For WT ProtL the contribution from charged surface groups is negligible, consistent with its mesophilic character and with the narrow distribution of individual values around zero (Figure R6). For the halotolerant Kx5E ProtL, however, the sum of all contributions results in a large electrostatic stabilization that accounts for 47% of the overall salt induced effect. Thus, while a single carboxyl group may slightly stabilize or destabilize a protein, a significant electrostatic stabilizing net contribution results from the cumulated unspecific weak interactions between solute cations and the many carboxylate groups in halophilic proteins. Stabilization arises mainly from the unspecific shielding of negative charges accumulated in the folded state. Lysine amine groups, on the other hand, contribute at a much lesser extent to salt stabilization at high salt concentrations.

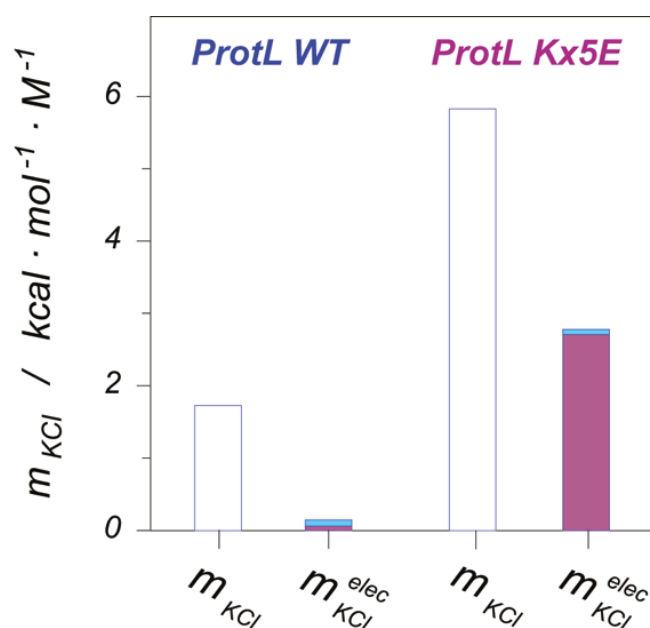


Figure R9. Contribution of electrostatics to global stabilization. Global KCl stabilization from chemical denaturation experiments (white bars) versus electrostatic contribution to haloadaptation ($m_{\text{KCl}}^{\text{elec}}$) from carboxyl groups (purple bars) and lysine amine groups (blue bars) for WT and Kx5E ProtL.

Complementary, absolute pK_a values at low salt concentration are informative about intrinsic protein stability. Introduction of new carboxylates modifies the charge network of the surface and, as expected, increases electrostatic repulsion between alike charges raising the pK_a values. In the absence of salt (0 M KCl) the average pK_a value for WT ProtL is 4.11, whereas it goes up to an average of 4.44 for Kx5E ProtL. The excess of 0.33 pH units can be interpreted as a destabilization by almost 4 kcal · mol⁻¹ (Equation R4) of the halophilic Kx5E

with respect to the WT ProtL. To a lesser extent, electrostatic repulsion also occurs in the unfolded conformations. The average pK_a raises 0.18 pH units (from 4.61 in uSRC to 4.79 in haloSRC). The equivalent calculated destabilization is $1.7 \text{ kcal}\cdot\text{mol}^{-1}$. Not surprisingly, electrostatic repulsion is lower in the unfolded than in the folded conformation, probably due to the entropic chain freedom that allows the protein to find less constrained orientations between side chains and / or more extended conformations. Our data show a $4.4 - 1.7 = 2.7 \text{ kcal}\cdot\text{mol}^{-1}$ difference in global stability between halophilic and mesophilic states due to electrostatic repulsion. This is consistent with previous studies on ProtL mutants which report a stability of about $5 \text{ kcal}\cdot\text{mol}^{-1}$ for Kx5E ProtL as compared to the $7 \text{ kcal}\cdot\text{mol}^{-1}$ of the WT ProtL (Tadeo, López-Méndez et al. 2009b). It is difficult to accurately predict the effect of cumulative mutations on protein stability and break it down to all the different contributions (side-chain packing, intramolecular interactions, hydrophobic effect...). Nevertheless, in our opinion our data support the notion that electrostatic repulsion is the major contributor to the lower intrinsic stability (in the absence of salt) of folded halophilic proteins (Polosina, Zamyatkin et al. 2002), eventually leading to protein unfolding and **obligate halophilicity**.

K^+ -carboxylate interactions are weak and unspecific.

The stabilizing effect arises from the screening of electrostatic repulsive forces.

Lysines do not show appreciable interactions with Cl^- anions.

Electrostatic interactions are not sufficient to account for all the salt stabilization.

4.3. STUDY OF THE HYDRATION SHELL OF HALOPHILIC PROTEINS

A recurring notion in the search for the molecular determinants of haloadaptation (halophilic solvation-stabilization model already discussed in Introduction 1.1.4) is that **halophilic proteins' hydration shell is tighter than that of their mesophilic counterparts**. This stronger binding of solvating water molecules is due to the high surface density of negatively charged aspartate and glutamate side-chains (Madern, Ebel et al. 2000, Mevarech, Frolow et al. 2000). Moreover, this extensive hydration is widely believed to protect halophilic proteins against unfolding and aggregation at high salinity.

To shed further light on the hydration properties of halophilic proteins as compared to mesophilic proteins, we have measured the ^{17}O nuclear spin relaxation rate of water molecules in protein solutions at different salt concentrations. Oxygen-17 spin relaxation directly and selectively monitors the rotational dynamics of water molecules, without complications from labile hydrogens or interferences from the protein, reporting on the single-molecule rotational correlation time without contributions from other relaxation sources or intermolecular couplings (Halle, Denisov et al. 2002).

Two different experimental approaches were followed:

- i. by recording the longitudinal spin relaxation rate (R_1) as a function of resonance frequency (the **magnetic relaxation dispersion** profile, MRD) information about long-lived and internal water molecules as well as about the more mobile water molecules interacting with the external protein surface was sought (Halle, Denisov et al. 2002),
- ii. R_1 at higher frequency (where internal water molecules do not contribute) was measured over a wide temperature range extending into the supercooled regime to characterize the surface hydration dynamics in more detail (Mattea, Qvist et al. 2008).

Here we report the ^{17}O **MRD profile and the temperature dependence of R_1** over a 60 K interval for WT ProtL and the Kx6E mutant with (folded) and without (unfolded) salt. This work was done in collaboration with Prof. Bertil Halle and Dr. Johan Qvist from Lund University (Lund, Sweden) and resulted in the publication:

“Hydration dynamics of a halophilic protein in folded and unfolded states” by Johan Qvist, Gabriel Ortega, Xavier Tadeo, Óscar Millet and Bertil Halle in *The Journal of Physical Chemistry. B* (2012) Vol 116(10): 3436-44.

4.3.1. Protein and internal water dynamics

Water ^{17}O longitudinal relaxation rate for a dilute solution of a protein containing N_i internal water molecules with residence time $\tau_i \ll 1 \mu\text{s}$ is determined by the expression (Halle, Denisov et al. 2002, Mattea, Qvist et al. 2008):

$$R_1(\omega_0) = R_1^0 \cdot \left(1 + \frac{v_{dyn}}{N_W}\right) + \frac{N_i S_i^2}{N_W} \cdot \frac{\omega_Q^2}{5} \cdot \left[\frac{\tau_C}{1 + (\omega_0 \tau_C)^2} + \frac{4 \tau_C}{1 + (2\omega_0 \tau_C)^2} \right] ;$$

Equation R9. Water ^{17}O longitudinal relaxation rate. Expressed as the contributions from molecules in the hydration layer (first term) and from the slower internal water molecules (second term).

where N_W is the known water/protein mole ratio, R_1^0 is the relaxation rate measured in a separate protein-free reference sample, ω_0 is the ^{17}O resonance frequency and $\omega_Q = 7.61 \cdot 10^6 \text{ s}^{-1}$ is the ^{17}O nuclear quadrupole frequency in an immobilized water molecule. *First term* in equation R9 contains the contributions from water molecules in the bulk solvent (unaffected by the protein) and in the hydration layer, characterized by the dynamic hydration number,

$$v_{dyn} = N_H \cdot (\xi - 1) ;$$

Equation R10. Dynamic hydration number. Behavior of water molecules in the hydration layer depends on the dynamic hydration number, which in turn depends on the total number of water molecules in the hydration layer N_H and their dynamic perturbation factor ξ .

where N_H is the number of water molecules in the hydration layer and ξ is the average dynamic perturbation factor (DPF), defined as the average rotational correlation time of the N_H hydration water molecules $\langle \tau_H \rangle$ divided by the same quantity for bulk water in the reference sample, τ_{bulk} :

$$\xi = \frac{\langle \tau_H \rangle}{\tau_{bulk}} ;$$

Equation R11. Dynamic Perturbation Factor (DPF). Expressed as the ratio between hydration and bulk water molecules average rotational correlation times.

Second term in equation R9 accounts for the contribution from the N_i internal water molecules with mean-square orientation order parameter S_i^2 and overall correlation time τ_C . The latter is determined by the rotational correlation of the protein τ_P and the mean residence time τ_i of the internal water molecules according to:

$$\tau_C = \left(\frac{1}{\tau_P} + \frac{1}{\tau_i} \right)^{-1} ;$$

Equation R12. Internal water molecules correlation time τ_C . Determined by both the protein rotational correlation time τ_P and the mean residence time τ_i .

The water ^{17}O longitudinal relaxation rate R_1 exhibits dispersion in the MHz frequency range only if the protein contains one or more internal water molecules with mean residence time τ_i in the ns- μs range, as characterized by the first term in equation R9. Within the context of halophilic proteins, water molecules tightly bound to ion binding sites could be expected to have residence times in the order of ns; therefore giving rise to larger dispersions than mesophilic proteins.

Figure R10 shows the ^{17}O MRD profiles recorded at 300 K from aqueous solution of ProtL (green) and the mutant Kx6E (blue) with and without 0.75 M NaCl. The dispersions from WT and salt-folded Kx6E are well described by equation R9. The values of the three adjustable parameters are given in Table R2. The values of 0.7 ± 0.1 (WT) and 0.8 ± 0.2 (Kx6E)

obtained for $N_i \cdot S_i^2$ (the product of the long-lived internal water molecules and their mean-square order parameter) are very similar and demonstrate that both proteins should contain, at least, one long-lived water molecule. Such long-lived water molecules are invariably buried in deep pockets or cavities within the native protein structure. No fully buried waters are seen in the crystal structure of WT ProtL (PDB id: 1hz6), but a water molecule resides hydrogen bonded to Thr-30 and Gly-55 in a deep surface pocket between the C-terminal beta hairpin and the N-terminal end of the helix (Figure R11). As a result, it may have residence times in the nanosecond range that could account for the observed ^{17}O dispersions. The correlation time deduced from the MRD ($\tau_C = 3.8 \pm 0.3$ ns) matches the correlation time of the protein ($\tau_P = 3.9$ ns). In view of equation R12, this implies that the residence time τ_i is, indeed, much longer than 4 ns. Conversely, the shorter correlation time deduced for Kx6E ($\tau_C = 2.3 \pm 0.5$ ns) implies that the residence time of the water molecule is shortened by the mutations ($\tau_i = 5.4 \pm 3.0$ ns assuming the same correlation time for the protein). This shortening of the residence time is plausible since one of the substituted residues, Lys-54, is close to the probable location of the water molecule (Figure R11). This mutation-induced reduction of the internal water residence time is the reason for the smaller amplitude of the ^{17}O dispersion step for Kx6E (Figure R10, blue) as compared to WT ProtL (green):

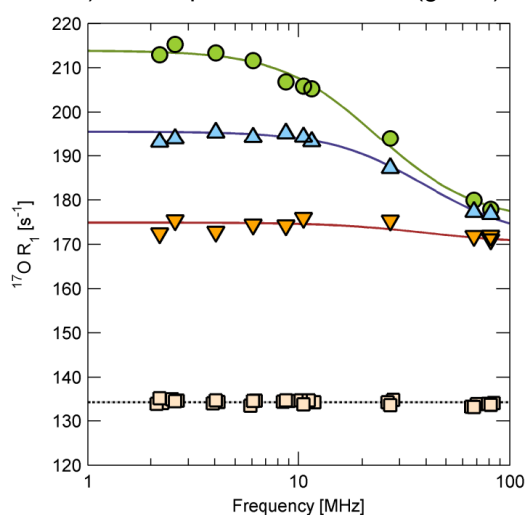


Figure R10. Relaxation–dispersion profiles. MRD profiles for WT (green, circles), Kx6E with (blue, triangles) and without (orange, inverted triangles) 0.75 M NaCl, and reference protein-free samples (pale yellow, squares) at 300 K. Lines represent least-square fitting to equation R9. The relaxation rates R_1^0 from the three different reference samples (squares) are slightly displaced in frequency for visual clarity. For Kx6E with salt and the corresponding salt-containing reference sample, the small frequency-independent salt induced R_1 enhancement (4.0 s^{-1}) has been subtracted.

Property	WT ProtL	Kx6E, folded (0.75 M NaCl)	Kx6E unf. (0 M NaCl)
C_P (mM)	12.4	8.2	7.4
N_W	4005	6332	7042
v_{dyn}	1216 ± 30	1102 ± 82	1138 ± 31
$N_i \cdot S_i^2$	0.70 ± 0.06	0.79 ± 0.23	0.15 ± 0.04
τ_C (ns)	3.38 ± 0.27	2.26 ± 0.47	$[2.26]^1$
χ^2_{red}	2.07	0.44	0.60

¹ Frozen during fit

Table R2. Summary of ^{17}O MRD experiments. Sample characteristics and results derived from fit of ^{17}O MRD data at 300 K to Equation R9.

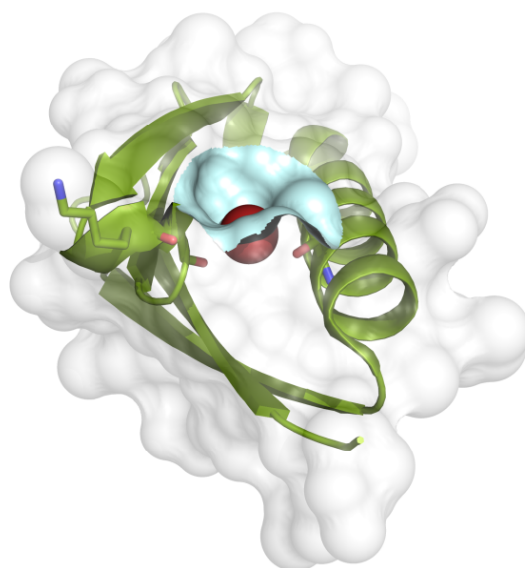


Figure R11. Structural interpretation of MRD data. The putative long-lived water molecule (red sphere) residing in a surface pocket (cyan surface) of ProtL (green cartoon, white transparent surface). Shown in stick representation are the two residues (Thr-30 and Gly-55) that are H-bonded to this water molecule and a nearby residue, Lys-54, which is mutated in Kx6E.

Unfolded proteins are not expected to give rise to a ^{17}O relaxation dispersion since they lack the persistent structure required to trap water molecules for times longer than ~ 1 ns. Accordingly, Kx6E without salt shows a barely significant dispersion (Figure R10, orange). Previous studies indicate that a minor fraction of the protein (10–20 %) remains folded in the absence of salt (Tadeo, López-Méndez et al. 2009b), which could explain the small dispersion observed. Therefore, it can be concluded that the water molecules interacting with the unfolded protein are very mobile, with sub-nanosecond correlation times, as expected.

Relevant conclusions to haloadaptation can be extracted from this analysis. *First*, there is **no evidence of a tighter hydration layer** for halophilic Kx6E in the presence of KCl as compared to the mesophilic WT ProtL. Actually, the analysis shows that the residence time of the water molecules must be very similar for both proteins. Moreover, introduction of the mutations *decreases* the observed residence time, but that is probably due to local effects on the surface hydrophobic pocket. *Second*, observed dispersions arise from a single water molecule trapped in a hydrophobic pocket, thus **ruling out the presence of a significant number of water molecules tightly bound** to carboxylic groups with residence times ≥ 1 ns. As a reference, water molecules trapped inside the protein and strongly H-bonded would show correlation times near protein correlation time with residence times $\gg 1$ ns (Modig, Liepinsh et al. 2004).

4.3.2. Dynamics of the hydration layer of halophilic proteins

A third parameter obtained from the ^{17}O MRD fits in figure R10 is the *dynamic hydration number* (ν_{dyn} , equation R10), which involves the number of water molecules in the hydration layer and the dynamic perturbation factor. The *hydration number* N_{H} was estimated by dividing the solvent-accessible surface area ASA, by the average solvent-accessible surface area occupied by a single water molecule (10.75 \AA^2), as inferred from molecular dynamics simulations (Mattea, Qvist et al. 2008). For the folded proteins, ASA was computed with the program GetArea (Fraczkiewicz and Braun 1998), using a probe radius of 1.7 \AA , the default set of van der Waals radii and the (first model of) solution structures *2ptl* (Wikstrom, Drakenberg et al. 1994) for WT ProtL and *2kac* (Tadeo, López-Méndez et al. 2009b) for Kx6E. The *dynamic perturbation factor*, DPF or ξ , is expressed in Equation R11 as the ratio of correlation times in the hydration layer and in the bulk solvent. It is, thus, a measure of the restriction (or increase) of the dynamics of solvating water molecules; i.e. the tighter the water molecules are in the protein hydration shell, the larger the ξ value.

The resulting N_{H} values (Table R3) were then combined with the experimentally derived ν_{dyn} values (Table R2) to obtain $\xi = 4.0 \pm 0.1$ for ProtL and 3.6 ± 0.2 for Kx6E. From previous ^{17}O MRD studies of 11 monomeric proteins, $\xi = 4.2 \pm 0.4$ (Halle 2004). WT ProtL is thus a typical protein with regard to hydration dynamics, whereas the Kx6E mutant has a slightly more *mobile* hydration layer (only a minor part of the 10 % reduction in DPF for Kx6E can be explained by the 3 % salt-induced increase of τ_{bulk} in the 750 mM NaCl solvent, as inferred from R_1^0 for the bulk solvent reference samples). Clearly, the large number of carboxylate groups and the large negative net charge density on the surface of Kx6E **does not result in a reduction of water dynamics**, as one might expect for a "highly ordered, multilayered solvation shell" of "immobile" and "tightly bound" water molecules proposed by the traditional halophilic model. In addition, we note that the DPF reported here is deduced under the well-founded assumption that the effect of the protein on water dynamics is essentially confined to the first layer of (N_{H}) water molecules covering the protein surface (Halle 2004, Mattea, Qvist et al. 2008, Marcus 2009). If the perturbation had a longer range, so that more water molecules were affected, then the dynamic hydration number obtained from the ^{17}O MRD data would yield an even smaller average DPF.

To better characterize the hydration dynamics, we also investigated the temperature dependence of the dynamic hydration number. To do so, R_1 was measured at high frequency (81.3 MHz) and the small internal-water contribution was subtracted off (obtained from equation R9 and the parameter values in table R2). At this high frequency, the internal-water correction is small at room temperature and negligible at the lower temperatures, therefore *only water molecules in the hydration shell are monitored*. Denoting the corrected R_1 value by R_1' , we then estimate the dynamic hydration number from equation R9 as $\nu_{\text{dyn}} = N_{\text{W}} \cdot (R_1' / R_1^0 - 1)$. The results are shown in figure R12.

The temperature dependence of the DPF (and according to equation R10, the hydration number ν_{dyn} as well) was analysed with a model that describes the broad distribution of rotational correlation times, τ_{H} , of the heterogeneous N_{H} hydration water molecules by assuming a power-law distribution for the rotational correlation time, $\xi = f(\tau_{\text{H}}) \propto \tau_{\text{H}}^{-\nu}$, as

suggested by molecular dynamics simulations. In this model (Mattea, Qvist et al. 2008) the DPF is determined by two parameters: the power-law exponent ν and the activation energy E_H^- for rotation of the most mobile water molecules in the hydration layer (Table R3). The DPF depends weakly on frequency at low temperatures, because τ_H for the slowest hydration waters is then comparable to $1/\omega_0$. The quantity $\xi(\omega_0)$ derived by means of equation R9 and R10 is therefore an apparent DPF, which in general is slightly smaller than the true DPF (Equation R11). For simplicity, the minor difference between the apparent and true DPFs will be disregarded in the following, although it was fully included in the quantitative data analysis. A maximum in DPF and ν_{dyn} is observed for all three samples. Neglecting any temperature dependence in N_H , the maximum occurs at temperature where the activation energy for water rotation is the same in the hydration layer as in the bulk solvent. At temperatures below the maximum, hydration layer thus rotates with lower activation energy than in the bulk solvent and *vice versa*.

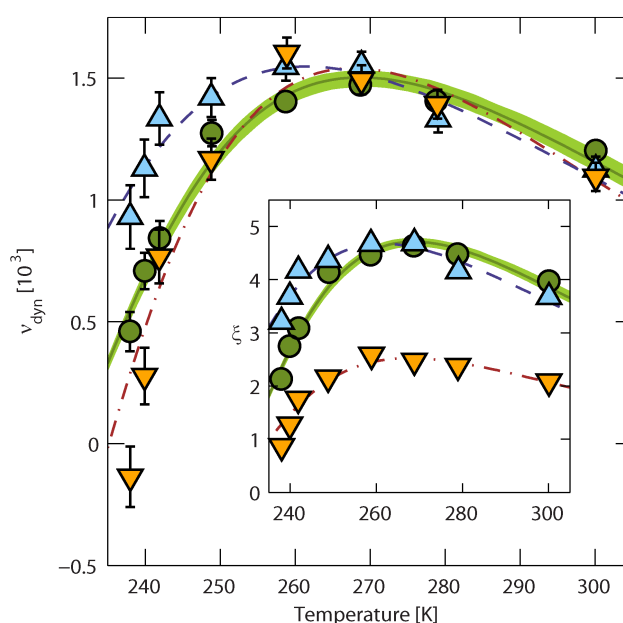


Figure R12. Temperature dependence of the dynamic hydration number, ν_{dyn} , for ProtL (green circles), Kx6E with salt (blue triangles) and without salt (orange inverted triangles). The curves are model fits to the three data sets (green solid, blue dash and brown dash-dot, respectively) with the parameter values given in Table R3. For ProtL, the 95 % confidence interval of fit is also shown (light green area). The inset shows the DPF, ξ , derived from ν_{dyn} with N_H estimated from the solvent-accessible surface area.

Property	WT ProtL	Kx6E folded (0.75 M NaCl)	Kx6E unf. (0 M NaCl)
N_H^*	405	421	972
ν	2.23 ± 0.01	2.27 ± 0.01	2.80 ± 0.03
E_H^- (kJ·mol ⁻¹)	27.0 ± 0.3	27.0 ± 0.3	25.6 ± 0.4
χ^2_{red}	1.32	0.56	3.50

* Estimated from the ASA

Table R3. Summary of the temperature dependence of the hydration number. Results derived from the fits of the temperature dependence of ν_{dyn} to a power-law distribution are similar for folded proteins.

For the folded proteins, the temperature dependence of $\nu_{\text{dyn}}(T)$ is well described by the model (Figure R12). The parameter values resulting from the fits (Table R3) are nearly the same for both proteins and they are also very close to the values found previously for the similarly sized proteins BPTI and ubiquitin (Mattea, Qvist et al. 2008). This finding implies that, in the examined 60 K temperature range, the average rotational correlation time, $\langle \tau_H \rangle$ (on Equation R11) of water molecules interacting with the highly charged surface of the halophilic Kx6E is indistinguishable from that of non-halophilic proteins. In other words, there is **no discernable halophilic signature in the hydration dynamics** even at temperatures near the practical limit of supercooling (238 K), where any differences might have been amplified. It should be noted that the difference in ν_{dyn} at low temperatures between ProtL and folded Kx6E is mainly a consequence of the weaker temperature dependence of the bulk water relaxation rate, R_1^0 , in the presence of 750 mM NaCl. If we consider instead the quantity $N_W \cdot (R_1 - R_1^0) = \nu_{\text{dyn}} \cdot R_1^0$, the difference between the two proteins almost disappears.

Thus, the *replacement of 6 lysines with glutamates has little or no effect on the global hydration dynamics*, consistent with the conclusion from previous ^{17}O MRD studies of numerous proteins that ν_{dyn} is not significantly correlated with the net protein charge, the total number of charged residues, the total number of carboxylate groups nor with any of these quantities normalized by the ASA of the protein (Halle 2004). By recording ^{17}O MRD profiles for ubiquitin as a function of pH, a small increase on ν_{dyn} was observed upon deprotonation of the 12 carboxyl groups at 300 K (Halle, Denisov et al. 2002). For the proteins examined here, this result would translate into a 10 % increase in ν_{dyn} upon introduction of 6 glutamate residues. However, our findings show a 9 ± 7 % *decrease* in ν_{dyn} at 300 K (Table R2). This result can be explained by the loss of 12 methylene groups upon mutation, which are known to slow down water dynamics significantly. All results from present and previous water ^{17}O relaxation studies are consistent with the picture emerging from molecular dynamics simulations, where protein hydration dynamics are governed chiefly by surface topography rather by the chemical structure, charge or polarity of the hydrated residues (Makarov, Andrews et al. 2000, Henschman and McCammon 2002).

^{17}O MRD data rules out the possibility of finding tightly bound water molecules with residence times ≥ 1 ns in the hydration shell of halophilic proteins

^{17}O relaxation data show no discernable halophilic signature in hydration dynamics

4.4. ROLE OF THE UNFOLDED STATE IN SALT ADAPTATION

Up to this point, we have investigated the contribution from the electrostatic interactions between salt ions and ionizable residues, as well as the halophilic hydration shell compared to that of mesophilic proteins. So far, our experimental data support a model in which salt ions (mainly K^+) screen the electrostatic repulsion of negatively charged residues by means of weak unspecific interactions. The low values of the individual interaction constants together with the absence of tightly bound water molecules in the hydration shell clearly contradict a model where “specific patches of carboxyl groups are stabilized by the interaction with cooperative hydrated cation networks, particularly potassium” (Zaccai, Cendrin et al. 1989, Mevarech, Frolov et al. 2000, Ebel, Costenaro et al. 2002). In conclusion, the electrostatic stabilizing contributions reported so far are clearly not sufficient to explain protein haloadaptation in its whole, i. e. the whole global m_{KCl} value. Hence, other contributions have to be investigated.

We further analysed the residual salt effect on different structural parameters: salt and temperature **chemical shift perturbation coefficients**, *CSP* (see Materials and Methods 3.4.6) on the native conformation of WT and Kx5E ProtL. Results are shown in Figure R13. In both cases, we *failed to find any differences between the mesophilic and the halophilic proteins even for the mutated residues* (marked in red in figure R13). The analysis of the magnitude of the **salt coefficients** (left) shows very low values almost identical for both proteins, reflecting the weak effect of ion interactions with the folded state. **Temperature coefficients** (right) are identical in both proteins, which supports the idea that the native structure of halophilic proteins is the same as compared to that of mesophilic proteins. Moreover, temperature coefficients show no dependence upon salt addition at all, which evidences the lack of effects of KCl in the strength of the protein hydrogen–bonding network (Rothemund, Weisshoff et al. 1996).

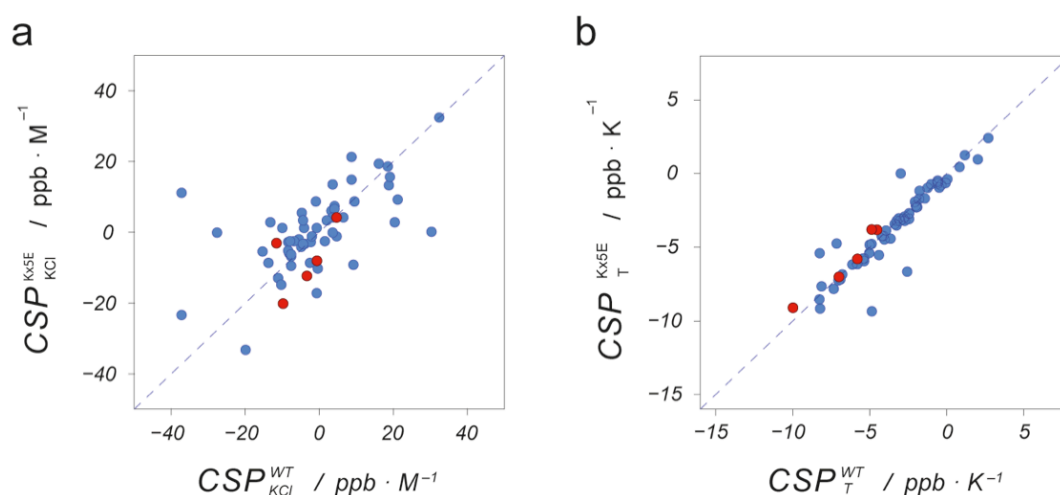


Figure R13. Chemical Shift Perturbation coefficients of ProtL. a) KCl and b) temperature coefficients for each residue of WT (abscise axis) and Kx5E (ordinate axis) ProtL. Mutated residues shown in red. Dashed lines correspond to $y = x$ and is just intended to serve as a guide to the eye.

No additional stabilizing mechanisms could be found from the study of the folded conformation. On the contrary, all our data confirm the similar structural properties of halophilic native states as compared to mesophilic states. Some works have outlined structural features stabilizing halophilic proteins, such as the insertion of certain structural elements or the presence of amphipathic helices (Frolow, Harel et al. 1996, Taupin, Hartlein et al. 1997, Premkumar, Greenblatt et al. 2005). Nonetheless, those are neither of general applicability nor seem related to the amino acidic halophilic signature. As a result, they are not considered to be universal adaptive features of halophilic proteins (Graziano and Merlino 2014).

A better understanding of the protein salt stabilization always implies analysing both the folded and the unfolded states. To describe haloadaptation, efforts have traditionally focused in understanding protein stability from the perspective of the native conformation, but it is known that the entropic penalty associated to denatured state solvation provides great part of the free energy favouring the native conformation. Particularly, some authors have shown how the effect of cosolutes on protein stability depends on the balance of the contributions between the folded and the unfolded states (Schellman 2003); therefore haloadaptation should certainly rely on the unfolded state at a great extent. Evolution has selected an optimal set of amino acids for halophilic proteins that, most likely, is also altering the properties of halophilic unfolded states.

Weak unspecific electrostatic K^+ -carboxylate interactions stabilize the folded conformation of halophilic proteins, but...

... no specific features in the structure, interactions or H-bonding pattern are found in the folded states of halophilic proteins.

4.4.1. The unfolded state becomes destabilized upon salt addition

Our hypothesis is that a relevant fraction of the salt stabilization of halophilic proteins occurs through effects on denatured conformations. In order to get a complete picture of how halophilic amino acid composition promotes haloadaptation, we have investigated the unfolded state. As a first approach, we monitored the effects of salt on the unfolded state at a thermodynamic level. For that, we used a **designed obligate halophilic protein: Kx7E ProtL** (Tadeo, López-Méndez et al. 2009a, Schlesinger, Wang et al. 2011). Kx7E is completely unfolded in the absence of salt, yet it folds progressively upon salt addition. At low salt concentration (< 1 M KCl) the ^{15}N - ^1H -HSQC of Kx7E shows significant populations of both native and denatured conformations (Figure R19).

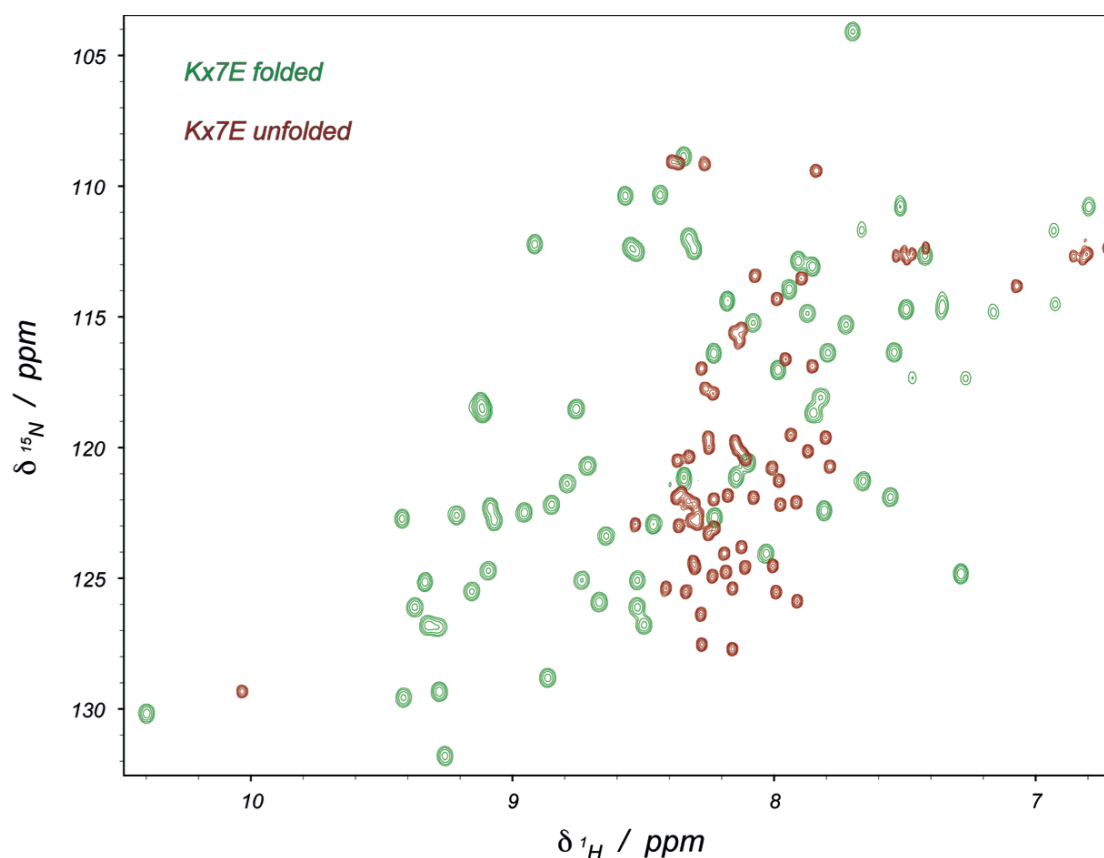


Figure R14. Kx7E ProtL is a strict halophile. Overlay of the ^{15}N , ^1H -HSQC from native (green, 1 M KCl) and denatured (red, 0 M KCl) Kx7E ProtL.

We measured the effect of KCl on the thermodynamic unfolding equilibrium constant as well as on the individual folding and unfolding kinetic rates of Kx7E ProtL. Kinetics of folding processes typically fall among the s–ms timescale (Ortega, Pons et al. 2013). Hence, folding and unfolding rates can be determined from an **EXSY** NMR experiment (Materials and Methods 3.4.5.3). In this experiment, the populations of native and denatured states are measured before and after a fixed exchange delay. Thus, information about the thermodynamic equilibrium as well as about the kinetic exchange between the two conformations can be obtained. A typical EXSY spectrum is displayed in Figure R20: the two *diagonal peaks* report on the folded and unfolded populations of the H_ϵ of Trp47, whereas *cross-peaks* report on the exchange between both conformations. Kinetic rate constants for folding and unfolding can be extracted from the relative intensity of the signals, as shown in Equation M20 in Materials and Methods.

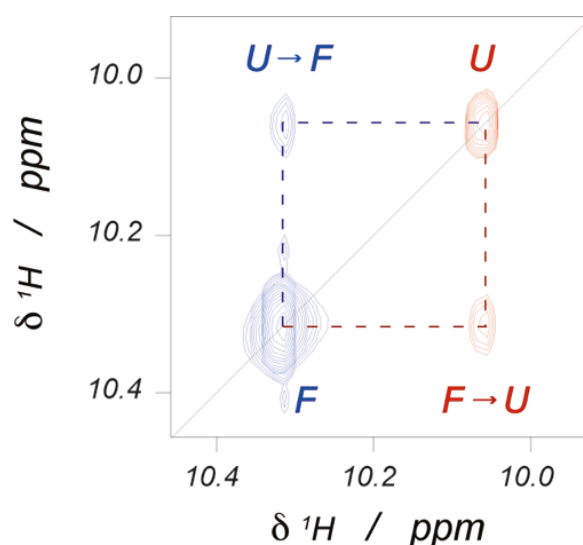
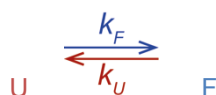


Figure R15. EXchange Spectroscopy. $^1\text{H}, ^1\text{H}$ -EXSY experiment for Trp-47 H_ϵ measured at 575 mM KCl. Peaks on the diagonal correspond to the folded (F) and unfolded (U) populations. Cross peaks correspond to the folding (blue) and unfolding (red) transitions.

Measurements in the presence of KCl show the modulation of each kinetic constant by salt. Assuming a simple two-states folding model (Equation R14) k_F and k_U are determined by the stability of the unfolded and the folded state respectively. Both kinetic constants change (Figure R21): while k_F increases, k_U is decreased, supporting the idea that the stabilities of both the folded and the unfolded states are affected by KCl. Remarkably, *the folding rate constant k_F increases significantly upon KCl addition, whereas the unfolding rate constant k_U is affected to a much lesser extent.* This indicates an important **destabilization of the unfolded state**.



Equation R14. Two-states equilibrium, determined by folding (k_F) and unfolding (k_U) rate constants.

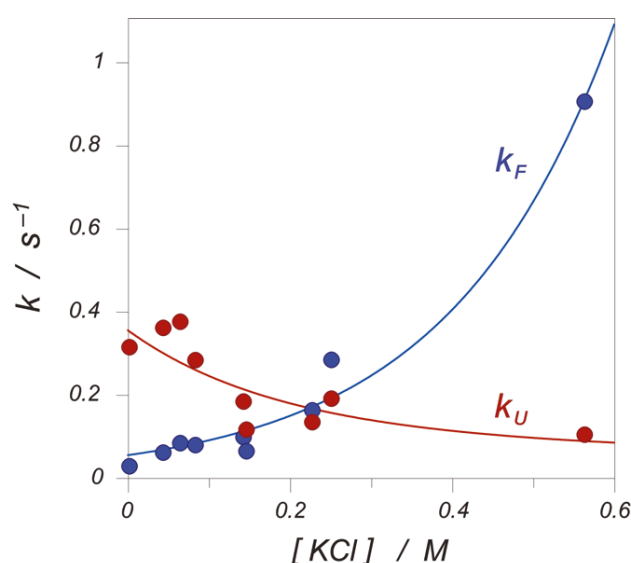


Figure R16. KCl modulation of folding and unfolding rates. Experimental kinetic rate constants (circles) for folding (k_F , blue) and unfolding (k_U , red) of the obligate halophile Kx7E ProtL obtained from

EXSY NMR experiments. Monoexponential fittings (lines) reveal that k_F depends more strongly on KCl than k_U .

The kinetic information obtained from EXSY experiments confirms that the unfolded state is a key player in salt stabilization. Moreover, the fact that the folding rate is modulated by salt to a greater extent than the unfolding rate indicates that *an important part of KCl stabilization occurs by mechanisms destabilizing the unfolded state*. Our data show that the contribution of those mechanisms is, at least, comparable to that of electrostatic interactions. Unfortunately, we lack an equivalent mesophilic system in order to study the behaviour of individual kinetic rates and compare the differences related to haloadaptation.

It is important to note that the interpretation of kinetic data is complex. The assumption of a simple two-states model might not be a good approximation. It has often been considered a major over-simplification, even when data agree well to a cooperative folding without any detectable intermediates (Sosnick and Barrick 2011). A more detailed discussion in light of current folding views will be carried out further ahead in this thesis.

Electrostatic stabilization of halophilic proteins is not enough to account for all the salt stabilization of halophilic proteins.

Our kinetic data show that an important contribution comes through an unfolded state destabilization.

4.4.2. Structural and dynamic analysis of the unfolded conformational ensemble

The characterization of the unfolded state is unfortunately restricted by the limited information that can be extracted from available experiments: since there is no defined structure, it is difficult to obtain adequate descriptors for its conformational ensemble (Forman-Kay and Tanja 2013). The study of the denatured state of stable proteins is particularly challenging because it requires extreme experimental conditions of temperature, pressure or denaturant concentration, being the latter the most common approach. Unfortunately, the use of high denaturant concentrations is problematic. First, it complicates things from an experimental point of view due to limited cosolubility. Second, it is likely to interfere with the studies of the effects of other cosolutes such as KCl. There are ways to bypass this issue based on Wyman linkage theory (Wyman 1964, Kloss and Barrick 2008). However, we were not assuming any interaction mechanisms *a priori*, so we couldn't be certain about the effects of both cosolutes being independently additive. In order to overcome these limitations, we decided to use **intrinsically disordered proteins (IDPs) as models for the denatured state**. IDPs are sequences that cannot acquire a native folded conformation but still have a relevant defined biological function, often involved in molecular recognition, regulation or signalling processes (Liu, Faeder et al. 2009). Similarly to the denatured state of stable proteins (Robic, Guzmán-Casado et al. 2003), they exhibit some *transient residual structure* (De Biasio, Ibáñez de Opakua et al. 2014). All together, it supports the use of IDPs as an adequate approach to the study of the unfolded state. Moreover, since they do not require any denaturants for unfolding, the effects of KCl can be monitored directly.

We have chosen the SH4 and *unique* domains of human SRC kinase, **uSRC** (see Materials and Methods 3.1.3), as a model of a **mesophilic unfolded state**. uSRC is a disordered protein with residual structure in regions involved in lipid binding and membrane anchoring. It also shows a well-defined NMR spectrum that has been previously assigned (Perez, Gairi et al. 2009). The number of proteins obtained from halophilic organisms is small, and no unfolded halophilic proteins are available to the best of our knowledge. Thus, a number of selected mutations were introduced on uSRC sequence in order to design a halophilic disordered homologue, **haloSRC** (see Materials and Methods 3.1.4), which is used as a model for the **unfolded state of halophilic proteins**. The assigned spectrum of the designed haloSRC can be found in figure R14 (red) overlaid to the uSRC spectrum (blue).

The $^{15}\text{N}, ^1\text{H}$ -HSQC spectrum of haloSRC is very similar to that of its homologous uSRC. The dispersion is very narrow as corresponds to a largely unfolded peptide. Nevertheless, nearly complete assignment (95 %) was achieved: 81 out of 85 non-proline residues (BMRB accession number 251961).

haloSRC is a disordered halophilic sequence intended as a model for halophilic unfolded states.

It is designed using uSRC as the starting template.

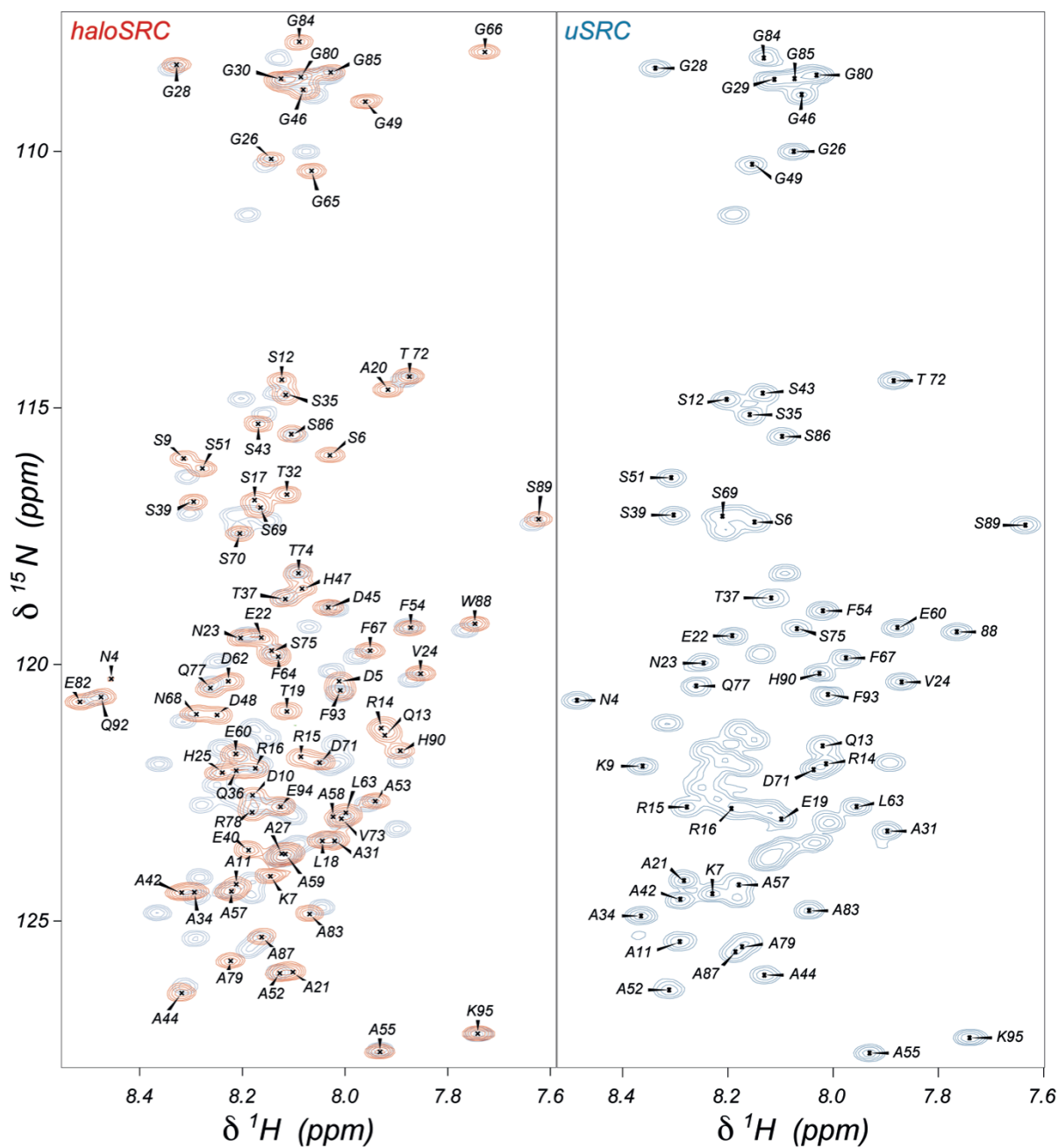


Figure R17. uSRC and haloSRC HSQC spectrum. Left panel shows the assigned $^{15}\text{N},^1\text{H}$ -HSQC spectrum at 280 K for haloSRC (red) overlaid to that of uSRC (blue). The dispersion of the signals is very low, as corresponds to a largely disordered peptide, but plenty of signals change upon introduction of the mutations. For comparison, right panel shows the assigned uSRC $^{15}\text{N},^1\text{H}$ -HSQC spectrum (in blue). The assignment of uSRC was performed by using ^{13}C -detected experiments in order to overcome the structural overlap (Perez, Gairi et al. 2009).

As already explained, the unfolded state plays an important role in protein stability, so it should be taken into account to understand KCl stabilization of proteins. Actually, the effect of cosolutes on protein stability has been already addressed and can be understood according to different interpretations, some of them explicitly considering the unfolded state. Particularly, **Schellman** proposed a simple thermodynamic and mathematical model that wraps up the accumulated knowledge and conceptual contributions from Pielak, Tanford, Timasheff... He considers two effects: *cosolute–protein interactions* and *intermolecular size–excluded volume* (Schellman 1990):

- i. **Cosolute molecules interact with protein** atoms. The strength of these interactions depends mainly on the fraction of *polar surface* of both counterparts (Street, Bolen et al. 2006), therefore interactions occur mainly with *backbone amides and carbonyls* (Liu and Bolen 1995). Since these groups are largely exposed in unfolded conformations, the number of interactions is larger with the unfolded state. As a result, interactions stabilize the unfolded state and have a net native–destabilizing effect. In fact, this enthalpy–driven mechanism is the dominant one in denaturant–induced unfolding (Tanford 1970). Stabilizing cosolutes, on the other hand, are preferentially excluded from the protein surface.
- ii. Meanwhile, **excluded volume** has a net stabilizing effect. It arises from the entropic penalty of creating a cavity in the solvent to host a protein molecule (Graziano and Merlino 2014), which is proportional to the concentration of both cosolute and macromolecule (Schellman 2003, Schlesinger, Wang et al. 2011). The volume occupied by the protein cannot be occupied by the cosolute molecule, which is *excluded* from the cavity (protein is also excluded from the volume occupied by the solute molecule, but the latter is so small that the effect is negligible). As a result, the system favours folded conformations because of the lower volume restrictions.

In the next pages, we will explain the experimental design, as well as the results and conclusions, of a set of experiments aimed to probe the contributions of protein–cosolute interactions and excluded volume in the unfolded states of halophilic and mesophilic proteins.

The conformational features and rigidity of the folded state are highly conserved in halophilic proteins, as demonstrated by several X–Ray and NMR structures (Graziano and Merlino 2014) – also supported by our similar temperature and salt coefficients (Figure R13) and the structural similarity of the designed halophilic variants (Tadeo, López–Méndez et al. 2009b). Thus, any *changes in the excluded volume contribution would arise from changes in the conformational ensemble of the unfolded state*. In disordered peptides, amide proton shifts (HN and HA) and ^{15}N relaxation rates (^{15}N R_1 , R_2 and NOE) (Bruschweiler, Liao et al. 1995) are the most valuable direct indicators of residual structure (Klein–Seetharaman, Oikawa et al. 2002). We measured the full ^{15}N **relaxation dataset** (Materials and Methods 3.4.4.1) and analysed it according to the *model–free formalism* (Lipari and Szabo 1982a) to characterize local dynamics of the polypeptide backbone by *per* residue order parameters, S^2 (*i*), and correlation times, τ_c (*i*) (Materials and Methods 3.4.4.2); which comprise both overall and internal motions on the nanosecond timescale (Figure R15). *If there were to be any significant differences in structure or compaction, they would be reflected on these structural parameters.*

Both uSRC and haloSRC show molecular tumbling features characteristic of an unfolded protein: faster than the motion modelled for a rigid spherical body of equal molecular weight (~ 11 ns for a 95 amino acids protein) but slower than fully autonomous, unrestricted side chain motions (Figure R15a). **No significant differences in τ_c (*i*) values** were observed

between halophilic and mesophilic unfolded peptides, indicating that *both conformational ensembles are similar* within the sensitivity range of the experiment. *Halophilic amino acid composition does not significantly alter the unfolded conformational ensemble*. Furthermore, no significant changes in $\tau_c(i)$ or $S^2(i)$ were detected upon KCl addition, indicating that both conformational ensembles **maintain their structural and dynamic properties** at the range of concentrations tested.

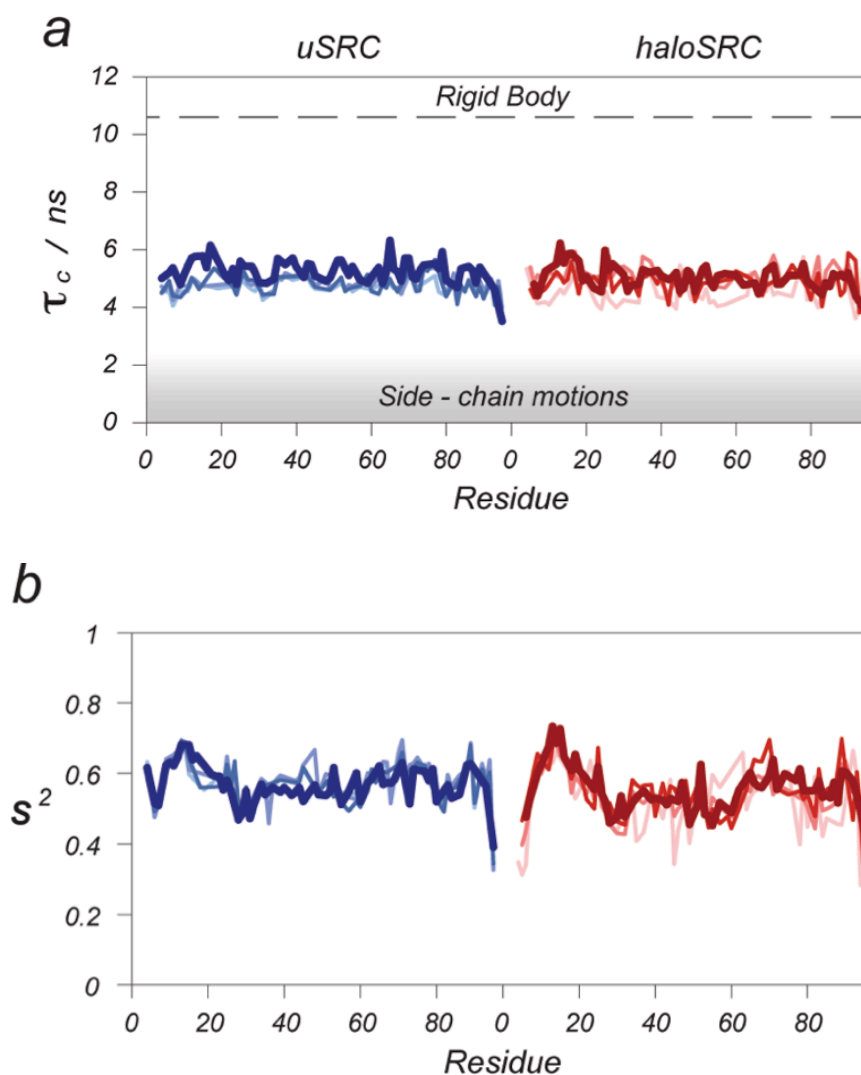


Figure R18. Model-free analysis of uSRC and haloSRC. Fast local dynamics obtained for two unfolded sequences: mesophilic uSRC (blue) and halophilic haloSRC (red). *Per residue* a) correlation times τ_c and b) order parameters S^2 were obtained from fitting to model-free formalism (Lipari and Szabo 1982a) of R_1 , R_2 and NOE data. Local correlation times in a) have intermediate values between modelled rigid body motion and side-chain unrestricted motions. Darker to lighter gradient colour correspond to 0, 0.5, 1 and 1.5 M KCl.

We then probed salt induced changes in **local dynamics** by *transverse ^{15}N relaxation rates in the rotating frame, $R_{1\rho}$* . Measurements were made at 0, 0.5, 1 and 2.5 M KCl (Figure R16). Results were analysed using a model that assumes unrestricted side chain motion, which is particularly suitable to identify hydrophobic or electrostatic clusters with residual structure in IDPs (Materials and Methods 3.3.4.2) (Klein-Seetharaman, Oikawa et al. 2002). For uSRC, a

rigid cluster identified within the 20 N-terminal residues corresponds to a previously reported lipid-binding region (Perez, Gairi et al. 2009). The fact that the *cluster structure remains invariant to KCl addition* reinforces the hypothesis of its hydrophobic nature (Figure R16, blue). The designed haloSRC protein also shows an equivalent cluster with reduced mobility (despite three amino acid mutations concentrating within this region) and very similar baseline values (Figure R16, red). Consistently, it also remains largely unaffected by salt. All these data corroborate the conclusion that **neither salt nor the halophilic sequence alter the conformational ensemble of the unfolded state**. Thus, our experimental data show no significant contribution of the unfolded state to haloadaptation through changes in the excluded volume (i. e. differential compaction or extension of halophilic unfolded state as compared to the mesophilic).

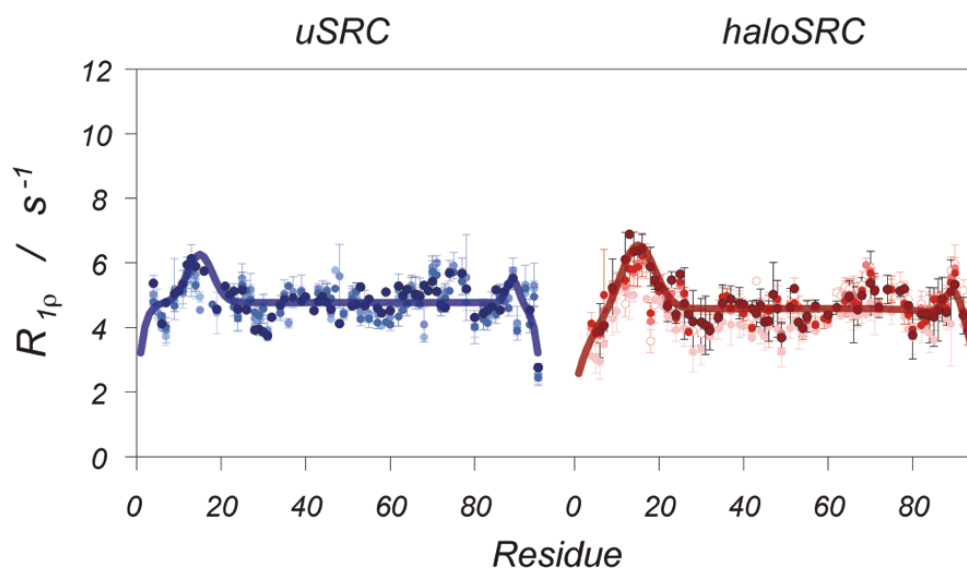


Figure R19. Local dynamics of uSRC and haloSRC. ^{15}N relaxation rates in the rotating frame $R_{1\rho}$ for uSRC (blue) and haloSRC (red) measured at KCl concentrations increasing (from dark to light colour: 0, 0.5, 1 and 1.5). Solid line represents the least square fitting of data at 0 M KCl to a model for segmental motion assuming reduced flexibility around two hydrophobic clusters near the N and C-termini.

Measurement of global and local dynamics of halophilic and mesophilic unfolded states show that...

... neither halophilic amino acid composition nor the addition of KCl alter the denatured conformational ensembles,

thus ruling out contributions from excluded volume to haloadaptation.

4.4.3. Analysis of the interactions between KCl and the unfolded state

Once ruled out changes in the excluded volume term, the role of other contributions to salt stabilization in halophilic proteins was tested. To probe the putative interactions between cosolute ions and the two unfolded models at a *per* residue level, we monitored the **KCl-induced chemical shift perturbation of the amide NMR signal**, CSP (i), (see Materials and Methods 3.4.6) for uSRC and haloSRC – in a similar way as it was previously done for folded WT and Kx5E ProtL (Figure R13). CSP coefficients monitor changes in the chemical environment of the NH amide bond, ultimately reflecting the **proximity of salt ions**.

Again, all residues show a linear dependence of their chemical shifts on KCl concentration. This time, though, the CSP (i) values are larger than for the folded proteins – as expected for the fully exposed backbone of an unfolded peptide – and more negative – as corresponds to an increase of the surrounding electron density. In addition, and more important within the context of haloadaptation, is that there are *noticeable differences between the mesophilic and halophilic sequences*: the chemical shifts of **haloSRC residues are significantly less affected by cosolute ions** than those of mesophilic uSRC. Figure 17 summarises these $CSP_{KCl}(i)$ experimental values, along with their averages and pairwise differences, where available (spectral overlap in unfolded peptides makes it troublesome to follow the shifts during the KCl titration). Changes occur **throughout the whole sequence**, even though haloSRC differs from uSRC in just eight point substitutions (marked as shaded bars).

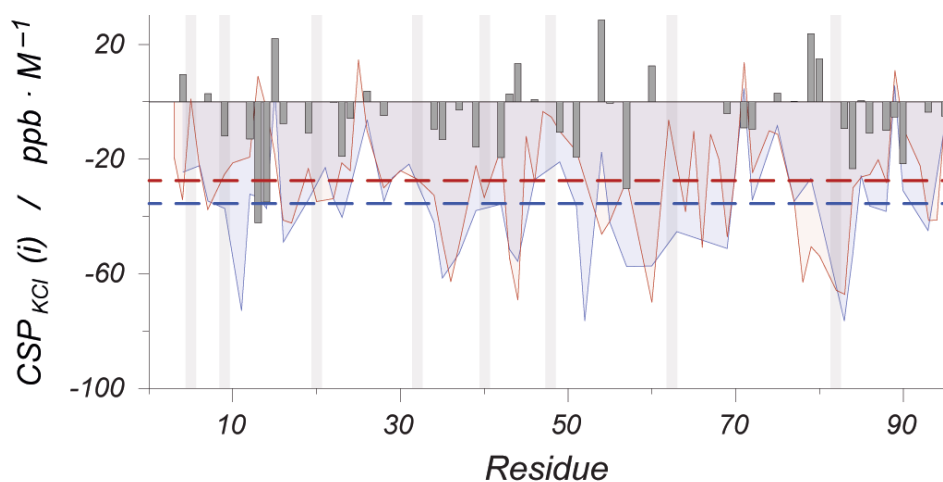


Figure R20. Chemical shift perturbation coefficients. $CSP_{KCl}(i)$ experimental values for uSRC (blue) and haloSRC (red) are represented as coloured areas, while dashed lines mark the averages. Grey bars correspond to the differences between the coefficients of the halophilic and mesophilic proteins, where available. Shaded columns outline the mutated positions.

Two conclusions can be extracted from the analysis of the CSP data:

- The lower absolute CSP values of the halophilic sequence are consistent with **weaker interactions** between KCl ions and the exposed backbone in unfolded conformations.
- The direction of the perturbation towards lower chemical shifts means that the NH bonds become more “shielded” from the magnetic field upon KCl addition. This is conventionally explained as an increase of the local magnetic field generated by surrounding electrons. As a result, our evidences point to **chloride** anion as the main ionic specie interacting with the amide NH. This is in agreement with studies suggesting that the adsorbability of anions determines their Hofmeister ordering (Nihonyanagi, Yamaguchi et al. 2014).

To confirm the lower interaction / preferential exclusion of salt ions from halophilic unfolded states we determined the KCl dependent *fast (ms timescale) amide HN/H₂O exchange rates* with **CLEANEX** experiments (see Materials and Methods 3.4.5.2) (Hwang, van Zijl et al.

1998). We obtained exchange rates $k_{\text{ex}}^{\text{CLEX}}(i)$ for 45 and 48 residues in uSRC and haloSRC respectively at 0, 0.5, 1 and 1.5 M KCl (Figure R18). In all cases, exchange rates showed dependence with salt concentration; which was well described by a single exponential **decay**:

$$k_{\text{ex}}^{\text{CLEX}}(\text{KCl}) = k_{\text{ex}}^{\text{CLEX}}(0) \cdot e^{-R_{\text{KCl}}^{\text{CLEX}} \cdot [\text{KCl}]} + C \quad ;$$

Equation R13. CLEANEX measurement of fast exchange. Kinetic rate constant measured by CLEANEX experiments at a given KCl concentration. It is described by a single exponential characterized by a decay constant $R_{\text{KCl}}^{\text{CLEX}}$ characteristic for each residue.

where $k_{\text{ex}}^{\text{CLEX}}(0)$ is the exchange rate in the absence of salt, $R_{\text{KCl}}^{\text{CLEX}}$ the exponential decay rate and C the empirical baseline constant. Examples of fitting of the experimental data to equation R13 are shown in Figure R18. The values of the exponential decay rates $R_{\text{KCl}}^{\text{CLEX}}$ are summarized in Figure R19.

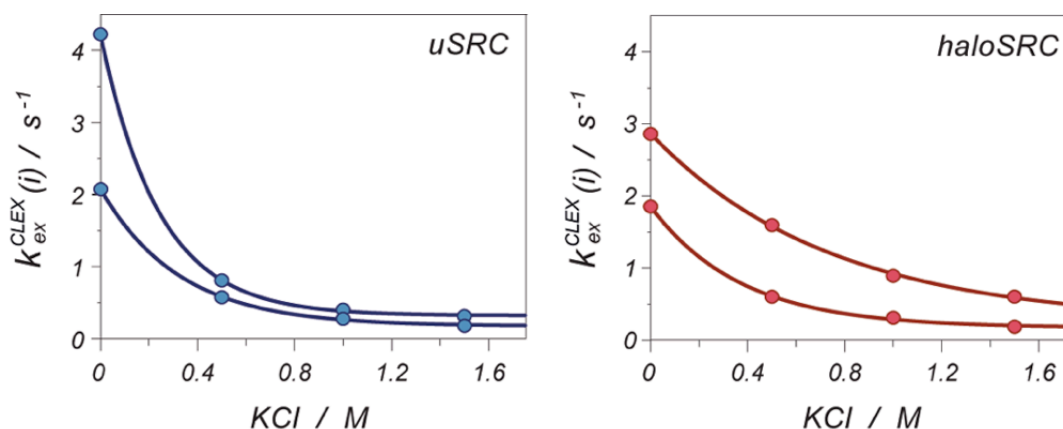


Figure R21. KCl slows exchange exponentially. Fast amide ^1H exchange rates, $k_{\text{ex}}^{\text{CLEX}}$, derived from CLEANEX experiments at different KCl concentrations are illustrated for two representative residues of uSRC and haloSRC. Upon KCl addition, $k_{\text{ex}}^{\text{CLEX}}$ rates decrease exponentially (solid lines) with characteristic decay constants $R_{\text{KCl}}^{\text{CLEX}}$.

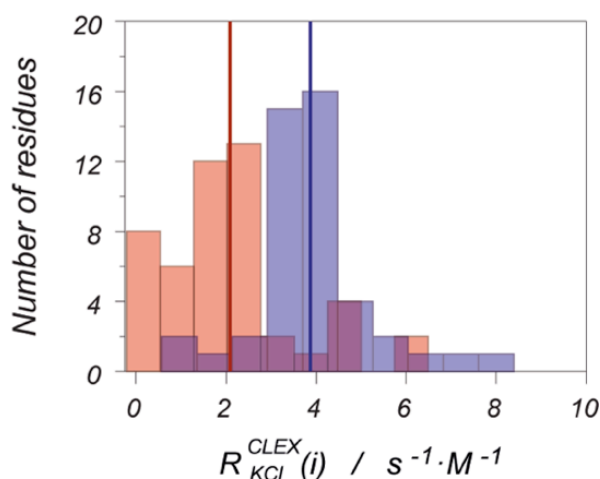


Figure R22. Rates of KCl-induced exchange decay. Distribution of $R_{\text{KCl}}^{\text{CLEX}}$ constants for uSRC (blue) and haloSRC (red) with their indicated averages (vertical lines) reveal that ^1H exchange rates of the halophilic protein are less sensitive to salt (*i.e.*, generally lower values for haloSRC).

Remarkably, there is a general **decrease of the exchange rate** upon KCl addition that can be accurately modelled by an exponential single decay. Interestingly, the rate of the decay is different for both proteins. Mesophilic uSRC exhibits a narrow distribution of decay rates centred around $\sim 4 \text{ s}^{-1} \cdot \text{M}^{-1}$, while halophilic haloSRC shows a broader distribution around ~ 2

$s^{-1}\cdot M^{-1}$. The slower exchange in the presence of KCl could be explained arguing that *ions interacting with the protein surface block water from exchanging with local amide groups*. In other words, salt ions compete with water molecules for surface binding, decreasing the effective exchange rate k_{ex}^{CLEX} . Thus, larger R_{KCl}^{CLEX} values indicate a stronger attenuating ion effect on local HN/H₂O exchange. The lower R_{KCl}^{CLEX} values reported for all residues in haloSRC reveal that this *halophilic construct is less sensitive to cosolute ion concentration* than the mesophilic uSRC, in full agreement with the chemical shift perturbation data. In addition, our results in the absence of salt (Figure R20) show slightly faster exchange rates for haloSRC than for uSRC. Provided the *conservation of the conformational ensemble and the similarity of hydration shell dynamics*, this presumably arises from a higher number of waters near the halophilic surface.

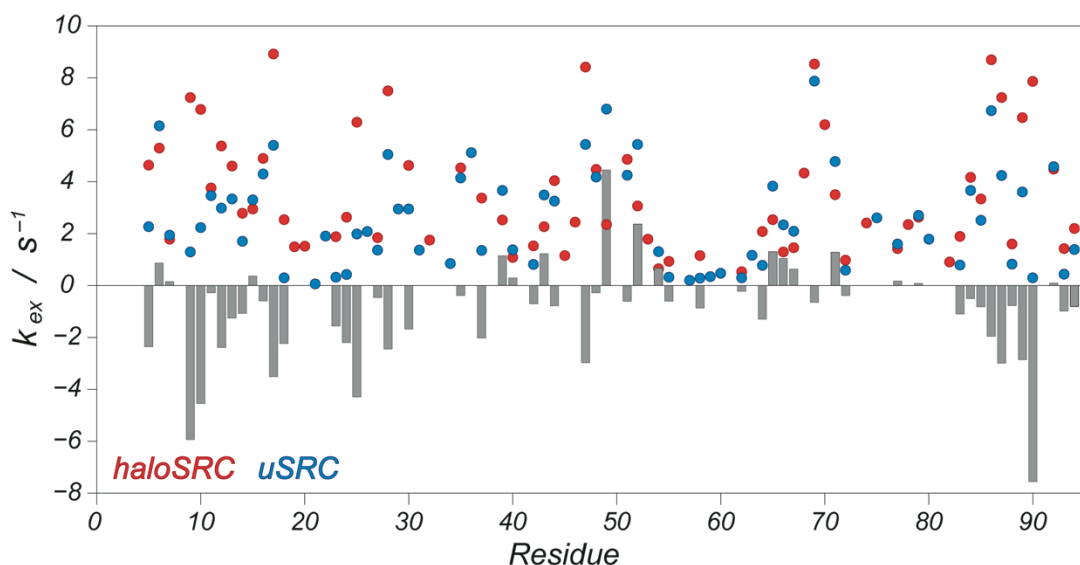


Figure R23. uSRC and haloSRC fast exchange rates measured by CLEANEX. Experimental k_{ex} values in the absence of salt (circles). Exchange is slightly faster for haloSRC (red) than uSRC (blue). Differences are indicated as grey bars, where available.

Thus, **disordered halophilic proteins preferentially exclude chloride and/or potassium ions** from their hydration layer as compared to their mesophilic counterpart. This result can probably be generalized to halophilic denatured states. According to Schellman model, the balance of cosolute–protein interactions results in a net enthalpic contribution that favours the unfolded state and destabilizes the native conformation. The decreased number of interactions between halophilic unfolded states and salt ions reduces this contribution; in line with a previously proposed mechanism for osmolyte stabilization (Arakawa and Timasheff 1982). As a result, KCl addition results in **lower stabilization of halophilic unfolded states** compared with mesophiles, ultimately contributing to a larger native state stabilization; i.e. halophilic amino acid selection triggers a differential destabilization of the unfolded state.

CSP and CLEANEX data show preferential ion exclusion from halophilic unfolded state,

which can account for the unfolded state destabilization found in the EXSY experiments.

4.4.4. Thermodynamic contribution of the unfolded state, m^{unf}

We have identified a new differential mechanism for salt stabilization of halophilic proteins, occurring mainly through an unfolded state destabilization. Unfortunately, accurate quantification of its thermodynamic contribution is extremely complicated. The characterization of weak interactions and competition with hydration is still dealt with by computer simulations and theory, being the current status of the field limited to basic systems such as methane (Graziano 2008, Baldwin 2014) or charged surfaces with simple determined geometries (Nihonyanagi, Yamaguchi et al. 2014). We have adopted a thermodynamic approach for the quantification of preferential ion exclusion energetics: measurement of the slow **H/D exchange** of backbone amides of β LAC, WT and Kx5E ProtL as a function of KCl concentration.

H/D exchange experiments allow probing local salt stabilization at the level of individual residues. Labile protons within the sample can exchange with molecules of protic solvents. In H/D exchange experiments, the exchange of protons from the protein with deuterons from the solvent is monitored by the decrease of signal intensity in a ^1H , ^{15}N -HSQC (see Materials and Methods 3.4.5.1). The exchange of backbone amide NH's from the protein occurs through local or global unfolding. Under the appropriate conditions – EX2 regime – the rate of exchange $k_{\text{ex}}^{\text{H/D}}(i)$ can be interpreted as a '*local protection factor*' and, in some cases, as local unfolding free energy, $\Delta G_{\text{U-F}}^0(i)$ (Equations M16 and M17 in Materials and Methods 3.4.5.1). Measurements at different KCl concentrations would return the **local salt stabilization** for each residue. Thus, we performed H/D exchange experiments for our three model folded systems: β LAC, WT and Kx5E ProtL. For all residues, we observed that the local stability depends **linearly** on salt concentration, therefore the slopes $m_{\text{KCl}}^{\text{H/D}}(i)$ can be used as reporters of local KCl stabilization (Equation M18). Positive values show a stabilizing contribution of a given residue i , whereas negative values indicate destabilization.

Figures R22 and R23 show the $m_{\text{KCl}}^{\text{H/D}}(i)$ values for halophilic β LAC, mesophilic WT and halotolerant Kx5E ProtL. When looking at individual residues, most favoured residues in halophilic sequences (Asp, Thr, Glu and Val) do not show larger $m_{\text{KCl}}^{\text{H/D}}(i)$ values. Particularly, in the case of β LAC some of these residues are even slightly destabilizing (e.g. residues like D5, D71, T120, E136, E166 and V255), whilst some of the residues showing larger stabilization do not correspond to the evolutionarily selected subset (e.g. A48, Q54, A95, L107, L118 and R123). Accordingly, positions mutated in Kx5E ProtL do not stand out as particularly stabilizing. If specific cation–carboxylate interactions were to be haloadaptation major driving forces, larger stabilization values would be expected from acidic and nearby residues. Instead, we find that *the effects of hypersaline adaptation are not exclusively restricted to the local environment of the evolutionarily selected amino acids, but instead they are found **spread along the entire sequence*** (lower panel on Figure R23), *supporting the hypothesis of some general unspecific mechanism*. There is, however, a certain correlation between $m_{\text{KCl}}^{\text{H/D}}(i)$ values and the exposure of the residue to the solvent. Surface residues exposed to solvent – therefore to cosolute ions – are the ones showing lower stabilization at high KCl concentrations. On the other hand, residues that are less exposed to the solvent – i.e. buried residues – show the largest stabilizing $m_{\text{KCl}}^{\text{H/D}}(i)$ values. This can be understood by considering that any mechanism of salt–induced stabilization results from a balance of the contributions between the folded and unfolded conformations (Schellman 2003), ultimately depending on the *changes* in solvent–exposed area upon unfolding (Davis-Searles, Saunders et al. 2001).

Traditionally, m -values have been obtained in the presence of denaturants (mainly guanidinium salts or urea). They reflect the amplitude of local/global unfolding events or limited structural fluctuations that result in exchange (Mayo and Baldwin 1993). Denaturant modulation of those events is directly proportional to the area exposed: local or global unfolding processes expose large amounts of area and are greatly enhanced by denaturants (i.e. show a large m -value), whereas small fluctuations show negligible dependence on denaturant concentration (i.e. small or null m -value). This is because the mechanism for protein destabilization and denaturation is totally dominated by strong *interactions of the protein backbone* with urea or guanidinium (Courtenay, Capp et al. 2001). These cosolutes bind tightly to the amide polar atoms, therefore favouring the unfolded conformation. The protein destabilization is directly proportional to the number of binding sites made available upon unfolding/opening, which at the same time are proportional to the surface newly exposed (Schellman 1990). For instance, deep buried residues exclusively exchange through large global unfolding events, so their m -values reflect the global thermodynamic slope measured in chemical denaturations. Individual m -values are, then, a measure of the effect of intermolecular protein–cosolute interactions in the unfolded state.

In the case of KCl, the individual $m_{KCl}^{H/D}(i)$ show **positive values**, opposite to the negative values (destabilizing) generally found for denaturants. Unlike urea or guanidinium chloride, *backbone contacts with KCl are less favourable than with water. The lesser and weaker interactions with the backbone are not enough to stabilize the unfolded state, as a result the unfolded state is destabilized* and the native conformation indirectly stabilized. The individual m -values are, then, a measure of the *preferential exclusion* of KCl from the backbone amides in the unfolded state. As it has been previously pointed out, there is a correlation between *per residue salt stabilization* and solvent accessibility. Since solvent–exposed residues contribute equally in both folded and unfolded states, their net contribution to salt–induced stabilization is negligible, as so are their $m_{KCl}^{H/D}(i)$ values. In other words, the changes in ASA are very small. Meanwhile, fully buried residues are protected from solvent interaction in the folded state, becoming exposed only upon protein unfolding. Thus, **buried residues report directly on the impact of ion interactions in the unfolded state**. *Their stabilizing $m_{KCl}^{H/D}(i)$ values are equivalent to the energetic destabilization of the unfolded state, $-m_{KCl}^U$, due to preferential exclusion from the backbone amides.*

To quantify the contribution of KCl to the thermodynamic stability of the *unfolded state*, we selected the five buried residues ($f_{ASA} > 0.9$) with highest local absolute stability $\Delta G_{U-F}^0(i)$. The overall m_{KCl}^U is estimated as the average of those five $m_{KCl}^{H/D}(i)$ values (dashed lines in Figure R22 and R23). Our results (Figure R24) show that, in the presence of KCl, the **unfolded state is marginally destabilized** for mesophilic WT ProtL but **strongly destabilized in both halophilic proteins**. The most straightforward interpretation is that halophilic unfolded states display fewer sites for KCl–protein interaction or, if preferred, that they exhibit more sites for preferential hydration and KCl preferential exclusion. This is consistent with our CLEANEX results, showing faster water–exchange for the halophilic sequence in the presence of KCl. Why halophilic sequences show more sites for hydration/ion exclusion? Our NMR relaxation data show that halophilic and mesophilic unfolded conformational ensembles are similar. Therefore we conclude that differences arise from the composition of the area, ultimately depending on amino acid composition, rather than from the amount of area exposed. This key idea will be further developed in the Model proposal (section 5).

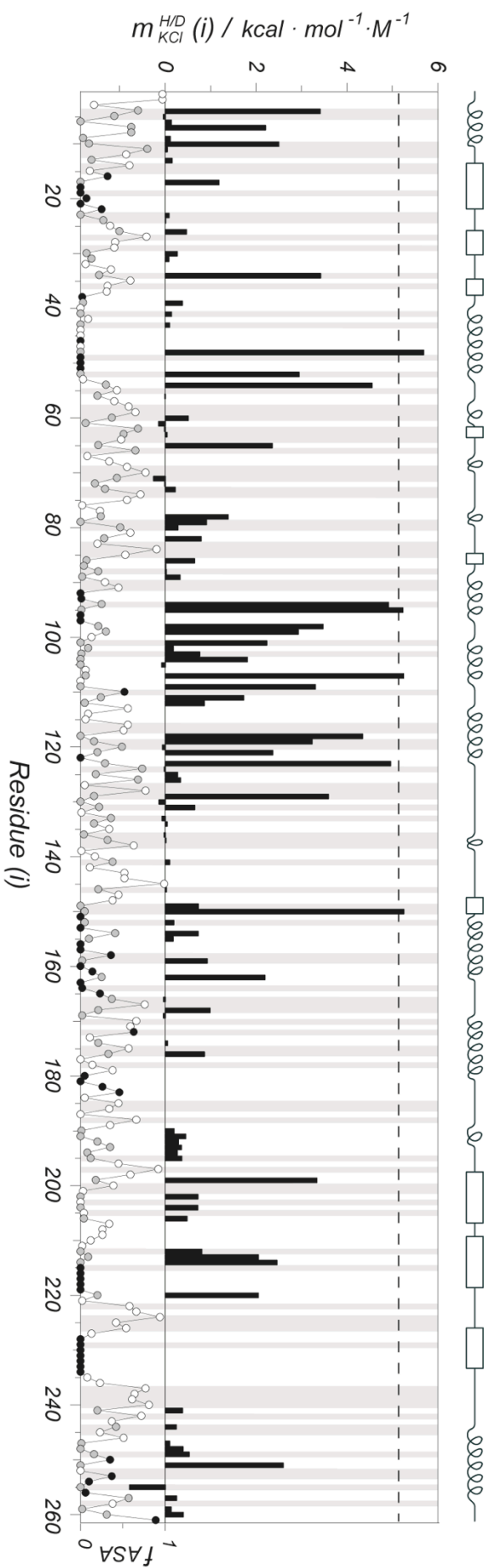


Figure R24. HD exchange profile of βLAC . Local KCl stabilization $m_{KCl}^{H/D}(i)$ derived from H/D exchange rates for each residue of natural halophilic βLAC are represented as black bars. Average value of five most protected residues is shown as a dashed line. Four most abundant residues in halophilic sequences (D, T, E, V) are highlighted in grey. Fractional solvent accessible surfaces, f_{SAS} , are shown below (right axis) and colour coded according to the timescale of the exchange: measurable timescales (grey circles), too short timescales for exposed residues (< 30 min, white circles) and too long timescales for deep buried residues (> 1 y, black circles). Schematic representation of the secondary structure is shown above.

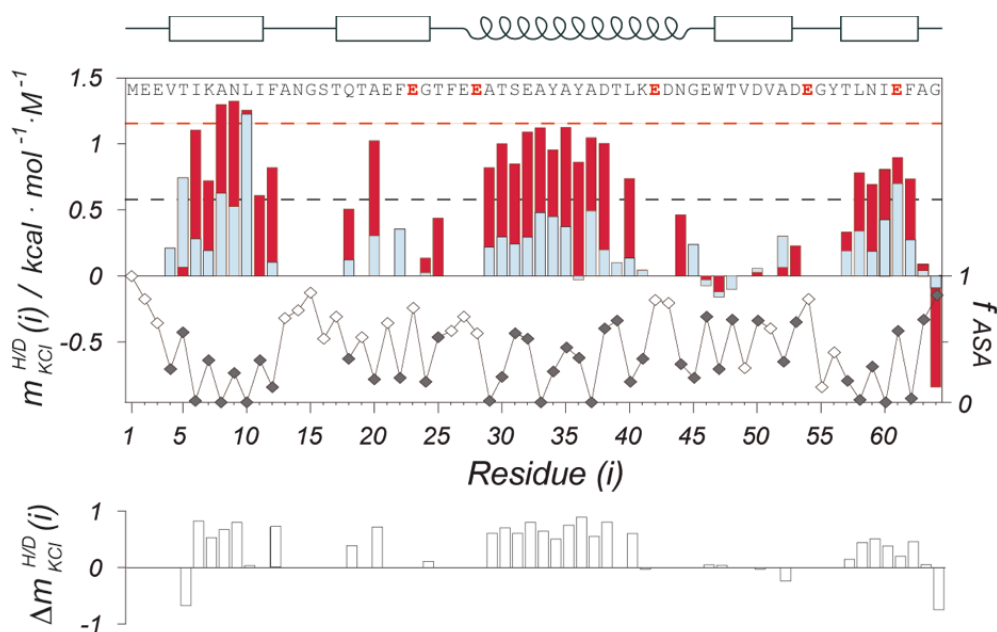


Figure R25. H/D exchange of ProtL. Barplot with local KCl stabilization $m_{KCl}^{H/D}(i)$ values derived from H/D exchange rates for each residue of natural mesophilic WT ProtL (light blue) and designed halophilic Kx5E mutant (red). Average values of five most protected residues are shown as colour-coded dashed lines. Fractional solvent accessible surfaces, f_{SASA} , are shown below (right axis) and colour coded according to the timescale of the exchange: measurable residues with intermediate timescales (grey circles) and surface residues with too short timescales (< 30 min, white circles). Schematic representation of the secondary structure is shown above the plot, as well as the sequence with mutated residues highlighted in red. Barplot below shows the pairwise differences (when available) in KCl stabilization, $\Delta m_{KCl}^{H/D}(i)$.

Remarkably, the individual m -values are always smaller than the global stabilization measured from global thermodynamic data (thermal denaturation experiments), especially for halophilic proteins. This is because, unlike denaturants, interactions with the unfolded state are just part of the total thermodynamic effect of salts on protein stability. Additional stabilization comes from the electrostatic screening of negatively charged carboxylates in the folded state. Not surprisingly, the agreement between the individual and global m -values is better for the mesophilic WT ProtL, where the electrostatic contribution is almost negligible.

In light of the CLEANEX results for fast exchange in the presence of KCl, one could argue that the intrinsic exchange rate constant, k_{rc} , might be changing with KCl, and that the stabilization seen in H/D exchange experiments is an apparent effect from the slower water dynamics in the presence of KCl. Experiments with Ala–Ala dipeptides in the presence of salts did not show a salt modulation of the intrinsic exchange rates (data not shown), but we have to concede that the error of those measurements was considerable. Nevertheless, pairwise comparison of exchange rates between halophilic and non-halophilic systems (haloSRC *versus* uSRC for CLEANEX; Kx5E *versus* WT ProtL for H/D exchange) largely neutralizes any effect unrelated to the haloadaptation process. In a way, the strategy is analogous to the thermodynamic and mutant cycles (Bosshard, Marti et al. 2004, W.B., Cohen et al. 2015). Therefore, differences shown in the lower barplot of figure R23 correspond purely to the *preferential ion exclusion and differential unfolded state destabilization* of the halophilic Kx5E *versus* the mesophilic ProtL.

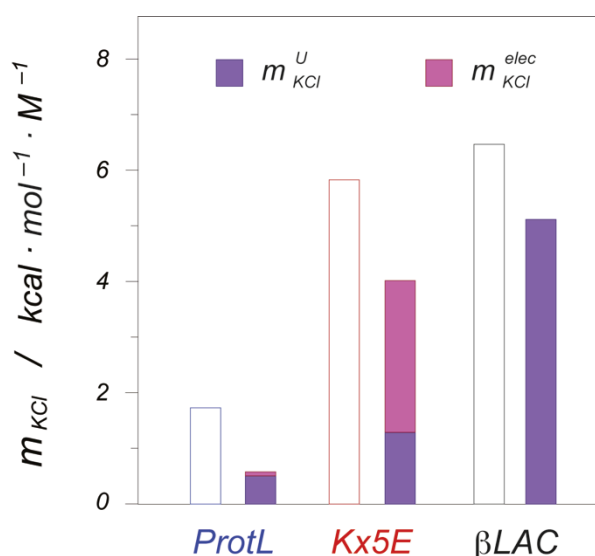


Figure R26. KCl dependent stabilization. Global stabilization m_{KCl} derived from thermal denaturations (white bars) is compared with the contributions from the unfolded state preferential exclusion m_{KCl}^U (purple bars) and from the electrostatic interactions with the carboxylic groups in the folded conformation m_{KCl}^{elec} (pink bars).

Preferential ion exclusion from the hydration layer has since long been linked to thermodynamic destabilization of the unfolded state (Arakawa and Timasheff 1984). Preferential exclusion can be rationalized by a too high entropic penalty associated to specific conformations for constructive ion–protein interactions in disordered states, not compensated by the weak binding enthalpy; instead of co-solubilising, both components compete for separate solvation.

In summary, our results show that the unfolded state of halophilic proteins cannot form the multiple stabilising cation–carboxylate interactions prevailing in the well–ordered surface of the folded conformation. On the contrary, *ions are preferentially excluded from the surface of halophilic unfolded conformations. The resulting destabilization of the unfolded state is comparable in magnitude to the stabilization of the folded state, being another distinctive driving force for salt stabilization of halophilic proteins and, hence, haloadaptation.*

H/D exchange points to a general unspecific mechanism, whose effects are spread throughout the whole sequence.

H/D exchange analysis is consistent with preferential ion exclusion from the unfolded state of halophilic proteins,

thermodynamic quantification of which shows a moderate contribution to salt stabilization and haloadaptation.

4.5. CONTRIBUTION FROM EXCLUDED VOLUME AND OTHER INTERACTIONS

We have so far identified and quantified two main contributions to the total salt-induced stabilisation of halophilic proteins: i) electrostatic stabilisation of the folded state through multiple weak K^+ -carboxylate interactions and ii) unfolded state destabilization due to preferential ion exclusion. Yet, both contributions cannot account for the entire global stabilisation. Additional contributions may derive from the mutually excluded volume effect and from ion-protein interactions other than the ones already considered (those with carboxylates and NH groups from backbone amides and lysines). The total m_{KCl} value then breaks down into (Schellman 2002):

$$m_{KCl} = m_{KCl}^{elec} + m_{KCl}^U + RT \cdot \Delta X + m_{KCl}^{other} \quad ;$$

Equation R14. Contributions to overall KCl stabilization. Global salt stabilization as the sum of the contributions from: electrostatic interactions, unfolded state destabilization through preferential exclusion, excluded volume and other interactions.

Excluded volume contribution arises from the entropic penalty of creating a cavity to host the protein molecule (Graziano and Merlino 2014). It *favours folded conformations* because denatured extended conformations occupy larger volumes and are thus penalized with respect to the native compact conformations. The *presence of cosolutes enhances the effect* because cosolute molecules are excluded from the volume occupied by the protein (Davis-Searles, Saunders et al. 2001). As a result, the energetic contribution of excluded volume to protein stability, $RT \cdot \Delta X$, is directly proportional to cosolute concentration, same as other m -values.

The ΔX contribution can be calculated as the volumetric difference (in \AA^3) between the native and denatured conformations. It can be estimated from a geometric description of the protein in the folded and unfolded states, as described in a protocol previously reported (Schellman 2003). We calculated the ΔX value for mesophilic and halophilic ProtL variants. The PDB files 1HZ6 (O'Neill, Kim et al. 2001) and 2JZP (Tadeo, López-Méndez et al. 2009b) were used as models for the WT and Kx5E ProtL respectively, and their volumes were computed using MSRoll (Connolly 1993). An extended model was employed as a descriptor for the unfolded state, using residual solvent accessible areas previously reported (Creamer, Srinivasan et al. 1997). The ionic radii employed in the calculations are 1.18 \AA for chloride and 1.33 \AA for potassium. The resulting calculated excluded volume contribution for the ProtL variants is equal to 3.2 $\text{kcal} \cdot \text{mol}^{-1} \cdot \text{M}^{-1}$ at 298 K (Figure R25). This value is consistent with a previous estimate for the effect of NaCl on a 50-residue globular protein ($\sim 4.6 \text{kcal} \cdot \text{mol}^{-1} \cdot \text{M}^{-1}$) based on the work of cavity creation by means of scaled particle theory and the volume packing density of NaCl solutions (Pica, Krauss et al. 2013).

Within calculation uncertainty (mainly arising from the estimates for the volume of the unfolded state), the value is the **same for both ProtL variants**. Our ^{15}N relaxation data on uSRC and haloSRC (Figures R15–16) suggest that both mesophilic and halophilic denatured conformational ensembles are similar. Together with the high structural homology found in folded halophilic proteins, this suggests that excluded volume contribution ΔX is the same for mesophilic and halophilic sequences. Moreover, our data on IDPs show that halophilic and mesophilic denatured conformational ensembles remain invariant upon salt addition. In a similar way, no conformational changes have been reported in folded proteins in the presence of molar salt concentrations (Tadeo, López-Méndez et al. 2009b). Therefore, **excluded volume contribution is invariant with salt concentration** and can be described as a constant contribution, $RT \cdot \Delta X$, to the m -value as shown in equation R14.

As shown in Figure R25, the addition of excluded volume contribution to those already reported is not enough to account for the whole m_{KCl} value. Besides, since the degree of stabilization due to the excluded volume is the same for halophilic and non-halophilic sequences, it is *not related to haloadaptation*. This is not surprising because the biological function of proteins, which should be conserved, is intimately ligated to the structure and topology of the native conformation. These are, in turn, ultimately determined by the sequence of the protein (Sosnick and Barrick 2011). Therefore, it seems that no structural changes occur in the evolution of salt-accumulating organisms due to the conservative nature of such a simplistic adaptive strategy.

According to Schellman model (Schellman 2003), remaining cosolute contributions would rely on **intermolecular interactions**. Interactions other than the ones already determined experimentally (carboxylates, lysines and backbone amides) result from the comparison between all estimated contributions and the global m_{KCl} value. They are represented by the term $m_{\text{KCl}}^{\text{other}}$ (Equation R14 and Figure R25). Such contribution accounts for the interactions of KCl ions with *other polar residues, the hydrophobic surface of the protein, etc.* It is negative for both ProtL variants – hence, destabilizing – due to the larger area exposed to the solvent in the denatured conformation. Remarkably, it is smaller for the halophilic Kx5E ProtL ($-1.1 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$) than for the WT ProtL ($-2.2 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$), *in line with the observed preferential ion exclusion from the unfolded state of halophilic proteins*.

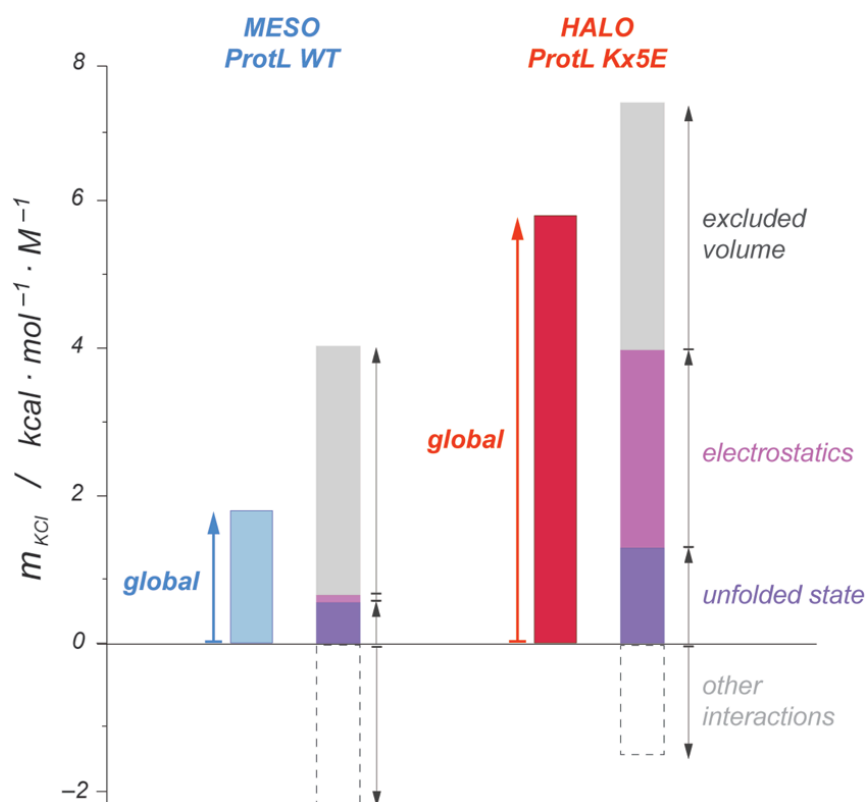


Figure R27. Contributions to KCl stabilization of ProtL. Bar plot showing the experimental global m -values for WT (blue) and Kx5E (red) ProtL as well as the different thermodynamic contributions of KCl to global protein stabilization: electrostatic interactions with the side chains ($m_{\text{KCl}}^{\text{elec}}$, pink), preferential ion exclusion from amide backbone in the unfolded state (m_{KCl}^u , purple), excluded volume (ΔX , grey) and the loss of other interactions upon folding ($m_{\text{KCl}}^{\text{other}}$, white), which is always destabilizing, but less for halophilic proteins. Both electrostatic interactions and preferential ion exclusion contribute at a similar extent to haloadaptation.

5. MODEL

"Science is the poetry of reality"

Richard Dawkins

5. A MODEL FOR PROTEIN HALOADAPTATION

Proteins from halophilic bacteria constitute an excellent example of how simple rules encoded in the protein sequence provide all the information required for the evolutionary adaptation of an organism to a hostile environment. In this case, the selection of certain groups of amino acids determines the adaptation to conditions of extreme salinity (Paul, Bag et al. 2008). This adaptive mechanism is universal since phylogenetically distant species have reached the same solution (Oren 2006), but it lacks a complete mechanistic explanation. Traditional models have focused on the acidic nature of the surface of halophilic proteins and the interaction thereof with solvent cations. Although these models could successfully explain a lot of experimental observations, they turn out to be insufficient when applied to more complex phenomena (Introduction 1.4.1).

This work aims to fill that gap. Our group has dedicated almost ten years to the study of the effects of salt on the stability of proteins with a particular focus on halophilic proteins. During that time, we made significant contributions to the field. The most important was probably the unveiling of the role of solvent exposed hydrophobic area as the main determinant for haloadaptation; which allowed us to artificially emulate halophilic evolution and tune the halophilic character of proteins (Tadeo, López-Méndez et al. 2009a, Tadeo, López-Méndez et al. 2009b). We have now applied the acquired know-how to study the main contributions to salt stabilization for a wide set of systems. *By means of **NMR** in combination with other biophysical techniques we have performed an unprecedented and comprehensive thermodynamic characterization at the level of **individual residues**. We have also expanded the study to the **denatured state** of halophilic sequences by using intrinsically disordered proteins as a system approximation. This has been, to our knowledge, the first reported approach to the halophilic unfolded state. All the information obtained has been integrated into a **new mechanistic model**. Some of the previously reported ideas are embraced, but restricting their reach and reinterpreting them on the basis of more general and fundamental principles.*

The new model proposed here also encompasses aspects that were not explained by traditionally established views. It accounts not only for protein stability and solubility, but also for the complex modulation of the **enzymatic activity** of halophilic enzymes, **Hofmeister** effect, the effect of salt in the **folding** landscape of halophilic proteins and the effect of organic **osmolytes**. As a result, its range of validity is expanded with respect to previous available models. In the following sections, the main ideas and implications of our proposed mechanistic model are carefully described in the frame of a semi-quantitative picture.

5.1. DISENTANGLING THE ENERGETIC CONTRIBUTIONS TO HALOADAPTATION

The main feature of salt adapted proteins is their high stability in environments with molar salt concentrations. In this study, we have analysed salt dependent protein stabilization in order to separate and quantify the energetic contributions of individual amino acids in both the folded and the unfolded states. Our results indicate that stabilization occurs throughout the whole sequence, thus pointing to some general effect mechanism. Accordingly, we show how *both protein states effectively determine the overall energetics of haloadaptation. Two distinct but **complementary mechanisms** alternate to achieve salt stabilization of halophilic proteins: i) favourable weak and unspecific **electrostatic interactions** between cations and carboxylates that stabilise the folded state and ii) unfavourable preferential **ion-exclusion** destabilizing the unfolded state.*

5.1.1. Non specific interactions in the folded state

Most attempts to explain haloadaptation have focused on the stability of the folded state, looking for specific structural features and side-chain arrangements of their prevalent acidic residues. The pioneer works from Mevarech, Zaccai, Madern, Ebel and others (see Introduction) invoked such highly specific constructive interactions. However, this and other studies show that synthetic halophilic proteins, incorporating no particular design other than altering the amino acid composition towards that found in halophilic proteins, are equally stabilized by KCl. Accordingly, our data reveal that *individual electrostatic contributions to protein stability are **not sufficient** to explain neither salt stabilization nor haloadaptation.*

At low salt concentrations, the electrostatic repulsion due to the abundance of acidic residues destabilizes halophilic proteins. The extent of this destabilization can explain the low intrinsic stability of halophilic proteins as compared to mesophilic proteins. Upon salt addition, the folded state of a halophilic protein is greatly stabilised by the *numerous constructive **intermolecular interactions** of its prevalent carboxylate groups with the salt cations.* This not only provides a favourable enthalpic contribution, but also enables optimal co-solvation of both components within a shared hydration shell. In contrast, such interactions are negligible in non-adapted proteins. The model is depicted in figure D1 for ProtL. At the extreme halophilic salt concentrations, halophilic folded states become more stable than mesophilic.

The nature and strength of such interactions is not consistent with previous models; which propose specific strong binding sites for hydrated cations formed by patches of carboxylic groups. We have performed, for the first time, a quantification of the energetics and the strength of these individual electrostatic interactions in halophilic proteins. Our results show that the *interactions are rather **weak** and do not require any specific arrangement.* Accordingly, our ¹⁷O-MRD experiments evidence that the hydration shell of halophilic proteins shows residence times and dynamic properties equivalent to that of mesophilic proteins, thus ruling out the presence of tightly bound hydrated cations. In turn, the accumulation of acidic residues is the main driving force for the attraction of cations to the protein surface. The *stabilizing contribution is consequently due to the **unspecific screening** of the electrostatic repulsion between negatively charged carboxylates.* Nevertheless, we show that the addition of all these electrostatic contributions is not enough to explain the

larger KCl stabilization of halophilic proteins. We, as well as others (Graziano and Merlino 2014), have tested other structural parameters of the folded state of halophilic proteins without finding any distinct features. We then decided to look down into the unfolded state in order to seek for novel determinants of halophilic adaptation.

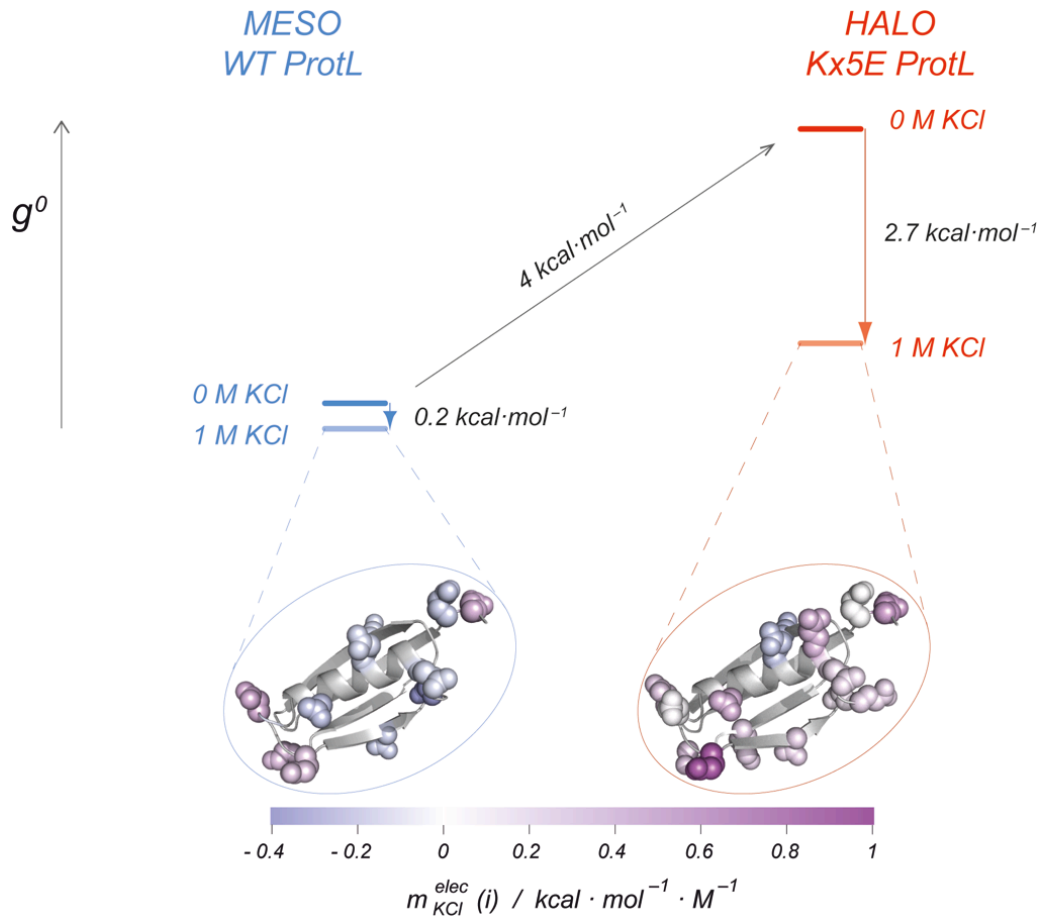


Figure D1. Folded state haloadaptation. The upper diagram shows the energetics of folded state stabilization from 0 to 1 M KCl for mesophilic ProtL (left, blue) and halophilic Kx5E (right, red). Mesophilic folded state is only marginally affected by salt. The electrostatic repulsion due to the increase in acidic residues destabilizes the halophilic mutant but, in turn, makes it greatly stabilized upon salt addition (lower structures).

Weak unspecific electrostatic K^+ -carboxylate interactions occurs in the folded conformation of halophilic proteins.

The stabilization arises from the screening of the electrostatic repulsion between negative charges in the surface.

Electrostatic contribution is not enough to explain the total KCl stabilization.

5.1.2. Preferential ion exclusion in the unfolded state

Protein thermodynamic stability results from the Gibbs free energy balance between folded and unfolded states. The unfolded state of folded proteins is hardly accessible. More often than not, its study requires extreme experimental conditions or denaturants that could interfere with the effect of ionic cosolutes. At the same time, the number of halophilic systems available is limited and no unfolded halophilic sequences have been described up to date. Thus, we have adopted a creative combination of intrinsically disordered proteins and rational design as an approach. This is, to our knowledge, the *first systematic study of the unfolded state of halophilic proteins*. We have collected a plethora of experimental data and performed a thorough thermodynamic and structural characterization. As a result, we managed to report and quantify new contributions to the effect of cosolutes in protein stability and, in particular, haloadaptation.

Our NMR relaxation data reveal that the *local and global dynamics of the unfolded states are very similar in both halophilic and mesophilic proteins*. Within the sensitivity of our experimental design, halophilic amino acid selection does not alter the overall compaction or the conformational ensemble of the unfolded state, as neither does the addition of KCl for none the halophilic or the mesophilic unfolded states. Thus, similarly to native conformations, no structural changes occur in the denatured states. As a result, the excluded volume contribution, known to stabilize folded and compact conformations, remains invariant along the halophilic evolutionary process.

Regarding intermolecular interactions, *no interactions between K^+ cations and negatively charged carboxylate moieties are detected*. Unlike native halophilic conformations, the contribution of electrostatic interactions in the unfolded state is almost negligible. In fact, the halophilic electrostatic profile is nearly identical to that of mesophilic unfolded states except for a slight destabilization due to repulsion – much lower than the one measured for the folded state (Figure D2). The large conformational freedom of the unfolded state probably allows the protein to find more energetically favourable positions in which the electrostatic repulsion is minimized. This entropic mechanism partially compensates the unfavourable enthalpic contribution of charge repulsion. As a result, the lower negative charge density is not enough to attract positively charged K^+ , which explains why the electrostatic stabilization mechanism is exclusive of the folded halophilic state.

Interactions other than those with carboxylate groups were also tested. Our chemical shift perturbation analysis and the exchange measurements (CLEANEX and H/D exchange experiments) at increasing KCl concentrations demonstrate the *preferential exclusion of KCl ions from the backbone of halophilic unfolded states*. In other words: **halophilic amino acid composition favours ion exclusion from the protein surface**. This results in preferential hydration and ion exclusion from the unfolded protein's solvent layer, with an entropic penalty that increases with salt concentration. Preferential exclusion has an energetically destabilizing effect (Lee and Timasheff 1981) because constructive protein–ion interactions are no longer occurring. Instead, both components have to compete for their separate solvation (Figure D2). This imposes an entropic penalty that is especially unfavorable at high salt concentrations, where the thermodynamic activity of water is decreased and the amount of ions a protein has to compete with for hydration is increased. The degree of preferential exclusion as well as the extent of the destabilization is also directly proportional to the area of protein exposed to solvent. By LeChatelier principle, the system will minimize the

thermodynamically unfavourable effect of ion exclusion by populating the state with the smallest surface area, i.e. the folded state. As a result, the unfolded state is largely destabilised at high salt concentrations, whereas the folded state is indirectly favoured (Figure D2).

Preferential exclusion has since long been described by Sergei Timasheff for neutral and organic cosolutes (Lee and Timasheff 1981, Kendrick, Chang et al. 1997, Xie and Timasheff 1997), but it has never been applied in the context of halophilic proteins and ionic cosolutes. What we report here is not only the quantification and confirmation of a key effect to understand the effects of cosolutes on proteins, but also the unanticipated finding that preferential ion exclusion is enhanced in halophilic proteins. The resulting concomitant salt destabilization of the unfolded state is the other main driving force behind haloadaptation. Halophilic amino acid selection enhances the preferential ion exclusion from the protein surface in the unfolded state, increasing protein sensitivity to the cosolutes. This constitutes a major advance in the field since it unveils a new adaptive mechanism not considered until now. Moreover, some of the previously unexplained features of halophilic adaptation and protein–salt interactions can be understood in the lights of this new idea.

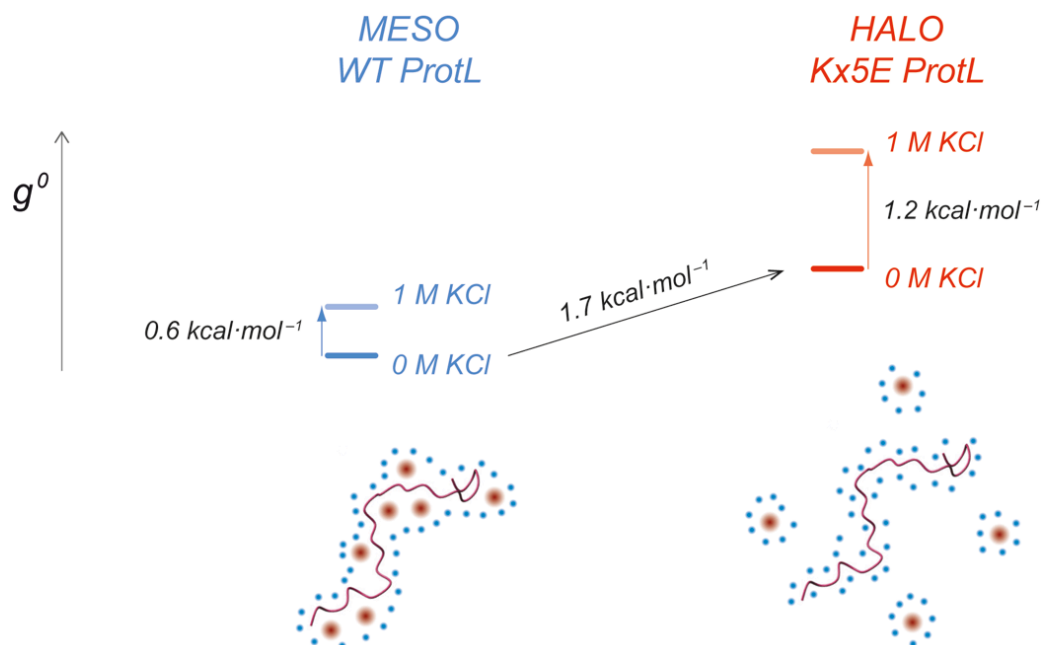


Figure D2. Unfolded state haloadaptation. The upper diagram shows the energetics of the unfolded state of mesophilic WT (left, blue) and halophilic Kx5E (right, red) ProtL. In the absence of salt, the halophilic unfolded state is slightly destabilized (much less than the unfolded state) due to electrostatic repulsion. Upon KCl addition, ion exclusion and preferential hydration (pictures below) cause a larger destabilization of the unfolded halophilic state.

No significant contributions to haloadaptation are made by electrostatics (interactions) or conformational changes (excluded volume).

For the first time, ion–exclusion from halophilic unfolded states is reported.

5.1.3. Towards a semi-quantitative model for protein haloadaptation

Our results indicate that both folded and unfolded states determine the overall energetics of haloadaptation. We report the two main mechanisms contributing to the salt stabilisation of halophilic proteins in high molar salt concentrations: i) *stabilization of the folded state by cationic non-specific screening of the electrostatic repulsion between acidic groups* and ii) *destabilization of the unfolded state by preferential ion exclusion*. Haloadaptation relies on these two alternate mechanisms that cooperatively affect both the folded and unfolded states, as illustrated in figure D3. The unfolded state excludes ions from the surface, therefore avoiding ion-protein interactions. This results in a high entropic penalty, unfolded state destabilization and the balancing of the equilibrium towards the folded state. Upon folding, the clustering of the acidic groups *switches on* K^+ -carboxylate interactions, which decrease the electrostatic repulsion and stabilise the folded state. Influencing work from John A. Schellman already stated that the balance of the interactions between the folded and the unfolded states was critical for the changes in protein stability induced by cosolvents (Schellman 1990, Schellman 2003). Haloadaptation is just a particular case in which halophilic composition modulates salt-protein interactions.

This **alternate switching** of interactions along the folding coordinate is represented in the *semi-quantitative* energetic diagram D3 for ProtL. In the absence of salt, the designed Kx5E mutant (Figure D3, right) is less stable than its mesophilic counterpart WT ProtL (Figure D3, left). The free energy difference is mainly attributed to the halophilic folded state because of the large intramolecular electrostatic repulsion between the acidic groups (and because according to our NMR relaxation data, the unfolded state ensemble remains mostly the same). At high salt concentration, the *folded state* of Kx5E is *greatly stabilized by pervasive and weak unspecific interactions* of the carboxylates with the K^+ cations. Complementary, the *unfolded state is destabilized by salt due to preferential exclusion* of ions from the surface. In turn, both the folded and unfolded states of WT ProtL are much less sensitive to salt.

A semi-quantitative bar-plot representation in figure D4 shows all the relevant measured contributions for WT and Kx5E ProtL. The stabilizing effects of electrostatics (pink) and the interactions with the unfolded state (purple) are greater for the halophilic protein. Despite the electrostatic term being larger, they are both comparable in magnitude. They can account for great part, but not all, of the total KCl-induced stabilization (m_{KCl}). The residual term can nonetheless be explained by the excluded volume contribution (grey) and the interactions with other parts of the protein (white). The latter are also decreased in halophilic sequences due to preferential ion exclusion and the subsequent loss of interactions in the unfolded state.

Two opposite mechanisms cooperate to synergistically stabilize halophilic proteins:

- i. Weak unspecific K^+ -carboxylate interactions in the folded state*
- ii. Preferential exclusion from the unfolded state*

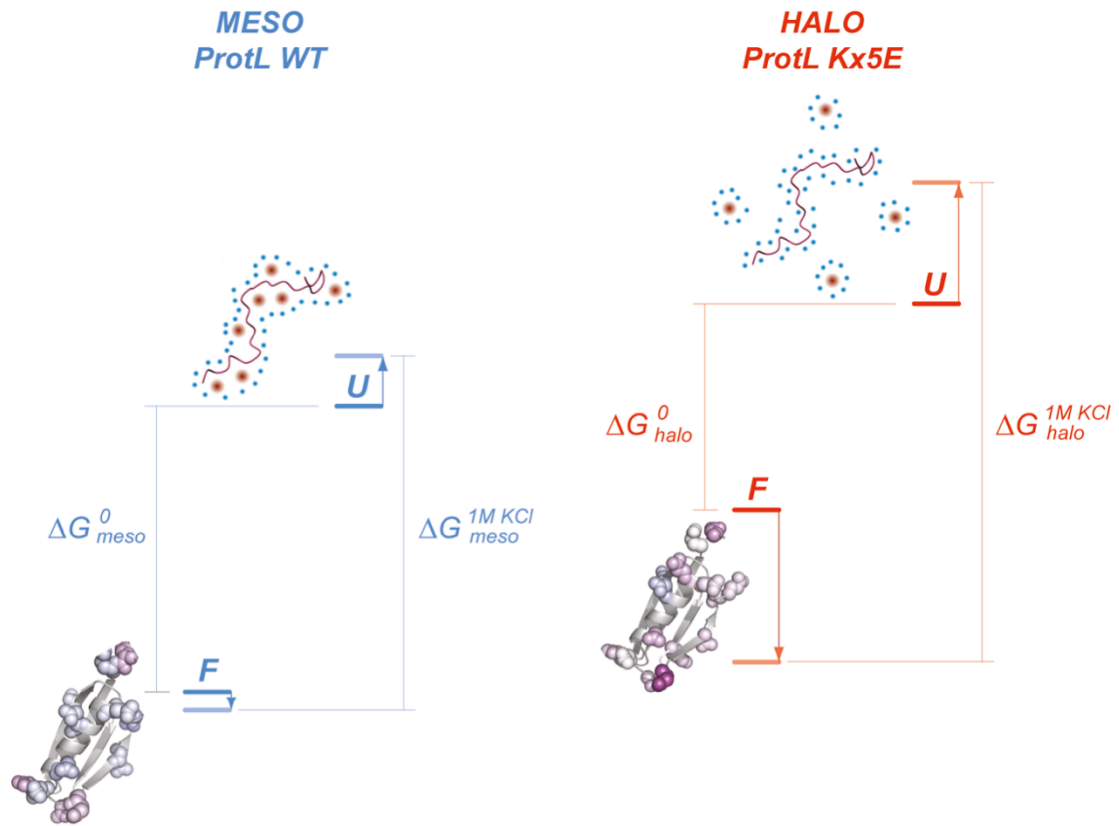


Figure D3. Energetics of protein adaptation to hypersaline environments. The scheme shows the energy levels in the absence (dark lines) and presence (faint line) of 1M KCl for halophilic Kx5E (right) and mesophilic WT (left) ProtL. In the absence of salt, mesophilic protein is more stable, but the situation is reversed at 1 M KCl. The depicted halophilic unfolded state reflects the destabilization due to the preferential exclusion of ions from its hydration layer, whereas the folded ProtL shows the electrostatic stabilization.

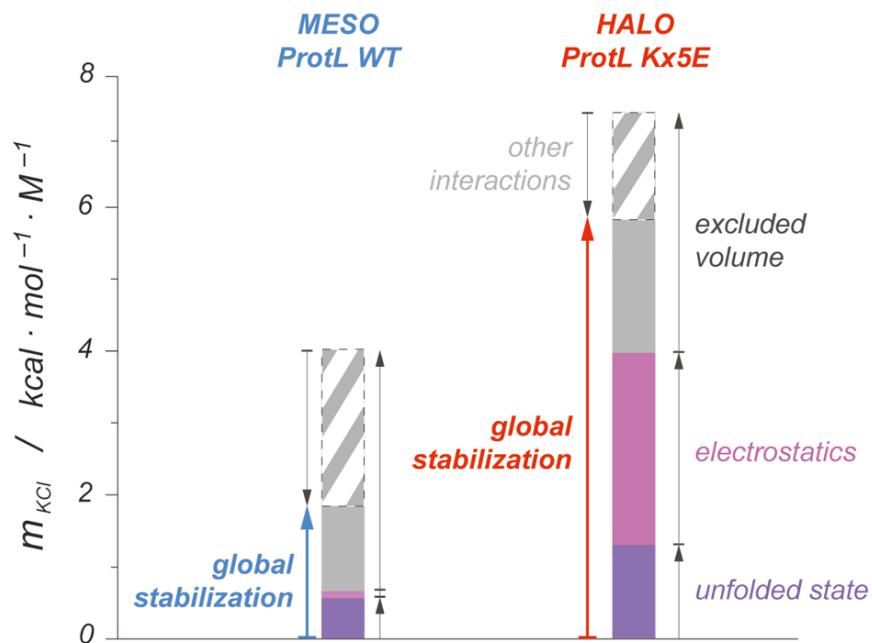


Figure D4. Decomposition of ProtL overall KCl stabilization. m_{KCl} is broken down to the distinct energetic contributions analyzed in this work: screening of electrostatic repulsion in the folded state, ion exclusion from the unfolded backbone, excluded volume and interactions with other parts of the protein.

In summary, we propose a model in which protein–KCl interactions in both folded and unfolded states intervene in salt stabilization and the process of haloadaptation, adapting former contributions from **Shellman & Timasheff** to the particular case of halophilic proteins and KCl. Two differential mechanisms for KCl stabilization are found with respect to mesophilic proteins: constructive weak unspecific protein–K⁺ interactions in the folded state and preferential ion–exclusion in the unfolded state. *Subsequent stabilization of the folded state and destabilization of the unfolded state cooperate to synergistically stabilize the native conformation. Halophilic amino acid selection amplifies this alternate switching of the interactions.* As a result, the thermodynamic free energy of unfolding of halophilic proteins is greatly increased in the presence of salt. A picturesque notion of the proposed model is displayed in Figure D5:

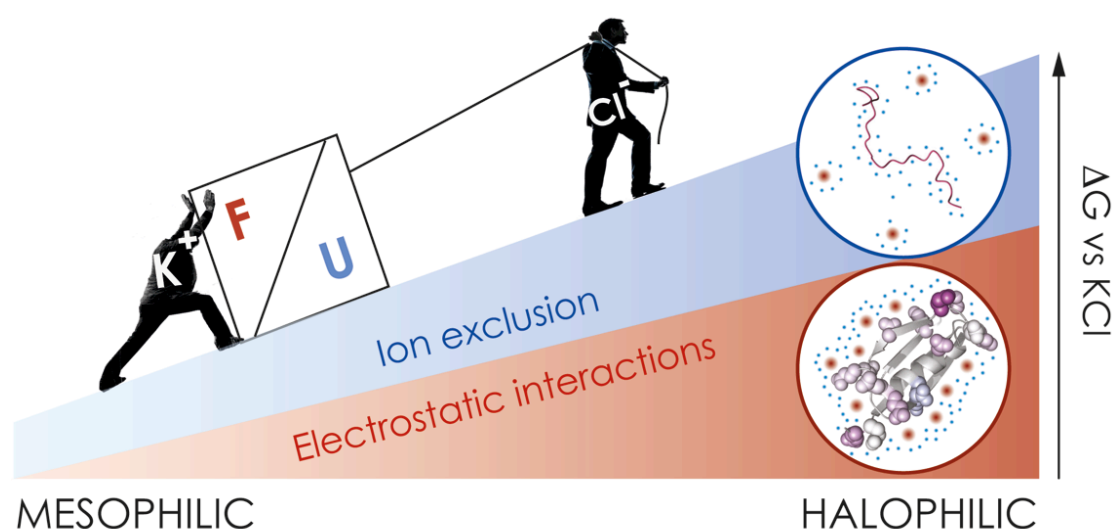


Figure D5. Haloadaptation mechanism. Alternate switching of unspecific protein–KCl interactions in the folded and unfolded states cooperate to stabilize halophilic proteins.

Halophilic amino acid composition enhances the alternate switching of ion–protein interactions

5.2. WHY THIS AMINO ACID SELECTION? STRICT HALOPHILICITY

Adaptation of proteins to high salt environments relies on the selection of a very specific set of amino acids. This evolutionary strategy is widespread among different taxonomically unrelated families, thus pointing to a unique universal mechanism. We have identified the basis and main contributions to haloadaptation, which happen to be intimately ligated to the interactions between the constituents of the ternary ion–protein–solvent system. The nature of these contacts is determined by the composition of the surfaces involved, which is ultimately modulated by the choice of the side–chains. Our results show that residues throughout the entire protein sequence contribute energetically to haloadaptation. Thus, the relative abundance of certain s, rather than their specific location within the protein, is the key for haloadaptation.

The mechanism proposed here requires a *minimum of charge density on the protein's surface: enough to stabilize multiple weak and unspecific interactions with solute cations in the folded state that, in turn, cannot form properly in the unfolded state*. **Aspartate**, and to a lesser extent **glutamate**, are optimal residues for this purpose. They provide *highly polar, negatively charged moieties on short side chains* that are readily hydrated without imposing high entropic costs when fixing a favorable conformation for cation interactions. The abundance of negatively charged residues has been proposed to be directly responsible of halophilic adaptation through the formation of strong K^+ –binding sites. We demonstrate, however, that such interactions are weak and unspecific. The subsequent protein stabilization arises from the screening of the intermolecular electrostatic repulsion between the negative charges in the surface of the protein. Thus, the *accumulation of acidic residues rather than their specific geometry triggers the cation–induced stabilization of the protein*. We also show that the increase in negative charge in the surface is not enough to explain haloadaptation, in line with previous work in our lab directly correlating salt–stabilization to the reduction in ASA rather than to negative charge (Tadeo, López-Méndez et al. 2009a). Then, why has evolution developed such a strong bias for acidic residues? Several reports associate overall charge to solubility (Madern, Ebel et al. 2000, Coquelle, Talon et al. 2010, Adrover, Mariño et al. 2014). Indeed, previous studies in our lab showed a correlation between the increase in negative charge and **improved solubility** (Tadeo, López-Méndez et al. 2009b). *Intermolecular attractive interactions leading to aggregation are enhanced at high salt concentrations due to a lowered water activity. Electrostatic repulsion between negatively charged surfaces, in turn, prevents aggregation and precipitation*. Thus, not only the highly abundant acidic residues cooperate to stabilize proteins in halophilic environments, but also contribute to favour solubilisation. In fact, extensive precipitation of halophilic proteins was observed upon neutralization of acidic groups during the pH titrations in the presence of KCl.

Acidic residues improve stability (interactions with K^+), solubility and avoid aggregation (electrostatic repulsion)

According to traditional interpretations, the use of KCl as an osmoprotectant requires interactions between K^+ cations and acidic side chains on the protein surface. *Mutations progressively increasing negative charge eventually lead to protein unfolding in the absence of salt due to the elevated intramolecular electrostatic repulsion*. Thus, obligate protein halophilicity would become a mandatory “price to pay” as a consequence of the resulting

acidic nature of the proteomes in extreme halophiles. We challenge this view and propose instead that **obligate protein halophilicity** is a *non-adaptive property resulting from a genetic drift*. In support of this interpretation, strongly acidic proteome in KCl accumulating *Halorhodospira halofila*, a bacterial halophile that grows optimally above 3 M NaCl, has been recently reported to remain fully functional even at cytoplasmic salt concentrations as low as those from *E. coli* (Deole, Challacombe et al. 2013). It is very unusual to find extreme-adapted KCl accumulating organisms capable of reducing its intracellular salt concentration, but it provides the biological evidence that an acidic proteome does not necessarily require KCl accumulation.

Electrostatic repulsion unfolds obligate halophiles at low [KCl]

The role of negative charges seems clear but still a couple of questions remain unresolved. First, *why aspartate is preferred over glutamate?* Glutamic acid would also improve solubility by mutual repulsive interactions simultaneously providing weak binding sites for K^+ binding. Secondly, *why is an acidic instead of an alkaline proteome chosen?* Positively charged surfaces could also increase protein solubility. The solutions to these questions rely on the ability of halophilic proteins to maintain an efficient solvation even in media with low water activity. As explained in the previous section, this has been accomplished by **reducing the area exposed to solvent**. Aspartate is preferred because its side chain is one methylene shorter than glutamate's. Among basic residues – histidine, arginine and lysine – *no equivalent to aspartic acid can be found because all three alkaline residues have bulky hydrophobic side chains*. Particularly, **lysine's drastic reduction in the composition of halophilic proteins** could be explained by the extraordinary entropic penalty accompanying the hydration of its long, disordered side chain. Its large hydrophobic area disrupts the creation of a well-ordered hydration network around the protein, with water molecules surrounding the methylenes probably adopting energetically unfavorable clathrate-like conformations. Accordingly, our experimental results provide no indication of interaction between surface lysines and Cl^- anions.

The reduction in hydrophobic content is further accomplished through a **decrease in amino acids with large apolar side chains**. Cysteine, methionine, leucine, tryptophan, isoleucine, phenylalanine and arginine are the most penalized residues after lysine. For instance, leucine-to-valine and isoleucine-to-valine are highly conservative mutations that decrease apolar side chains without altering essential hydrophobic contacts. In fact, they rank amongst the top-ten substitutions when going from mesophilic to halophilic proteins.

This large decrease in hydrophobic area is accompanied by an **increase in polar side chains**, threonine and aspartate being the most favored amino acids. Aspartate abundance can be explained by its negative charge, but not threonine. Our results show how halophilic proteins preferentially exclude KCl ions from their hydration layer, especially in the unfolded state. As a result, unfolded state is destabilized and folded state is indirectly favored in the presence of KCl. Ion exclusion, together with the parallel preferential hydration, occur mainly in polar surfaces, where strong competition with the salt ions for solvation results in a local ion concentration lower than that in the bulk solvent. *Halophilic adaptation, through the decrease in apolar hydrophobic area and the coupled selection of amino acids with highly hydrophilic side chains, enhances ion exclusion and, as a result, the stabilization mechanism.*

Introduction of the most halophilic amino acids aspartate, threonine and glutamate, with their short and often negatively charged hydrophilic side chains, thus enhances preferential ion exclusion and protein solvation for the halophilic proteins studied. **Threonine**, in particular, *provides a short and highly polar surface that helps solvation and preferential hydration*, thus facilitating protein solubilization and stabilization.

Halophilic amino acid selection increases the fraction of polar surface and decreases hydrophobic content in order to provide sites for ion exclusion and preferential hydration

Our model provides a link between halophilic amino acid composition and the mechanisms for salt stabilization and haloadaptation. Negatively charged groups, residues with low hydrophobic content and small polar side chains are preferred in order to facilitate protein solubilization and the *synergistic switch between ion–protein interactions in the folded state and exclusion in the unfolded state*; which ultimately confer halophilic proteins its great stability at high KCl concentrations.

5.3. CONTRIBUTION OF THE HYDRATION OF SOLVENT EXPOSED HYDROPHOBIC AREA

Previous work from our group revealed that the **reduction in solvent accessible area** results in an increase in salt stabilization (Tadeo, López-Méndez et al. 2009b), constituting a key factor in halophilic adaptation even more determinant than overall negative charge. Inspection of figure D6 shows a clear correlation between the changes in solvent exposed area introduced upon mutation and the NaCl (A) or KCl (B) induced stabilization for a set of case studies. We provide now the mechanistic basis to understand the role of the reduction of hydrophobic area in terms of preferential hydration and ion-exclusion.

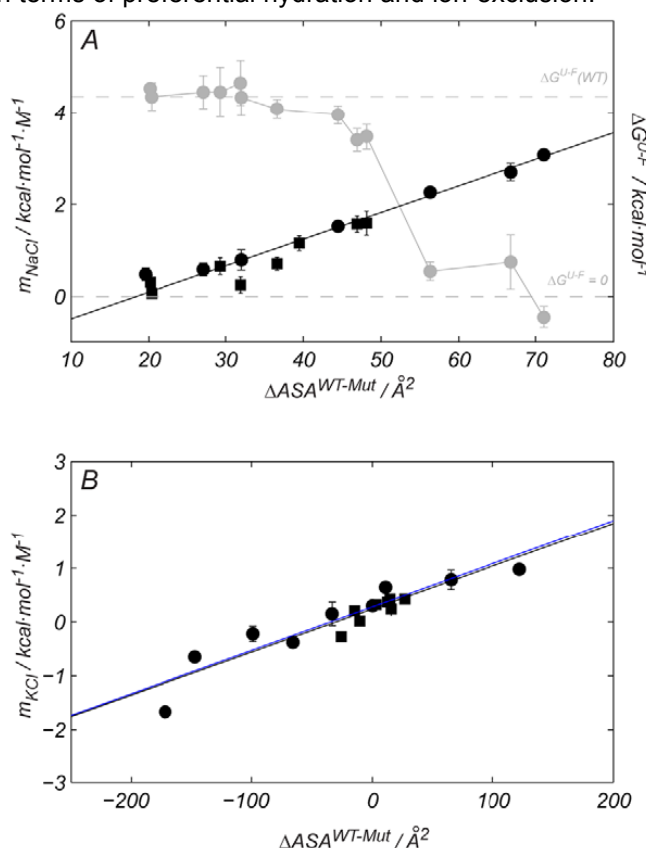


Figure D6. (A) A mesophilic protein ($\Delta G^{U-F} > 0$, $m_{NaCl} = 0$) can be converted into an obligate halophile ($\Delta G^{U-F} < 0$, $m_{NaCl} > 0$). This transition is a continuous process as demonstrated by the empirical correlation found between the m_{NaCl} and ΔASA^{WT-Mut} for ProtL mutants: K for Q (black squares) and K for E (black circles). For each protein, the standard unfolding free energy is shown in gray circles (right ordinate axis). Dashed gray lines mark the ΔG^{U-F} values from 0 to that of WT ProtL. Left ordinate axis for the m_{NaCl} units. (B) An equivalent correlation can also be found for KCl and the D for E and E for D mutants *Haloferax volcanii* (black circles) and *Escherichia coli* (black squares) 1ALigN. Solid black line corresponds to the best linear fit of the data. In (B), black line corresponds to the linear fit including *Hv* and *Ec* 1ALigN whereas the fit of *Hv* 1ALigN alone is depicted with a blue line.

Protein **solvation** is particularly demanding at high salt concentrations, where water activity is reduced and a competition for water molecules takes place between ions and protein. As a result, processes favouring the reduction of exposed area – such as protein aggregation and precipitation – are favoured. Halophilic proteins adapt by diminishing their solvent-exposed hydrophobic area, solvation of which is entropically penalized. The increase in polar groups and the drop in methylenes from lysines favour water-protein interactions, which enhance protein hydration, solubilisation and, at the same time, the unfolded state destabilization by the aforementioned ion-exclusion mechanism. As opposed to non-adapted proteins, halophilic proteins can achieve an efficient solvation even in molar salt concentrations, where water activity is drastically reduced. In support of this idea, the ^{17}O relaxation data show comparable hydration properties for low salt-WT and high salt-Kx6E ProtL.

5.4. HALOPHILIC ADAPTATION AND FOLDING LANDSCAPE

Haloadaptation can be equally understood as the mechanism for folding an obligate halophilic protein. Current views recognize that reduction of the folding problem to thermodynamics, i.e. the free energies of the folded and unfolded states, is often an inaccurate oversimplification, even when experimental data agree well with a two-state folding model (Sali, Shakhnovich et al. 1994). Instead, folding and stability appear mainly governed by kinetics, which are in turn determined by the energetic barriers of the different transitions involved (Dill and Chan 1997, Sosnick and Barrick 2011):

- i. The folding rate (k_F) is defined by the early collapse of the denatured state ensemble (DSE) and foldon (local folding units) formation on the path towards the transition state (TS);
- ii. For the reverse process, the kinetic rate of unfolding (k_U) is governed by the transition going from the native state conformational ensemble (NSE) to the TS.

In line with this interpretation, our results clearly show that the combination of a biased amino acid composition and high KCl concentration significantly alters the protein's energetic landscape at both sides of the TS.

In light of this folding perspective, the EXSY experiments (Figure R21, results section 4.4.1) are reinterpreted (Figure D7, A) in order to extract kinetic information. The kinetic rate constants correspond to the *rate limiting transitions* (DSE to TS for k_F , in blue, and NSE to TS for k_U , in red). The experimental data are reasonably well reproduced by monoexponential fittings, thus ruling out the presence of any detectable intermediates – as confirmed by the existence of only two sets of signals in the $^{15}\text{N}, ^1\text{H}$ -HSQC (Figure R19) and by previous work on ProtL (Scalley, Yi et al. 1997). Logarithmic representation of the calculated kinetic constants (Figure D7, B) shows a typical *Chevron Plot* (Skinner, Wookyung et al. 2014). The linear limbs correspond to the folding and unfolding rate constants, whereas the slopes thereof correspond to the salt stabilization of each transition:

$$m_{eq} = m_{TS-DSE} - m_{NSE-TS} \quad ;$$

Equation D1. Kinetic analysis of KCl stabilization. Overall salt stabilization, m_{eq} is the result of the KCl modulation of both folding and unfolding kinetic processes. In turn, the former results from the DSE to TS transition and the latter results from the NSE to TS transition.

Analysis of the Chevron Plot reports a reduction of the folding barrier of $3.22 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$, which favours the DSE to TS transition. On the other hand, the barrier for the transition from the NSE to the TS increases in $2.37 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$. The resulting KCl total stabilization is $5.58 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$, consistent with the value of $5.62 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$ obtained from independent CD and NMR KCl titrations of Kx7E. As already discussed, both changes go in the direction of favouring the population of the native conformation, but a quantitative estimate of the contribution from each transition can be now obtained. The quantity introduced by Ferhst and co-workers: $\Phi = \Delta\Delta G_{TS-DSE} / \Delta\Delta G_{NSE-DSE}$ (Matouschek, Kellis et al. 1989), serves as a measure of where in the folding trajectory are the changes occurring. A value of $\Phi = 1$ means that all changes occur in the early folding steps leading to the transition state, whereas native state and transition state are equally affected. Alternatively, a value of $\Phi = 0$ would mean that all the stabilization occurs in the final route going from the transition state to the final native ensemble; which implies that new stabilizing interactions are formed in the NSE but not in the TS. The Φ value for the KCl stabilization of Kx7E is 0.58, which means that almost two thirds of the stabilization occurs during the *early collapse of the denatured extended conformations* and foldon formation.

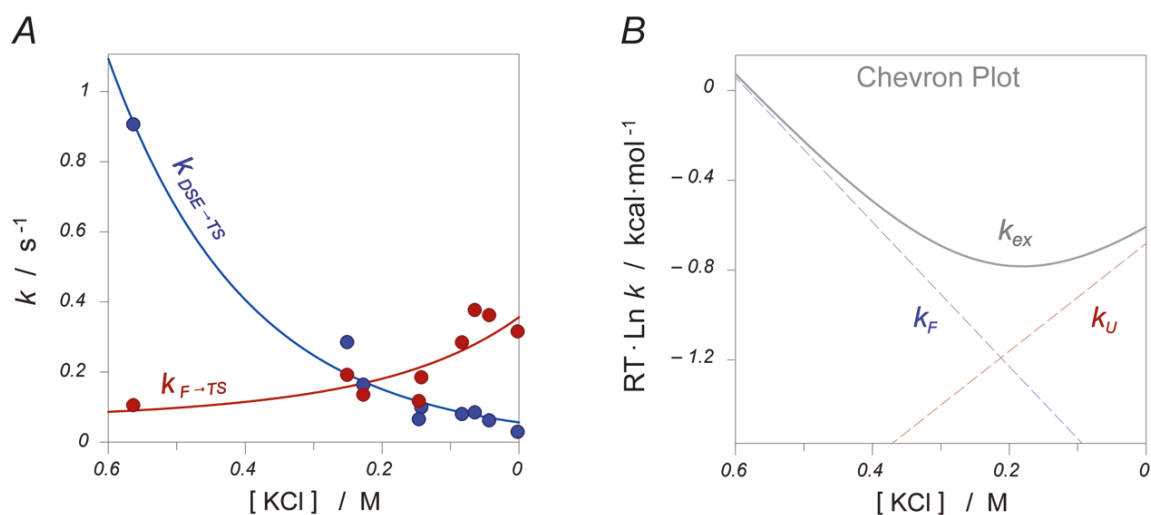


Figure D7. Chevron plot representation of kinetic EXSY data. A) EXSY experimental points (circles) and monoexponential fittings (lines). Folding (blue) and unfolding (red) kinetic rates are respectively determined by the DSE to TS and the NSE to TS rate limiting steps. B) Logarithmic scale shows linear calculated folding and unfolding rates (dashed lines). The observed exchange constant, k_{ex} (grey line) arises from the addition of $k_{\text{U}} + k_{\text{F}}$. Notice that KCl concentrations in abscise axis are inverted in order to maintain the traditional disposition of Chevron Plots with the folding (unfolding) arm on the left (right).

Current views on folding suggest that foldons and main secondary structure are already formed in the TS; therefore most of the backbone area is already buried forming native contacts. Posterior transition from the TS to the NSE proceeds through the burial of side chains, the formation of side chain contacts and of the remaining backbone secondary and tertiary structure (Sosnick and Barrick 2011). Particularly for ProtL, the first and second β -sheets as well as the central α -helix are already formed in the TS, with about two thirds of the surface to be hidden already buried (Guinn, Kontur et al. 2013, Skinner, Wookyung et al. 2014). These data are consistent with our H/D exchange data, which show higher local stability values for first β -hairpin and central α -helix.

Figure D8 shows a hypothetical semi-quantitative energy landscape for WT ProtL (blue). The DSE is represented by a fractional relative buried area $\theta_{\text{ASA}} = 0$, indicating that all the area eventually buried in the NSE is still fully exposed. The DSE of WT ProtL is well described by extended polypeptide conformations since this model predicts a heat capacity change upon unfolding ($\Delta C_{\text{p,calc}} = 937 \text{ cal} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$) (Guinn, Kontur et al. 2013) that is in reasonably good agreement with the experimental result ($\Delta C_{\text{p,exp}} = 876 \pm 50 \text{ cal} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$) (Tadeo, López-Méndez et al. 2009a). Folding progresses through early collapsed conformations with higher θ_{ASA} values. Although most experimental data suggests a simple two-state cooperative folding for WT ProtL (Yi and Baker 1996), several illustrative foldons have been introduced into the folding pathway in Figure D8 (Englander and Mayne 2014) for generalization purposes. The progress along the folding coordinate then leads to the transition state around $\theta_{\text{ASA}} = 0.66$, as suggested by a recent interpretation of the kinetic folding data for ProtL (Guinn, Kontur et al. 2013). This number is in line with our Φ analysis. Since our stabilizing mechanisms occur mainly through interactions and are thus proportional to the area exposed, the Φ value of 0.58 can be given an approximate structural interpretation: about 58 % of the area buried in the NSE is already buried in the TS. Further burial of side chains and formation of all hydrophobic contacts finally produces the native state ensemble of the protein, at $\theta_{\text{ASA}} = 1$.

Figure D8 also includes a plausible folding landscape for the obligate halophilic Kx7E ProtL variant in the absence of salt (dashed red line). Our ^{15}N relaxation data analysis indicates that mesophiles and halophiles share similar conformational landscape of their DSE, in contrast to the unique residual structure found in thermophilic ribonuclease H (Robic, Guzmán–Casado et al. 2003). Nevertheless, the halophilic DSE is expected to be less stable than its mesophilic equivalent due to electrostatic repulsion between the enriched Asp and Glu residues (offset arrow ΔG_1 in Figure D8). We have shown that this destabilization can be estimated from the net difference between corresponding pK_a values for the unfolded mesophilic uSRC and halophilic haloSRC in the absence of salt ($\sim 1.7 \text{ kcal}\cdot\text{mol}^{-1}$). A similar electrostatic mechanism was observed for the acidic denaturation of several ferrocycytochrome c variants with different numbers of acetylated lysine ϵ -amino groups (Goto and Nishikiori 1991). In the absence of salt, folding of the obligate halophile along the reaction coordinate θ_{ASA} is energetically forbidden since the destabilizing electrostatic repulsion is expected to increase with the conformational compaction (dashed red line in Figure D8).

At high salt concentrations ($> 1 \text{ M KCl}$), obligate halophilic proteins adopt a stable fold (Lanyi 1974). Our CLEANEX and chemical shift perturbation data on intrinsically disordered model proteins prove that the DSE of an obligate halophile in the presence of KCl is further destabilized due to preferential ion exclusion from the hydration layer (offset arrow ΔG_2 in Figure D8). During the initial collapse, carboxylate approximation allows cation inclusion, which decreases electrostatic repulsion (offset arrow ΔG_3 in Figure D8). A comparable mechanism has been reported for H^+ association during acid induced protein folding (Goto, Linda et al. 1990). Electrostatic screening of repulsive forces allows the protein to explore more compact productive conformations and progress down the folding funnel. As these two opposite yet complementary mechanisms are proportional to the ASA, their combined effect primarily destabilises the DSE and secondly enables the initial folding collapse.

After the initial surface charge compensation from cation inclusion, the folding pathways for mesophile (blue curve) and obligate halophile (red curve in Figure D8) are expected to be very similar. Actually, transition state and foldons are mainly determined by the native topology encoded in the sequence (Sosnick and Barrick 2011), henceforth the conformational space should remain unchanged. Our results as well as previous studies (Madern, Ebel et al. 2000) demonstrate that K^+ cations constructively interact with carboxylate groups on the surface of halophilic proteins, thus actively stabilising their folded conformation. E.g., a similar mechanism has been reported for the highly acidic repeat protein YopM (Kloss and Barrick 2008), for which considerable stabilization up to 500 mM NaCl was attributed to the unspecific screening of long-range intramolecular repulsion in the folded state. Since folding implies the successive formation of native molecular structure as θ_{ASA} approaches 1 (Englander and Mayne 2014), a progressive cation-induced stabilization is expected for the halophilic protein, represented by the increasing energy offset arrows ΔG_4 in Figure D8, which reaches a maximum of $2.5 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$ for the NSE of Kx5E ProtL.

Depending on the nature of the folding energy landscape, a protein can be classified as a cooperative folder (Type I, folding through a well defined rate-limiting TS) or a downhill folder (Type 0, continuous folding without a clear TS) (Bryngelson, Onuchic et al. 1995). An interesting suggestion arises from the model depicted in Figure D8 for ProtL: the biased amino acid composition of halophilic proteins in combination with the stabilization mechanisms observed for KCl (both proportional to θ_{ASA}) may push the folding landscape from a cooperative pathway towards a more gradual, downhill folder – like scenario.

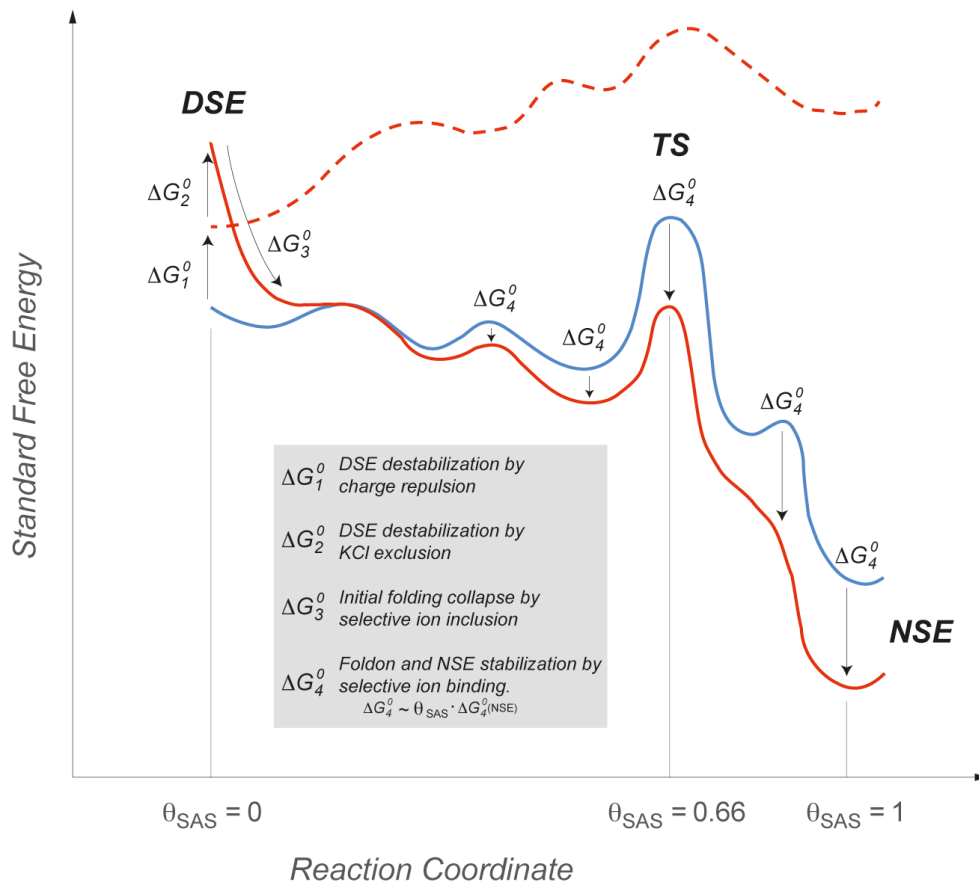


Figure D8. Folding landscape modulation in protein adaptation to hypersaline environments. Representation of the folding energy landscape for mesophilic ProtL (blue) and an obligate halophilic homologue in the absence (red dashed line) and presence (red solid line) of molar KCl concentrations. Horizontal axis shows the folding coordinate as the fraction of ASA buried with respect to the native conformation, θ_{ASA} . In contrast to mesophilic proteins, electrostatic repulsion in the absence of salt destabilizes obligate halophiles at an extent proportional to θ_{ASA} , thus avoiding population of more compact structured conformations (ΔG_1^0). Upon salt addition, ion exclusion penalizes extended conformations (ΔG_2^0); complementary, KCl electrostatic screening helps the protein early collapse along the reaction coordinate (ΔG_3^0), ultimately leading to further energy reduction in the transition state and the folded conformation (ΔG_4^0). The electrostatic stabilization due to K^+ cation – carboxylates interactions is proportional to the compactness of the conformation, as represented by the increasing arrow length.

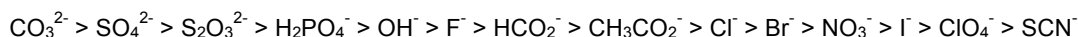
Electrostatic repulsion impedes the early collapse of the denatured conformations.

In halophilic proteins, KCl destabilizes unfolded conformations and screens the electrostatic repulsion, favouring more compact conformations.

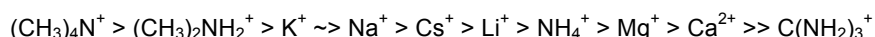
Salt affects the folding energy landscape of halophilic proteins at both sides of the transition state, at an extent proportional to ASA and compactness.

5.5. COMPARISON WITH OTHER IONS AND THE HOFMEISTER EFFECT

Several reports associate haloadaptation with the Hofmeister series (Pundak, Aloni et al. 1981, Ebel, Faou et al. 1999, Ortega, Laín et al. 2011). Originally developed as the minimal concentration of various ions required to precipitate a given protein from aqueous solution (Hofmeister 1888), this typical ordering is now related to some other properties such as surface tension, cavity creation, colloidal assembly and more relevant for us, protein stability. Hofmeister series are dominated by **anion's** properties, since *the effects are more pronounced for anions than for cations*. For a given cation, the anion series is generally written as (Zhang and Cremer 2006):



A less well developed series exist among cations for a given anion:



Although Hofmeister effects on macromolecules in aqueous solution are ubiquitous, the molecular level–mechanisms by which ions operate are only beginning to be unravelled. Over last decades, it has been considered that salt's influence on macromolecular properties was caused by 'making' or 'breaking' bulk water structure by *kosmotrope* or *chaotrope* ions respectively. Certainly, experimental evidences demonstrate the salt–induced increase/decrease in water structure in first and second solvation shells, at least in dilute, isotropic solutions (Marcus 2009). The controversy still lies in the water beyond the immediate hydration shell, where claims have been made about the *"negligible effects of ions on the hydrogen bond structure in liquid water"* (Omta, Kropman et al. 2003). Nonetheless, water molecules are already non–isotropically ordered on more extensive surfaces such as polymers, macroscopic air–solution interfaces or proteins. When dealing with ion effects near surfaces, careful thermodynamic analysis and experimental evidence revealed that making/breaking properties of ions in bulk solutions do not exactly match the Hofmeister series (Batchelor, Olteanu et al. 2004), and no straightforward correlation was found between the bulk water structure and the Hofmeister series (Zhang and Cremer 2006). Instead, influencing work from Paul S. Cremer and other studies have contributed to change the current paradigm in scientific community towards an *understanding of Hofmeister phenomena based on direct interactions between ions and macromolecules* (Zhang and Cremer 2006, Marcus 2009, Nihonyanagi, Yamaguchi et al. 2014).

The dominant role of interactions is particularly true for **anions**, which show a *higher propensity to interact with protein surface* (Zhang and Cremer 2006). Cremer *et al.* use vibrational sum-frequency spectroscopy (VSFS), which provides interface-selective vibrational spectrum, to monitor the order of hydrophobic layers and interfacial water structure. They show in different studies how anion penetration in hydrophobic monolayer followed Hofmeister ordering (Gurau, Lim et al. 2004), with destabilizing anions effectively penetrating in non-polar environments (ClO_4^- , I^- , SCN^-) whilst stabilizing anions (SO_4^{2-} , Cl^- , NO_3^-) penetrate less (Figure D9). These conclusions are further confirmed in other experiments with neutral and charged surfaces (Chen, Yang et al. 2007), as well as with water-protein interfaces (Chen, Flores et al. 2010, Flores, Kherb et al. 2012). Accordingly, they could also explain the Hofmeister dependent solubility of a thermoresponsive polymer in terms of direct anions' interactions with the macromolecule amides and its first hydration shell (Zhang, Furyk et al. 2005). Independent studies from Tahei Tahara and coworkers confirmed the adsorbability of anions to aqueous interfaces and its agreement with the Hofmeister series (Nihonyanagi, Yamaguchi et al. 2014). By systematically probing different surfaces and

ions they managed to *discriminate two distinct interaction mechanisms* for anions and cations: while anions tend to form direct interactions by ion–pairs or hydrophobic contacts with positively-charged or apolar surfaces respectively, cations form indirect interactions through water hydrogen–bonding (solvent ion–pairs) with negatively charged surfaces. This recent avalanche of data settles the new foundations for a preliminary understanding of the Hofmeister effects on proteins and macromolecules. Interactions at the solvent–macromolecule interface determine the ion-specific Hofmeister effects and the compositely roles of anions and cations could help to discriminate the cation and anion series.

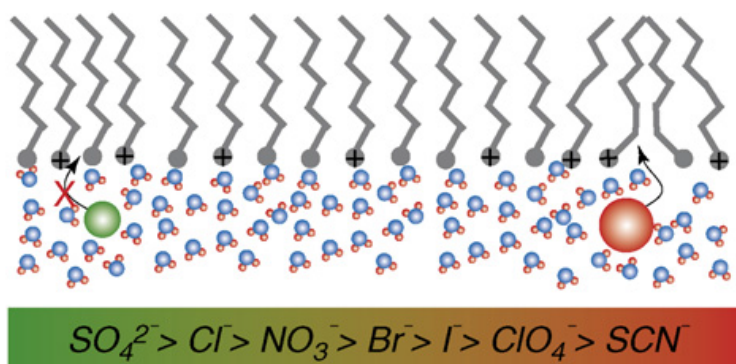


Figure D9. An octadecylamine monolayer at the air/salt solution interface. The arrows compare the penetration of kosmotropic (green) and chaotropic (red) anions into the headgroup region of the monolayer. The red cross indicates that the kosmotrope does not effectively penetrate into the headgroup region (Gurau, Lim et al. 2004).

The biased amino acid composition of halophilic proteins amplifies the ion impact on the stability and folding of halophilic proteins (Tadeo, López-Méndez et al. 2009a). That is, stabilizing ions according to Hofmeister series have a larger stabilizing effect on halophilic than in mesophilic proteins. Conversely, destabilizing salts also show a stronger destabilizing effect on halophilic proteins. Up to date, this connexion between Hofmeister and halophilicity has not been successfully rationalized. For instance, a model based in Poisson–Boltzmann equation for K^+ –carboxylate electrostatic interactions considers ions as point surfaces, therefore ion specificity of the Hofmeister effects is lost. This theory works rather well up to 0.1 M concentrations, but loses its validity at relevant biological concentrations (Zhang and Cremer 2006). Additionally, it would not be able to explain anion stabilisation effects; which happen to be more significant than cation effects (Tadeo, Pons et al. 2007).

Halophilic amino acid selection amplifies Hofmeister effects, especially for anions.

As we have already discussed, salt stabilization of proteins is driven by mutual volume exclusion and weak unspecific interactions between protein and cosolute ions (Baldwin 1996). The former is always stabilizing because it favours the smallest folded conformation, whilst the latter is usually destabilizing due to the higher number of interactions with the unfolded state (Schellman 2003). In the work presented here, we demonstrate for the first time that the **modulation of ion–protein interactions by means of amino acid selection** plays a central role in salt stabilization and halophilic adaptation. This is a very suggestive link to the ultimate determinants of the properties of Hofmeister series. We propose that *the same model of halophilic adaptation postulated in the previous sections can shed some light into the variety of effects of different salts on the stability of proteins, and particularly halophilic proteins.*

As analogous to halophilic *versus* mesophilic proteins, the *ion-specificity of the stabilizing effects of Hofmeister salts corresponds to a larger exclusion from the protein surface*, which is more pronounced for stabilizing anions, e.g. SO_4^{2-} and Cl^- . As a result, the interaction component – usually destabilizing – is decreased and the folded conformation becomes stabilized (Xie and Timasheff 1997, Tadeo, Pons et al. 2007). In turn, destabilizing ions, e.g. SCN^- and ClO_4^- , are preferentially in contact with the protein surface; therefore the conformations with larger exposed area (unfolded state) are favoured due to the increased number of interactions. Halophilic amino acid composition amplifies this performance by reducing the hydrophobic content and increasing the fraction of polar area. The resulting *increase in the number of sites available for preferential hydration exaggerates the competition for preferential interaction/exclusion from the protein surface* and, as a consequence, *amplifies the energetic contribution of interactions*. Stabilizing anions with lesser surface binding capacity than water (Cl^-) are further excluded; whereas destabilizing anions, which tend to displace water from protein surface (Jungwirth and Tobias 2006, Nihonyanagi, Yamaguchi et al. 2014), find more sites to do so. Hence, the halophilic character of proteins could be interpreted as a parallel classification complementary to the Hofmeister ion series. In support of this idea, the negative sign of the chemical shift perturbation coefficients points to the *anions as the main specie interacting with the molecular surface* (except for the K^+ -carboxylate interactions, of course), consistent with the higher influence of anions in the Hofmeister effects. CLEANEX results also show a decrease of water molecules at the protein surface upon KCl addition. Presumably, Cl^- is competing with water molecules (Zhang, Furyk et al. 2005) progressively displacing them from the regions near backbone amides and slowing down the exchange rate. The **reduced hydrophobic content and the increased polar surface in halophilic sequence haloSRC increases the number of sites for preferential hydration, favouring water over Cl^- anions at the protein-solution interface** and accelerating the exchange with respect to mesophilic uSRC.

Anions (more than cations) compete with water for protein surface binding.

Halophilic proteins show more hydration / ion-binding sites, enhancing the competition and amplifying the effect of interactions.

Several reports outline the *predilection of anions for hydrophobic surfaces* (Gurau, Lim et al. 2004, Jungwirth and Tobias 2006), which is quite surprising for a charged particle. Computational studies have introduced an ionic dispersive potential to current interparticle interaction theories and successfully applied this model to examine specific ion effects (Bostrom, Williams et al. 2003, Zhang and Cremer 2006). In the context of halophilic adaptation, these anion-apolar dispersive interactions would also help to understand the ubiquitous reduction of hydrophobic content from a different perspective: anion competition for direct hydrophobic or ionic contacts with the protein surface would ultimately lead to protein dehydration at large salt concentrations, such as those found inside KCl-accumulating cells. The equilibrium of the ternary water-cosolute-protein system is complex. As a very naïve intuition, *Cl^- competes not only with the protein for water molecules, but also with water molecules for protein surface*. In order to maintain a **functional solvation shell** around the protein, the hydrophobic surface is reduced (Lys reduction, conservative Leu and Ile to Val mutations) or substituted by polar side chains (Thr or Asp). This mechanism would successfully explain the increase in salt stabilization achieved by methylene removing mutations such as Glu to Asp; or even Glu to Asn, which additionally remove negative charge (Tadeo, López-Méndez et al. 2009b). Moreover, it adds another argument against lysines and positively charged residues in halophilic proteins, which would contribute to attract negatively charged anions to the protein surface.

Regarding cations, a relevant question remains unanswered: *why K^+ is preferred instead of Na^+* ? If the contributions from cations were purely electrostatic, any small positively charged cation would do the job. However, K^+ is strongly favoured, even at the expense of selectively accumulating large intracellular K^+ concentrations despite of Na^+ being the major cation in the extracellular media. Most of the data available for Na^+ and K^+ in solution and their interaction with other macromolecules comes from computational studies. As opposed to anions, ion-specific effects for cations at hydrophobic water–air interface are negligible, consistent with the present picture indicating that cation specificity originates from local interactions with charged and polar groups: ion–ion interactions with negatively charged side chains and ion–dipole interactions with backbone amide carbonyls (Jungwirth and Tobias 2006). However, unlike anions, the interactions are not based on direct contact ion pairs. Instead, solvent–ion pairs dominate the thermodynamic properties of cation–carboxylate solutions (Nihonyanagi, Yamaguchi et al. 2014). *Na^+ binds water more strongly than K^+* . The better **solvation properties** (higher dehydration energetic barrier) of Na^+ over K^+ makes the former to develop better solvent–ion pairs with carboxylates, as shown for formate, acetate and carboxylic groups in protein's surface (Hess and van der Vegt 2009). Experimental as well as theoretical studies have addressed the pairing of Na^+ and K^+ with anionic carboxylate groups in simple electrolyte solutions but also with charged carboxylic groups (aspartate and glutamate) on protein surfaces (Aziz, Ottosson et al. 2008, Uejio, Schwartz et al. 2008). Specifically, Pavel Jungwirth's lab (Vrbka, Vondrasek et al. 2006) performed molecular dynamics and conductivity measurements on various proteins. They reported that *Na^+ binds at least twice as strong as K^+ to protein surface*, with aspartate and glutamate side chains playing the most prominent role. However, unlike K^+ , considerable Na^+ binding was also found to occur with the backbone atoms. Substantial differential binding of Na^+ versus K^+ to the **backbone carbonyls** of short model peptides was also independently reported by other researchers (Dzubiella 2009).

The consistent observation is that *Na^+ pairs stronger with carboxylate groups than K^+ does*. Therefore, Na^+ would seem a more suitable option for salt-stabilization of proteins through ion–carboxylate interactions. However, the stronger water binding and solvation of Na^+ makes it an *unfriendly travel companion for proteins in media with low water activity*. In a scenario in which both molecules need to compete for water molecules, the weaker hydration properties of K^+ would be more helpful for the protein to get properly solvated. Moreover, the propensity of Na^+ to form interactions with carbonyl backbone energetically favors unfolded conformations, thus destabilizing the folded conformation. In summary, *K^+ exhibits a sufficiently strong hydration shell as to form solvent ion pairs with surface carboxylates but, unlike Na^+ , weak enough as not to be able to form solvent mediated destabilizing interactions with backbone carbonyls and not to compete with the protein for a proper solvation*.

K^+ hydration properties make this ion more suitable than Na^+ in environments with low water activity: high salt concentrations.

A preliminary understanding of the influence of ions on the physical properties of macromolecules is merely a first step towards understanding the effects of solution conditions on biological systems. We provide a mechanistic link between Hofmeister effects and protein properties in the novel frame of ion–macromolecule interactions, and extend the range of tools available by proposing halophilic character as a modulation of Hofmeister effects on the macromolecule side. Much work still remains; a complete picture will inevitably involve an integrated understanding of the role of cations, anions and osmolytes as well.

5.6. MECHANISTIC COMPARISON BETWEEN HALOADAPTATION & OSMOPROTECTION

Polyhydric compounds and sugars are among the most prevalent molecules used by nature to protect organisms against the stresses of extreme conditions, including high osmotic pressure (for a detailed description of osmolytes and compatible solutes, see Introduction 1.2.1) (Yancey, Clark et al. 1982, Yancey 2005). The same compounds have often been found to be effective **stabilizers** of the native conformations of globular proteins and biological assemblies when added at high concentrations (~ 1 M) (Arakawa and Timasheff 1982). Studies aimed at understanding the mechanism by which these compounds stabilize globular proteins have found that all of these solvent additives are **preferentially excluded** from the surface of the protein; which in turn means that the proteins are preferentially hydrated in their presence. This is true of sugars (sucrose, glucose, lactose, trehalose), polyhydric alcohols (glycerol, sorbitol) and many others compounds (TMAO, sarcosine) (Lee and Timasheff 1981, Timasheff 2002).

Same as KCl, there is neither evidence nor need to invoke any specific conformational effect or other reactions induced in the protein by osmolytes (Pais, Lamosa et al. 2009). Instead, stabilization of proteins by most of these organic cosolutes can be fully accounted by *unspecific general effects: excluded volume contribution and weak nonspecific interactions* at the protein surface (Schellman 2003). The excluded volume contribution can be reasonably well estimated by means of physical descriptors (see Results 4.5), whereas the mechanistic determinants for protein–cosolute interactions are less clear in spite of extensive thermodynamic characterization (Schellman 1990). *Generally, an important fraction of the stabilizing action is driven by a stronger exclusion from the unfolded protein than from the native structure* (Xie and Timasheff 1997). The degree of preferential exclusion is proportional to (Kendrick, Chang et al. 1997):

- i. The protein surface **area**; i.e. the unfolded state has a larger exposed area and therefore a higher degree of preferential exclusion.
- ii. The number of **polar hydrophilic sites** available for preferential hydration, which are more numerous in unfolded states where the backbone atoms are exposed.

Timasheff and Lee established that the preferential exclusion of organic cosolutes from the protein surface increases the protein free energy (Lee and Timasheff 1981). Hence, by LeChatelier Principle, the system will minimize the thermodynamically unfavourable effect of preferential exclusion by favouring the protein state with the lowest i) surface and ii) hydration sites exposed: the native conformation.

Particularly, **sucrose** has recurrently been considered to be a paradigm of a stabilizing osmolyte due to the preferential exclusion of this sugar from the protein surface (Kendrick, Chang et al. 1997, Xie and Timasheff 1997). Since this mechanism shows considerable similarities to the one proposed in this work for KCl in halophilic adaptation, we have tested the effect of sucrose on halophilic proteins. To that end, we have measured the melting temperatures of different halophilic and non–halophilic related systems in the presence of increasing sucrose concentrations. Similarly to KCl, the increase in stability induced by sucrose is linear and can be reported as the slope $m_{SUCROSE}$ (Table D1). As we initially expected, **halophilic proteins are considerably more stabilized by sucrose than mesophilic proteins**. In most cases, sucrose is actually a better stabilizer than KCl.

Protein (MW / KDa)	χ^2	m_{KCl} kcal·mol ⁻¹ ·M ⁻¹	$m_{SUCROSE}$ kcal·mol ⁻¹ ·M ⁻¹ *
βLAC (29)	3.9	5.7 ± 0.5	12.6 ± 0.3
HvCopG (5)	4.9	6.5 ± 1.5	8.3 ± 1.2
Kx5E ProtL (7)	8.7	5.8 ± 0.9	7.8 ± 0.1
WT ProtL (7)	20.7	1.7 ± 0.7	4.3 ± 0.5
HvLigN_ade (77)	5.5	4.1 ± 0.5	4.3 ± 0.1
HvLigN_dea (77)	5.5	1.3 ± 0.5	3.5 ± 0.1

Table D1. KCl and sucrose stabilization. Global stabilization upon KCl addition m_{KCl} and the same quantity with sucrose $m_{SUCROSE}$ obtained from thermodynamic data of global thermal unfolding for various halophilic and non-halophilic proteins. HvCopG: translational repressor from halophilic archaea *Haloferax volcanii*.

* Strict thermodynamic theory works in molal units or partial volumes. However, molar units are used here as an approximation, discarding changes in activity coefficients or partial molar volumes. It is a common approach found in most of the literature consulted because the deviation is small compared to other factors and the trends become linear, simplifying the calculations (Collins and Washabaugh 1985, Schellman 2003).

Natural halophilic proteins β LAC and HvCopG show an outstanding increase in stability induced by sucrose. In addition, HvLigN_ade, which is more KCl stabilized than HvLigN_dea (as will be discussed in the next section), is also more stabilized by sucrose. It seems that *halophilic evolutionary process not only generates proteins adapted to high salt environments, but also proteins more efficiently stabilized by organic compatible solutes such as sucrose*. For instance, halophilic Kx5E mutant is more stabilized than the original mesophilic WT ProtL.

Thus, sucrose is capable of inducing a considerable stabilization of halophilic proteins. We have already shown how KCl stabilization of halophilic proteins arises from the screening of the electrostatic repulsion by KCl and the preferential ion exclusion in the unfolded state. Being sucrose a **neutral molecule**, it seems unlikely that it would form appreciable electrostatic interactions. Therefore, based on the previous literature about osmolyte stabilization mechanisms and the extrapolation of our model for haloadaptation, our hypothesis is that *halophilic adaptation enhances osmolyte-stabilizing mechanisms (i.e. preferential hydration and cosolute exclusion from the protein surface) that increase the unfolded state free energy and ultimately stabilize the folded state*.

Halophilic proteins are stabilized by sucrose, an uncharged osmolyte

In order to investigate further the interplay between these two mechanisms (electrostatics and preferential exclusion), we tested sucrose effect on the folding of Kx7E ProtL by integrating the signals in ¹⁵N,¹H-HSQC spectra (Figure D2). Kx7E ProtL is a designed obligate halophile. Due to the large number of surface negatively charged residues, the repulsive forces unfold the protein at salt concentrations below 1 M (Tadeo, López-Méndez et al. 2009b). *In this system, in which the electrostatic repulsion plays a clear dominant role, sucrose was found to be less efficient than KCl*. Still, there is a considerable stabilization induced by sucrose. Due to the increase in the viscosity of the solution, we could only take measurements at sucrose concentrations below 1 M, which were not enough to completely fold the protein. We then added 100 mM KCl in order to partially compensate the electrostatic forces. The effects of both cosolutes seem to be additive, probably indicating complementary unspecific mechanisms. The two sucrose titrations show a sigmoidal curve typical of a two state equilibrium, although the transition is much less cooperative than that with KCl alone, which is always much more efficient. Still, sucrose continues folding the protein at almost 70 % folded

fraction, which led us to estimate that roughly 100 % folded fraction could be achieved at sucrose concentrations near 2 M. Our interpretation is that the early collapse in the folding reaction is greatly hampered by the electrostatic repulsion. The lack of net charge of sucrose makes it very inefficient at screening those repulsive forces, yet the contributions from preferential exclusion and excluded volume progressively destabilize the unfolded conformations at increasing cosolute concentrations, ultimately favouring the native state. Thus, *the action of alternative mechanisms in the unfolded state partially overcomes the electrostatic penalty imposed by the acidic residues, making sucrose capable of folding our model strict halophile and generally stabilizing halophilic proteins.*

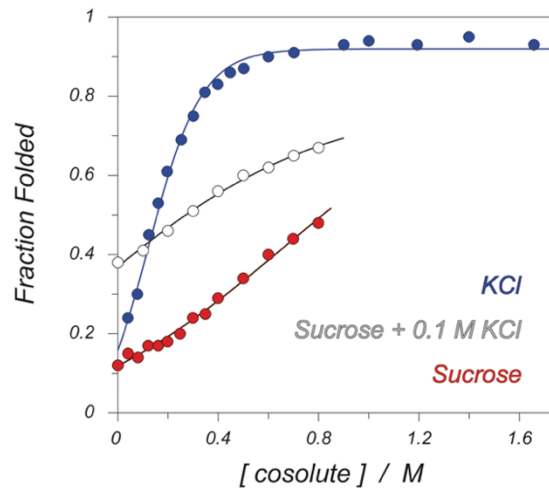


Figure D10. Folding titration of Kx7E ProtL. The relative populations of Kx7E ProtL folded and unfolded states as a function of sucrose and KCl concentration are shown as circles. Lines represent the least square fitting to a sigmoidal two-states equilibrium.

If preferential exclusion was playing a prominent role in the sucrose stabilization of halophilic proteins, its effect should reflect mainly in the unfolded state. We repeated the H/D exchange experiments in the presence of sucrose. For β LAC, the profile of sucrose stabilization (Figure D11a) mirrors that of KCl (Figure D11b): *core regions with buried residues show a larger stabilization* than exposed residues, whereas four most favoured amino acids in halophilic proteins (Asp, Thr, Glu and Val, shaded in grey) do not show a particularly impressive stabilization. As opposed to global data, here KCl stabilization is slightly larger than sucrose's, probably reflecting the additional electrostatic contribution. Again, only buried residues with large Δ ASA upon unfolding are sensitive to the effect of the cosolute, so this experiment mainly reports on the unfolded state destabilization. The fact that both profiles looked nearly identical, led us to conclude that *the same mechanisms destabilizing the unfolded state in the presence of KCl are activated by sucrose.*

The same is also observed for ProtL (Figure D12). In both WT (white) and Kx5E (red) mutant, residues less exposed to the solvent show a larger sensitivity to the cosolute in terms of stabilization. Similarly to KCl, *upon introduction of halophilic substitutions there is a general rise in sucrose stabilization throughout the whole sequence* (lower chart), *pointing to a general non-specific mechanism.* Accordingly, no differential stabilizing contributions are found at KxE positions (but for tyrosine-34, which could be due to a specific sucrose-binding site). A direct comparison between the individual KCl and sucrose m -values for both WT and Kx5E ProtL (Figure D13) shows a clear correlation between them: residues more stabilized by KCl are equally more stabilized by sucrose, both in halophilic and mesophilic sequences. Furthermore, the halophilic mutations introduced cause a general increase in the individual stabilization, similar for all residues, but maintain the correlation. In our opinion, this demonstrates that the **preferential exclusion stabilization mechanism is mainly the same for both sucrose and KCl, and is similarly modulated by halophilic composition.**

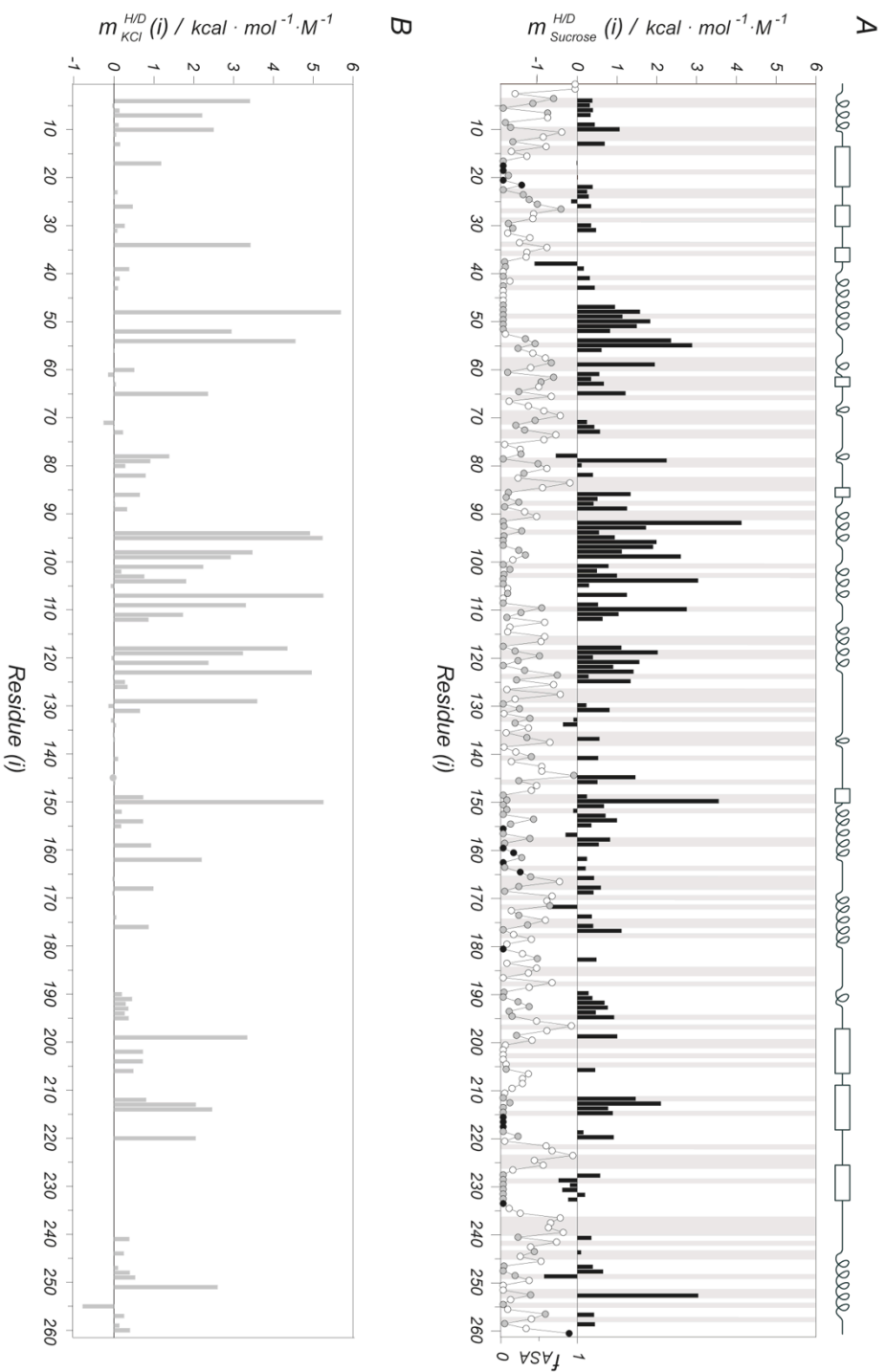


Figure D11. Sucrose stabilization of βLAC. A) Local sucrose stabilization of βLAC $m_{\text{Sucrose}}^{\text{H/D}}(i)$ calculated from H/D exchange rates represented as black bars. Fractional solvent accessible surfaces, f_{ASA} , are shown below (right axis) and colour coded according to the timescale of the exchange: measurable timescales (grey circles), too short timescales for exposed residues (< 30 min, white circles) and too long timescales for deep burried residues (> 1 y, black circles). Four most abundant amino acids in halophilic proteins (aspartate, threonine, glutamate and valine) are shaded in grey. No preferential stabilization is found for these selected amino acids. Remarkably, the overall profile resembles that of KCl stabilization (B), with burried regions showing a higher stabilization.

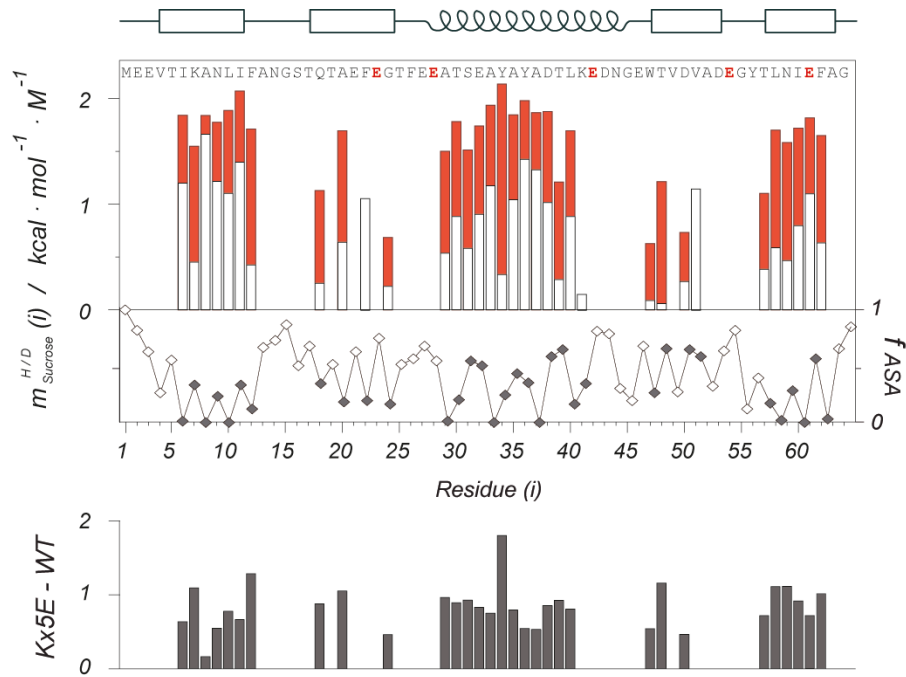


Figure D12. Sucrose stabilization of ProtL. Upper bar-plot: *per residue* stabilization upon sucrose addition calculated from H/D exchange experiments for natural mesophilic WT (white) and designed halophilic Kx5E (red) ProtL. Lower bar-plot: differences between the halophilic mutant Kx5E and the non-halophilic WT ProtL. Stabilization is larger for sucrose as compared to KCl, although some similarities are observed: first, stabilization occurs preferentially for buried residues, which only “feel” the osmolyte in unfolded conformations; second, mutations cause a general increase in the *per residue* stabilizing effect upon osmolyte addition throughout the whole protein, not only in the new acidic residues (same as for KCl, but less surprising for a neutral osmolytes).

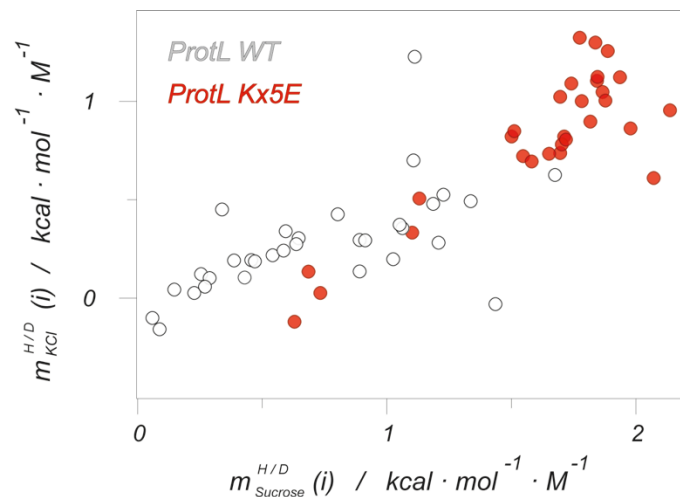


Figure D13. Comparison of sucrose and KCl stabilization. Plot representation of the *per residue* cosolute stabilization from H/D exchange experiments for KCl and sucrose. Each point corresponds to a residue of WT (white) or halophilic mutant Kx5E (red) ProtL. Coordinates mark the individual *m*-values for sucrose (abscissa axis) and KCl (ordinate axis). Residues exhibit a similar behaviour with both cosolutes, although *sucrose is always a better stabilizer than KCl*. Remarkably, *not only introduction of halophilic residues not only increases the KCl stabilization for all residues, but it also provokes a greater increase in sucrose stabilization*.

Examples of mechanistic studies on sucrose and other osmolytes stabilizing properties can be found in the literature. For example, the effect of sucrose on interleukin 1 receptor agonist (rhIL-1ra) was tested by Carpenter, Arakawa and co-workers (Kendrick, Chang et al. 1997). The results from this protein-solvent interaction study indicate that sucrose is **preferentially excluded** from the protein domain, increasing the free energy of the system. Thermodynamically, this leads to protein stabilization since *the unfolded state of the protein becomes thermodynamically even less favorable in the presence of sucrose*. Preferential exclusion/hydration involves water-cosolute exchange at thermodynamically non-neutral sites. The exclusion of sucrose from the protein domain seems to be related to the higher cohesive force of the sucrose-water and protein-water solvent systems as compared to the sucrose-protein system (Lee and Timasheff 1981). In addition, previous H/D exchange experiments with sucrose were carried out on Ribonuclease A (Wang, Robertson et al. 1995). A larger effect on buried residues experiencing a larger Δ ASA was also reported. Their interpretation was that the cosolute was mainly affecting residues exchanging through the unfolded state ensemble.

Similar conclusions arise from the study of other cosolutes. For instance, detailed work showed that the stabilizing action of **sorbitol** on Ribonuclease A is driven by stronger exclusion from the unfolded protein than from the native structure (Xie and Timasheff 1997). Xie and Timasheff carried out a careful thermodynamic analysis and dialysis equilibrium measurements of the preferential binding of sorbitol to native and denatured RNase A. The results of all these measurements returned the *preferential binding parameter* of sorbitol to the surface of the protein. They showed that sorbitol was preferentially excluded from the denatured conformations and, to a lesser extent, from the native state (where, on the contrary, at large concentrations some molecules could bind very weakly to folded RNase A). In other words, *the denatured protein was preferentially hydrated, whereas the native protein could bind the osmolyte*. The stabilization of RNase A by sorbitol was ultimately determined by the change in preferential interactions during the protein unfolding reaction and along the osmolyte titration coordinate. The amount of sorbitol present within the solvation layer of the protein increased with increasing co-solvent concentration; which implies that preferential hydration decreased with sorbitol concentration. Here lies the molecular description of the source of the stabilization. *At all co-solvent concentrations, contact of sorbitol with the denatured state was thermodynamically more unfavourable than with the native state*.

Sucrose – as well as other stabilizing osmolytes – is preferentially excluded from the protein surface.

Why was sorbitol (and equivalent osmolytes) preferentially excluded? RNase has a high ratio of polar to nonpolar residues and a very low hydrophobicity. With such a high polarity, RNase A exhibits a very **hydrophilic surface**. Similarly, sorbitol molecules, being hydrophilic co-solvents, should have an affinity for the polar residues on the protein surface. However, because protein interactions with sorbitol would be weaker than with water molecules, the exchange reaction would favour preferential hydration. The remaining fraction of the protein surface is nonpolar and is occupied by atoms that have no ability to form hydrogen bonds. Sorbitol molecules should be excluded preferentially from these *loci* to form water-osmolyte interactions. On denaturation, the exposed protein surface increases. Non-polar residues buried in the interior of a globular protein become exposed, as well as hydrophilic sites previously involved in intramolecular peptide-peptide hydrogen bonds. The resultant increase in the nonpolar residues exposed to solvent and the hydration of peptide bonds and polar

area is translated into a *greater cosolute exclusion (or as equivalent, a greater preferential hydration) from the denatured protein* relative to the native protein. As already explained, preferential exclusion is especially unfavourable at high cosolute concentrations and the entropic penalty destabilizes the unfolded state. *The relative scarcity of apolar surface and the abundance of polar residues in halophilic proteins would presumably make them efficiently stabilized by sorbitol as well.*

All these observations are also consistent with other studies showing that in sugars (sucrose, lactose and glucose) and polyhydric alcohol–water mixtures (glycerol, ethylene glycol, xylitol, mannitol, sorbitol and inositol) all the proteins examined were preferentially hydrated (Arakawa and Timasheff 1982, Kendrick, Chang et al. 1997, Xie and Timasheff 1997). *Henceforth, stabilization is caused by general nonspecific interactions. The effect is of a strictly nonspecific thermodynamic nature and can be accounted for by the weak interactions of the protein with the water–sorbitol system, expressed as preferential hydration (exclusion).* Accordingly, it was also reported that the change in the ionization of carboxyl groups had no effect on the preferential exclusion of sorbitol from the protein surface, as expected for a neutral molecule (Xie and Timasheff 1997).

Halophilic increase in polar surface and reduction of hydrophobic area favour cosolute exclusion and enhance ion and osmolyte stabilization.

The study of osmolyte modulation of protein stability has attracted a lot of attention from the scientific community. C. Tanford, John A. Schellman, Wayne Bolen, Robert Baldwin, Thomas Record, George Rose and Gary Pielak are just a few of the names involved in the field. The issue has been addressed from two main types of analyses: excluded volume and intermolecular interactions (Patel, Nobel et al. 2002). Several other types of solvent interactions affecting protein stability in osmolyte solutions are also important and are still being studied; e.g. compaction and flexibility or protein rigidification upon cosolute addition (Wang, Robertson et al. 1995, Doan-Nguyen and Loria 2007, Pais, Lamosa et al. 2009). For instance Gary Pielak has performed extensive work on the effect of water structure (Batchelor, Olteanu et al. 2004) or crowding (Davis-Searles, Saunders et al. 2001), particularly focusing on intracellular conditions and *in cell* NMR (Schlesinger, Wang et al. 2011). It is out of the scope of this thesis to describe all the literature about osmolytes. As this work is framed within the context of haloadaptation, only protein–cosolute interactions will be discussed in depth.

The solution thermodynamics of protein/osmolyte mixtures have been well characterized: **protecting osmolytes** *interact unfavourably with the unfolded state*, resulting in preferential depletion and increase of unfolded state free energy; whereas **denaturing osmolytes** *interact favourably with the unfolded state*, resulting in accumulation of osmolyte proximate to the protein surface and enthalpic stabilization of the unfolded state. However, *thermodynamics is a descriptive science, deliberately devoid of mechanistic information.* As yet, there is no universal molecular theory that can explain the mechanism by which osmolytes interact with the protein to affect its stability.

Groundworks for such a theory rely on influencing work from Tanford (Tanford 1970). He established the *Molecular Transfer Model*, an approach based on the proposition that, thermodynamically, whether a co–solvent acts as a stabilizer of the native structure or not, is

determined by the **transfer free energies** from water to the solvent system, Δg_{tr} , of the protein in the folded and unfolded states. The transfer of a native or denatured protein from water to osmolyte solutions could be quantified simply by summing the group transfer free energies of the constituent solvent–exposed parts: backbone and side chain surface. Wayne Bolen and co-workers calculated the contributions from backbone and side chains for the folded state and unfolded states, and managed to accurately predict the osmolyte–dependent energetics of proteins; i.e. the m –values (Auton and Bolen 2005). They found that the **peptide backbone contribution** was the largest and more dominant contribution, exceeding by far that of side chains. Although side chains play a role, *it is primarily the backbone Δg_{tr} that determines the extent to which osmolytes either stabilize or destabilize a protein.* Our findings, nonetheless, show how *side chain mutations can provoke important changes in the m –values of organic and ionic cosolutes.* We don't doubt the backbone is the main determinant to explain the stabilizing or destabilizing effect of a given cosolute. In fact, in our hands halophilic mutations enhanced the effect of salts and osmolytes, but never altered the Hofmeister order or the stabilizing/destabilizing character of a given cosolute, which is most likely determined ultimately by the backbone composition. However, **the role of side chains might have been largely underestimated.** After all, the authors recognize their initial surprise for backbone making such a disproportionately large contribution to free energy since it just represents roughly a quarter of the total area newly exposed on unfolding (Auton and Bolen 2005).

Osmolyte–induced behaviour has been well characterized in thermodynamic terms, but the specific molecular interactions in protein–osmolyte aqueous solution that stabilize or destabilize the unfolded state of proteins remain elusive. An important clue came from a recent work (Street, Bolen et al. 2006) showing that the Δg_{tr} for protein transfer from water to osmolyte solution are negatively correlated with their fractional polar surface area, $f_{polar\ ASA}^{osmolyte}$ (Figure D14). That is, the more fraction of polar area the cosolute has, the more favourable is the transfer from water to the osmolyte solution (the lower the Δg_{tr}). As a result, the denatured ensemble is relatively more favoured with respect to the native state, probably due to the increased number of osmolyte–protein interactions.

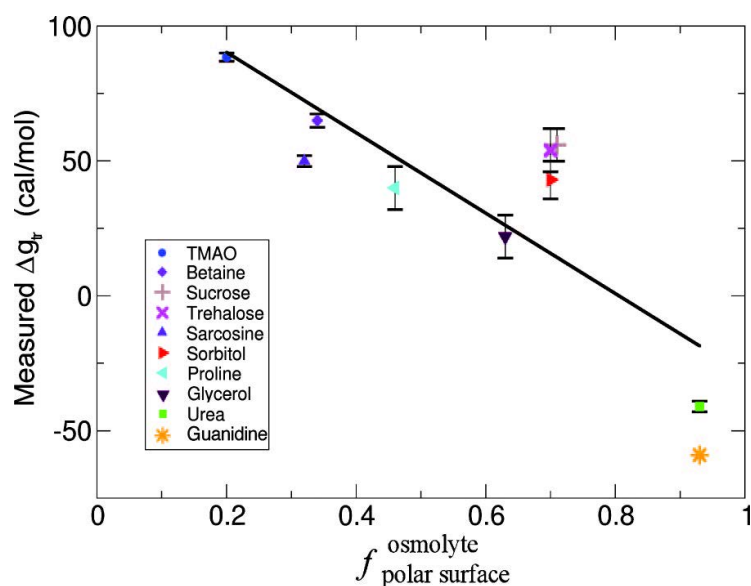


Figure D14. Polar fraction of osmolyte surface correlates with measured Δg_{tr} values. Fractional polar ASA, $f_{polar\ ASA}^{osmolyte}$, is plotted against Δg_{tr} values for the ten osmolytes indicated in the inset. Linear regression has a negative slope with a correlation coefficient of 0.81, indicating that backbone/osmolyte interactions become increasingly favourable as osmolytes become increasingly polar.

This correlation suggests that polar and nonpolar osmolyte surfaces interact with the protein at different energies, being the interaction between the protein backbone and osmolyte polar groups more favourable than the corresponding interaction with nonpolar groups. Such interpretation immediately suggests the existence of a universal mechanism involving osmolyte, protein and water. The analysis focuses on backbone interactions and backbone Δg_{tr} , obliterating the role of side chains. Still, they built a computational molecular model for osmolyte–backbone interactions as a function of the polarity of the surfaces involved as compared to water. In spite of the model simplicity, they accurately predicted backbone Δg_{tr} values for a series of osmolytes, with reasonably good agreement to the experimental data. Their model is based on weak protein/osmolyte interactions, which successfully explain the linear stabilization effects of osmolytes. In this weak–binding regime, free energy effects are ostensibly additive because neither type of osmolyte occupies a significant fraction of the backbone surface. Essentially, there is no competition for backbone binding sites.

This idea can be readily extrapolated to the effect of ionic cosolutes on halophilic proteins, where our data show the weak and unspecific nature of the ion–protein interactions. The similarities do not end here. The mechanism we propose for salt stabilization of halophilic proteins is based in protein–cosolute (ionic or organic) interactions, which can be modulated by halophilic signature. Mutations present in halophilic proteins increase the fraction of polar surface and reduce the nonpolar area, which reflects on a larger ion – or osmolyte – exclusion and preferential hydration. This is clear evidence of the role of side chains in modulating cosolute–protein interactions, including those with backbone atoms. The use of halophilic proteins could therefore provide a complementary scale of polarity on the protein side, constituting a powerful tool for detailed studies of protein–cosolute interactions and, especially, the role of side chains.

6. ENZYMATIC ACTIVITY

"Progress occurs when ideas have sex"

Matt Ridley

6. ENZYME ACTIVATION INDUCED BY SALTS AS A SUBROGATE OF PROTEIN HALOADAPTATION

So far, the accepted hypothesis is that **haloadaptation has occurred through optimization of protein stability** in high salt environments. However, biological function is very often determined at the thermodynamic protein stability level, but ultimately through features such as binding to other macromolecules, oligomerization and catalysis. It is important to understand how the selection of certain amino acids affects other aspects of protein function to complete the picture of halophilic adaptation. Particularly, we will discuss through this section the effect of high salt concentrations in the activity of selected halophilic enzymes.

Since the function must be preserved, so must be the overall protein fold (Paul, Bag et al. 2008). Moreover, key catalytic residues are often maintained, specially in the case of the largely penalized lysines (Fukuchi, Yoshimune et al. 2003). Despite all these features common to mesophilic proteins, enzymatic activity of halophilic proteins is heavily modulated by salt.

6.1. COMPLEX SALT MODULATION OF ACTIVITY

The monotonic stabilization of halophilic proteins by KCl and NaCl contrasts with the complex modulation of their catalytic function (as already discussed in section 1.3.3. of the Introduction). Whereas some proteins are activated by salt, others show a decrease in their catalytic activity at high salt concentrations. Alternatively, the maximal activity of some enzymes occurs at an optimum range of concentrations and, more strikingly, other enzymes show a heavy ion-dependent activity modulation. To name just a few representative examples:

- **Salt-activated enzymes:** β -galactosidase from *Haloferax alicantei* (Holmes, Scopes et al. 1997). Dihydrofolate reductase from *Haloferax volcanii* (Wright, Banks et al. 2002).
- **Enzymes with activity decreased by salt:** glucose dehydrogenase from *Haloferax mediterranei* (Esclapez, Pire et al. 2007). Catabolic ornithine transcarbamylase of *Halobium salinarum* (Madern, Ebel et al. 2000).
- **Enzymes showing a complex modulation:** the two activities of catalase-peroxidase from *Halobacterium halobium* (Brown-Peterson and Salin 1993) show different dependencies on NaCl addition.

3-hydroxy-3-methylglutaryl-coenzyme A reductase from *Haloferax volcanii* is activated by KCl and, in turn, is deactivated by NaCl (Madern, Ebel et al. 2000).

DNA ligase N from *Haloferax volcanii* (**Hv LigN**) shows a complex behavior: addition of KCl activates the enzyme, whereas it becomes inactive at all NaCl concentrations (Poidevin and MacNeill 2006).

Thus, the variable salt effects on the activity of halophilic enzymes have been demonstrated for many case studies and tested over different ionic species. However, a unified molecular mechanism still remains elusive. We have chosen *Hv LigN* as a case study to investigate the complex salt modulation of enzymatic activity. This halophilic enzyme has been widely studied and its mechanism of action is well characterized.

6.2. NAD⁺-DEPENDENT DNA LIGASE FROM *HALOFERAX VOLCANII* LIGN

DNA ligases play key roles in all forms of cellular life, e.g. joining of DNA molecules in DNA repair processes or joining of Okazaki fragments during DNA replication. Along with RNA ligases and messenger RNA capping enzymes, DNA ligases constitute the nucleotidyl transferase superfamily. These enzymes interact with a nucleotide cofactor to form a covalent enzyme–nucleotide monophosphate moiety. Particularly, DNA ligases can utilize either NAD⁺ (nicotinamide adenine dinucleotide) or ATP (adenosine triphosphate). These two families, differing in their cofactor specificity, are characteristic of the two major branches of evolution: the eubacteria and the eukarya/archaea. **NAD⁺-dependent** ligases are encoded predominantly by eubacteria, but also by certain eukaryotic viruses and some bacteriophages. In contrast, ATP-dependent DNA ligases are characteristic of eukaryotic and archaeal cells, but can also be encoded by certain eukaryotic viruses, bacteriophages and eubacteria (Tomkinson, Vijayakumar et al. 2006).

The mechanism of DNA ligation is similar for both types of enzymes and, to a lesser extent, also to the other nucleotidyl transferases. Substantial differences are only found in the first step (**adenylation**) by virtue of their distinct cofactor requirements. Both types of ligases form a covalent enzyme–adenylate intermediate between the catalytic lysine and the adenosine monophosphate moiety (*AMP*), releasing nicotinamide mononucleotide (*NMN*) in the case of NAD⁺-dependent ligases or pyrophosphate for ATP-dependent ligases. The controlled hydrolysis of NAD⁺ or ATP provides the energy required to form the new DNA bond (Tomkinson, Vijayakumar et al. 2006).

There are two distinct DNA ligases in extreme halophilic archaea *Haloferax volcanii*: NAD⁺-dependent ligase N (*Hv* LigN) and ATP-dependent ligase A. Although none of them is required for organism viability, simultaneous deletion of both enzymes is lethal, so they probably share an essential ligation function (Zhao, Gray et al. 2006). *Hv* LigN is a 699 amino acid protein (75.7 kDa) with a typical halophilic amino acid signature (149 acidic residues for just 11 lysines, sequence in the *Annex*). Phylogenetic analysis reveals that *Hv* LigN was probably acquired at some point of *Haloferax volcanii* evolution by lateral gene transfer from the eubacterial class δ -proteobacterium (Zhao, Gray et al. 2006). *It shares structure and sequence homology with other bacterial NAD⁺ dependent ligases*, typically displaying a toroidal form and a highly modular architecture. Five different domains conform *Hv* LigN (Figure A1) (Lee, Chang et al. 2000):

- **Adenylation domain** (aa 1 – 336) is the catalytic domain. It is divided in two distinct subdomains linked by a long flexible loop: domain **1A** (red, aa 1 – 96) and the larger subdomain **1B** (blue, aa 97 – 336).
- **OB** (gray, aa 337 – 418): Oligomer Binding fold domain. Typically, OB-fold domains exhibit a β -barrel structure. It is involved in DNA binding. Along with the adenylation domain, they form the minimal unit for functional bacterial DNA ligases.
- **Zinc finger** (yellow, aa 419 – 445): formed by four Cys residues that coordinate a Zn²⁺ ion. Also found in other DNA interacting proteins, it has been suggested to play a role in DNA binding and recognition of the nick in the duplex DNA.
- **HhH** (green, aa 446 – 575): a helix-hairpin-helix motif domain. It is also present in a number of DNA repair enzymes and implicated in non sequence-specific DNA binding.

- *BRCT* (purple, aa 576 – 699): a member of the BRCA1 C-terminus domain superfamily. It shows a high mobility and partial disorder. It has been suggested to perform critical functions in cell cycle control and act as a signal transducer. It may transmit signal from DNA damage sensors to other components of the DNA repairing machinery by means of protein-protein interactions.

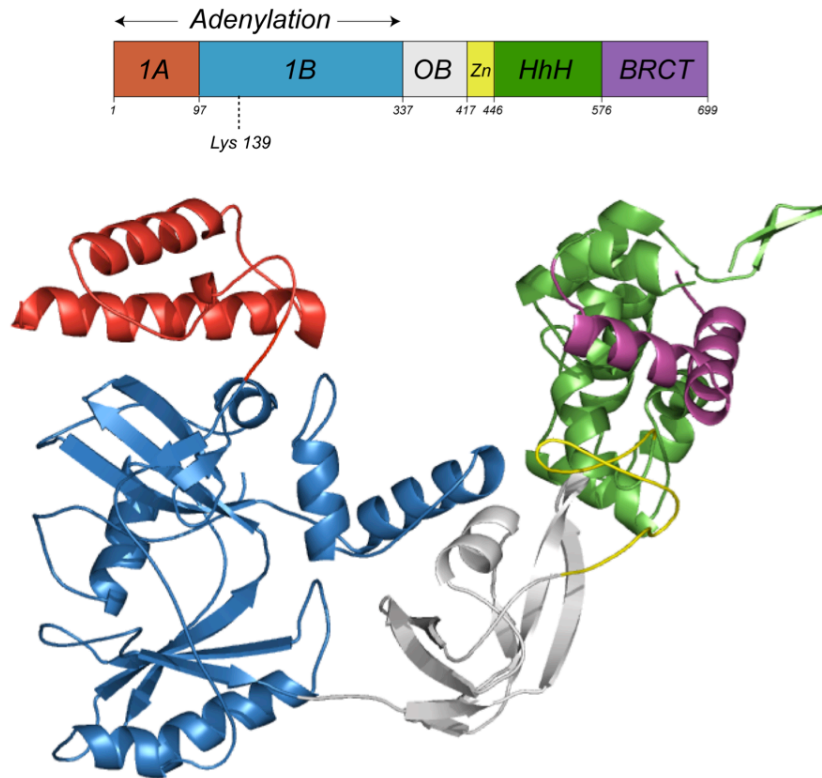


Figure A1. Structural homology model for the *Hv* LigN enzyme. The adenylation domain is divided in two subdomains: *1A* (red) and *1B* (blue). *OB* (gray), *zinc finger* (yellow), *HhH* (green) and part of the disordered *BRCT* (purple) domains are also shown.

Using NAD^+ as a cofactor, ligase N catalyzes the formation of a phosphodiester bond between the 5' and 3' ends of two DNA helices in a three-step reaction (Poidevin and MacNeill 2006):

- In a first step, NAD^+ binds to the adenylation domain and the **1A moiety closes over domain 1B**. Catalytic Lysine 139 acts as a nucleophile and attacks the phosphorous of the NAD^+ , resulting in covalent binding of the AMP half of the NAD^+ to the amino group of the catalytic residue and NMN release. *Subdomain closure and concomitant adenylation of the enzyme is the very first step in the catalytic cycle and is required for the reaction to proceed.*
- In the second step of the reaction, the AMP moiety is transferred from the enzyme to the 5'-phosphate end of one of the DNA strands.
- Finally, ligation between 5' and 3' ends of the DNA strands occurs coupled to the release of the AMP group. Hydrolysis of NAD^+ into NMN and AMP fulfills the energetic requirement for DNA bond formation (Figure A2).

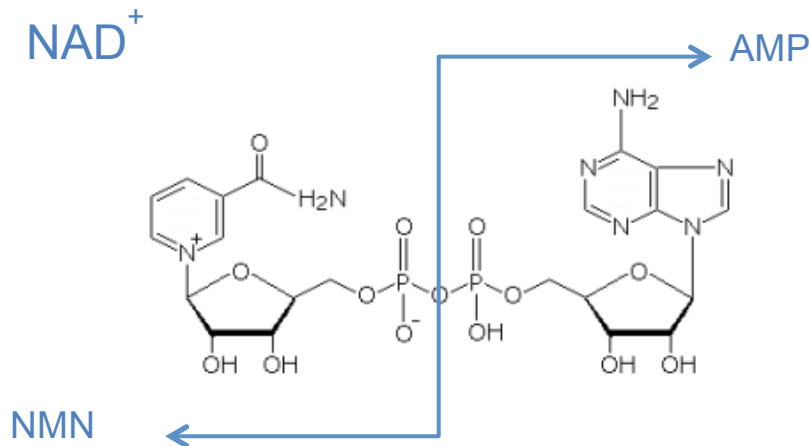


Figure A2. NAD^+ hydrolysis into NMN and AMP. NAD^+ is the cofactor for the ligation reaction. *Hv* LigN couples the hydrolyzation of the molecule into NMN and ATP with the formation of a new phosphodiester bond.

Whereas the last two steps are essentially the same for both eukaryotic and bacterial ligases, the adenylation step differs substantially for the two classes of enzymes by virtue of their distinct cofactor requirements, as previously explained.

Open – closure dynamics of the adenylation domain and subsequent *transition from open NMN-bound conformation to close NAD^+ -bound conformation* was shown by Gajiwala et al. for the NAD^+ -dependent DNA ligases from *Enterococcus faecalis*, *Bacillus stearothermophilus* and *Thermus filiformis* (Gajiwala and Pinko 2004). Crystal structures showed the movement of almost 180° and 80 \AA of 1A domain upon closure over 1B domain. In the same work, binding sites for NMN and AMP are described at near atomic resolution. 1A domain shows affinity for the NMN half of the NAD^+ molecule whilst 1B is the domain recognizing the AMP part. Formation of the binding site for the whole NAD^+ at domain 1B requires the domain 1A to accommodate the NMN half of the molecule; therefore, a conformational change is involved to alternatively host the NMN, AMP and NAD^+ molecules along the ligation reaction.

Authors propose that *overall reaction coordinate dictates the conformational switch between the open and closed forms* of the enzyme. The open conformation (Figure A2, left) reflects the state of the enzyme before adenylation, where the adenylation site is exposed and accessible to incoming NAD^+ . This conformation is analogous to that of the enzyme bound to NMN. In order to continue with the reaction, the enzyme must take up the closed conformation to constitute the NAD^+ binding site and catalyze the self-adenylation step. Post-adenylation conformational rearrangement would make the adenylated lysine accessible to incoming DNA molecule, which undergoes next two steps towards ligation of the two DNA strands.

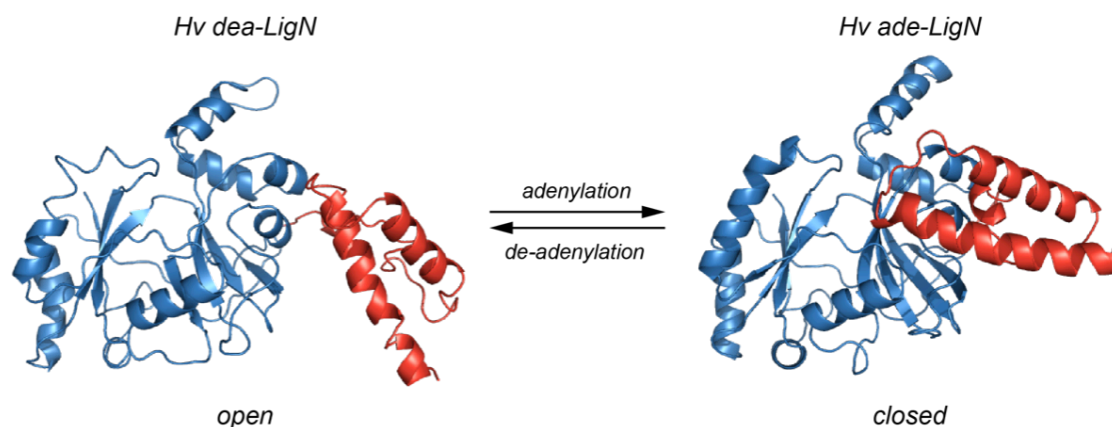


Figure A3. Open–closure dynamics and adenylation. Homology model with SWISS–PROT web server (Arnold, Bordoli et al. 2006) of the adenylation domain of *Hv* LigN based on the crystal structures from *Enterococcus faecalis* bacterial ligase (Gajiwala and Pinko 2004). The open and closed conformations along the reaction coordinate are represented here. Domain 1A is shown in red and domain 1B in blue. Both structures share the same orientation of domain 1B. NAD^+ binding causes the domain 1A to close, represented by the protein structure on the right. Deadenylation of the adenylation site with excess NMN leaves the adenylation site exposed to solvent, resulting in the open NMN-bound conformation shown here on the left.

Hv LigN is a paradigm for halophilic enzymes, and its activity and stability are differently modulated by KCl as reported in several studies (Poidevin and MacNeill 2006). Taking into account the lack of satisfactory models, our hypothesis is that both mechanisms are coupled, and that the key element to understand the activation of the enzyme by KCl is the *differential effect of salt on two distinct enzyme conformations with different exposed area, both part of the reaction pathway*: open NMN-bound (*Hv* dea–LigN) and closed NAD^+ -bound (*Hv* ade–LigN) conformations. In order to collect evidences supporting that putative interplay, we carried out extensive characterization of the enzymatic activity of *Hv*LigN and the stabilities of both isoforms in the presence of different concentrations of various cosolutes. Hopefully, that would allow a deeper understanding of the salt modulation of enzymatic activity and the nature of protein–solvent–cosolute interactions.

Results from this work were published in Scientific Reports (2011) 1:6. “Halophilic enzyme activation induced by salts” by Gabriel Ortega, Ana Laín, Xavier Tadeo, Blanca López–Méndez, David Castaño and Óscar Millet.

*The enzymatic activity of *Hv* LigN shows a complex modulation by salt: it is only active in the presence of KCl, but NaCl inactivates it.*

The enzyme undergoes a conformational rearrangement along the reaction coordinate that buries solvent-exposed area.

6.3. MODULATION OF ACTIVITY BY KCl AND NaCl

Hv LigN selectively joins fragments I (8454 bp) and IV (5686 bp) of *Bst*EII-digested λ DNA to produce a new oligomer that runs in an agarose gel at a molecular weight equal to the sum of the two fragments I+IV (Figures M6 and A3). This reaction has been used to evaluate the relative enzymatic activity upon addition of increasing cosolute concentrations with respect to that at 3.2 M KCl (see Materials and Methods 3.1.5.2), which is the *in vivo* KCl concentration measured for *Haloferax volcanii*. Initially, the two most common ionic species in halophilic cells are tested: KCl and NaCl.

As previously reported (Poidevin and MacNeill 2006), ligase activity of *Hv* LigN is fully dependent on the KCl concentration: *the enzyme cannot ligate DNA in the absence of salt, it activates upon KCl addition over 0.75 M KCl and the activity increases steadily until 3.2 M*, the maximum concentration tested (Figure A4-a). Larger salt concentrations could not be reached because of solubility issues and the smirring on the gel lanes. Surprisingly, when potassium is substituted by sodium, the enzyme is mostly inactive, even at high NaCl molar concentrations (Figure A4-b).

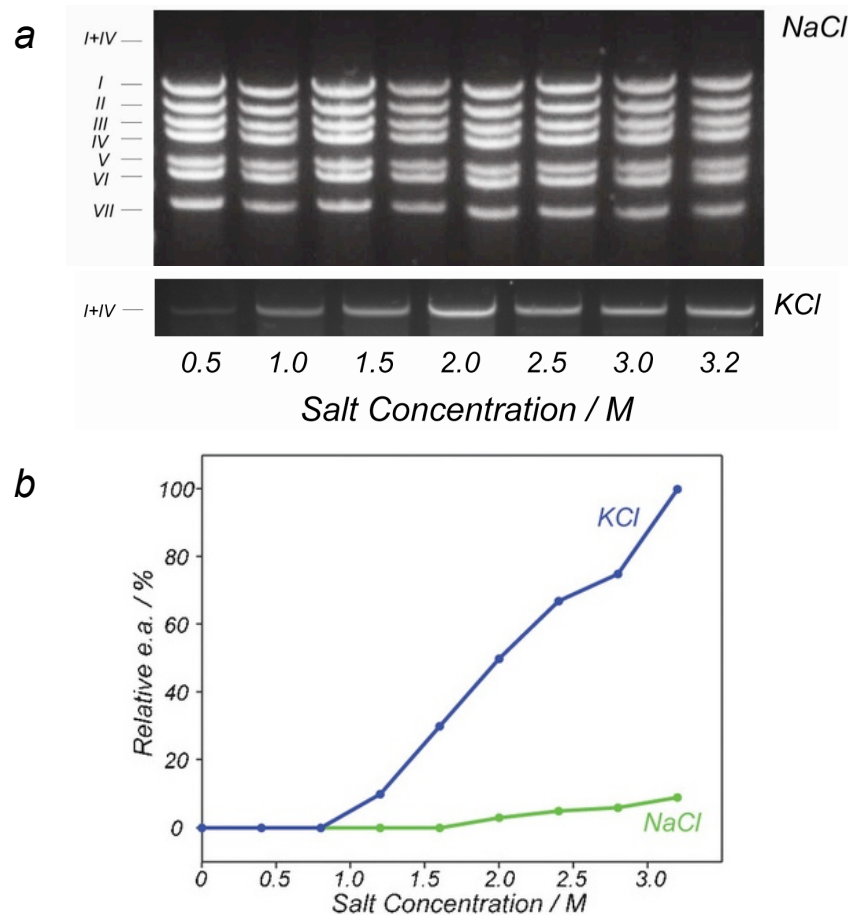


Figure A4. Ligase activity in the presence of KCl and NaCl. a) In the presence of salts, *Hv* LigN joins fragments I and IV to produce the adduct (band I+IV). Comparison of the I+IV band at different concentrations of NaCl (top) or KCl (bottom). All enzyme activities have been referred to the maximum activity measured in KCl (3.2 M). b) Relative enzyme activity as a function of the salt concentration and referenced to the catalytic activity in the presence of 3.2 M KCl.

How can the effect of two such similar salts be so dramatically different? How could the complex effect of KCl be explained? We will see how these questions are intimately related to the same processes that determine salt stabilization of halophilic proteins and haloadaptation.

6.4. SALT DEPENDENCE OF STABILITY AND COUPLING WITH ACTIVITY

We explored the putative interplay between enzyme activity and protein stability. It has been previously shown that the NAD^+ -dependent ligases undergo a conformational rearrangement upon adenylation, resulting in the closure of subdomains 1A and 1B. Interestingly, the adenylated protein has less exposed area, burying 331 \AA^2 to solvent as compared to *Hv* dea-LigN. Adenylation of LigN is required as the first step for DNA ligation to occur, so the two conformations correspond to two instances in the enzymatic reaction coordinate. Both isoforms can be isolated (see adenylation and deadenylation assays in Material and Methods 3.1.5.1) to obtain samples of fully deadenylated (inactive *Hv* dea-LigN) and fully adenylated (active *Hv* ade-LigN) ligase. We explored the stability of each conformation separately in the presence of various concentrations of NaCl and KCl.

Hv LigN unfolds reversibly in response to chemical agents. We performed chemical denaturations with guanidinium chloride (see Materials and Methods 3.3.1) and found that the unfolding was clearly not following a two-state model. Instead, experimental data fit well to a three-state model with one detectable intermediate, which was validated by means of a statistical F-test analysis (Figure A5; for a detailed description of the equations used in the analysis see Materials and Methods 3.3.1.2). This is not surprising given that *Hv* LigN is a multidomain protein of high molecular weight. Previous thermodynamic characterization of the NAD^+ -dependent ligase N from *Thermus scotoductus* also showed a complex folding landscape with detectable intermediates (Georlette, Blaise et al. 2004).

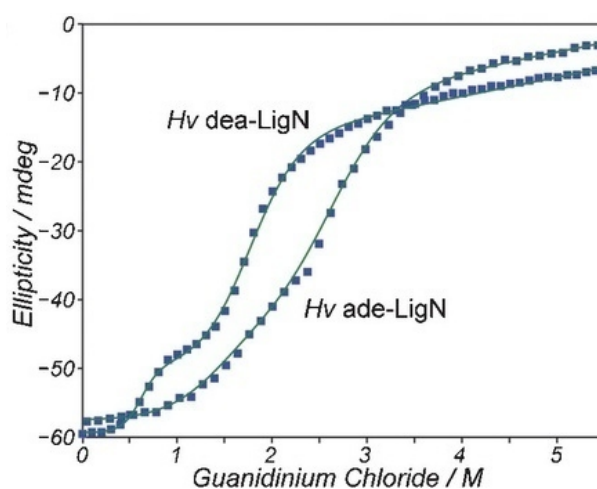


Figure A5. Chemical denaturation of LigN. Two representative examples of the chemical denaturation profiles monitored by CD for *Hv* ade-LigN and *Hv* dea-LigN. Blue squares mark the experimental ellipticity values measured at 222 nm and green lines correspond to the best fit to a three-states model.

Free energies of *Hv* ade-LigN and *Hv* dea-LigN were determined at increasing KCl concentration (Figure A6-a). At low KCl concentrations ($< 0.75 \text{ M}$), deadenylated protein is more stable than adenylated conformation, meaning that the latter is not significantly populated. As a result, the binding site for NAD^+ cannot be formed and adenylation, which is the first step of the ligation reaction, does not occur. This provides an explanation for the lack of activity at low salt concentrations. In contrast, non-halophilic ligases become stabilized upon adenylation, which helps the reaction to proceed (Georlette, Blaise et al. 2004). KCl

addition monotonically stabilizes both conformations showing a linear dependence with molar concentration. Surprisingly, the slope of the stabilization m_{KCl} is larger for *Hv* ade-LigN than for the deadenylated counterpart. The differences in free energy between both isoforms can be translated into *apparent* populations ($\Delta G = -RT \cdot \ln K_{eq}$) as a function of KCl concentration (Figure A6-b). We can see a very good agreement between the availability of the closed–active conformation and the values of enzymatic activity: at KCl concentrations below ~ 0.75 M there is not detectable ligase activity, consistent with *Hv* dea-LigN being the only populated conformation. At higher KCl concentrations, as *Hv* ade-LigN gets thermodynamically more stable, closed conformation becomes accessible and the enzyme starts the ligase activity. Therefore, our data suggest that the closure of subdomain 1A over 1B to form NAD^+ binding site and subsequent adenylation is the rate–limiting step at low salt concentrations. *The fact that the enzymatic activity continues to rise with KCl concentration supports the idea that the energetic barrier between the open–inactive and the closed–active conformations ultimately determines the kinetics of the reaction.*

The same thermodynamic characterization is reproduced in the presence of NaCl (Figure A6, lower plots). Interestingly, adenylylated and deadenylated *Hv* LigN are equally stabilized throughout the whole concentration range tested, as seen by the almost parallel lines in the Figure A6-c. As a result, *Hv* dea-LigN is the major dominant specie populated and the equilibrium is never displaced towards the closed conformation. To our knowledge, this was the first report for differential stabilization of a protein induced by KCl as compared to NaCl, even though computer simulations had predicted a minor stabilizing effect of Na^+ due to its higher affinity for the carbonyl groups of the protein (Vrbka, Vondrasek et al. 2006, Dzubiella 2009, Hess and van der Vegt 2009).

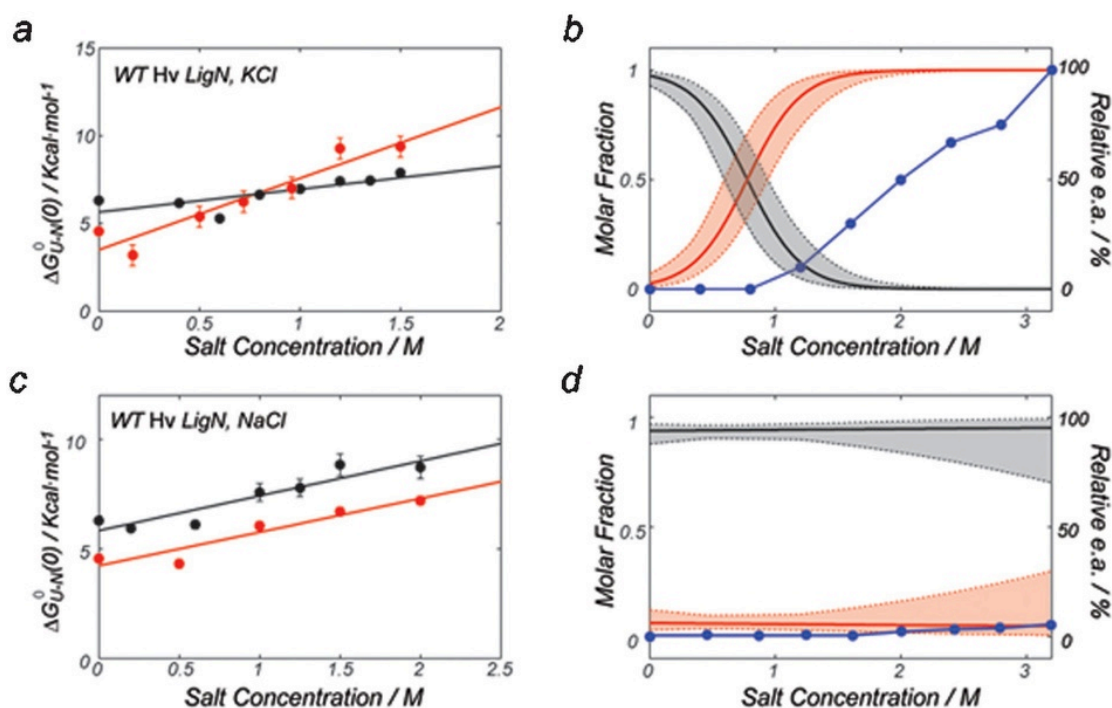


Figure A6. Coupling of stability with enzymatic activity. Plots of the unfolding free energy (ΔG_{U-F}^0) versus molar salt concentration in a and c panels for KCl and NaCl. *Hv* dea-LigN and *Hv* ade-LigN are represented as black and red, respectively. Circles mark the experimental values whereas solid lines correspond to the linear regressions. The apparent populations are plotted in the b and d panels (left ordinate axis) using the same color code. Shaded areas correspond to the estimated uncertainties in the populations calculated by a Montecarlo simulation. The blue circles show the experimental enzymatic activity normalized to that at 3.2 M KCl (right ordinate axis).

6.4.1. Stabilizing cosolutes select conformations with lower solvent exposed hydrophobic area

Why does KCl preferentially stabilize the *Hv* ade-LigN species? The effect of salt on protein stability is intimately ligated to the interactions between the protein surface and the solvent (Schellman 2003). Hydration of salt ions competes with protein solvation for the available water molecules. Therefore, the presence of ionic cosolutes enhances the hydrophobic effect, penalizing the exposure of hydrophobic surface. Previous works have shown that a reduction in the area accessible to solvent (ASA) results in higher stabilization induced by the salt (Pegram and Record 2007, Tadeo, López-Méndez et al. 2009a, Tadeo, López-Méndez et al. 2009b). Indeed, overall reduction in hydrophobic area is a universal adaptive feature of halophilic proteins. According to our structural model, *Hv* ade-LigN buries 331 Å² of its surface to the solvent, adopting a more closed conformation than the deadenylated specie. Thus, KCl preferentially stabilizes the more compact *Hv* ade-LigN, resulting in the activation of the enzyme. This result confirms the emerging picture in which halophilic signature enhances the effects of salt in proteins, altering the populations of the different states favoring those with lower exposed area: folded *versus* unfolded state, closed *versus* open conformations and oligomeric *versus* monomeric form.

This mechanism can also be extrapolated to explain the salt-modulation of catalytic activity of other halophilic enzymes. Malate dehydrogenase (*Hm* MalD) and lipase C (*Hm* LipC) from *Haloarcula marismortui* also require KCl to be active (Dym, Mevarech et al. 1995, Rao, Zhao et al. 2009). It has been reported that salt favours the formation of oligomers (tetramers in the case of *Hm* MalD and dimers for *Hm* LipC), the only forms catalytically active for these enzymes. In cases where the enzyme reaction coordinate implies a conformational change, differential stabilization of the species in the reaction pathway (as in *Hv* LigN) will also result in indirect modulation of the catalytic efficiency. The fact that this is an apparent effect is consistent with the very different modulations that KCl (or NaCl) can exert on the activity of halophilic enzymes.

6.4.2. Design of mutations reactivate the enzyme in the presence of NaCl

We tested our model by engineering changes in the ASA of *Hv* LigN and measuring its effect on salt induced stabilization. Our hypothesis is that mutations reducing the hydrophobic solvent-exposed surface would give the protein a more halophilic character, increasing the sensitivity of its stability to salt concentration and enhancing the differential stabilization of *Hv* ade-LigN and *Hv* dea-LigN. This would ultimately result in the recovery of the enzymatic activity in the presence of NaCl. Thus, we introduced two very conservative substitutions (E41D and E45D) in the 1A domain of *Hv* LigN. These mutations slightly decrease the hydrophobic character of the protein by removing two methylene groups without affecting the overall charge. Circular dichroism spectrum for the mutant was identical to wild-type *Hv* LigN, indicating that no structural perturbation occurred upon mutation. Enzyme activity for the mutant was tested as a function of NaCl and KCl concentration. In the

presence of NaCl, there was a clear recovery of catalytic activity upon mutation (Figures A7 and A8-b). A similar increase in the activity of the mutated enzyme was also observed when KCl was added instead, with the enzyme becoming catalytically active at a much lower concentration of this salt. Free energies of the *Hv* ade-LigN and *Hv* dea-LigN isoforms for the engineered protein showed that the two species were stabilized by NaCl with different slopes (Figure A8-a) progressively favouring the adenylated form required to trigger the reaction. Thus, very subtle modifications – removal of just two methylene groups from a protein of nearly 700 residues – in the protein surface can effectively tune the salt modulation of enzymatic activity.

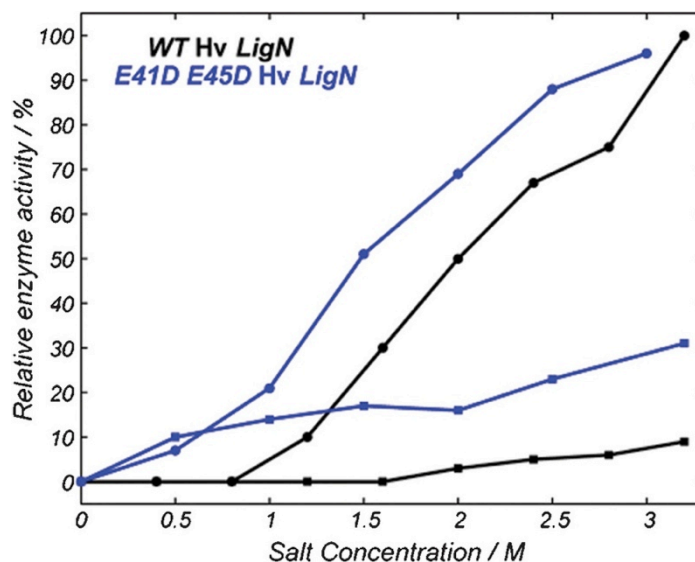


Figure A7. Ligase activity of E41D E45D LigN in the presence of KCl and NaCl. Dependence of the enzyme activity of the double ExD *Hv* LigN mutant (blue) with the molar salt concentration referenced to the wild type catalytic activity in the presence of 3.2 M KCl. Wild type activity is also shown (black) for comparison. Squares and circles represent the enzyme activity in the presence of NaCl and KCl respectively.

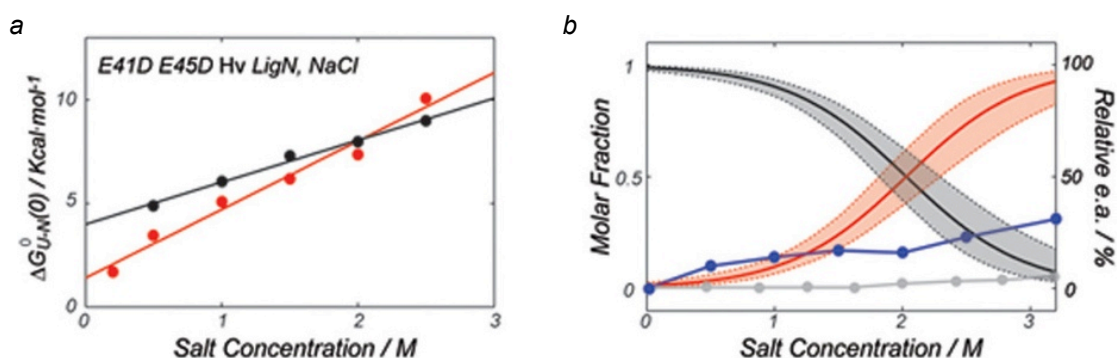


Figure A8. Mutations restore activity in the presence of NaCl. Unfolding free energy (ΔG_{U-F}^0) of E41D E45D *Hv* LigN versus NaCl concentration in a shows a differential stabilization of adenylated with respect to deadenylated form. As a result, closed conformation becomes more accessible and enzymatic activity is restored.

6.5. EFFECT OF OTHER HOFMEISTER IONS

To account for the effect of other salts in the modulation of enzymatic activity of wild type *Hv* LigN, ligation assays were performed in the presence of different salts. The effects of salt in proteins are usually more sensitive to the chemical nature of the anion (Tadeo, Pons et al. 2007). Therefore, acetate, sulphate and phosphate are selected as examples of stabilizing anions; according to the order established by Hofmeister (Hofmeister 1888). In turn, bromide, nitrate and thiocyanate are chosen as destabilizers. Chloride anion, already discussed, falls between the two categories and is often considered a weak stabilizer. Additionally, both sodium and potassium salts were tested in order to disentangle the role of the cation.

6.5.1. Stabilizing ions

Hv LigN gets activated in the presence of stabilizing anions. Results are shown in figure A9. In fact, sulphate, phosphate and acetate trigger more efficiently the enzymatic activity than KCl: the ligase activity starts at earlier salt concentrations and reaches higher values. In turn, no significant differences are found between sodium and potassium (except for chloride, of course).

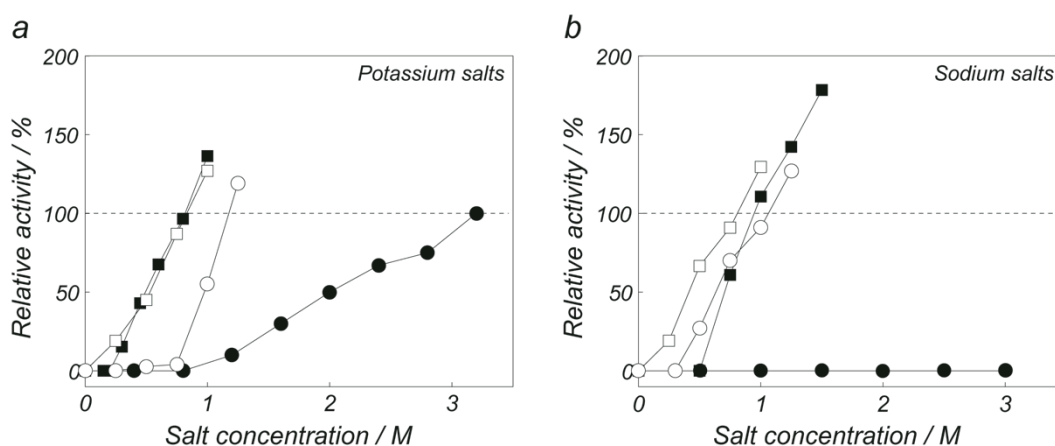


Figure A9. Hofmeister anions modulate catalysis. Catalytic activity of *Hv* LigN as a function of salt concentration normalized by the natural activity at 3.2 M KCl. Panels *a* and *b* relate to potassium and sodium salts respectively. Legend: phosphate (open squares), sulphate (black squares), acetate (open circles) and chloride (black circles). The dashed lines correspond to the maximum activity found at 3.2 M KCl.

We also explored the stability profile of wild type *Hv* *ade*-LigN and *Hv* *dea*-LigN in the presence of these salts. Sodium was chosen as counterion instead of potassium due to the higher solubility of guanidinium chloride in the presence of sodium salts. Results are shown in Figure A10 for sodium sulfate and sodium phosphate. In both cases the cosolute preferentially stabilizes *Hv* *ade*-LigN. There is also a good agreement between the relative population of the two species and the activity of the enzyme: when sodium sulfate is added, the lack of activity at concentrations below 0.5 M is explained by the higher stability of *Hv*

dea-LigN over this concentration range (Figure A10-a,b); whereas in the presence of sodium phosphate, ligation is observed even at low concentrations consistent with its free energy profile (Figure A10-c,d). Again, the salt dependent stability of both isoforms results in a complex modulation of catalytic activity. Moreover, the extent of the differential stabilization depends on the stabilizing character of the cosolute following the qualitative order in the Hofmeister series.

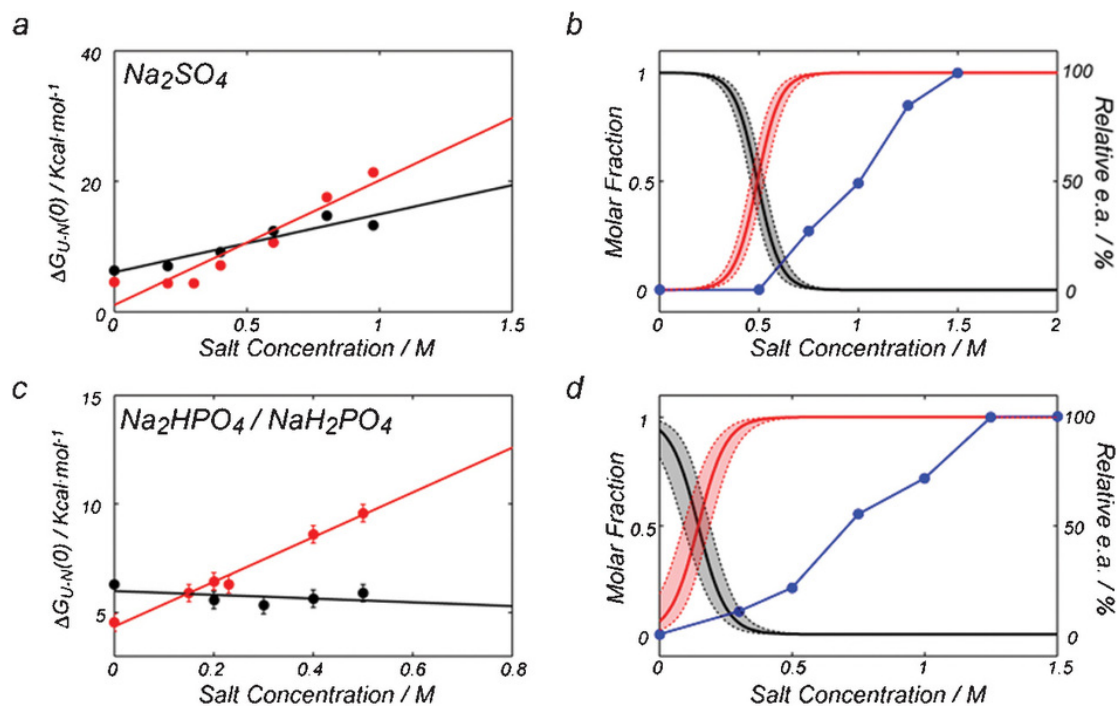


Figure A10. Coupling of enzymatic activity and stability for Hofmeister anions. Plots of the free energy upon unfolding (ΔG_{U-F}^0) for wild type *Hv* LigN versus the molar concentration of salt are shown in a and c for the salts indicated in the panels. *Hv* dea-LigN and *Hv* ade-LigN are represented by black and red circles, respectively. A linear dependence between stability and the molar concentration of salt was found and the solid lines correspond to the linear regressions. The populations of *Hv* dea-LigN and *Hv* ade-LigN are plotted in panels b and d using the same colour code (left ordinate axes). The shaded areas correspond to the uncertainties in the populations estimated by a Montecarlo propagation of the experimental standard deviations. The blue circles represent the enzyme activity normalized with respect to 3.2 M KCl (right ordinate axes).

6.5.2. Destabilizing ions

The enzyme is completely inactive in the presence of destabilizers. Similarly to NaCl, the band corresponding to the ligated *I+IV* fragment is below our detection limit (Figure A11). We pursued further characterization of destabilizing anions and repeated the enzymatic assays in the presence of activating buffer with 3.2 M KCl. Results in figure A12 show a decrease in enzymatic activity proportional to the concentration of salt. Again, good qualitative agreement between the effect on catalytic activity and the order in the Hofmeister series is found.

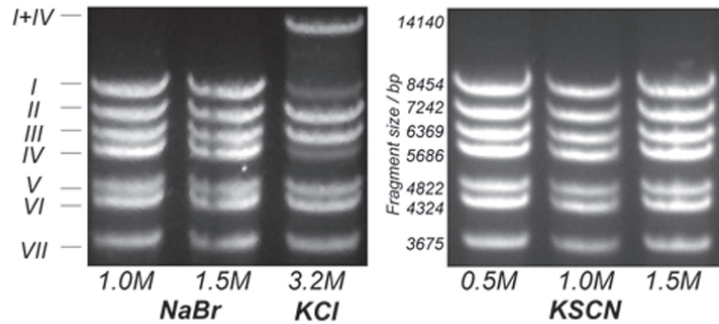


Figure A11. Absence of ligation with destabilizing anions. Enzymatic assays of wild type *Hv* LigN in the presence of the destabilizing anions bromide and thiocyanate, as compared to activity at 3.2 M KCl. No significant ligation activity is detected.

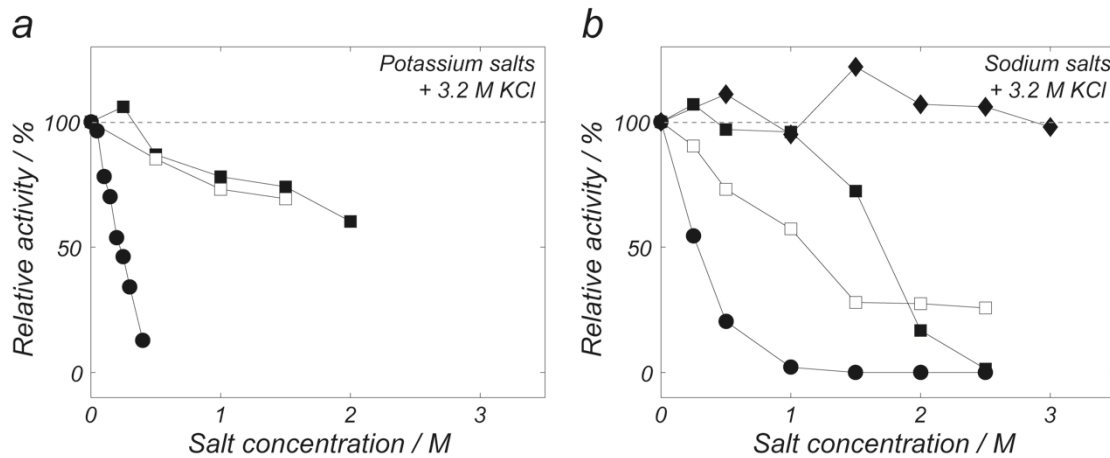


Figure A12. Enzyme inactivation by Hofmeister destabilizing anions. Catalytic activity of *Hv* LigN as a function of salt concentration normalized by the natural activity at 3.2 M KCl. Panels *a* and *b* relate to potassium and sodium salts respectively. Legend: thiocyanate (black circles), nitrate (open squares), bromide (black squares) and sodium chloride (black diamonds). The dashed lines correspond to the maximum activity found at 3.2 M KCl.

These results can be easily explained on the basis of previously exposed ideas. Salts, and mainly anions, linearly affect protein stability. Complementarily, the halophilic character of the protein (i.e. surface charge and ASA) amplifies the extent of such stabilizing/destabilizing effect. Closed *Hv ade*-LigN has lower exposed area, therefore being more sensitive to salt concentration. As a result, the stabilizing (or destabilizing) effect of salt is more pronounced. As opposed to stabilizing salts, destabilizing salts preferentially *destabilize* *Hv ade*-LigN with respect to *Hv dea*-LigN, blocking the initial adenylation step. Moreover, experiments in the presence of 3.2 M KCl demonstrate a certain additivity of the effects of different solutes in enzymatic activity, reflecting the additive contribution of cosolutes to protein stability (Schellman 1990, Tadeo, Pons et al. 2007). This ultimately reinforces a model based on weak, unspecific interactions.

7. CONCLUDING REMARKS

*"Our paths lead us from the tree of knowledge
to the control of our destiny"*

Jean Marie Lehn

7. CONCLUDING REMARKS

After years of research, our lab unveiled the role of solvent exposed hydrophobic area – and not negative charge – as the key determinant for salt stabilization of proteins. This new finding led us to seek a novel mechanistic model based on general unspecific contributions:

- ✓ By using **NMR** in combination with other biophysical techniques, we have addressed halophilic adaptation at the level of **individual residues**.
- ✓ We have performed an exhaustive characterization of the **electrostatic contribution** of acidic residues to the salt stabilization of halophilic proteins, as well as a detailed analysis of the effect of ions on each residue of halophilic and non-halophilic sequences.
- ✓ Additionally, we have carried out an unprecedented study of the halophilic **unfolded state**, using intrinsically disordered proteins and rational design as an approximation.

Our results demonstrate that:

- ✓ The nature of electrostatic **hydrated cation–carboxylate interactions** is not consistent with a tight and specific binding of heavily hydrated potassium cations to acidic clusters.
- ✓ We quantify the extent of such **electrostatic interactions** and show that their energetic contribution is **not enough** to completely explain haloadaptation or salt stabilization.
- ✓ Accordingly, characterization of the **hydration shell** of halophilic proteins by ^{17}O MRD showed no significant differences in the hydration dynamics of halophilic and mesophilic proteins. No water molecules bound with residence time longer than 1 ns were found.
- ✓ In turn, we reveal new important contributions arising from a preferential **ion-exclusion** ultimately **destabilizing the unfolded state** of halophilic proteins.

The finding of a preferential destabilization of the unfolded state of halophilic proteins due to preferential ion exclusion is a novel mechanism. Our results are consistent with a model in which:

- ✓ **Halophilic amino acid** set selects residues that **favour ion exclusion**.
- ✓ The reduction in **hydrophobic area** (lysines and residues with bulky hydrophobic side chains) together with the increase in **polar surface** (small polar residues such as aspartates, threonines and glutamates) makes more sites available for the competition between water and ions for the proteins surface.
- ✓ As a result, the effect of salts in halophilic proteins is increased. In particular, **KCl exclusion** and **preferential hydration** destabilize the unfolded state of halophilic proteins and stabilize the native conformation.

We propose a new mechanistic model that provides a fresh perspective to explain some problems related to *halophilia*, as well as some more general issues:

- ✓ The complex modulation of the **catalytic activity** of halophilic enzymes as a function of the differential stabilization of isoforms with different solvent-exposed surface.
- ✓ The correlation between salt stabilization of halophilic proteins and the changes in **solvent-exposed area**.
- ✓ The dependence of salt stabilization of proteins with the **Hofmeister** series, especially for halophilic proteins and anions.
- ✓ The influence of ions in the **folding** landscape of halophilic proteins, particularly **obligate halophilic proteins**, which are unfolded in the absence of salt due to the electrostatic repulsion.
- ✓ The stabilization of halophilic proteins by neutral organic cosolutes, as well as the role of side chains in the context of **osmolyte stabilization** of proteins.

8. SUMMARY

*"La única materia prima verdaderamente
inagotable es la materia gris"*

Aquilino López Díez

8. SUMMARY

Salt Adaption of Extreme Halophilic Organisms

Adaptation of organisms to extreme halophilic environments ($> 1 - 2$ M) occurs through the accumulation of large intracellular concentrations of KCl. Their major adaptive feature is the extensive modification of the constituting proteome. A biased set of amino acids is selected in order to improve the stability and solubility of halophilic proteins: large hydrophobic residues are penalized, specially lysines, whereas small, polar and often negatively charged residues are favoured, such as aspartic acid, threonine and glutamic acid (Figure I4). The modifications occur mainly at the surface, so the overall structure is conserved. The molecular determinants for such a selection remain elusive despite of considerable efforts. Previous models based on weak unspecific K^+ -carboxylate interactions have proved themselves insufficient to explain some features of haloadaptation, such as the complex salt-modulation of enzymatic activity or the dependence of protein stability with Hofmeister anions.

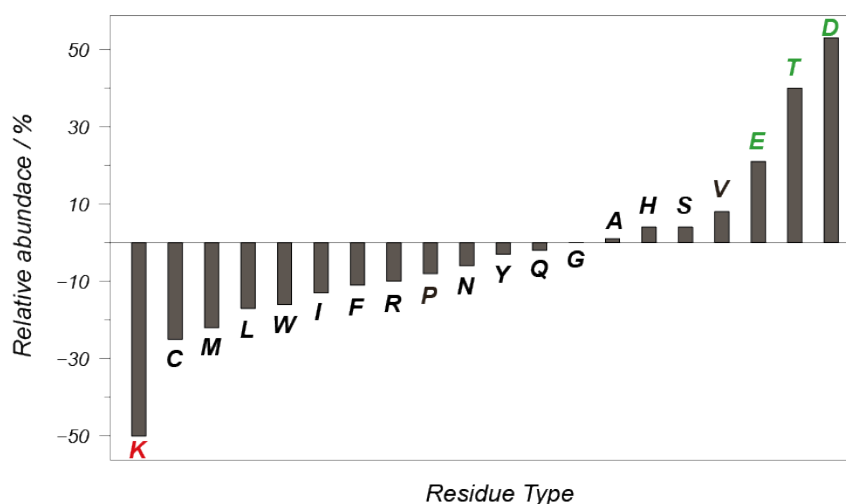


Figure I4. Halophilic amino acid composition. Residue abundance in halophilic proteins compared to mesophilic proteins expressed as the percentage of relative variation in the average amino acid composition.

Working Hypothesis: Correlation with Solvent-Exposed Area

KCl, and to a lesser extent NaCl, monotonically stabilize halophilic proteins. The salt-induced stabilization of proteins – the *m*-value – is considered the main reported for the halophilic character of a protein. Previous work in our lab showed that the reduction in solvent exposed hydrophobic area held a strong correlation with the *m*-values with KCl and NaCl for a series of case-study proteins: halophilic and non-halophilic, natural and rationally designed. Thus, a relevant feature in halophilic adaptation, which had been largely ignored by the established models, was brought to the front line for the study of halophilic adaptation. Based on this novel ground-breaking result, our working hypothesis is that halophilic adaptation depends on general unspecific mechanisms involving solvent-protein-cosolute interactions rather than on specific electrostatic interactions. Protein stability and its modulation by cosolutes rely on the balance of contributions between the folded and the unfolded states. Hence, in order to get a complete mechanistic picture, we must also tackle the unfolded state of halophilic proteins.

Since the adaptation occurs through a selection of certain types of amino acids, we have pursued the characterization of KCl influence on each individual residue. NMR provides information at atomic resolution, thus being a particularly adequate technique to monitor at

the level of individual residues how salt affects different the properties of halophilic proteins. This information was complemented with a biophysical characterization by means of thermal and chemical denaturations monitored by circular dichroism, assays of enzymatic activity, etc.

Proteins under Study

Several protein systems have been employed. As a natural halophilic protein we have chosen a *beta*-lactamase from *Oceanobacillus iheyensis*, β LAC, a globular enzyme with typical halophilic sequence able to remain stable and functional at high molar salt concentrations. As a natural mesophilic (non-halophilic) protein we selected ProtL; a small globular protein often used in biophysical and NMR studies, whose stability changes only marginally upon KCl or NaCl addition. Mutations were engineered in the surface of ProtL in order to make it more halophilic: lysines were substituted by glutamates. Accumulation of five substitutions makes Kx5E ProtL a halophilic protein considerably stabilized by NaCl and KCl. Further substitutions turn Kx6E and Kx7E ProtL into obligate halophilic proteins, which only fold in the presence of molar quantities of salt. Both ProtL and β LAC exhibit long-term stability and reversible thermal unfolding, which make them ideal candidates for NMR and biophysical studies.

As an approximation for the study of the unfolded state of stable proteins we employed intrinsically disordered peptides (IDPs). IDPs are protein sequences with no native structure but with a certain degree of residual structure (similarly to the unfolded state of some globular proteins). Human uSRC was chosen as an example of a mesophilic unstructured protein. No available halophilic IDP was found, so we designed a new one using uSRC as a template. A number of mutations were introduced in order to make its amino acid sequence match that of halophilic proteins. The resulting haloSRC is also an intrinsically disordered protein that exhibits common structural features with uSRC.

Protein Stabilization upon KCl Addition

The amino acid sequence is ultimately responsible for the halophilic character of proteins. Thermal melts monitored by CD show a larger global KCl stabilization (m_{KCl}) for halophilic than for mesophilic proteins (Table R1).

Protein (MW / KDa)	χ^2	m_{KCl} kcal·mol ⁻¹ ·M ⁻¹
β LAC (29)	3.9	5.7 ± 0.5
HvCopG (5)	4.9	6.5 ± 1.5
HvLigN_ade (77)	5.5	4.1 ± 0.5
HvLigN_dea (77)	5.5	1.3 ± 0.5
Kx5E ProtL (7)	8.7	5.8 ± 0.9
WT ProtL (7)	20.7	1.7 ± 0.7
haloSRC (10)	4.3	n.a.
uSRC (10)	15.9	n.a.

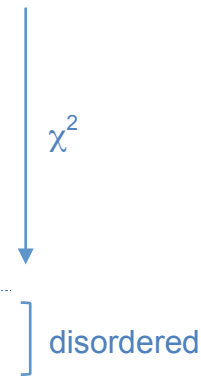


Table R1. Stability and halophilicity. Deviation from a consensus average halophilic amino acid composition, χ^2 , and global stabilization upon KCl addition, m_{KCl} , obtained from thermodynamic data of global protein unfolding for a series of protein systems (Molecular Weight in kDa). HvCopG: translational repressor from halophilic archaea *Haloferax volcanii*

The Role of Charged Groups & Electrostatic Interactions

The great excess of acidic residues in the surface of halophilic proteins has been the main inspiration for traditional halophilic models. We pursued a detailed characterization of cation–carboxylate interactions measuring by NMR the pK_a values for each acidic residue. Salt induces a shift in pK_a as a consequence of a differential stabilization in the charged specie as compared to the neutral one. Our results show a negligible contribution for WT ProtL, as expected, and a stabilizing contribution from electrostatic interactions in β LAC and Kx5E ProtL. The nature of the interactions, however, differs from tight, specific binding of heavily hydrated K^+ cations to particular arrangements of carboxylates. Instead, our data show a very weak binding ($< 1 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$) occurring throughout the whole protein; in support of a model in which the unspecific screening of electrostatic repulsion between negatively charged carboxylates by K^+ cations leads to the stabilization of the folded, compact conformation. The overall contribution to protein stability can be estimated from the sum of the individual contributions for all the acidic groups in the protein. This contribution is almost null for WT ProtL, but much larger for Kx5E (pink bars). However, it is clearly insufficient to explain global KCl stabilization (white bars) and haloadaptation (Figure R9).

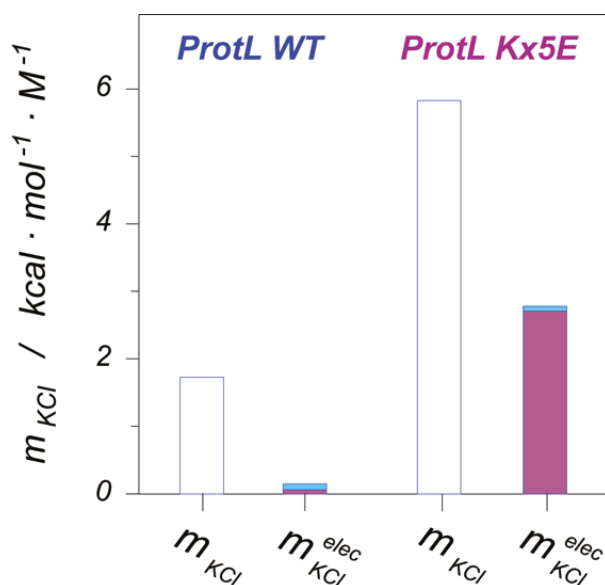


Figure R9. Contribution of electrostatics to global stabilization. Global KCl stabilization from chemical denaturation experiments (white bars) versus electrostatic contribution to haloadaptation (m_{KCl}^{elec}) from carboxyl groups (purple bars) and lysine amine groups (blue bars) for WT and Kx5E ProtL.

Study of the Hydration Shell of Halophilic Proteins

A recurring paradigm in traditional halophilic models is that, as a result of the strong interactions between hydrated cations and carboxylate clusters, the hydration shell of halophilic proteins is tighter than that of their mesophilic counterparts. We challenged this hypothesis by means of ^{17}O NMR, which allows selective observation of the water molecules in the solvation shell. Measurements were performed on mesophilic WT ProtL in the absence of salt and on halophilic Kx6E mutant in the presence of 750 mM NaCl. Two experimental approaches were adopted.

Magnetic relaxation dispersion experiments report on internal water molecules and long-lived waters; i.e. water molecules interacting with the surface of the proteins with residence times larger than 1 ns. No significant differences were detected in the halophilic hydration shell as compared to that of the mesophilic protein. Thus, the existence of water molecules forming part of tight clusters of hydrated K^+ cations and carboxylate arrangements can be ruled out at timescales \geq ns.

Additionally, NMR longitudinal relaxation of water molecules in the solvation shell was measured in order to monitor the dynamics of the hydration shell at a wide range of temperatures. Clearly, the large number of carboxylate groups and the large negative net charge density on the surface of Kx6E does not result in a reduction of water dynamics, as it might have been expected for the "highly ordered, multilayered solvation shell" of "quasi-immobile" and "tightly bound" water molecules proposed by the traditional halophilic model.

Results from this study imply that there is not discernable halophile signature in the hydration dynamics: the correlation time of water molecules in the solvation shell is similar for WT ProtL (no salt) and Kx6E ProtL (750 mM NaCl) and no additional tightly bound water molecules with residence time above 1 ns are detected. This is consistent with the model proposed here based on weak, unspecific ion-protein interactions driven by the screening of electrostatic repulsion.

Role of the Unfolded State in Salt Adaptation

The structures of halophilic proteins are conserved with respect to their mesophilic counterparts. Accordingly, no evidences for compaction or increased rigidity could be found for the native structures of halophilic proteins. We found differences in the electrostatic fingerprint, but those do not suffice to explain salt stabilization of halophilic proteins. Cosolute effect on protein stability depends on the balance between the folded and unfolded states of the contributions from excluded volume and intermolecular interactions. Therefore, we focused on the unfolded state through the study of models for unfolded proteins.

We selected an artificial obligate halophilic mutant, Kx7E ProtL. Below 1 M KCl, it shows detectable populations of folded and unfolded state. Exchange Spectroscopy experiments (EXSY) measures the folded and unfolded populations (thermodynamics) and the rates of exchange between them (kinetics). The data agree well with a two-states equilibrium between folded and unfolded state. KCl displaces the population towards the folded state by greatly increasing the folding rate constant; that is, destabilizing the unfolded state. This constitutes strong evidence for a novel mechanism pointing towards the destabilization of the unfolded state as the main driving force for salt stabilization and, most likely, also haloadaptation.

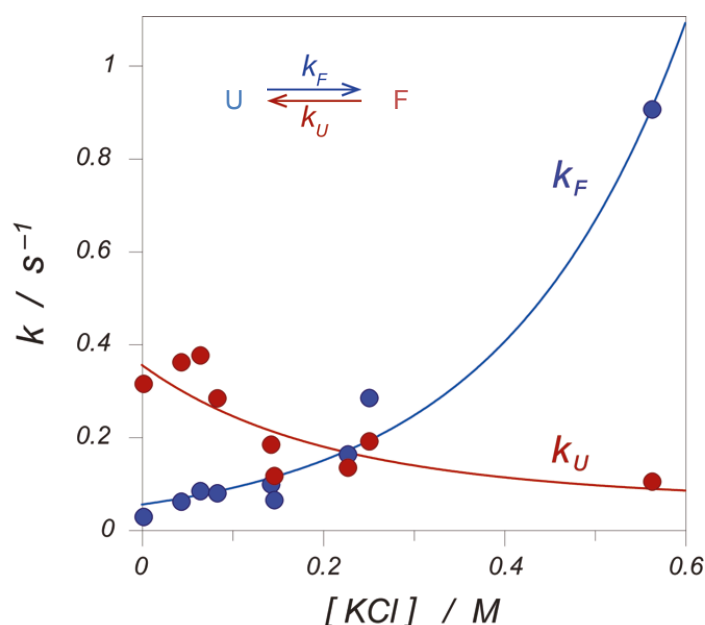


Figure R15. KCl modulation of folding and unfolding rates. Experimental kinetic rate constants (circles) for folding (k_F , blue) and unfolding (k_U , red) of the obligate halophile Kx7E ProtL obtained from EXSY experiments. Monoexponential fittings (lines) show that k_F depends more strongly on KCl than k_U .

We then looked into the particular mechanisms for this salt-induced destabilization of the unfolded state. Variations in compaction or dynamics of unfolded halophilic proteins leading to changes in the excluded volume contribution to protein stability were probed by NMR. Full relaxation datasets (R_1 , R_2 and NOE) were acquired for unfolded uSRC and haloSRC at increasing molar quantities of KCl and the data were analysed following Lipari–Szabo model-free formalism. The resulting parameters showed no evidence for changes in the global dynamic behaviour – i.e. the amplitude (S^2) and frequency (τ_c) of the dynamics – between halophilic and mesophilic unfolded states or upon salt addition.

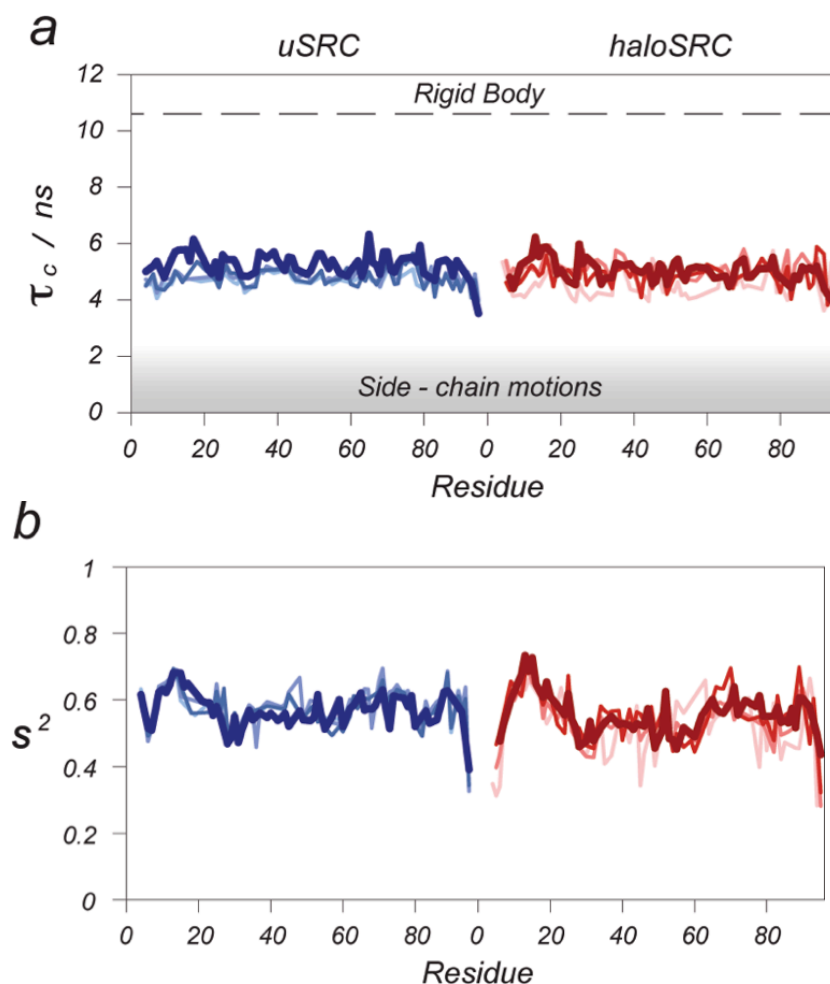


Figure R18. Fast global dynamics obtained for uSRC (blue) and haloSRC (red). Per residue correlation times τ_c (a) and order parameters S^2 (b) were obtained from fitting to Lipari–Szabo model-free formalism of R_1 , R_2 and NOE data. Local correlation times in (a) have intermediate values between modelled rigid body motion and side-chain unrestricted motions. Darker to lighter gradient colour correspond to 0, 0.5, 1 and 1.5 M KCl.

In parallel, analysis of the $R_{1\rho}$ showed that local dynamics (hydrophobic clusters related to lipid binding regions) were also similar for uSRC and haloSRC. Again, in both cases dynamics did not change much upon KCl addition (Figure R16).

Our relaxation data demonstrate that, within the timescale-sensitivity of the experiments, the denatured conformational ensemble of halophilic proteins is similar to that of mesophilic proteins and remains invariant to KCl concentration. Considering the high structural conservation of the folded state, we can conclude that haloadaptation cannot be attributed to changes in the excluded volume contribution.

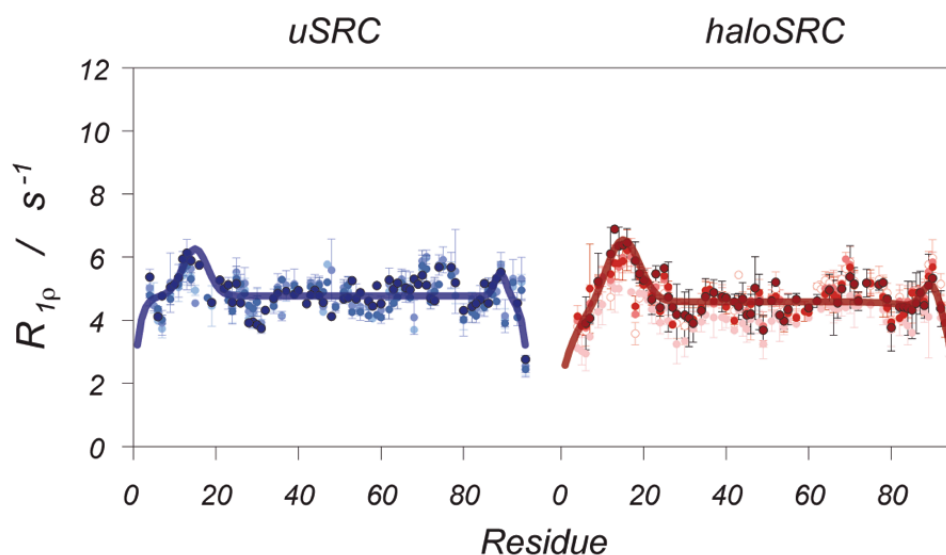


Figure R19. Local dynamics of uSRC and haloSRC. ^{15}N relaxation rates in the rotating frame $R_{1\rho}$ for uSRC (blue) and haloSRC (red) measured at KCl concentrations increasing (from dark to light colour: 0, 0.5, 1 and 1.5). Solid line represents the least square fitting of data at 0 M KCl to a model for segmental motion assuming reduced flexibility around two hydrophobic clusters near the N and C-termini.

Preferential Ion Exclusion from Halophilic Unfolded States

Once discarded excluded volume contributions to haloadaptation, we moved on to test the contribution from intermolecular protein-cosolute interactions. We measured the chemical shift perturbation induced by KCl (CSP_{KCl}) for uSRC and haloSRC. It serves as an indicator of the effect of KCl in the chemical environment of individual backbone amide bonds. As opposed to folded proteins, where no significant differences were found between halophilic and mesophilic CSP_{KCl} coefficients, our results show a *lower* sensitivity of halophilic unfolded state to the presence of ions in solution:

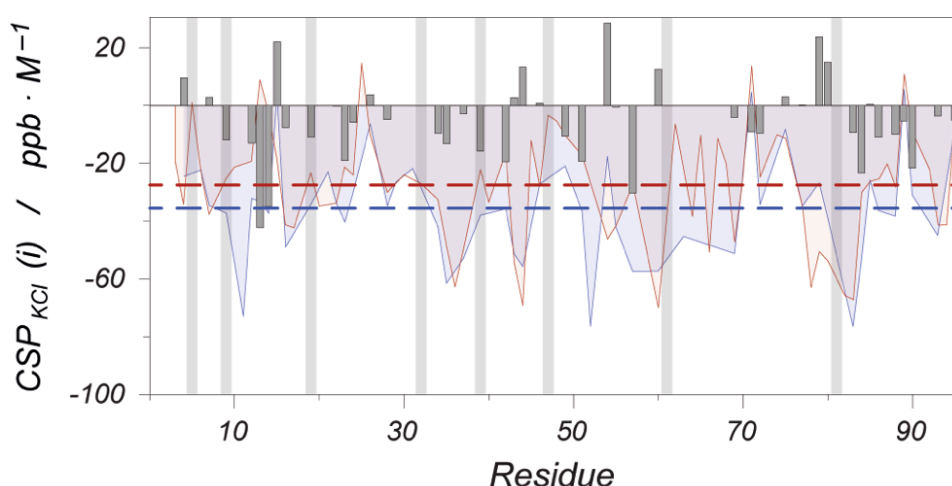


Figure R20. Chemical shift perturbation coefficients. $\text{CSP}_{\text{KCl}}(i)$ experimental values for uSRC (blue) and haloSRC (red) are represented as coloured areas, while dashed lines mark the averages. Grey bars correspond to the differences between the coefficients of the halophilic and mesophilic proteins, where available. Shaded columns outline the mutated positions.

There is an overall decrease in the absolute value of the CSP_{KCl} upon increasing the halophilic character of the protein. The effect is not limited to the mutation sites. Rather, the whole protein becomes less sensitive to salt ions as a consequence of the introduction of

halophilic aminoacids. This result is further confirmed by CLEANEX experiments, which measure the fast exchange of backbone amides' protons with the solvent. Upon salt addition, both exchange rates slow down with molar KCl concentration probably due to salt ions occluding binding sites at the backbone and obstructing water molecules. As shown in figure R18, the decay can be adjusted to a monoexponential function characterized by a single decay rate constant, R_{KCl}^{CLEX} (summarized in Figure R19). Remarkably, halophilic haloSRC shows lower decay rates R_{KCl}^{CLEX} ; i.e., a faster exchange at high salt concentrations. This is totally consistent with a lower presence of ions near the surface of halophilic unfolded states.

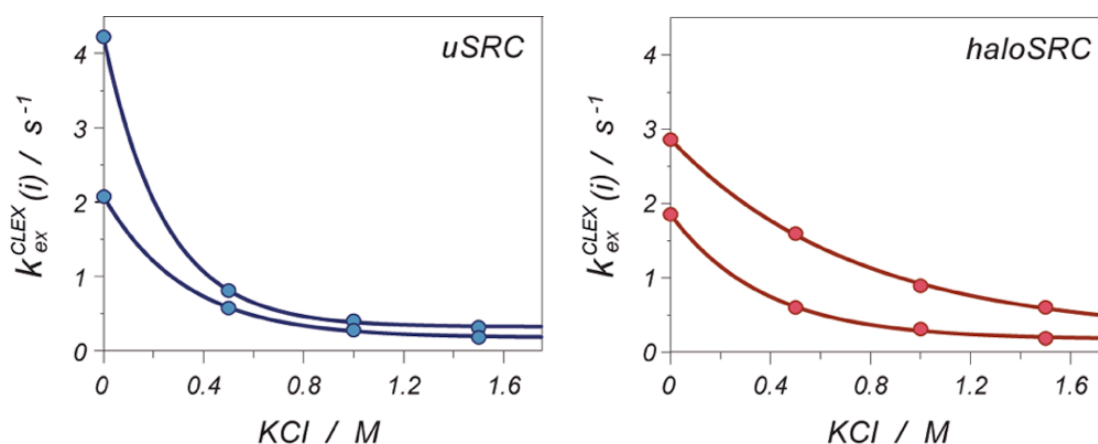


Figure R21. KCl slows exchange exponentially. Fast amide ^1H exchange rates, $k_{\text{ex}}^{\text{CLEX}}$, derived from CLEANEX experiments at different KCl concentrations are illustrated for two representative residues of uSRC and haloSRC. Upon KCl addition, $k_{\text{ex}}^{\text{CLEX}}$ rates decrease exponentially (solid lines) with characteristic decay constants R_{KCl}^{CLEX} .

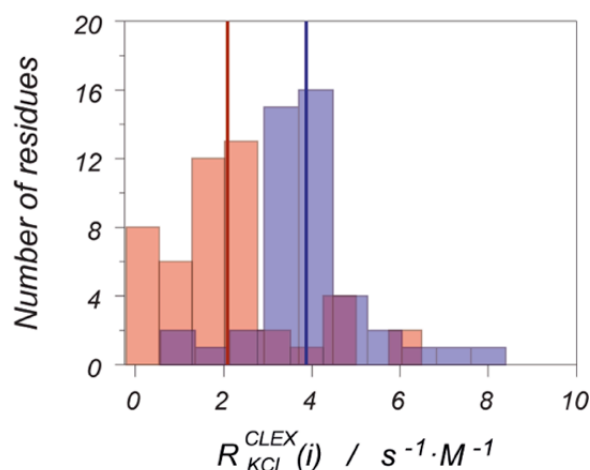


Figure R22. Rates of KCl-induced exchange decay. Distribution of R_{KCl}^{CLEX} constants for uSRC (blue) and haloSRC (red) with their indicated averages (vertical lines) reveal that ^1H exchange rates of the halophilic protein are less sensitive to salt (i.e., generally lower values for haloSRC).

Preferential cosolute exclusion from the protein surface has been attributed a destabilizing effect. Therefore, our experimental results provide evidence for a novel important contribution to haloadaptation: the preferential ion exclusion from the unfolded state of halophilic proteins destabilizes the unfolded state, ultimately stabilizing the native state.

Quantification of Unfolded State Destabilization: H/D Exchange

Accurate quantification of the contribution to energetics of weak interactions and preferential exclusion is extremely complicated. The characterization of weak interactions and competition with hydration is still limited to theoretical studies, computer simulations and experiments on simple surface-models. We have adopted a thermodynamic approach for the quantification of preferential ion exclusion energetics: measurement of the slow H/D exchange of backbone amides of β LAC, WT and Kx5E ProtL as a function of KCl concentration. H/D exchange experiments allow probing local salt stabilization at the level of individual residues. The linear KCl dependence of individual local stability, $\Delta G_{U-F}^0(i)$, can be reported as the slope $m_{KCl}^{H/D}(i)$. Figure R23 shows the results for WT and Kx5E ProtL:

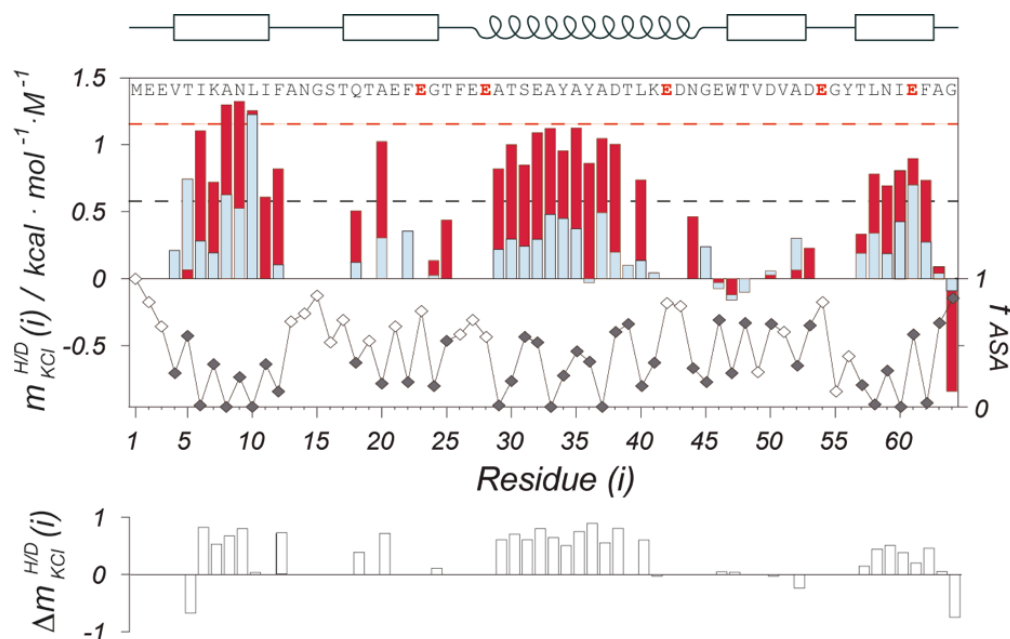


Figure R25. H/D exchange of ProtL. Bar-plot with local KCl stabilization $m_{KCl}^{H/D}(i)$ values derived from H/D exchange rates for each residue of natural mesophilic WT ProtL (light blue) and designed halophilic Kx5E mutant (red). Average values of five most protected residues are shown as colour-coded dashed lines. Fractional solvent accessible surfaces, f_{SAS} , are shown below (right axis) and colour coded according to the timescale of the exchange: measurable residues with intermediate timescales (grey circles) and surface residues with too short timescales (< 30 min, white circles). Schematic representation of the secondary structure is shown above the plot, as well as the sequence with mutated residues highlighted in red. Bar-plot at the bottom (white bars) shows the pairwise differences (when available) in KCl stabilization, $\Delta m_{KCl}^{H/D}(i)$.

There is a general increase in KCl-stabilization spread throughout the whole Kx5E sequence, supporting the idea of a general unspecific mechanism. Remarkably, mutated positions do not show outstanding $m_{KCl}^{H/D}(i)$ values. Instead, buried residues are the ones showing larger KCl-stabilization and larger increments upon introduction of the halophilic amino acids: there is a certain correlation between the $m_{KCl}^{H/D}(i)$ value of a certain amino acid and its solvent-exposure. This profile is similar to the one of β LAC (Figure R22). Protein cosolute-stabilization depends on a balance of the contribution of interactions in the folded and the unfolded states. Exposed residues contribute the same on both states, showing almost negligible stabilization values. In turn, fully buried residues show the largest differences between folded and unfolded states. As they are protected from solvent interaction in the folded state, their $m_{KCl}^{H/D}(i)$ values report directly on the impact of ion interactions in the unfolded state; i.e. the energetic stabilization due to the preferential exclusion and the resultant destabilization of the unfolded state. Our results show that KCl strongly (slightly) destabilizes the unfolded state of halophilic (mesophilic) proteins (Figure R24).

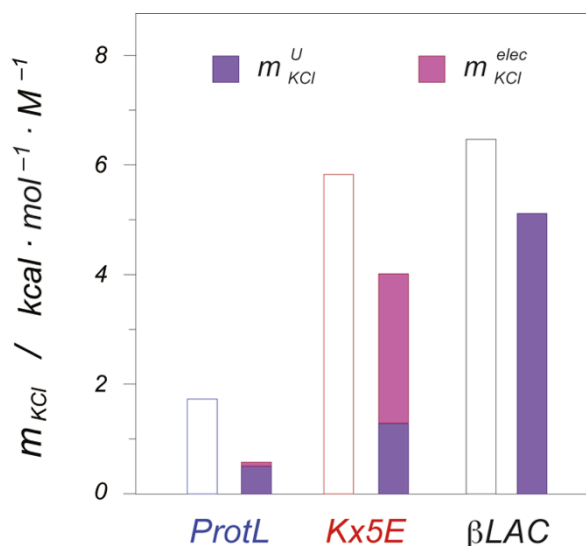


Figure R26. KCl dependent stabilization. Global stabilization m_{KCl} derived from thermal denaturations (white bars) is compared with the contributions from the unfolded state preferential exclusion m_{KCl}^U (purple bars) and from the electrostatic interactions with the carboxylic groups in the folded conformation m_{KCl}^{elec} (pink bars).

In summary, our results show that the unfolded state of halophilic proteins cannot form the multiple stabilising cation–carboxylate interactions prevailing in the well–ordered surface of the folded conformation. On the contrary, ions are preferentially excluded from the surface of halophilic unfolded conformations. The resulting destabilization of the unfolded state is comparable in magnitude to the stabilization of the folded state, being another distinctive driving force for salt stabilization of halophilic proteins and, hence, haloadaptation.

Contribution from Excluded Volume and Other Interactions

Additional contributions to KCl-stabilization derive from the mutually excluded volume effect and from ion–protein interactions other than the ones already considered (those with carboxylates and NH groups from backbone amides and lysines). The total m_{KCl} value then breaks down into:

$$m_{KCl} = m_{KCl}^{elec} + m_{KCl}^U + RT \cdot \Delta X + m_{KCl}^{other} \quad ;$$

Equation R14. Contributions to overall KCl stabilization. Global salt stabilization as the sum of the contributions from: electrostatic interactions, unfolded state destabilization through preferential exclusion, excluded volume and other interactions.

Excluded volume contribution arises from the entropic penalty of creating a cavity to host the protein molecule. It is a stabilizing contribution because it favours more compact folded conformations. It can be estimated from the volumetric difference between native and denatured conformations (ΔX in \AA^3). The high structural homology between halophilic and mesophilic native states has since long been described, same as the lack of conformational rearrangements upon salt addition. Our results for the unfolded uSRC and haloSRC models suggest the same for the denatured ensemble conformation. Therefore, no relevant changes are expected between halophilic and mesophilic proteins. Indeed, within calculation uncertainty, the ΔX value is the same for both ProtL variants. Thus, excluded volume explains great part of KCl-stabilization, but not haloadaptation.

As shown in Figure R25, the addition of excluded volume contribution to those already reported is not enough to account for the whole m_{KCl} value. Remaining cosolute contributions would rely on **intermolecular interactions**. Interactions other than the ones already

determined experimentally result from the comparison between all estimated contributions and the global m_{KCl} value. They are represented by the term $m_{\text{KCl}}^{\text{other}}$. It is negative for both ProtL variants – hence, destabilizing – due to the larger area exposed to the solvent in the denatured conformation. Remarkably, it is smaller for the halophilic Kx5E ProtL than for the WT ProtL, in line with the observed preferential ion exclusion from the unfolded state of halophilic proteins.

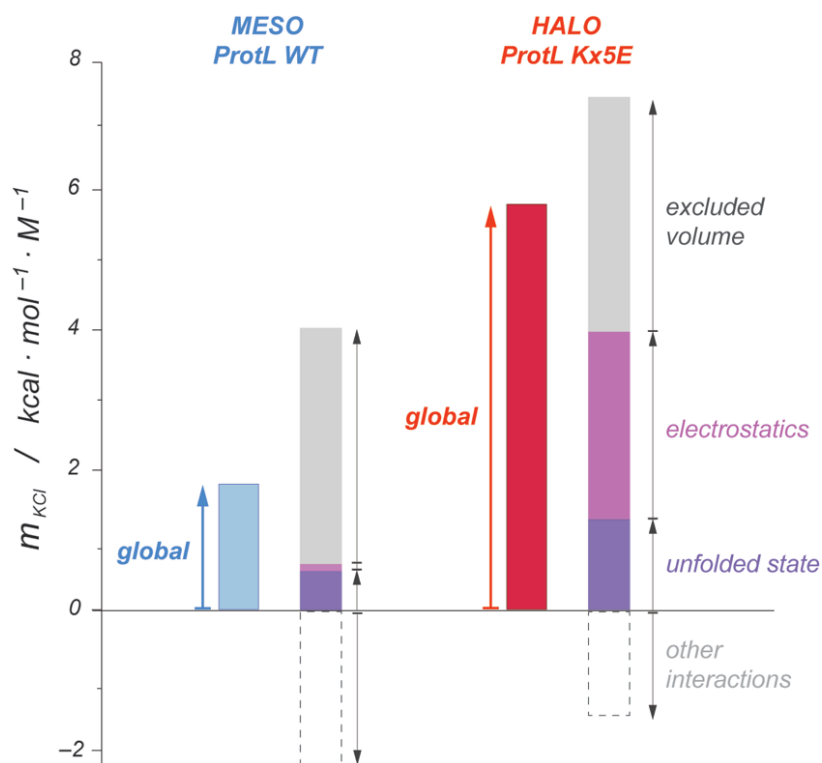


Figure R25. Contributions to KCl stabilization of ProtL. Bar plot showing the experimental global m -values for WT (blue) and Kx5E (red) ProtL as well as the different thermodynamic contributions of KCl to global protein stabilization: electrostatic interactions with the side chains ($m_{\text{KCl}}^{\text{elec}}$, pink), preferential ion exclusion from amide backbone in the unfolded state ($m_{\text{KCl}}^{\text{u}}$, purple), excluded volume (ΔX , grey) and the loss of other interactions upon folding ($m_{\text{KCl}}^{\text{other}}$, white), which is always destabilizing, but less for halophilic proteins. Both electrostatic interactions and preferential ion exclusion contribute at a similar extent to haloadaptation.

Contributions to Haloadaptation: Interactions in the Folded State

Alternatively to traditional models based on tight and highly specific interactions between heavily hydrated K^+ cations and carboxylate clusters, we propose a model in which the weak and unspecific screening of electrostatic repulsion provides a considerable salt-induced stabilization. Then, it is the accumulation of acidic groups rather than their particular geometry what triggers the favourable interactions. Accordingly, such interactions are negligible in non-adapted proteins. This stabilizing enthalpic contribution, however, is not enough to explain the whole KCl-stabilization or haloadaptation.

Contributions to Haloadaptation: Interactions in the Unfolded State

Using IDP's as an approach and NMR, we have studied the unfolded state of halophilic versus non-halophilic proteins. Our NMR relaxation data reveal that the local and global dynamics of the unfolded states are very similar in halophilic and mesophilic proteins, and no changes are detected upon KCl addition. Thus, no evidences for changes in the compaction or the conformational ensemble of halophilic proteins are found.

Regarding intermolecular interactions, no significant interactions between K^+ cations and carboxylate groups are detected. Unlike native halophilic conformations, the contribution of electrostatics in the unfolded state is almost negligible. When probing interactions with the backbone of the protein (CSP_{KCl} and CLEANEX), ions were found to be preferentially excluded from the halophilic unfolded state. Preferential exclusion has an energetically destabilizing effect because constructive protein–ion interactions are no longer occurring. Instead, both components have to compete for their separate solvation. This entropic penalty, not compensated by the diminished enthalpic contribution, is especially unfavourable at high salt concentrations, where less water is available. The extent of the destabilization is directly proportional to the area exposed to solvent. By LeChatelier principle, the system minimizes the thermodynamically unfavourable ion-exclusion by populating the more compact folded state. Thus, preferential ion-exclusion from the unfolded halophilic state ultimately stabilizes the native conformation. Its magnitude is comparable to that of electrostatic interactions and therefore constitutes the other main driving force behind haloadaptation.

Our results indicate that both folded and unfolded states determine the overall energetics of haloadaptation. We report the two main mechanisms contributing to the salt stabilisation of halophilic proteins in high molar salt concentrations: i) *stabilization of the folded state by cationic non-specific screening of the electrostatic repulsion between acidic groups* and ii) *destabilization of the unfolded state by the preferential ion exclusion*. Haloadaptation relies on these two alternate mechanisms that cooperatively affect both the folded and unfolded states. This proposed model is summarized in the semi-quantitative energetic diagram in Figure D3.

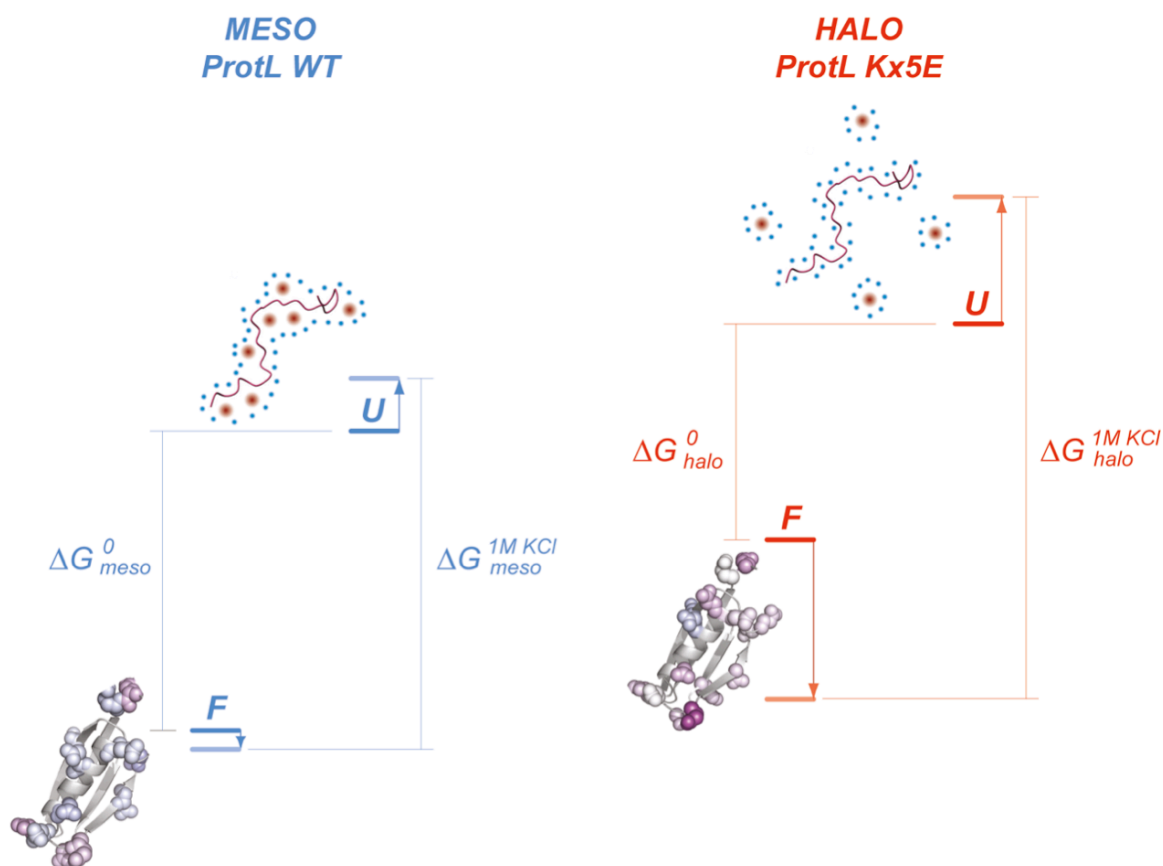


Figure D3. Energetics of protein adaptation to hypersaline environments. The scheme shows the energy levels in the absence (dark lines) and presence (faint line) of 1M KCl for halophilic Kx5E (right) and mesophilic WT (left) ProtL. In the absence of salt, mesophilic protein is more stable, but the situation is reversed at 1 M KCl. The depicted halophilic unfolded state reflects the destabilization due to the preferential exclusion of ions from its hydration layer, whereas the folded ProtL shows the electrostatic stabilization.

Why This Amino Acid Composition?

The mechanism proposed here requires a minimum of charge density on the protein's surface: enough to stabilize multiple weak and unspecific interactions with solute cations in the folded state that, in turn, cannot form properly in the unfolded state. Aspartate and, to a lesser extent, glutamate provide highly polar, negatively charged moieties on short side chains. This makes them to be readily hydrated without imposing high entropic costs when fixing a favorable conformation for cation interactions.

Unlike folded state, halophilic unfolded states exclude ions from the surface avoiding ion-protein interactions. This alternate switching of interactions along the folding coordinate, which is the basis of our mechanistic model, is enhanced by halophilic amino acid selection. Preferential exclusion occurs mainly at polar hydrophilic surfaces, where water binds more effectively than ions. The large decrease of hydrophobic area in halophilic proteins – lysine reduction and decrease of residues with large hydrophobic side chains – together with the increase in short, polar residues – aspartate, threonine – enhances ion-exclusion, protein solvation and the concomitant unfolded state destabilization mechanism.

In summary, negatively charged groups, residues with low hydrophobic content and small polar side chains are preferred in order to promote the synergistic switch between ion-protein interactions in the folded state and exclusion in the unfolded state, which ultimately confers halophilic proteins its great stability at high KCl concentrations.

Obligate Halophilicity *versus* Solubility

The electrostatic repulsion between negatively charged residues is the main responsible for protein obligate halophilicity. In the absence of salt, the repulsive forces are too strong and unfold the protein. The unfolded state has a larger conformational freedom that allows the protein to find more favourable conformations and reduce the unfavourable repulsion. Then, why negatively charged residues are favoured? The presence of a high number of acidic residues provides an enhanced solubility in media where the requirements of ions for solvation reduce the availability of water molecules for protein hydration. As a result, processes reducing the exposed surface, such as aggregation, are favoured. Coating of the protein surface with negative charges provides an intermolecular repulsion that avoids aggregation. Additionally, it helps protein-ion co-solvation. Thus, we propose that obligate protein halophilicity is a non-adaptive property resulting from a genetic drift in the seek of increased solubility, not a mandatory “fee” in order to enhance protein stability in concentrated salt solutions.

Contribution of the Hydration of Solvent Exposed Hydrophobic Area

Previous work from our group revealed that the reduction in solvent accessible area results in an increase in salt stabilization, constituting a key factor in halophilic adaptation even more determinant than overall negative charge. Protein solvation is particularly demanding at high salt concentrations, where water activity is reduced and a competition for water molecules takes place between ions and proteins. Halophilic proteins adapt by diminishing their solvent-exposed hydrophobic area, solvation of which is entropically penalized. The increase in polar groups and the drop in methylenes from lysines favour water-protein interactions, which enhance protein hydration, solubilisation and, at the same time, the unfolded state destabilization by the aforementioned ion-exclusion mechanism. As opposed to non-adapted proteins, halophilic proteins can achieve an efficient solvation even in molar salt concentrations, where water activity is drastically reduced.

Halophilic Adaptation and Folding Landscape

Haloadaptation can be equally understood as the mechanism for folding a halophilic protein. Some of our data allow us to overcome the two-states folding model and evaluate how salt and halophilic adaptation affects each step in the folding energetic landscape (Figure D8). Particularly, EXSY experiments show that most of the stabilization occurs before the transition state is reached. Early collapse of denatured state ensemble is usually the rate-limiting step for this folding transition. Electrostatic repulsion hampers this first collapse (dashed line), which is allowed by KCl electrostatic screening (red solid line). In addition, preferential ion-exclusion from denatured state also penalizes extended conformations, thus favouring more compact states along the folding pathway. In summary, we propose a model in which electrostatic screening and ion-exclusion progressively stabilize more compact conformations in halophilic proteins. Therefore, KCl exerts a driving force for folding proportional to the fraction of native ASA buried (θ_{ASA}); that is, progressively favouring progress towards the native conformation.

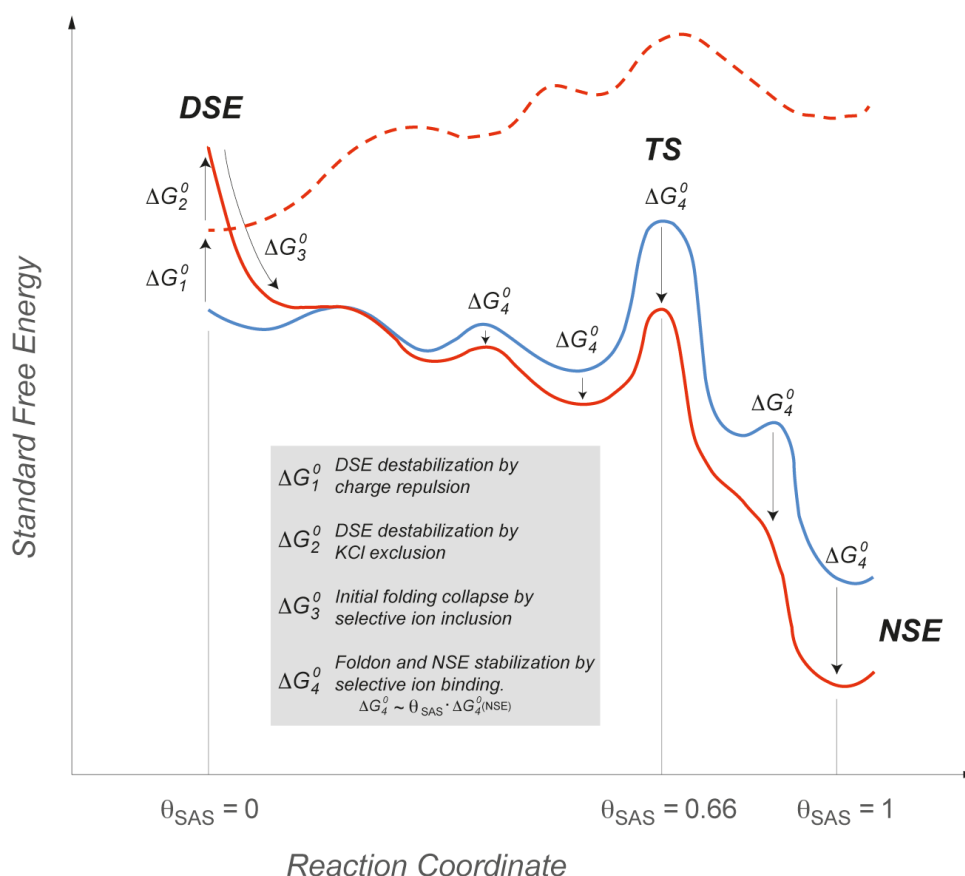


Figure D8. Folding landscape modulation in protein adaptation to hypersaline environments. Representation of the folding energy landscape for mesophilic ProtL (blue) and an obligate halophilic homologue in the absence (red dashed line) and presence (red solid line) of molar KCl concentrations. Horizontal axis shows the folding coordinate as the fraction of ASA buried with respect to the native conformation, θ_{ASA} . In contrast to mesophilic proteins, electrostatic repulsion in the absence of salt destabilizes obligate halophiles at an extent proportional to θ_{ASA} , thus avoiding population of more compact structured conformations (ΔG_1^0). Upon salt addition, ion exclusion penalizes extended conformations (ΔG_2^0); complementary, KCl electrostatic screening helps the protein early collapse along the reaction coordinate (ΔG_3^0), ultimately leading to further energy reduction in the transition state and the folded conformation (ΔG_4^0). The electrostatic stabilization due to K^+ cation – carboxylates interactions is proportional to the compactness of the conformation, as represented by the increasing arrow length.

Comparison with Other Ions and the Hofmeister Effect

Several reports associate halophilia with Hofmeister effects. Halophilic proteins are more sensitive to Hofmeister ions, especially anions. Since long-time known, the molecular mechanisms behind Hofmeister series are still being unravelled. Current views point to weak interactions with the protein surface. We propose that halophilic amino acid selection amplifies the ions' impact on the stability and function of halophilic proteins by modulating ion–protein interactions. The increase of polar and hydrophilic area provides more sites for the competition between hydration and ion binding, therefore enhancing ion exclusion for stabilizing ions or ion binding for destabilizing ions. This effect is particularly relevant for anions, which in general are stronger modulators of stability than cations. Regarding cations, K^+ is largely preferred over Na^+ even though the latter is usually much more abundant in the extracellular milieu. This is probably because of weaker K^+ hydration properties, which make it a better choice in environments where protein has to compete with cations for hydration.

Mechanistic Comparison between Haloadaptation and Osmoprotection

We tested the behaviour of our halophilic and non-halophilic systems in the presence of a stabilizing, organic osmolyte: sucrose. We found that halophilic proteins are more stabilized by sucrose than mesophilic proteins. In fact, the extent of the stabilization was generally larger than with KCl (Table D1).

Protein (MW / KDa)	χ^2	m_{KCl} kcal·mol ⁻¹ ·M ⁻¹	$m_{SUCROSE}$ kcal·mol ⁻¹ ·M ⁻¹
βLAC (29)	3.9	5.7 ± 0.5	12.6 ± 0.3
HvCopG (5)	4.9	6.5 ± 1.5	8.3 ± 1.2
Kx5E ProtL (7)	8.7	5.8 ± 0.9	7.8 ± 0.1
WT ProtL (7)	20.7	1.7 ± 0.7	4.3 ± 0.5
HvLigN_ade (77)	5.5	4.1 ± 0.5	4.3 ± 0.1
HvLigN_dea (77)	5.5	1.3 ± 0.5	3.5 ± 0.1

Table D1. Sucrose stabilization. Global stabilization upon KCl addition m_{KCl} and the same quantity with sucrose $m_{SUCROSE}$ obtained from thermodynamic data of global thermal unfolding for various halophilic and non-halophilic proteins. HvCopG: translational repressor from *Haloferax volcanii*.

Sucrose is a neutral molecule; therefore no electrostatic stabilization is expected. Instead, several reports attribute sucrose stabilizing effect to preferential exclusion. Preferential exclusion is proportional to cosolute concentration and the amount of area exposed. Since the unfolded state exposes more area to the solvent, it is preferentially more destabilized at increasing cosolute concentrations. This is confirmed by additional experimental evidences:

- Sucrose is capable of folding obligate halophilic Kx7E ProtL. However, it is less efficient due to the inability to screen electrostatic repulsion.
- The H/D exchange profile is similar to that of KCl for the three proteins studied. The individual stabilization values are proportional to the ΔASA^{U-F} , indicating a destabilizing effect of osmolyte in the unfolded state.
- There is a strong correlation between individual $m_{KCl}^{H/D}$ (i) and $m_{Sucrose}^{H/D}$ (i) values, clearly pointing to a common mechanism of KCl and sucrose.

This mechanism is common to other osmolytes such as other sugars (glucose, lactose), polyhydric alcohols (glycerol, sorbitol, mannitol, ethylene glycol) and other compounds (TMAO, sarcosine...).

Halophilic amino acid sequence increases polar, hydrophilic surface and decreases apolar, hydrophobic area in order to increase preferential hydration and subsequent ion exclusion. The exact mechanism for the interactions is not yet fully understood. Current theories are based in transfer free-energy models, in which the fraction of polar area of a given osmolyte determines its stabilizing/destabilizing character. Our data demonstrate that tuning the polar fraction of a protein by means of side chain mutations can effectively modulate osmolytes' effect on proteins. Thus, halophilic mutations provide an additional tool to probe the effect of osmolytes on proteins and the role of area composition. Moreover, we vindicate the importance of side chains, which may have been largely underestimated in favour of the more predominant backbone exposed area.

Catalytic Activity of Halophilic Enzymes

The monotonic salt stabilization of halophilic proteins contrasts with the complex modulation of their catalytic function. Whereas some enzymes are activated by salt, others show a decrease in their catalytic activity at high salt concentrations. Alternatively, the maximal activity of some enzymes occurs at an optimum range of concentrations. More strikingly, other enzymes show a heavy ion-dependent activity modulation. NAD^+ -dependent DNA Ligase N from *Haloferax volcanii*, *HvLigN*, belongs to this last group: it is only active at KCl concentration above 0.5 M. In contrast, it is virtually inactive in the presence of NaCl.

HvLigN exhibits a typical halophilic amino acid sequence and maintains a high structural homology with other DNA ligases. Its mechanism of action is well characterized. Similarly to other ligases, it involves a conformational rearrangement. The initial adenylation step requires 1A moiety closure over domain 1B in order to conform the NAD^+ binding site. Subdomain open – closure dynamics of the adenylation domain and subsequent transition from open (deadenylated) to close (adenylated) conformation are thus essential to start the catalytic cycle. Subdomain closure buries 331 \AA^2 . Our hypothesis is that the apparent complex modulation of enzymatic activity simply reflects the salt differential stabilization of two protein isoforms with different ASA, which are part of the reaction coordinate.

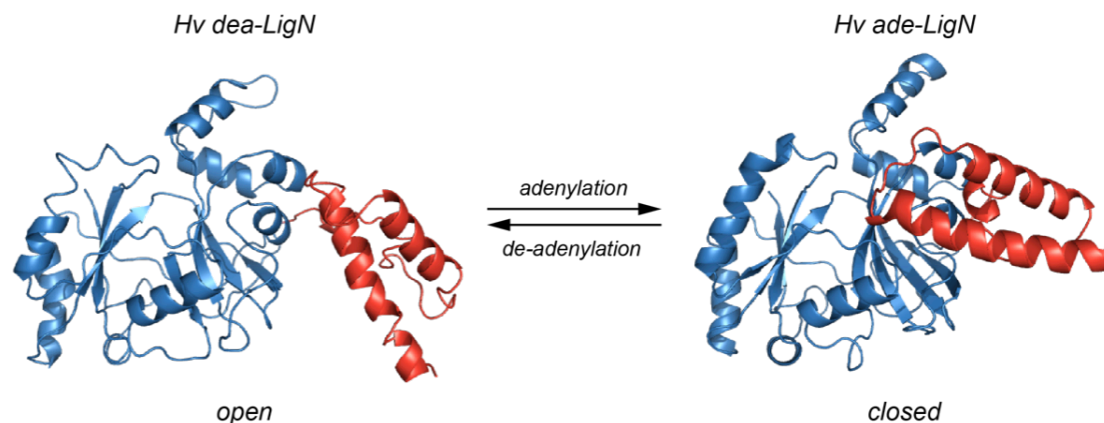


Figure A3. Open–closure dynamics and adenylation. Model of the adenylation domain of *Hv LigN*. The open and closed conformations along the reaction coordinate are represented here. Domain 1A is shown in red and domain 1B in blue. NAD^+ binding causes the domain 1A to close (right). Deadenylation of the adenylated protein with excess NMN leaves the adenylated site exposed to solvent, resulting in the open NMN-bound conformation (left).

Both isoforms were isolated, their stability measured at increasing concentrations of different salt ions and collated to the enzymatic activity (Figure A6). Our results show that, in the absence of salt, closed-adenylated conformation (the active form) is less stable than open-deadenylated conformation (inactive form). As a result, the adenylation step cannot occur. Upon KCl addition, the closed conformation becomes differentially stabilized and, thus,

accessible. As a result, adenylation takes place and triggers ligation activity. In contrast, NaCl equally stabilizes both isoforms, therefore the active form never gets accessible and adenylation does not occur. This coupling between differential stabilization and enzymatic activity is confirmed for other stabilizing (sulphate, phosphate, acetate) and destabilizing (thiocyanate, nitrate and bromide) ions.

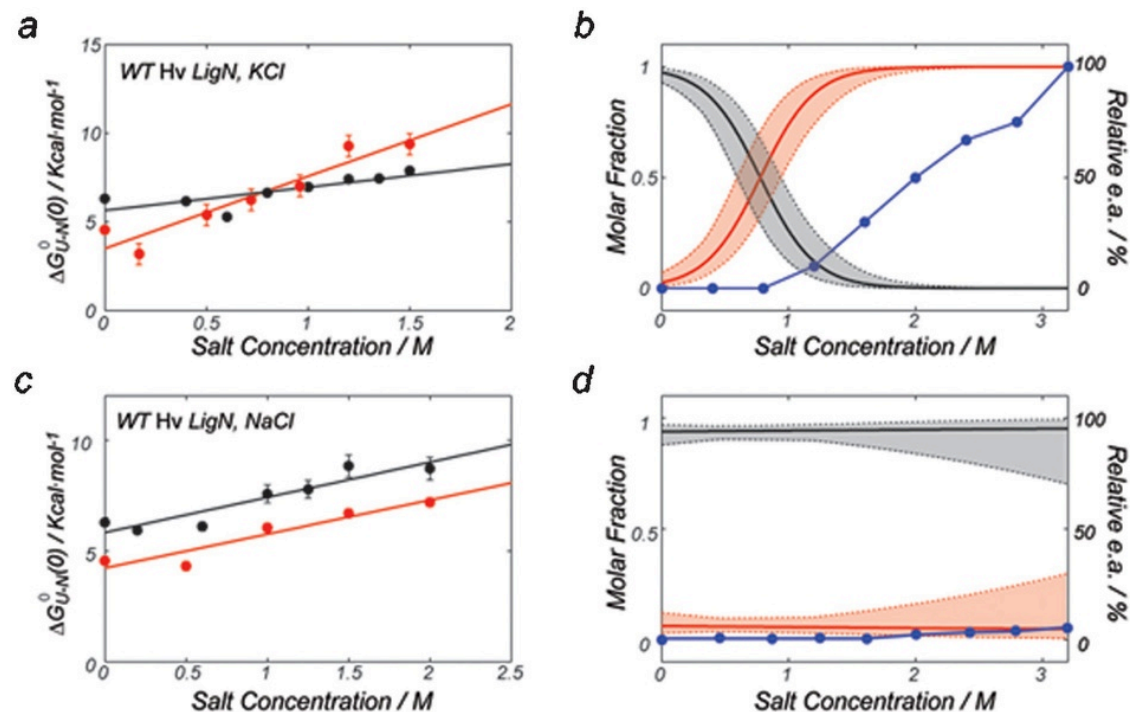


Figure A6. Coupling of stability with enzymatic activity. Plots of the unfolding free energy (ΔG_{U-F}^0) versus molar salt concentration for KCl (a) and NaCl (c). *Hv* dea-LigN and *Hv* ade-LigN are represented as black and red, respectively. Circles mark the experimental values whereas solid lines correspond to the linear regressions. The apparent populations are plotted in panel b and d (left ordinate axis) using the same color code. Shaded areas correspond to the estimated uncertainties in the populations calculated by a Montecarlo simulation. The blue circles show the experimental enzymatic activity normalized to that at 3.2 M KCl (right ordinate axis).

We also introduced two conservative halophilic mutations (ExD) in order to enhance salt-sensitivity of *HvLigN*. We managed to trigger a differential stabilization of both isoforms in the presence of NaCl, thus recovering catalytic activity in the presence of NaCl.

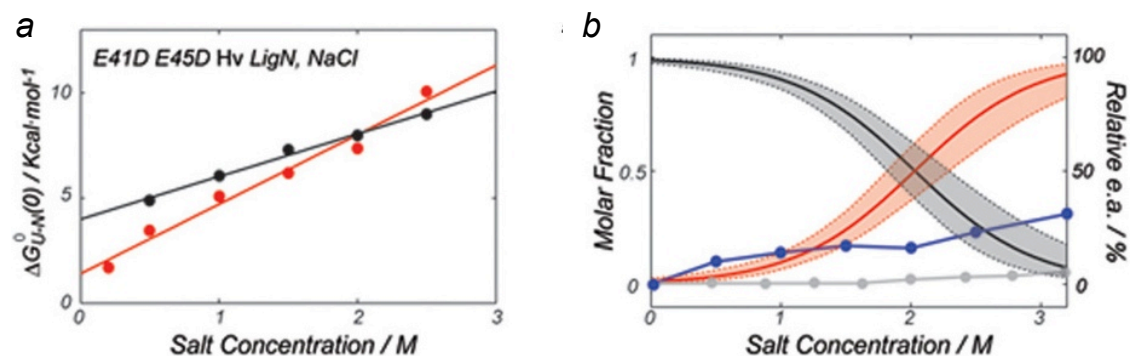


Figure A8. Mutations restore activity in the presence of NaCl. Unfolding free energy (ΔG_{U-F}^0) of E41D E45D *Hv LigN* versus NaCl concentration (a) shows a differential stabilization of adenylated with respect to deadenylated form. As a result, the former becomes accessible and activity is restored.

This mechanism of differential salt stabilization of isoforms with different ASA can be extrapolated to other halophilic enzymes. In cases where the enzymatic catalysis involves conformational changes or oligomerization, differential stabilization of the species in the reaction pathway will also result in an indirect – and apparently complex – salt modulation of catalytic activity.

Conclusions

- ✓ Rational redesign allowed us to emulate the process of haloadaptation. In combination with NMR and biophysical characterization, we have managed to study at the level of individual residues the main contributions to halophilic protein stability at extreme salt concentrations. Particularly, we have tackled the unfolded state of halophilic proteins using IDPs as an approach.
- ✓ We confirm, quantify and redefine the role of electrostatic interactions, based on the weak, unspecific electrostatic screening of abundant negatively charged carboxylates, without the requirement for any specific arrangement. However, they are clearly not sufficient to explain haloadaptation.
- ✓ The study of the unfolded state allowed us to unveil a novel contribution arising from enhanced ion-exclusion. Halophilic amino acid sequence increases the polar surface and decreases the hydrophobic content. As a result, more sites for preferential hydration and concomitant preferential ion–exclusion become available.
- ✓ We propose a model, semi-quantitative for ProtL, in which an alternate switching of interactions in the folded (electrostatic screening) and unfolded states (preferential exclusion) synergistically stabilizes halophilic proteins.
- ✓ We apply this model to a case study of complex salt modulation of enzymatic activity: *HvLigN*. We show how differential salt-stabilization of two isoforms with different ASA, part of the reaction coordinate, provoke an apparent complex salt-modulation of enzymatic activity. Additionally, we manage to engineer a variant that recovers catalytic activity in the presence of NaCl.
- ✓ Our model could bring insightful perspectives into the study of osmolyte–protein and ion–protein interactions (i.e. Hofmeister effects). In general, it could be very helpful in the study of protein–water–cosolute ternary system.

THESIS PUBLICATIONS

“Halophilic protein adaptation results from synergistic residue-ion interactions in the folded and the unfolded states”

Ortega G, Diercks T, Millet O.

Submitted 2015.

“Hydration dynamics of a halophilic protein in folded and unfolded states”

Qvist J, **Ortega G**, Tadeo X, Millet O, Halle B.

Journal of Physical Chemistry B. 2012 Mar 15;116(10):3436-44.

“Halophilic enzyme activation induced by salts”

Ortega G, Laín A, Tadeo X, López-Méndez B, Castaño D, Millet O.

Scientific Reports. 2011;1:6.

OTHER PUBLICATIONS

“The mitochondrial intermembrane space oxireductase Mia40 funnels the oxidative folding pathway of the cytochrome c oxidase assembly protein Cox19”

Fraga H, Bech-Serra JJ, Canals F, **Ortega G**, Millet O, Ventura S.

Journal of Biological Chemistry. 2014 Apr 4;289(14):9852-64.

“Protein functional dynamics in multiple timescales as studied by NMR spectroscopy”

Ortega G, Pons M, Millet O. (Review)

Advances in Protein Chemistry and Structural Biology. 2013;92:219-51.

“Carbohydrate affinity for the glucose-galactose binding protein is regulated by allosteric domain motions”

Ortega G, Castaño D, Diercks T, Millet O.

Journal of the American Chemical Society. 2012 Dec 5;134(48):19869-76.

“Uroporphyrinogen III synthase mutations related to congenital erythropoietic porphyria identify a key helix for protein stability”

Fortian A, Castaño D, **Ortega G**, Laín A, Pons M, Millet O.

Biochemistry. 2009 Jan 20;48(2):454-61.

REFERENCES

- Adrover, M., L. Mariño, P. Sanchis, K. Pauwels, Y. Kraan, P. Lebrun, B. Vilanova, F. Muñoz, K. Broersen and J. Donoso (2014). "Mechanistic Insights in Glycation-Induced Protein Aggregation." Biomacromolecules **15**(9): 3449-3462.
- Antón, J., A. Oren, S. Benlloch, F. Rodríguez-Valera, R. Amman and R. Rosselló-Mora (2002). "Salinibacter ruber gen. nov., sp. nov., a novel, extremely halophilic member of the Bacteria from saltern crystallizer ponds." Int. J. Syst. Evol. Microbiol. **52**: 485-491.
- Arakawa, T. and S. N. Timasheff (1982). "Stabilization of protein structure by sugars." Biochemistry **21**(25): 6536-6544.
- Arakawa, T. and S. N. Timasheff (1984). "Mechanism of protein salting in and salting out by divalent cation salts: balance between hydration and salt binding." Biochemistry **23**(25): 5912-5923.
- Arnold, K., L. Bordoli, J. Kopp and T. Schwede (2006). "The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling." Bioinformatics **22**(2): 195-201.
- Auton, M. and D. W. Bolen (2005). "Predicting the energetics of osmolyte-induced protein folding/unfolding." Proc Natl Acad Sci U S A **102**(42): 15065-15068.
- Aziz, E. F., N. Ottosson, S. Eisebitt and W. Eberhardt (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**(12567–70).
- Bai, Y., J. S. Milne, L. Mayne and S. W. Englander (1993). "Primary structure effects on peptide group hydrogen exchange." Proteins **17**(1): 75-86.
- Baker, N. A., D. Sept, S. Joseph, M. J. Holst and J. A. McCammon (2001). "Electrostatics of nanosystems: Application to microtubules and the ribosome." Proceedings of the National Academy of Sciences **98**(18): 10037-10041.
- Baldwin, R. L. (1996). "How Hofmeister ion interactions affect protein stability." Biophys J **71**(4): 2056-2063.
- Baldwin, R. L. (2014). "Dynamic hydration shell restores Kauzmann's 1959 explanation of how the hydrophobic factor drives protein folding." Proceedings of the National Academy of Sciences **111**(36): 13052–13056.
- Batchelor, J. D., A. Olteanu, A. Tripathy and G. J. Pielak (2004). "Impact of protein denaturants and stabilizers on water structure." J Am Chem Soc **126**(7): 1958-1961.
- Becktel, W. J. and J. A. Schellman (1987). "Protein stability curves." Biopolymers **26**(11): 1859-1877.
- Bosshard, H. R., D. N. Marti and I. Jelesarov (2004). "Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings." J. Mol. Recognit. **17**: 1-16.
- Bostrom, M., D. R. Williams, P. R. Stewart and B. W. Ninham (2003). "Hofmeister effects in membrane biology: the role of ionic dispersion potentials." Phys Rev E Stat Nonlin Soft Matter Phys **68**(4 Pt 1): 041902.

Britton, K. L., P. J. Baker, M. Fisher, S. Ruzheinikov, D. J. Gilmour, M. J. Bonete, J. Ferrer, C. Pire, J. Esclapez and D. W. Rice (2006). "Analysis of protein solvent interactions in glucose dehydrogenase from the extreme halophile *Haloferax mediterranei*." Proc Natl Acad Sci U S A **103**(13): 4846-4851.

Brown-Peterson, N. J. and M. L. Salin (1993). "Purification of a catalase-peroxidase from *Halobacterium halobium*: characterization of some unique properties of the halophilic enzyme." J Bacteriol **175**(13): 4197-4202.

Bruschweiler, R., X. Liao and P. E. Wright (1995). "Long-range motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling." Science **268**(5212): 886-889.

Bryngelson, J. D., J. N. Onuchic, N. D. Socci and P. G. Wolynes (1995). "Funnels, Pathways, and the Energy Landscape of Protein Folding: A Synthesis." Proteins: Structure, Function and Genetics **21**: 167-195.

Cacace, M. G., E. M. Landau and J. J. Ramsden (1997). "The Hofmeister series: salt and solvent effects on interfacial phenomena." Q Rev Biophys **30**(3): 241-277.

Castaño, D. and O. Millet (2010). "Backbone chemical shifts assignments of D -allose binding protein in the free form and in complex with D-allose." Biomol NMR Assign: Advanced online publication.

Chen, X., S. C. Flores, S.-M. Lim, J. Zhang, T. Yang, J. Kherb and P. S. Cremer (2010). "Specific Anion Effects on Water Structure Adjacent to Protein Monolayers." Langmuir **26**(21): 16447-16454.

Chen, X., T. Yang, S. Kataoka and P. S. Cremer (2007). "Specific ion effects on interfacial water structure near macromolecules." J Am Chem Soc **129**(40): 12272-12279.

Christian, J. H. and J. A. Waltho (1962). "Solute concentrations within cells of halophilic and non-halophilic bacteria." Biochimica et Biophysica Acta **17**(65): 506-508.

Collins, K. D. and M. W. Washabaugh (1985). "The Hofmeister effect and the behaviour of water at interfaces." Quarterly Review of Biophysics **18**(4): 323-422.

Connolly, M. L. (1993). "The molecular surface package." J Mol Graph **11**: 139-141.

Coquelle, N., R. Talon, D. H. Juers, E. Girard, R. Kahn and D. Madern (2010). "Gradual adaptive changes of a protein facing high salt concentrations." Journal of Molecular Biology **404**: 493-505.

Courtenay, E. S., M. W. Capp and M. T. Record, Jr. (2001). "Thermodynamics of interactions of urea and guanidinium salts with protein surface: relationship between solute effects on protein processes and changes in water-accessible surface area." Protein Sci **10**(12): 2485-2497.

Creamer, T. P., R. Srinivasan and G. D. Rose (1997). "Modeling unfolded states of proteins and peptides. II. Backbone solvent accessibility." Biochemistry **36**(10): 2832-2835.

DasSarma, P., J. A. Coker, V. Huse and S. DasSarma (2010). Halophiles, industrial applications. Encyclopedia of Industrial Biotechnology: Bioprocess, Bioseparation and Cell Technology. M. C. Flickinger, John Wiley & Sons, Inc.

Davis-Searles, P. R., A. J. Saunders, D. A. Erie, D. J. Winzor and G. J. Pielak (2001). "Interpreting the effects of small uncharged solutes on protein-folding equilibria." Annu Rev Biophys Biomol Struct **30**: 271-306.

- De Biasio, A., A. Ibáñez de Opakua, T. N. Cordeiro, M. Villate, N. Merino, N. Sibille, M. Lelli, T. Diercks, P. Bernadó and P. J. Blanco (2014). "p15PAF is an Intrinsically Disordered Protein with Nonrandom Structural Preferences at Sites of Interaction with other Proteins." Biophysical Journal **106**: 1–10.
- Dennis, P. P. and L. C. Shimmin (1997). "Evolutionary divergence and salinity-mediated selection in halophilic archaea." Microbiology and Molecular Biology Reviews **61**(1): 90-104.
- Deole, R., J. Challacombe, J. W. Raiford and W. D. Hoff (2013). "An extremely halophilic Proteobacterium combines a highly acidic proteome with a low cytoplasmic potassium content." Journal of biological chemistry **288**(1): 581-588.
- Dill, K. A. and H. S. Chan (1997). "From Levinthal to pathways to funnels." Nature Structural Biology **4**(1): 10–19.
- Doan-Nguyen, V. and J. P. Loria (2007). "The effects of cosolutes in protein dynamics: the reversal of denaturant-induced protein fluctuations by trimethylamine N-oxide." Protein Sci **16**: 20–29.
- Dym, O., M. Mevarech and J. L. Sussman (1995). "Structural Features That Stabilize Halophilic Malate Dehydrogenase from an Archaeobacterium." Science **267**(5202): 1344-1346.
- Dzubiella, J. (2009). "Salt-specific stability of short and charged alanine-based alpha-helices." J Phys Chem B **113**(52): 16689-16694.
- Ebel, C., L. Costenaro, M. Pascu, P. Faou, B. Kernel, F. Proust-De Martin and G. Zaccai (2002). "Solvent interactions of halophilic malate dehydrogenase." Biochemistry **41**(44): 13234-13244.
- Ebel, C., P. Faou, B. Kernel and G. Zaccai (1999). "Relative role of anions and cations in the stabilization of halophilic malate dehydrogenase." Biochemistry **38**(28): 9039-9047.
- Elazari-Volcani, B. (1936). "Life in the Dead Sea." Nature **138**: 467.
- Englander, S. W. and R. Hiller (2001). "Dynamics and thermodynamics of hyperthermophilic proteins by hydrogen exchange." Methods Enzymol **334**: 342-350.
- Englander, S. W. and L. Mayne (2014). "The nature of protein folding pathways." Proceedings of the National Academy of Sciences **111**: 15873–15880.
- Esclapez, J., C. Pire, V. Bautista, R. M. Martinez-Espinosa, J. Ferrer and M. J. Bonete (2007). "Analysis of acidic surface of *Haloferax mediterranei* glucose dehydrogenase by site-directed mutagenesis." FEBS Lett **581**(5): 837-842.
- Farrow, N. A., R. Muhandiram, A. U. Singer, S. M. Pascal, C. M. Kay, G. Gish, S. E. Shoelson, T. Pawson, J. D. Forman-Kay and L. E. Kay (1994). "Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation." Biochemistry **33**(19): 5984-6003.
- Fayos, R., M. Pons and O. Millet (2005). "On the origin of the thermostabilization of proteins induced by sodium phosphate." J Am Chem Soc **127**(27): 9690-9691.
- Flores, S. C., J. Kherb and P. S. Cremer (2012). "Direct and reverse Hofmeister effects on interfacial water structure." J Phys Chem C **116**(27): 14408–14413.
- Forman-Kay, J. D. and M. Tanja (2013). "From Sequence and Forces to Structure, Function, and Evolution of Intrinsically Disordered Proteins." Structure **21**: 1492–1499.

- Fraczkiewicz, R. and W. J. Braun (1998). "Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules." J Comput Chem **19**: 319-333.
- Frolow, F., M. Harel, J. L. Sussman, M. Mevarech and 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**(5): 452-458.
- Fukuchi, S., K. Yoshimune, M. Wakayama, M. Moriguchi and K. Nishikawa (2003). "Unique amino acid composition of proteins in halophilic bacteria." J Mol Biol **327**(2): 347-357.
- Gajiwala, K. S. and C. Pinko (2004). "Structural rearrangement accompanying NAD⁺ synthesis within a bacterial DNA ligase crystal." Structure **12**(8): 1449-1459.
- Georlette, D., V. Blaise, F. Bouillenne, B. Damien, S. H. Thorbjarnardottir, E. Depiereux, C. Gerday, V. N. Uversky and G. Feller (2004). "Adenylation-dependent conformation and unfolding pathways of the NAD⁺-dependent DNA ligase from the thermophile *Thermus scotoductus*." Biophys J **86**(2): 1089-1104.
- Goto, Y., J. C. Linda and A. L. Fink (1990). "Acid-induced folding of proteins." Proceedings of the National Academy of Sciences **87**: 573-577.
- Goto, Y. and S. Nishikiori (1991). "Role of Electrostatic Repulsion in the Acidic Molten Globule of Cytochrome c." Journal of Molecular Biology **222**: 679-686.
- Gottschalk, M. and B. Halle (2003). "Self-Association of Lysozyme as Seen by Magnetic Relaxation Dispersion." J Phys Chem B **107**: 7914-7922.
- Graziano, G. (2008). "Salting out of methane by sodium chloride: A scaled particle theory study." The Journal of Chemical Physics **129**(084506): 1-10.
- Graziano, G. and A. Merlino (2014). "Molecular bases of protein halotolerance." Biochimica et Biophysica Acta **1844**: 850-858.
- Greene, R. F., Jr. and C. N. Pace (1974). "Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin." J Biol Chem **249**(17): 5388-5393.
- Guinn, E. J., W. S. Kontur, O. V. Tsodikov, I. Shkel and M. T. Record, Jr. (2013). "Probing the protein-folding mechanisms using denaturant and temperature effects on rate constants." Proceedings of the National Academy of Sciences **110**(42): 16784-16789.
- Gurau, M. C., S. M. Lim, E. T. Castellana, F. Albertorio, S. Kataoka and P. S. Cremer (2004). "On the mechanism of the Hofmeister effect." J Am Chem Soc **126**(34): 10522-10523.
- Halle, B. (2004). "Protein hydration dynamics in solution: a critical survey." Phil Trans R Soc Lond **359**: 1207-1224.
- Halle, B., V. P. Denisov and K. Venu (2002). Multinuclear Relaxation Dispersion Studies of Protein Hydration. Biological Magnetic Resonance. N. R. Krishna and L. J. Berliner, Springer. **17**: 419-484.
- Harrison, F. C. and M. E. Kennedy (1922). "The red discoloration of cured codfish." Trans. Roy. Soc. Canad. Sect III **16**: 101-152.
- Henchman, R. H. and J. A. McCammon (2002). "Structural and dynamic properties of water around acetylcholinesterase." Protein Sci **12**: 2080-2090.

Hess, B. and N. F. van der Vegt (2009). "Cation specific binding with protein surface charges." Proc Natl Acad Sci U S A **106**(32): 13296-13300.

Hofmeister, F. (1888). "Zur lehre der Wirkung der Salze. Zweite mittheilung." Arch. Exp. Pathol. Pharmacol. **24**: 247-260.

Holmes, M. L., R. K. Scopes, R. L. Moritz, R. J. Simpson, C. Englert, F. Pfeifer and M. L. Dyall-Smith (1997). "Purification and analysis of an extremely halophilic beta-galactosidase from *Haloferax alicantei*." Biochim Biophys Acta **1337**(2): 276-286.

Hwang, T. L., P. C. van Zijl and S. Mori (1998). "Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme." J Biomol NMR **11**(2): 221-226.

Isom, D. G., C. A. Castañeda, B. R. Cannon and B. García-Moreno E. (2011). "Large shifts in pKa values of lysine residues buried inside a protein." Proceedings of the National Academy of Sciences **108**(13): 5260-5265.

Jeener, J., B. H. Meier, P. Bachmann and R. R. Ernst (1979). "Investigation of exchange processes by two - dimensional NMR spectroscopy." The Journal of Chemical Physics **71**(11): 4546-4553.

Jungwirth, P. and D. J. Tobias (2006). "Specific ion effects at the air/water interface." Chemical Reviews **106**: 1259–1281.

Keeler, J. (2002). Understanding NMR Spectroscopy, Wiley.

Kendrick, B. S., B. S. Chang, T. Arakawa, B. Peterson, T. W. Randolph, M. C. Manning and J. F. Carpenter (1997). "Preferential exclusion of sucrose from recombinant interleukin-1 receptor antagonist: Role in restricted conformational mobility and compaction of native state." Proceedings of the National Academy of Sciences **94**: 11917–11922.

Klebahn, H. (1919). "Die Schädlinge des Klippfisches." Mitt. Inst. Allg. Botanik Hamburg **4**: 11-69.

Klein-Seetharaman, J., M. Oikawa, S. B. Grimshaw, J. Wirmer, E. Duchardt, T. Ueda, T. Imoto, L. J. Smith, C. M. Dobson and H. Schwalbe (2002). "Long-range interactions within a nonnative protein." Science **295**(5560): 1719-1722.

Kloss, E. and D. Barrick (2008). "Thermodynamics, Kinetics, and Salt dependence of Folding of YopM, a Large Leucine-rich Repeat Protein." Journal of Molecular Biology **383**: 1195–1209.

Lanyi, J. K. (1974). "Salt-dependent properties of proteins from extremely halophilic bacteria." Bacteriol Rev **38**(3): 272-290.

Lanyi, J. K. and S. J. (1969). "Effect of salts and inorganic solvents on the activity of the *Halobacterium cutirubrum* catalase." Journal of bacteriology **98**: 611-616.

Lee, J. C. and S. N. Timasheff (1981). "The stabilization of proteins by sucrose." J Biol Chem **256**(14): 7193-7201.

Lee, J. Y., C. Chang, H. K. Song, J. Moon, J. K. Yang, H. K. Kim, S. T. Kwon and S. W. Suh (2000). "Crystal structure of NAD⁺-dependent DNA ligase: modular architecture and functional implications." The EMBO Journal **19**: 1119-1120.

Lipari, G. and A. Szabo (1982a). "Model free approach to the interpretation of nuclear magnetic relaxation in macromolecules: 1. Theory and range of validity." J Am Chem Soc **104**: 4546-4559.

- Lipari, G. and A. Szabo (1982b). "Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results." Journal of the American Chemical Society **104**(17): 4559-4570.
- Liu, J., J. R. Faeder and C. J. Camacho (2009). "Towards a quantitative theory of intrinsically disordered proteins and their function." Proceedings of the National Academy of Sciences **106**(47): 19819-19823.
- Liu, Y. and D. W. Bolen (1995). "The peptide backbone plays a dominant role in protein stabilization by naturally occurring osmolytes." Biochemistry **34**(39): 12884-12891.
- 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 U S A **110**(6): 2135-2139.
- Madern, D., C. Ebel and G. Zaccai (2000). "Halophilic adaptation of enzymes." Extremophiles **4**(2): 91-98.
- Madern, D., C. Pfister and G. Zaccai (1995). "Mutation at a Single Acidic Amino Acid Enhances the Halophilic Behaviour of Malate Dehydrogenase from Haloarcula Marismortui in Physiological Salts." European Journal of Biochemistry **230**(3): 1088-1095.
- Makarov, V. A., B. K. Andrews, P. E. Smith and B. M. Pettitt (2000). "Residence times of water molecules in the hydration sites of myoglobin." Biophysical Journal **79**: 2966-2974.
- Marcus, Y. (2009). "Effect of Ions on the Structure of Water: Structure Making and Breaking." Chemical Reviews **109**(3): 1346-1370.
- Matouschek, A., J. T. Kellis, Jr., L. Serrano and A. R. Fersht (1989). "Mapping the transition state and pathway of protein folding by protein engineering." Nature **340**: 122-126.
- Mattea, C., J. Qvist and B. Halle (2008). "Dynamics at the Protein-Water Interface from 17O Spin Relaxation in Deeply Supercooled Solutions." Biophysical Journal **95**(6): 2951-2963.
- Mayo, S. L. and R. L. Baldwin (1993). "Guanidinium Chloride Induction of Partial Unfolding in Amide Proton Exchange in RNase A." Science **262**(5135): 873-876.
- Meury, J. and M. Kohiyama (1989). "ATP is required for K⁺ active transport in the archaeobacterium Haloferax volcanii." Archives of Microbiology **151**: 530-536.
- Mevarech, M., H. Eisenberg and E. Neumann (1977). "Malate dehydrogenase isolated from extremely halophilic bacteria of the Dead Sea. 1. Purification and molecular characterization." Biochemistry **16**: 3781-3785.
- Mevarech, M., F. Frolow and L. M. Gloss (2000). "Halophilic enzymes: proteins with a grain of salt." Biophys Chem **86**(2-3): 155-164.
- Modig, K., E. Liepinsh, G. Otting and B. Halle (2004). "Dynamics of Protein and Peptide Hydration." Journal of the American Chemical Society **126**: 102-114.
- Mullankhanbhai, M. F. and H. Larsen (1975). "Halobacterium volcanii spec. nov., a Dead Sea halobacterium with a moderate salt requirement." Archives of Microbiology **104**: 207-214.
- Myers, J. K., C. N. Pace and J. M. Scholtz (1995). "Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding." Protein Sci **4**(10): 2138-2148.

- Nihonyanagi, S., S. Yamaguchi and T. Tahara (2014). "Counterion Effect on Interfacial Water at Charged Interfaces and Its Relevance to the Hofmeister Series." Journal of the American Chemical Society **136**(17): 6155–6158.
- O'Neill, J. W., D. E. Kim, D. Baker and K. Y. Zhang (2001). "Structures of the B1 domain of protein L from *Peptostreptococcus magnus* with a tyrosine to tryptophan substitution." Acta Crystallogr D Biol Crystallogr **57**(Pt 4): 480-487.
- Omta, A. W., M. F. Kropman, S. Woutersen and H. J. Bakker (2003). "Negligible effect of ions on the hydrogen-bond structure in liquid water." Science **301**(5631): 347-349.
- Oren, A. (2006). The Order Haloanaerobiales. The Prokaryotes. M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer and E. Stackebrandt, Springer US. **4**: 809-822.
- Oren, A. (2006). The Order Halobacteriales. The Prokaryotes. M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer and E. Stackebrandt, Springer US. **3**: 113-164.
- Ortega, G., A. Laín, X. Tadeo, B. López-Méndez, D. Castaño and O. Millet (2011). "Halophilic enzyme activation induced by salts." Sci Rep **1**: 6.
- Ortega, G., M. Pons and O. Millet (2013). "Protein functional dynamics in multiple timescales as studied by NMR spectroscopy." Adv Protein Chem Struct Biol **92**: 219-251.
- Pais, T. M., P. Lamosa, B. García-Moreno E., D. L. Turner and H. Santos (2009). "Relationship between Protein Stabilization and Protein Rigidity Induced by Mannosylglycerate." Journal of Molecular Biology **394**: 237–250.
- Patel, C. N., S. M. Nobel, G. T. Weatherly, A. Tripathy, D. J. Winzor and G. J. Pielak (2002). "Effects of molecular crowding by saccharides on α -chymotrypsin dimerization." Protein Sci **11**: 997–1003.
- Paul, S., S. K. Bag, S. Das, E. T. Harvill and C. Dutta (2008). "Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes." Genome Biol **9**(4): R70.
- Pegram, L. M. and M. T. Record, Jr. (2007). "Hofmeister salt effects on surface tension arise from partitioning of anions and cations between bulk water and the air-water interface." J Phys Chem B **111**(19): 5411-5417.
- Perez, Y., M. Gairi, M. Pons and P. Bernado (2009). "Structural characterization of the natively unfolded N-terminal domain of human c-Src kinase: insights into the role of phosphorylation of the unique domain." J Mol Biol **391**(1): 136-148.
- Pica, A., I. R. Krauss, I. Castellano, F. La Cara, G. Graziano, F. Sica and A. Merlino (2013). "Effect of NaCl on the conformational stability of the thermophilic γ -glutamyltranspeptidase from *Geobacillus thermodenitrificans*: Implication for globular protein halotolerance." Biochimica et Biophysica Acta **1834**: 149–157.
- Pietila, M. K., P. Laurinmaki, D. A. Russell, K. Ching-Chung, D. Jacobs-Sera, S. J. Butcher, D. H. Bamford and R. W. Hendrix (2013). "Insights into Head-Tails Viruses Infecting Extremely Halophilic Bacteria." Journal of virology **87**(6): 3248-3260.
- Poidevin, L. and S. A. MacNeill (2006). "Biochemical characterisation of LigN, an NAD⁺-dependent DNA ligase from the halophilic euryarchaeon *Haloferax volcanii* that displays maximal in vitro activity at high salt concentrations." BMC Mol Biol **7**: 44.
- Polosina, Y. Y., D. F. Zamyatkin, A. S. Kostyukova, V. V. Filimonov and O. V. Fedorov (2002). "Stability of *Natrialba magadii* NDP kinase: comparisons with other halophilic proteins." Extremophiles **6**(2): 135-142.

- Premkumar, L., H. M. Greenblatt, U. K. Bageshwar, T. Savchenko, I. Gokhman, J. L. Sussman and A. Zamir (2005). "Three-dimensional structure of a halotolerant algal carbonic anhydrase predicts halotolerance of a mammalian homolog." Proc Natl Acad Sci U S A **102**(21): 7493-7498.
- Pundak, S., H. Aloni and H. Eisenberg (1981). "Structure and activity of malate dehydrogenase from the extreme halophilic bacteria of the Dead Sea. 1 & 2." European Journal of Biochemistry **118**: 463-477.
- Pundak, S., H. Aloni and H. Eisenberg (1981). "Structure and activity of malate dehydrogenase from the extreme halophilic bacteria of the Dead Sea. 2. Inactivation, dissociation and unfolding at NaCl concentrations below 2 M. Salt, salt concentration and temperature dependence of enzyme stability." Eur J Biochem **118**(3): 471-477.
- Ramos, C. H. and R. L. Baldwin (2002). "Sulfate anion stabilization of native ribonuclease A both by anion binding and by the Hofmeister effect." Protein Sci **11**(7): 1771-1778.
- Rao, L., X. Zhao, F. Pan, Y. Li, Y. Xue, Y. Ma and J. R. Lu (2009). "Solution behavior and activity of a halophilic esterase under high salt concentration." PLoS One **4**(9): e6980.
- Rasmussen, D. H. and A. P. MacKenzie (1973). "Clustering in supercooled water." J. Chem. Phys. **59**: 5003-5013.
- Robic, S., M. Guzmán-Casado, J. M. Sánchez-Ruiz and S. Marqusee (2003). "Role of the residual structure in the unfolded state of a thermophilic protein." Proceedings of the National Academy of Sciences **100**(20): 11345-11349.
- Rothmund, S., H. Weisshoff, M. Beyermann, E. Krause, M. Bienert, C. Mügge, B. D. Sykes and F. D. Sönnichsen (1996). "Temperature coefficients of amide proton NMR resonance frequencies in trifluoroethanol: A monitor of intramolecular hydrogen bonds in helical peptides?" Journal of Biomolecular NMR **8**: 93-97.
- Rule, G. S. and T. K. Hitchens (2006). Fundamentals of Protein NMR Spectroscopy, Springer.
- Sali, A., E. Shakhnovich and M. Karplus (1994). "How does a protein fold?" Nature **369**(248-51).
- Santoro, M. M. and D. W. Bolen (1988). "Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants." Biochemistry **27**(21): 8063-8068.
- Scalley, M. L., Q. Yi, H. Gu, A. McCormack, J. R. Yates, 3rd and D. Baker (1997). "Kinetics of folding of the IgG binding domain of peptostreptococcal protein L." Biochemistry **36**(11): 3373-3382.
- Schellman, J. A. (1990). "A simple model for solvation in mixed solvents. Applications to the stabilization and destabilization of macromolecular structures." Biophys Chem **37**(1-3): 121-140.
- Schellman, J. A. (2002). "Fifty years of solvent denaturation." Biophys Chem **96**(2-3): 91-101.
- Schellman, J. A. (2003). "Protein stability in mixed solvents: a balance of contact interaction and excluded volume." Biophys J **85**(1): 108-125.
- Schlesinger, A. P., Y. Wang, X. Tadeo, O. Millet and G. J. Pielak (2011). "Macromolecular crowding fails to fold a globular protein in cells." J Am Chem Soc **133**(21): 8082-8085.

- Siglioccolo, A., A. Paiardini, M. Piscitelli and S. Pascarella (2011). "Structural adaptation of extreme halophilic proteins through decrease of conserved hydrophobic contact surface." BMC Struct Biol **11**: 50.
- Skinner, J. J., Y. Wookyung, E. K. Gichana, M. C. Baxa, J. R. Hinshaw, K. F. Freed and T. R. Sosnick (2014). "Benchmarking all-atom simulations using hydrogen exchange." Proceedings of the National Academy of Sciences: 6.
- Sosnick, T. R. and D. Barrick (2011). "The folding of single domain proteins – have we reached a consensus?" Current Opinion in Structural Biology **21**(Folding and Binding): 12–24.
- Street, T. O., D. W. Bolen and G. D. Rose (2006). "A molecular mechanism for osmolyte-induced protein stability." Proc Natl Acad Sci U S A **103**(38): 13997-14002.
- Tadeo, X., B. López-Méndez, D. Castaño, T. Trigueros and O. Millet (2009a). "Protein stabilization and the Hofmeister effect: the role of hydrophobic solvation." Biophys J **97**(9): 2595-2603.
- Tadeo, X., B. López-Méndez, T. Trigueros, A. Laín, D. Castaño and O. Millet (2009b). "Structural basis for the amino acid composition of proteins from halophilic archaea." PLoS Biol **7**(12): e1000257.
- Tadeo, X., M. Pons and O. Millet (2007). "Influence of the Hofmeister anions on protein stability as studied by thermal denaturation and chemical shift perturbation." Biochemistry **46**(3): 917-923.
- Tanford, C. (1970). "Protein denaturation. C. Theoretical models for the mechanism of denaturation." Adv Protein Chem **24**: 1-95.
- Taupin, C. M., M. Hartlein and R. Leberman (1997). "Seryl-tRNA synthetase from the extreme halophile *Haloarcula marismortui*—isolation, characterization and sequencing of the gene and its expression in *Escherichia coli*." Eur J Biochem **243**(1-2): 141-150.
- Timasheff, S. N. (2002). "Protein hydration, thermodynamic binding, and preferential hydration." Biochemistry **41**(46): 13473-13482.
- Timson, D. J. and D. B. Wigley (1999). "Functional domains of an NAD⁺-dependent DNA ligase." J Mol Biol **285**(1): 73-83.
- Tomkinson, A. E., S. Vijayakumar, J. M. Pascal and T. Ellenberger (2006). "DNA ligases: structure, reaction mechanism, and function." Chem Rev **106**(2): 687-699.
- Toth, M., C. Smith, H. Frase, S. Mobashery and S. Vakulenko (2010). "An antibiotic-resistance enzyme from a deep-sea bacterium." J Am Chem Soc **132**(2): 816-823.
- Uejio, J. S., C. P. Schwartz and J. Saykally (2008). "Characterization of selective binding of alkali cations with carboxylate by X-ray absorption spectroscopy." Proceedings of the National Academy of Sciences **105**(6809–12).
- Vrbka, L., J. Vondrasek, B. Jagoda-Cwiklik, R. Vacha and P. Jungwirth (2006). "Quantification and rationalization of the higher affinity of sodium over potassium to protein surfaces." Proc Natl Acad Sci U S A **103**(42): 15440-15444.
- W.B., M., R. D. Cohen, A. E. Smith, E. Guzman–Cisneros and G. J. Pielak (2015). "Quinary structure modulates protein stability in cells." Proceedings of the National Academy of Sciences **112**(6): 1739–1742.
- Wang, A., A. D. Robertson and D. W. Bolen (1995). "Effects of a naturally occurring compatible osmolyte on the internal dynamics of ribonuclease A." Biochemistry **34**(46): 15096-15104.

- Wikstrom, M., T. Drakenberg, S. Forsen, U. Sjobring and L. Bjorck (1994). "Three-dimensional solution structure of an immunoglobulin light chain-binding domain of protein L. Comparison with the IgG-binding domains of protein G." Biochemistry **33**(47): 14011-14017.
- Wojtas, M. N., M. Mogni, O. Millet, S. D. Bell and N. G. Abrescia (2012). "Structural and functional analyses of the interaction of archaeal RNA polymerase with DNA." Nucleic Acids Res **40**(19): 9941-9952.
- Wright, D. B., D. D. Banks, J. R. Lohman, J. L. Hilsenbeck and L. M. Gloss (2002). "The effect of salts on the activity and stability of Escherichia coli and Haloferax volcanii dihydrofolate reductases." J Mol Biol **323**(2): 327-344.
- Wyman, J., Jr. (1964). "Linked Functions and Reciprocal Effects in Hemoglobin: a Second Look." Adv Protein Chem **19**: 223-286.
- Xie, G. and S. G. Timasheff (1997). "Mechanism of the stabilization of ribonuclease A by sorbitol: Preferential hydration is greater for the denatured than for the native protein." Protein Sci **6**: 211-221.
- Yancey, P. H. (2005). "Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses." Journal of Experimental Biology **208**(15): 2819-2830.
- Yancey, P. H., M. E. Clark, S. C. Hand, R. D. Bowlus and G. N. Somero (1982). "Living with water stress: evolution of osmolyte systems." Science **217**: 1214-1222.
- Yi, Q. and D. Baker (1996). "Direct evidence for a two-state protein unfolding transition from hydrogen-deuterium exchange, mass spectrometry, and NMR." Protein Sci **5**(6): 1060-1066.
- Yi, Q., M. L. Scalley, K. T. Simons, S. T. Gladwin and D. Baker (1997). "Characterization of the free energy spectrum of peptostreptococcal protein L." Fold Des **2**(5): 271-280.
- Zaccai, G., F. Cendrin, Y. Haik, N. Borochoy and H. Eisenberg (1989). "Stabilization of halophilic malate dehydrogenase." J Mol Biol **208**(3): 491-500.
- Zhang, Y. and P. S. Cremer (2006). "Interactions between macromolecules and ions: The Hofmeister series." Curr Opin Chem Biol **10**(6): 658-663.
- Zhang, Y., S. Furyk, D. E. Bergbreiter and P. S. Cremer (2005). "Specific ion effects on the water solubility of macromolecules: PNIPAM and the Hofmeister series." J Am Chem Soc **127**(41): 14505-14510.
- Zhao, A., F. C. Gray and S. A. MacNeill (2006). "ATP- and NAD⁺-dependent DNA ligases share an essential function in the halophilic archaeon Haloferax volcanii." Mol Microbiol **59**(3): 743-752.

ANEXO

β LAC

MGSEDLKKLE	EEFDVRLGVY	AIDTGADKEI	SYRENERFAY	TSTFKPLAVG	AVLQTKSDEE	60
LEETITYSEE	DLVTYSPITE	QHVDEGMTLV	EIADAAIRYS	DNTAGNLLLE	AMGGPDELET	120
ILRDIGDETI	EMDRYETELN	EAKPGDIRDT	STAKAMATTL	QQYVLEDVLD	ADRREVLTNM	180
LINNTTGDAL	IRAGVPDGWT	VGDKTGAGGY	GTRNDIGLIW	PEGDEEPIVI	AIMSSRDEED	240
ADYDDKLIK	ATEIVLQELR	N				261

Residue	Number Found
A Ala Alanine	21
R Arg Arginine	13
N Asn Asparagine	9
D Asp Aspartate	28
Q Gln Glutamine	5
E Glu Glutamate	34
G Gly Glycine	20
H His Histidine	1
I Ile Isoleucine	20
L Leu Leucine	23
K Lys Lysine	10
M Met Methionine	7
F Phe Phenylalanine	3
P Pro Proline	7
S Ser Serine	10
T Thr Threonine	24
Y Tyr Tyrosine	10
V Val Valine	14
W Trp Tryptophan	2
C Cys Cysteine	0

Isotopically Averaged Molecular Weight = 28975.8418

Estimated pI = 4.20

Estimated charge over pH range

pH	Charge
4.00	6.5
5.00	-25.6
6.00	-36.7
7.00	-38.7
8.00	-39.7
9.00	-41.7
10.00	-50.1

Extinction Coefficient

Extinction coefficient estimated by the method of Gill and von Hippel (*Analytical Biochemistry*, **182**: 319-326, 1989) where lyophilized proteins were used to establish an absorbance curve based on the number of typtophans, tyrosines, and disulfide bonds. Units are in M-1 cm-1.

Wavelength	Molar Extinction w/o Disulfides	Molar Extinction w/ All Disulfides
280	24180	24180

HvLigN

```

MSDADVDAES NPYLRDPPTE FEPAESLSRE AAEGQAALLR EAVREHDHRY YVAADPLVSD 60
AAYDALFSRL VALEDAFDLD TTNSPTNRVG GEPIDALETV EHVAPMLSID QSTDADDLRE 120
FDERVRREVG AVDYVCEPKF DGLSVEVVYE DGEFVRAATR GDGRRGDDVS AQVKTIPTVP 180
LSLRGDHPDR LAVRGEIYMP KSDFSDLNAR RVEAGEDAFA NPRNAAAGTL RNLDPSSVAD 240
RPLAVFFYDI LDASARPDSQ WAALDRLREW GLRVTDRIER AEDVAEAIDY RDRMQAARDD 300
LDYEIDGTVI KVDSRDARER LGEKSRSVRW AFAYKFPARH EVTTVRDIVV QVGRTRGLTP 360
VAILDVVDVG GVTVSRTLH NPDERAALGV AVGDRVRVCR AGDVIPQVVE VTEDGGGCYE 420
FPDECPVCGS AVDRDGPLAF CSGGLSCPAQ REASIGHFAV KGAMDIDGLG EERVAQLVDA 480
GLVETVADLY DLTADDLAEEL EGWGETSAEN LVAAVENAKH PSLDSFLVGL SIPEVGEATA 540
RGLAREFGSI EAFPIEADAE EDEFDAFEER LTTVPDVGET VARRVRDFFE NADNRAVIRA 600
LLDRGVDPEP VESGGDELDT LTFVVTGTLA ASRSDVTELV ESHGGNVTGS VSGNTDYLTV 660
GENPGRSKRD DAEANDVPTL AETEFEALLA ERGVAYPPE 699

```

	Residue	Number Found
A	Ala Alanine	85
R	Arg Arginine	62
N	Asn Asparagine	16
D	Asp Aspartate	77
Q	Gln Glutamine	9
E	Glu Glutamate	68
G	Gly Glycine	54
H	His Histidine	9
I	Ile Isoleucine	18
L	Leu Leucine	55
K	Lys Lysine	10
M	Met Methionine	5
F	Phe Phenylalanine	24
P	Pro Proline	36
S	Ser Serine	37
T	Thr Threonine	34
Y	Tyr Tyrosine	15
V	Val Valine	75
W	Trp Tryptophan	4
C	Cys Cysteine	6
U	SeC Selenocysteine	0

Isotopically Averaged Molecular Weight = 75811.2266

Estimated pI = 4.50

Estimated charge over pH range

pH	Charge
4.00	39.8
5.00	-35.2
6.00	-62.6
7.00	-70.8
8.00	-74.9
9.00	-80.8
10.00	-92.9

Extinction Coefficient

Extinction coefficient estimated by the method of Gill and von Hippel (*Analytical Biochemistry*, **182**: 319-326, 1989) where lyophilized proteins were used to establish an absorbance curve based on the number of typtophans, tyrosines, and disulfide bonds. Units are in M⁻¹ cm⁻¹.

Wavelength	Molar Extinction w/o Disulfides	Molar Extinction w/ All Disulfides
278	43400	43781
279	42815	43175
280	41960	42320
282	40400	40700

ProtL

MEEVTIKANL IFANGSTQTA EFKGTFEKAT SEAYAYADTL KKDNGEYTVD VADKGYTLNI 60
KFAG 64

Amino Acid Count

	Residue	Number Found
A	Ala Alanine	9
R	Arg Arginine	0
N	Asn Asparagine	4
D	Asp Aspartate	4
Q	Gln Glutamine	1
E	Glu Glutamate	6
G	Gly Glycine	5
H	His Histidine	0
I	Ile Isoleucine	3
L	Leu Leucine	3
K	Lys Lysine	7
M	Met Methionine	1
F	Phe Phenylalanine	4
P	Pro Proline	0
S	Ser Serine	2
T	Thr Threonine	8
Y	Tyr Tyrosine	4
V	Val Valine	3
W	Trp Tryptophan	0
C	Cys Cysteine	0
U	SeC Selenocysteine	0

Isotopically Averaged Molecular Weight = 6991.7310

Estimated pI = 4.77

Estimated charge over pH range

pH	Charge
4.00	4.3
5.00	-1.0
6.00	-2.8
7.00	-3.1
8.00	-3.6
9.00	-4.9
10.00	-9.5

WARNING: pI estimate assumes all residues have pKa values that are equivalent to the isolated residues. For a folded protein this is **not** valid. However, this rough value can be useful for planning protein purifications. pKa values for the individual amino acids from Stryer Biochemistry, 3rd edition.

Extinction Coefficient

Extinction coefficient estimated by the method of Gill and von Hippel (*Analytical Biochemistry*, **182**: 319-326, 1989) where lyophilized proteins were used to establish an absorbance curve based on the number of typtophans, tyrosines, and disulfide bonds. Units are in M⁻¹ cm⁻¹.

Wavelength	Molar Extinction w/o Disulfides	Molar Extinction w/ All Disulfides
278	5600	5600
279	5380	5380
280	5120	5120
282	4800	4800

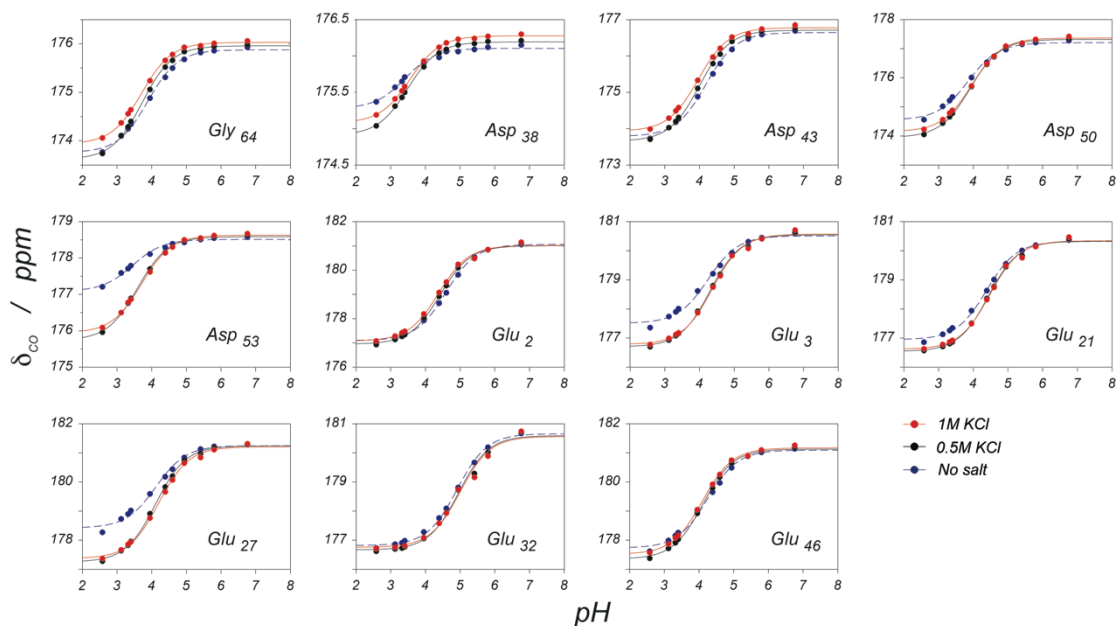


Figure S1. pH titrations for all the acidic residues in natural mesophilic WT ProtL at different salt concentrations (0, 0.5 and 1 M KCl). Circles mark the experimental values, whereas lines represent the best linear fitting to the Henderson–Hasselbach equation (Equation M11).

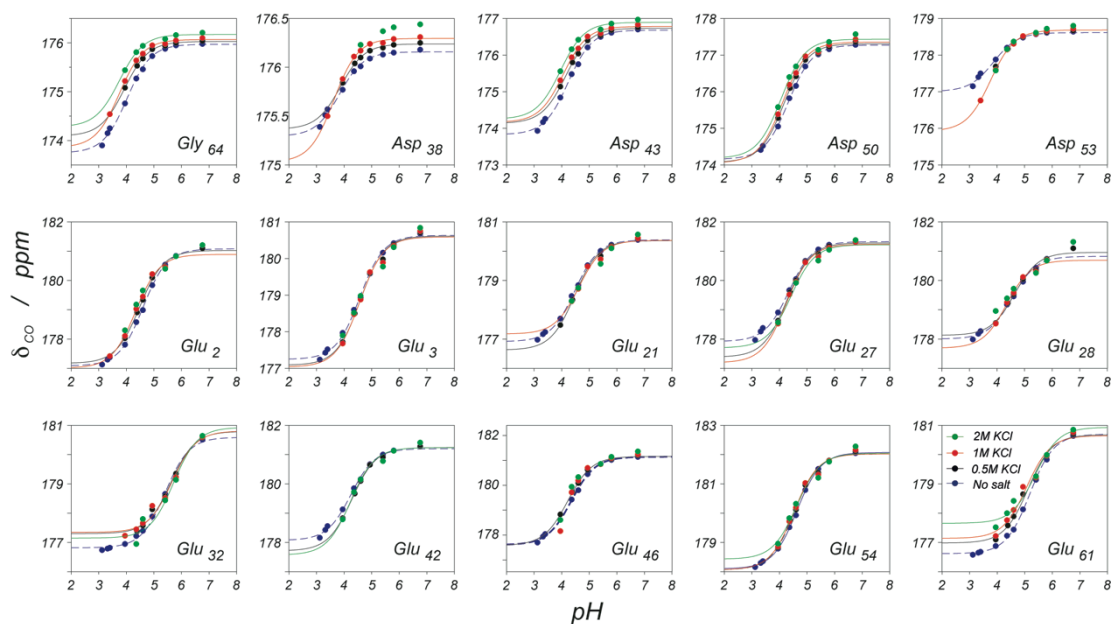


Figure S2. pH titrations for all the carboxylate groups in the designed halotolerant Kx5E ProtL at different salt concentrations (0, 0.5, 1 and 2 M KCl). Circles mark the experimental values, whereas lines represent the best linear fitting to the Henderson–Hasselbach equation (Equation M11).

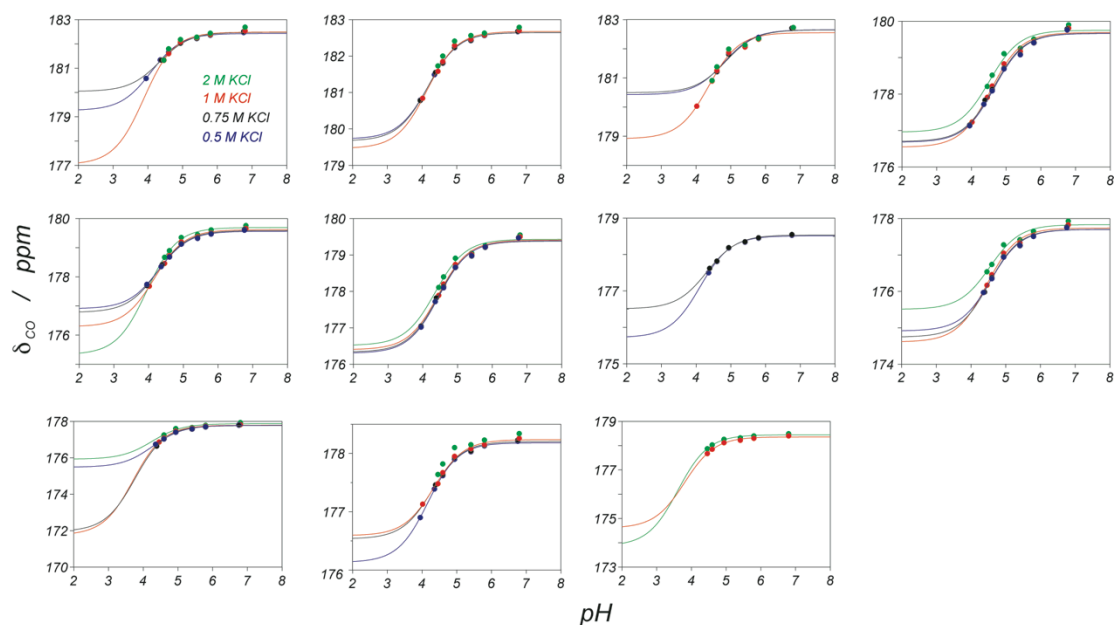


Figure S3. pH titrations for 11 unidentified acidic groups in natural halotolerant β LAC at different salt concentrations (0.5, 0.75, 1 and 2 M KCl). Residue numbers are not available because only α and β -carbons are assigned. Circles mark the experimental measurements and lines represent the best linear fitting to Henderson–Hasselbach expression (Equation M11).

WT ProtL			Kx5E ProtL			β LAC		
Res #	pK _a (0M KCl)	m _{KCl} ^{elec} (i) kcal·mol ⁻¹ ·M ⁻¹	Res #	pK _a (0M KCl)	m _{KCl} ^{elec} (i) kcal·mol ⁻¹ ·M ⁻¹	Res # ¹	pK _a (0M KCl)	m _{KCl} ^{elec} (i) kcal·mol ⁻¹ ·M ⁻¹
2 E	4.59	0.29	2 E	4.61	0.36	1	4.09	0.56
3 E	4.25	-0.16	3 E	4.57	0.01	2	4.19	0.10
21 E	4.38	-0.12	21 E	4.47	-0.20	3	4.81	0.88
27 E	4.12	-0.11	27 E	4.34	0.12	5	4.66	0.13
32 E	4.91	-0.12	28 E	4.55	0.27	6	4.29	0.32
38 D	3.41	-0.19	32 E	5.41	-0.18	7	4.45	0.09
43 D	4.19	0.30	38 D	3.84	0.28	8	4.12	-1.07
46 E	4.24	0.20	42 E	4.25	0.02	9	4.57	0.32
50 D	3.85	-0.15	43 D	4.23	0.19	10	4.28	1.53
53 D	3.45	-0.34	46 E	4.35	0.90	11	4.17	-0.47
64 C _t	3.89	0.24	50 D	4.32	0.16	12	n.a.	0.26
			53 D	3.87	0.08			
			54 E	4.64	0.15			
			61 E	5.19	0.15			
			64 C _t	4.02	0.19			

¹ Arbitrary residues numbers not corresponding to the real sequence

Table S1. pK_a and m_{KCl}^{elec} (i) calculated from experimental pH titrations for aspartic (D) and glutamic (E) acids in ProtL and β LAC. Residue 64 in ProtL corresponds to the C–terminal carboxylic group.

uSRC			haloSRC		
Res # ¹	pK _a (0M KCl)	m _{KCl} ^{elec} (i) kcal·mol ⁻¹ ·M ⁻¹	Res # ¹	pK _a (0M KCl)	m _{KCl} ^{elec} (i) kcal·mol ⁻¹ ·M ⁻¹
1	4.84	-0.09	1	4.89	-0.26
2	4.91	-0.04	2	5.10	-0.02
3	4.77	-0.08	3	4.87	-0.20
4	4.96	0.01	4	4.99	-0.14
5	4.23	0.09	5	5.12	0.07
6	4.03	-0.28	6	5.21	0.11
7	4.56	0.02	7	4.41	-0.24
			8	4.57	-0.11
			9	4.33	-0.23
			10	4.61	0.10
			11	4.66	0.05
			12	4.73	0.16

¹ Arbitrary residues numbers not corresponding to the real sequence

Table S2. pK_a and m_{KCl}^{elec} (i) calculated from experimental pH titrations for ProtL and βLAC.

WT ProtL		
Res #	pK _a (0.5 M KCl)	m _{KCl} ^{elec} (i) kcal·mol ⁻¹ ·M ⁻¹
7	11.84	0.02
23	11.91	0.04
28	11.89	-0.03
41	11.91	0.02
42	11.88	0.03
54	11.87	0.04
61	11.82	0.00

Table S3. pK_a and m_{KCl}^{elec} (i) calculated from experimental pH titrations for the Lysines in ProtL WT.

Abbreviations

ASA	Accesible solvent área
AMP	Adenine mononucleotide
ATP	Adenine triphosphate
β LAC	betha-lactamase from <i>Oceanobacillus iheyensis</i>
AUC	Analytical Ultracentrifugation
CD	Circular Dichorism
CSA	Chemical Shift Anisotropy
CSP	Chemical Shift Perturbation
DMSP	Dymethyl sulphoxide
DNA	Deoxyribonucleic acid
DSC	Differential Scanning Calorimetry
DSE	Denatured State Ensemble
DPF	Dinamyc Perturbation factor (ξ)
EDTA	Ethylen diamine tetraacetate
EXSY	Exchange Spectroscopy
GuHCl	Guanidinium chloride
<i>Hm</i>	<i>Haloarcula marismortui</i>
HSQC	Heteronuclear Single Quantum Correlation
Hv	<i>Haloferax volcanii</i>
IPTG	isopropyl- β -D-thiogalactopyranoside
LigN	NAD ⁺ -dependent DNA Ligase N
LS	Lipari-Szabo
MalDH	Malate dehydrogenase
MRD	Magnetic Relaxation Dispersion
NAD ⁺	Nicotin Adenin Dinucleotide cofactor
NMN	Nicotine mononucleotide
NMR	Nuclear Magnetic Resonanca
NOE	Nuclear Overhauser Effect
ProtL	Protein L
PDB	Protein Data Bank
PIC	Protease Inhibitor Cocktail
PMSF	Phenylmethylsulfonyl fluoride
RMSD	Root mean square deviation
SANS	Small Angle Neutron Scattering
SAXS	Small Angle X-ray Scattering
TAE	Tris-acetate-ethanol buffer
TMAO	Trymethyl ammonium oxide
TMS	Trymethyl Sylil
uSRC	Intrinsically disordered N-terminal domain of the human c-SRC kinase
haloSRC	Halozhilic homologous from uSRC
UV	Ultraviolet radiation

Figures List

- Figure I1. Salinas de Añana.
Figure I2. Halophiles and mesophiles.
Figure I3. Adaptive strategy to avoid osmotic shock.
Figure I4. Halophilic amino acid composition.
Figure I5. Surface negative charge.
Figure I6. The structure is conserved.
Figure I7. Hofmeister series.
- Figure M1. ProtL: Immunoglobulin G binding domain of Protein L from *Peptostreptococcus magnus*.
Figure M2. Structural overlay of β -lactamases
Figure M3. uSRC versus haloSRC sequence.
Figure M4. Structural homology model for *Hv* LigN enzyme.
Figure M5. Adenylation dynamics.
Figure M6. Example of an assay of ligation activity
Figure M7. Size-exclusion chromatography.
Figure M8. Anion-exchange chromatography.
Figure M9. Elliptical polarization after the interaction with a chiral element.
Figure M10. Far UV CD spectra associated with various types of secondary structure.
Figure M11. Linear Extrapolation Method.
Figure M12. Chemical denaturation with guanidinium chloride.
Figure M13. Protein stability curves as a function of temperature.
Figure M14. Thermal denaturation.
Figure M15. Spin precession and energetic diagram.
Figure M16. Magnetic moment oscillation.
Figure M17. Example of a ^1H monodimensional spectrum of a protein.
Figure M18. ^1H , ^{15}N -HSQC spectrum of ProtL.
Figure M19. Schematic diagram of the connectivity established by the 3D-experiments for protein assignments.
Figure M20. Molecular dynamics and NMR experiments.
Figure M21. H/D exchange.
Figure M22. Scheme for the H/D exchange process.
- Scheme R1. Changes in accessible surface area modulate the halophilic character of proteins.
- Figure R1. Global stabilization upon KCl addition.
Figure R2. pH titration for WT ProtL.
Figure R3. Close-up of the pH titration for two residues of Kx5E ProtL.
Figure R4. Stability of individual charged groups calculated from the pK_a values.
Figure R5. Structural analysis of the electrostatic stabilization of ProtL.
Figure R6. Electrostatic contribution of carboxylic groups.
Figure R7. Electrostatic contribution of individual carboxylic groups in the unfolded state.
Figure R8. Lysine electrostatic contribution.
Figure R9. Contribution of electrostatics to global stabilization.
Figure R10. Relaxation-dispersion profiles.
Figure R11. Structural interpretation of MRD data.
Figure R12. Temperature dependence of the dynamic hydration number
Figure R13. Chemical Shift Perturbation coefficients of ProtL.
Figure R14. Kx7E ProtL is a strict halophile.

Figure R15. EXchange Spectroscopy.
Figure R16. KCl modulation of folding and unfolding rates.
Figure R17. uSRC and haloSRC HSQC spectrum.
Figure R18. Model-free analysis of uSRC and haloSRC.
Figure R19. Local dynamics of uSRC and haloSRC.
Figure R20. Chemical shift perturbation coefficients.
Figure R21. KCl slows exchange exponentially.
Figure R22. Rates of KCl-induced exchange decay.
Figure R23. uSRC and haloSRC fast exchange rates measured by CLEANEX.
Figure R24. H/D exchange profile of β LAC.
Figure R25. H/D exchange of ProtL.
Figure R26. KCl dependent stabilization.
Figure R27. Contributions to KCl stabilization of ProtL.

Figure D1. Folded state haloadaptation.
Figure D2. Unfolded state haloadaptation.
Figure D3. Energetics of protein adaptation to hypersaline environments.
Figure D4. Decomposition of ProtL overall KCl stabilization.
Figure D5. Haloadaptation mechanism.
Figure D6. Hydrophobic area and halophilic stabilization.
Figure D7. Chevron plot representation of kinetic EXSY data.
Figure D8. Folding landscape modulation in protein adaptation to hypersaline environments.
Figure D9. An octadecylamine monolayer at the air/salt solution interface.
Figure D10. Folding titration of Kx7E ProtL.
Figure D11. Sucrose stabilization of β LAC.
Figure D12. Sucrose stabilization of ProtL.
Figure D13. Comparison of sucrose and KCl stabilization.
Figure D14. Polar fraction of osmolyte surface correlates with measured Δg_{tr} values.

Tables List

Table I1. Compatible solutes.
Table I2. Halophilic organisms.
Table I3. Intracellular salt concentrations.
Table I4. Comparison between the structures of proteins from halophilic and non-halophilic organisms.

Table M1. Amino acid compositions for all the proteins under consideration.
Table M2. Lysine to glutamate substitutions introduced to WT ProtL
Table M3. Amino acidic content in *uSRC* and *haloSRC*.
Table M4. Summary of protein expression.
Table M5. Composition of M9 minimal media.
Table M6. Anion-exchange *versus* negative charge.

Table R1. Stability and halophilicity.
Table R2. Summary of ^{17}O MRD experiments.
Table R3. Summary of the temperature dependence of the hydration number.

Table D1. KCl and sucrose stabilization.

Equations List

Equation I1. Average relative abundance in halophilic proteins of a given amino acid as compared to mesophilic proteins.

Equation M1. Design of a halophilic sequence composition.

Equation M2. Fitting of CD data to two-states model.

Equation M3. Equation used for fitting of chemical denaturation data to a three-states model.

Equation M4. Total free energy difference upon unfolding of Hv LigN.

Equation M5. Gibbs–Helmholtz equation.

Equation M6. Thermal denaturation analysis.

Equation M7. ΔC_p calculation.

Equation M7. Stabilization estimates from thermal melts.

Equation M9. Simplified expression for the calculation of salt stabilization measured as increments in melting temperature.

Equation M10. Nuclear energetic transitions.

Equation M11. Chemical shift scale in *parts per million* (ppm).

Equation M12. Chemical shift as a probe of protonation states.

Equation M13. Spectral density function proposed by Lipari and Szabo for the model-free analysis.

Equation M14. General expressions for NMR relaxation derived from theory and Solomon equations

Equation M15. Expression for model of correlated local motion along a polypeptide chain.

Equation M16. Thermodynamic analysis of H/D exchange.

Equation M17. Local stability.

Equation M18. Local salt stabilization.

Equation M19. Fast water-exchange rate constant.

Equation M20. Kinetic and thermodynamic analysis of EXSY experiments.

Equation R1. Stabilization calculated from changes in T_m .

Equation R2. Salt stabilization of proteins.

Equation R3. Halophilic character of a sequence.

Equation R4. Changes in the stability of the neutral and charged species in an acid – base equilibrium

Equation R5. Electrostatic potential of interactions.

Equation R6. Electrostatic KCl stabilization.

Equation R7. Electrostatic contribution to native stabilization.

Equation R8. Electrostatic contribution to global stability.

Equation R9. Water ^{17}O longitudinal relaxation rate.

Equation R10. Dynamic hydration number.

Equation R11. Dynamic Perturbation Factor (DPF).

Equation R12. Internal water molecules correlation time τ_c .

Equation R13. CLEANEX measurement of fast exchange.

Equation R14. Contributions to overall KCl stabilization.

Equation D1. Kinetic analysis of KCl stabilization.

*“Caminante, son tus huellas
el camino, y nada más;
Caminante, no hay camino,
se hace camino al andar.*

*Al andar se hace el camino,
y al volver la vista atrás
se ve la senda que nunca
se ha de volver a pisar.*

*Caminante, no hay camino,
sino estelas en la mar.”*

Antonio Machado

Caminante no hay camino