Doctoral thesis

On the effect of cosolutes and crowders in the stability and kinetic properties of proteins

Presented by

Nicanor Zalba Olaizola

Supervised by

Dr. Óscar Millet Aguilar-Galindo

Faculty of Science and Technology
Department of Biochemistry and Molecular Biology
Doctoral Program in Molecular Biology and Biomedicine

Leioa, 2022

Agradecimientos

En todo trabajo científico, y especialmente en la actualidad, es esencial la colaboración. Como no puede ser de otra manera, el presente trabajo es el resultado de la contribución de muchas personas. Estas líneas de agradecimiento van dirigidas a ellas.

- Al Doctor Óscar Millet, quien concibió las ideas que han dado forma a esta tesis, y que ha llevado a cabo un seguimiento constante de la misma que ha hecho posible obtener los excelentes resultados cosechados.

- A los Doctores Borja Mateos, Ganeko Bernardo y Gabriel Ortega por su aportación crucial en la adquisición y análisis de datos, y en la elaboración de los manuscritos publicados y pendientes de publicación, y por su influencia y guía para la redacción de esta memoria.

- Al Doctor Xavi Tadeo, cuya investigación precursora, trabajo y esfuerzo han sido imprescindibles para la consecución de los resultados mostrados en esta memoria.

- A todos los compañeros y amigos del laboratorio de Medicina de Precisión y de otros grupos de investigación del CIC bioGUNE por su apoyo continuado durante el desarrollo de este trabajo y por su influencia y su crítica de los resultados obtenidos.

- A todos mis amigos de Pamplona por su apoyo y su cariño.

- A mi madre, María del Mar, y a mi hermana, Flavia, por haber creído incondicionalmente en mí.

- A Judith, por su paciencia, su apoyo, su influencia y su cariño.

Esta memoria sólo es una recopilación de una enorme cantidad de trabajo, que hubiera sido imposible de haber carecido de la aportación de cualquiera de las personas citadas.

Resumen/Summary

Resumen

Se sabe desde hace mucho que la vida se encuentra en todos los entornos de la Tierra, sin embargo, este hecho resulta sorprendente, puesto que algunos de esos entornos están caracterizados por unas condiciones que parecen incompatibles con la supervivencia. Para evitar la muerte a causa de las condiciones extremas, la vida se ha visto obligada a adaptarse, aunque el mecanismo para lograrlo depende de cada condición específica y, en general, aún no se comprende por completo. A pesar de ello, algo resulta evidente cuando se estudia cómo se han adaptado los organismos que viven en ambientes extremos; que las condiciones extremas obligan a la adaptación del organismo a nivel molecular, forzando a optimizar la estabilidad de las biomoléculas, y en particular la de las proteínas.

Las proteínas desempeñan un gran número de funciones dentro de las células, desde estructural hasta de producción de energía. Debido a que en su gran mayoría las proteínas requieren de estructura tridimensional para poseer función y a que la propiedad que garantiza que preserven dicha estructura es su estabilidad, la pérdida de esta última supone un grave problema para las células. Con el objetivo de adaptar la estabilidad de las proteínas a su rol biológico (proteostasis), la naturaleza ha desarrollado múltiples estrategias, que implican una compleja maquinaria capaz de regular concentración efectiva de las proteínas en la célula. Dicha maquinaria surge de la interconexión entre un amplio sistema de reacciones y procesos muy finamente regulados, y esta sostenida por un nivel de organización estructural basado en interacciones intermoleculares débiles.

Dicho todo esto, buena parte de la homeostasis de una proteína es, en última instancia, función de su composición de aminoácidos. En este contexto, la superficie es la responsable de interaccionar con el medio exterior, incluyendo al resto de biomoléculas. Atendiendo a este hecho fundamental, parece razonable considerar a la superficie proteica como responsable de la sensibilidad al entorno y, en consecuencia, atribuirle un papel de relevancia en los mecanismos de adaptación.

De entre todos los mecanismos de adaptación a entornos extremos el de adaptación a ambientes hipersalinos (haloadaptación) constituye un claro ejemplo de cómo se han remodelado las superficies proteicas con el único objetivo de preservar la estabilidad en presencia de grandes cantidades de sal (KCl y otras sales inorgánicas) y/u otros cosolutos. Este ejemplo adaptativo, puede servir de guía para la construcción de un modelo general de la contribución del entorno a la estabilidad de la proteína (estructura quinaria).

Por otra parte, como parte de este trabajo también se ha considerado relevante estudiar qué influencia tiene el propio entorno sobre algunos procesos de pérdida de la homeostasis de la proteína, en particular sobre los de oligomerización mediante intercambio de dominio. Dichos procesos son de especial relevancia debido a su implicación en algunas funciones celulares principales y por su papel en la aparición de algunas enfermedades.

En resumen, este trabajo está orientado a tratar de extender el mecanismo de haloadaptación y extraer las claves necesarias para constituir un modelo para evaluar la contribución del entorno en la estabilidad de las proteínas (estructura quinaria), así como a estudiar la influencia del propio entorno sobre los procesos de oligomerización mediante intercambio de dominio.

Summary

It is known since long time that life is present in every environment on Earth, however, this fact results surprising, since some of such environments are characterized by conditions that seem to be incompatible with the survival. To avoid death due to extreme conditions, life has been obligated to adapt, although the mechanism to achieve this depends on the specific condition and, in general, it is still not completely understood. Despite this, something results evident when studying how the organisms that thrive in extreme environments have adapted; that the extreme conditions force the adaptation of the organism at a molecular level, forcing to optimize the stability of biomolecules, and in particular that of the proteins.

Proteins perform a great number of functions inside the cells, from structural to production of energy. Because the proteins, for the most part, require a three-dimensional structure to be active, and because the property that guarantees that they preserve such structure is its stability, the loss of the later poses a severe problem for the cells. Aiming to adapt the stability of the proteins to their biological role, Nature has developed multiple strategies that involve a complex machinery capable of modulating the effective concentration of the proteins in the cell. Such machinery emerges from the interconnexion between a wide system of reactions and processes tightly regulated and is sustained by a structural organization level based on weak intermolecular interactions.

Having said all this, great part of the homeostasis of a protein is ultimately a function of its amino acid composition. In this context, the surface is the responsible of interacting with the external medium, including the rest of biomolecules. Attending to this fundamental fact, it seems rationale to consider the protein surface as responsible of the sensitivity to the environment and, in consequence, to attribute it a role of relevance in the adaptation mechanisms.

Out of all the adaptation mechanisms to extreme environments, that of the adaptation to hypersaline environments (haloadaptation) constitutes a clear example of how protein surfaces have been remodeled exclusively to preserve the stability in presence of high amounts of salt (KCl and other inorganic salts) and/or other cosolutes. This adaptive example can serve as guide for building of a general model of the contribution of the environment to the stability of the proteins (quinary structure).

Apart from this, as part of this work, it has also been considered relevant to study what influence poses the own environment on some processes of protein stability loss, particularly the oligomerization reactions through domain swapping. These processes are of special interest due to their implication in some principal cell functions and because of their role in the appearance of several diseases.

In summary, this work is oriented to try to extend the haloadaptation mechanism and to extract the keys required to build a model to evaluate the contribution of the environment to the stability of the proteins (quinary structure), as well as to study the influence of the own environment on the oligomerization processes that occur via domain swapping.

Publications

Mateos, B., Bernardo, G., Dietrich, V., **Zalba, N.**, Ortega, G., Peccati, F., Jiménez-Oses, G., Konrat, R., Tollinger, M., and Millet, O., "Cosolute modulation of protein oligomerization reactions in the homeostatic timescale." *Biophysical Journal* 120.10 (2021): 2067-2077.

Chapter I

Introduction

1. Introduction

1.1. Molecular basis of the adaptation of the organism to the environment

Deserts, glaciers, volcanic landscapes, hydrothermal vents, marine trenches, and salt lakes are examples of environments whose main trait are their physicochemical conditions, considered as nearly incompatible with life. Contrary to what happens at mild environmental conditions, at extreme conditions biomolecules tend to loss their thermodynamic stability (understood as the conservation of the folded conformation) and can no longer sustain the processes of life. As a result, in the past it was assumed that no living-form could be found in such places. Nowadays, however, it is known that life has made its way to colonize most of them. As in any other environment, in extreme environments life has been forced to adapt for surviving (Zierenberg, R., et al. 2000; Siddiqui, K., et al. 2013; Wiegel, J., and Adams, M., 1998, Ma, Y., et al. 2010), although the exact mechanisms that Nature has developed to achieve these adaptations are still not completely understood. As a common trait, it has been observed that the main driving force that guides the adaptation mechanisms is the need to retain the stability of the biomolecules, particularly proteins (Jaenicke, R., and Böhm, G., 1998).

Many detailed studies performed along past decades (Tsukamoto, T., et al. 2016; Cacciapuoti, G., et al. 1994; Chan, C., et al. 2011; Lee, C., et al. 2005; Ortega, G., et al. 2015; Huston, A., and Feller, G., 2008) have deeply investigated how the organisms that thrive in extreme environments manage to preserve the stability of their proteins. The main conclusion extracted from these studies is that the adaptation to each extreme condition demands its own compensation mechanism, although there are some similarities between such mechanisms. As it is pointed by these works, it seems concomitant with each adaptation strategy that the protein stability is preserved by remodeling their amino acid composition. However, each adaptation is effectively implemented in a different way and shows a different pattern in each case.

The stress that menaces extremophile organisms commonly arises from physical or chemical agents, that is, temperature, pressure, or environmental composition among others. These agents shape the direction that the adaptation mechanisms can adopt, conditioning the changes that through evolution are introduced in the proteins of the organisms that thrive under such extreme conditions. For example, the proteins from thermophilic organisms show an increased number of disulfide bonds (Cacciapuoti, G., et al. 2012) and salt bridges (Karshikoff, A., and Landstein, L., 2001), probably because the desolvation penalty and the entropic cost of this strategy is reduced at high temperatures (Lee, C., et al. 2005), and the proteins from halophilic organisms show a large amount of acidic amino acids on their surfaces (Ortega, G., et al. 2015) because the carboxylic acids of their side chains can compete with salt ions for being in contact with water molecules, and because negatively charged surfaces prevent aggregation when high amounts of salt are present.

According to these ideas, it seems that it exists a tight relationship between the environmental conditions and the molecular traits that determine the extent of adaptation of a certain species, and that the changes in the environmental conditions unchain protein remodeling processes oriented to preserve such thermodynamic stability.

1.1.1. Proteostasis and protein stability

Life is possible, to a great extent, thanks to proteins, which perform a vast number of biological functions, from structural scaffolding (Downing, K., and Nogales, E., 1998) or movement (Theriot, J., and Mitchison, T., 1991) to chemical reaction catalysis (Fierke, C., et al. 1987) or environment sensing (Lee, J., and Pilch, P., 1994). For most folded and stable proteins, their biological function requires a three-dimensional structure, the stability of which determines their lifespan. Loss-of-protein function due to premature unfolding, or due to unexpected aggregation, can lead to disruption of the biological functions or produce toxic effects that compromise life (Goh, K., et al. 2007; Braun, P., et al. 2008). To avoid this, Nature has developed a complex system of tightly regulated processes (Figure I1), generically called proteostasis network, intended to preserve protein folding (Balch, W., et al. 2017), and consequently to exquisitely tune-up the protein function to its homeostasis. This system is capable of preventing protein unfolding or activating protein degradation in those cases in which the maintenance of the active structure is not possible, and results from a delicate equilibrium established between several processes (Ron, D., and Walter, P., 2007; Werner, E., et al. 1996; Cohen, E., et al. 2006; Morimoto, R., 1998; Bukau, B., et al. 2006). Proteostasis is the mechanisms that regulates whether a protein remains folded or undergoes unfolding, and it depends, in last term, on the thermodynamic stability of proteins. However, an open question is what is the exact relationship between the stability of a given protein and the amino acid composition that folds into its biologically active conformation.

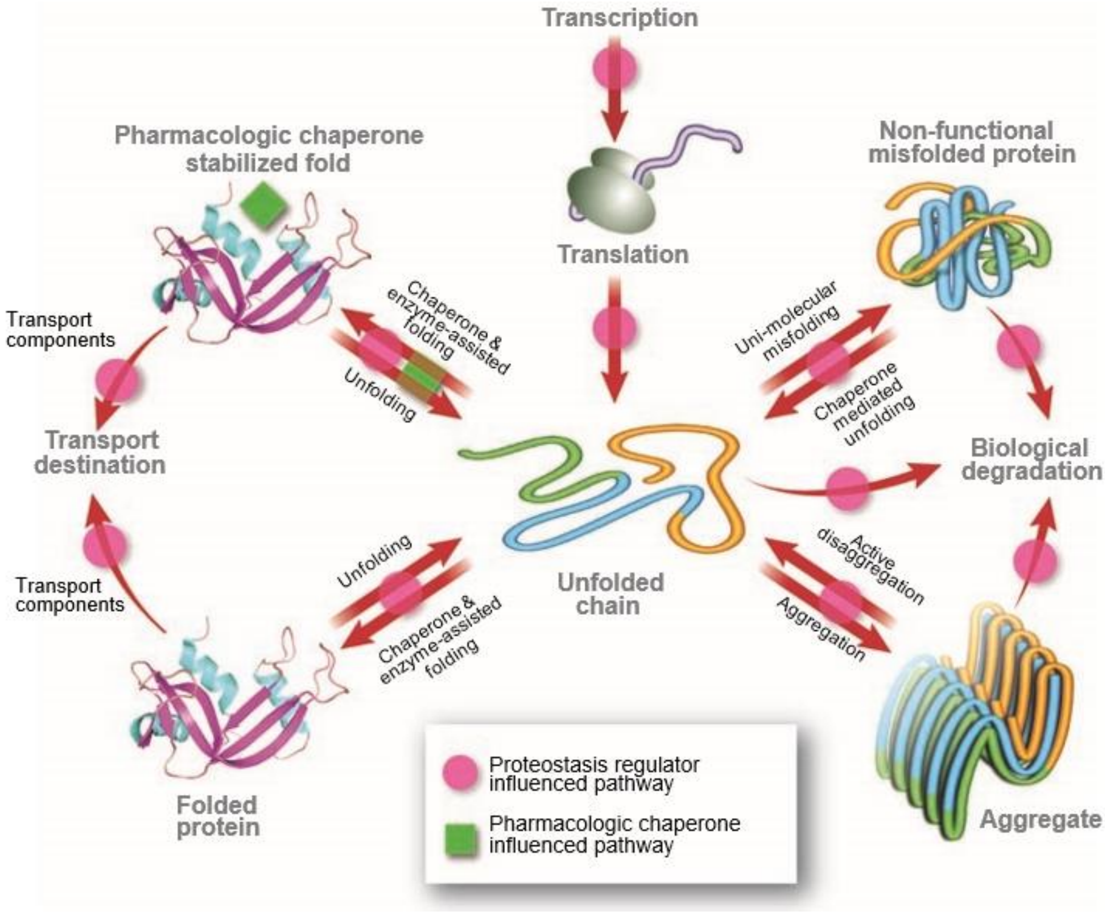


Figure I1. Processes that conform the proteostasis network. From the work of Balch, W., et al. (2007).

As already mentioned, protein stability can be understood as the capability of a protein to preserve its unique biologically active conformation. It was first stated by Anfinsen (Anfinsen, C., 1973) that the three-dimensional structure of a native protein in its normal physiological milieu corresponds to that in which the Gibbs free energy of the protein-solvent system is the lowest. According to this hypothesis, the energy of the native conformation is determined by the totality of atomic interactions, and therefore by the amino acid sequence, in certain environmental conditions. Since the establishment of the so called "Thermodynamic hypothesis" by Anfinsen and coworkers, the contributions from the structure to the Gibbs free energy of folding have been extensively investigated (Dill, K., 1990; Newberry, R., and Raines, R., 2019) to identify the major responsible of protein folding. Nowadays, it is considered that one of the main causes of the folding is the packing of the hydrophobic core (Vlassi, M., et al. 1999; Pace, C., et al. 2011; Ratnaparkhi, G., and Varadarajan, R., 2000) due to hydrophobic interactions and the hydrophobic effect, although the electrostatic interactions stablished between the protein surface and the solution also play an important role.

The formation of the hydrophobic core of a protein is the result of a process known as "hydrophobic collapse". In this process the non-polar side chains of the amino acids that compose the protein get packed together to avoid contact with the solvent. These side chains may be further stabilized by van der Waals interactions. The formation of the hydrophobic core also implies the minimization of the interactions with the solvent and is central for achieving the biologically active conformation (Lins, L., and Brasseur, R., 1995), and consequently nature tends to preserve the composition of the hydrophobic core during evolution (Sikosek, T., Chan, H., 2014).

In turn, the surface of the protein is also partially responsible of the stability, since it is the protein's interface with the solution. Such interaction is stablished through exposed polar groups and charged amino acids, and therefore it depends on the composition of the surface. The electrostatic interactions between the surface and the environment can produce a stabilizing or destabilizing effect (Dill, K., 1990) that can even lead to unfolding. Surface residues will also tend to minimize the hydrophobic solvation. Altogether, protein's stability is very sensitive to the cellular environment through the surface amino acid composition.

Surface residues are very plastic and can be mutated without jeopardizing the overall protein's fold. On the other hand, the conserved composition of the hydrophobic core is the product of costly years of protein's evolution, and it will ensure preservation of the native structure. Thus, it is plausible that, during the adaptation to extreme environments, Nature tends to preferentially remodel the protein surfaces to preserve the biologically active conformation while optimizing stability for a given environment.

1.1.2. The contribution of the protein surface

As has already been pointed, the protein surface is the more exposed part of the protein and it mediates the interaction with the environment, that is, the solvent and the soluble components present in the media. Due to this, it is responsible of a supramolecular organization level.

1.1.2.1. Proteostasis, quinary structure and cellular organization

Many cellular functions, such as intracellular signaling, metabolism or even protein synthesis, depend on the intermolecular interaction between biomolecules and/or with metabolites. For example, dihydrofolate reductase enzyme requires the binding of its substrate and a cofactor molecule for undergoing the conversion of 7,8-dihydrofolate into 5,6,7,8-tetrahydrofolate (Ainavarapu, S., et al. 2005), insulin receptor depends on the binding of insulin to trigger an intracellular response (Ward, C., and Lawrence, M., 2009), the synthesis of proteins requires the interaction between tRNAs, G-proteins and other factors and the ribosome, or some types of calcium channels require binding of metabolites such as inositol 1,4,5-triphosphate to allow calcium release. Consequently, the interior of a cell presents an organization level that overcomes that of the structural organization of its molecular components, and from which the main cellular functions arise. This organization level is referred to as quinary structure and is generally based on weak interactions (McConkey, E., 1982; Guin, D., Gruebele, M., 2019). Proteostasis results from the interaction between the components of the proteostasis network, and therefore it results from the quinary structure.

Furthermore, the interior of the cell is divided into organelles, which enclose different cellular environments and make possible the isolation of different cellular processes. The enclosure of the cellular interior provokes that the available space inside the cell is limited, and because macromolecules are bulky, the interior of the cell is completely crowded. Molecular crowding itself produces an effect on the stability of proteins (Miklos, A., 2011).

1.1.2.2. Solutes, cosolutes and the complexity of the cytosol

As mentioned, the composition of the environment influences the stability and the homeostasis of a protein. In most cases this environment is the cytosol, which is a very complex mixture of biomolecules, membranes, organelles, and metabolites. Moreover, the cytosol changes spatially and with time, evidencing an enormous complexity, almost impossible to grasp. For this reason, the composition of the solvent is often studied by means of reductionist approaches in which certain model compounds are individually tested (Record M., *et al.*, 1978; Arakawa, T., and Timasheff, S., 1985; Auton, M., and Bolen, D., 2005).

Many of such studies have focused on the effect of osmolytes on protein stability. In a simple way, the osmolytes are molecules that affect the osmosis, this is, the flux of the solvent through semipermeable membranes. It has been demonstrated that some compounds such as trimethylamine N-oxide (TMAO) (Ganguly. P., *et al.* 2020), sarcosine (Di Domeico, R., and Lavecchia, R., 2000), trehalose (Baptista, R., *et al.* 2000), taurine (Bruździak, P. *et al.* 2018), sucrose (Lee, J., and Timasheff, S., 1981), glycine (Stadmiller, S., *et al.* 2017) and the salts from Hofmeister series (Baldwin, R., 1996) affect protein stability when they are present in the media. Despite these model compounds show evident differences in their physicochemical properties, their modulating effect of the stability can be explained from common simplified theoretical framework (Schellman, J., 2003) for which preferential interaction and excluded volume effect represent the main contributions.

Preferential interaction refers to the tendence of the protein surface to preferentially be in contact with a certain solute (Figure I2) as a result of a favorable interaction. The opposite mechanism is referred to as preferential hydration, and it occurs when the concentration of the cosolute in the vicinity of the protein is lower than that in the bulk solvent. For preferential interaction to occur water molecules must be removed from the hydration shell of the protein. In consequence, preferential interaction results from the balance between entropic and enthalpic contributions in which both, protein and solvent must be considered. Assuming a favorable interaction between the protein surface and a given molecule, preferential interaction will tend to stabilize the unfolded conformation, because the number of interactions for this state is maximal due to the larger area exposed to the solvent.

Excluded volume (Minton, A., 1980) is an entropic effect that accounts for the fact that two molecules (i.e., protein and cosolute) cannot occupy some space at the same time (Figure I2). In this sense, excluded volume is totally related to the entropy of the protein-solvent system. The volume occupied by the protein and its immediate surroundings are forbidden for that molecule, which in last term leads the protein to adopt its most compact state. Since the unfolded conformation is bulkier than the folded one, excluded volume effect tends to stabilize the folded state.

The overall stability of a protein may be qualitatively understood as the balance between preferential interaction and excluded volume (Schellman, J., 2003).

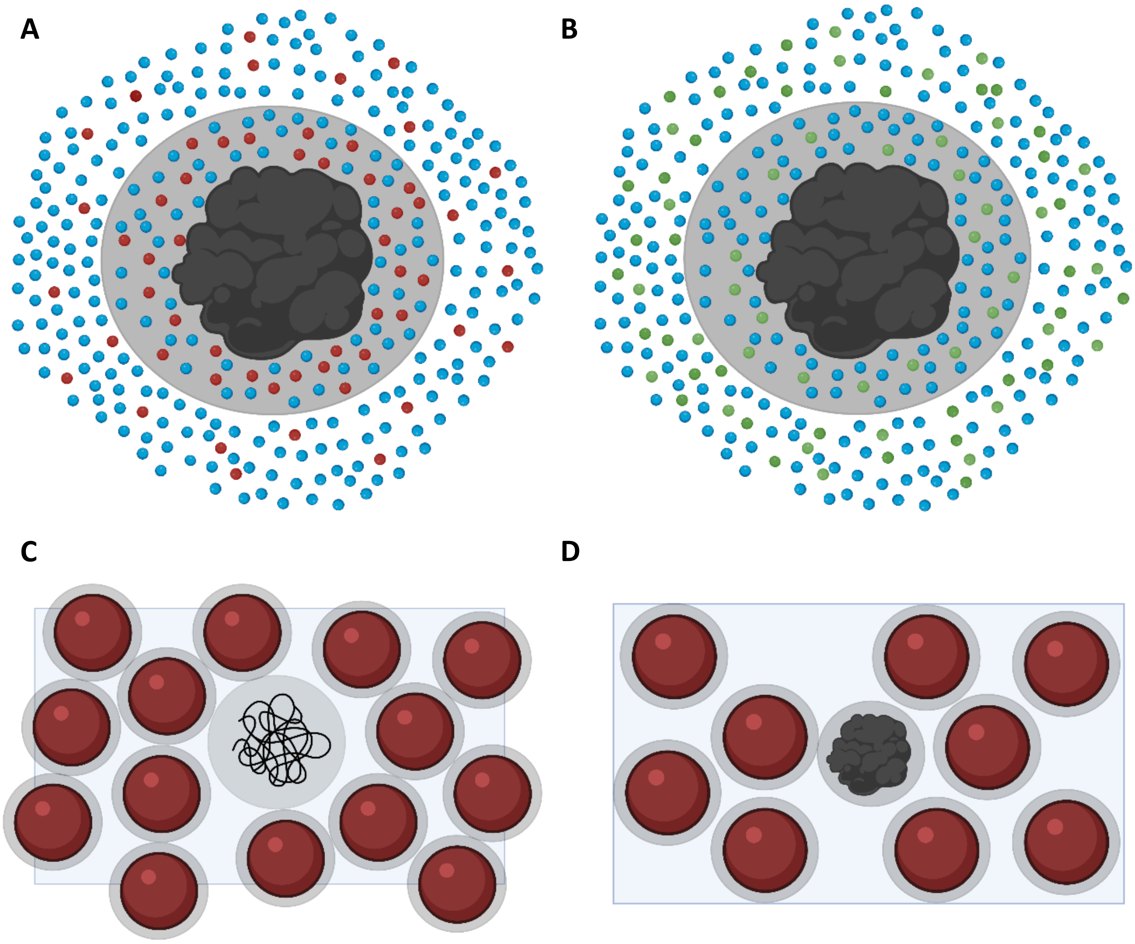


Figure I2. Preferential interaction and excluded volume. (A) Preferential interaction between the protein (black figure) and a solute (red spheres). (B) Preferential exclusion of a solute (green spheres). (C) Excluded volume for the unfolded state of the protein (black chain). (D) Excluded volume for the folded state (black figure). In all cases the grey shaded region represents the vicinity of the corresponding molecule. *Figure created with BioRender.

Despite much has been discovered up to present date, it is still not completely clear how evolution manages to enhance the protein-protein and protein-solvent interactions that are in the base of quinary structure through changes in protein surfaces, neither the exact mechanism that nature has chosen to direct evolution towards an optimal quinary structure. This is likely due to the difficulty in separating the factors that determine the composition of an optimal protein surface, in terms of stability and interaction with the environment, from the rest of the contributions to evolution. Despite this, the natural examples of adaptation to extreme conditions commented at the beginning of this introduction can serve as guides to identify the molecular traits of the general mechanism for the adaptation to the environment.

1.1.3. Haloadaptation of proteins

The adaptation to hypersaline conditions (haloadaptation) is a clear and remarkable example of how evolution has adapted proteins to optimize interactions with their environment. In particular, it is an example in which the composition of the surface has been chosen exclusively for optimized solubility and stability in presence of high amounts of salt in the media (Tadeo, X., *et al.* 2009b). Thus, studying haloadaptation could help understanding some of the general principles for the adaptation to the environment.

1.1.3.1. Strategies for the adaptation to hypersaline environments

The main trait of the hypersaline environments is, as their own name indicates, the high amounts of solved salts that they present (mostly KCl, NaCl and others). Under hypersaline stress water tends to abandon the interior of the cells for equilibrating the osmotic pressure at both sides of the plasma membrane and, in consequence, less water molecules are available for the solvation of proteins, leading to the desiccation of the cell (osmotic shock). To fight this problem, the organisms that live in these places have developed mainly two strategies: to produce small organic molecules, or to raise salt concentration in the interior of the cell.

Commonly, production of small organic molecules is a strategy observed in the organisms that live in environments with salt concentrations up to 1.5 M, it is to say, about 1000 milliosmoles (mOsm) per liter, which is a quantity far above the 300-400 mOsm produced by the solutes found in most cells (K^+, Na^+, Cl^-, etc.) (Ortega, G., 2015). The rationale for this strategy is that the small organic molecules, as well as salt ions, direct the movement of the solvent through the plasma membrane to preserve the osmotic pressure, thus compensating the water stress. Some of the most commonly found cosolutes are saccharides such as sucrose, trehalose or mannose, polyols such as glycerol, sorbitol or mannitol, amino acids such as glycine, taurine or β-alanine, or methylamines such as TMAO, betaine, ectoine or sarcosine (Yancey, P., *et al.* 1982). Surprisingly, the small organic molecules usually found in the organisms that make use of this strategy produce a neutral effect on the native structure of the proteins, reason by which they are also called compatible-solutes or cosolutes. As a result, this strategy avoids the need of modifying all the proteins. Yet, this strategy is feasible only at relatively low salt concentrations due to the energetic cost of producing the own cosolutes.

Alternatively, when salt concentration in the media exceeds 1.5 M, increasing salt concentration of the cytosol shows a clear advantage from the energetic point of view. By means of this alternative strategy some organisms have adapted to survive in environments with salt concentrations between 2 and 5 M. As well as the production of cosolutes, the selective accumulation of K^+ and Cl^- found in halophilic organisms (Christian, J. and Waltho J., 1962) compensates the osmotic stress. However, since the synthesis of cosolutes is no longer required, these organisms avoid changes in their metabolic pathways, which results in a more effective use of the nutrients and energy, thus leading to a simpler adaptive strategy (Ortega, G., 2015).

Despite the apparent simplicity of this last adaptation mechanism, the presence of high amounts of intracellular KCl compromise the native conformation of proteins

(Dennis, P., and Shimmin, L., 1997). In consequence, evolution has reengineered the protein composition of the halophilic organisms towards a composition prepared to be active in these conditions.

1.1.3.2. Molecular features of haloadaptation

An extensive analysis (Paul, S., *et al.* 2008) of the proteins from halophilic organisms showed that the proteins from halophilic organisms present a strongly biased amino acid composition. Specifically, the frequency of acid amino acids such as aspartic and glutamic acids, and polar amino acids such as threonine is increased in detriment of that of bulky and non-polar amino acids such as lysins, as compared to the proteins from the organisms that live in mild environmental conditions (mesophiles) (Figure I3). This characteristic protein composition found in proteins from certain halophilic organisms is usually denominated the halophilic signature, and it has been observed to be exclusive of the organisms that accumulate KCl in their cytoplasm.

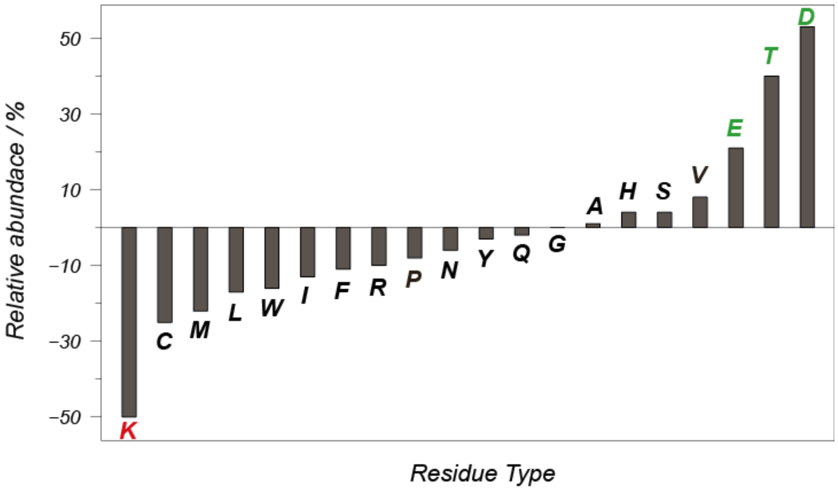


Figure I3. Comparison between the halophilic and mesophilic proteomes. From the work of Paul, S., *et al.* (Paul, S., *et al.* 2008).

These alterations in the protein composition mainly occur on the surface (Fukuchi, S., *et al.* 2003), which further supports the idea that the adaptation to the environment, and in particular haloadaptation, relies on the protein-solvent interactions. According to this, proteins from halophilic organisms show low hydrophobic content and a large negative charge density (Figure I4). Despite the changes in composition, the folding of these proteins is preserved (see Chapter IV), consistent with its key role for the biological function.

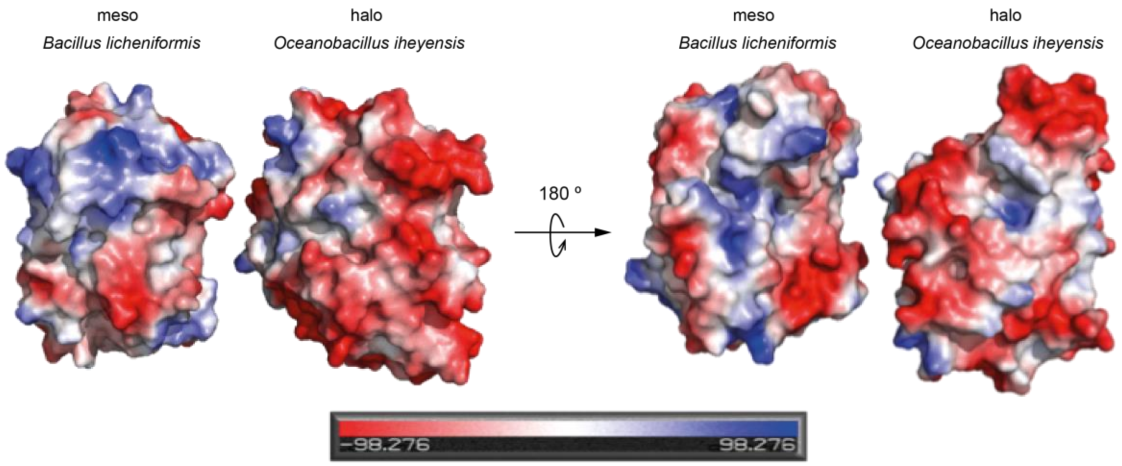


Figure I4. Surface potentials of halophilic (*Oceanobacillus iheyensis*, PDB: 3LEZ) and mesophilic (*Bacillus licheniformis*, PDB: 1NRF) β-lactamases. Negative potential is indicated in red, whereas positive potential is indicated in blue. From the work of Quintanilla (Quintanilla. G., 2015).

As a result of such biased amino acid composition, halophilic proteins become monotonically stabilized by increasing amounts of KCl and, to a lesser extent, NaCl (Polosina, Y., *et al.* 2002). Other ionic species also modulate the stability of halophilic proteins (Ebel, C., *et al.* 1999), following the ranking established in the Hofmeister series (Hofmeister, F., 1888). Furthermore, the stabilizing/destabilizing effect has been observed to be more pronounced on halophilic proteins (Tadeo, X., *et al.* 2009a) and to show a linear dependence on the molar concentration of cosolutes (Tadeo, X., *et al.* 2007). This fact can be expressed saying that proteins from halophilic organisms are more sensitive to their environment.

1.1.3.3. The general mechanism beneath the adaptation to hypersaline conditions

The haloadaptation mechanism has not been completely elucidated despite the extensive workload that has been gathered up to present days.

Nature often works by adapting general mechanisms to specific situations, and the main hypothesis, applied throughout this work, is that halophilic adaptation uses the same underlying mechanisms that determines the interactions of proteins with the environment, that is, the quinary structure. In support of this statement, it stands the argument that the accumulation of salts in the cytosol of halophilic organisms is often accompanied by the accumulation of other osmolytes, supporting the idea of a common shared mechanism of action at the protein level.

It is important to emphasize that haloadaptation is an evolutionary process completely driven by the need of increasing the thermodynamic stability of the entire proteome of the organism without altering the specific folds. It is also a rather simple mechanism that can be adapted to a plethora of proteins simultaneously. Thus, it can be understood as a metric for evolution towards adaptation to the environment, and ultimately as a quantitative metric for the study of quinary structure. This work will explore and develop this idea.

1.1.4. The oligomerization processes in the cellular context

The assembly of the multiproteic complexes that make possible many principal functions of the cell is crucial for such complexes to be functional and, by extension, for the cell to survive. Some examples of important multiproteic complexes are the enzymes of some metabolic reactions (Timm, D., *et al.* 1999), or the machinery that carries out protein degradation (Budenholzer, L., *et al.* 2017) or protein folding maintenance (Braig, K., *et al.* 1994). The assembly processes by which these components get constituted from more simple parts are generically known as oligomerization reactions (Figure I5) and are often the result of the hydrophobic interactions that emerge between the surfaces of the assembled proteins (Ali, M., and Imperiali, B., 2005). As any other reaction that takes place in the interior of the cell, the oligomerization reactions are influenced by the composition of the media (Sluchanko, N., *et al.* 2015). In consequence, changes in the cytosolic composition provoked by situations of cellular stress can trigger undesired oligomerization processes which, in last term, can lead to diseases such as amyloidogenic pathologies (Ellis, R., and Minton, A., 2006; Dobson, C., 2003; Breydo, L., 2014; Wang, W., *et al.* 2010).

Protein oligomerization reactions are of great interest for the study of how the composition of the environment influences quinary structure. Furthermore, this kind of systems are very convenient, since they often involve oligomerization reactions with well-defined architectures paving the way for a rationalization of the effect of model compounds such as cosolutes on the reaction kinetics and thermodynamics.

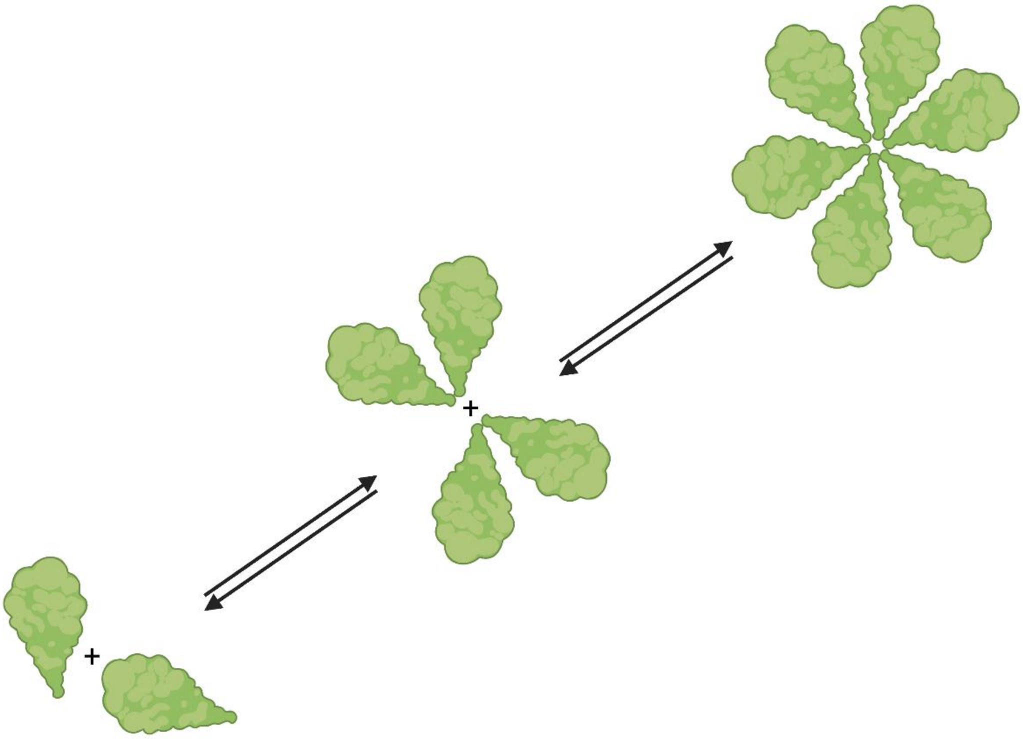


Figure I5. Schematic oligomerization reaction. *Figure created with BioRender.

1.1.4.1. Protein oligomerization through domain swapping

Among all the protein oligomerization processes, those that occur through domain swapping are considered of special relevance, since some important proteins acquire their quaternary structure this way (Hirota, S., 2019). These kind of oligomerization processes are characterized by the exchange of structural elements between identical monomers that get intertwined (Bennett, M., *et al.*, 1995). The common observation that the disassembly of many oligomeric proteins results in irreversible unfolding and aggregation of their constituents has led to think that domain swapping oligomerization reactions occur through the partial unfolding of the monomers. Thus, the domain swapping process is possible thanks to the hydrophobic interactions between the regions that become exposed (Liu, L., *et al.* 2012; Liu, Y., and Eisenberg, D., 2002; Bennet, M., *et al.* 1995) (Figure I6).

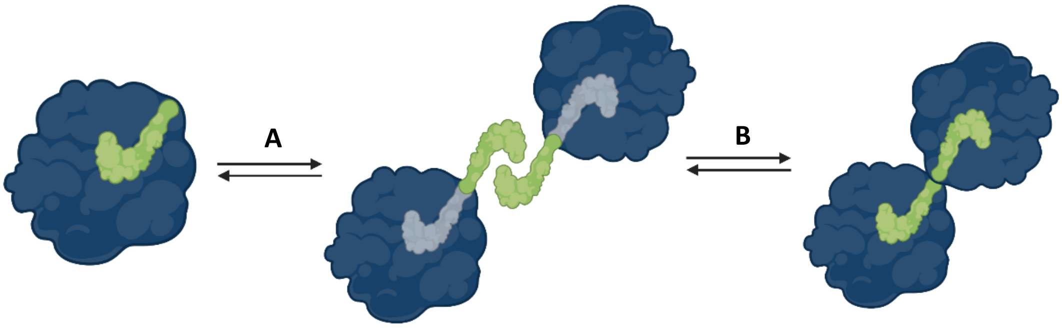


Figure I6. Schematic oligomerization reaction through domain swapping. (A) Each monomer underoes partial unfolding and exhibits a structural element. (B) The exposed structural element of one monomer then substitues that of its oligomerization partner and stablishes hydrophobic interactions that stabilize the oligomeric form. *Figure created with BioRender.

Domain swapping oligomerization reactions have been observed to be responsible of the structure of many common proteins (Janowski, R., *et al.* 2001; Bennett, M., and Eisenberg, D., 1994; Crestefield, A., *et al.* 1962), but the relevance of this particular case of oligomerization reactions comes from the fact that they seem to be implicated in the formation of some cytotoxic species such as amyloids (Perutz, M., 1997; Booth, D., *et al.* 1997).

1.2. Nuclear Magnetic Resonance spectroscopy

Nuclear Magnetic Resonance (NMR) is a spectroscopical technique that exploits the magnetic properties of matter. In its origins at the beginning of past century, when the phenomenon of magnetic resonance was discovered, the NMR spectroscopy was just a curiosity developed in the laboratory interesting only for the physicists who studied the Zeeman effect (Zeeman, P., 1896). Nevertheless, the technical limitations in creating magnetic fields strong enough to produce considerable Zeeman effects leaded physicists to abandon this technique. Despite this, it was early discovered that the resonance frequency of a substance is strongly influenced by its environment, fact that awoke great interest of the chemists on NMR spectroscopy. Thereafter, the NMR experienced an incredible development that allowed its application to the identification of small organic molecules first, and then to the study of biological macromolecules such as DNA or proteins. Nowadays NMR constitutes a reference technique for the study of many biological processes and has successfully implemented a wide variety of experiments that confer it a great versatility.

An in-depth explanation of all the concepts required to fully understand the NMR spectroscopy would require extending this introduction far more than what is reasonable, and therefore only a few basic aspects of the NMR spectroscopy are briefly explained in this work. However, any interested reader could find nice comprehensive descriptions of the technique in the following texts: Abragam, A., 1961; Campbell, I., 2012; Claridge, T., 2009; Doucleff, M., *et al.*, 2011; Ernst, R., *et al.*, 1987; Felli, I., and Pierattelli, R., 2015; Hore, P., *et al.*, 2000; Keeler, J., 2011; Levitt, M., 2001; Neuhaus, D., and Williamson, M., 1956; Rule, G., and Hitchens, T., 2006.

1.2.1. The angular momentum of nuclei

Atomic nuclei possess a property known as nuclear angular momentum, I, that can only adopt values in a discrete way, including integers, semi-integers and zero. I is directly related to the particle composition of the nucleus and is a function of the number of protons and neutrons. Any nuclei with an I-value higher than zero is expected to be sensitive to the magnetic resonance phenomenon, being its simpler manifestation for those nuclei with $I = \frac{1}{2}$. In this latter group fall most of the interesting nuclei for biomolecular NMR, including: ^{1}H, ^{13}C, ^{15}N, and ^{31}P.

1.2.2. The magnetic resonance phenomenon

In the presence of an external magnetic field and for nuclei with $I > 0$, the degenerate energy levels dedifferentiate according to the Zeeman effect. The magnetic moment, μ, can be calculated as shown in the following equation:

$$\mu = \gamma \cdot I$$

Equation I1. Calculation of the nuclear magnetic moment. γ is the gyromagnetic ratio, a proportionality constant that reflects the sensitivity of a given nuclei to the external magnetic field.

According to Equation M1 the higher the gyromagnetic ratio, the greater the magnetic moment. The gyromagnetic ratios of some common nuclei are shown in Table I1. For a nucleus with $I = \frac{1}{2}$ there are two possible spatial orientations for the angular momentum, which are represented by the quantic number m_I ($m_I = \frac{1}{2}$ or $m_I = -\frac{1}{2}$). The energy that corresponds to each orientation can be calculated as follows:

$$E = \mu \cdot B_0 = \gamma \cdot \hbar \cdot m_I \cdot B_0$$

Equation I2. Energy of the magnetic moment at a given magnetic field. ħ is the Planck constant, h, divided by 2π.

Table I1. Gyromagnetic ratios of common nuclei used in NMR.

Nucleus	$\boldsymbol{\gamma\ (10^6\ rad \cdot s^{-1} \cdot T^{-1})}$
^1H	267.522
^2H	41.065
^{13}C	67.283
^{15}N	-27.116
^{17}O	-36.264
^{19}F	251.662
^{31}P	108.291

The difference between both energy levels, expressed in frequency units, is called Larmor frequency, and is described by the following equation:

$$\nu = \frac{\gamma}{2\pi} \cdot B_0$$

Equation I3. Larmor frequency calculation.

An NMR sample will absorb any irradiating radiofrequency radiation with the same energy as that of the difference between the energy levels, altering the population associated to each energy level in a process known as magnetic resonance. NMR measures the return of the populations to the equilibrium, recording the free induction decay (FID) (Figure I7A).

Once acquired, the data from the FID are mathematically manipulated to express them as function of frequency instead of as function of time. This manipulation is known as Fourier transformation (FT). FT peak signals can be of different shape depending on diverse factors (Figure I7B).

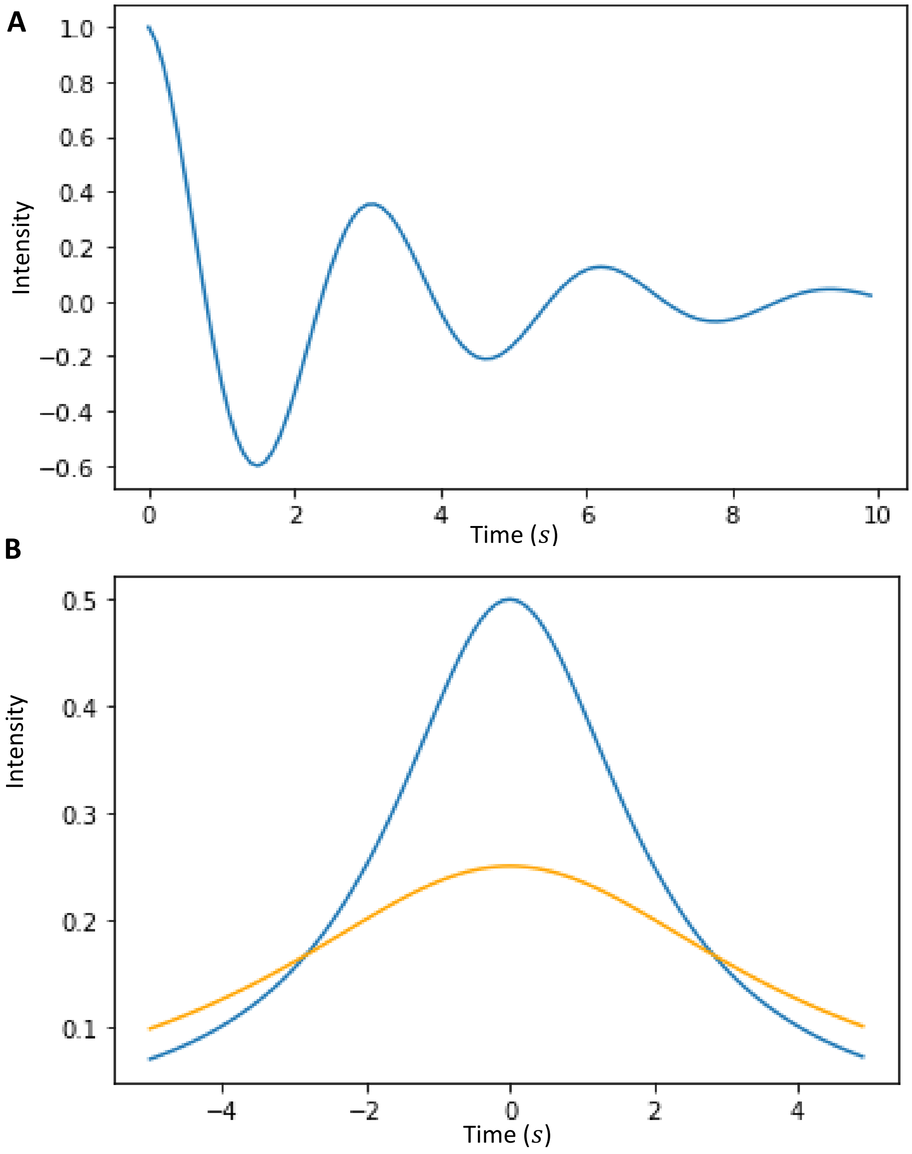


Figure I7. NMR signal. (A) FID signal over time. (B) Signal widening and intensity loss.

1.2.3. The chemical shift scale

The electron cloud partially shield the magnetic resonance effect experienced by a given nucleus, ultimately providing information on the chemical nature of the atom. As a consequence, not all the nuclei of the same species will resonate at the same frequency, but they experience a chemical shift (δ).

$$E = \frac{\gamma}{2\pi} \cdot B_0 \cdot (1 - \delta)$$

Equation I4. Difference between energy levels for the angular momentum with electron shielding.

1.2.4. Scalar coupling and intensity of the spectroscopic signal

Just in the same way that the nuclear magnetic moment interacts with the external magnetic field, when coupled through the electron cloud, two nuclei can interact with each other. As a consequence of this interaction, the signals of the NMR spectra split accordingly. The distance between split bands in the spectra, called the coupling constant, J, which is proportional to the number of bonds existing between the coupled nuclear magnetic moments, and is helpful to identify the distance or the angular orientation between two nuclei in a molecule.

Additionally, the J-coupling also allows the transference of magnetization between the nuclei of bonded atoms. This fact is the basis of the multinuclear experiments, in which the pulses are applied on highly sensitive nuclei, such as ^{1}H, and the magnetization is transferred to less sensitive nuclei (those with low γ values, Table I1), such as ^{15}N or ^{13}C.

1.2.5. The Heteronuclear Single Quantum Coherence experiment

One of the most commonly used two-dimensional NMR experiment is the Heteronuclear Single Quantum Coherence (HSQC)-experiment. It consists of a combination of pulses that induces resonance on a very sensitive nuclear spices (usually ^{1}H) and thereafter drives the transference of the resulting magnetization to a less sensitive nucleus (typically ^{13}C or ^{15}N).

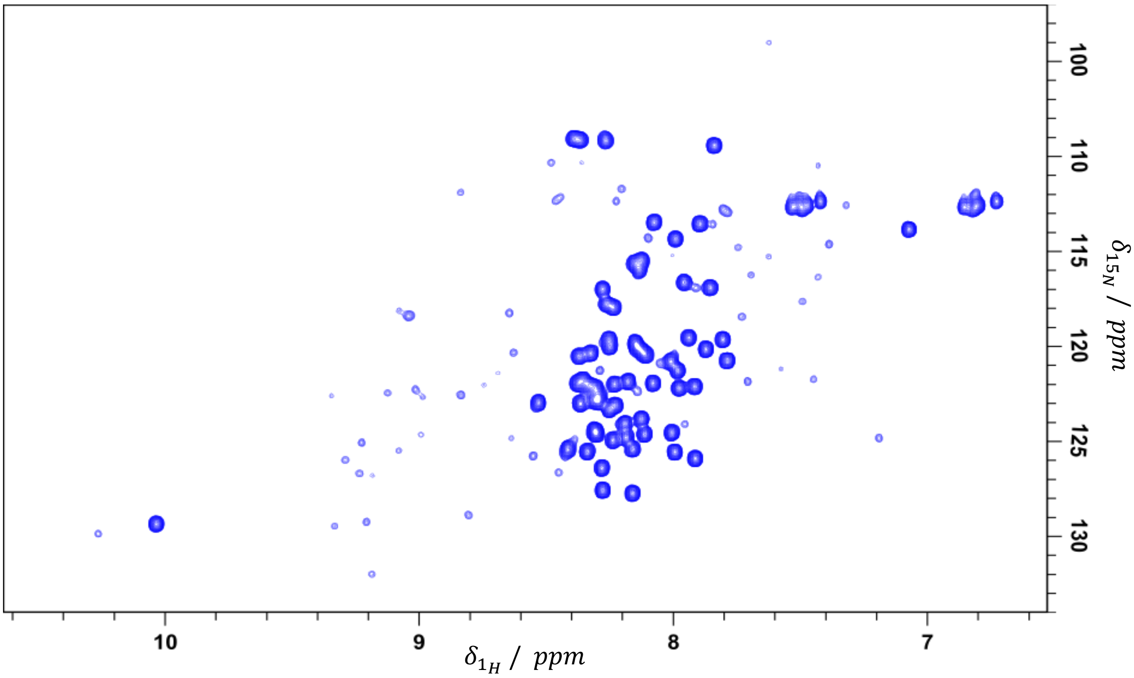


Figure I8. ^{1}H ^{15}N HSQC spectrum of PL_Kx7E. The signals corresponding to the folded and unfolded species appear with different intensity. The signals corresponding to the folded (red) and unfolded (green) forms of glycine 55 are highlighted.

This experiment provides information about the chemical shift and the intensity of the signals and therefore it is ideal for kinetic and thermodynamic analysis. However, signal overlapping sometimes limits its usefulness.

The importance of the HSQC experiment resides in the fact that it allows to monitor the chemical environment at atomic level, by measuring intensity changes and chemical shift of the signals, which is useful for kinetic and structural studies (Moschen, T., and Tollinger, M., 2014).

1.2.6. Experiments for the assignment of spectroscopic signals

In the three-dimensional experiments HNCA, HNCB, HN(CA)CB, HN(CO)CA and HN(COCA)CB the magnetization is transferred from a sensitive nucleus to a less sensitive nucleus (Figure I9). The key point about these experiments is that in HNCA and HNCB experiments the magnetization is transferred from the amide NH group of one amino acid to the amide α/β carbon of the same and the previous amino acids in the chain. Complementarily, in the HN(CO)CA and HN(COCA)CB the magnetization is transferred only to the α/β carbon of the previous amino acid. The information provided by all these experiments can be interpreted in terms of correlations between chemical groups to deduce the order relationship between the aminoacids in the primary sequence. Furthermore, since the chemical shift allows the identification of each type of aminoacid, all this information together makes possible to assign each signal of the HSQC spectrum to the aminoacid producing it.

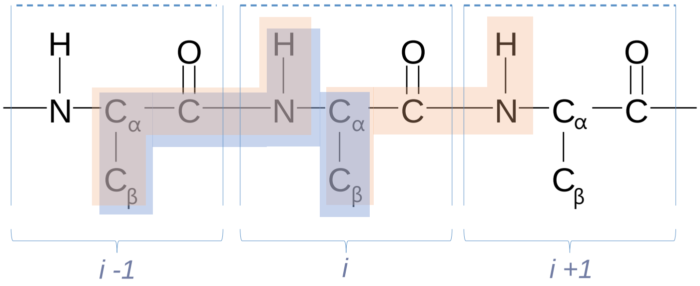


Figure I9. Connection between coupled magnetic moments. C_α and C_β are the alpha and beta carbons respectively. The magnetization is transferred from the proton of the -NH group of the aminoacid *i* to alpha and beta carbons of the aminoacids *i* or *i-1*. Figure adapted from the work of Ortega, G. (Ortega, G., 2015).

1.3. Circular Dichroism

1.3.1. A brief introduction to Circular Dichroism

The origin of Circular Dichroism (CD) relies on how the light interacts with matter, and it be can informative about some physicochemical processes such as protein unfolding. The light can be understood as the propagation of an oscillating electromagnetic field through all directions of the space, and it is known from Maxwell laws that the magnetic field is orthogonal to the electric field in each point of its trajectory. The interaction with matter may induce linear or circular polarization of light. When the light is linearly polarized the electromagnetic wave stops propagating in all directions of the space and start propagating only in one direction, whereas when it is circularly polarized the electric field becomes constant and describes a helix along the direction of propagation (Figure I10).

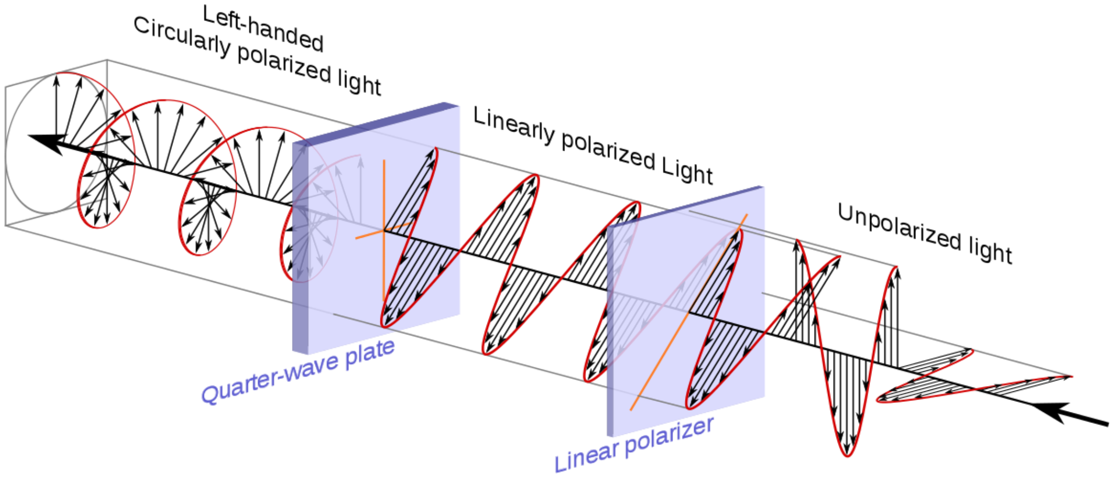


Figure I10. Light linear and circular polarization processes. The electric field oscillates in all directions of the space in unpolarized light (right), only in one direction in linearly polarized light (middle) and becomes constant and describes a helix in circularly polarized light (left).

A linearly polarized light beam can be decomposed into the sum of two circularly polarized light parts, one right-handed and the other left-handed, that correspond to the clockwise or counterclockwise rotation of the electric field in the circularly polarized light. Optically active substances interact with light by absorbing it. If, in addition, they are also chiral, they interact differently with right-handed and left-handed circularly polarized light, and as a consequence when linearly polarized light passes through an optically active sample one of its components is preferentially absorbed over the other and becomes circularly polarized. This distinct interaction can be described by a parameter called *ellipticity* angle, θ, which corresponds to the inverse tangent of the ratio between the difference and addition vectors depicted in Figure I11 following the formula:

$$\theta = tan^{-1}\frac{E_{f,R-L}}{E_{f,R+L}} = tan^{-1}\frac{E_{f,R} - E_{f,L}}{E_{f,R} + E_{f,L}}$$

Equation I5. Calculation of the ellipticity angle θ. $E_{f,R-L}$ and $E_{f,R+L}$ respectively represent the module of the vectors resulting from the difference and the addition of the right and left-handed electric field vectors after interaction with an optically active compound. $E_{f,R}$ and $E_{f,L}$ respectively represent the right and left-handed electric field vectors after interaction with an optically active compound.

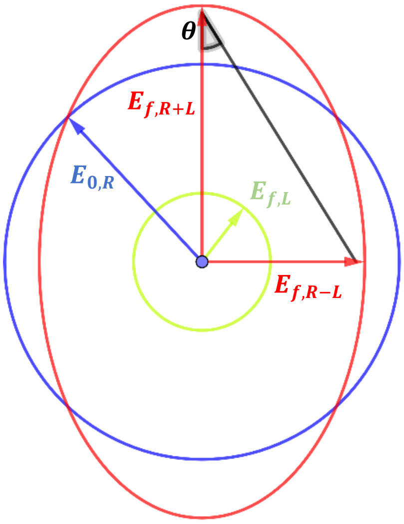


Figure I11. The origin of ellipticity. $E_{0,R}$, (blue), represents the electric field vector of right-handed circularly polarized light before interaction with matter, $E_{f,L}$, (green), represents the electric field vector of left-handed circularly polarized light after interaction with matter, $E_{f,R+L}$ and $E_{f,R-L}$, represent the vectors resulting from addition and difference of right and left-handed electric field vectors after interaction with matter (red).

Like some other spectroscopic properties, the ellipticity depends on the wavelength of the light used to irradiate, and therefore it is possible to obtain CD spectra of chiral molecules. Elements placed in a three-dimensional chiral environment, as well as chiral molecules, also produce CD signal. According to this, secondary and tertiary protein structures such as α-helixes or β-sheets exhibit characteristic spectra (Figure I12). This last fact converts CD in a reference technique for the obtention of qualitative information about the structure of proteins and for monitoring the conformational changes induced by denaturing agents.

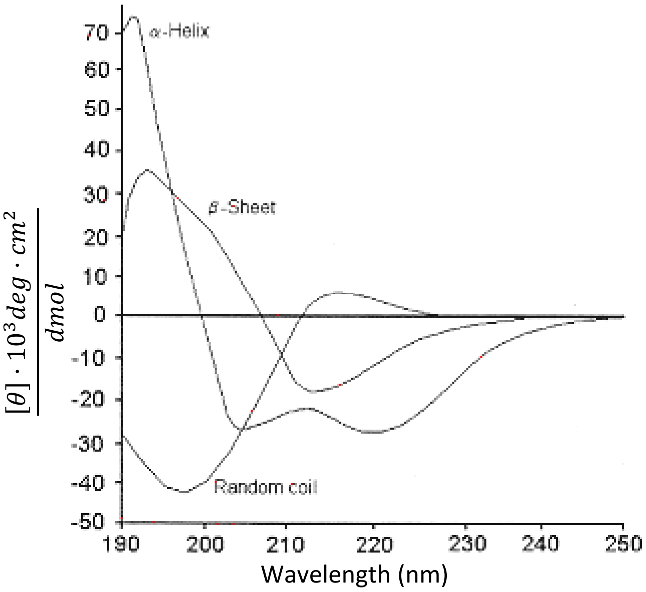


Figure I12. Far ultraviolet CD spectra of α-helix, β-sheet and random coil protein secondary structural elements.

Chapter II

Hypothesis and Objectives

Hypothesis and Objectives

In this work it has been managed the hypothesis that the adaptation mechanism observed in the organisms that thrive in hypersaline environments represents a natural example of a wider and more general mechanism of adaptation to the changes in the composition of the environment that can explain, at least in part, the quinary structure of a protein, understood as the composition of its surface optimized for the interaction with the environment.

This hypothesis will be expanded not only to protein stability, but to other functionally relevant observables. Given the importance of the oligomerization reactions that occur through domain swapping and taking in consideration that some stages of these reactions are highly dependent on the composition of the media, it has also been hypothesized that the effect of the composition of the environment affects asymmetrically the different stages of the oligomerization reactions, thus altering their kinetics and thermodynamics.

The objectives of this work, conceived under the abovementioned hypotheses, are the following:

1. To design a quantitative metric based on the haloadaptation mechanism, that takes into account the composition of the surface, to rationalize the different sensitivity of proteins to a given environment.

2. To test the usefulness of the metric for discriminating between different complex mixtures representing the cytosols of different organisms.

3. To use the metric for rationalizing the effect of a set of small molecules and ions on the stability of proteins.

4. To characterize kinetically and thermodynamically the oligomerization reaction through domain swapping of the mutant G55A from protein L in presence of a set of small organic molecules and ions under different initial conditions.

Chapter III

Material and Methods

3. Material and methods

3.1. Studied Proteins

All the proteins in this work were characterized by NMR spectroscopy or by CD depending on their physicochemical and spectroscopical properties. According to this criterion they are shown in two separate tables (Tables M1-M3).

3.1.1. Proteins studied by Circular Dichroism

Table M1. List of the proteins studied by CD and some of their more relevant properties (part1).

Acronym	**Theoretical pI**	**Mw (kDa)**	**Reference**
PL_WT	4.77	7.0	(O'Neill, J., *et al.*, 2001)
PL_Kx5E	3.65	7.0	(Tadeo, X., *et al.* 2009)
PL_Kx6E	3.37	7.0	(Tadeo, X., *et al.* 2009)
PL_Kx5Q	3.84	7.0	(Tadeo, X., *et al.* 2009)
PL_Ex4D	4.47	6.94	(Tadeo, X., *et al.* 2009)
PL_Ex5D	4.77	6.92	(Tadeo, X., *et al.* 2009)
PL_Ex6D	4.77	6.91	(Tadeo, X., *et al.* 2009)
PL_Kx4E	3.86	7.0	(Tadeo, X., *et al.* 2009)
PL_K42E	4.49	7..0	(Tadeo, X., *et al.* 2009)
PL_Kx2E	4.26	7.0	(Tadeo, X., *et al.* 2009)
PL_Kx3E	4.07	7.0	(Tadeo, X., *et al.* 2009)
PL_DNx8EQ	4.77	7.1	(Tadeo, X., *et al.* 2009)
PL_DNx7EQalt	4.77	7.1	(Tadeo, X., *et al.* 2009)
PL_DNx7EQ	4.77	7.1	(Tadeo, X., *et al.* 2009)
PL_DNx6EQ	4.77	7.1	(Tadeo, X., *et al.* 2009)
PL_DEx5K	9.78	7.1	(Tadeo, X., *et al.* 2009)
PL_DEx6K	9.95	7.1	(Tadeo, X., *et al.* 2009)
PL_Kx7R	4.77	7.1	(Tadeo, X., *et al.* 2009)

Table M2. List of the proteins studied by CD and some of their more relevant properties (part 2).

Acronym	**Theoretical pI**	**Mw (kDa)**	**Reference**
UROIIIS	6.72	31.2	(Fortian, A., *et al.* 2009; Fortian, A., *et al.*, 2011; Mathews, M., *et al.*, 2001)
RBP	6.39	28.7	(Ortega, G., *et al.*, 2012)
YmoA	7.23	8.8	(McFeeters, R., *et al.*, 2007; Cordeiro, T., et al., 2011)
OIH2	4.20	29.0	(Toth, M., *et al.*, 2009)
GGBP	5.49	33.5	(Ortega, G., *et al.*, 2012)
hvCOPG	9.60	5.2	(Gomis-Rüth, F., *et al.*, 1998)
1ALigN	6.60	75.8	(Ortega, G., *et al.* 2011)
btGH99	6.53	43.2	(Sobala, L., *et al.*, 2020; Thompson, A., *et al.*, 2012)
CAM	5.22	16.8	(Bernardo-Seisdedos, G., *et al.*, 2018)
FAH	7.38	49.0	(Macias, I., *et al.*, 2019; Gil-Martinez, J., *et al.*, 2021)
GSTZ1	8.31	24.2	(Macias, I., *et al.*, 2019)
HGD	6.99	50.0	(Macias, I., *et al.*, 2019)
UROD	6.80	40.8	(Urquiza, P., *et al.*, 2018)
FXN	6.52	13.8	Roman E., *et al.* 2012
CPOX	6.79	39.2	(Urquiza, P., *et al.*, 2018)
MBP	7.06	41.5	(Gardner, K., *et al.*, 1998)
GST	7.28	26.1	(Hiller, N., *et al.*, 2006)
Nras	8.69	22.0	(Taparowsky, E., *et al.*, 1983)
ALAD	6.79	36.3	(Urquiza, P., *et al.*, 2018)
RNaseA	7.46	13.7	(delCardayré *et al.*, 1995)

3.1.2. Proteins studied by Nuclear Magnetic Resonance

Table M3. List of the proteins studied by NMR and some of their more relevant properties.

Acronym	Theoretical pI	Mw (kDa)	Reference
PL_G55A	4.77	7.0	(Moschen, T., *et al.* 2014)
PL_Kx7E	2.97	6.98	(Tadeo, X., *et al.* 2009)
drk_WT	4.75	7.94	(Zhang, O., *et al.* 1994)
drk_T12K (drk_m1)	4.91	7.97	(Zhang, O., *et al.* 1994)
drk_T12K_G43D (drk_m2)	4.80	8.02	(Zhang, O., *et al.* 1994)
drk_T12K_T22D_G43D (drk_m3)	4.70	8.04	(Zhang, O., *et al.* 1994)
drk_D42V (drk_m4)	4.87	7.92	(Zhang, O., *et al.* 1994)

3.2. Preparation of samples

The next sections explain the procedures followed for the obtention of the proteins shown in Tables M1-M3. All the overexpression procedures were carried out in either an Innova®44 Incubator Shaker from New Brunswick Scientific or a Certomat® BS-T Incubator Shaker from Sartorius. All the purification procedures were performed in either an ÄKTA Purifier or an ÄKTA Pure, both from General Electric Healthcare (GE Healthcare).

3.2.1. Obtention of surface remodeled proteins

PL mutants were kindly provided by Dr. Tadeo (Tadeo, X., *et al.*, 2009). All the mutants derived from drk were cloned into the corresponding plasmid (see Table M4) between Bgl II – xhoI restriction sites (Figure M1) and their synthesis was ordered to the GenScript company. The cloning vectors, antibiotic resistances, and overexpression conditions of all the proteins under consideration are shown in Table M4.

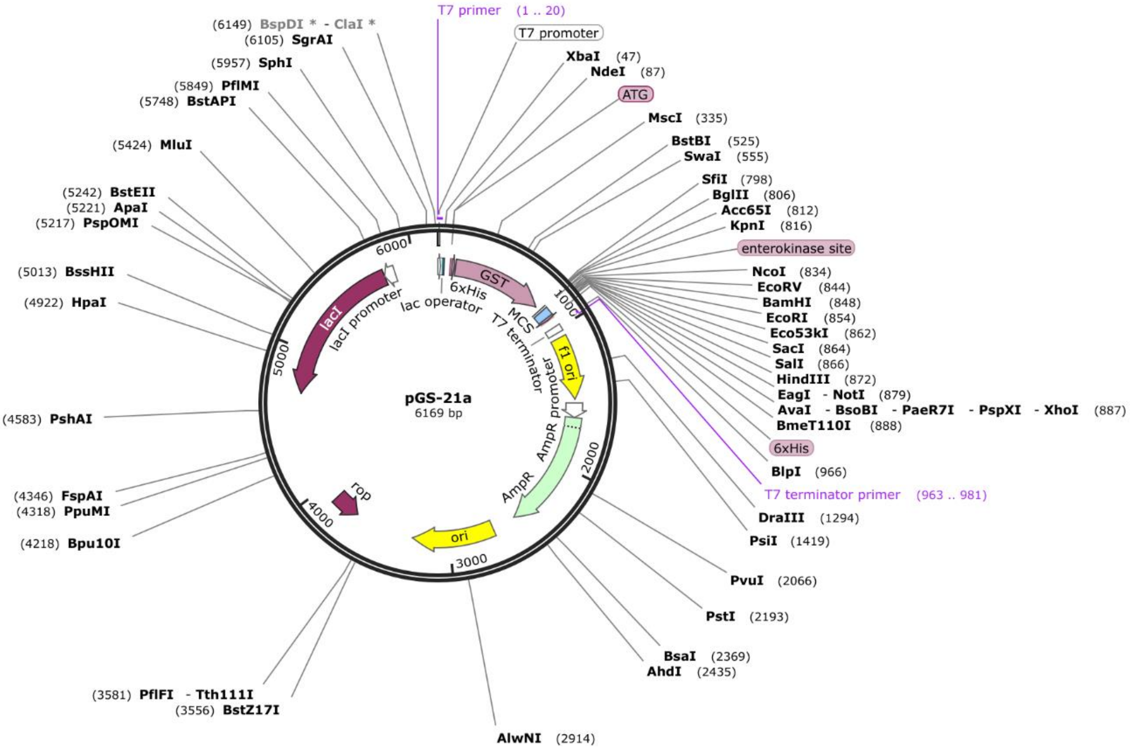


Figure M1. pGS-21a plasmid map. Obtained from SnapGene webpage.

3.2.2. Culture growth and isotopic labelling

All the proteins were overexpressed following the same strategy with only specific overexpression conditions varying from one protein to another. First, the plasmids encoding for the corresponding protein were cloned into the BL21(DE3) *Escherichia coli* expression-enhanced strain (Thermo Fisher Scientific) following the adaptation of the protocol from Stratagene QuickChange Mutagenesis Kit® described below:

1. A vial of 50 mL of BL21(DE3) cells was thawed on ice for 5 min.
2. Around 100 ng of plasmid was added to cells, and then they were then incubated on ice for 30 min.
3. Thermal shock was applied for 1 minute at 42 °C in a Thermomixer Comfort from Eppendorf. Then, cells were immediately incubated on ice for 2 min
4. Cells were incubated at 37 °C for 1 h with gentle shaking.
5. Cells were seeded in a LB (Lysogeny Broth, Pronadisa) agar plate supplemented with antibiotic.
6. Cell plate was cultured at 37 °C overnight in a Function Line incubator from Heraeus.

Transformed colonies were selected over non-transformed by their resistance to the corresponding antibiotic. Next, a single colony of transformed cells was inoculated in 5-10 mL of LB medium and cultured with stirring until turbidity was detected (typically after 4 h). Then, depending on whether isotopic enrichment was required or not, two procedures were applied. For overexpression without isotopic labelling, the following protocol was applied:

1. 100 mL of cell pre-inoculum was added to a 500 mL-flask containing 100 mL of sterile LB media supplemented with corresponding antibiotic and cultured overnight at 37 °C with stirring for each liter of final culture. Typically, an optical density (O.D.) of ~2.0 was achieved after this step.
2. Then, 100 mL of cells were centrifuged at 2000 *g* for 20 min at room temperature (RT), resuspended in 10 mL of sterile LB media, added to a 3 L-Fernbach flask containing 1 L of sterile LB media supplemented with antibiotic and cultured until an O.D. of 0.6 - 0.8 was achieved.
3. Protein overexpression was then induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM and incubating overnight at 20 °C with stirring.
4. Cells were finally centrifuged for 30 min at 4 °C and 5000 *g*, fast-frozen with liquid nitrogen after removing the supernatant and stored at -20 °C until required.

On the other hand, when isotopic enrichment was required, the following procedure was applied instead of the previous one:

1. For each liter of final culture three 500 mL-flasks containing 100 mL of sterile M9 minimal growth media supplemented with antibiotic were inoculated with 20 mL, 40 mL and 100 mL of cells respectively and cultured overnight at 37 °C with stirring. Typically, an O.D. of 0.8-1.8 was achieved after this step.
2. Thereafter, 100 mL of cells were centrifuged at 2000 *g* for 20 min at RT, resuspended in 10 mL of sterile M9 media, then added to a 3 L-Fernbach flask containing 1 L of sterile M9 media supplemented with antibiotic and cultured until optimal O.D. was achieved (see Table M4).
3. Protein overexpression was induced by addition of IPTG to a variable final concentration and incubating at variable conditions depending on the protein (see Table M4).
4. Finally, cells were harvested by centrifugation at 5000 *g* and 4 °C for 30 min, fast-frozen with liquid nitrogen after removing the supernatant and stored at -20 °C until required.

Isotopic enrichment was achieved by substituting either the nitrogen or carbon sources by $^{15}NH_4Cl$ or ^{13}C-D-Glucose respectively in the M9 media from step 2 of the preceding protocol depending on what kind of enrichment was required. In those cases in which full isotopic labelling was required both sources were substituted.

Table M4. List of the overexpression conditions for the all the proteins used in this work.

Acronym	Plasmid	Antibiotic resistance	Optimal O.D.	IPTG induction concentration (mM)	Temperature (ºC)	Time (h)
PL_WT	pET-16b	Ampicillin	0.45 - 0.6	1	37	4
PL_Kx5E	pET-16b		0.45 - 0.6	1	37	4
PL_Kx6E	pET-16b		0.45 - 0.6	1	37	4
PL_Kx5Q	pET-16b		0.45 - 0.6	1	37	4
PL_Ex4D	pET-16b		0.45 - 0.6	1	37	4
PL_Ex5D	pET-16b		0.45 - 0.6	1	37	4
PL_Ex6D	pET-16b		0.45 - 0.6	1	37	4
PL_Kx4E	pET-16b		0.45 - 0.6	1	37	4
PL_K42E	pET-16b		0.45 - 0.6	1	37	4
PL_Kx3E	pET-16b		0.45 - 0.6	1	37	4
PL_DNx8EQ	pET-16b		0.45 - 0.6	1	37	4
PL_DNx7EQalt	pET-16b		0.45 - 0.6	1	37	4
PL_DNx7EQ	pET-16b		0.45 - 0.6	1	37	4
PL_DNx6EQ	pET-16b		0.45 - 0.6	1	37	4
PL_DEx5K	pET-16b		0.45 - 0.6	1	37	4
PL_DEx6K	pET-16b		0.45 - 0.6	1	37	4
PL_Kx7R	pET-16b		0.45 - 0.6	1	37	4
UROIIIS	pET-16		0.7 - 0.8	0.5	20	18
YmoA	pKM974	Kanamycin	0.45 - 0.6	0.5	30	4
OIH2	pET-16b	Ampicillin	0.45 - 0.6	1	25	20
GGBP	pET11d		0.45 - 0.6	0.5	20	20
hvCOPG	pMV158	Tetracycline	0.45 - 0.6	0.5	37	18
1ALigN	pET-47b	Kanamycin	0.45 - 0.6	1	25	20
btGH99	pET28a		0.45 - 0.6	0.5	37	16
CAM	pICB1	Ampicillin	0.45 - 0.6	0.5	20	20
FAH	pET-16b		0.6 - 0.8	0.5	20	18
GSTZ1	pQE30		0.6 - 0.8	0.5	37	18
HGD	pET-16b		0.6 - 0.8	0.5	20	18
UROD	pET-16b		0.6 - 0.7	0.5	20	18
FXN	pGS-21a		0.7 - 0.8	0.5	37	6
CPOX	pET16b		0.7 - 0.8	0.5	20	18
MBP	pKM974	Kanamycin	0.45 - 0.6	0.5	30	4
GST	pGS-21a	Ampicillin	0.7 - 0.8	0.5	37	6
Nras	pET-26b	Kanamycin	0.6 - 0.8	0.5	20	18
ALAD	pET-16b		0.6 - 0.8	0.5	20	18
RNaseA	pET-22b	Ampicillin	0.5 - 0.7	1	37	4
PL_G55A	pET-16b		0.45 - 0.6	1	37	4
PL_Kx7E	pET-16b		0.45 - 0.6	1	37	4
drk_WT	pGS-21a		0.7 - 0.8	0.5	37	6
drk_T12K	pGS-21a		0.7 - 0.8	0.5	37	6
drk_T12K_G43D	pGS-21a		0.7 - 0.8	0.5	37	6
drk_T12K_T22D_G43D	pGS-21a		0.7 - 0.8	0.5	37	6
drk_D42V	pGS-21a		0.7 - 0.8	0.5	37	6

3.2.3. Protein purification

All the proteins in this work were purified following one of the three strategies described below. In all cases, cell pellets from the overexpression step were thawed at RT and protease inhibitor cocktail (PIC) was added in a 1:100 dilution before cell lysis. Cells were then lysed either by sonication with pulses of 10 s and 60% amplitude followed by 20 s of rest or by cell disruption with up to 3 cycles of pressure pulses at 50 kPSI. Cellular debris were separated from the soluble components by ultracentrifugation at 125000 *g* and 4 °C for 30 min and the purification of the proteins of interest from the soluble fraction was performed by different chromatography steps. Protein purity and yield were checked by SDS-PAGE and Coomassie staining assays respectively.

3.2.3.1. Purification by thermal shock

This method was applied to proteins that present a reversible thermal unfolding equilibrium. For each of these proteins the supernatant obtained from cell lysis was heated to 75 °C for 5 min to cause irreversible thermal unfolding, precipitation and/or aggregation on contaminating proteins. Recovery of the soluble folded form of the protein of interest was achieved by incubating at RT for 30 min with gentle stirring. Insoluble or aggregated proteins were separated from the soluble fraction by ultracentrifugation at 125000 *g* for 30 min at 4 °C. The pellet was discarded whereas the soluble fraction was further processed.

Note: further processing consisted of loading the sample to a Superdex75® (GE Healthcare) size exclusion chromatography column. Of note, some proteins were coeluted with DNA after the size exclusion purification step and they were further purified using an additional anion exchange chromatography with a HiTrap Q Sepharose (Cytiva) column followed by another size exclusion chromatography with a Superdex75 column.

3.2.3.2. Purification by histidine tag

This method was applied to the proteins that were fused with a six-histidine tag during their overexpression process. For all of them, the supernatant from cell lysis was incubated in a nickel-nitrilotriacetic (Ni-NTA, Invitrogen) affinity column for 0.5-2 h at 4 °C with gentle stirring. Subsequently the protein was eluted with variable concentrations of imidazole (see able M4) and the elution pool was loaded to a Superdex75® column.

3.2.3.3. Purification by GST-His tag

This method is a variation of the His-tag purification method and was employed for proteins that encoded for a Glutathione S-transferase (GST) protein tag fused with a six-histidine tag (GST-His tag). The supernatant from cell lysis was loaded to a HiTrap TALON crude (GE Healthcare) and eluted with 300 mM imidazole. The fraction pool was dialyzed against the corresponding buffer (see Table M6) to remove the imidazole and at the same time a thrombin cut was induced to get rid of the GST-his tag. Thereafter, the product of the previous step was loaded to a HiTrap TALON crude column to remove the GST-his tag fragments, and the flow-through of this step was collected and loaded to a Superdex75® column.

3.2.3.4. Summary of purification conditions, columns, and buffers

The purification conditions of all the proteins under consideration in this work, including the employed buffers and chromatography columns used are summarized in Tables M5-M7.

Table M5. List of the purification buffers used in this work.

Buffer Code	**Composition**
A	20 mM Sodium phosphate (NaP), pH 6.0
B	20 mM (hydroxymethyl)-1,3-propanediol (TRIS), pH 8.0
C	20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.5
D	A + 0.5 M NaCl + 5 mM imidazole
E	D (500 mM imidazole)
F	B (50 mM TRIS) + 0.12 M NaCl + 2 mM imidazole
G	F (300 mM imidazole)
H	A (50 mM NaP) + 0.3 M NaCl + 5% glycerol + 5 mM imidazole
I	H (500 mM imidazole)
J	B + 0.12 M NaCl + 2 mM imidazole
K	J (300 mM imidazole)
L	C + 0.15 M NaCl + 1 mM $MgCl_2$ + 1 mM TCEP
M	A (pH 8.0) + 2 M KCl

Table M6. List of the chromatographic columns used in this work.

Column code	**Name**	**Type**
1	Superdex 75	Size Exclusion
2	HiTrap Talon Crude	Affinity
3	Ni-NTA	Affinity
4	HiTrapQ Sepharose	Anion Exchange
5	HiTrap SP	Cation Exchange

Table M7. Summary of the purification conditions for all the proteins under consideration in this work. Buffer and column codes are the same as specified in Tables M4 and M5. (a) OIH2 was purified loading the supernatant from cell lysis to a HiTrap Q Sepharose column and then loading the elution of this step to a Superde75 column. (b) RNaseA was lysed in buffer B + 6 M urea + 1 mM EDTA at pH 8.0, then underwent a chemical refolding in B + 6 M urea + 1 mM EDTA + 0.4 M NaCl + 20 mM DTT at pH 8.0, it was dialyzed against buffer B + 0.1 M NaCl at pH 5.0 and purified by both, an anion and a cation exchange chromatography.

Acronym	Lysis method	Purification Method	Column 1	Loading buffer	Elution buffer	Column 2	Loading buffer	Elution buffer	Column 3	Loading/Elution buffer
PL_WT	sonication	thermal shock	1	A	A	-	-	-	-	-
PL_Kx5E	sonication	thermal shock	1	A	A	4	A	A + 1 M KCl	1	A
PL_Kx6E	sonication	thermal shock	1	A	A	4	A	A + 1 M KCl	1	A
PL_Kx5Q	sonication	thermal shock	1	A	A	-	-	-	-	-
PL_Ex4D	sonication	thermal shock	1	A	A	-	-	-	-	-
PL_Ex5D	sonication	thermal shock	1	A	A	-	-	-	-	-
PL_Ex6D	sonication	thermal shock	1	A	A	-	-	-	-	-
PL_Kx4E	sonication	thermal shock	1	A	A	4	A	A + 1 M KCl	1	A
PL_K42E	sonication	thermal shock	1	A	A	4	A	A + 1 M KCl	1	A
PL_Kx3E	sonication	thermal shock	1	A	A	4	A	A + 1 M KCl	1	A
PL_DNx8EQ	sonication	thermal shock	1	A	A	4	A	A + 1 M KCl	1	A
PL_DNx7EQalt	sonication	thermal shock	1	A	A	4	A	A + 1 M KCl	1	A
PL_DNx7EQ	sonication	thermal shock	1	A	A	4	A	A + 1 M KCl	1	A
PL_DNx6EQ	sonication	thermal shock	1	A	A	4	A	A + 1 M KCl	1	A
PL_DEx5K	sonication	thermal shock	1	A	A	4	A	A + 1 M KCl	1	A
PL_DEx6K	sonication	thermal shock	1	A	A	4	A	A + 1 M KCl	1	A
PL_Kx7R	sonication	thermal shock	1	A	A	4	A	A + 1 M KCl	1	A
UROIIIS	sonication	histidine tag	3	D (pH 7.4)	E (pH 7.4)	1	B + 0.15 M NaCl	B + 0.15 M NaCl	-	-
OIH2	cell disruption	(a)	4	A (pH 7.5) + 0.25 M KCl	A (pH 7.5) + 1 M KCl	1	A (ph 7.5) + 0.5 M KCl	A (ph 7.5) + 0.5 M KCl	-	-
1ALigN	cell disruption	histidine tag	3	M + 5 mM imidazole	M + 500 mM imidazole	1	M	M	-	-
FAH	cell disruption	histidine tag	3	D (pH 7.4)	E (pH 7.4)	1	A (pH 8.0) + 0.3 M KCl	A (pH 8.0) + 0.3 M KCl	-	-
GSTZ1	sonication	histidine tag	3	H (pH 7.5)	I (pH 7.5)	1	C (5 mM HEPES, pH 7.0) + 5% glycerol	C (5 mM HEPES, pH 7.0) + 5% glycerol	-	-
HGD	sonication	histidine tag	3	D (pH 7.4)	E (pH 7.4)	1	B (pH 7.0) + 0.5 M NaCl	B (pH 7.0) + 0.5 M NaCl	-	-
UROD	sonication	histidine tag	3	D (pH 7.4)	E (pH 7.4)	1	B (pH 8.0) + 0.15 M NaCl	B (pH 8.0) + 0.15 M NaCl	-	-
FXN	sonication	GST-His tag	2	J	K	2	J	-	1	C (pH 8.0) + 0.12 M NaCl
GST	sonication	GST-His tag	2	F (pH 6.0)	G (pH 6.0)	2	F (pH 6.0)	G (pH 6.0)	-	-
Nras	sonication	histidine tag	3	B + 0.5 M NaCl + 25 mM imidazole	B + 0.5 M NaCl + 250 mM imidazole	1	L	L	-	-
ALAD	sonication	histidine tag	3	D (pH 7.4)	E (pH 7.4)	1	B (pH 7.4) + 50 mM ZnSO4	B (pH 7.4) + 50 mM ZnSO4	-	-
RNaseA	(b)		4	B	B + 1 M NaCl	5	B	B + 1M NaCl	-	-
PL_G55A	sonication	thermal shock	1	A	A	-	-	-	-	-
PL_Kx7E	sonication	thermal shock	1	A	A	4	A	A + 1 M KCl	1	A
drk_WT	sonication	GST-His tag	2	F (pH 6.0)	G (pH 6.0)	2	F (pH 6.0)	-	1	A (50 mM NaP)
drk_T12K	sonication	GST-His tag	2	F (pH 6.0)	G (pH 6.0)	2	F (pH 6.0)	-	1	A (50 mM NaP)
drk_T12K_G43D	sonication	GST-His tag	2	F (pH 6.0)	G (pH 6.0)	2	F (pH 6.0)	-	1	A (50 mM NaP)
drk_T12K_T22D_G43D	sonication	GST-His tag	2	F (pH 6.0)	G (pH 6.0)	2	F (pH 6.0)	-	1	A (50 mM NaP)
drk_D42V	sonication	GST-His tag	2	F (pH 6.0)	G (pH 6.0)	2	F (pH 6.0)	-	1	A (50 mM NaP)

Note: after purification, drk_WT and derived mutants, and PL_Kx7E were aliquoted to a final concentration of 150 μM and 200 μM respectively, and lyophilized. Monomeric and dimeric forms of PL_G55A were separated as soon as they eluted from the column, aliquoted to a final concentration of 500 μM and immediately lyophilized. Previous to NMR measurement, lyophilized samples were resuspended in either 10% D_2O with or without cosolutes, or cell extracts. In the case of *Human Embryonic Kidney* (HEK) cell extract samples, also d_6-sodium trimethylsilylpropanesulfonate (DSS-d_6) was added to a final concentration of 1 mM for protein quantity determination.

3.2.4. Cytosol mimic buffer and cellular extracts

This section is aimed to explain the preparation procedure of the cytosol mimic buffer used for the studies of thermal stability by either CD or NMR, and the *Escherichia coli* (*E. coli*) and (*HEK*) cell extracts used for the thermodynamic stability determinations by NMR.

3.2.4.1. Cytosol mimic buffer

Table M8. Composition of the cytosol mimic buffer. *When prepared for thermal stability measurement studies by CD also D-Glutamic acid (Sigma-Aldrich, G1001) was added at half the total concentration to racemize the solution.

Compound	**Concentration (mM)**	**Quantity (g/L)**	**Company**	**Code**
K_2HPO_4	200	7.698	Sigma-Aldrich	P5655
KH_2PO_4		24.952	Sigma-Aldrich	P8281
NaCl	5	0.292	Thermo Fisher Scientific	S/3120/60
$MgCl_2$	1	0.095	Sigma-Aldrich	63063
$CaCl_2$	0.1	0.011	Sigma-Aldrich	6991
$ZnCl_2$	0.1	0.014	Thermo Fisher Scientific	7646-85-7
**L-Glutamic Acid*	96	14.124	Thermo Fisher Scientific	1149257
Glutathione	17	5.334	Sigma-Aldrich	G4251
D-Fructose-1,6-bisphosphate	15	5.102	Sigma-Aldrich	F6803
Adenosine triphosphate (ATP)	9,66	4.869	Sigma-Aldrich	A2383

3.2.4.2. *E. coli* extracts

E. coli cell extracts were prepared following the adaptation of the protocol described in the work from Kwon (Kwon, Y., Jewett, M., 2015) shown below:

1. A single colony or 5 mL of liquid culture of DH5α *E. coli* strain were inoculated to 5 mL of sterile LB medium and grown overnight at 37 °C with stirring.
2. 1 mL of overnight culture was added to a 500 mL-Erlenmeyer flask containing 200 mL of sterile LB medium and incubated for 4 h at 37 °C with stirring.
3. Culture growth was arrested by cooling down to 4 °C on ice, then the cells were pelleted by centrifugation at 5000 *g* and 4 °C for 20 min.
4. Supernatant was discarded and the cell pellets were resuspended in 50 mM sodium phosphate pH 6.0 buffer.
5. Cells were pelleted again by centrifugation at 5000 *g* and 4 °C for 10 min. The supernatant was discarded, and the cell were flash-frozen with liquid nitrogen.
6. 1 mL of 50 mM sodium phosphate pH 6.0 was added per gram of cell pellet and the cells were sonicated for a total time of 10 min with 0.5 s of pulse and 0.5 s of rest at 60% amplitude.
7. Cell lysates were centrifuged at 12000 *g* and 4 °C for 10 min, then the supernatant was separated from the pellet and incubated at 37 °C for 1.5 h with stirring.
8. Cell lysates were then centrifuged at 12000 *g* and 4 °C for 10 min and fast-frozen in aliquots of 1 mL until required.

Protein concentration of the *E. coli* extracts was determined by the Bradford assay. First, a calibration line that covered protein concentration values from 0.125 mg/mL to 2 mg/mL was constructed with standards and Bradford reagent from the Thermo Fisher Scientific kit. Then, the absorbance values of 1:10, 1:100 and 1:1000 diluted *E. coli* extract samples were measured and interpolated on the calibration line. The protein concentration determined this way was $\sim$40 mg/mL for all the prepared extracts.

3.2.4.3. *HEK* extracts

HEK cell extracts were prepared following the adaptation of the protocol described in the work from Theillet (Theillet, F. *et al.*, 2013) shown below:

1. Between $3.3 \cdot 10^8$ and $4.4 \cdot 10^8$ *HEK* cells were cultured in complete Gibco™ Dulbecco's Modified Eagle Medium (Gibco™ DMEM, Thermo Fisher Scientific) in a humidified incubator at 37 °C and 5% CO_2 using culture plates treated for adherent cells.
2. After growth to optimal number (i.e., $2.4 \cdot 10^8$ cells), the culture medium was removed, the cells were then washed with ice-cold phosphate buffer saline (PBS) and detached with the help of a cell scraper.
3. The cell suspension was centrifuged at 500 *g* for 10 min at RT, the supernatant was discarded, and the cell pellets were resuspended in an equivalent volume of 50 mM phosphate buffer, 3% (vol/vol) NP-40 detergent and 1% (vol/vol) PIC.
4. Cell lysis was induced by incubation for 10 min on ice.
5. The cell lysate was then centrifuged at 14500 *g* for 5 min at 4 °C, and the pellet was discarded.
6. Supernatant was flash-frozen with liquid nitrogen, the thawed at RT and centrifuged at 14500 *g* for 5 min at 4 °C.
7. Final cell extracts were flash-frozen with liquid nitrogen and stored at -20 °C until required.

Before its use for NMR measurements, HEK extracts were thawed at RT and mixed with pure D_2O to a final concentration of 10% and DSS to a final concentration of 1 mM.

3.3. Nuclear Magnetic Resonance Spectroscopy

In this work NMR spectroscopy has been used for the determination of the thermodynamic stability of proteins that present a reversible unfolding equilibrium, for the assignment of all the signals in the ^{1}H ^{15}N-HSQC spectra of the folded and unfolded species of PL_Kx7E, for the characterization of the dimerization reaction kinetics and thermodynamics of PL_G55A, for the determination of the free energy of unfolding of drk_WT, its related mutants; and PL_Kx7E; and for the determination of the hydrodynamical radii of a rationally chosen group of compounds, hereinafter referred to as cosolutes, through diffusion experiments. The following sections describe the acquisition and processing conditions of the data obtained by means of the experiments mentioned above.

3.3.1. Backbone assignment experiments

For the assignment of all the signals corresponding to the folded and unfolded forms in the ^{1}H ^{15}N HSQC spectra of both, drk_WT and PL_Kx7E, the HNCO/HN(CA)CO and HNCACB/HN(CO)CACB experiment pairs, as well as the ^{1}H ^{15}N HSQC experiments were performed. The acquisition conditions are summarized below (Table M9):

Table M9. NMR acquisition parameters for PL_Kx7E assignment.

Experiment	Dimension of acquired data			Spectral width (ppm)			Number of acquired scans	Relaxation delays (s)	% NUS
	t$_1$	t$_2$	t$_3$	F$_1$	F$_2$	F$_3$			
BT-HSQC	128 (^{15}N)	2048 (^{1}H)	-	35	16	-	8	0,3	-
BT-HNCACB	128 (^{13}C)	128 (^{15}N)	1024 (^{1}H)	80	35	13.6	16	0,3	25
BT-HN(CO)CACB	128 (^{13}C)	128 (^{15}N)	1024 (^{1}H)	80	35	13.6	16	0,3	30
BT-HNCO	128 (^{13}C)	128 (^{15}N)	1024 (^{1}H)	14	35	13.6	4	0,3	25
BT-HN(CA)CO	128 (^{13}C)	96 (^{15}N)	1536 (^{1}H)	14	35	13.6	8	0,3	35

The assignment data of both, the folded and unfolded signals of PL_Kx7E were collected in a 600-MHz Bruker Avance NEO (Bruker, Rheinstetten, Germany) equipped with a RT TXI probehead. The backbone assignments were performed using BEST-TROSY type versions of HNCACB, HN(CO)CACB, HNCO, and HN(CA)CO experiments (Solyom, Z., *et al.* 2013). The analysis of the data was conducted with the CcpNmr software package (Vranken W., *et al.* 2005). The secondary structure propensity score (SSP)of the unfolded form was estimated using random-coil shifts extracted from POTENCI (Nielsen, J., and Frans, M., 2018). The assignment data were deposited in the Biological Magnetic Resonance Bank (BMRB) with code 50954.

3.3.2. Osmolyte titrations

The obtention of the data required for the determination of the free energy of unfolding of drk_WT, all its related mutants and PL_Kx7E was performed by acquisition of ^{1}H ^{15}N-HSQC spectra at different cosolute concentrations. The data for the acquisition are shown below (Table M10):

Table M10. NMR acquisition parameters for monitoring the unfolding of drk and PL_Kx7E systems.

Experiment	**Dimension of acquired data**		**Spectral width (ppm)**		**Number of acquired scans**	**Recovery delay (s)**	**Experimental time (mm:ss)**	**Magnetic field intensity (T)**
	t_1 (^{15}N)	t_2 (^{1}H)	F_1 (^{15}N)	F_2 (^{1}H)				
BEST-HSQC	140/200/256	800	30/35/38	16	32	0.2	12:09	14.1
BEST-HSQC	280	800	38	16	16	0.2	23:34	18.8

All the data were acquired on an 800-MHz Bruker Avance III spectrometer equipped with a cryoprobe and on a 600-MHz Bruker Avance III US2 spectrometer. Acquired unfolding data were processed with NMRPipe, and the spectral analysis was performed with CcpNmr using 46 and 42 residues with well-separated signals as reporters for PL_Kx7E and drk respectively.

3.3.3. Kinetic studies

The data required for the determination of the kinetic rates of the direct and the reverse reactions of the dimerization reaction of PL_G55A were obtained by measuring the intensity changes of residues E2, I6, K7, F12, A13, T19, A20, G24, A33, A37, D38, K42, Y47, L58, I60, K61, and F62 in series of ^{1}H ^{15}N HSQC experiments measured at different times. The acquisition parameters and measured experiments are summarized below (Tables M11-M13):

Table M11. NMR acquisition parameters for monitoring the real-time oligomerization reactions of PL_G55A.

Experiment	**Dimension of acquired data**		**Spectral width (ppm)**		**Number of acquired scans**	**Recovery delay (s)**	**Experimental time (mm:ss)**	**Magnetic field intensity (T)**
	t_1 (^{15}N)	t_2 (^{1}H)	F_1 (^{15}N)	F_2 (^{1}H)				
BEST-HSQC	140	800	30	16	16	0.2	12:09	14.1
BEST-HSQC	280	800	38	16	8	0.2	11:52	18.8

Processing of the acquired data was performed with the help of CcpNmr software package under the parameters described below (Table M12):

Table M12. Summary table of the real-time NMR experiments performed. All measurements were performed at 298 K.

	Initial state	[Cosolute] (M)	B_0 (MHz)	Exp. Time (h)
Reference	Monomer	-	800	17.3
	Dimer	-	800	14.3
	Dimer	-	600	22.1
	Dimer	-	600	13.8
	Dimer	-	600	12.8
	Dimer	-	600	12.8
	Dimer	-	600	13.8
KCl	Monomer	0.5	600	23.1
	Dimer	0.5	600	22.5
	Dimer	0.5	800	15.4
	Dimer	0.25	600	14.6
K_2SO_4	Monomer	0.5	800	19.6
	Dimer	0.5	600	24.1
KSCN	Monomer	0.5	800	13.9
	Dimer	0.5	800	10.9
Taurine	Monomer	0.5	800	19.8
	Dimer	0.5	800	21.0
	Dimer	0.5	600	23.5
	Dimer	0.25	600	13.8
Sarcosine	Monomer	0.5	800	14.7
	Dimer	0.5	800	20.8
Glycine	Monomer	0.5	800	16.0
	Dimer	0.5	800	21.2
	Dimer	0.25	600	10.9
Trehalose	Monomer	0.5	800	21.2
	Dimer	0.5	800	24.2
Sucrose	Monomer	0.5	800	16.0
	Dimer	0.5	800	24.4
Urea	Monomer	0.5	800	20.4
	Dimer	0.5	800	12.9
	Dimer	0.25	600	11.7
GuHCl	Monomer	0.5	800	14.3
	Dimer	0.5	800	11.5
TMAO	Monomer	0.5	600	25.1
	Dimer	0.5	800	20.8
	Dimer	0.25	600	17.6

Table M13. NMR processing parameters of real-time experiments.

Experiment	Real Points		Spectral width (ppm)		Zero filling	-off (rad)	-end (rad)	Exponent of the sine bell	First-point scaling
	t_1 (^{15}N)	t_2 (^{1}H)	F_1 (^{15}N)	F_2 (^{1}H)					
BEST-HSQC (B_0 = 14.1 T)	70	400	EA	DQD	1	0.5	0.95	2	0.5
BEST-HSQC (B_0 = 18.8 T)	140	400	EA	DQD	1	0.5	0.95	2	0.5

All the data were collected on an 800-MHz Bruker Avance III spectrometer equipped with a cryoprobe and on a 600-MHz Bruker Avance III US2 spectrometer.

3.3.4. Diffusion experiments

The hydrodynamic radii, R_h, of the rationally chosen group of cosolutes considered in this work (see Table M12) were extracted from diffusion NMR experiments for measuring the translational diffusion (D_T). To this aim, a DOSY spectrum was recorded with 5 mM of cosolute in 20 mM phosphate buffer at pH 6.0 with 1 mM sodium azide and 40 mM of DSS-d_6 as internal reference. The estimated viscosity ratio of H_2O/D_2O 0.93:0.07 at 25 °C was calculated as proposed by Hardy and Cottington (Hardy, R., and Cottington R., 1949). The singlet at 0 ppm of DSS-d_6 was used as internal reference in the diffusion dimension (D_T = -9.1412 log(m^2/s)). As described in the work from Wang *et al.* (Wang, C., *et al.*, 2014), the DSS has the advantage of keeping a relatively constant D_T over a wide range of deuterium content and pH. R_h was calculated by solving the Stokes-Einstein equation (Equation M1):

$$R_h = \frac{k_B T}{6\pi\eta D_T}$$

Equation M1. Stokes-Einstein equation for the determination of the hydrodynamic radii of the cosolutes used in this work. k_B is the Boltzmann constant (1.3806 x 10^{-23} m^2·kg/s^2·K), and η is the viscosity.

For ions, the thermodynamical radii were extracted from previously reported values in bibliography (Roobottom, H., *et al.*, 1999; Creamer, T., *et al.*, 2018) (K^+ = 1.33 Å, Na^+ = 1.02 Å, Cl^- = 1.81 Å, SO_4^{2-} = 2.18 Å, and SCN^- = 2.09 Å), and for water, if it assumed no water shell around the molecule (i.e., no clathrates), then the reported value is 1.64 Å at 25 °C (Price, W., *et al.*, 2000). For the stick condition, the $H_2{}^{17}O$ radius has been reported to be 1.21 Å, however, the calculation taking in consideration this does not drastically change. The acquisition parameters and the excluded volume differences calculated for all the cosolutes shown in Table M12 are summarized below (Tables M14 to M16):

Table M14. NMR DOSY acquisition parameters for R_h calculation. [a]Number of variable gradient echoes. [b]Length of the gradient pulse.

Experiment	Dimension of acquired data		Spectral width	Number of scans	Recovery delay	Experimental time (mm:ss)	Magnetic field (T)	Separation between gradients	d[b] (s)
	t_1[a]	t_2 (^{1}H)	F_2 (^{1}H)						
DOSY	32	16384	22	32	1.5	13:19	18.8	0.125	0.0015

Table M15. Excluded volume and accessible surface area (ASA) differences for the cosolutes tested on PL_G55A.

Cosolute	R_h (Å)	ASA_{mon} - ASA_{H2O}	ASA_{dim} - ASA_{H2O}	Ratio ASA	V_{mon} - V_{H2O}	V_{dim} - V_{H2O}	Ratio V	Ratio differences ASA/V
Sucrose	4.79	1321.9	1483.4	1.12	2033.5	4389.7	2.16	0.52
Trehalose	4.64	1237.8	1374.5	1.11	1955.9	4211.9	2.15	0.52
Betaine	3.52	683.8	728.3	1.06	1231.9	2575.0	2.09	0.51
DSS	3.38	616.4	651.0	1.06	1139.7	2334.7	2.05	0.52
TMAO	3.12	491.5	511.6	1.04	946.4	1927.6	2.04	0.51
Glycine	2.90	398.7	403.3	1.01	793.7	1624.4	2.05	0.49
Sarcosine	2.72	328.7	315.1	0.96	676.2	1373.1	2.03	0.47
Taurine	2.69	316.9	296.3	0.94	656.3	1338.9	2.04	0.46
H_2O	1.64	-	-	-	-	-	-	-
K^+	1.81	28.9	12.9	0.45	118.9	213.6	1.80	0.25
Cl^-	1.33	-15.5	10.2	-0.66	-227.4	-378.8	1.67	-0.40

Table M16. List of parameters for the calculation of R_h of the group of cosolutes shown in Table M12. For those cosolutes that presented more than one signal, the averaged R_h in also shown.

Cosolute	D_T (log (m²/s⁻¹))	d(¹H) (ppm)	R_h (Å)	$\overline{R_h}$ (Å)
Betaine	-9.1378	3.89	3.35	3.52
	-9.1788	3.25	3.69	
DSS	-9.1412	0.00	3.38	3.38
Glycine	-9.9741	3.55	2.90	2.90
Sarcosine	-9.0432	3.60	2.70	2.72
	-9.0499	2.73	2.74	
Sucrose	-9.3342	5.40	5.27	4.79
	-9.2583	4.21	4.43	
	-9.2718	4.04	4.57	
	-9.2782	3.88	4.64	
	-9.3060	3.81	4.94	
	-9.2853	3.75	4.71	
	-9.3059	3.67	4.94	
	-9.2922	3.55	4.79	
	-9.2922	3.46	4.79	
Taurine	-9.0392	3.34	2.67	2.69
	-9.0459	3.25	2.72	
Trimethylamine N-oxide (TMAO)	-9.1062	3.26	3.12	3.12
Trehalose	-9.2712	5.18	4.56	4.64
	-9.2837	3.85	4.70	
	-9.2779	3.76	4.63	
	-9.2788	3.64	4.64	
	-9.2834	3.45	4.69	

The excluded volume differences shown in Table M15 were calculated using the ProteinVolume tool (Chen, C., and George, I., 2015), and the ASA were calculated using Pymol.

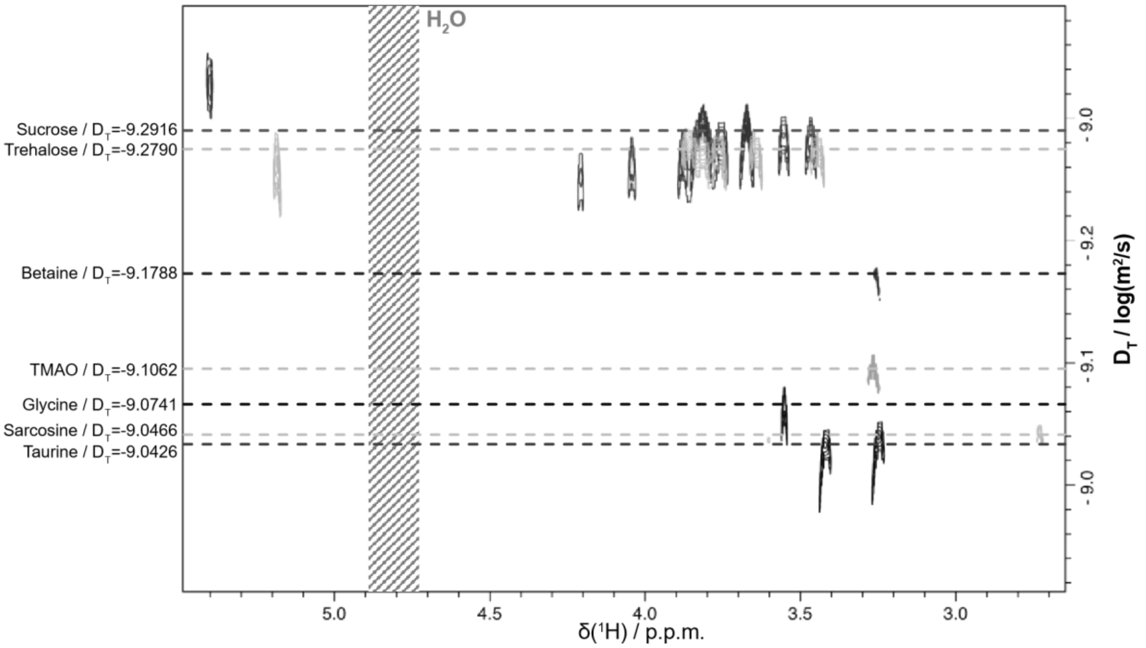


Figure M2. DOSY spectra for the cosolutes shown in Table M12.

3.3.5. Thermodynamic stability determination by NMR spectroscopy

Tables M17 and M18 show the concentration ranges and the number of points acquired for the determination of the stability for each protein and each cosolute under consideration. Figure M3 shows the intensity changes experienced by the signals in the ^{1}H ^{15}N-HSQC spectra corresponding to the cosolute titration of PL_Kx7E with KCl.

Table M17. List of the concentration ranges and number of points acquired for the determination of ΔG_{F-U} of the proteins shown in Table M3 (part 1). (a)Concentration range. (b)Number of acquired points. (c)Did not present folded signals at any concentration. For all the cosolutes tested, except for the cytosol mimic buffer and *E. coli* extracts, the concentration ranges are expressed in molar units. The cytosol mimic buffer and *E. coli* extracts' concentration ranges correspond to the proportion (per unit) of a stock solution prepared as explained in sections 3.2.3.1 and 3.2.3.2. respectively. Acquired points were equally distributed along the concentration range.

Acronym	**Concentration (mM)**	**Cosolute**											
		$CaCl_2$		Cito. mim.		*E. coli* extract		Glycine		KCl		K_2SO_4	
		(a)cr	(b)np	cr	np	cr	np	cr	np	cr	np	cr	np
PL_Kx7E	200	-	-	0-1.0	6	0-1.0	6	0-2.0	6	0-1.7	20	-	-
drk_WT	150	0-0.1	5	0-1.0	5	0-1.0	6	0-2.0	6	0-2.0	8	0-0.5	5
drk_T12K		0-0.1	5	0-1.0	5	0-1.0	7	0-2.0	5	0-2.0	9	0-0.5	5
drk_T12K_G43D		(C)	(C)	0-1.0	6	0-1.0	5	0-2.0	5	0-3.0	8	0-0.5	5
drk_T12K_T22D_G43D		0-0.1	5	0-1.0	6	0-1.0	5	0-2.0	5	0-2.0	6	0-0.5	5
drk_D42V		(C)	(C)	0-1.0	5	0-1.0	7	0-2.0	4	0-2.0	5	0-0.5	5

Table M18. List of the concentration ranges and number of points acquired for the determination of ΔG_{F-U} of the proteins shown in Table M3 (part 2). $^{(a)}$Concentration range. $^{(b)}$Number of acquired points. For all the cosolutes tested, except for the HEK extracts, the concentration ranges are expressed in molar units. The HEK extracts' concentration range corresponds to the proportion (per unit) of a stock solution prepared as explained in section 3.2.3.3. Acquired points were equally distributed along the concentration range.

Acronym	**Concentration (mM)**	**Cosolute**											
		HEK cell extract		Sarcosine		Sucrose		Taurine		Trehalose		TMAO	
		$^{(a)}$cr	$^{(b)}$np	cr	np	cr	np	cr	np	cr	np	cr	np
PL_Kx7E	200	0 - 1.0	6	0-2.0	6	0-1.0	16	0-0.5	6	0-1.0	6	0-2.0	5
drk_WT	150	0 - 1.0	5	0-0.5	5	0-1.0	5	0-0.5	6	0-1.0	5	0-2.0	6
drk_T12K		0 - 1.0	8	0-2.0	6	0-1.0	5	0-0.5	5	0-1.0	6	0-1.5	5
drk_T12K_G43D		0 - 1.0	6	0-1.0	5	0-0.75	5	0-0.5	6	0-1.0	5	0-2.0	5
drk_T12K_T22D_G43D		0 - 1.0	5	0-1.0	5	0-0.5	5	0-0.5	5	0-1.0	5	0-0.5	5
drk_D42V		0 - 1.0	2	0-2.0	5	0-1.35	4	0-0.5	5	0-1.0	5	0-2.0	5

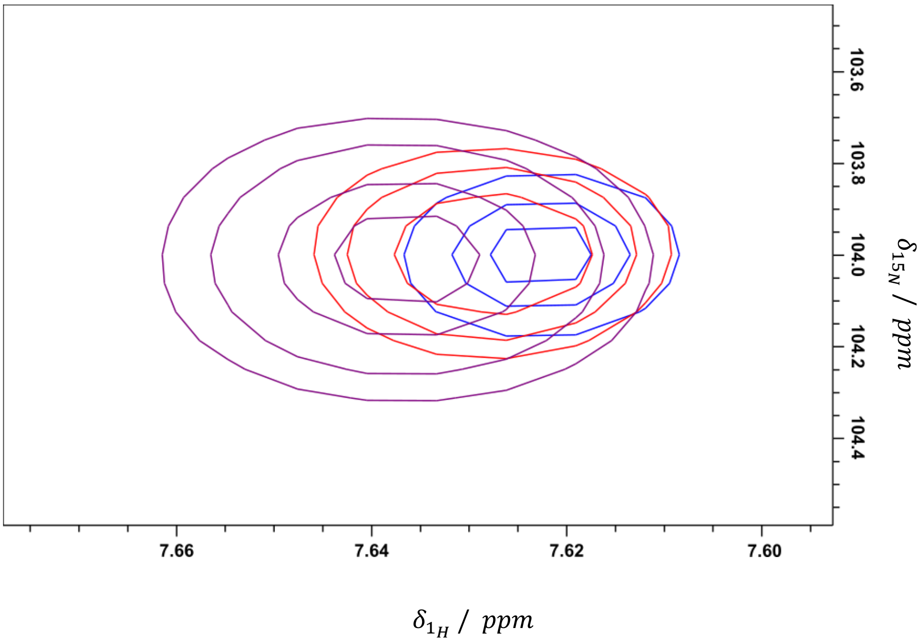


Figure M3. Signal intensity changes of glycine 55 from PL_Kx7E at 0 mM (blue), 78 mM (red), and 168 mM (purple) of KCl respectively.

For each protein at each cosolute concentration the peak volumes of all the signals in the ^{1}H ^{15}N-HSQC spectra obtained from CcpNmr analysis were converted into relative populations of the species and used for the calculation of the free energy of unfolding. First, the FF of each residue was calculated by dividing the peak volume of its folded form by the sum of the peak volumes of its folded and unfolded forms. Thereafter, the folded fractions were used to calculate the corresponding free energies of unfolding following the next equations:

$$\Delta G_{U-F} = \Delta G^{\circ}_{U-F} + RT \cdot \ln FF_i \qquad \Delta G^{\circ}_{U-F} = -RT \cdot \ln FF_i$$

Equations M2 and M3. Calculation of free energy of unfolding of residue *i* at a given cosolute concentration (left) and calculation of the free energy of unfolding of residue *i* in absence of cosolute. ΔG_{U-F} and ΔG°_{U-F} are the free energies of unfolding at a given cosolute concentration and in absence of cosolute respectively, and R in the ideal gas constant ($1.9872 \cdot 10^{-3}$ kcal·mol^{-1}·K^{-1}), and FF_i is the folded fraction of residue *i*.

3.4. Circular Dichroism Spectroscopy

In this thesis, CD has been used for the determination of the melting temperature, T_m, of a given protein in presence or absence of cosolutes. The proteins investigated by this technique are shown in Tables M1-M3.

All the CD spectra were collected in a JASCO J-810 spectropolarimeter using a quartz cuvette of 1 cm of path length. Depending on the protein, the concentration of the samples oscillated between 50 nM and 10 μM. Melting curves were acquired at 1 °C/min, monitored at 222 nm, with a bandwidth of 1 nm and collecting data every 0.2 °C. The temperature range of the experiment was optimized for every sample, assuring the proper determination of the baselines for the folded and unfolded states. For the determination of the experimental error each melting curve was duplicated. For each cosolute and protein the entire solubility range was covered taking between 4 and 9 measurements at different cosolute concentrations. All the measurements were performed in 20 mM sodium phosphate at pH 6.0.

Since the proteins whose thermal stability was measured by CD were observed not to present a reversible thermal unfolding equilibrium or did not present the spectroscopical properties required for the evaluation of their thermodynamic stability by NMR, their stability could not be directly measured and their T_m was determined instead.

3.4.1. Thermal denaturation experiments by Circular Dichroism

Tables M19 and M20 show the concentration ranges and the number of points acquired for each protein shown in Tables M1 and M2 and each cosolute.

Table M19. List of the concentration ranges and number of points acquired for the determination of the T_m of all the proteins shown in Tables M1-M3 (part 1). (a)Concentration range. (b)Number of acquired points. For all the cosolutes tested, except for the cytosol mimic buffer, the concentration ranges are expressed in molar units. The cytosol mimic buffer concentration range corresponds to the proportion of a stock solution prepared as explained in section 3.2.3.1. Acquired points were equally distributed along the concentration range.

Acronym	Concentration (mM)	Cosolute									
		$CaCl_2$		Cyt. mim.		Glycine		KCl		K_2SO_4	
		(a)cr	(b)np	cr	np	cr	np	cr	np	cr	np
PL_WT	150	0 - 2.0	6	0 - 0.8	5	0 - 0.8	5	0 - 3.0	7	0 - 0.5	6
PL_Kx5E		0 - 2.0	6	0 - 1.0	6	0 - 0.8	6	-	-	0 - 0.5	6
PL_Kx6E		-	-	-	-	0 - 1.0	6	-	-	-	-
PL_Kx5Q		-	-	-	-	0 - 1.0	6	-	-	-	-
PL_Ex4D		0 - 1.6	6	0 - 0.4	5	0 - 1.0	6	-	-	0 - 1.6	5
PL_Ex5D		-	-	0 - 1.0	6	0 - 1.0	6	0 - 2.0	6	0 - 0.5	6
PL_Ex6D		-	-	-	-	0 - 1.0	5	0 - 2.0	5	-	-
PL_Kx4E	500	0 - 1.6	5	0 - 0.6	5	0 - 1.0	6	0 - 2.0	5	0 - 0.47	6
PL_K42E		0 - 0.8	5	0 - 0.8	5	0 - 1.5	5	0 - 2.0	5	0 - 0.5	6
PL_Kx3E		-	-	-	-	0 - 1.0	6	0 - 2.0	5	0 - 0.5	6
PL_DNx8EQ	14	-	-	-	-	0 - 1.0	6	0 - 2.0	5	-	-
PL_DNx7EQalt	24	-	-	-	-	-	-	0 - 1.5	6	-	-
PL_DNx7EQ	22	0 - 0.4	5	0 - 0.9	6	0 - 1.0	6	0 - 1.5	8	0 - 1.8	6
PL_DNx6EQ	100	-	-	-	-	-	-	0 - 1.5	6	-	-
PL_Kx2E	500	0 - 2.0	6	0 - 1.0	6	0 - 1.5	5	0 - 2.0	5	0 - 0.5	6
PL_DEx5K	70	0 - 1.2	5	-	-	0 - 0.8	5	0 - 1.6	6	0 - 0.4	5
PL_DEx6K	64	0 - 1.6	5	0 - 0.6	6	0 - 0.8	5	0 - 1.6	6	0 - 0.4	5
PL_Kx7R	40	0 - 1.0	5	-	-	0 - 0.4	5	0 - 0.4	5	0 - 0.4	5
UROIIIS	50	0 - 0.4	5	-	-	0 - 2.0	8	0 - 1.125	6	0 - 0.4	5
RBP	1000	0 - 1.2	6	0 - 1.5	5	0 - 1.0	6	0 - 3.0	5	0 - 0.5	6
YmoA		-	-	-	-	0 - 1.5	6	0 - 3.0	5	-	-
OIH2		0 - 2.0	6	0 - 1.5	5	0 - 1.5	6	-	-	0 - 0.5	6
GGBP	500	0 - 1.2	7	0 - 1.0	6	0 - 2.0	6	0 - 3.0	6	0 - 0.5	6
hvCopG	1000	-	-	-	-	0 - 1.0	6	-	-	-	-
1ALigN											
btGH99		-	-	-	-	0 - 2.0	7	0 - 2.0	5	-	-
CAM		-	-	-	-	0 - 1.5	5	0 - 2.0	5	-	-
FAH	100	0 - 2.0	6	0 - 0.8	6	0 - 2.0	7	0 - 2.0	5	0 - 0.5	6
GSTZ1	49	0 - 1.6	5	-	-	0 - 2.0	8	0 - 2.0	6	0 - 0.4	5
HGD	80	0 - 0.8	5	-	-	0 - 1.5	5	0 - 2.0	5	0 - 0.4	5
UROD	70	0 - 0.1	5	-	-	0 - 2.0	6	0 - 3.0	8	0 - 0.4	5
FXN	150	0 - 2.0	6	0 - 0.9	6	0 - 1.5	5	0 - 2.0	7	0 - 0.5	6
CPOX	37	0 - 2.0	8	0 - 0.8	5	0 - 1.0	6	0 - 3.0	6	0 - 0.5	5
MBP	1000	0 - 1.6	5	0 - 1.5	5	0 - 1.0	6	0 - 3.0	6	0 - 0.5	6
GST	200	0 - 0.4	5	0 - 1.5	5	0 - 1.0	6	0 - 3.0	6	0 - 0.4	5
Nras	60	1 - 0.4	5	-	-	0 - 1.0	6	0 - 2.0	7	0 - 0.5	6
ALAD	40	2 - 0.4	6	-	-	0 - 1.0	6	0 - 2.0	6	0 - 0.8	5
RNaseA	300	3 - 0.4	6	0 - 0.94	6	0 - 2.0	6	0 - 2.0	6	0 - 0.5	6

Table M20. List of the concentration ranges and number of points acquired for the determination of the T_m of all the proteins shown in Tables M1-M3 (part 2). $^{(a)}$Concentration range. $^{(b)}$Number of acquired points. For all the cosolutes tested the concentration ranges are expressed in molar units. Acquired points were equally distributed along the concentration range.

Acronym	**Concentration (mM)**	**Cosolute**									
		Sarcosine		Sucrose		Taurine		Trehalose		TMAO	
		cr	np	cr	np	cr	np	cr	np	cr	np
PL_WT	150	0 - 1.0	6	-	-	0 - 0.5	6	0 - 1.0	5	0 - 1.0	8
PL_Kx5E		0 - 1.5	6	-	-	0 - 0.5	6	0 - 1.0	6	-	-
PL_Kx6E		0 - 1.0	6	-	-	0 - 0.5	6	0 - 1.0	6	0 - 1.0	6
PL_Kx5Q		0 - 1.0	6	0 - 0.5	6	0 - 0.5	7	0 - 1.0	6	0 - 1.0	6
PL_Ex4D		0 - 1.5	6	-	-	0 - 0.4	6	0 - 1.0	6	0 - 1.0	6
PL_Ex5D		0 - 1.0	6	0 - 1.0	6	0 - 0.5	6	0 - 1.0	6	0 - 1.0	6
PL_Ex6D		0 - 1.5	6	0 - 1.0	6	0 - 0.5	6	0 - 1.0	5	0 - 1.0	6
PL_Kx4E	500	0 - 1.0	5	0 - 1.0	6	0 - 0.5	6	0 - 1.0	6	0 - 1.0	6
PL_K42E		0 - 1.0	5	0 - 1.0	6	-	-	0 - 1.0	6	0 - 1.0	6
PL_Kx3E		0 - 1.5	6	0 - 1.0	6	0 - 0.5	6	0 - 1.0	7	0 - 1.0	6
PL_DNx8EQ	14	0 - 1.0	6	0 - 1.0	6	0 - 0.45	6	0 - 0.5	6	0 - 1.0	6
PL_DNx7EQalt	24	-	-	0 - 1.0	6	-	-	-	-	0 - 1.0	6
PL_DNx7EQ	22	0 - 1.0	6	0 - 1.0	6	0 - 0.42	6	-	-	-	-
PL_DNx6EQ	100	-	-	0 - 1.0	6	-	-	-	-	0 - 1.0	6
PL_Kx2E	500	0 - 1.5	5	0 - 1.0	6	0 - 0.5	6	0 - 0.6	5	0 - 1.0	6
PL_DEx5K	70	0 - 0.8	5	0 - 0.8	5	0 - 0.4	5	0 - 0.8	5	0 - 0.4	5
PL_DEx6K	64	0 - 0.8	5	0 - 0.8	5	0 - 0.4	5	0 - 0.8	5	0 - 0.4	5
PL_Kx7R	40	0 - 0.6	5	0 - 1.0	5	0 - 0.4	5	0 - 0.6	6	0 - 0.4	5
UROIIIS	50	0 - 2.0	5	0 - 0.75	6	0 - 0.5	6	0 - 1.0	5	0 - 0.75	6
RBP	1000	0 - 1.0	6	0 - 1.0	5	0 - 0.5	6	0 - 1.0	6	0 - 1.0	6
YmoA		0 - 1.5	6	0 - 1.0	5	0 - 0.5	6	0 - 1.0	6	0 - 1.0	6
OIH2		0 - 1.5	7	-	-	0 - 0.4	7	0 - 2.0	5	0 - 1.0	6
GGBP	500	0 - 2.0	6	0 - 1.0	5	0 - 0.5	6	0 - 1.0	8	0 - 1.0	6
hvCopG	1000	0 - 1.0	6	-	-	0 - 0.5	6	0 - 1.0	6	-	-
1ALigN											
btGH99		0 - 2.0	6	0 - 1.0	5	0 - 0.5	7	0 - 1.0	7	0 - 1.0	6
CAM		0 - 2.0	5	0 - 1.0	6	0 - 0.5	6	0 - 1.0	6	0 - 1.0	6
FAH	100	0 - 1.0	6	0 - 1.0	6	0 - 0.5	6	0 - 0.8	5	0 - 1.0	6
GSTZ1	49	0 - 2.0	8	0 - 1.0	6	0 - 0.5	8	0 - 1.0	7	0 - 1.0	6
HGD	80	0 - 1.2	5	0 - 1.0	6	0 - 0.5	6	0 - 1.0	6	0 - 1.0	6
UROD	70	0 - 2.0	6	0 - 2.0	5	0 - 0.5	6	0 - 1.0	7	0 - 2.0	7
FXN	150	0 - 1.5	5	0 - 2.0	6	0 - 0.5	6	0 - 1.0	5	0 - 2.0	5
CPOX	37	0 - 1.0	6	0 - 1.0	6	0 - 0.5	6	0 - 2.0	5	0 - 1.0	6
MBP	1000	0 - 2.0	6	0 - 1.0	5	0 - 0.5	6	0 - 1.0	6	0 - 2.0	6
GST	200	0 - 1.0	6	0 - 2.0	6	0 - 0.5	6	0 - 1.0	6	0 - 1.0	6
Nras	60	0 - 1.2	5	0 - 1.2	5	0 - 0.5	6	0 - 1.0	6	0 - 0.5	6
ALAD	40	0 - 0.8	6	0 - 0.8	5	0 - 0.5	6	0 - 1.0	6	0 - 0.5	5
RNaseA	300	0 - 1.6	5	0 - 1.6	5	0 - 0.5	6	0 - 1.0	6	0 - 0.5	5

3.4.1.1. Linear extrapolation method: two-states fitting

The extraction of T_m values from the experimental measurements (Figure M4) was conducted as follows: for each curve, least-squares fitting of Equation 11 to the CD experimental data was applied under the assumption that all the proteins analyzed by CD presented a two-states unfolding.

As explained in the work from Santoro and Bolen (Santoro, M., Bolen, D., 1988), it can be assumed that the observable (in this case the ellipticity) reporting on protein unfolding behaves linearly in both, the folded and the unfolded regions, and can be described by the following equations:

$$\theta_F = \theta_F^{\circ} + m_f \cdot T \qquad\qquad \theta_U = \theta_U^{\circ} + m_U \cdot T$$

Equations M4 and M5. Linear dependence of the ellipticity on temperature for the folded (left) and unfolded (right) states. θ_F and θ_U are the ellipticities of the folded and unfolded states at a given temperature, θ_F° and θ_U° are the intercepts at zero temperature, m_f and m_U are the slopes and T is the temperature.

Furthermore, this method also assumes that the observable at any point of the transition region is a combination of two contributions, coming from the folded and unfolded states. On the other hand, Becktel and Schellman (Becktel, W., and Schellman, J., 1987) demonstrated that if it is assumed a two-states unfolding, that the protein is stable at some temperature range, and that the variation of the heat capacity at constant pressure, ΔC_P, is constant and positive, then the variation of enthalpy, ΔH, entropy, ΔS, and free energy, ΔG, can be calculated by ideal dependence conditions:

$$\Delta H_{F-U} = \Delta H_{F-U}^{\circ} - \Delta C_P \cdot (T_m - T) \qquad\qquad \Delta S_{F-U} = \frac{\Delta H_{F-U}^{\circ}}{T_m} - \Delta C_P \cdot ln\frac{T_m}{T}$$

Equations M6 and M7. Relationship between the unfolding enthalpy (left) and entropy (right) differences with heat capacity at constant pressure and temperature. ΔH_{F-U} and ΔS_{F-U} are the differences of enthalpy and entropy of unfolding, ΔH_{F-U}° and ΔH_{F-U}° are enthalpy and entropy differences of unfolding at standard conditions respectively, and ΔC_p is the heat capacity difference at constant pressure.

$$\Delta G_{F-U} = \Delta H_{F-U} - T \cdot \Delta S_{F-U}$$

Equation M8. Calculation of the unfolding free energy difference from enthalpy and entropy differences.

The unfolding free energy difference can be used for the calculation of the equilibrium constant of the unfolding process, which in las term can be used to calculate the protein folded fraction, FF, as described below:

$$k_{F-U} = e^{-\frac{\Delta G_{F-U}}{RT}} = \frac{U}{F} \qquad\qquad FF = \frac{1}{1+k_{F-U}} = \frac{F}{F+U}$$

Equations M9 and M10. Calculation of the unfolding equilibrium constant (left) and relationship between the folded fraction and the equilibrium constant of the unfolding. F and U represent the abundances of the folded and unfolded species respectively.

Thus, taking in consideration all these aspects and the linear extrapolation assumptions, the ellipticity at any temperature value can be calculated as explained in the following formula:

$$\theta_{F-U} = FF \cdot \left(\theta_F^{\circ} + m_F \cdot T\right) + (1 - FF) \cdot (\theta_U^{\circ} + m_U \cdot T)$$

Equation M11. Calculation of the ellipticity from the folded fraction. θ_{F-U} is the ellipticity at any temperature value.

All the equations of section 3.4.1.1 were computed using in-house Matlab© scripts. For each thermal unfolding curve, the T_m (Figure M4) was obtained as the temperature value that minimized the squared error function between Equation M11 and the experimental data. The function minimization routine applied was the Levenberg-Marquardt simplex algorithm (Moré, J. 1978), and the initial values passed to the function were a standard value of ΔC_P for protein unfolding and a standard value of ΔH_{F-U}° obtained from the bibliography (Myers, J., Pace, N., and Scholtz, M., 1995; Seeling, J., Schönfeld, H., 2016), and the slopes and intercepts of the Equations M4 and M5 fitted to the (in average) initial 60 and final 40 experimental points of the corresponding melting curve.

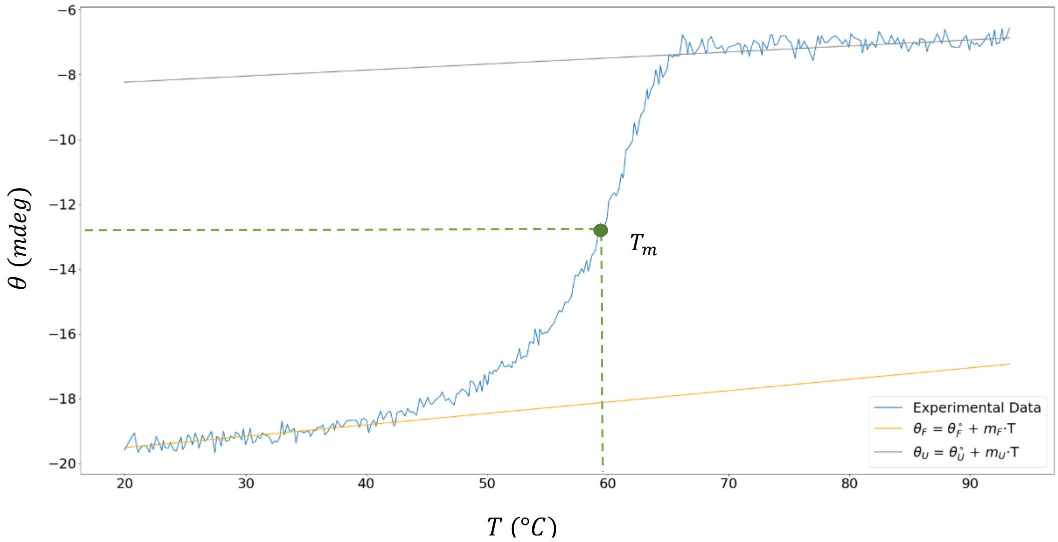


Figure M4. Thermal unfolding curve of OIH2 in absence of cosolutes. The T_m, corresponding to the inflection point of the curve, and least-squares fitting lines of the folded and unfolded regions are highlighted.

3.4.2. Data analysis

For a given protein and cosolute, the resulting T_m dataset (Figure M5) was fit to the general equation:

$$T_m = T_m^0 + m_{T_m} \cdot [C] + a \cdot e^{-b \cdot [C]}$$

Equation M12. Calculation of the T_m at a given cosolute concentration. T_m and T_m° are the melting temperature at a certain cosolute concentration and the intercept at zero molar concentration of cosolute respectively, $[C]$ is the molar concentration of cosolute and a and b are coefficients that account for an eventual additional non-linear dependence.

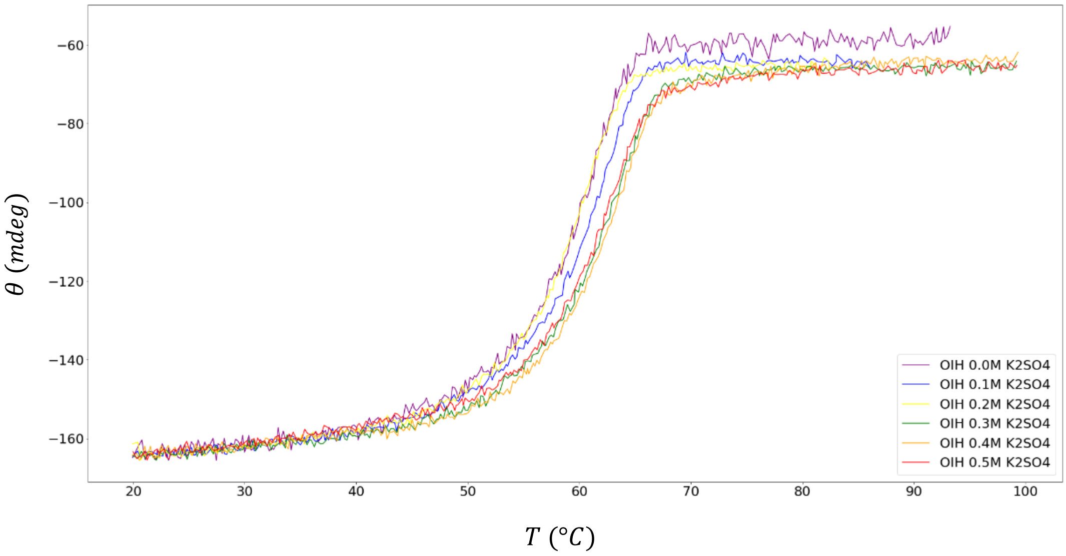


Figure M5. Thermal unfolding titration of OIH2 protein with variable concentrations of K_2SO_4. The ellipticity (y axis) variation at different cosolute concentration over a temperature range (x axis) is shown.

In most cases, the T_m dataset fitted well to a linear version of Equation M2, it is to say, non-linear term was unnecessary and could be neglected from the analysis. However, in other cases it must be considered (Figure M6).

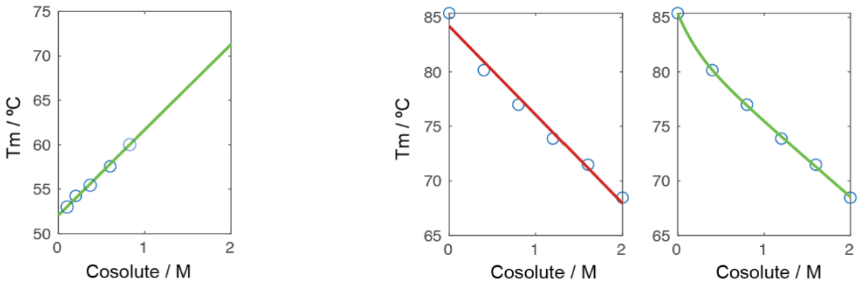


Figure M6. Fitting of the T_m data set of PL_Kx3E in presence of sucrose to the linear version of Equation M2 (left)., and fitting of the T_m data set of ALAD in presence of $CaCl_2$ to the linear (red) and exponential (green) versions of Equation M2 (right)

As demonstrated by Becktel and Schellman (Becktel, W., and Schellman, J., 1987), the melting temperature at a given cosolute concentration could be converted into changes in free energy, $\Delta\Delta G_{[C]}^{\circ,U-F}$, using the following equation:

$$\Delta\Delta G_{[C]}^{0,U-F} = \Delta H_{m}^{0,U-F} \cdot \left(1 - \frac{T_m^0}{T_{m,[C]}}\right)$$

Equation M13. Calculation of the changes in the free energy from melting temperature. $\Delta H_m^{0,U-F}$ is the unfolding enthalpy of PL_WT at T_m and in the absence of cosolute (53 ± 4.8 kcal/mol) described in the work from Tadeo (Tadeo, X. *et al.* 2009).

According to this, the T_m of PL_WT and all related mutants shown in Tables M1-M3 was converted to $\Delta\Delta G_{[C]}^{0,U-F}$.

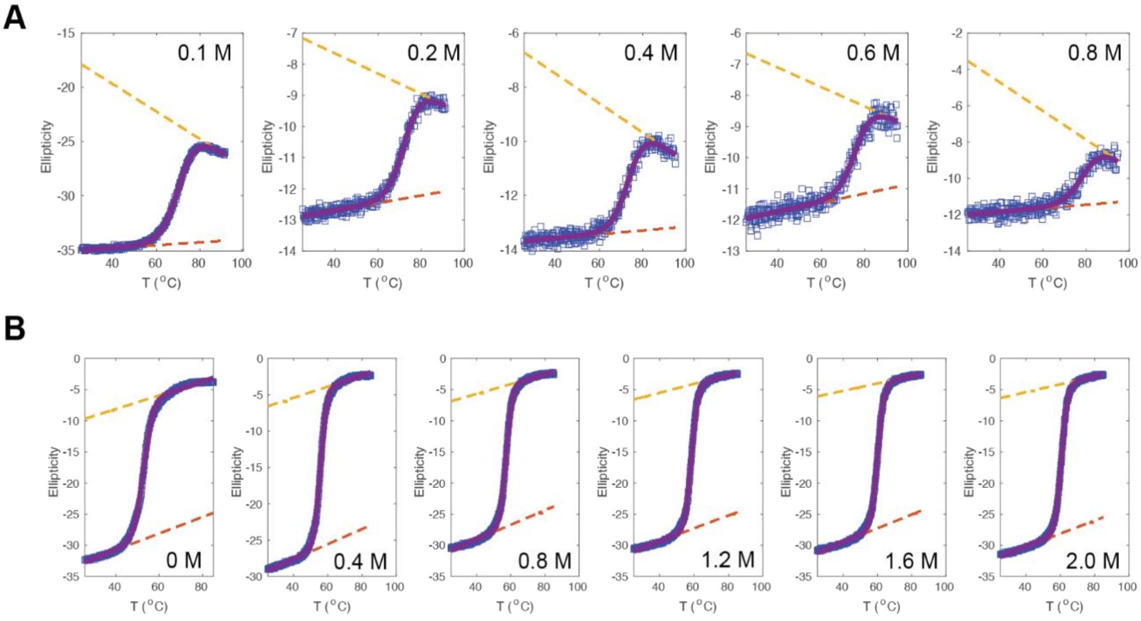


Figure M7. Thermal denaturation curves of PL_Kx3E in presence of sucrose (A), and ALAD in presence of CaCl2.

3.5. Error analysis

Experimental error was estimated from duplicate measurements (Table M21).

Table M21. List of the replicates used for the estimation of the experimental error.

	Protein	GST	PL_Kx2E	UROIIIS
Cosolute concentration	$CaCl_2$	-	1.6 M	-
		-	2.0 M	-
	ECMIM	-	0.2 X	-
		-	0.4 X	-
	KCl	0.5 M	0.5 M	0.225 M
		1.0 M	1.0 M	0.45 M
	K_2SO_4	-	0.1 M	-
		-	0.2 M	-
	Glycine	0.2 M	0.25 M	0.25 M
		0.4 M	0.5 M	0.5 M
	Sarcosine	0.2 M	0.5 M	0.5 M
		0.4 M	1.0 M	1.0 M
	Sucrose	0.5 M	0.2 M	0.15 M
		0.75 M	0.4 M	0.3 M
	Taurine	0.1 M	0.1 M	0.1 M
		0.2 M	0.2 M	0.2 M
	Trehalose	-	0.2 M	0.2 M
		-	0.4 M	0.4 M
	TMAO	-	0.2 M	0.225 M
		-	0.4 M	-

Linear regression models' statistical significance was checked by the regression diagnostic plots from the linda::gg_diagnose function package (version 0.9). For the cosolutes that presented a significant correlation ($|R|>0.5$ and $p<0.05$) the linearity, homoscedasticity, independence between observations and normality were evaluated.

3.6. Surface calculations

Residue-specific solvent accessible areas were calculated using Pymol, with a probe radius of 1.4 Å, and referenced to the total area of the residues (Creamer,T., *et al.* 1997). Exposed residues were selected according to a cutoff value of 60% of area exposed to the solvent. Similar results were obtained for a cutoff of 70%.

3.7. Accelerated molecular dynamic simulation

This section describes the conditions under which the molecular dynamic simulation was executed to study the transition state of the dimerization reaction of PL_G55A.

The dynamic simulation was performed using the x-ray diffraction structures of the B1 domain of protein L from *Peptostreptococcus magnus* (PDB: 1Hz5) and a variant in which the G55A mutation induces B1 domain swapping (PDB: 1K51) as initial geometries. Molecular dynamic simulations were run with the AMBER suite (Lee, T., *et al.* 2018) using the *ff14SB* (Maier, J., *et al.* 2015) force field. Initial structures were first neutralized with either Na^+ or Cl^- and located at the center of a cubic TIP3P water box (Jorgensen, W., *et al.* 1983) with a buffering distance between solute and box of 10 Å. A two-stage geometry optimization approach was followed for each complex: the first stage minimized only the positions of solvent molecules and ions, and the second stage was an unrestrained of all the atoms in the simulation cell. Temperature was then raised from 0 to 300 K along 1 ns under a constant pressure of 1 atm and periodic boundary conditions. The Andersen temperature coupling scheme (Andersen, H., 1980; Andrea, T., *et al.* 1983) was followed to apply harmonic restrains of 10 kcal·mol^{-1} to the solute. The SHAKE algorithm (Miyamoto, S., and Kollman, P., 1992) was used to treat water molecules such that the angle between hydrogen atoms was kept fixed along the simulation. Long-range electrostatic effects were modeled using the particle mesh Ewald method (Darden, T., *et al.* 1993) and a cutoff of 8 Å was applied to the Lennard-Jones interactions. Each system was equilibrated for 2 ns with a 2 fs time step at a constant volume and temperature of 300 K. Then, a Gaussian accelerated molecular dynamics (GaMD) protocol (Miao, Y., *et al.* 2014; Miao, Y., *et al.* 2015) designed to enhance conformational sampling by addition of a harmonic boost potential to smoothen the system potential energy surface was applied after equilibration using conventional molecular dynamics (cMD). This protocol consisted of an initial cMD pre-equilibration stage of 2 ns in which the boost potential was applied and the boost parameters were not updated, and a second pre-equilibration stage of 50 ns in which the boost potential was applied and boost parameters were updated (*igamd* = 3, *iE* = 1, *irest_gamd* = 0, *sigmaOP* = 6.0, *sigmaOD* = 6.0, *ntave* = 200000, *ntcmdprep* = 200000, *ntcmd* = 1000000, *ntebprep* = 800000, *nteb* = 25000000). In the final production stage (90-100 ms), the boost potential was applied, and the boost parameters obtained from de GaMD equilibration were not updated (*igamd* = 3, *iE* = 1, *irest_gamd* = 1, *sigmaOP* = 6.0, *sigmaOD* = 6.0, *ntave* = 200000, *ntcmdprep* = 0, *ntcmd* = 0, *ntebprep* = 0, *nteb* = 0). The GaMD trajectories were run in the NTV ensemble at 300 K with a time step of 2 fs using the Andersen thermostat. The applied potential was a dual boost on both dihedral and total potential energy. Representative geometries, molecular surfaces, and volumes were extracted from the trajectories using the *cpptraj* module of AMBER.

3.8. Kinetic rates determination

The dimerization reaction of PL_G55A can be schematically expressed as follows:

$$M + M \underset{k_{dis}^{D\to M}}{\overset{k_{ass}^{M\to D}}{\leftrightarrows}} D$$

Equation M14. Schematic representation of the dimerization reaction of PL_G55A. $k_{ass}^{M\to D}$ and $k_{dis}^{D\to M}$ are the association and dissociation kinetic rates respectively.

The kinetic process of domain-swapping is bi-exponential (Moschen, T., and Tollinger, T., 2014) and the homodimerization kinetic equations show the following expression:

$$C_M(t) = C_0 - 2\frac{\lambda_1\lambda_2(1 - e^{-(4k_{ass}(\lambda_1-\lambda_2)t)})}{\lambda_1 - \lambda_2 e^{-(4k_{ass}(\lambda_1-\lambda_2)t)}}$$

$$C_D(t) = \frac{\lambda_1\lambda_2(1 - e^{-(4k_{ass}(\lambda_1-\lambda_2)t)})}{\lambda_1 - \lambda_2 e^{-(4k_{ass}(\lambda_1-\lambda_2)t)}}$$

Equations M15 and M16. Time-course monomer and dimer populations for the association pseudoequilibrium of PL_G55A. $C_D(t)$ and $C_M(t)$ are the normalized concentrations over time of the dimer and the monomer respectively, C_0 is the initial concentration of protein, and k_{ass} is the association kinetic rate. $\lambda_{1/2}$ is described in Equation M17.

$$C_M(t) = C_0 - 2\frac{\lambda_2(1-\lambda_1) + \lambda_1(\lambda_2 - 1)e^{-(4k_{ass}(\lambda_1-\lambda_2)t)}}{(1-\lambda_1) + (\lambda_2 - 1)e^{-(4k_{ass}(\lambda_1-\lambda_2)t)}}$$

$$C_D(t) = \frac{\lambda_2(1-\lambda_1) + \lambda_1(\lambda_2 - 1)e^{-(4k_{ass}(\lambda_1-\lambda_2)t)}}{(1-\lambda_1) + (\lambda_2 - 1)e^{-(4k_{ass}(\lambda_1-\lambda_2)t)}}$$

Equations M17 and M18. Time-course monomer and dimer populations for the dissociation pseudoequilibrium of PL_G55A.

$$\lambda_{1/2} = \frac{1}{2}\left(\frac{k_{dis}}{4k_{ass}} + C_0\right) \pm \sqrt{\frac{1}{2}\left(\frac{k_{dis}}{4k_{ass}} + C_0\right)^2 - \frac{1}{4}(C_0)^2}$$

Equation M19. λ parameter of Equations M15, M16, M17 and M18. k_{dis} is the dissociation kinetic rate.

For each of the Equations M15-M18, a Matlab function was built for the determination of both, the association and dissociation kinetic rates of the PL_G55A dimerization reaction. This was achieved by applying a least-squares fitting of these equations to the real-time NMR peak volumes extracted from CcpNmr analysis. Since the times required for the separation of both species by SEC and acquirement of the first ^{1}H ^{15}N HSQC spectrum were unknown, the first data points of the time series were normalized by the extrapolated value at t = 0 to ensure the correct fitting.

Chapter IV

A Quantitative Metric to Investigate the Relationship Between Surface Composition and Quinary Structure

4. A Quantitative Metric to Investigate the Relationship Between Surface Composition and Quinary Structure

4.1. A metric that relates protein's surface composition to quinary structure

As an introduction for this part, the following lines summarize the section 1.1 of the introduction and connect it with the results presented in this section.

Life has adapted to thrive in very diverse conditions, some of them considered as extreme. Some extreme environments are characterized by their hyperosmotic compositions, for example places with high molar concentrations of salts. To survive in these places, Nature has developed strategies to compensate the osmotic stress that the biomolecules such as proteins undergo. For example, some of the organisms that survive in hypersaline environments reduce the osmotic stress on their proteins by increasing their cytosolic salt concentration up to 4-5 M (Janos, L., 1974). Consequently, proteins belonging to these organisms present a biased aminoacid composition in which residues with negatively charged and short side chains are preferred over bulkier and neutral or positively charged residues. This trait, named haloadaptation, involve in great part the residues at the surface of the protein and does not affect the protein's fold, thus keeping the biomolecular function unaltered. Investigation of the haloadaptation during past decades has demonstrated that this mechanism is intended exclusively to optimize protein stability either directly or by increasing solubility. Remarkably, the haloadaptation mechanism is triggered not only by high concentrations of salts in the cytosol, but also by the accumulation of other molecules (Yancey, P., *et al.* 1982; Somero, G., 1986), and at the molecular level it seems not to act through a single mechanism, but to be general enough to allow the adaptation to the presence of other salts via the Hofmeister effect (Baldwin, R., 1996; Tadeo, X., *et al.* 2009; Tadeo, X., Pons, M., and Millet, O., 2007; Tadeo, X., Castaño, D., and Millet, O., 2007). Furthermore, it can be qualitatively explained in terms of excluded volume and preferential interaction, which govern the effect of other cosolutes (Record, M., Anderson, F., and Lohman, T., 1978; Miklos, A., *et al.* 2011; Bai, J., *et al.* 2017).

Taking in consideration all the ideas previously discussed and the generality of the haloadaptation mechanism, it was proposed to use it as model to investigate the sensitivity of proteins to the environment and the quinary structure.

The simplest approach would be to build a metric considering the whole sequence of a protein (Equation R1):

$$\sigma_{all} = \frac{1}{n_T} \cdot \sum_{i=1}^{20} n_i \cdot \Delta n_i$$

Equation R1. Metric considering the whole sequence of a protein. Δn_i is the difference between the average number of residues of type i in the halophilic and mesophilic proteomes (Table R1).

Table R1. Differences between the amino acid compositions of halophilic and mesophilic organisms considering the whole protein (Δn_i) or simply the residues at the surface that expose more than 60% surface area to the solvent.

Residue	Δn_i	$\Delta n_{i,SASA}$
A	0.06	-0.07
R	-0.15	0.11
N	-0.11	-0.01
D	0.58	0.65
Q	-0.07	0.00
E	0.26	0.12
G	-0.05	-0.05
H	0.09	0.06
I	-0.18	-0.04
L	-0.22	-0.01
K	-0.55	-0.47
M	-0.27	-0.01
F	-0.16	-0.03
P	-0.13	-0.08
S	0.09	-0.02
T	0.45	-0.07
Y	-0.08	0.02
V	0.13	-0.12
W	-0.21	0.03
C	-0.30	-0.01

As an improved version of the correlation (see section 4.2.2 for a comparison), we selected a set of 49 proteins belonging to halophilic organisms for which high-resolution structures were available (Table R2). A mirror set of proteins was created with the mesophilic orthologs of each of the proteins included in the halophilic set of proteins (Table R3), trying to include proteins from organisms as diverse as possible (i.e., plants, animals, viruses, and bacteria) with the objective of keeping a broad definition of mesophilic organism. The analysis and comparison between the structures of all the proteins shown in Table R2 and their corresponding mesophilic counterparts demonstrated that the fold was always preserved even at the quaternary level (Figures R1 and R2), which reinforced the idea that the evolutionary pressure on halophilic organisms has reshaped the protein's surface in order to optimize its interactions with the environment. This observation suggested a greater relevance of the surface in the haloadaptation mechanism, which prompted a refined definition of the halophilic composition in which only residues at the surface that expose more than a 60% of their surface to the solvent were considered.

Table R1 shows both, the frequency differences between halophilic and mesophilic proteomes when considering the whole protein, Δn_i, and the equivalent differences when considering only the surface, $\Delta n_{i,SASA}$.

Table R2. List of the 49 halophilic proteins considered for the calculation of the amino acid composition differences shown in Table R1. The code of each protein coincides with the code of its corresponding mesophilic counterpart shown in Table R3.

Protein Type	Code	PDB ID	N° of Residues	Halophilic Organism
Alcohol dehydrogenase	1	5yvr	409	Haloferax volcanii
Alkaline phosphatase	2	2x98	431	Halobacterium salinarum
Alkaline serine protease	3	3cp7	218	Nesterenkonia sp
Alpha-amylase	4	1jd7	453	Pseudoalteromonas haloplanktis
Amidase	5	3hkx	283	Nesterenkonia sp
Amidase	6	4izs	283	Nesterenkonia sp
Amidase	7	5jqn	266	Nesterenkonia sp
Amidase	8	5ny2	271	Nesterenkonia sp
Bacterorhodopsin	9	5vn9	262	Halobacterium salinarium
Catalase-peroxidase 2	10	3vli	737	Haloarcula marismortui
Cell division protein Cetz2	11	4b45	349	Halobacterium salinarum
Cell division protein Cetz1	12	4b46	395	Halobacterium salinarum
Cellulase	13	1tvn	293	Pseudoalteromonas haloplanktis
Chemotaxis protein CheC	14	3qta	233	Haloarcula marismortui
D-2-hydroxyacid dehydrogenase	15	5mh5	308	Haloferax mediterranei
D-2-hydroxyacid dehydrogenase	16	5mha	308	Haloferax mediterranei
DNA binding protein	17	6qil	116	Halobacterium salinarium
Dihydrofolate reductase	18	2ith	162	Haloferax volcanii
Dodecin	19	2vxa	72	Halorhodospira halophila
Ectoine hydrolase DoeA	20	6yo9	399	Halomonas elongata
Endo-1,4-beta-xylanase	21	2b4f	405	Pseudoalteromonas haloplanktis
Ferredoxin	22	1e0z	128	Halobacterium salinarum
Glucose dehydrogenase	23	2b5v	357	Haloferax medterranii
Glutathione S-transferase	24	6gzf	341	Natrialba magadii
H. volcanii putative methyltransferase	25	6f5z	231	Haloferax volcanii
Halorhodopsin	26	3a7k	291	Natrosomonas pharaonis
Malate dehydrogenase	27	1o6z/4bgu	303	Haloferax volcanii
Malate dehydrogenase	28	2x0r	304	Haloarcula marismortui
Malate synthase	29	5tao/3pug	433	Haloferax volcanii
Nitrogen regulatory protein P-II	30	4ozj	143	Haloferax mediterranei
Nucleoside diphosphate kinase	31	2az1	181	Halobacterium salinarum
Nucleoside diphosphate kinase	32	3vgs	141	Halomonas sp
Nucleoside diphosphate kinase	33	2zua	174	Haloarcula marismortui
Periplasmic substrate binding protein	34	2vpn	316	Halomonas elongata
Periplasmic substrate binding protein	35	3gyy	341	Halomonas elongata
Phage integrase/site-specific recombinase	36	3nrw	117	Haloarcula marismortui
Proliferating cell nuclear antigen	37	3hi8	247	Haloferax volcanii
Ribulose bisphosphate carboxylase large chain	38	6uew	506	Halothiobacillus neapolitanus
S-formylglutathione Hydrolase	39	3ls2	280	Pseudoalteromonas haloplanktis
Sensor protein	40	3bwl	126	Haloarcula marismortui
Superoxide dismutase [Fe]	41	4L2B	192	Pseudoalteromonas haloplanktis
Transcription regulator	42	3crj	199	Haloarcula marismortui
Ubiquitin-like archaeal modifier protein 1	43	3po0	89	Haloferax volcanii
Ubiquitin-like archeal modifier protein 2	44	2m19	106	Haloferax volcanii
Uncharacterized (gene locus rrnAC0354)	45	2lnu	190	Haloarcula marismortui
Universal stress protein TeaD	46	3hgm	147	Halomonas elongata
Vng1086c	47	2gf4	100	Halobacterium salinarum
b-carbonic anhidrase	48	2fgy	542	Halothiobacillus neapolitanus
katG catalase-peroxidase	49	3vlm	737	Haloarcula marismortui

Table R3. List of the 49 mesophilic counterparts considered for the calculation of the amino acid composition differences shown in Table R1. The code of each protein coincides with the code of its corresponding halophilic counterpart shown in Table R2.

Mesophilic Counterpart	**Code**	**PDB ID**	**N° of Residues**	**Mesophilic Organism**
Alcohol dehydrogenase	1	1ADC	374	Equus caballus
Alkaline phosphatase	2	3BDG	458	Escherichia coli
M-protease	3	1WSD	269	Bacillus clausii
Alpha-amylase	4	1BVN	496	Sus scrofa
AMPDH2	5	4BJ4	246	Pseudomonas aeruginosa
Amidase	6	3A1I	521	Rhodococcus sp.
Aliphatic amidase	7	2UXY	341	Pseudomonas aeruginosa
N-Acetylmuramoyl-L-alanine amidase	8	2BH7	261	Escherichia coli
Chloride pumping rhodopsin	9	6JYB	296	Nonlabens marinus
Catalase-peroxidase 2	10	5jhx	764	Pyricularia oryzae
FTSZ	11	1W5E	364	Methanocaldococcus jannaschii
Cell division protein FTSZ	12	1W5F	353	Thermotoga maritima
Cellulase	13	1CEN	343	Hungateiclostridium thermocellum
Chemotaxis protein CheC	14	2f9z	205	Thermotoga maritima
D-2-hydroxyacid dehydrogensase protein	15	4LSW	318	Ketogulonicigenium vulgare
NAD+-dependent (R)-2-Hydroxyglutarate Dehydrogenase	16	1XDW	331	Acidaminococcus fermentans
DNA-binding protein Fis	17	5DTD	98	Escherichia coli
Dihydrofolate reductase	18	2INQ	159	Escherichia coli
Dodecin	19	2V18	68	Thermus thermophilus
Ectoine hydroxylase	20	4MHU	314	Sphingopyxis alaskensis
Endo-1,4-beta-xylanase	21	1XYF	436	Streptomyces olivaceoviridis
Ferredoxin-thioredoxin reductase	22	2PVO	110	Spinacia oleracea
Glucose 1-dehydrogenase	23	2CDB	366	Saccharolobus solfataricus
Glutathione S-transferase theta-2	24	4MPG	266	Homo sapiens
C-methyltransferase CouO	25	5M58	230	Streptomyces rishiriensis
Halorhodopsin	26	6NWF	241	Mastigocladopsis repens
Malate dehydrogenase	27	1IB6	312	Escherichia coli
Malate dehydrogenase	28	5mdh	333	Sus scrofa
Malate synthase G	29	2GQ3	729	Mycobacterium tuberculosis
Nitrogen regulatory protein P-II	30	2XBP	113	Synechococcus elongatus
Nucleoside diphosphate kinase	31	5V6D	149	Neisseria Gonorrhoeae
Nucleoside diphosphate kinase	32	2HUR	142	Escherichia coli
Nucleoside diphosphate kinase I	33	1u8w	149	Arabidopsis thaliana
ABC transporter, periplasmic substrate-binding protein	34	1TWY	290	Vibrio cholerae
Maltose-binding periplasmic protein	35	3HPI	372	Escherichia coli
Integrase	36	2oxo	103	Bacteriophage sp.
Proliferating cell nuclear antigen	37	1VYJ	261	Homo sapiens
Ribulose bisphosphate carboxylase large chain	38	1UZH	475	Chlamydomonas reinhardtii
S-formylglutathione hydrolase	39	6JZL	279	Shewanella frigidimarina
Sensor protein pfeS	40	3kyz	125	Pseudomonas aeruginosa
Superoxide dismutase [Fe]	41	1WB8	210	Saccharolobus solfataricus
HTH-type transcriptional regulator RutR	42	4xk4	212	Escherichia coli
Modifier protein 1	43	1AP0	73	Mus musculus
Molybdopterin converting factor subunit 1	44	1NVI	81	Escherichia coli
-	45	-	-	-
-	46	-	-	-
-	47	-	-	-
-	48	-	-	-
Catalase-peroxidase	49	5L02	728	Burkholderia pseudomallei

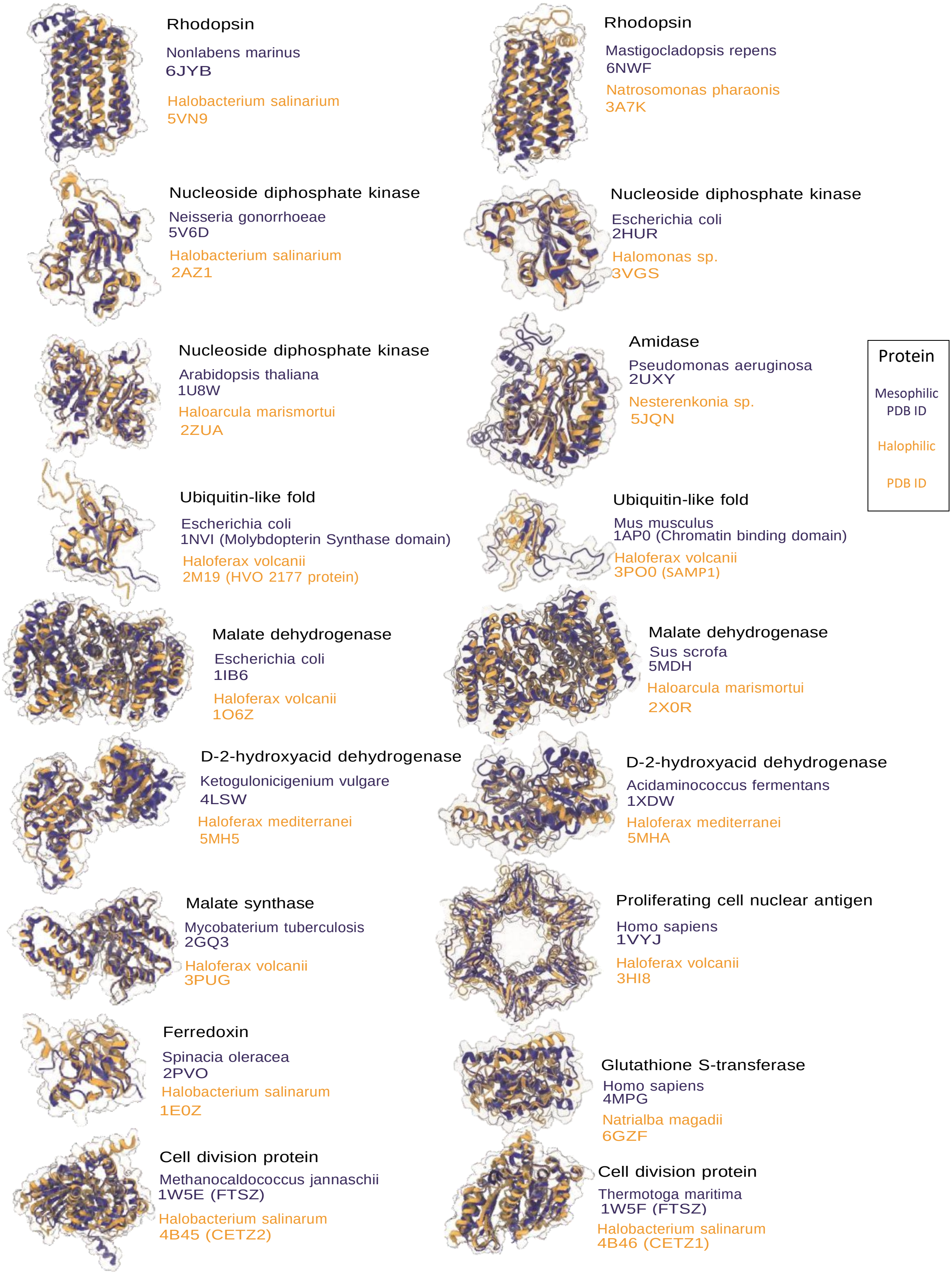


Figure R1. Comparison between the structures of the proteins shown in Table R2 (blue) and their corresponding mesophilic counterparts (orange).

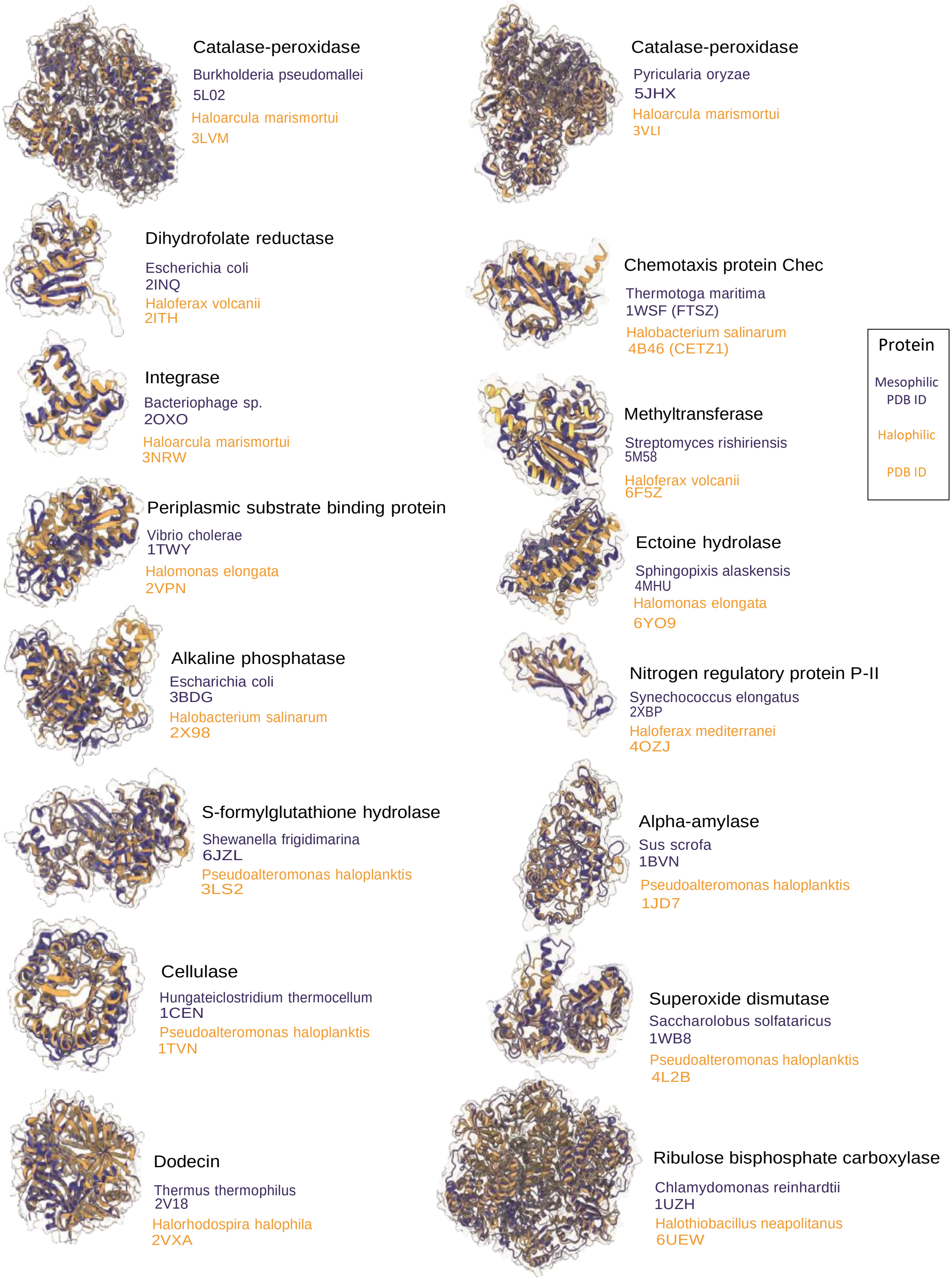


Figure R2. Comparison between the structures of the proteins shown in Table R2 (blue) and their corresponding mesophilic counterparts (orange).

Figure R3 shows the frequency differences that result from the comparison between the proteomes of halophilic and mesophilic organisms performed in agreement with the work from Paul, S., *et al.* (Paul, S., *et al.* 2008). Both groups were previously separated by means of a phylogenetic analysis (Figures R4 and R5). Figure R3 also shows the results of the equivalent analysis performed with the set of 49 proteins shown in Table 2 and their mesophilic counterparts.

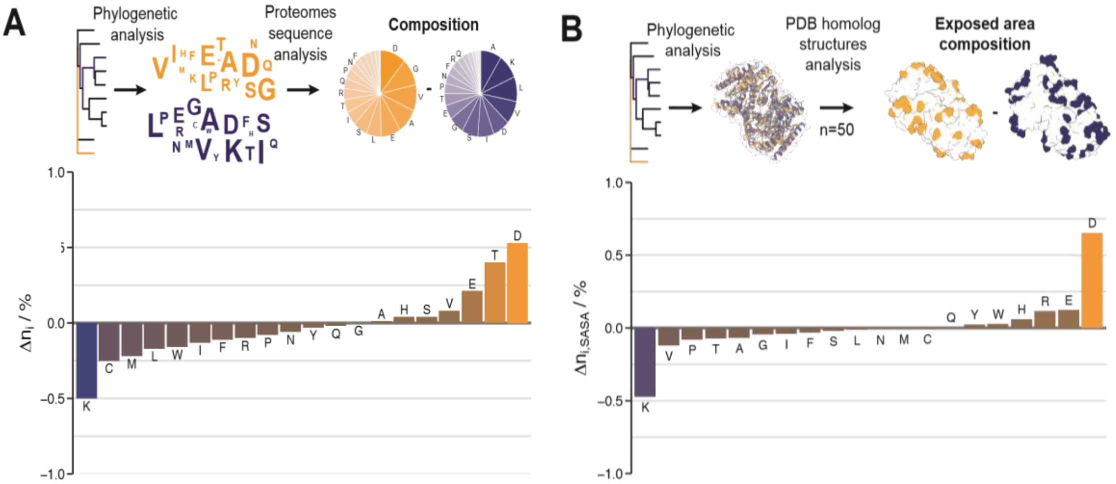


Figure R3. Relative frequency differences, Δn_i, for each aminoacid type obtained as result of the comparison between the proteomes of halophilic (orange) and mesophilic (blue) organisms previously separated by means of a phylogenetic analysis when considering the whole protein (A), and the equivalent frequency differences, $\Delta n_{i,SASA}$, performed with the set of 49 selected protein when considering only residues at the surface that expose more than 60% of surface to the solvent (B).

It can be observed from Figure R3 how residues such as leucines, isoleucines, methionines or tryptophans, which are penalized in halophilic sequences and mainly localized in the protein core, have no relevance when only residues at the surface are considered. In contrast, negatively charged residues such as aspartates or glutamates are enriched in detriment of lysins in protein's surfaces, which supports the idea that they play a significant role in the interaction with the environment.

According to previous observations, it was hypothesized that the differences in surface composition as compared to a consensus halophilic surface could be appropiate for the construction of a metric that allowed the evaluation of the contribution of the protein's surface to the cosolute-induced effect on protein stability. In consequence, the differences in surface composition, $\Delta n_{i,SASA}$, were used for the definition of such metric under the formula below:

$$\sigma_{Prot} = \frac{1}{n_T} \cdot \sum_{i=1}^{20} n_i \cdot \Delta n_{i,SASA}$$

Equation R2. Definition of a metric reporting on the halophilicity degree of a protein based on its primary sequence and weighted by the relative abundances of each residue type in halophilic surfaces, $\Delta n_{i,SASA}$. n_i and n_T respectively represent the aminoacid of type i and the total number of aminoacis of the protein.

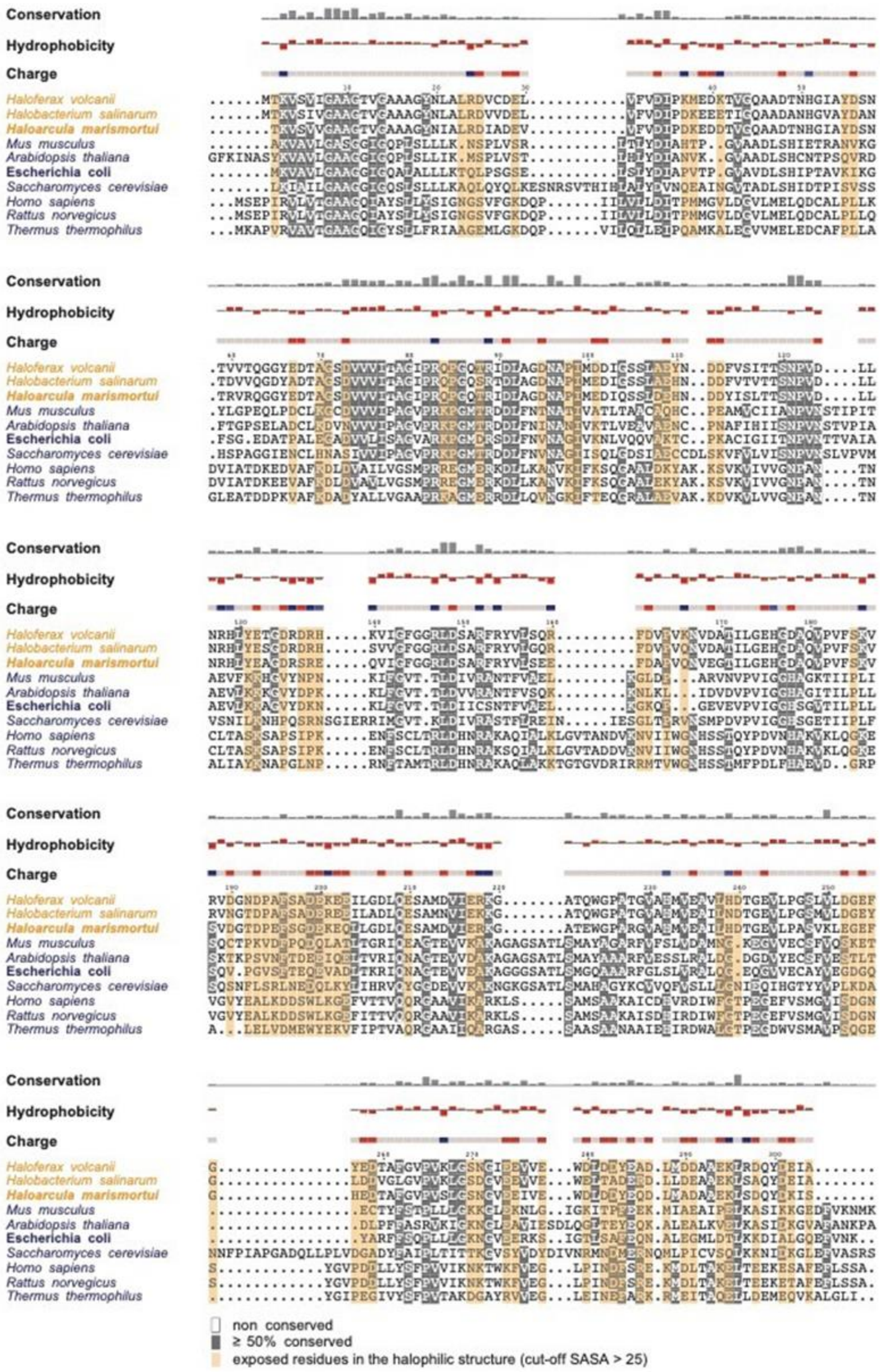


Figure R4. Sequence alignment for malate dehydrogenase from mesophilic (blue) and halophilic (orange) species. Residues highlighted in grey correspond to those with a high similarity degree (> 50%) among the two groups. Residues highlighted in orange correspond to those exposed in the three-dymensional structure of the halophilic protein.

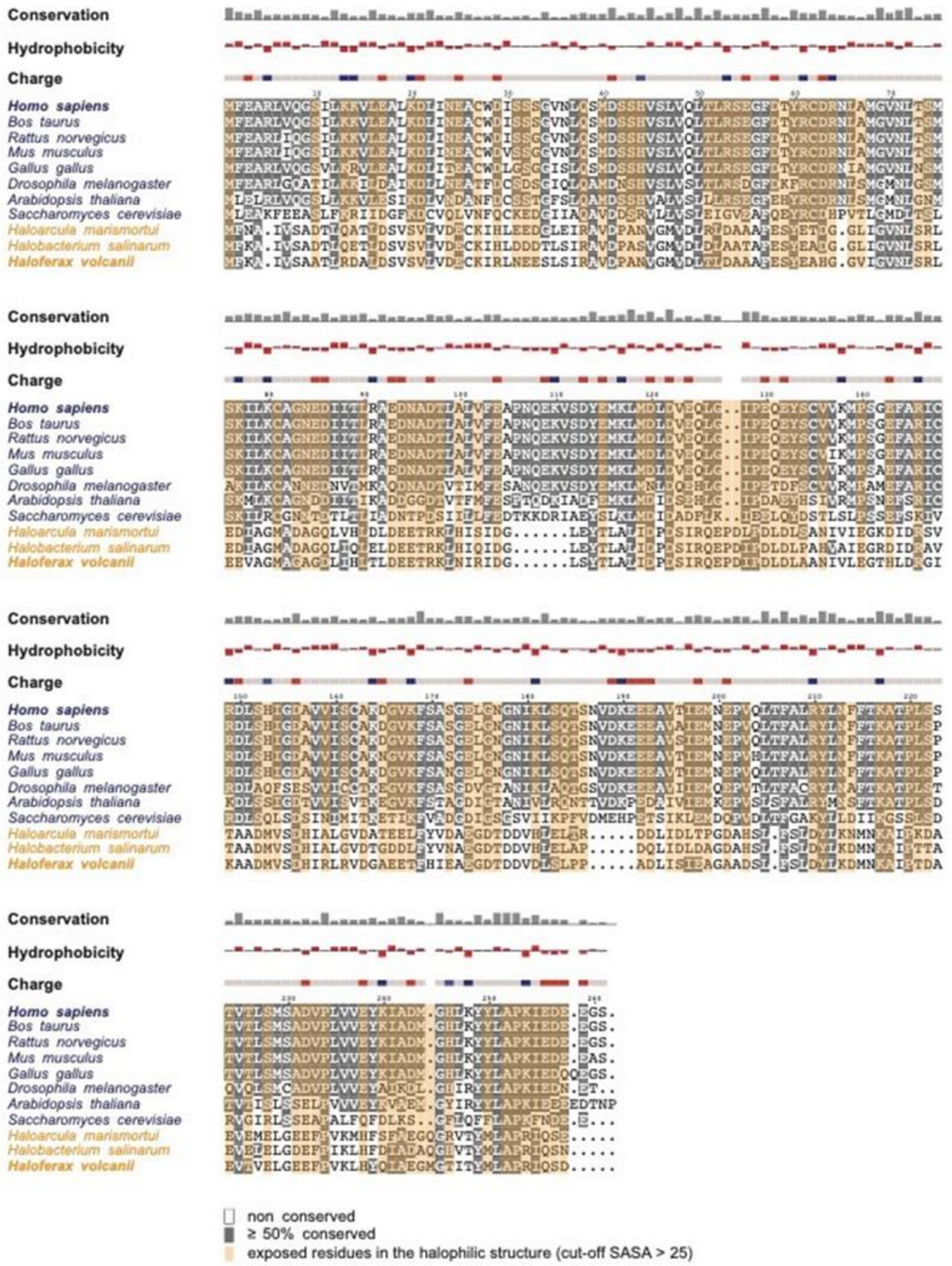


Figure R5. Sequence alignment for proliferating cell nuclear antigen from mesophilic (blue) and halophilic (orange) species. Residues highlighted in grey correspond to those with a high similarity degree (> 50%) among the two groups. Residues highlighted in orange correspond to those exposed in the three-dymensional structure of the halophilic protein.

The only requisite to calculate the value of the metric defined in Equation R1 is a knowledge of the primary sequence of a protein. Since this information is available for many proteins belonging to different organism, the distributions of the σ_{Prot} values for a set of proteins extracted from halophilic and mesophilic proteomes were calculated (Figure R6), filtering out those protein sequences that have been annotated to form oligomers, in order to avoid the contribution of the composition of the surfaces that are implicated in the formation of supramolecular complexes, as has been described in the work from García-Seisdedos, H. *et al.* (García-Seisdedos, H., *et al.* 2017). It was observed from these distributions that the average σ_{Prot} value was centered around a value of zero for mesophilic organisms, wheras deviations towards positive or negative values were found for those organism whose proteome's have evolved for adaptation.

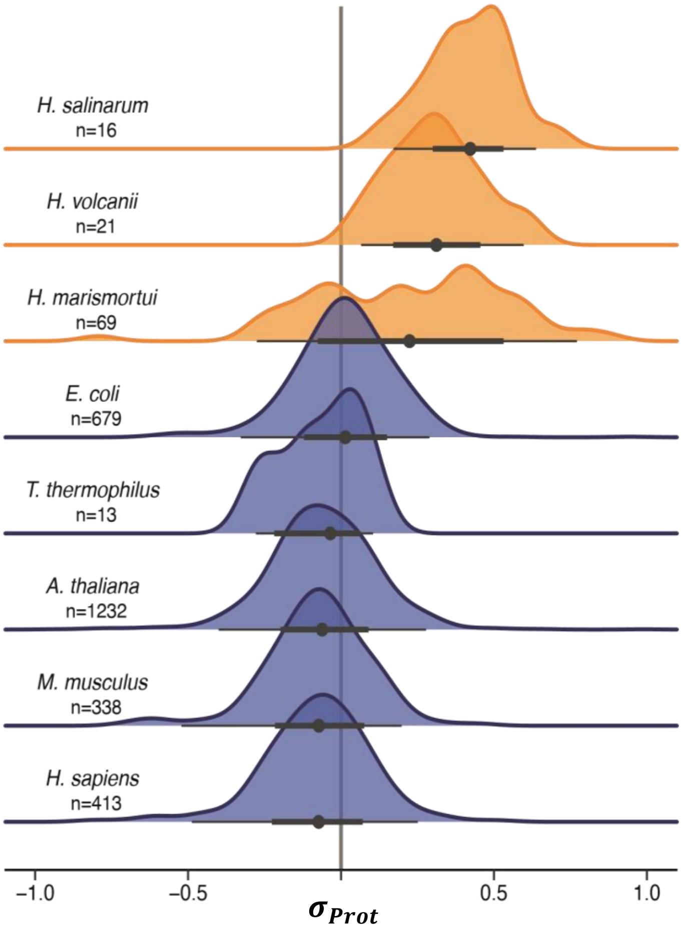


Figure R6. Distribution of the σ_{Prot} values calculated for a pool of monomer proteins extracted from the proteomes of different organisms. The number of proteins, n, considered for the construction of each distribution, and the organisms are indicated.

The σ_{Prot} metric takes in consideration only the surface composition

The average values of the distributions of σ_{Prot} obtained from the proteomes of different organisms reflect their adaptation to the environment

4.2. The effect of cosolutes on protein stability

In the following subsections it is explored how the metric coined in previous section can be applied to rationalize the effect that the cosolutes have on the stability of proteins. First, it is elaborated on construction of the library of proteins used for the study of the influence that a small set of cosolutes (Table R4) has on the stability. Then, it is discused how it can be found a relationship between the composition of the surface of each protein in the library and the observed effect of the set of cosolutes on their stability. Finally, it is developed on how this metric can also be applied to find an equivalent relationship between the effect that cell extracts have on the stability and the composition of the surface.

Table R4. List of the cosolutes used to study the relationship between the metric σ_{Prot} and the effect of cosolutes on protein stability. HS: Hofmeister salt; AA: amino acid; PS: polyol sugar; OS: osmoprotectant.

Cosolute	**Type**	**Reference**
$CaCl_2$	HS	(Tadeo, X., *et al.* 2009; Ortega, G., *et al.* 2015; Tadeo, X., *et al.* 2009)
KCl	HS	
K_2SO_4	HS	
Glycine	AA	(Somero, G., 1986)
Sarcosine	AA	(Somero, G., 1986)
Sucrose	PS	(Keith, S. and Winzor, D., 1988)
Trimethylamine N-oxide	OS	(Jiangiang, M., *et al.* 2014; Pincus, D., *et al.* 2008)
Taurine	AA	(Somero, G., 1986)
Trehalose	PS	(Keith, S. and Winzor, D., 1988)

4.2.1. Building the study protein library

The library of proteins referred at the biginning of this section was created from two separate sets and intended for different analysis. The first set (S1) was composed of proteins, non-necessarily showing reversible unfolding, and whose thermal denaturation was studied by circular dichroism (Figure M4), whereas the second set (S2) was formed by proteins with reversible unfolding and whose thermodynamic stability was studied by NMR (Figure R7). In the design of S1, up to 17 surface variants of the B1 domain of protein L from *Peptostreptococcus magnus*, obtained by site-directed mutagenesis (Tadeo, X., *et al.* 2009) were considered (Table M1, Figure R9). Aditionally, 20 unrelated proteins selected to cover a range as diverse in size (between 5.2 and 75.8 kDa), function (e.g., enzymes, transcription factors, periplasmic binding proteins, signalling pathway comonents, etc.) and origin (from up to 8 different organisms) as possible were added (Table M2). S2 was composed by the mutants of protein L in which seven lysins were subtituted by aspartic acids, and 5 mutants designed to reshape the surface of the SH3 domain of the adapter protein Drk from *Drosophila melanogaster* (Figure R8) (Bezsonova, I., *et al.* 2005)(Table M3). All the mutants obtained by directed mutagenesis preserved the wild type's tertiary structure, in agreement with what was previously observed in the work from Der, B., *et al.* (Der, B., *et al.* 2013).

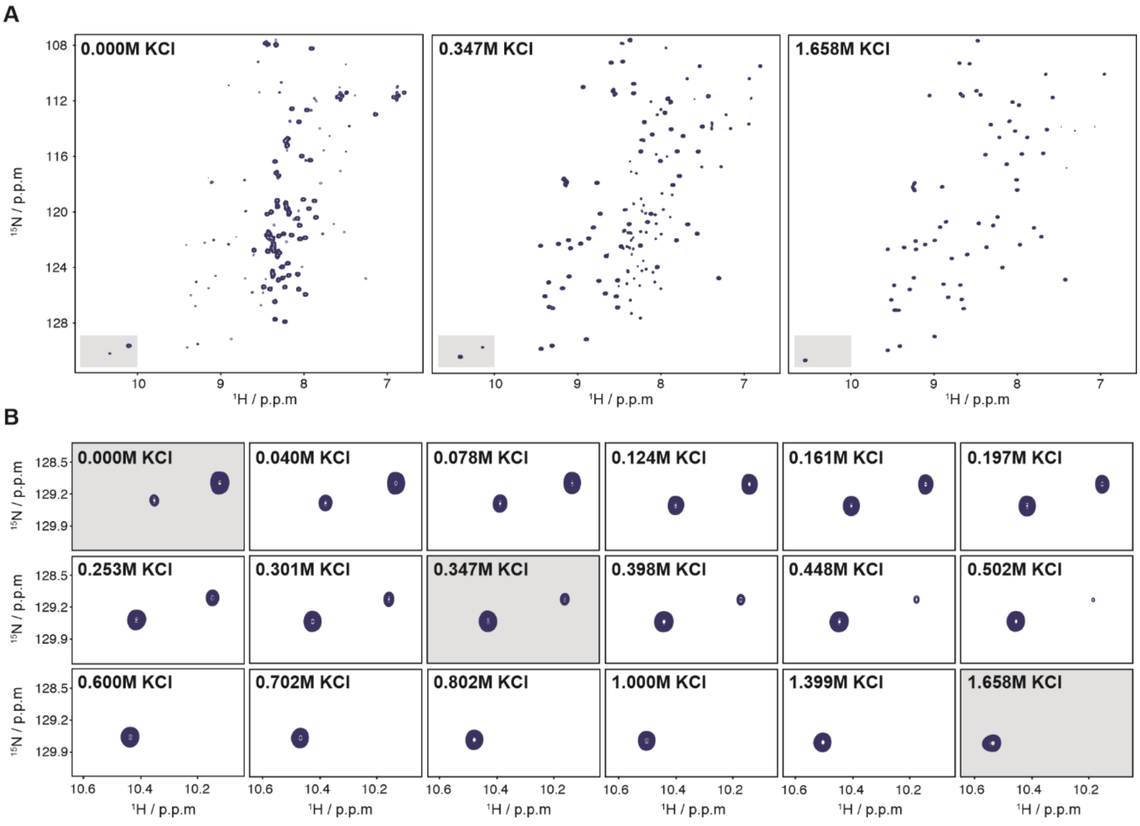


Figure R7. Example of folding monitorization of PL_Kx7E in presence of increasing concentrations of KCl. (A) Intensity changes for the whole ^{1}H ^{15}N HSQC spectra. (B) Intensity changes of the tryptophan in position 47.

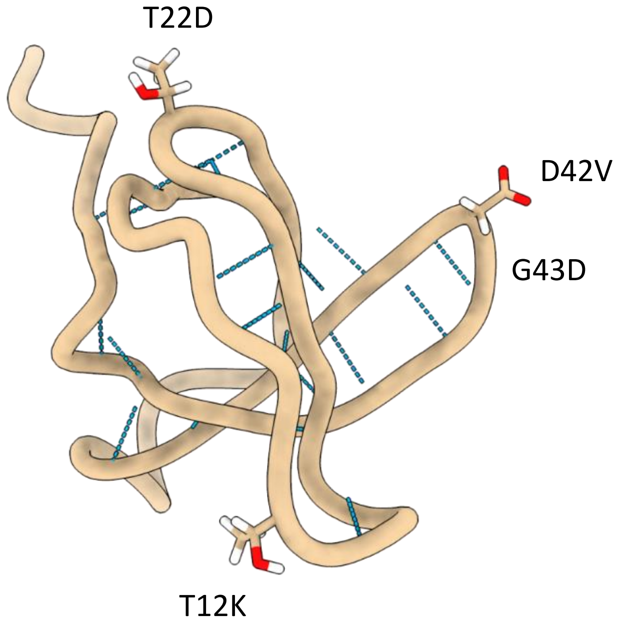


Figure R8. Structure of the SH3 domain of the adapter protein drk from *Drosophila melanogaster*. The residues mutated for the obtention of the 4 different variants used in this work are highlichted.

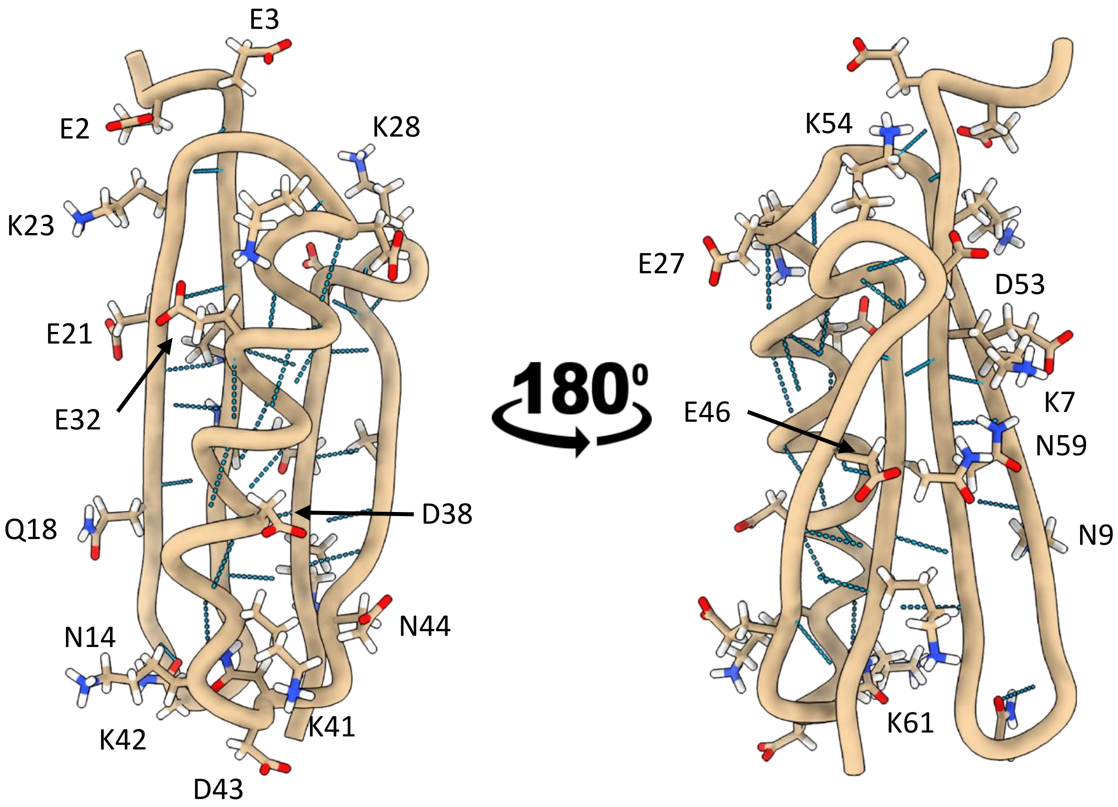


Figure R9. Structure of the B1 domain of PL from Peptostreptococcus magnus. The residues mutated for the obtention of the 17 different variants used in this work are highlichted.

4.2.2. Modulation of the protein's by added cosolutes

The stability of each of the proteins in the library was measured as described in section 3.7 (Tables R5 and R6). Depending on the dataset, thermal (S1) or thermodinamic stability (S2) was measured in presence of increasing cosolute concentrations (Tables M17 and M18). In agreement with prior studies (Schellman, J., 2003), the effect of the cosolutes on the protein's thermal denaturation curve inflection's point, T_m, or unfolding free energy, ΔG°_{F-U}, was observed to almost always follow a linear correlation on cosolute concentration. However, in some cases it was observed a more complex behaviour which was better described by the sum of an exponential and a linear dependences.

For each protein titration with each cosolute, the chosen reporter for the S1 dataset was the slope of the linear correlation of the T_m with the cosolute concentration, $m_{\Delta T_m} = \partial \mathrm{T}_m / \partial [\mathrm{C}]$. Equivalently, for the S2 dataset, the slope of the linear correlation of the unfolding free energy with the cosolute concentration, $m_{\Delta G} = \partial \Delta \mathrm{G}^{\circ}_{F-U} / \partial [\mathrm{C}]$ was used as reporter for protein stability.

Although not necessarily expected, the σ_{Prot} metric showed a linear correlation with the experimental dataset of $m_{\Delta T_m}$-values obatined for KCl (Figures R10 and R11). Different statistical tests (see section 3.5 and Appendix) were used for each correlation to stablish whether the slope was significantly different from zero. In this regard, it was found significant for $CaCl_2$, KCl, TMAO and taurine. It is worth to remark that the correlations between the $m_{\Delta T_m}$-values or $m_{\Delta G}$ show equivalent results as shown in Figures 10 and 11.

Table R5. Experimental thermal denaturation $m_{\Delta T_m}$-values determined by CD. Bold number indicate a non-linear fitting to the experimental data. Error bars were obtained from duplicates.

Acronym	**KCl**	**Sucrose**	**TMAO**	**Sarcosine**	**Glycine**	**Taurine**	**Trehalose**	**K_2SO_4**	**$CaCl_2$**
PL_WT	**1.3 ± 0.1**	6.7 ± 0.4	**5.6 ± 0.3**	5.3 ± 0.3	9.7 ± 0.5	3.8 ± 0.2	11.8 ± 0.7	11.2 ± 0.3	-13.7 ± 0.3
PL_Kx5E	**4.3 ± 0.6**	10.0 ± 0.5	8.3 ± 0.2	8.2 ± 0.6	8.0 ± 0.4	**4.9 ± 0.4**	12.4 ± 0.8	31.4 ± 0.4	4.3 ± 0.2
PL_Kx6E	**2.0 ± 0.5**	n.a.	8.8 ± 0.3	11.3 ± 0.8	4.5 ± 0.3	-18.2 ± 0.7	10.2 ± 0.7	n.a.	n.a.
PL_Kx5Q	1.5 ± 0.8	8.2 ± 0.3	6.2 ± 0.5	7.0 ± 0.4	6.6 ± 0.4	12.6 ± 0.2	12.9 ± 0.8	n.a.	n.a.
PL_Ex4D	**3.1 ± 0.2**	7.4 ± 0.4	**5.8 ± 0.2**	6.2 ± 0.3	7.9 ± 0.4	-27.6 ± 0.5	10.7 ± 0.6	0.5 ± 0.6	-27.9 ± 0.4
PL_Ex5D	**2.4 ± 0.3**	12.5 ± 0.8	5.7 ± 0.2	4.5 ± 0.2	5.4 ± 0.3	**1.4 ± 0.4**	9.7 ± 0.6	9.5 ± 0.3	n.a
PL_Ex6D	**2.8 ± 0.4**	9.3 ± 0.3	4.1 ± 0.2	2.8 ± 0.2	3.9 ± 0.2	0.9 ± 0.9	9.7 ± 0.6	n.a	n.a
PL_Kx4E	**4.3 ± 0.2**	**8.1 ± 0.2**	**6.7 ± 0.2**	9.6 ± 0.7	8.0 ± 0.4	-2.0 ± 0.8	**7.9 ± 0.5**	34.0 ± 0.4	-14.6 ± 0.7
PL_K42E	**0.4 ± 0.2**	7.6 ± 0.5	6.6 ± 0.2	4.0 ± 0.2	1.9 ± 0.2	**-5.3 ± 0.3**	6.9 ± 0.3	19.5 ± 0.2	-14.7 ± 0.3
PL_Kx3E	4.0 ± 0.1	7.7 ± 0.2	4.4 ± 0.1	**7.4 ± 0.3**	**7.7 ± 0.4**	6.97 ± 0.2	10.8 ± 0.8	4.0 ± 0.6	n.a.
PL_DNx8EQ	**-3.0 ± 0.4**	4.2 ± 0.1	-0.7 ± 0.4	5.5 ± 0.3	n.a.	n.a.	n.a.	n.a.	n.a.
PL_DNx7EQalt	-4.4 ± 0.6	7.3 ± 0.6	1.9 ± 0.3	7.0 ± 0.4	6.8 ± 0.4	21.0 ± 0.3	7.5 ± 0.4	n.a.	n.a.
PL_DNx7EQ	-3.8 ± 0.5	7.2 ± 0.4	1.5 ± 0.2	5.0 ± 0.4	6.1 ± 0.3	19.0 ± 0.3	6.8 ± 0.4	5.3 ± 0.7	-33.8 ± 0.9
PL_DNx6EQ	-3.0 ± 0.4	7.9 ± 0.4	3.3 ± 0.3	6.0 ± 0.4	n.a.	n.a.	n.a.	n.a.	n.a.
PL_Kx5S	7.4 ± 0.5	7.6 ± 0.5	6.6 ± 0.1	5.8 ± 0.4	7.5 ± 0.4	10.6 ± 0.2	8.2 ± 0.6	n.a.	n.a.
PL_Kx5R	4.0 ± 1.0	6.7 ± 0.3	6.7 ± 0.2	8.5 ± 0.5	6.0 ± 0.3	1.0 ± 0.3	11.4 ± 0.8	n.a.	n.a.
PL_Kx2E	**1.3 ± 0.4**	7.4 ± 0.5	5.6 ± 0.2	**4.0 ± 0.2**	5.7 ± 0.3	-8.7 ± 0.5	9.5 ± 0.5	15.7 ± 0.3	-17.7 ± 0.3
PL_Kx7R	6.5 ± 0.3	10.1 ± 0.1	13.4 ± 0.3	4.4 ± 0.5	4.5 ± 0.6	-31.5 ± 0.5	n.a.	15.5 ± 0.5	15.1 ± 0.3
PL_DEx5K	**-5.8 ± 0.4**	6.3 ± 0.4	2.0 ± 0.2	7.0 ± 0.4	3.7 ± 0.8	33.1 ± 0.8	7.7 ± 0.7	16.2 ± 0.8	-52.5 ± 0.9
PL_DEx6Kalt	-3.7 ± 0.4	6.9 ± 0.5	-8.5 ± 0.7	2.4 ± 0.4	-0.2 ± 0.4	26.0 ± 0.7	9.5 ± 0.4	8.3 ± 0.6	**-30.0 ± 1.5**
UROIIIS	3.6 ± 1.1	4.9 ± 0.1	3.0 ± 0.1	4.7 ± 0.3	5.7 ± 0.3	15.7 ± 0.6	3.3 ± 0.3	**4.9 ± 0.4**	-40.0 ± 0.5
RBP	1.5 ± 0.5	8.7 ± 0.5	8.3 ± 0.2	8.0 ± 0.5	6.6 ± 0.4	-2.1 ± 0.8	11.3 ± 0.7	16.3 ± 0.8	-34.8 ± 0.6
YmoA	**6.4 ± 0.5**	4.1 ± 0.2	**0.0 ± 0.4**	4.7 ± 0.3	4.5 ± 0.3	-35.9 ± 0.4	**2.0 ± 0.4**	n.a.	n.a.
OIH2	**6.3 ± 0.3**	8.0 ± 0.5	**5.7 ± 0.6**	4.8 ± 0.4	4.0 ± 0.2	**-4.3 ± 0.2**	6.3 ± 0.5	3.9 ± 0.4	12.9 ± 0.6
GGBP	2.0 ± 0.4	6.1 ± 0.3	5.3 ± 0.1	5.2 ± 0.3	**5.9 ± 0.4**	-21.0 ± 0.6	12.3 ± 0.8	10.4 ± 0.3	-20.3 ± 0.5
HvCopG	3.5 ± 0.3	8.3 ± 0.4	10.0 ± 0.4	4.2 ± 0.2	**6.2 ± 0.4**	-16.2 ± 0.9	10.5 ± 1.1	n.a.	n.a.
1ALigN	4.0 ± 0.5	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
BtGH99	**1.0 ± 0.4**	3.6 ± 0.3	**3.0 ± 0.5**	-1.0 ± 0.4	4.4 ± 0.2	-15.7 ± 0.3	4.2 ± 0.4	n.a.	n.a.
CAM	**4.6 ± 0.8**	**7.0 ± 0.4**	11.6 ± 0.2	3.4 ± 0.2	**5.7 ± 0.3**	-19.9 ± 0.6	**10.0 ± 0.5**	n.a.	n.a.
FAH	2.6 ± 0.4	4.2 ± 0.2	3.7 ± 0.2	9.5 ± 0.3	3.3 ± 0.2	**-9.6 ± 0.5**	5.1 ± 0.3	0.4 ± 0.2	-9.0 ± 0.7
GSTZ1	4.5 ± 0.2	7.0 ± 0.6	6.4 ± 0.3	7.1 ± 0.4	**6.4 ± 0.4**	**12.0 ± 0.4**	4.8 ± 0.4	8.6 ± 0.5	-15.8 ± 0.5
HGO	1.0 ± 0.3	4.3 ± 0.2	6.8 ± 0.2	3.4 ± 0.2	3.8 ± 0.2	-10.4 ± 0.8	5.8 ± 0.4	7.8 ± 0.5	-15.8 ± 0.3
UROD	**2.2 ± 0.3**	**6.5 ± 0.3**	1.4 ± 0.4	2.6 ± 0.2	**3.2 ± 0.2**	-16.4 ± 0.5	**8.7 ± 0.6**	14.6 ± 0.3	**-20.0 ± 0.6**
FTX	5.1 ± 0.2	5.8 ± 0.3	2.6 ± 0.3	3.1 ± 0.2	3.9 ± 0.2	**-22.4 ± 0.4**	4.2 ± 0.4	17.8 ± 0.2	-5.8 ± 0.6
CPOX	-0.1± 0.2	5.5 ± 0.3	4.0 ± 0.4	**6.9 ± 0.5**	**5.8 ± 0.3**	**-14.0 ± 0.6**	**10.3 ± 0.7**	-1.0 ± 0.4	-10.1 ± 0.5
MBP	**-1.2 ± 0.2**	11.6 ± 0.7	3.6 ± 0.3	2.6 ± 0.2	8.0 ± 0.4	13.4 ± 0.9	8.3 ± 0.6	10.1 ± 0.5	-20.4 ± 0.7
GST	**1.0 ± 0.4**	4.3 ± 0.1	2.3 ± 0.2	5.5 ± 0.3	5.6 ± 0.4	-4.5 ± 0.4	1.9 ± 0.2	3.3 ± 0.3	**7.0 ± 0.5**
Nras	1.5 ± 0.5	**14.8 ± 0.8**	6.6 ± 0.3	2.6 ± 0.2	-3.4 ± 0.3	-5.3 ± 0.3	12.8 ± 0.8	**-3.5 ± 0.2**	4.5 ± 0.3
ALAD	4.7 ± 0.5	8.1 ± 0.4	10.7 ± 0.4	8.9 ± 0.5	7.3 ± 0.4	-31.8 ± 0.6	7.4 ± 0.4	7.5 ± 0.7	-8.1 ± 0.4
RNaseA	2.3 ± 0.2	7.0 ± 0.4	6.4 ± 0.5	2.9 ± 0.2	**3.7 ± 0.8**	**5.4 ± 0.5**	11.7 ± 0.9	14.4 ± 0.3	-6.2 ± 0.4

Table R6. Experimental thermodynamic $m_{\Delta G_{F-U}}$-values determined by either circular dichroism or NMR spectroscopy. Error bars were obtaines from duplicates.

Acronym	**KCl**	**Sucrose**	**TMAO**	**Sarcosine**	**Glycine**	**Taurine**	**Trehalose**	**K_2SO_4**	**$CaCl_2$**
PL_WT	0.20 ± 0.06	1.04 ± 0.05	0.87 ± 0.10	0.81 ± 0.05	1.49 ± 0.13	0.59 ± 0.05	1.81 ± 0.12	1.73 ± 0.60	-2.12 ± 0.50
PL_Kx5E	0.73 ± 0.20	1.71 ± 0.05	1.40 ± 0.22	1.39 ± 0.09	1.36 ± 0.10	0.83 ± 0.12	2.11 ± 0.14	5.33 ± 0.40	0.73 ± 0.40
PL_Kx6E	0.35 ± 0.14	n.a.	1.54 ± 0.30	1.98 ± 0.12	0.79 ± 0.07	-3.17 ± 0.24	1.78 ± 0.12	n.a.	n.a.
PL_Kx5Q	-0.24 ± 0.30	1.30 ± 0.10	0.98 ± 0.25	1.10 ± 0.05	1.04 ± 0.05	1.99 ± 0.04	2.03 ± 0.13	n.a.	n.a.
PL_Ex4D	0.48 ± 0.07	1.15 ± 0.20	0.90 ± 0.10	n.a.	n.a.	-4.26 ± 0.18	n.a.	0.07 ± 0.20	-4.20 ± 0.60
PL_Ex5D	0.39 ± 0.11	2.02 ± 0.30	0.91 ± 0.12	0.73 ± 0.05	0.87 ± 0.09	0.23 ± 0.08	1.57 ± 0.09	1.53 ± 0.30	n.a.
PL_Ex6D	0.45 ± 0.16	1.51 ± 0.06	0.66 ± 0.07	0.46 ± 0.03	0.63 ± 0.08	0.14 ± 0.03	1.56 ± 0.10	n.a.	n.a.
PL_Kx4E	0.69 ± 0.08	1.28 ± 0.07	1.06 ± 0.11	1.52 ± 0.11	1.27 ± 0.12	-0.32 ± 0.05	1.25 ± 0.08	5.39 ± 0.30	-2.31 ± 0.50
PL_K42E	0.06 ± 0.09	1.18 ± 0.04	1.02 ± 0.17	0.63 ± 0.05	0.60 ± 0.07	-0.83 ± 0.06	1.08 ± 0.08	3.03 ± 0.60	-2.29 ± 0.30
PL_Kx3E	0.66 ± 0.03	1.26 ± 0.08	0.72 ± 0.03	1.22 ± 0.07	1.26 ± 0.09	1.14 ± 0.12	1.78 ± 0.09	0.66 ± 0.20	n.a.
PL_DNx8EQ	-0.47 ± 0.19	0.65 ± 0.04	0.50 ± 0.12	n.a	n.a.	n.a.	n.a.	n.a.	n.a.
PL_DNx7EQt	-0.69 ± 0.31	1.12 ± 0.03	0.30 ± 0.09	1.08 ± 0.05	1.05 ± 0.07	3.24 ± 0.24	1.16 ± 0.08	n.a.	n.a.
PL_DNx7EQ	-0.59 ± 0.25	1.11 ± 0.03	0.23 ± 0.05	0.77 ± 0.05	0.94 ± 0.09	2.93 ± 0.20	1.05 ± 0.07	n.a.	-5.20 ± 0.50
PL_DNx6EQ	-0.47 ± 0.20	1.25 ± 0.05	0.52 ± 0.13	n.a.	n.a.	n.a.	n.a.	0.84 ± 0.40	n.a.
PL_Kx5S	0.79 ± 0.28	1.20 ± 0.04	1.00 ± 0.03	0.92 ± 0.08	1.18 ± 0.12	1.67 ± 0.15	1.29 ± 0.09	n.a.	n.a.
PL_Kx5R	-0.08 ± 0.40	1.06 ± 0.05	1.06 ± 0.04	1.36 ± 0.12	0.95 ± 0.06	0.15 ± 0.05	1.81 ± 0.12	n.a.	n.a.
PL_Kx2E	0.21 ± 0.20	1.17 ± 0.05	0.88 ± 0.06	0.63 ± 0.02	0.90 ± 0.06	-1.37 ± 0.09	1.49 ± 0.09	2.49 ± 0.30	-2.80 ± 0.70
PL_Kx7E	0.58 ± 0.18	0.85 ± 0.02	1.20 ± 0.11	1.43 ± 0.07	1.21 ± 0.20	1.30 ± 0.05	1.60 ± 0.15	n.a.	n.a.
PL_Kx7R	1.26 ± 0.09	1.55 ± 0.08	2.05 ± 0.12	0.69 ± 0.09	0.69 ± 0.08	-4.85 ± 0.40	n.a.	2.38 ± 0.30	2.32 ± 0.40
PL_DEx5K	-0.90 ± 0.50	0.98 ± 0.20	0.31 ± 0.06	1.09 ± 0.08	0.58 ± 0.11	5.15 ± 0.30	1.19 ± 0.04	2.52 ± 0.10	-8.18 ± 0.60
PL_DEx6EKa	-0.57 ± 0.90	1.07 ± 0.15	-1.32 ± 0.21	0.37 ± 0.10	-0.03 ± 0.08	4.03 ± 0.18	1.47 ± 0.20	1.28 ± 0.20	-4.65 ± 0.80
drk_WT	0.97 ± 0.30	0.84 ± 0.02	0.88 ± 0.03	0.42 ± 0.03	1.50 ± 0.12	-0.67 ± 0.03	0.56 ± 0.04	0.60 ± 0.70	5.00 ± 0.90
drk_m1	0.95 ± 0.35	1.30 ± 0.06	1.52 ± 0.12	1.09 ± 0.08	1.32 ± 0.05	-1.42 ± 0.14	1.17 ± 0.08	2.75 ± 0.50	n.a.
drk_m2	1.10 ± 0.20	0.87 ± 0.04	1.21 ± 0.14	0.89 ± 0.05	0.76 ± 0.04	-1.45 ± 0.12	1.51 ± 0.10	2.18 ± 0.10	n.a.
drk_m3	0.96 ± 0.30	1.25 ± 0.04	1.63 ± 0.20	0.70 ± 0.05	1.45 ± 0.07	-2.85 ± 0.20	1.30 ± 0.06	1.31 ± 0.30	9.00 ± 0.80
drk_m4	0.36 ± 0.20	1.08 ± 0.03	1.07 ± 0.11	0.75 ± 0.05	0.78 ± 0.04	-0.70 ± 0.07	1.66 ± 0.12	2.51 ± 0.40	7.00 ± 0.90

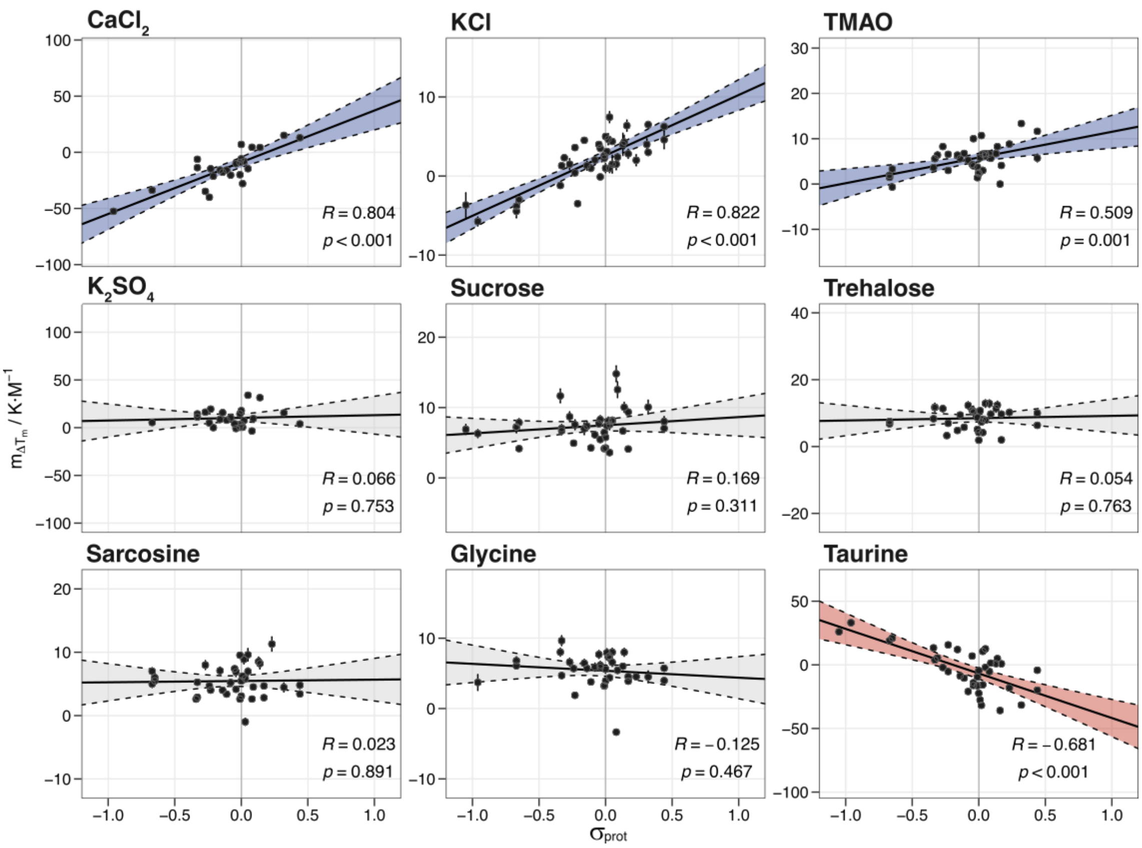


Figure R10. Correlation between the thermal stability, $m_{\Delta T_m}$, and σ_{Prot} for each of the cosolutes shown in Table R4. Linear correlations are showed as dark lines. Shaded regions correspond to the 95% confidence interval. Blue and red shaded regions correspond to possitive or negative correlations ($|R| > 0.5$ and $p < 0.5$), respectively. Grey circles represent the experimental points. Error bars represent the experimental variability, and were calculated as described in section 3.5. For each correlation, the corresponding Pearson coefficient and its associated p-value are shown. The correlation slopes, intercepts and statistics are shown in Table R7.

Table R7. Satatistic parameters for the correlations corresponding to each cosolute shown in Table R4 obtained in thermal denaturation experiments, the corresponding slopes, η_{cosol}, and intercepts, m_0, with their corresponding p-values.

Cosolute	**Pearson R (p-value)**	**Sepearman R (p-value)**	**η_{cos} (p-value)**	**p < a**	**m_0 (p-value)**	**p < a**
$CaCl_2$	$0.804 (2.16 \cdot 10^{-6})$	$0.689 (1.97 \cdot 10^{-1})$	46.10 ± 7.25 $(2.16 \cdot 10^{-6})$	***	-8.85 ± 2.20 $(6.33 \cdot 10^{-11})$	***
KCl	$0.821 (2.60 \cdot 10^{-11})$	$0.709 (1.45 \cdot 10^{-7})$	7.64 ± 0.84 $(2.60 \cdot 10^{-11})$	***	2.61 ± 0.30 $(5.67 \cdot 10^{-4})$	***
K_2SO_4	$0.066 (7.53 \cdot 10^{-1})$	$-0.047 (8.22 \cdot 10^{-1})$	2.79 ± 8.78 $(7.53 \cdot 10^{-1})$	-	10.30 ± 2.00 $(3.30 \cdot 10^{-5})$	***
Glycine	$-0.125 (4.67 \cdot 10^{-1})$	$-0.176 (3.06 \cdot 10^{-1})$	-0.98 ± 1.34 $(4.67 \cdot 10^{-1})$	-	5.37 ± 0.39 $(1.84 \cdot 10^{-15})$	***
Sarcosine	$0.023 (8.91 \cdot 10^{-1})$	$0.028 (8.69 \cdot 10^{-1})$	0.21 ± 1.50 $(8.91 \cdot 10^{-1})$	-	5.47 ± 0.43 $(8.43 \cdot 10^{-15})$	***
Sucrose	$0.169 (1.29 \cdot 10^{-1})$	$0.251 (1.28 \cdot 10^{-1})$	1.16 ± 1.13 $(3.11 \cdot 10^{-1})$	-	7.48 ± 0.41 $(7.94 \cdot 10^{-20})$	***
Trimethylamine N-oxide	$0.509 (1.29 \cdot 10^{-3})$	$0.421 (9.38 \cdot 10^{-3})$	5.67 ± 1.62 $(1.29 \cdot 10^{-3})$	***	5.85 ± 0.46 $(1.04 \cdot 10^{-14})$	***
Taurine	$-0.681 (3.44 \cdot 10^{-6})$	$-0.520 (9.65 \cdot 10^{-4})$	-35.00 ± 6.36 $(3.44 \cdot 10^{-6})$	***	-6.76 ± 2.12 $(3.02 \cdot 10^{-3})$	**
Trehalose	$0.053 (7.63 \cdot 10^{-1})$	$0.074 (6.77 \cdot 10^{-1})$	0.69 ± 2.28 $(7.63 \cdot 10^{-1})$	-	8.48 ± 0.56 $(3.54 \cdot 10^{-16})$	***

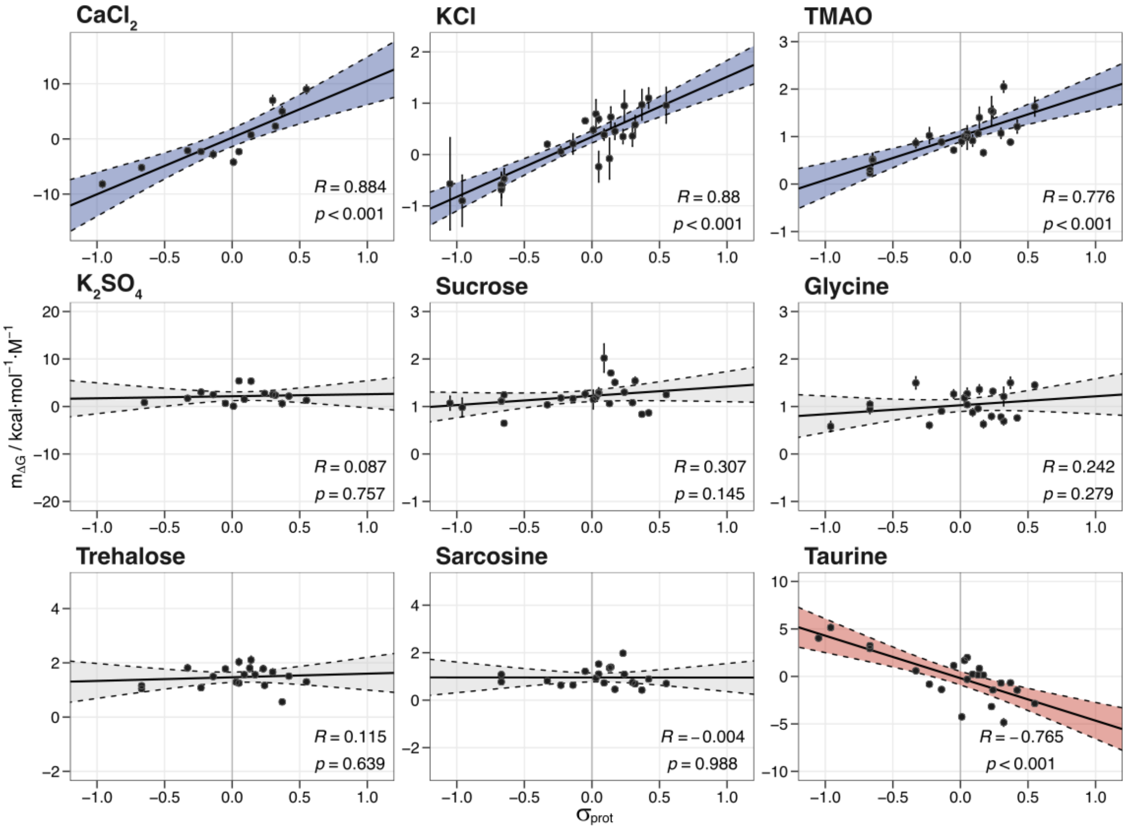


Figure R11. Correlation between the thermodynamic stability, $m_{\Delta G}$, and σ_{Prot} for each of the cosolutes shown in Table R4. Linear correlations are showed as dark lines. Shaded regions correspond to the 95% confidence interval. Blue and red shaded regions correspond to possitive or negative correlations ($|R| > 0.5$ and $p < 0.5$), respectively. Grey circles represent the experimental points. Error bars represent the experimental variability, and were calculated as described in section 3.5. For each correlation, the corresponding Pearson coefficient and its associated p-value are shown. The correlation slopes, intercepts and statistics are shown in Table R8.

Table R8. Satatistic parameters for the correlations corresponding to each cosolute shown in Table R4 obtained in thermodynamic stability experiments, and the corresponding slopes, η_{cosol}, and intercepts, m_0, with their corresponding p-values.

Cosolute	**Pearson R (p-value)**	**Sepearman R (p-value)**	**η_{cos} (p-value)**	**p < a**	**m_0 (p-value)**	**p < a**
$CaCl_2$	0.884 ($1.36 \cdot 10^{-4}$)	0.860 ($5.97 \cdot 10^{-4}$)	10.30 ± 1.72 ($1.36 \cdot 10^{-4}$)	***	0.25 ± 0.74 ($7.42 \cdot 10^{-1}$)	-
KCl	0.880 ($6.71 \cdot 10^{-9}$)	0.807 ($1.07 \cdot 10^{-6}$)	1.17 ± 0.13 ($6.71 \cdot 10^{-9}$)	***	0.35 ± 0.06 ($5.26 \cdot 10^{-6}$)	***
K_2SO_4	0.087 ($7.57 \cdot 10^{-1}$)	-0.004 ($9.95 \cdot 10^{-1}$)	0.43 ± 1.35 ($7.76 \cdot 10^{-1}$)	-	2.16 ± 0.43 ($2.24 \cdot 10^{-4}$)	***
Glycine	0.242 ($2.79 \cdot 10^{-1}$)	0.155 ($4.91 \cdot 10^{-1}$)	0.19 ± 0.17 ($2.75 \cdot 10^{-1}$)	-	1.03 ± 0.06 ($6.14 \cdot 10^{-13}$)	***
Sarcosine	-0.004 ($9.88 \cdot 10^{-1}$)	-0.138 ($5.60 \cdot 10^{-1}$)	0.01 ± 0.28 ($9.88 \cdot 10^{-1}$)	-	0.96 ± 0.09 ($3.73 \cdot 10^{-9}$)	***
Sucrose	0.307 ($1.45 \cdot 10^{-1}$)	0.304 ($1.49 \cdot 10^{-1}$)	0.20 ± 0.13 ($1.45 \cdot 10^{-1}$)	-	1.22 ± 0.06 ($4.59 \cdot 10^{-16}$)	***
Trimethylamine N-oxide	0.776 ($1.36 \cdot 10^{-5}$)	0.771 ($1.67 \cdot 10^{-5}$)	0.92 ± 0.16 ($1.36 \cdot 10^{-5}$)	***	1.01 ± 0.06 ($6.55 \cdot 10^{-14}$)	***
Taurine	-0.765 ($2.09 \cdot 10^{-5}$)	-0.684 ($3.23 \cdot 10^{-4}$)	-4.47 ± 0.82 ($2.09 \cdot 10^{-5}$)	***	-0.19 ± 0.35 ($5.97 \cdot 10^{-1}$)	-
Trehalose	0.115 ($6.39 \cdot 10^{-1}$)	0.111 ($6.52 \cdot 10^{-1}$)	0.13 ± 0.28 ($6.39 \cdot 10^{-1}$)	-	1.47 ± 0.10 ($9.26 \cdot 10^{-12}$)	***

Figure R11 includes the values of unfolding free energy for each of the protein L mutants shown in Table M1, which were calculated from their previously measured heat capacity (Tadeo, X., *et al.* 2009). On the other hand, the unfolding free energy for drk_WT and all derived mutants was determined as described in section 3.7. As previusly commented, the correlations that resulted statistically significant in Figures R10 and R11 were debided to a definition of the σ_{Prot} metric in which only the surface was considered was reinforced by the fact that an equivalent correlation analysis performed using similar metric that took in consideration the whole protein resulted in the absence of correlation ($|R| < 0.5$) (Figure R12).

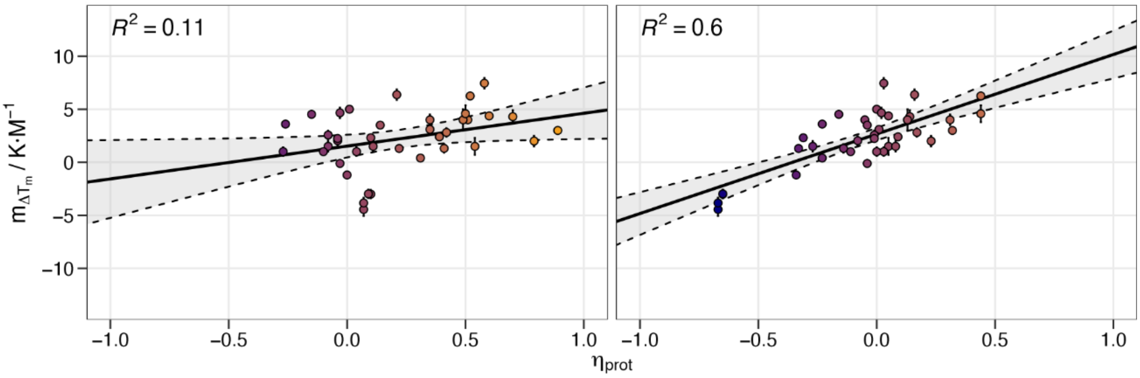


Figure R12. Comparison between the correlation analsys obtained for a metric considering the whole protein, Δn_i, (left) and the metric considering only the surface, $\Delta n_{i,SASA}$, (right).

One indication on the putative generality of the adaptation mechanism of the surface is the wide range of σ_{Prot} values obtained for the experimental library considered in this work. It can be observed from Figure R13 how the range of σ_{Prot} values of the library includes the ranges observed for the proteomes of different organisms.

The σ_{Prot} metric linearly correlates with the effect of a set of cosolutes on the stability of proteins

This correlation arises as a consequence of the surface composition, and significantly weakens when the whole protein composition is considered

The protein datasets considered in this work (S1 and S2) covers the biologically relevant range of values of the σ_{Prot} metric

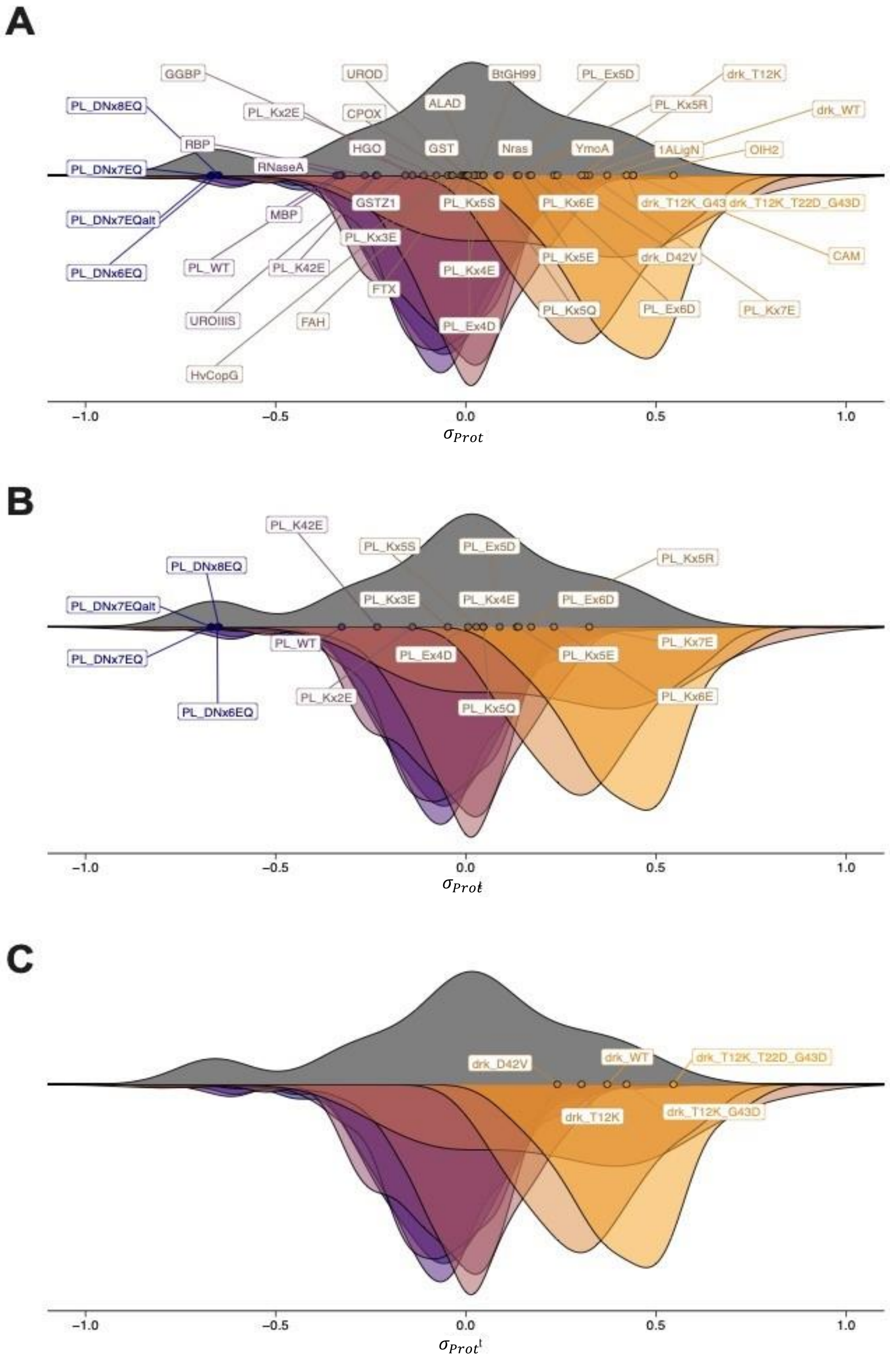


Figure R13. Distributions of the σ_{Prot} metric obtained for the protein library used in this work and the organisms considered in Figure R3 (*H. sapiens*, *A. thaliana*, *T. termophilus*, *E. coli*, *H. marismortui*, *H. volcanii*, and *H. salinarum*).

The following step was to statistically validate the regression model obtained for TMAO, taurine, KCl and $CaCl_2$. The results of the analysis of homocedasticity and normality distribution of the residuals for each of these cosolutes is showin in the Appendix. For each of these analysis the experimental points, represented as green points, should be inside the grey-shaded regions, and in the linearity analysis the experimental residuals should also be aligned with the black dashed line. Furthermore, the distribution of the residuals, shown as a blue-shaded region in the last plot of each analysis, should be fitted to the solid green line.

Complementary, in the independence observation analysis the standard deviations of the residuals should be inside the region delimited by the dashed green line, and their fitting curve, represented as a solid green line, shoud be straight and aligned with the dashed grey line, which represents the mean of the distribution.

Despite the deviations with respect to the theoretical models in each of the previous analysis, it was considered reasonable to accept the regression models obtained for each of the corresponding cosolutes. Consequently, it was accepted that KCl, $CaCl_2$ and TMAO present positive correlations with the σ_{Prot} metric, whereas taurine shows a negative correlation, and that the regression models corresponding to the correlation between the metric and the m-values from either the thermal or thermodynamic stability measurements take the general form shown below:

$$m_{\Delta T_m} = m_0 + \eta_{cos} \cdot \sigma_{Prot} \qquad\qquad m_{\Delta G} = m_0 + \eta_{cos} \cdot \sigma_{Prot}$$

Equations R3 and R4. Regression models for the correlations between the thermal (left) or thermodynamic (right) stability and the σ_{Prot} metric. m_0 and η_{cos} are, respectively, the intercept for an average mesophilic protein ($\sigma_{Prot} = 0$) and the slope.

4.2.3. Influence of cosolutes on quinaty structure

For the four cosolutes, for which $\eta_{cos} \neq 0$, the intercept can be understood as the stabilization produced by the cosolute on an average mesophilic protein, whereas the slope indicates that the cosolute modulates the protein stability through a mechanism dependent on the composition of the surface. As expected from their possitive correlations, the η_{cos} slope was possitive for KCl and $CaCl_2$ (Tables R7 and R8), which means that these cosolutes have a greater stabilizing effect on proteins with a more halophilic character. This fact can be rationalized by arguing that these cosolutes probably exert their stabilizing effect through a non-specific and electrostatically-driven preferential interaction mechanism, mainly debided to the cation. This hypothesis also explains the fact that $\eta_{CaCl_2} \gg \eta_{KCl}$ if the Hofmeister series are taken in consideration (Baldwin, R., 1996), and if it is taken into account that the charge density of Ca^{2+} is greater than that of K^+. Similarly, TMAO also showed a possitive value of η_{cos}, which could be expected from the fact that nature regularly uses it to compensate osmotic stress in halophilic and other extremophilic cytosols (Yancey, P., *et al.* 1982; Somero, G., 1986). According to this, the stabilizing effect of TMAO also depends on the composition of the surface, and similarly to the case of KCl and $CaCl_2$, it was hypothesized based on previous works (Jiangiang, M., and Pazos, I., 2014; Liao, Y., *et al.* 2017) that its stabilizing mechanism is also due to preferential interactions with the protein. On contrary, taurine presented a negative value of η_{cos}, which in this case indicated that its stabilizng effect is grater on more mesophilic-like proteins. Since it is an amino acid with a pK_a value between 0 and 1.5, at physiological pH it only exists in its zwitterionic form, and therefore it would be expected it to better interact with protein whose surfaces are proximal to electroneutrality (Bruździak, P., *et al.* 2018).

The dependence between the effect of the cosolutes on protein stability and the σ_{Prot} metric can be described by linear regression models
The stabilizing effect of KCl, $CaCl_2$ and TMAO is greater the more halophilic the protein, whereas the stabilizing effect of taurine is grater the more mesophilic-like the protein
The relationship between the effect of these cosolutes on protein stability and the composition of the surface can be explained from preferential interaction

4.3. Exploring the use of the σ_{Prot} metric on cell extracts

The cellular millieu is a complex mixture of different components with time and spacial-dependent concentrations. Due to this characteristic, working with pure cytosols is a challenge. Some plausible alternatives are the use of mimetic formulations (Theillet, F., *et al.* 2014), cell extracts as additives (Sharkar, M., *et al.* 2013), or even by measuring protein stability directly *in cellulo* (Smith, A., *et al.* 2016; Monteith, W., *et al.* 2015; Bai, J., *et al.* 2017).

Aiming to contextualize in a biological frame the previous results of this section, the effect on the stability of a formulation mimicking *E. coli* cellular environment (hereinafter referred to as ECMIM), and cellular extracts from *E. coli* and *Human embryonic kidney* cells (HEK) was measured by NMR. In order to guarantee that the comparison between cell extracts was possible, both of them were normalized to 50 mg/mL of protein.

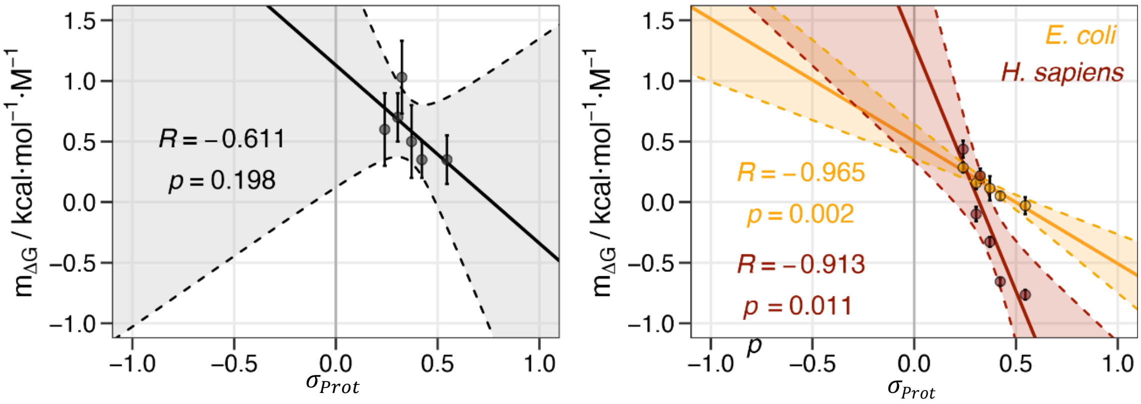


Figure R14. Correlations between σ_{Prot} and the m-values for ECMIM (left) and cell extracts (right). For each correlation, the corresponding Pearson correlation coefficient and its p-value are shown. Shaded regions correspond to the 95% confident intervals, and dark lines correspond to the fitting line. Error bar indicate the experimental variability.

Figure R14 shows the correlations between the m-values of the proteins shown in Table M3 and the corresponding values of the σ_{Prot} metric that were obtained in presence of ECMIM and cell extracts. For both, the *E. coli* and HEK exctracts the values obtained for the η_{cos} slope were not negligible, which indicated that the cytosol affects the stability in a way that depends on surface composition. Furthermore, it was observed that $|\eta_{HEK}| \cong 4 \cdot |\eta_{E.coli}|$, which underlined the the greater effect exerted on the stability by eukaryotic extracts, and consequently on the quinary structure. It should be alsto noted from Figure R3 that the mean of the σ_{Prot} distribution for *H. sapiens* is moderatly more "mesophilic", it is to say, more proximal to a value of zero, than that of the *E. coli* distribution, which could explain why for both species the effect of cell extracts is moderately stabilizing.

The effect of *E. coli* and *HEK* extracts shows a dependence on the σ_{Prot} metric and, in both cases, it tends to stabilize more mesophilic-like proteins

4.4. On the cosolutes that show weak correlation with the metric

Sucrose, sarcosine, glycine, trehalose and K_2SO_4 showed no significant correlation between the m-values and σ_{Prot}. For these cosolutes the p-values obtained (Tables R7 and R8) were, in all cases, largely greater than 5%, which prompted to consider the hypothesis that there was no correlation between the m-values and σ_{Prot}. However, despite the fact that for these p-values it could be possible to observe a Pearson's correlation coefficient of zero, any other value among those inside the confident region could be also possible. Therefore, since the confident interval of $R = 0$ with a confident coefficient of 95% for $n \geq 15$ includes values between -0.51 and 0.51, it could be assumed that, if it existed correlation for these cosolutes, this correlation was weak enough to consider it negligible. According to this, it was assumed that the effect of sucrose, sarcosine, glycine, trhalose and K_2SO_4 on the stability is independent from the composition of the surface.

Thus, for these surface-insensitive cosolutes the regression model (Figures R10 and R11) would consist of a line with a slope $\eta_{cos} = 0$, and the intercept, m_0, could be interpreted as the average effect of the corresponding cosolute on the stability of proteins. In fact, under this interpretation the resulting m_0 parameters were in good agreement with the effect expected for these cosolutes from the current theoretical framework, $m_{trehalose} > m_{sucrose} \cong m_{K_2SO_4} > sarcosine \cong glycine$ (Auton, M., and Bolen, D., 2005; Ibarra-Molero, B., *et al.* 2000). In addition, variations in these average effects would be expected, especially if proteins with very different sizes are compared. However, some theoretical approaches allow to predict the effect of the cosolutes on protein stability and several models have demonstrated to be useful for rationalizing these effects (Record, M., *et al.* 1978; Auton, M. and Bolen, D., 2005). It can be argued that they exert their effect on the stability mainly by a mechanism insensitive to the composition of the protein's surface. This hypothesis is consistent with the mechanism based on excluded volume commonly attributed to sucrose, trehalose and potassium sulfate, and to a lesser extent to glycine and sarcosine (Miklos, A., *et al.* 2011; Patel, C., *et al.* 2002).

No statistically significant correlation between the m-values and σ_{Prot} can be assumed for sucrose, sarcosine, trehalose, glycine and K_2SO_4. Therefore, no regression model stands for these cosolutes

However, a linear regression model with a slope near zero agrees well with the expected effect of these cosolutes on protein's stability

The intercepts of such hypothetical regression models represent the average effect of the cosolutes on protein's stability and also agree well with theory

4.5. The relevance of the σ_{Prot} metric

As already mentioned, the protein's quinary structure is considered the structural organization level that comprises the formation of supramolecular complexes, and is responsible of the cellular organization (Cohen, R., and Peilak, G., 2015; Cohen, R., and Pielak, G., 2017). The main role of the protein's surface has long been relegated to the role of a merely solubility guaranteer. Tet, it seems evident that in order to conform multiple complexes, and with the interaction with other biomolecules the only region exposed to the exterior is the surface. Some recent works (Monteith, W., *et al.* 2015) point towards the idea that the protein's surface plays a more relevant role and that modulates the stability in the cellular environment. Thus, it seems plausible to believe that the protein's surface is responsible of modulating the homeostasys of the biomolecule through the optimization of its interaction with other molecules. Assuming that Nature has selected this as one of the criteria to drive protein evolution, the modifications found in the surfaces of the proteins belonging to extremophilic organisms can be interpreted as a natural strategy to preserve quinary structure.

In this work the analysis of the evolutionary pressure observed in haloadaptation has lead to the construction of a metric that allows to clasify a protein's surface, ultimately providing a measurement of the contribution of the later to the quinary structure. It is important to highlight that the definition of this metric is not based on a sequence alingment, but on the comparison between the composition of the surfaces of halophilic proteins and the surfaces of proteins from other organisms, and that the definition of the surface has been based exclusively on structural parameters. Thus, the validity of the σ_{Prot} metric is supported by the premise that the haloadaptation has exerted evolutive pressure mainly on protein's surface with the objective of maintaining the stability, an idea that becomes strengthened by the overwhelming structural similarity between the halophilic proteins and their mesophilic orthologs. This metric is straightforward to calculate, and only requires to know the amino acid composition of the protein.

Related to the concept of evolution used in the context of this work, perhaps it is illustrative to comment on an interpretation that can be extracted from Figure R3. It can be observed there that the mean of the distribution of the σ_{Prot} values for the proteins from mesophilic organisms is slightly different from zero. This fact can be understood as that the stability of an average mesophilic protein is marginally sensitive to the composition of the environment. This is not surprising, since the stability of a protein is tightly related to its homeostasis (Balchin, D., *et al.* 2016), a process which in last term is regulated by many other specific mechanisms, and as the σ_{Prot} metric highlights, a high sensitivity to the media could compromise the stability.

This work also analyzed the effect of a set of cosolutes on the stability of a library of proteins, and as have been shown, for a given cosolute the $m_{\Delta G}$-values and $m_{\Delta T_m}$-values keep a linear correlation with the σ_{Prot} metric characterized by its slope ($\eta_{cos,\Delta G}$ or $\eta_{cos,\Delta T_m}$) and its intercept ($m_{0,\Delta G}$ or $m_{0,\Delta T_m}$). Figure R15 shows a plot in which the slopes are represented against their corresponding intercepts. For each case, the η_{cos} slope can be interpreted as the response of the protein's quinary structure to the cosolute, and in the cases in which it is significantly different from zero, it indicates a direct or indirect interaction between the cosolute and the protein's surface. For example, this interpretation agrees well with the mechanism attributed to $CaCl_2$, taurine, KCl and TMAO. Furthermore, for these cosolutes the sign of η_{cos} indicates the surface composition that is primarily stabilized, it is to say, halophilic-like surfaces for $CaCl_2$ and non-halophilic-like surfaces for taurine. On the other hand, as was commented before, for the cosolutes for which the $\eta_{cos} \cong 0$, the intercepts represent the average stabilization produced on a protein. For instance, sulfate, trehalose, and sucrose present large possitive values for the m_0 intercept, wich is consistent with hypothesis that excluded volume is their main stabilization mechanism (Patel, C., *et al.* 2002).

In addition to the analysis performed with pure cosolutes, this work also demonstrates that proteins' quinary structure can be investigated in environments more relevant from the physiological point of view. It is of great interest to highlight that the $m_{\Delta G}$-values obtained by NMR spectroscopy for the subset of 7 partially unfolded proteins revealed that cell extracts can also stabilize proteins in a way dependent on protein's surface composition, a result that was remarkably similar for the designed cytosolic mixture (ECMIM). Thus, the σ_{Prot} metric captures how evolution has selected the optimal protein's surface composition for interacting with the cytosol, or in other words, it provides a clear insight of the coevolution of the cytosolic and the protein's surface composition.

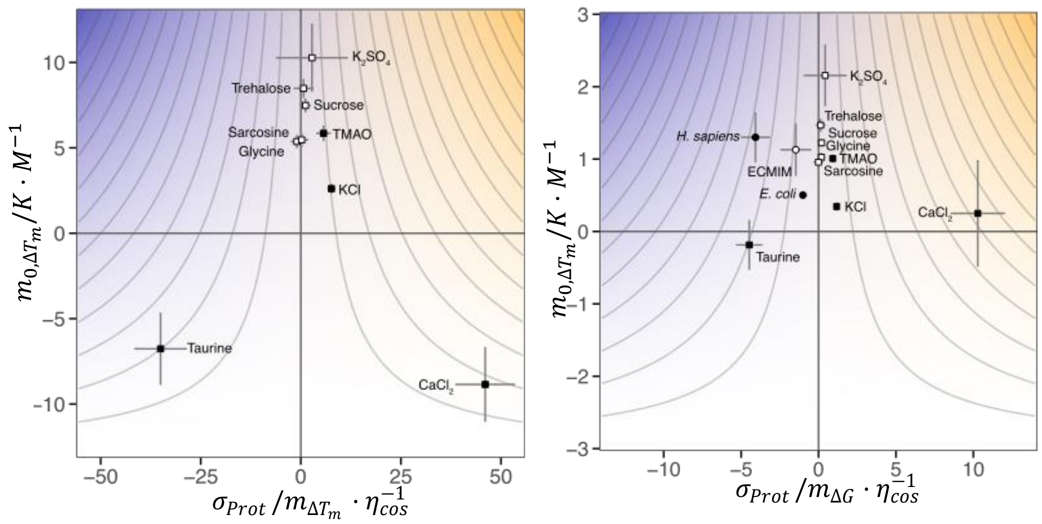


Figure R15. m_0 values versus η_{cos} extracted from thermal stability (left), and thermodynamic stability (right) analysis. Squares and circles represent pure cosolute, and complex mixtures, respectively. Black filling indicates significative correlation ($p < 0.05$). Contour lines and color gradients represent the stabilizing effect of a given cosolute or cytosolic mixture (sarker colour indicates higher stabilization). Stabilization of more halophilic-like or mesophilic-like surface's composition are represented in orange and blue, respectively.

In summary, it is presented here a quantitative metric that takes advantage of the extremophile evolution as a tool to probe the contribution protein's surface and cosolutes to the quinary structure. This metric, inspired by the natural adaptation mechanism of halophilic organisms allow to order proteins by their sensitivity to the environment, as well as cosolutes by the mechanisms by which they affect protein stability. In the light of the results presented in this section, the σ_{Prot} metric is expected to be useful, for example, to optimize the stability of protein formulations in industrial or pharmaceutical applications, for investigating the sensitivity of IDPs to the environment and their conformational landscapes (Moses, D., *et al.*, 2020), understand in other in vivo phenomena such as liquid-liquid phase separation or protein aggregation, and the relationship between protein stability, quinary structure and proteostasis.

Chapter V

Cosolute modulation of the dimerization reaction of protein L G55A

5. Cosolute modulation of the dimerization reaction of protein L G55A

Many protein oligomerization processes take place inside the cell and therefore are affected by the surrounding environment, normally the cytosol, which can be considered as a complex and dynamic mixture of cosolutes. This section presents the results corresponding to the study of the effect of a large set of cosolutes of diverse physicochemical nature on the domain-swapping dimerization reaction of the model protein PL_G55A. All the information provided in the following pages was published in the work from Mateos *et al.* (Mateos, B., *et al.* 2021).

5.1 PL_G55A homo-dimerization process under reference conditions

PL in its monomeric form folds into a four-stranded β-sheet and a single α-helix packed together in a $\beta\beta\alpha\beta\beta$ fold, with $\beta 1/\beta 2$ and $\beta 3/\beta 4$ each forming β-hairpins. The first β -turn, stablished between strands $\beta 1$ and $\beta 2$, is a classic type I turn, whereas three out of four residues in the β-turn formed by strands $\beta 3$ and $\beta 4$ have consecutive φ-angles, thus causing conformational strain in this part of the structure (Moshen, T., and Tollinger, M., 2014). Mutation of the glycine at position 55 by an alanine increases the strain of the second β-turn, thus leading to the formation of a domain swapped dimer in which the second β-turn straightens and the C-terminal β-strands of two protein molecules are mutually inserted into the β-sheet of their dimerization partner (Figure R16) (O'Neill, J., *et al.* 2001).

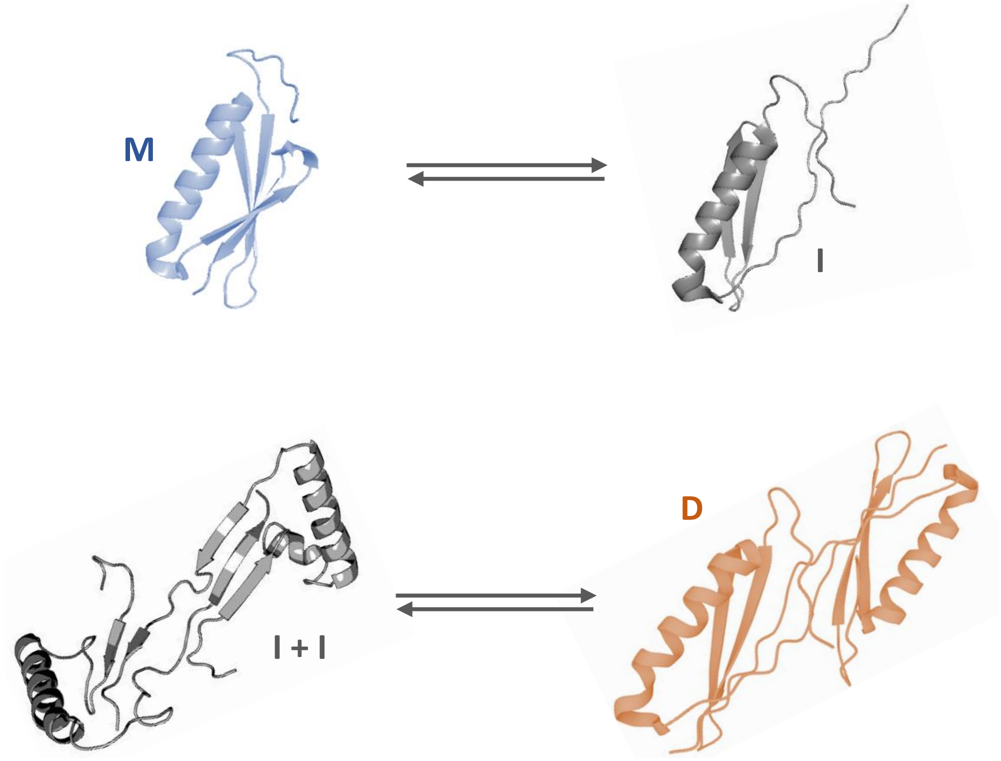


Figure R16. Domain swapping of PL_G55A. Schematic representation of the oligomerization reaction starting from monomer (blue) and dimer (orange). Intermediates (gray) of both processes are also represented.

The interconversion between the monomer and the dimer presents a half-time of around 2 hours in the reference buffer (see section 3.2.2.4.), which is slow enough to allow the separation and purification of both forms by size-exclusion chromatography. Gaussian Fitting to the experimental data of the SEC's profile obtained from a sample equilibrated for more than 48 hours at RT and subsequent integration showed that the monomer and dimer populations are present in the equilibrium in a proportion of 0.657 and 0.343 respectively (Figure R17).

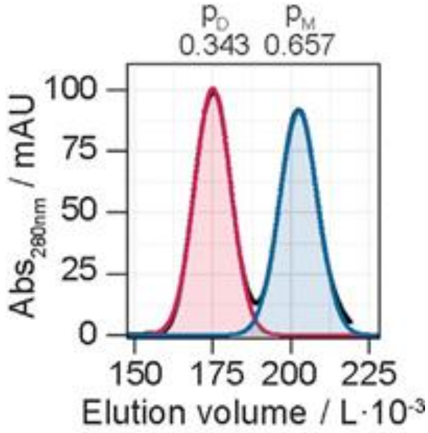


Figure R17. SEC profile of a sample equilibrated for > 48 h at RT. P_M and P_D are monomer and dimer populations respectively. Approximately one third of the protein molecules dimerize thus allowing independent purification due to chromatographic separation.

Taking advantage of the fact that the monomer and the dimer can be independently purified, the kinetics of the association and dissociation reactions were studied starting from both, pure dimer and pure monomer. The progression of both reactions could be followed by monitoring the intensity changes of the signals of up to 17 reporter residues that showed nonoverlapped monomer and dimer resonances in the ^{1}H ^{15}N HSQC spectrum (Figures R18 and R19).

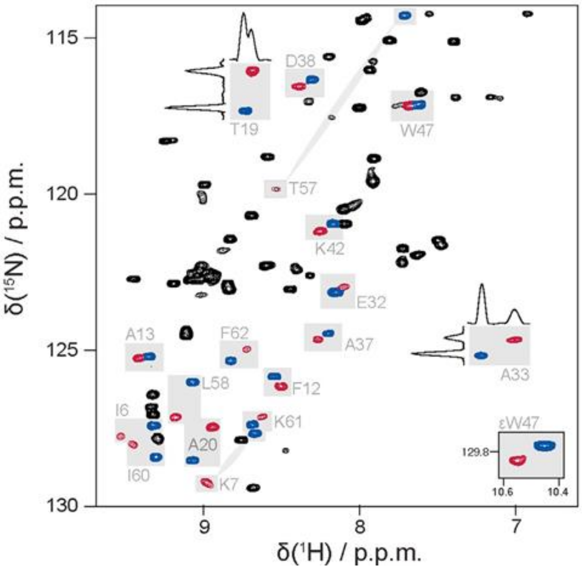


Figure R18. ^{1}H ^{15}N HSQC spectrum of ^{15}N-labelled PL_G55A. Monomer and dimer signals are indicated in blue and red, respectively.

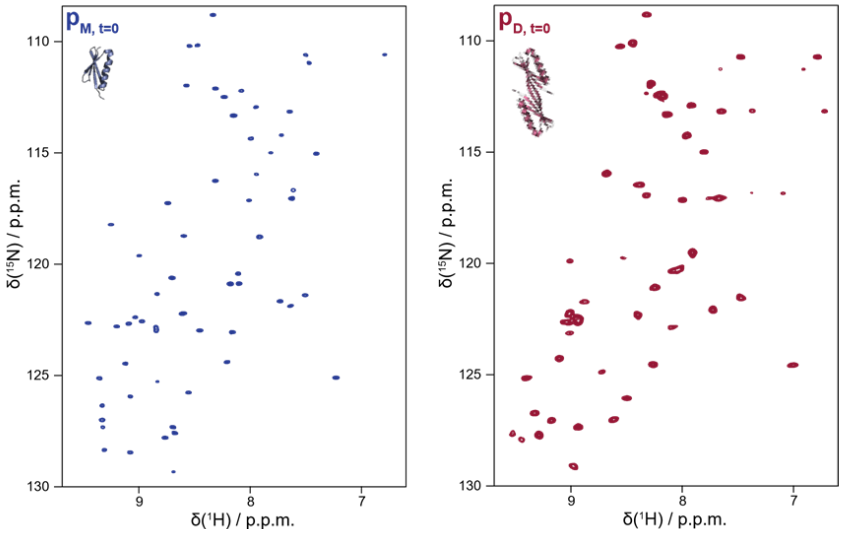


Figure R19. Initial ^{1}H ^{15}N HSQC starting from monomer (blue) and dimer (red).

Since the peak intensities, and consequently the peak volumes of the signals in a ^{1}H ^{15}N HSQC spectrum are proportional to the chemical activities, a, the peak volumes over time report on the reaction chemical potentials, μ. Taking this in consideration, the equilibrium populations in the reference buffer could be determined by extrapolation of m at very long times ($\lim_{t\to\infty} \mu(t)$ = 0.39 kcal·mol^{-1}) as described in the work from Moschen and Tollinger (Moschen, T., and Tollinger, M., 2014). Regarding this, a very similar free energy value was obtained after 35 hours of reaction (ΔG = 0.38 kcal·mol^{-1}), to which corresponded a monomer population of 0.68 ± 0.06, in good agreement with the value obtained from SEC data.

To a good approximation, the peak volumes of the NMR resonances can be fitted to a biexponential curve from which pure association, k_{asc}, and dissociation, k_{diss}, kinetic rates can be extracted (see section 3.8). According to this idea, both kinetic rates were evaluated starting the reaction from either the monomeric, M, and dimeric, D, species. Following this procedure, it was possible to extract both reaction rates (Figure R20) with high accuracy.

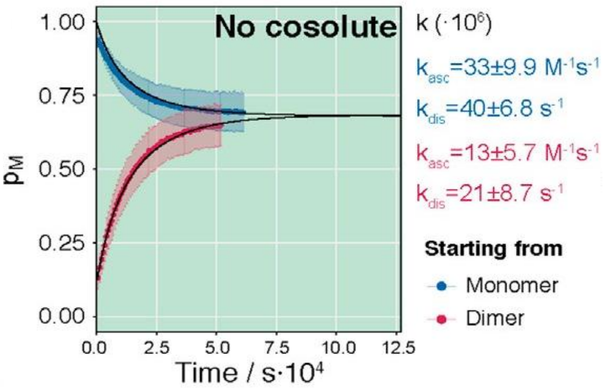


Figure R20. Monomer populations along time when starting the reaction from monomer (blue), and dimer (red) in the absence of cosolutes. Curves fitted to the experimental data are represented in black. Association and dissociation kinetic rates for both reactions are presented on the right side of the plot.

In agreement with what could be expected from a theoretical point of view, the kinetic rates of the association and dissociation processes presented similar values in the absence of cosolutes, it is to say, $k_{asc,D} = k_{asc,M}$ and $k_{dis,D} = k_{dis,M}$, where $k_{asc,D}$ and $k_{dis,D}$ are the association and dissociation kinetic rates when starting from dimer, and $k_{ass,M}$ and $k_{dis,M}$ are the association and dissociation kinetic rates when starting from monomer, respectively. Consequently, the equivalent time evolution of the chemical potentials also showed the same tendence, thus leading to essentially identical estimations of the equilibrium thermodynamics. An important aspect regarding this that is worth to note is that the equiliribum that will be referred to hereinafter corresponds to the apparent equilibrium stablished in a temporal window of ~35 hours, which is a biologically relevant time-frame and which in this case does not differ from the equilibrium expected from extrapolation at $t \rightarrow \infty$. Following these ideas, it was stablished that this equilibrium obtained in absence of cosolutes is achieved in refence conditions. According to this, its reaction's free energy and kinetic rates were defined as reference parameters:

$$\Delta G^{\circ}_{M\leftrightarrow D} = \Delta G^{\circ}_{D\leftrightarrow M} \qquad k^{\circ}_{asc} = k_{asc,D} = k_{ass,M} \qquad k^{\circ}_{dis} = k_{dis,D} = k_{dis,M}$$

Equations R5, R6 and R7. Reference free energy, association, and dissociation kinetic rates definition.

Dimerization reaction is slow enough to allow monomer and dimer separation by SEC and/or 2D NMR spectroscopy

Kinetic rates and populations at the apparent equilibrium (~35 h) are taken as reference conditions

5.2. PL_G55A domain swapping in presence of cosolutes

Complementarily, the kinetics of the two reactions were also monitored in the presence of several cosolutes at 0.5 M concentration covering a hydrodynamic radius range from 1.33 Å to 4.79 Å and with a volume occupancy that belonged to the dilute regime. The protein was observed to be completely folded at all tested concentrations of cosolutes (even in presence denaturants), what indicated that the effect of the cosolutes chosen on the stability of the protein was negligible. All the cosolutes considered were rationally selected, looking for them to be small, relatively inert, and highly soluble organic molecules (Figure R21).

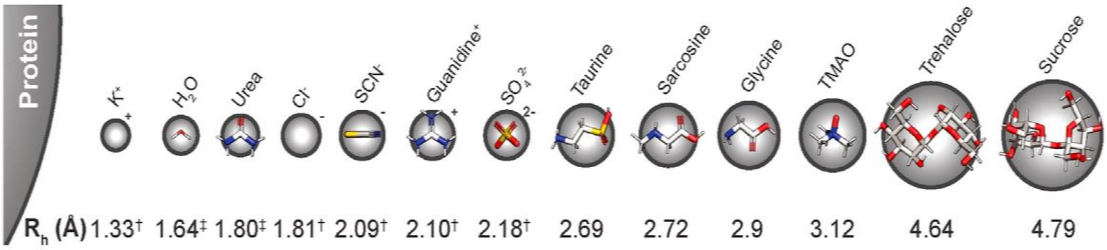


Figure R21. Schematic representation of the cosolutes used for the study of the kinetics of the dimerization reaction of PL_G55A. The size of the sphere is proportional the the hydrodinamic radius (R_h) calculated from diffusion-ordered NMR spectroscopy in 20 mM phosphate buffer pH 6.0 and 298 K (reference conditions). For ions, the thermochemical radii (indicated with a dagger (†)) reported in the literature were used (Yizhak, M., 2012; Roobotom, H., *et al.* 1999). The R_h of water and urea are reported to be 1.64A° (Price, W., *et al.* 2000) and 1.80A° (Schultz, S., and Solomon, A., 1961), respectively, and are indicated with a double dagger (‡).

Since the idea was to conform a group of cosolutes as diverse in nature as possible, the group of cosolutes considered in this work was composed of sugar alcohols (sucrose and trehalose), amino acids, trimethylamine N-oxide, guanidinium chloride (GuHCl), urea, and inorganic ions. Sugar alcohols have historically been used to study the deviations of protein dimerization from thermodynamically ideal behavior (Sherwin, K., and Winzor, D., 1988), the amino acids glycine, its methylamine sarcosine, and the amino sulfonic acid taurine are produced as compensatory compounds to preserve cytoplasm homeostasis under stress conditions (Somero, G. 1986), and TMAO has demonstrated to possess protective properties against urea, GuHCl (Ma, J., *et al.* 2014; Pincus, D., *et al.* 2008) and inorganic ions (K^+, Cl^-, SCN^-, SO_4^{2-}), which also contribute to protein stability (Ortega, G., *et al.* 2015; Tadeo, X., *et al.* 2009) and oligomerization by the Hofmeister effect (Tadeo, X., *et al.* 2009).

As well as in the case of absence of cosolutes, the thermodynamic stability and kinetics of the dimerization reaction of PL_G55A were monitored in presence of each of the cosolutes shown in Figure R21 (Figures R22, R23 and Table R9) as described in section 3.3.3.

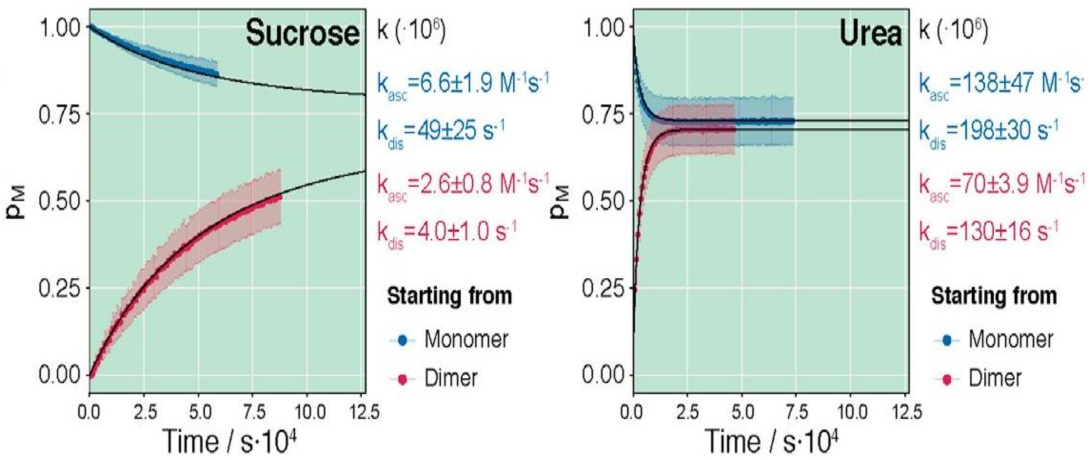


Figure R22. Monomer populations along time when starting the reaction from monomer (blue), and dimer (red) in the presence of cosolutes. Curves fitted to the experimental data are represented in black. Association and dissociation kinetic rates for both reactions are presented on the right side of the plot.

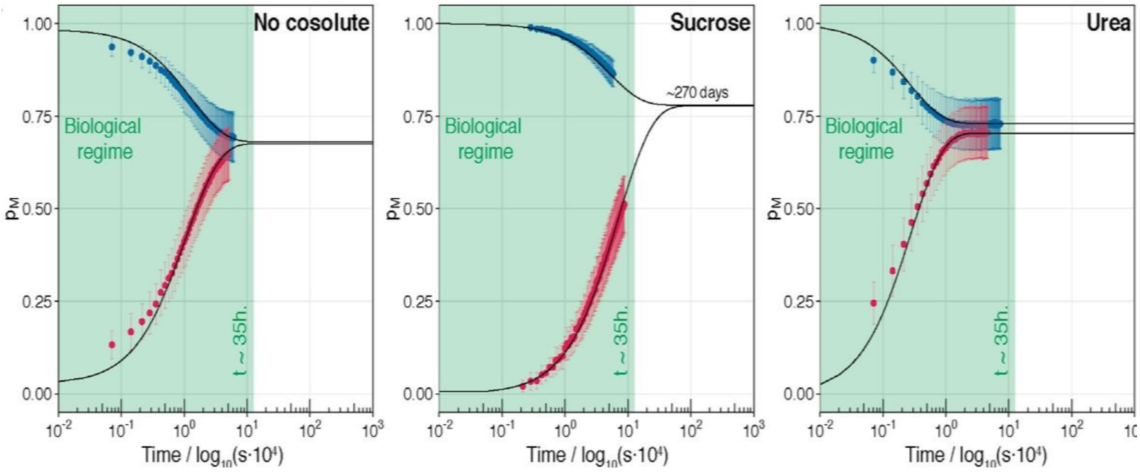


Figure R23. Monomer populations shown in Figure R22 plotted against time in logarithmic scale in absence and presence of cosolutes. As shown, equilibrium is always reached. However, presence of certain cosolutes such as sucrose slows down the reaction so that the equilibrium is stablished in times out of the biologically relevant timeframe (~270 days).

Table R9. Averaged values of the kinetic rates extracted from fitting, the number of points considered for their calculation, n, their standard deviations, σ, and the value of the corresponding statistical estimator, χ2.

	Initial state	[cosol.] (M)	Replica	n	Avgd. k_{dis} (s^{-1})	$\sigma_{k_{dis}}$ (s^{-1})	Avgd. k_{asc} ($M^{-1}\cdot s^{-1}$)	$\sigma_{k_{asc}}$ ($M^{-1}\cdot s^{-1}$)	Avgd. χ2
reference	monomer	0	No	14	$4.05\cdot10^{-5}$	$6.84\cdot10^{-6}$	$3.28\cdot10^{-5}$	$9.87\cdot10^{-6}$	0.033
	monomer	0	Yes	14	$4.05\cdot10^{-5}$	$6.89\cdot10^{-6}$	$3.28\cdot10^{-5}$	$9.81\cdot10^{-6}$	0.033
	dimer	0	No	13	$2.10\cdot10^{-5}$	$8.70\cdot10^{-6}$	$1.27\cdot10^{-5}$	$5.71\cdot10^{-6}$	0.085
KCl	monomer	0.5		18	$2.86\cdot10^{-5}$	$5.37\cdot10^{-6}$	$6.78\cdot10^{-5}$	$1.52\cdot10^{-5}$	0.108
	dimer	0.5		14	$1.60\cdot10^{-5}$	$1.12\cdot10^{-5}$	$9.68\cdot10^{-6}$	$6.35\cdot10^{-6}$	0.144
K_2SO_4	monomer	0.5		13	$1.21\cdot10^{-5}$	$6.85\cdot10^{-6}$	$1.60\cdot10^{-5}$	$7.22\cdot10^{-6}$	0.413
	dimer	0.25		14	$6.28\cdot10^{-6}$	$1.29\cdot10^{-6}$	$4.10\cdot10^{-6}$	$8.85\cdot10^{-7}$	0.241
	dimer	0.5		9	$5.19\cdot10^{-7}$	$8.71\cdot10^{-8}$	$2.79\cdot10^{-7}$	$4.02\cdot10^{-8}$	0.022
Taurine	monomer	0.5		14	$1.95\cdot10^{-5}$	$2.57\cdot10^{-6}$	$2.64\cdot10^{-5}$	$6.84\cdot10^{-6}$	0.044
	dimer	0.25		9	$6.10\cdot10^{-6}$	$1.55\cdot10^{-6}$	$3.92\cdot10^{-6}$	$9.09\cdot10^{-7}$	0.734
	dimer	0.5		14	$5.87\cdot10^{-6}$	$1.64\cdot10^{-6}$	$3.48\cdot10^{-6}$	$9.31\cdot10^{-7}$	0.138
Sarcosine	monomer	0.5		14	$2.38\cdot10^{-5}$	$3.97\cdot10^{-6}$	$1.53\cdot10^{-5}$	$4.69\cdot10^{-6}$	0.034
	dimer	0.5		11	$5.65\cdot10^{-6}$	$2.04\cdot10^{-6}$	$3.21\cdot10^{-6}$	$1.18\cdot10^{-6}$	0.038
Glycine	monomer	0.5		14	$1.76\cdot10^{-5}$	$2.68\cdot10^{-6}$	$1.77\cdot10^{-5}$	$5.74\cdot10^{-6}$	0.039
	dimer	0.25		10	$6.76\cdot10^{-6}$	$1.83\cdot10^{-6}$	$4.34\cdot10^{-6}$	$1.02\cdot10^{-6}$	0.728
	dimer	0.5		11	$4.50\cdot10^{-6}$	$9.27\cdot10^{-7}$	$2.63\cdot10^{-6}$	$5.75\cdot10^{-7}$	0.044
Trehalose	monomer	0.5		14	$2.74\cdot10^{-5}$	$1.11\cdot10^{-5}$	$5.06\cdot10^{-6}$	$2.26\cdot10^{-6}$	0.039
	dimer	0.5		9	$3.18\cdot10^{-6}$	$7.34\cdot10^{-7}$	$1.91\cdot10^{-6}$	$4.67\cdot10^{-7}$	0.040
Sucrose	monomer	0.5		9	$4.90\cdot10^{-5}$	$2.54\cdot10^{-6}$	$6.63\cdot10^{-6}$	$1.93\cdot10^{-6}$	0.020
	dimer	0.5		9	$4.04\cdot10^{-6}$	$1.01\cdot10^{-6}$	$2.64\cdot10^{-6}$	$7.9\cdot10^{-7}$	0.034
Urea	monomer	0.5		13	$1.98\cdot10^{-4}$	$2.97\cdot10^{-5}$	$1.38\cdot10^{-4}$	$4.69\cdot10^{-5}$	0.016
	dimer	0.5		12	$1.30\cdot10^{-4}$	$1.63\cdot10^{-5}$	$7.05\cdot10^{-5}$	$3.96\cdot10^{-6}$	0.050
TMAO	monomer	0.5		11	$1.72\cdot10^{-5}$	$3.87\cdot10^{-6}$	$1.01\cdot10^{-5}$	$3.32\cdot10^{-6}$	0.093
	dimer	0.25		12	$5.78\cdot10^{-6}$	$1.69\cdot10^{-6}$	$3.67\cdot10^{-6}$	$1.01\cdot10^{-6}$	0.771
	dimer	0.5		9	$6.01\cdot10^{-6}$	$1.61\cdot10^{-6}$	$3.81\cdot10^{-6}$	$1.24\cdot10^{-6}$	0.027

Independently from the starting point of the reaction, the kinetics observed for each of the 17 probe residues selected were very similar. This fact reinforced the idea that the cosolutes exert their effect mainly on the backbone, and due to this, they mainly affect the global thermodynamics of the protein (Auton, M., and Bolen, D., 2005).

It can be clearly observed from Figure R23 how when 0.5 M sucrose was added the apparent equilibrium was still not reached after ~35 hours, and that the observed monomer and dimer populations were different from those obtained in absence of cosolutes. On contrary, when 0.5 M urea was added then the apparent equilibrium was achieved faster, but the monomer and dimer populations still differed from that of the equilibrium at standard conditions. These results lead to the interpretation that the presence of cosolutes significantly alter both, the kinetics of the dimerization reaction and the populations of monomer and dimer reached in the final apparent equilibrium in a way that depends not only on the nature of the cosolute, but also on the initial reaction conditions. Since these results were counterintuitive from a theoretical point of view, they required a deeper analysis in view of that the tested model protein presented similar characteristics of its initial (monomer) and final (dimer) states in terms of solvent-accessible area ($\Delta ASA \approx 0$) and volume ($\Delta V \approx 0$), and according to current theoretical models it would be expected no differences from what is observed at standard conditions. With this last idea, the reactions started from the two initial conditions were monitored separately aiming to clarify the origin of the observed differences.

The presence of cosolutes alter the kinetics of the dimerization reaction and the populations at the apparent equilibrium

5.3. Reaction starting from dimer

When started from the dimer, the reaction progressed until it achieved an equilibrium in which the populations of the monomer and dimer, and consequently the chemical potential, did not significantly differ from that obtained under reference conditions (Figure R24), and were equivalent to those obtained from extrapolation to $t \to \infty$.

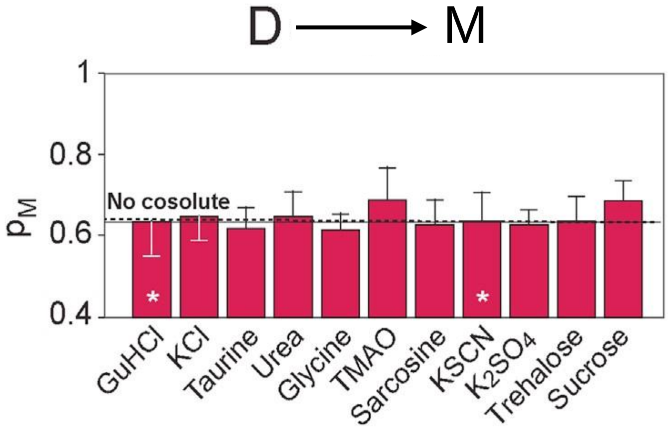


Figure R24. Average monomer populations in the apparent equilibrium of the reaction started from dimer for each of the cosolutes considered in Figure R21. Monomer population at reference conditions is indicated as a dashed line. White asterisks indicate the data sets in which monomer population was directly calculated from peak volumes at the final spectra.

As has already been pointed, this result was expected, taking in consideration that the changes in excluded volume and accessible surface area calculated between two monomers and the domain-swapped dimer (Table M16) are minimal. In contrast, when cosolutes were added it was observed a significant alteration of both, the association and dissociation kinetic rates, with urea and sulfate producing the fastest and slowest rates, respectively (Figure R25). Due to these alterations, the kinetics of GuHCl and KSCN could not be determined because the equilibrium was achieved so fast it was already stablished when the first points were measured. However, as it is shown in Figure R25 despite the changes in the values of the kinetic rates, for each cosolute the ratio between the association and dissociation kinetic rates remained fixed at the same value that was observed in absence of cosolutes, and the slope of the corresponding correlation yielded a $\Delta G_{D \rightleftharpoons M}$ value of 0.38 kcal·mol^{-1}, which was essentially identical to the equivalent value determined from the equilibrium populations in absence of cosolutes (0.38 kcal·mol^{-1}). This result was also expected, taking in consideration that $\Delta G_{D \to M} \propto k_{diss,D}^{cosol.} / k_{ass,D}^{cosol.}$. In line with these findings, it was also observed that changing the cosolute concentration provoked variations in the values of the kinetic rates, but the populations at equilibrium remained unaltered.

All these results support the interpretation that the reaction started from dimer is under thermodynamic control, and in agreement with the existing theoretical framework (Schreiber, G., 2002; Minton, A., 1998; Ming, J., *et al.* 2010), when cosolutes are present the free energy of this reaction, $\Delta G_{D \to M}$, can be calculated from the equation below:

$$\Delta G_{D\to M} = \Delta G^{\circ}_{D\to M} + RTln(\frac{\gamma_M^2}{\gamma_D})$$

Equation R8. Calculation of the free energy difference for the dimerization reaction in presence of cosolutes. γ_D and γ_M are the activity coefficients of the dimer and monomer, respectively.

Data shown in Figure R22 allow to claim that $\gamma_D \cong \gamma_M^2$, that together with the fact the changes in volume and solvent-accessible area between the two species considered are almost non-existent lead to the hypothesis that $\gamma_D \cong \gamma_M \cong 1$.

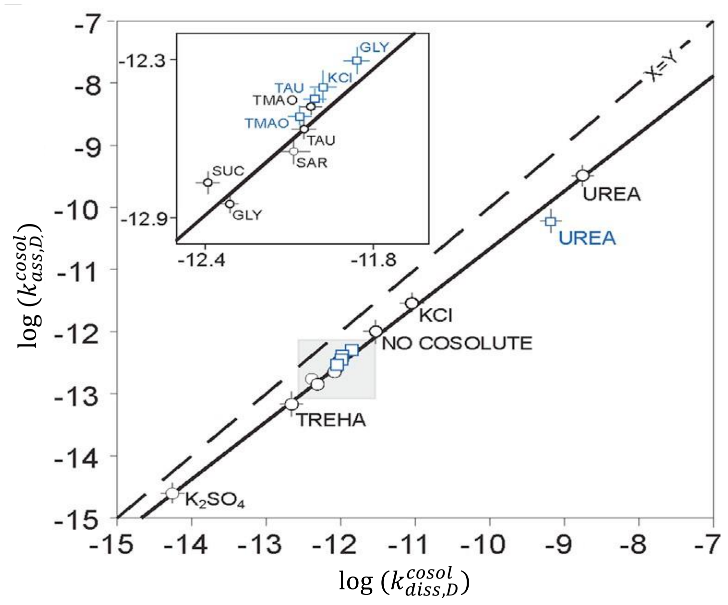


Figure R25. Direct, $k^{cosol.}_{ass,D}$, and reverse, $k^{cosol.}_{diss,D}$, kinetic rates in presence of 0.5 M of cosolutes for the dimerization reaction. The presence of 0.25 M concentration for a representative subset of cosolutes is indicated with blue squares. Error bars indicate standard deviations. Dashed line indicates a theoretical equilibrium in which $k^{cosol.}_{ass,D} = k^{cosol.}_{diss,D}$.

On the other hand, the kinetic rates of association and dissociation obey the following equations:

$$k^{cosol.}_{dis,D} = k^{\circ}_{dis} \cdot \frac{\gamma_D}{\gamma_{TS}} \qquad\qquad k^{cosol.}_{asc,D} = k^{\circ}_{asc} \cdot \frac{\gamma_M^2}{\gamma_{TS}}$$

Equations R9 and R10. Relationship between the kinetic coefficients of association and dissociation and the activity coefficients. γ_{TS} is the chemical potential of the transition state, which depends on the nature and concentration of the cosolute.

If it is assumed that $\gamma_D \cong \gamma_M \cong 1$, according to Equations R9 and R10 the strong dependence of the kinetic rates on the concentration of cosolutes arises from the chemical potential of the transition state. This suggests that the structure of the transition state may significantly differ from that of the folded species.

Table R10. Activity coefficients of the transition state and their estimated errors. n.a. (not applicable) indicates that the corresponding value could not be determined.

Cosolute	γ_{TS} **(0.5M)**	γ_{TS} **(0.25M)**
GuHCl	n.a.	n.a.
KCl	0.6 ± 0.03	1.0 ± 0.02
Taurine	1.7 ± 0.03	1.6 ± 0.01
Urea	0.1 ± 0.07	0.1 ± 0.06
Glycine	2.3 ± 0.05	1.4 ± 0.03
TMAO	1.6 ± 0.04	1.7 ± 0.08
Sarcosine	1.8 ± 0.02	n.a.
KSCN	n.a.	n.a.
K_2SO_4	20.4 ± 0.14	n.a.
Trehalose	3.2 ± 0.06	n.a.
Sucrose	2.3 ± 0.04	n.a.

Table R10 shows the estimated values of γ_{TS} for each of the cosolutes under consideration. For each case, the value of the calculated activity coefficient for the transition state was virtually the same, regardless of whether it was calculated from the association or dissociation rate. It is worth noting that urea and high concentrations of KCl showed a value of $\gamma_{TS} < 1$, fact that can be interpreted as that these cosolutes stablishes favorable interactions with the largely unfolded transition state, thus resulting in an acceleration of the reaction rate. On contrary, sugars, amino acids, and stabilizing ions exhibited values of $\gamma_{TS} > 1$, leading to a slowdown of the reaction. This last fact suggests that it exists a contributing excluded volume term from the transition state that could arise from a bulkier structure compared to the folded species. This interpretation gains strength if it is considered that the dominant mechanism of these cosolutes is excluded volume (Patel, C., *et al.* 2002; Tadeo, X., *et al.* 2007), even though other mechanisms may also play a role.

Aiming to gain insight on the nature of the transition state, and trying to obtain a structural model of it, the dissociation reaction starting from dimer was computed be means of Gaussian Molecular Dynamics (GaMD). As it is shown in Figure R26, the system required a period of around 25 ms to vanish some of the interactions between the interlaced β-sheets, and then the trajectories became stable for more than 75 ms.

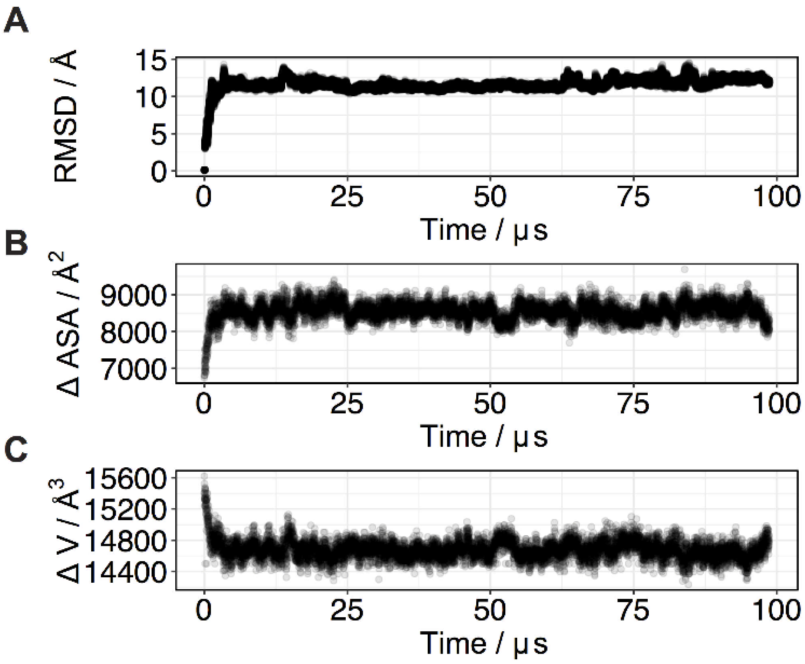


Figure R26. Molecular dynamic trajectory of the dimerization reaction starting from a PL_G55A dimer. (A) Root mean square deviation from the original structure. (B) Accessible surface area difference. (C) Volume difference.

The structural model that corresponded to the stabilized trajectories (Figure R27) consisted of an extended dimer in which the intermolecular interactions involving T48, K61 and F62 were localized in the interlaced β-sheets and were responsible of the stabilizing force. Regarding this, when the reaction was monitored for PL_G55A_K61A the monomer population raised up to $p_M = 0.71$ (Figure R28), which pointed towards a weaker interaction in the dimer interface for the double mutant. In line with this, it was observed that the structure shown in Figure R27 exposed 26% more area to the solvent than the completely formed dimer, which could mechanistically explain the dependence of the kinetic rates on the cosolute concentration. However, since near complete unfolding of the monomer is required for domain swapping (Liu, L., *et al.* 2012), the value of exposed area was expected to be higher.

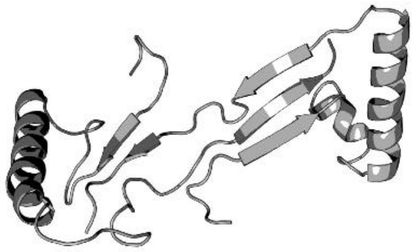


Figure R27. Structural model for the stabilized conformation in the trajectory (likely prior to the transition state).

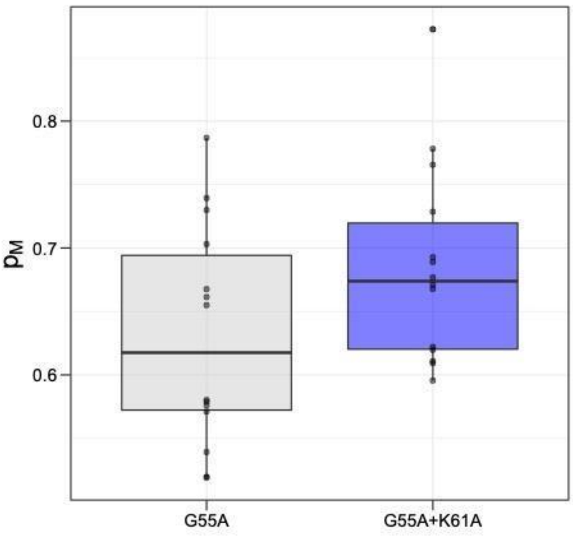


Figure R28. Equilibrium monomer populations calculated from peak volumes of ^{1}H ^{15}N HSQC of PL_G55A (grey) and PL_G55A_K61A (blue).

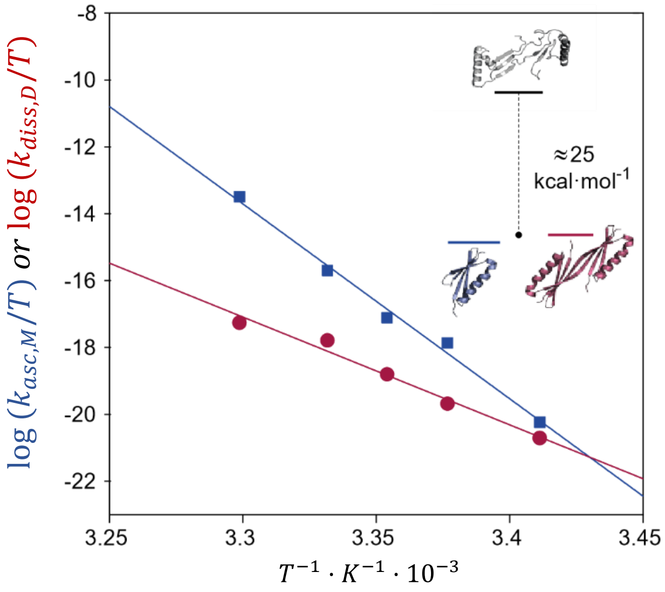


Figure R29. Dependence of kinetic rates on temperature (Eyring plot). Enthalpy and entropy differences were extracted from the slopes and intercepts of the fitting to experimental data, respectively. The free energy barrier for the transition state is indicated in the figure.

Figure R29 shows the dependence of the association and dissociation rates on temperature. Fitting of the Eyring's equation to the experimental data of dimerization reaction at different temperatures allowed the calculation of the activation enthalpy and entropy differences, which in last term were used to determine a value of the activation free energy of the reaction of $\Delta G \cong 25\ kcal \cdot mol^{-1}$.

Cosolutes do not alter the monomer and dimer populations at the true nor the apparent equilibrium when the reaction is started from dimer

Kinetic rate values are altered by cosolutes, but their ratio remains constant, thus explaining the unaltered populations at equilibrium

5.4. Reaction starting from monomer

The dimerization reaction was also studied starting from monomer. As it is shown in Figures R22 and R23, when the reaction was started from monomer, the kinetic rates, as well as the equilibrium populations showed a dependence on the nature and concentration of the cosolute added. If it is considered that the association reaction is bimolecular, consequently two partially unfolded intermediates of PL_G55A have to encounter previous to the formation of the dimer. With this approach, the experimental data and the information from molecular dynamics allow to claim that this reaction is kinetically trapped, this is, that it is dependent on the diffusion rates of the participating species. As it is shown in Figures R23 and R30, the thermodynamic equilibrium populations could be extrapolated to longer times, $t \to \infty$. However, these timeframes become meaningless from the point of view of a biological timescale (i.e., > 100 days).

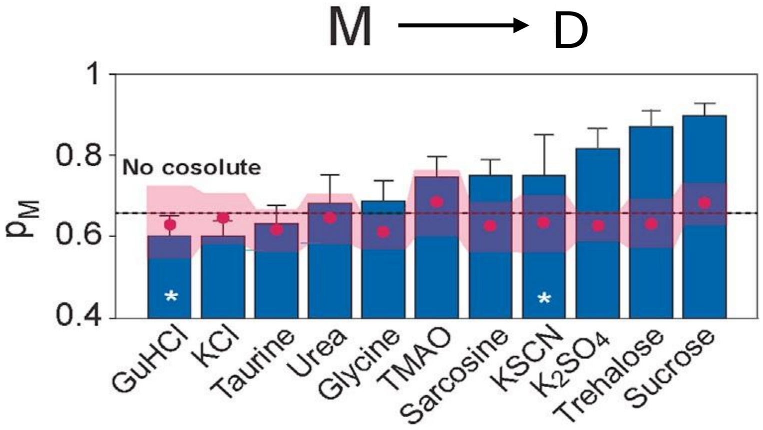


Figure R30. Average monomer populations in the apparent equilibrium for the reaction starting from monomer (blue bars). The reference dimer population values and their corresponding deviations are indicated as red points and shades, respectively. Monomer population at reference conditions is indicated as a dashed line. White asterisks indicate the data sets in which monomer population was directly calculated from peak volumes at the final spectra.

When considering the system in the presence of additives, it has already been described that the association rate shows an exponential dependence on the concentration of cosolutes following the relationship presented below (Equation R11) (Minton, A., 1998). Despite the definition of the association rate is purely kinetic, it also determines the final populations at apparent equilibrium that become kinetically accessible. In the present case, the modulation of the association rate by the encounter rate was very accentuated, even at moderate cosolute concentrations such as 0.5 M, and provoked the deviation of the monomer population from the value of 0.65 found in absence of cosolute up to the value of 0.9 ± 0.05 observed for sucrose or down to 0.6 ± 0.08 observed for GuHCl, as it is shown in Figure R30.

$$k_{asc,M} \cong k^{\circ}_{asc,M} \cdot e^{-g\cdot[cosol.]}$$

Equation R11. Dependence of the association rate on cosolute concentration. g is a proportionality factor that accounts for the encounter rate, in this case of the intermediate species, in presence of cosolutes.

The g factor that appears in Equation R11 has been described to be function of the relative shapes and sizes of the protein and cosolutes (Minton, A., 1998). Since any molecule can be modeled as a sphere whose size depends on its radius, determining the hydrodynamic radius of a molecule provides an approximate idea of its size and shape. According to this, diffusion-ordered spectroscopy (Sinnaeve, D., 2012) was used to determine the translational diffusion coefficients of the cosolutes considered in Figure R21, which in last term were used to calculate the hydrodynamic radii (see section 3.3.4).

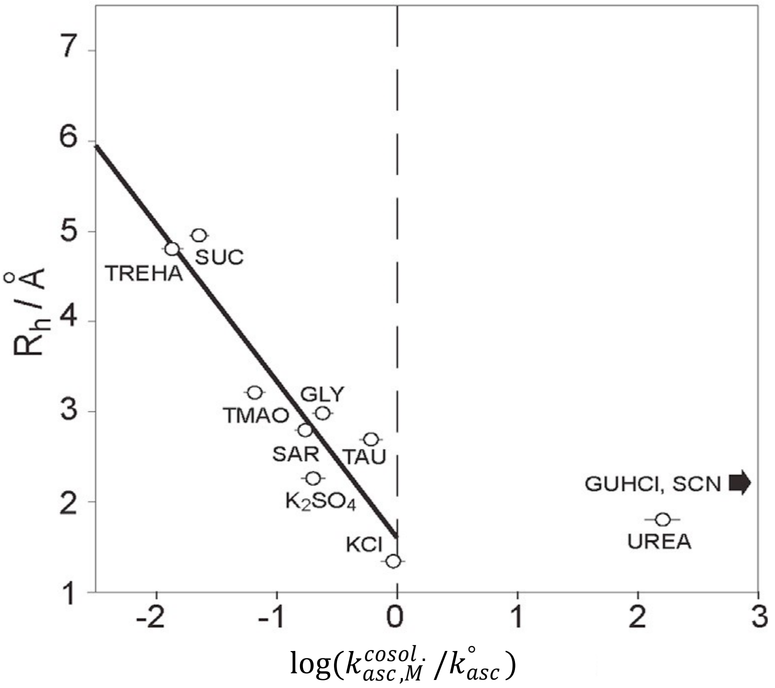


Figure R31. Relationship between the hydrodynamic radius of the cosolutes, R_h, and $k_{asc,M}^{cosol.}$ normalized by the association rate at standard conditions.

Figure R31 shows the relationship between the association rates in presence of cosolutes normalized by the association rate at standard conditions with the hydrodynamic radii of the cosolutes. It can be observed that there was a good negative linear correlation between both parameters that agreed with the hypothesis that the formation of the encounter complex plays a rate-limiting role in the $M \rightarrow D$ reaction.

Monomer populations at apparent equilibrium are function of the type and concentration of the cosolutes

The association reaction depends on the encounter rate and is kinetically trapped in presence of cosolutes

There is a correlation between the normalized association rate in presence of a given cosolute and its hydrodynamic radius

---

5.5. Model and examples

The large free energy barrier found for the transition state that resulted from Eyring's analysis (Figure R29) suggested that the conformation obtained by molecular dynamics could not reflect the rate-limiting step of the reaction, which would be characterized by the partial unfolding of the monomer. Therefore, these data pointed towards the appearance of a partially unfolded intermediate state before (or after) the largely unfolded transition state. Regardless of if the surface exposed to the solvent by the intermediate is greater than that of the transition state, the reaction progresses through the formation of a specie that requires partial unfolding, which in consequence should be more sensitive to the excluded volume effects introduced by the cosolutes. Thus, according to the ideas discussed above, the working model is that the dimerization reaction started from monomer seems to require the encounter of two partially unfolded monomer molecules (intermediates), which get trapped together by solvent cage effect and bind to form an encounter complex. This encounter complex ultimately evolves reversibly to produce the dimer through the formation of a largely unfolded transition state. The chemical equation below provides a schematic representation of the dimerization reaction according to this model:

$$M \underset{k_{I\to M}}{\overset{k_{M\to I}}{\rightleftharpoons}} I \qquad\qquad I + I \underset{k_{EC\to I}}{\overset{k_{I\to EC}}{\rightleftharpoons}} EC \underset{k_{D\to EC}}{\overset{k_{EC\to D}}{\rightleftharpoons}} D$$

Equations R12 and R13. Schematic representation of the stage in which the unfolded intermediate forms (left), and the dimerization reaction starting from intermediates (right). $k_{M\to I}$ and $k_{I\to M}$ respectively represent the kinetic rates of the initial stage, $k_{M\to EC}$, $k_{EC\to M}$, represent the direct and reverse encounter rates, respectively, EC represents the encounter complex, and I represents the intermediate.

According to this model, data shown in Figure R31 allowed to interpret that the cosolutes tend to oppose the formation of the encounter complex, and in consequence they decelerate the reaction. On contrary, it must be noted that denaturing agents such as GuHCl, SCN$^-$ or urea not only do not oppose the formation of the encounter complex, but even increase $k_{asc,M}^{cosol.}$. This apparently contradictory fact can be rationalized arguing that despite denaturants also alter passive diffusion, they preferentially bind to the protein, causing unfolding on it and therefore accelerating the path to the transition state.

Taken altogether, the results obtained from the reactions started from monomer and dimer emphasize the idea that the global reaction is dependent on an initial stage in which two partially unfolded intermediates must encounter, that a largely unfolded transition state must exist, and that impeded diffusion and stabilization of the folded conformations act against the domain-swapping oligomerization. Focusing now on the intermediate, it is important to note that since it is partially unfolded, it exposes more area to the solvent. Due to this, the cosolutes that act mainly by preferential interaction, such as urea, increase the association rate, thus accelerating the reaction and leading to a faster attainment of the equilibrium (Figure R32). Contrarily, cosolutes whose main acting mechanism is excluded volume, for example sucrose, tend to decelerate the reaction because they stabilize the folded conformations, hindering or even preventing the appearance of the intermediate required in the first stage of the association reaction, and because they reduce the diffusion rate, thus kinetically trapping the reaction (Figure R32). Following this reasoning, the effect of all the cosolutes considered could be explained as a balance between their capacity to preferentially stabilize/destabilize the intermediate and their modulating effect on diffusion, the later determined by their hydrodynamic radii.

An example that could illustrate all these ideas is the comparison between sulfate ion and TMAO. It can be observed from Table R2 that the association rate for the reaction starting from monomer is greater for sulfate than for TMAO ($1.60\cdot10^{-5} > 1.01\cdot10^{-5}$). According to the excluded volume effect, both of them are expected to reduce the diffusion and to have a destabilizing effect on the intermediate, which is bulkier compared to the monomer. However, the R_h of TMAO is greater than that of sulfate ion, which explains why the association rate of the former is smaller than that of the later.

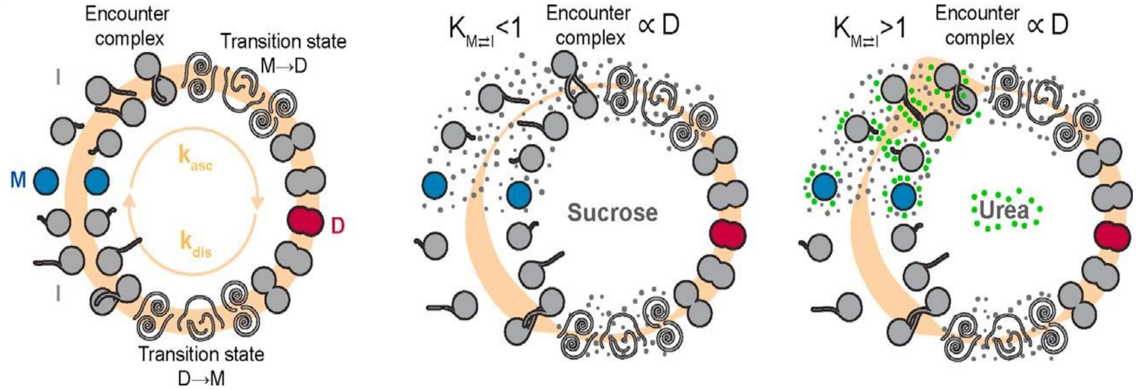


Figure R32. Diagram of the dimerization reaction of PL_G55A. In absence of cosolutes (left) two monomer molecules undergo partial unfolding to generate an intermediate that, once encountered a dimerization partner, evolves to generate a considerably unfolded transition state, which in last term progresses to dimer. Similarly, the dimer undergoes the opposite path. Orange line width indicates the contribution of the kinetic rates. In the case of sucrose (middle) and urea (right) small gray circles emphasize the steps in which the cosolute has a substantial impact on the equilibrium kinetics. In the case urea, small green circles represent preferential interactions that help the monomer in its unfolding process, thus facilitating the encounter between monomer units.

Finally, it is worth to mention that the data obtained in this work support the existence of the unfolded intermediates and the transition state, and it is assumed that they indeed exist in order to explain the experimental data. However, these assumptions could be confirmed by studying the stages in which both, the intermediates and the transition state are formed, but observing such species is hard to achieve due to their low populations and the timescale in which they exist. Due to this, the exact kinetic rates of the reactions that lead to their formation cannot be easily determined. Maybe some of the available approaches to quantify low-populated states (Baldwin, J., and Kay, L., 2009) would help solving this issue, but this aspect has been relegated to future work.

Chapter VI

Discussion

Discussion

In the context of the study about how nature has managed to adapt for surviving in extreme environments, the exhaustive analysis of the haloadaptation mechanism has shed light on the molecular traits required to overcome the loss of stability of the proteins that results from the adaptive strategy of increasing the cytosolic salt concentration. Out of all the molecular features identified as essential for the adaptation to the hypersaline conditions, the changes found in the protein surface composition are the most remarkable, since changing the surface composition results to be a simple, although effective and energetically efficient strategy to preserve the stability without altering the overall folding required for function. Taking these ideas into account, and also considering that the proteins from the halophilic organisms present an increased number of polar and acid amino acids in detriment of bulky and nonpolar amino acids, in this work it has been designed a metric that quantifies the degree of adaptation of a protein to the composition of the environment.

Throughout this work the results obtained in the analysis of the σ_{Prot} metric have demonstrated that this metric allows to discriminate organisms by the different degree of adaptation of their proteins to the environment (Figure R6), and to express the sensitivity of a protein to the composition of the environment as function of the composition of its surface (Figures R10 and R11). In fact, to some extent the σ_{Prot} metric reports on how an organism has adapted its quinary structure to a certain environment composition, since the protein surface is responsible of the transient interactions from which quinary structure results. Hence, this metric is useful to study the degree of adaptation of a protein to a certain environment, and by extension, of the organism that owns it.

Despite it seems to be a quite general descriptor of adaptation to the environment, the σ_{Prot} metric cannot be proposed as an absolute metric, but only as one of the possible metrics that can discriminate proteins by their tendence to be stabilized/destabilized by the composition of the environment. In fact, the σ_{Prot} metric results from the comparison between the composition of the surfaces of a small group of halophilic proteins (n = 49) and their mesophilic counterparts, but the resulting differences in the average frequencies for each amino acid (Table R1), used in the calculation of the σ_{Prot} metric, rather correspond to the sample means observed for the particular group of proteins considered in this work (Tables R2 and R3). Maybe a statistical estimation of the population values corresponding to such sample values would lead to the construction of a metric with the same generality but of a more absolute character.

Independently form these considerations, the σ_{Prot} metric proposed in this work is straightforward to calculate and only requires knowing the primary sequence of a protein, as can be observed from Equation R2. Due to this, the σ_{Prot} metric may be useful for optimizing the stability of protein formulations for industrial or pharmaceutical applications, for investigating the sensitivity to the environment of IDPs and their conformational landscapes (Moses, D., et al. 2020), or understand phenomena that occur in vivo, such as liquid-liquid phase separation or protein aggregation and, in a broader context, the relationship between protein stability, proteostasis and quinary structure.

As an example, in this work it has been studied the effect of the environment, represented by a group of rationally chosen cosolutes, on the kinetics and thermodynamics of the oligomerization process of PL_G55A. The results obtained in the context of such study, and the existing theoretical framework, allow to conclude that the composition of the medium modulates the thermodynamic equilibrium of an oligomerization reaction not only by changes in the protein stability, something that could be deduced from the results obtained for urea, but also by modifying the hydrodynamic properties of the solution, which impacts some critical stages of the reaction such as the formation of encounter dimers. Despite it does not provide new insight, this clear example provides experimental evidence on the mechanism by which the oligomerization processes are regulated by the composition of the environment. According to this, the tools that provide information about the sensitivity of a protein to the medium such as the σ_{Prot} metric could become of great value, since some main cellular function such as the maintenance of proteostasis depend on oligomerization processes.

To conclude, all the results and ideas already discussed throughout this work can be summarized in a model that explains the adaptation of the proteins to the medium. In this model, the adaptation to the exact composition of a certain environment is achieved through changes in the composition of the protein surface, although the surface composition chosen depends on the most concentrated component that is present in the medium and on its dominating mechanism. This model results from the fact that solutes affect the stability of proteins mainly by preferential interaction and excluded volume, two effects that are dependent on the concentration and physicochemical properties. Finally, it has been proven that the environment exerts a determinant effect on the kinetics and thermodynamics of the oligomerization processes that occur through domain swapping. Maybe the σ_{Prot} metric that derives from this model of adaptation could be a valuable tool for better understanding how nature manages to preserve the quinary structure and, in particular, the proteostasis network, since the later depends on the oligomerization processes. This final consideration could be the object of a future research.

Conclusions

Conclusions

- A quantitative metric (σ_{Prot}) can be calculated from the primary structure taking in consideration only the comparison between the surface compositions of halophilic proteins and their mesophilic orthologs.

- The σ_{Prot} metric allows to distinguish and classify organisms according to their degree of adaptation to the environment without performing any previous genomic analysis nor sequence alignment.

- The σ_{Prot} metric reports on the sensitivity of the proteins to the presence of a small group of biologically relevant cosolutes. Similarly, the σ_{Prot} metric reports on the sensitivity of the proteins to the cellular milieu.

- The presence of cosolutes alter the kinetics and the populations at the apparent equilibrium of the dimerization reaction of PL_G55A.

- When the oligomerization reaction of PL_G55A is started from the dimer, the monomer and dimer populations at apparent equilibrium do not change. Despite the kinetic rates are altered by the cosolutes, their ratio remains constant.

- When the same reaction is started from the monomer, the monomer and dimer populations depend on the type and concentration of the cosolute present. This is a consequence of the dependence of the association stage on the encounter rate, which gets reduced in presence of cosolutes thus kinetically trapping the reaction.

References

- Abragam, Anatole. *The principles of nuclear magnetism*. No. 32. Oxford university press, 1961.
- Ainavarapu, Sri Rama Koti, et al. "Ligand binding modulates the mechanical stability of dihydrofolate reductase." *Biophysical journal* 89.5 (2005): 3337-3344.
- Ali, Mayssam H., and Barbara Imperiali. "Protein oligomerization: how and why." *Bioorganic & medicinal chemistry* 13.17 (2005): 5013-5020.
- Andersen, Hans C. "Molecular dynamics simulations at constant pressure and/or temperature." *The Journal of chemical physics* 72.4 (1980): 2384-2393.
- Andrea, Tariq A., William C. Swope, and Hans C. Andersen. "The role of long ranged forces in determining the structure and properties of liquid water." *The Journal of chemical physics* 79.9 (1983): 4576-4584.
- Anfinsen, Christian B. "Principles that govern the folding of protein chains." *Science* 181.4096 (1973): 223-230.
- Arakawa, Tsutomu, and Serge N. Timasheff. "Mechanism of polyethylene glycol interaction with proteins." *Biochemistry* 24.24 (1985): 6756-6762.
- Arakawa, Tsutomu, and Serge N. Timasheff. "Preferential interactions of proteins with salts in concentrated solutions." *Biochemistry* 21.25 (1982): 6545-6552.
- Auton, Matthew, and D. Wayne Bolen. "Predicting the energetics of osmolyte-induced protein folding/unfolding." *Proceedings of the National Academy of Sciences* 102.42 (2005): 15065-15068.
- Bai, Jia, et al. "Macromolecular and Small Molecular Crowding Have Similar Effects on α-Synuclein Structure." *ChemPhysChem* 18.1 (2017): 55-58.
- Balch, William E., et al. "Adapting proteostasis for disease intervention." *science* 319.5865 (2008): 916-919.
- Balchin, David, Manajit Hayer-Hartl, and F. Ulrich Hartl. "In vivo aspects of protein folding and quality control." *Science* 353.6294 (2016).
- Baldwin, Andrew J., and Lewis E. Kay. "NMR spectroscopy brings invisible protein states into focus." *Nature chemical biology* 5.11 (2009): 808-814.
- Baldwin, Robert L. "How Hofmeister ion interactions affect protein stability." *Biophysical journal* 71.4 (1996): 2056-2063.
- Baptista, R. P., J. M. S. Cabral, and E. P. Melo. "Trehalose delays the reversible but not the irreversible thermal denaturation of cutinase." *Biotechnology and bioengineering* 70.6 (2000): 699-703.
- Becktel, Wayne J., and John A. Schellman. "Protein stability curves." *Biopolymers: Original Research on Biomolecules* 26.11 (1987): 1859-1877.
- Bennett, M. J., and David Eisenberg. "Refined structure of monomelic diphtheria toxin at 2.3 Å resolution." *Protein Science* 3.9 (1994): 1464-1475.
- Bennett, Melanie J., Michael P. Schlunegger, and David Eisenberg. "3D domain swapping: a mechanism for oligomer assembly." *Protein science* 4.12 (1995): 2455-2468.

- Bernardo-Seisdedos, Ganeko, et al. "Structural basis and energy landscape for the Ca2+ gating and calmodulation of the Kv7. 2 K+ channel." *Proceedings of the National Academy of Sciences* 115.10 (2018): 2395-2400.
- Bezsonova, Irina, et al. "Structural comparison of the unstable drkN SH3 domain and a stable mutant." *Biochemistry* 44.47 (2005): 15550-15560.
- Booth, David R., et al. "Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis." *Nature* 385.6619 (1997): 787-793.
- Braig, Kerstin, et al. "The crystal structure of the bacterial chaperonln GroEL at 2.8 Å." *Nature* 371.6498 (1994): 578-586.
- Braun, Pascal, Edward Rietman, and Marc Vidal. "Networking metabolites and diseases." *Proceedings of the National Academy of Sciences* 105.29 (2008): 9849-9850.
- Breydo, Leonid, et al. "The crowd you're in with: effects of different types of crowding agents on protein aggregation." *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1844.2 (2014): 346-357.
- Bruździak, Piotr, et al. "Taurine as a water structure breaker and protein stabilizer." *Amino acids* 50.1 (2018): 125-140.
- Budenholzer, Lauren, et al. "Proteasome structure and assembly." *Journal of molecular biology* 429.22 (2017): 3500-3524.
- Bukau, Bernd, Jonathan Weissman, and Arthur Horwich. "Molecular chaperones and protein quality control." *Cell* 125.3 (2006): 443-451.
- Cacciapuoti, Giovanna, et al. "Purification and characterization of extremely thermophilic and thermostable 5'-methylthioadenosine phosphorylase from the archaeon Sulfolobus solfataricus. Purine nucleoside phosphorylase activity and evidence for intersubunit disulfide bonds." *Journal of Biological Chemistry* 269.40 (1994): 24762-24769.
- Cacciapuoti, Giovanna, et al. "Role of disulfide bonds in conformational stability and folding of 5′-deoxy-5′-methylthioadenosine phosphorylase II from the hyperthermophilic archaeon Sulfolobus solfataricus." *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1824.10 (2012): 1136-1143.
- Campbell, Iain. *Biophysical techniques*. Oxford University Press, 2012.
- Chan, Chi-Ho, Tsz-Ha Yu, and Kam-Bo Wong. "Stabilizing salt-bridge enhances protein thermostability by reducing the heat capacity change of unfolding." *PloS one* 6.6 (2011): e21624.
- Chen, Calvin R., and George I. Makhatadze. "ProteinVolume: calculating molecular van der Waals and void volumes in proteins." *BMC bioinformatics* 16.1 (2015): 1-6.
- Christian, J. H. B., and Judith A. Waltho. "Solute concentrations within cells of halophilic bacteria." (1962).
- Claridge, Timothy DW. *High-resolution NMR techniques in organic chemistry*. Vol. 27. Elsevier, 2016.
- Cohen, Ehud, et al. "Opposing activities protect against age-onset proteotoxicity." *Science* 313.5793 (2006): 1604-1610.
- Cohen, Rachel D., and Gary J. Pielak. "A cell is more than the sum of its (dilute) parts: A brief history of quinary structure." *Protein Science* 26.3 (2017): 403-413.

- Cohen, Rachel D., and Gary J. Pielak. "Electrostatic contributions to protein quinary structure." *Journal of the American Chemical Society* 138.40 (2016): 13139-13142.
- Cordeiro, Tiago N., et al. "Indirect DNA readout by an H-NS related protein: structure of the DNA complex of the C-terminal domain of Ler." *PLoS pathogens* 7.11 (2011): e1002380.
- Creamer, Trevor P., Rajgopal Srinivasan, and George D. Rose. "Modeling unfolded states of proteins and peptides. II. Backbone solvent accessibility." *Biochemistry* 36.10 (1997): 2832-2835.
- Crestefield, Arthur M., William H. Stein, and Stanford Moore. "On the aggregation of bovine pancreatic ribonuclease." *Archives of biochemistry and biophysics* (1962): 217-222.
- Darden, Tom, Darrin York, and Lee Pedersen. "Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems." *The Journal of chemical physics* 98.12 (1993): 10089-10092.
- delCardayré, Stephen B., et al. "Engineering ribonuclease A: production, purification and characterization of wild-type enzyme and mutants at Gln11." *Protein Engineering, Design and Selection* 8.3 (1995): 261-273.
- Dennis, Patrick P., and Lawrence C. Shimmin. "Evolutionary divergence and salinity-mediated selection in halophilic archaea." *Microbiology and Molecular Biology Reviews* 61.1 (1997): 90-104.
- Der, Bryan S., et al. "Alternative computational protocols for supercharging protein surfaces for reversible unfolding and retention of stability." *PloS one* 8.5 (2013): e64363.
- Di Domenico, Roberto, and Roberto Lavecchia. "Thermal stability of human haemoglobin in the presence of sarcosine and sorbitol." *Biotechnology letters* 22.5 (2000): 335-339.
- Dill, Ken A. "Dominant forces in protein folding." *Biochemistry* 29.31 (1990): 7133-7155.
- Doucleff, Michaeleen, Mary Hatcher-Skeers, and Nicole J. Crane. *Pocket guide to biomolecular NMR*. Springer Science & Business Media, 2011.
- Downing, Kenneth H., and Eva Nogales. "Tubulin and microtubule structure." *Current opinion in cell biology* 10.1 (1998): 16-22.
- Ebel, Christine, et al. "Relative role of anions and cations in the stabilization of halophilic malate dehydrogenase." *Biochemistry* 38.28 (1999): 9039-9047.
- Ellis, R. John, and Allen P. Minton. "Protein aggregation in crowded environments." (2006): 485-497.
- Ernst, Richard R., Geoffrey Bodenhausen, and Alexander Wokaun. *Principles of nuclear magnetic resonance in one and two dimensions*. Vol. 14. Oxford: Clarendon press, 1987.
- Felli, Isabella C., and Roberta Pierattelli, eds. *Intrinsically disordered proteins studied by NMR spectroscopy*. Vol. 870. Springer, 2015.
- Fierke, Carol A., Kenneth A. Johnson, and Stephen J. Benkovic. "Construction and evaluation of the kinetic scheme associated with dihydrofolate reductase from Escherichia coli." *Biochemistry* 26.13 (1987): 4085-4092.

- Fortian, Arola, et al. "Intracellular rescue of the uroporphyrinogen III synthase activity in enzymes carrying the hotspot mutation C73R." *Journal of Biological Chemistry* 286.15 (2011): 13127-13133.
- Fortian, Arola, et al. "Uroporphyrinogen III synthase mutations related to congenital erythropoietic porphyria identify a key helix for protein stability." *Biochemistry* 48.2 (2009): 454-461.
- Fukuchi, Satoshi, et al. "Unique amino acid composition of proteins in halophilic bacteria." *Journal of molecular biology* 327.2 (2003): 347-357.
- Ganguly, Pritam, et al. "Protein stability in TMAO and mixed urea–TMAO solutions." *The Journal of Physical Chemistry B* 124.29 (2020): 6181-6197.
- Garcia-Seisdedos, Hector, et al. "Proteins evolve on the edge of supramolecular self-assembly." *Nature* 548.7666 (2017): 244-247.
- Gardner, Kevin H., et al. "Solution NMR Studies of a 42 KDa Escherichia C oli Maltose Binding Protein/β-Cyclodextrin Complex: Chemical Shift Assignments and Analysis." *Journal of the American Chemical Society* 120.45 (1998): 11738-11748.
- Gil-Martínez, Jon, et al. "Therapeutic targeting of fumaryl acetoacetate hydrolase in hereditary tyrosinemia type i." *International journal of molecular sciences* 22.4 (2021): 1789.
- Goh, Kwang-Il, et al. "The human disease network." *Proceedings of the National Academy of Sciences* 104.21 (2007): 8685-8690.
- Gomis-Rüth, F. Xavier, et al. "The structure of plasmid-encoded transcriptional repressor CopG unliganded and bound to its operator." *The EMBO journal* 17.24 (1998): 7404-7415.
- Guin, Drishti, and Martin Gruebele. "Weak chemical interactions that drive protein evolution: crowding, sticking, and quinary structure in folding and function." *Chemical reviews* 119.18 (2019): 10691-10717.
- Hardy, Robert C., and Robert L. Cottington. "Viscosity of deuterium oxide and water in the range 5 to 125 C." *J. Res. Natl. Bur. Stand* 42.1934 (1949): 573.
- Hiller, Nicole, et al. "Plasmodium falciparum glutathione S-transferase—Structural and mechanistic studies on ligand binding and enzyme inhibition." *Protein Science* 15.2 (2006): 281-289.
- Hirota, Shun. "Oligomerization of cytochrome c, myoglobin, and related heme proteins by 3D domain swapping." *Journal of inorganic biochemistry* 194 (2019): 170-179.
- Hoffman, Laurel, et al. "Relative cosolute size influences the kinetics of protein-protein interactions." *Biophysical journal* 109.3 (2015): 510-520.
- Hofmeister, F. "About the science of the effects of salts: About the water withdrawing effect of the salts." *Arch. Exp. Pathol. Pharmacol.* 24 (1888): 247-260.
- Hore, Peter J., Jonathan A. Jones, and Stephen Wimperis. *NMR: The toolkit: How pulse sequences work*. Vol. 92. Oxford University Press, USA, 2015.
- Huston, Adrienne L., Jesper Z. Haeggström, and Georges Feller. "Cold adaptation of enzymes: Structural, kinetic and microcalorimetric characterizations of an aminopeptidase from the Arctic psychrophile Colwellia psychrerythraea and of human leukotriene A4 hydrolase." *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1784.11 (2008): 1865-1872.

- Ibarra-Molero, Beatriz, et al. "The sarcosine effect on protein stability: a case of nonadditivity?." *Protein Science* 9.4 (2000): 820-826.
- Jaenicke, Rainer, and Gerald Böhm. "The stability of proteins in extreme environments." *Current opinion in structural biology* 8.6 (1998): 738-748.
- Janowski, Robert, et al. "Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping." *Nature structural biology* 8.4 (2001): 316-320.
- Jiao, Ming, et al. "Attractive protein-polymer interactions markedly alter the effect of macromolecular crowding on protein association equilibria." *Biophysical journal* 99.3 (2010): 914-923.
- Jorgensen, William L., et al. "Comparison of simple potential functions for simulating liquid water." *The Journal of chemical physics* 79.2 (1983): 926-935.
- Karshikoff, Andrey, and Rudolf Ladenstein. "Ion pairs and the thermotolerance of proteins from hyperthermophiles: a 'traffic rule'for hot roads." *Trends in biochemical sciences* 26.9 (2001): 550-557.
- Keeler, James. *Understanding NMR spectroscopy*. John Wiley & Sons, 2011.
- Kwon, Yong-Chan, and Michael C. Jewett. "High-throughput preparation methods of crude extract for robust cell-free protein synthesis." *Scientific reports* 5.1 (2015): 1-8.
- Lanyi, Janos K. "Salt-dependent properties of proteins from extremely halophilic bacteria." *Bacteriological reviews* 38.3 (1974): 272-290.
- Lee, Chi-Fung, George I. Makhatadze, and Kam-Bo Wong. "Effects of charge-to-alanine substitutions on the stability of ribosomal protein L30e from Thermococcus celer." *Biochemistry* 44.51 (2005): 16817-16825.
- Lee, James C., and Serge N. Timasheff. "The stabilization of proteins by sucrose." *Journal of Biological Chemistry* 256.14 (1981): 7193-7201.
- Lee, Jongsoon, and Paul F. Pilch. "The insulin receptor: structure, function, and signaling." *American Journal of Physiology-Cell Physiology* 266.2 (1994): C319-C334.
- Lee, Tai-Sung, et al. "GPU-accelerated molecular dynamics and free energy methods in Amber18: performance enhancements and new features." *Journal of chemical information and modeling* 58.10 (2018): 2043-2050.
- Levitt, Malcolm H. *Spin dynamics: basics of nuclear magnetic resonance*. John Wiley & Sons, 2013.
- Liao, Yi-Ting, et al. "Trimethylamine N-oxide stabilizes proteins via a distinct mechanism compared with betaine and glycine." *Proceedings of the National Academy of Sciences* 114.10 (2017): 2479-2484.
- Lins, Laurence, and Robert Brasseur. "The hydrophobic effect in protein folding." *The FASEB journal* 9.7 (1995): 535-540.
- Liu, Lin, et al. "Domain swapping proceeds via complete unfolding: a 19F-and 1H-NMR study of the Cyanovirin-N protein." *Journal of the American Chemical Society* 134.9 (2012): 4229-4235.
- Liu, Yanshun, and David Eisenberg. "3D domain swapping: as domains continue to swap." *Protein science* 11.6 (2002): 1285-1299.

- Ma, Jianqiang, Ileana M. Pazos, and Feng Gai. "Microscopic insights into the protein-stabilizing effect of trimethylamine N-oxide (TMAO)." *Proceedings of the National Academy of Sciences* 111.23 (2014): 8476-8481.
- Ma, Yanhe, et al. "Halophiles 2010: life in saline environments." *Applied and environmental microbiology* 76.21 (2010): 6971-6981.
- Macias, Iratxe, et al. "Hereditary tyrosinemia type I–associated mutations in fumarylacetoacetate hydrolase reduce the enzyme stability and increase its aggregation rate." *Journal of Biological Chemistry* 294.35 (2019): 13051-13060.
- Maier, James A., et al. "ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB." *Journal of chemical theory and computation* 11.8 (2015): 3696-3713.
- Marcus, Yizhak. "The guanidinium ion." *The Journal of Chemical Thermodynamics* 48 (2012): 70-74.
- Mateos, Borja, et al. "Cosolute modulation of protein oligomerization reactions in the homeostatic timescale." *Biophysical Journal* 120.10 (2021): 2067-2077.
- Mathews, Michael AA, et al. "Crystal structure of human uroporphyrinogen III synthase." *The EMBO journal* 20.21 (2001): 5832-5839.
- McConkey, Edwin H. "Molecular evolution, intracellular organization, and the quinary structure of proteins." *Proceedings of the National Academy of Sciences* 79.10 (1982): 3236-3240.
- McFeeters, Robert L., et al. "The high-precision solution structure of Yersinia modulating protein YmoA provides insight into interaction with H-NS." *Biochemistry* 46.49 (2007): 13975-13982.
- Miao, Yinglong, et al. "Improved reweighting of accelerated molecular dynamics simulations for free energy calculation." *Journal of chemical theory and computation* 10.7 (2014): 2677-2689.
- Miao, Yinglong, Victoria A. Feher, and J. Andrew McCammon. "Gaussian accelerated molecular dynamics: Unconstrained enhanced sampling and free energy calculation." *Journal of chemical theory and computation* 11.8 (2015): 3584-3595.
- Miklos, Andrew C., et al. "Protein crowding tunes protein stability." *Journal of the American Chemical Society* 133.18 (2011): 7116-7120.
- Minton, Allen P. "Excluded volume as a determinant of protein structure and stability." *Biophysical journal* 32.1 (1980): 77.
- Minton, Allen P. "[7] Molecular crowding: analysis of effects of high concentrations of inert cosolutes on biochemical equilibria and rates in terms of volume exclusion." *Methods in enzymology* 295 (1998): 127-149.
- Miyamoto, Shuichi, and Peter A. Kollman. "Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models." *Journal of computational chemistry* 13.8 (1992): 952-962.
- Monteith, William B., et al. "Quinary structure modulates protein stability in cells." *Proceedings of the National Academy of Sciences* 112.6 (2015): 1739-1742.
- Moré, Jorge J. "The Levenberg-Marquardt algorithm: implementation and theory." *Numerical analysis*. Springer, Berlin, Heidelberg, 1978. 105-116.

- Morimoto, Richard I. "Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators." *Genes & development* 12.24 (1998): 3788-3796.
- Moschen, Thomas, and Martin Tollinger. "A kinetic study of domain swapping of Protein L." *Physical Chemistry Chemical Physics* 16.14 (2014): 6383-6390.
- Moses, David, et al. "Revealing the hidden sensitivity of intrinsically disordered proteins to their chemical environment." *The Journal of Physical Chemistry Letters* 11.23 (2020): 10131-10136.
- Myers, Jeffrey K., C. Nick Pace, and J. Martin Scholtz. "Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding." *Protein Science* 4.10 (1995): 2138-2148.
- Neuhaus, David. "Nuclear overhauser effect." *eMagRes* (2007).
- Newberry, Robert W., and Ronald T. Raines. "Secondary forces in protein folding." *ACS chemical biology* 14.8 (2019): 1677-1686.
- Nielsen, Jakob Toudahl, and Frans AA Mulder. "POTENCI: prediction of temperature, neighbor and pH-corrected chemical shifts for intrinsically disordered proteins." *Journal of biomolecular NMR* 70.3 (2018): 141-165.
- O'Neill, Jason W., et al. "Single-site mutations induce 3D domain swapping in the B1 domain of protein L from Peptostreptococcus magnus." *Structure* 9.11 (2001): 1017-1027.
- Ortega, Gabriel, et al. "Carbohydrate affinity for the glucose–galactose binding protein is regulated by allosteric domain motions." *Journal of the American Chemical Society* 134.48 (2012): 19869-19876.
- Ortega, Gabriel, et al. "Halophilic enzyme activation induced by salts." *Scientific reports* 1.1 (2011): 1-6.
- Ortega, Gabriel, Tammo Diercks, and Oscar Millet. "Halophilic protein adaptation results from synergistic residue-ion interactions in the folded and unfolded states." *Chemistry & biology* 22.12 (2015): 1597-1607.
- Pace, C. Nick, et al. "Contribution of hydrophobic interactions to protein stability." *Journal of molecular biology* 408.3 (2011): 514-528.
- Patel, Chetan N., et al. "Effects of molecular crowding by saccharides on α-chymotrypsin dimerization." *Protein science* 11.5 (2002): 997-1003.
- Paul, Sandip, et al. "Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes." *Genome biology* 9.4 (2008): 1-19.
- Perutz, M. F. "Mutations make enzyme polymerize." *Nature* 385.6619 (1997): 773-775.
- Phillip, Yael, and Gideon Schreiber. "Formation of protein complexes in crowded environments–from in vitro to in vivo." *FEBS letters* 587.8 (2013): 1046-1052.
- Pincus, David L., Changbong Hyeon, and D. Thirumalai. "Effects of trimethylamine N-oxide (TMAO) and crowding agents on the stability of RNA hairpins." *Journal of the American Chemical Society* 130.23 (2008): 7364-7372.
- Polosina, Yaroslava Y., et al. "Stability of Natrialba magadii NDP kinase: comparisons with other halophilic proteins." *Extremophiles* 6.2 (2002): 135-142.

- Price, William S., Hiroyuki Ide, and Yoji Arata. "Translational and rotational motion of isolated water molecules in nitromethane studied using 17 O NMR." *The Journal of Chemical Physics* 113.9 (2000): 3686-3689.
- Quintanilla, Gabriel Ortega. *Structural studies of the hipersaline adaptation of proteins belonging to halophilic archaea*. Diss. Universidad del País Vasco-Euskal Herriko Unibertsitatea, 2015.
- Ratnaparkhi, Girish S., and R. Varadarajan. "Thermodynamic and structural studies of cavity formation in proteins suggest that loss of packing interactions rather than the hydrophobic effect dominates the observed energetics." *Biochemistry* 39.40 (2000): 12365-12374.
- Record, M. Thomas, Charles F. Anderson, and Timothy M. Lohman. "Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity." *Quarterly reviews of biophysics* 11.2 (1978): 103-178.
- Roman, Ernesto A., et al. "Protein stability and dynamics modulation: the case of human frataxin." (2012): e45743.
- Ron, David, and Peter Walter. "Signal integration in the endoplasmic reticulum unfolded protein response." *Nature reviews Molecular cell biology* 8.7 (2007): 519-529.
- Roobottom, Helen K., et al. "Thermochemical radii of complex ions." *Journal of chemical education* 76.11 (1999): 1570.
- Rule, Gordon S., and T. Kevin Hitchens. *Fundamentals of protein NMR spectroscopy*. Vol. 5. Springer Science & Business Media, 2006.
- Santoro, Marcelo M., and D. W. Bolen. "Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl. alpha-chymotrypsin using different denaturants." *Biochemistry* 27.21 (1988): 8063-8068.
- Sarkar, Mohona, Austin E. Smith, and Gary J. Pielak. "Impact of reconstituted cytosol on protein stability." *Proceedings of the National Academy of Sciences* 110.48 (2013): 19342-19347.
- Schellman, John A. "Protein stability in mixed solvents: a balance of contact interaction and excluded volume." *Biophysical journal* 85.1 (2003): 108-125.
- Schreiber, Gideon. "Kinetic studies of protein–protein interactions." *Current opinion in structural biology* 12.1 (2002): 41-47.
- Schultz, Stanley G., and A. K. Solomon. "Determination of the effective hydrodynamic radii of small molecules by viscometry." *The Journal of general physiology* 44.6 (1961): 1189-1199.
- Seelig, Joachim, and Hans-Joachim Schönfeld. "Thermal protein unfolding by differential scanning calorimetry and circular dichroism spectroscopy Two-state model versus sequential unfolding." *Quarterly reviews of biophysics* 49 (2016).
- Sherwin, Keith E., and Donald J. Winzor. "Effect of sucrose on the dimerization of α-chymotrypsin allowance for thermodynamic nonideality arising from the presence of a small inert solute." *Biophysical chemistry* 31.3 (1988): 287-294.
- Siddiqui, Khawar S., et al. "Psychrophiles." *Annual Review of Earth and Planetary Sciences* 41 (2013): 87-115.

- Sinnaeve, Davy. "The Stejskal–Tanner equation generalized for any gradient shape—an overview of most pulse sequences measuring free diffusion." *Concepts in Magnetic Resonance Part A* 40.2 (2012): 39-65.
- Sikosek, Tobias, and Hue Sun Chan. "Biophysics of protein evolution and evolutionary protein biophysics." *Journal of The Royal Society Interface* 11.100 (2014): 20140419.
- Sluchanko, Nikolai N., Natalia A. Chebotareva, and Nikolai B. Gusev. "Quaternary structure of human small heat shock protein HSPB6 (Hsp20) in crowded media modeled by trimethylamine N-oxide (TMAO): Effect of protein phosphorylation." *Biochimie* 108 (2015): 68-75.
- Smith, Austin E., et al. "In-cell thermodynamics and a new role for protein surfaces." *Proceedings of the National Academy of Sciences* 113.7 (2016): 1725-1730.
- Sobala, Lukasz F., et al. "An epoxide intermediate in glycosidase catalysis." *ACS central science* 6.5 (2020): 760-770.
- Solyom, Zsofia, et al. "BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins." *Journal of biomolecular NMR* 55.4 (2013): 311-321.
- Somero, George N. "Protons, osmolytes, and fitness of internal milieu for protein function." *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 251.2 (1986): R197-R213.
- Stadmiller, Samantha S., et al. "Osmotic shock induced protein destabilization in living cells and its reversal by glycine betaine." *Journal of molecular biology* 429.8 (2017): 1155-1161.
- Tadeo, Xavier, David Castaño, and Oscar Millet. "Anion modulation of the 1H/2H exchange rates in backbone amide protons monitored by NMR spectroscopy." *Protein Science* 16.12 (2007): 2733-2740.
- Tadeo, Xavier, et al. "Protein stabilization and the Hofmeister effect: the role of hydrophobic solvation." *Biophysical journal* 97.9 (2009a): 2595-2603.
- Tadeo, Xavier, et al. "Structural basis for the aminoacid composition of proteins from halophilic archea." *PLoS biology* 7.12 (2009b): e1000257.
- Tadeo, Xavier, Miquel Pons, and Oscar Millet. "Influence of the Hofmeister anions on protein stability as studied by thermal denaturation and chemical shift perturbation." *Biochemistry* 46.3 (2007): 917-923.
- Taparowsky, Elizabeth, et al. "Structure and activation of the human N-ras gene." *Cell* 34.2 (1983): 581-586.
- Theillet, Francois-Xavier, et al. "Physicochemical properties of cells and their effects on intrinsically disordered proteins (IDPs)." *Chemical reviews* 114.13 (2014): 6661-6714.
- Theillet, Francois-Xavier, et al. "Site-specific NMR mapping and time-resolved monitoring of serine and threonine phosphorylation in reconstituted kinase reactions and mammalian cell extracts." *Nature protocols* 8.7 (2013): 1416-1432.
- Theriot, Julie A., and Timothy J. Mitchison. "Actin microfilament dynamics in locomoting cells." *Nature* 352.6331 (1991): 126-131.

- Thompson, Andrew J., et al. "Structural and mechanistic insight into N-glycan processing by endo-α-mannosidase." *Proceedings of the National Academy of Sciences* 109.3 (2012): 781-786.
- Timm, David E., et al. "Crystal structure and mechanism of a carbon–carbon bond hydrolase." *Structure* 7.9 (1999): 1023-1033.
- Toth, Marta, Sergei B. Vakulenko, and Clyde A. Smith. "Purification, crystallization and preliminary X-ray analysis of the β-lactamase Oih-1 from Oceanobacillus iheyensis." *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* 65.6 (2009): 582-585.
- Tsukamoto, Takashi, et al. "X-ray crystallographic structure of thermophilic rhodopsin: implications for high thermal stability and optogenetic function." *Journal of Biological Chemistry* 291.23 (2016): 12223-12232.
- Urquiza, Pedro, et al. "Repurposing ciclopirox as a pharmacological chaperone in a model of congenital erythropoietic porphyria." *Science translational medicine* 10.459 (2018).
- Vlassi, Metaxia, Gianni Cesareni, and Michael Kokkinidis. "A correlation between the loss of hydrophobic core packing interactions and protein stability." *Journal of molecular biology* 285.2 (1999): 817-827.
- Vranken, Wim F., et al. "The CCPN data model for NMR spectroscopy: development of a software pipeline." *Proteins: structure, function, and bioinformatics* 59.4 (2005): 687-696.
- Wang, Conan K., et al. "Translational diffusion of cyclic peptides measured using pulsed-field gradient NMR." *The Journal of Physical Chemistry B* 118.38 (2014): 11129-11136.
- Wang, Wei, Sandeep Nema, and Dirk Teagarden. "Protein aggregation—pathways and influencing factors." *International journal of pharmaceutics* 390.2 (2010): 89-99.
- Ward, Colin W., and Michael C. Lawrence. "Ligand-induced activation of the insulin receptor: a multi-step process involving structural changes in both the ligand and the receptor." *Bioessays* 31.4 (2009): 422-434.
- Werner, Eric D., Jeffrey L. Brodsky, and Ardythe A. McCracken. "Proteasome-dependent endoplasmic reticulum-associated protein degradation: an unconventional route to a familiar fate." *Proceedings of the National Academy of Sciences* 93.24 (1996): 13797-13801.
- Wiegel, Juergen, and Adams WW Michael, eds. *Thermophiles: the keys to the molecular evolution and the origin of life?*. CRC Press, 1998.
- Yancey, Paul H., et al. "Living with water stress: evolution of osmolyte systems." *Science* 217.4566 (1982): 1214-1222.
- Zeeman, Pieter. "Over de invloed eener magnetisatie op den aard van het door een stof uitgezonden licht; Over de invloed eener magnetisatie op den aard van het door een stof uitgezonden licht; On the influence of magnetism on the nature of the light emitted by a substance." *Verslagen en Mededeelingen der Kon. Academie van Wetenschappen, Afd. Natuurkunde* 5 (1896): 181.
- Zhang, Ouwen, et al. "Backbone 1 H and 15 N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques." *Journal of biomolecular NMR* 4.6 (1994): 845-858.

- Zierenberg, Robert A., Michael WW Adams, and Alissa J. Arp. "Life in extreme environments: Hydrothermal vents." *Proceedings of the National Academy of Sciences* 97.24 (2000): 12961-12962.

Appendix

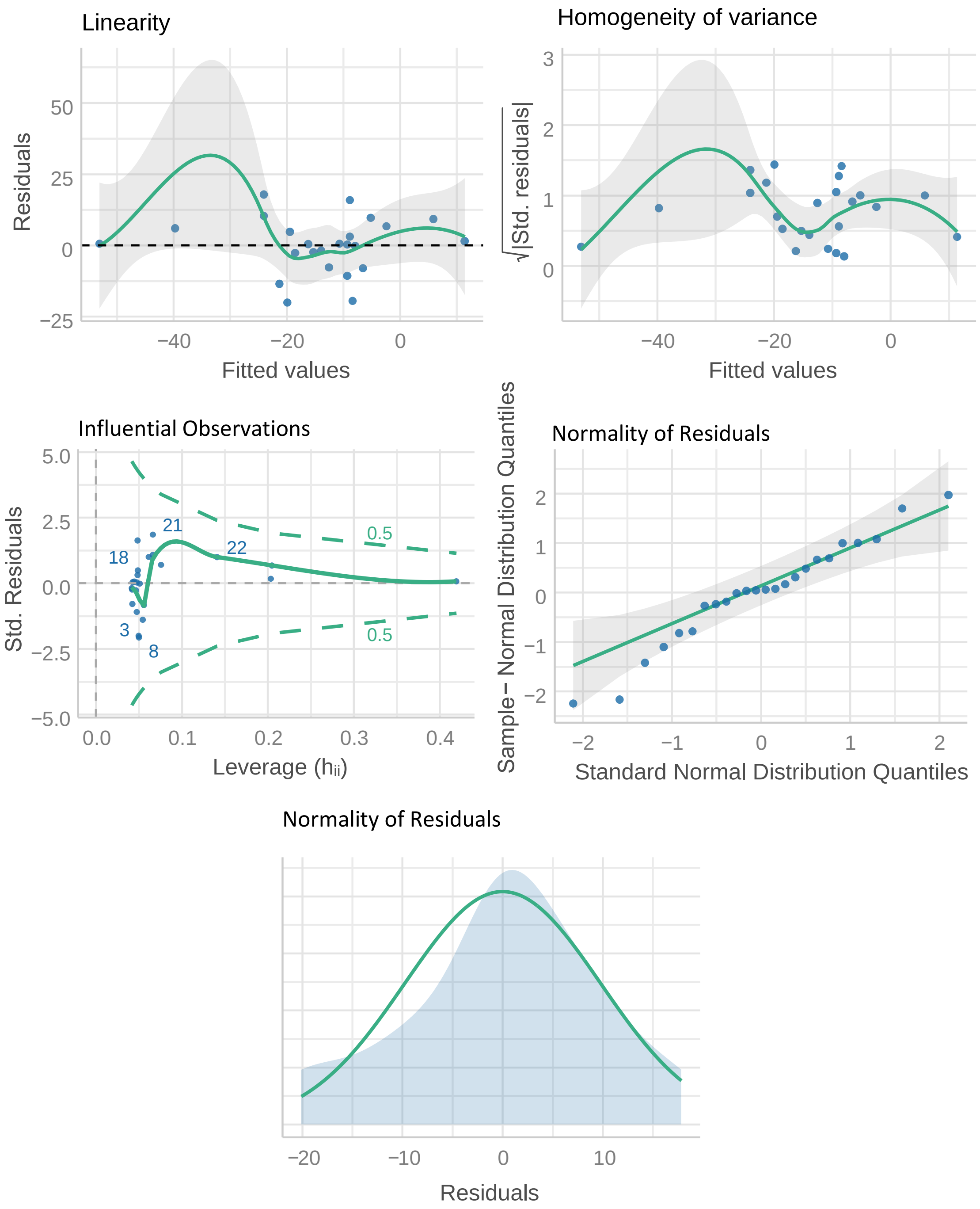


Figure A1. Statistical validation of the linear regression model for the thermal stability of $CaCl_2$. Blue points correspond to the experimental residuals observed, grey-shaded regions correspond to the confident interval, blue-shaded region corresponds to the experimental distribution of the residuals, solid green lines correspond to the fitting except for the normality plot, for which it corresponds to the theoretical normal distribution expected, and dashed green line encloses the region of the normal distribution to which corresponds a standard deviation of 0.5.

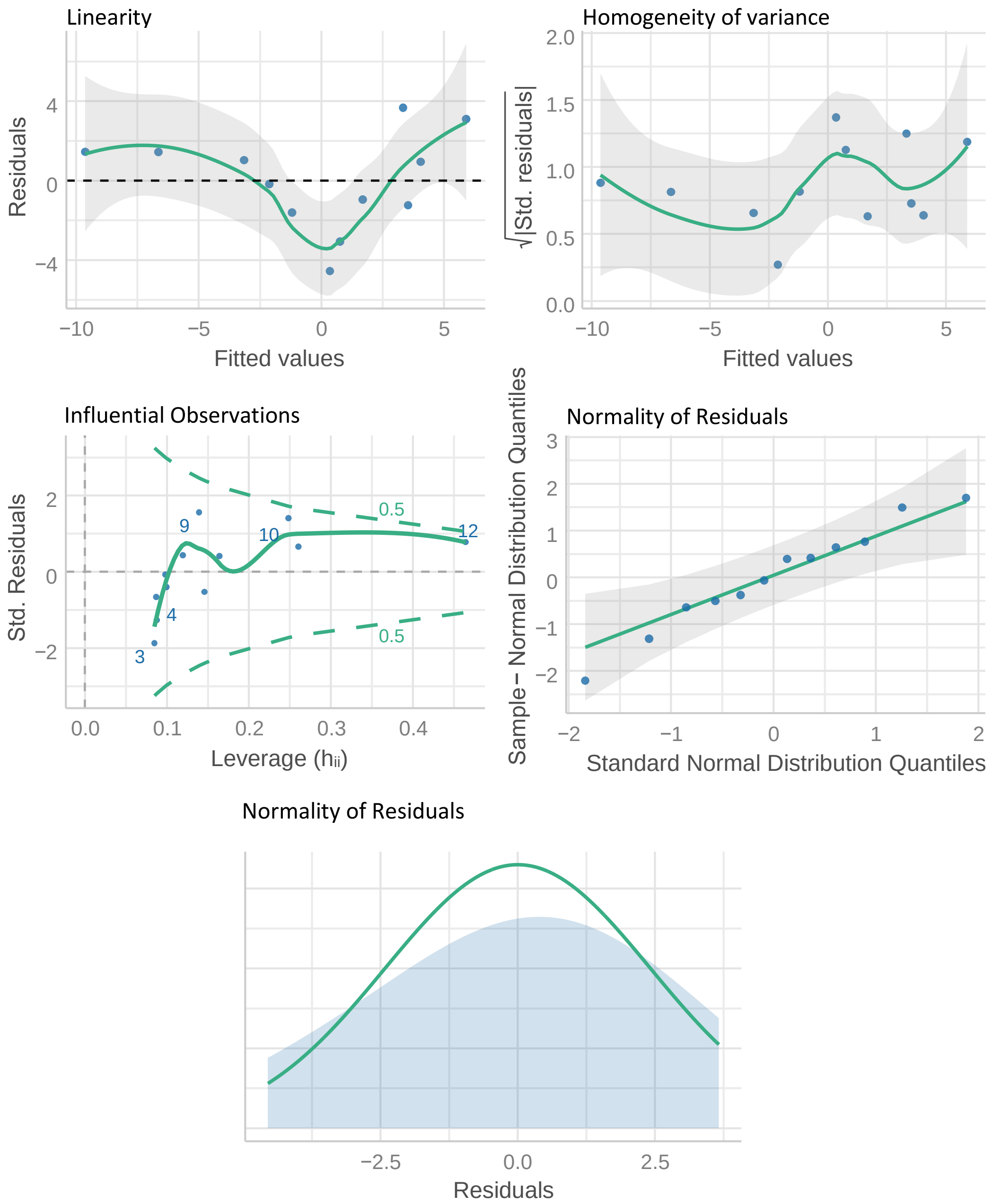


Figure A2. Statistical validation of the linear regression model for the thermodynamic stability of $CaCl_2$. Blue points correspond to the experimental residuals observed, grey-shaded regions correspond to the confident interval, blue-shaded region corresponds to the experimental distribution of the residuals, solid green lines correspond to the fitting except for the normality plot, for which it corresponds to the theoretical normal distribution expected, and dashed green line encloses the region of the normal distribution to which corresponds a standard deviation of 0.5.

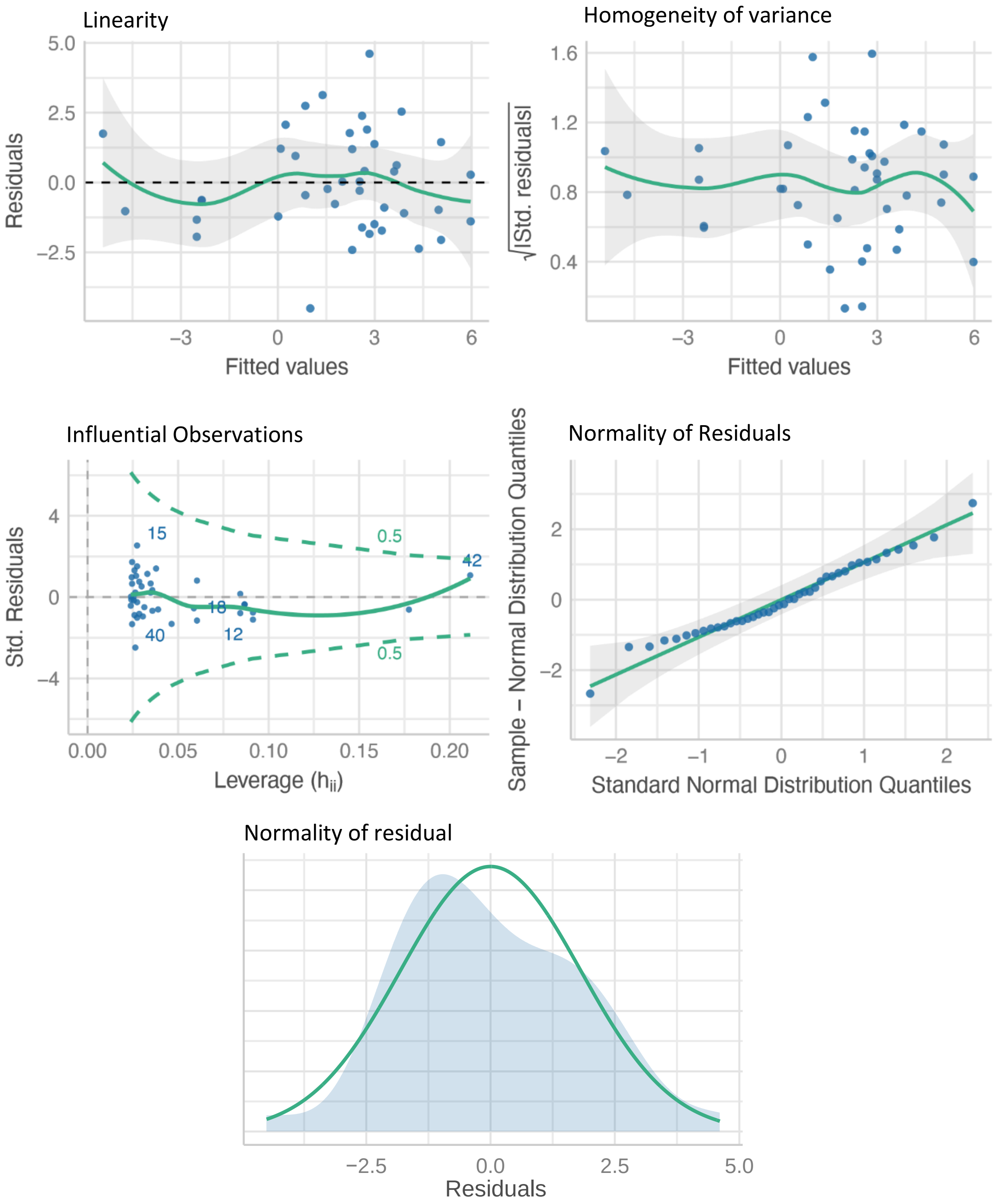


Figure A3. Statistical validation of the linear regression model for the thermal stability of KCl. Blue points correspond to the experimental residuals observed, grey-shaded regions correspond to the confident interval, blue-shaded region corresponds to the experimental distribution of the residuals, solid green lines correspond to the fitting except for the normality plot, for which it corresponds to the theoretical normal distribution expected, and dashed green line encloses the region of the normal distribution to which corresponds a standard deviation of 0.5.

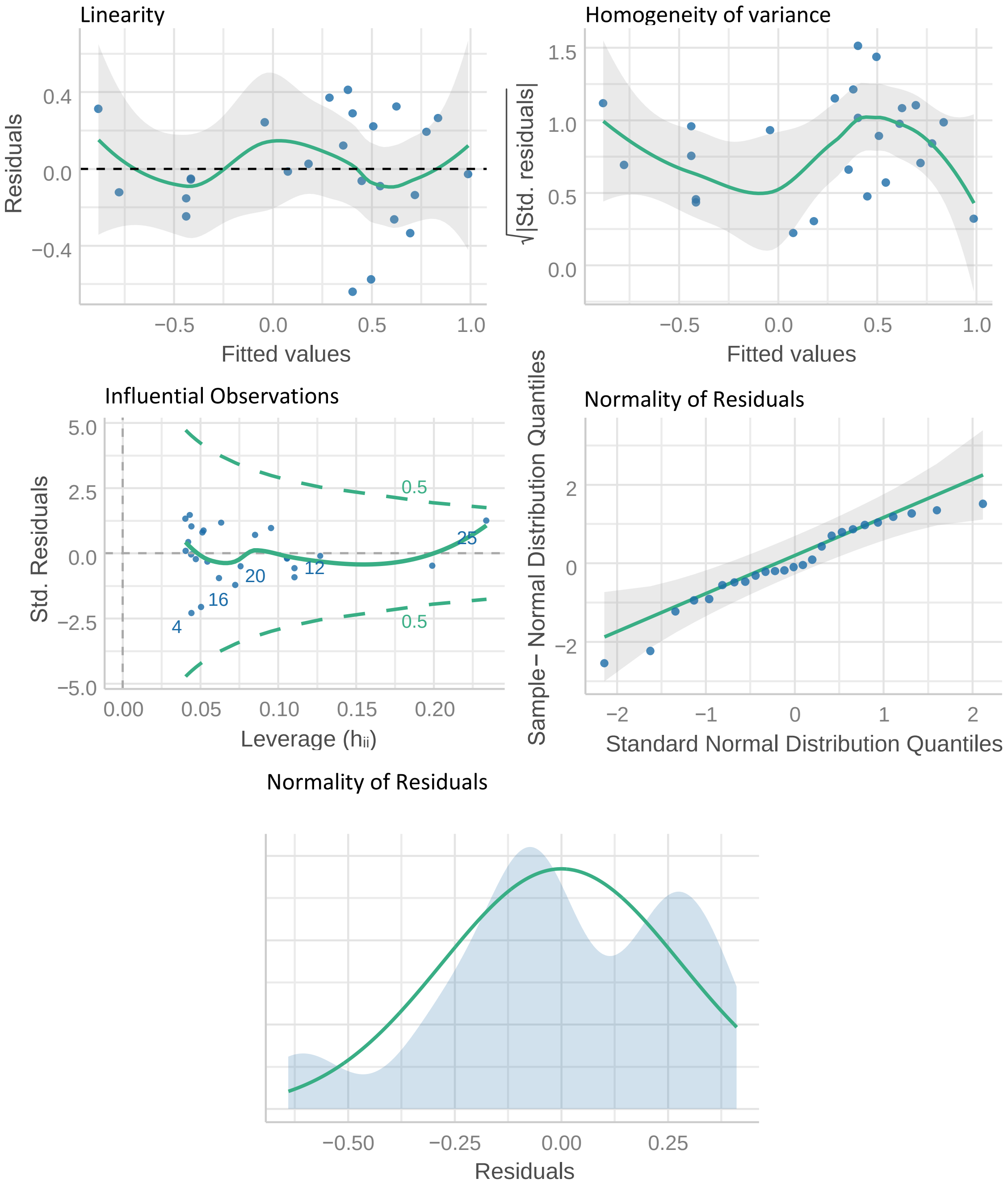


Figure A4. Statistical validation of the linear regression model for the thermodynamic stability of KCl. Blue points correspond to the experimental residuals observed, grey-shaded regions correspond to the confident interval, blue-shaded region corresponds to the experimental distribution of the residuals, solid green lines correspond to the fitting except for the normality plot, for which it corresponds to the theoretical normal distribution expected, and dashed green line encloses the region of the normal distribution to which corresponds a standard deviation of 0.5.

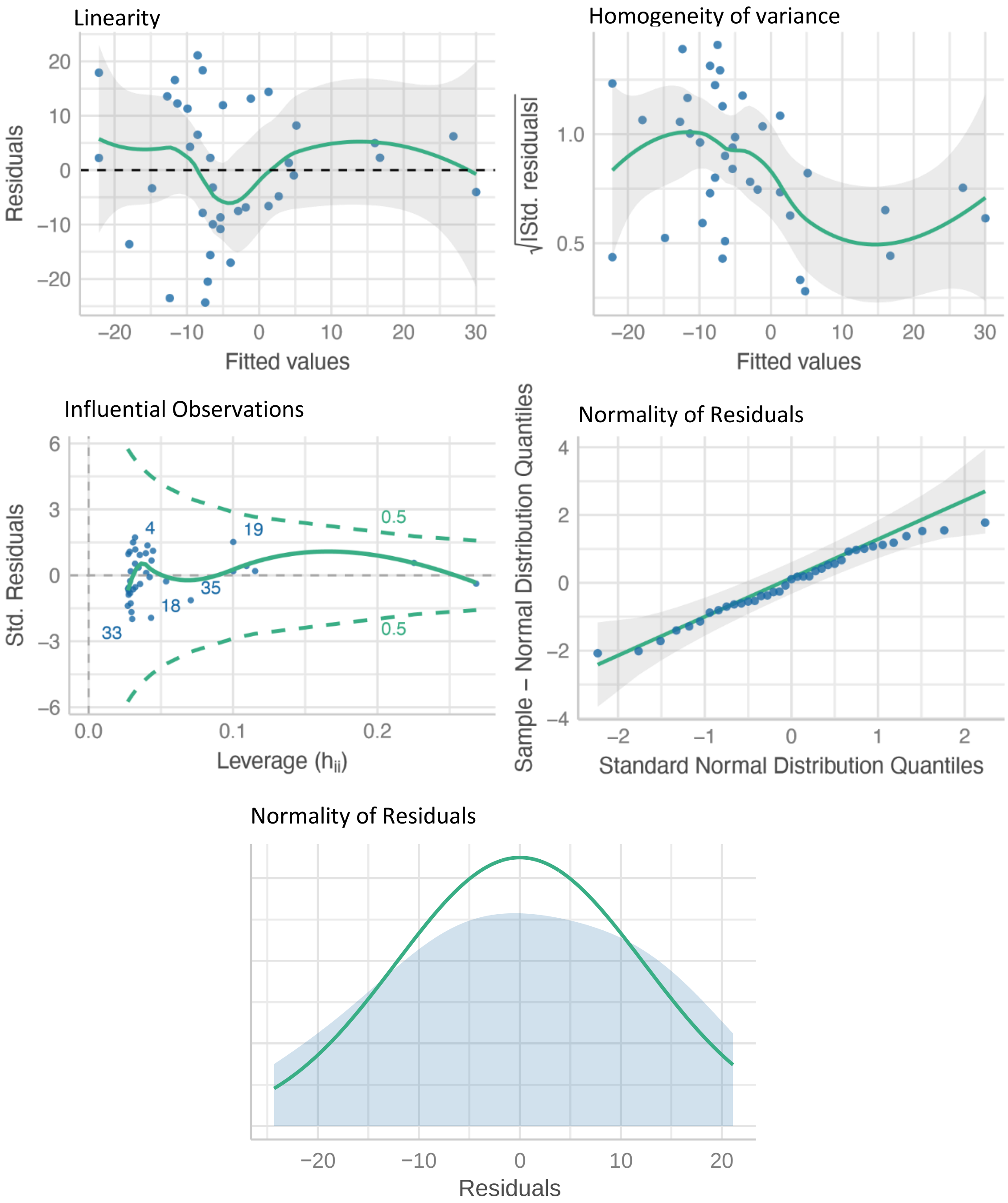


Figure A5. Statistical validation of the linear regression obtained for the thermal stability of taurine. Blue points correspond to the experimental residuals observed, grey-shaded regions correspond to the confident interval, blue-shaded region corresponds to the experimental distribution of the residuals, solid green lines correspond to the fitting except for the normality plot, for which it corresponds to the theoretical normal distribution expected, and dashed green line encloses the region of the normal distribution to which corresponds a standard deviation of 0.5.

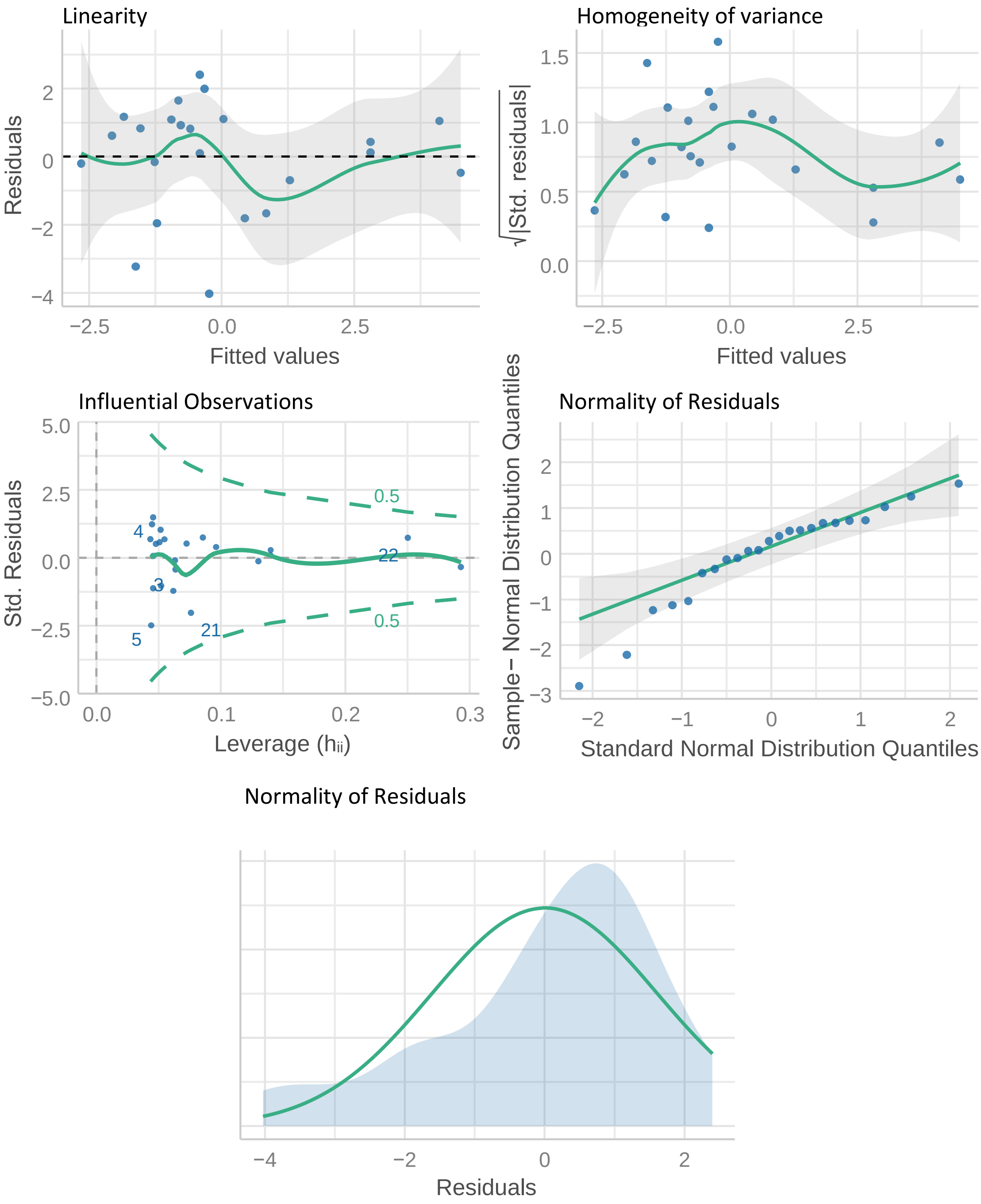


Figure A6. Statistical validation of the linear regression model for the thermodynamic stability of taurine. Blue points correspond to the experimental residuals observed, grey-shaded regions correspond to the confident interval, blue-shaded region corresponds to the experimental distribution of the residuals, solid green lines correspond to the fitting except for the normality plot, for which it corresponds to the theoretical normal distribution expected, and dashed green line encloses the region of the normal distribution to which corresponds a standard deviation of 0.5.

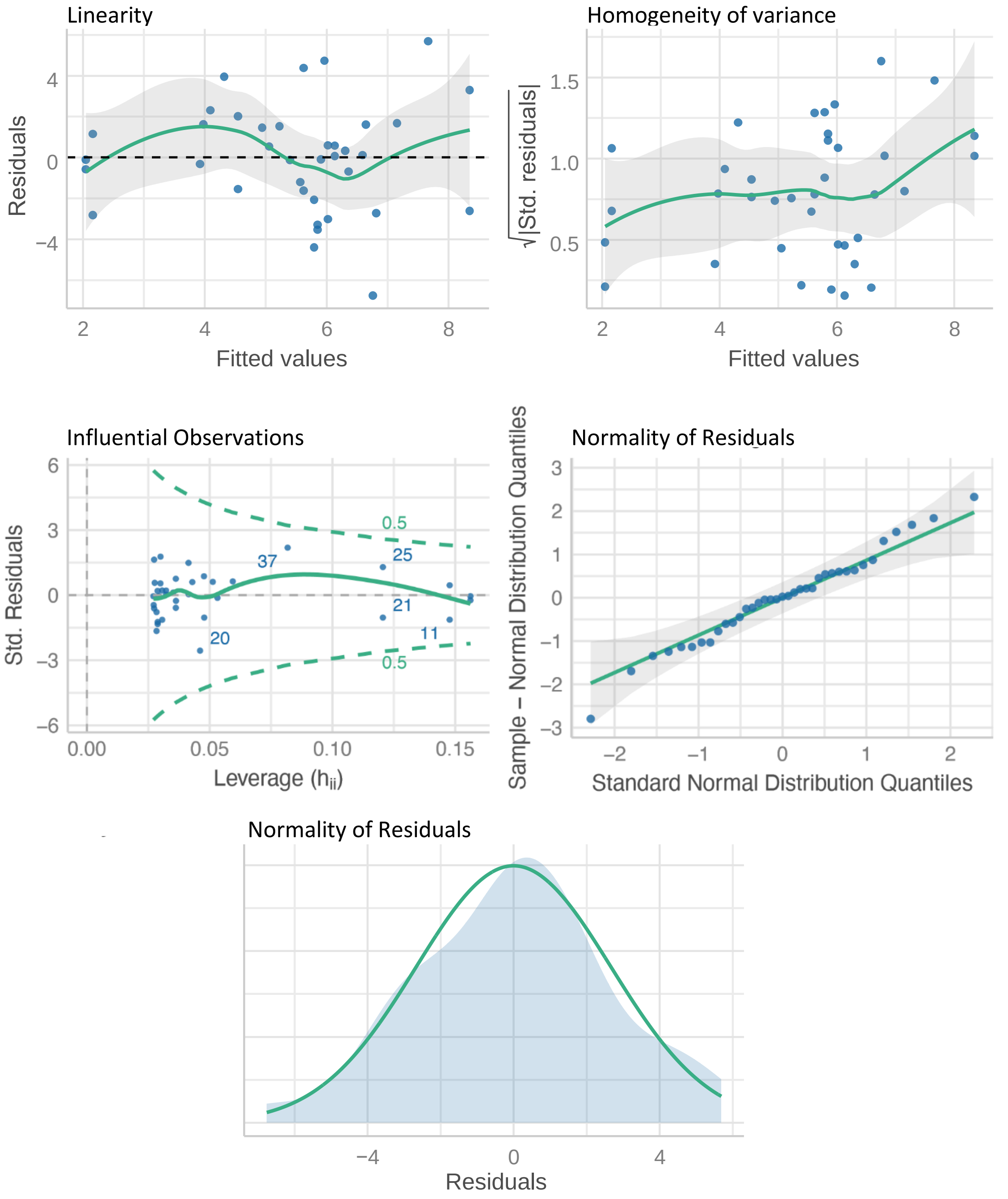
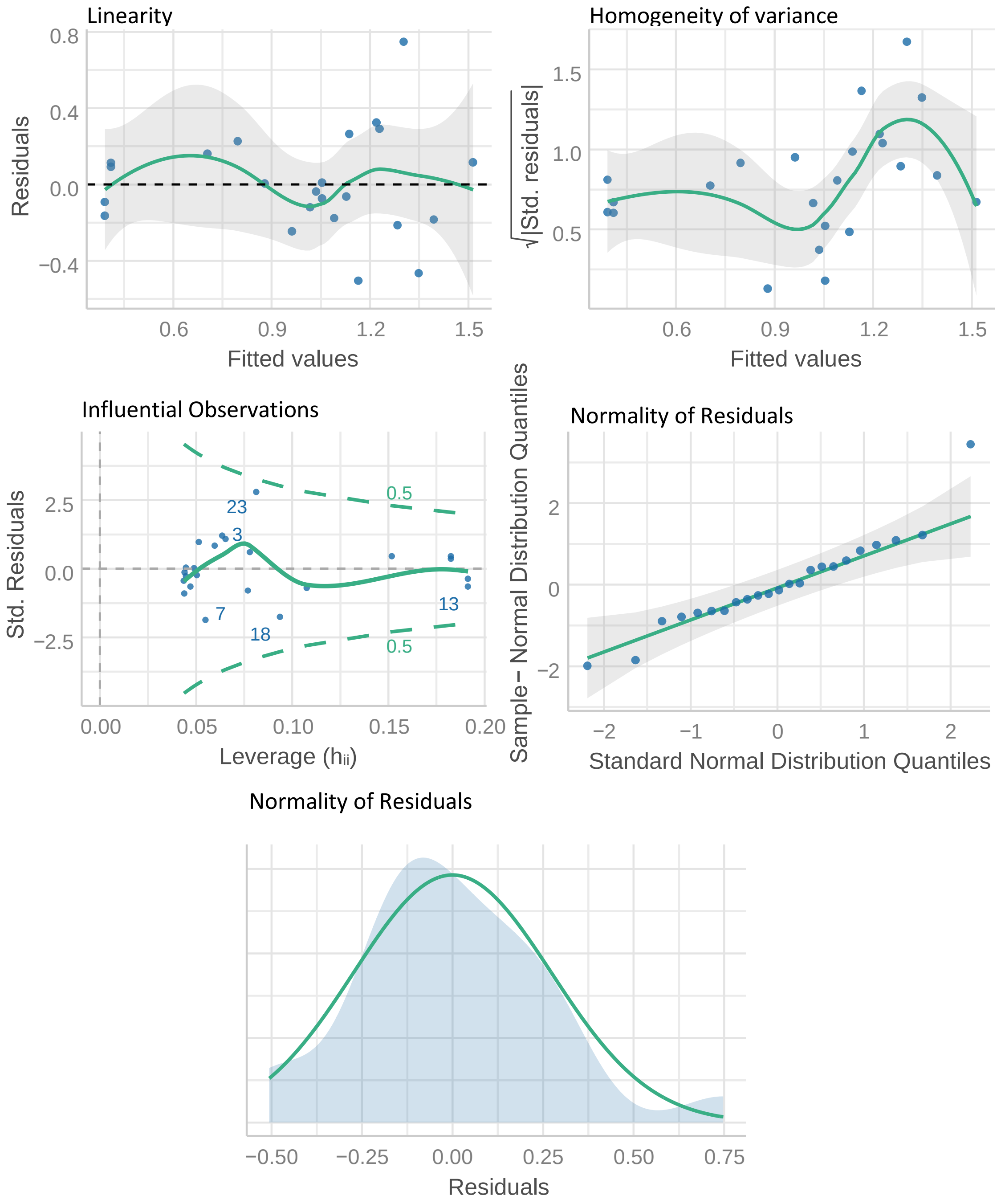


Figure A7. Statistical validation of the linear regression model for the thermal stability of TMAO. Blue points correspond to the experimental residuals observed, grey-shaded regions correspond to the confident interval, blue-shaded region corresponds to the experimental distribution of the residuals, solid green lines correspond to the fitting except for the normality plot, for which it corresponds to the theoretical normal distribution expected, and dashed green line encloses the region of the normal distribution to which corresponds a standard deviation of 0.5.



Linearity

Homogeneity of variance

Influential Observations

Normality of Residuals

Normality of Residuals

Figure A8. Statistical validation of the linear regression obtained for the thermodynamic stability of TMAO. Blue points correspond to the experimental residuals observed, grey-shaded regions correspond to the confident interval, blue-shaded region corresponds to the experimental distribution of the residuals, solid green lines correspond to the fitting except for the normality plot, for which it corresponds to the theoretical normal distribution expected, and dashed green line encloses the region of the normal distribution to which corresponds a standard deviation of 0.5.

List of Figures

List of Tables

List of Equations

Abbreviations

ASA	Accessible Surface Area
BMRB	Biological Magnetic Resonance Bank
CD	Circular Dichroism
DMEM	Dulbecco's Modified Eagle Medium
DNA	Desoxyribonucleic Acid
DSS	Sodium Trimethylsilylpropanesulfonate
ECMIM	E. coli Mimetic
FF	Fraction Folded
FID	Free Induction Decay
FT	Fourier Transformation/Transformed
GaMD	Gaussian Molecular Dynamics
GST	Glutathione S-transferase
GuHCl	Guanidinium Hydrochloride
HEK	Human Embryonic Kidney
HSQC	Heteronuclear Single Quantum Coherence
IPTG	Isopropyl β-D-thiogalactoside
LB	Lysogeny Broth
NMR	Nuclear Magnetic Resonance
O.D.	Optical Density
PBS	Phosphate Buffer Saline
PDB	Protein Data Bank
PIC	Protease Inhibitor Cocktail
RT	Room Temperature
SDS-PAGE	Sodium Dodecyl Sulfate – Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SSP	Secondary Structure Propensity
TROSY	Transverse Relaxation-Optimized Spectroscopy
TMAO	Trimethylamine N-oxide