# Molecular bases of proteome adaptation to high pressure in extremophilic Archaea

Antonino Caliò

**THESE de DOCTORAT DE**
**L'UNIVERSITE CLAUDE BERNARD LYON 1**

**Ecole Doctorale N° 341**
**Évolution Écosystème Microbiologie Modélisation**

**Discipline** : **Biophysique**

Soutenue publiquement le 08/12/2022, par :
**Antonino Caliò**

---

# Molecular bases of proteome adaptation to High Pressure in extremophilic Archaea

---

Devant le jury composé de :

| | |
|---|---|
| BROCHIER-ARMANET, Céline, Professeure, Université de Lyon | Présidente |
| BORDALLO, Heloisa Nunes, Professeure, University of Copenhagen | Rapporteure |
| GIRARD, Eric, Chargé de Recherche, CEA | Rapporteur |
| KALBITZER, Hans Robert, Professeur, Universität Regensburg | Rapporteur |
| COLLOC`H, Nathalie, Chargée de Recherche, Université Caen-Normandie | Examinatrice |
| OGER, Philippe, Directeur de Recherche, INSA Lyon, CNRS | Directeur de thèse |
| PETERS, Judith, Professeure, Université Grenoble-Alpes | Co-directrice de thèse |
| ROUMESTAND, Christian, Professeur, Université de Montpellier | Invité |

# CONTENTS

# LIST OF FIGURES

# ACRONYMS

| | | |
|---|---|---|
| AI | Artificial Intelligence | 128 |
| CS | Conformational Substates | 19 |
| EINS | Elastic Incoherent Neutron Scattering | 39 |
| EISF | Elastic Incoherent Structure Factor | 39 |
| ENS | Elastic Neutron Scattering | 38 |
| FFT | Fast Fourier Transform | 54 |
| HHP | High Hydrostatic Pressure | 3 |
| HP-NMR | High Pressure Nuclear Magnetic Resonance | 22 |
| HSQC | Heteronuclear Single Quantum Coherence | 46 |
| INS | Inelastic Neutron Scattering | 38 |
| IPTG | Isopropyl-$\beta$-D-1-thiogalactopyranoside | 27 |
| L-J | Lennard-Jones | 54 |
| MD | Molecular Dynamics | 22 |
| MR | Molecular Replacement | 52 |
| MSD | Mean Square Displacement | 41 |
| NMR | Nuclear Magnetic Resonance | 11 |
| NOE | Nuclear Overhauser Effect | 46 |

| | | |
|---|---|---|
| PME | Particle Mesh Ewald | 54 |
| ppm | parts per million | 45 |
| QENS | Quasielastic Neutron Scattering | 38 |
| RBS | Ribosome Binding Site | 25 |
| RF | Radio-frequencies | 45 |
| SEC | Size-Exclusion Chromatography | 28 |
| VdW | Van der Waals | 54 |

# OUTLINE

This thesis aims to answer the long-standing question of pressure adaptation, focusing in particular on microorganisms and on their proteome.

It is organized in four chapters and contains two research articles that embody all the main results of this work. The first one has been published while the second one is in the submission stage.

**Chapter 1** consists of an introduction, explaining the concept of adaptation in a broad sense, and then focusing on high pressure adaptation. It then describes the background this work is founded on as well as the main objectives of the thesis.

**Chapter 2** is a general survey on the experimental methods employed in this work. It presents the theoretical basis of the different techniques and explains the reasons behind their choice. It also contains references to textbooks and review articles where these techniques are described in more detail, for the interested reader.

**Chapter 3** contains the results, in the form of two research articles. The first one focuses more on the dynamical properties of the studied proteins, discussing their implication on pressure adaptation. The second one presents the NMR results as well as another neutron scattering experiment, putting all the findings into context with the amino acid substitutions found between the two proteins, and proposing the basis of a substitution pattern explaining high pressure adaptation.

**Chapter 4** draws the conclusions of this study, summarizing all the main results and discussing how they advance the present knowledge on the

subject. It also contains the future perspectives in this field and how these results can be seen as the basis for the development of a comprehensive model to describe high pressure adaptation.

# INTRODUCTION

## 1.1 Adaptation of life forms

Since the dawn of time, when the first life forms appeared on Earth, they have had to adapt to their environments. Adaptation has always been a constant in the evolution of life : organisms have had to face changing environmental conditions, especially when our planet was also frequently modified due to its considerably greater geological activity than today. Throughout history, living organisms have then found a way to thrive by adjusting to the available nutrients in their habitat or even by collaborating between them in symbiosis, and those that could not do it have disappeared. However, external physical conditions such as temperature, pressure, pH or salinity have also had a major impact on their evolution.

Nowadays, we know that life is ubiquitously present on our planet, and that organisms populate even the most extreme habitats. Figure 1.1 shows some examples of the most extreme environments that can be found on Earth. Deep-sea hydrothermal vents are usually located on oceanic ridges at considerable depths, and they emit hot gases coming from beneath the crust, therefore they are characterized by both high temperature and pressure. The deep sea in general is characterized by high pressure, but its temperature is markedly lower[1] (3°C). High-temperature environments exist also on the surface, at atmospheric pressure, such as in hot springs or volcanic formations. A number of other extreme conditions can be encountered in salty lakes, soda lakes or acidic rivers, where salinity and pH (high or low) reach extreme values.

*Figure 1.1 – Examples of extreme environments : a deep-sea hydrothermal vent (top left), the Yellowstone Grand Prismatic Spring, a hot spring (top right), the Lake Hillier, a hypersaline lake (bottom left), and the Rio Tinto, a highly acidic river (bottom right).*

These are only some examples of extreme environments that would be extraordinarily challenging to survive for the life forms that we are accustomed to study, and some of those environments even present combined stress factors. Nonetheless, a myriad of different organisms have been found in such environments, as a result of the evolutionary pathways that enabled them to adapt.

Among all the living creatures that populate these environments, microorganisms, prokaryotes in particular, are the most ubiquitous, owing to their impressive adaptive capabilities[1–4]. The domain of Archaea presents a very high number of species that are adapted to diverse environments[4,5], and indeed they are considered the *champions* of polyextremophiles.

When considering microorganisms, the main focus in adaptation is on proteins. They are the most abundant and less stable macromolecules present in the cell, they constitute its machinery and are needed to perform all the functions that an organism needs in order to survive[6]. While other cellular components, such as the membrane lipids, are also impacted by external conditions and are subject to adaptation[7–9], their synthesis still relies on the preservation of enzymatic activity in the proteins that build them.

With the astonishing progress that has been recently made in genome sequencing, genomic studies have become pivotal to uncover the mechanisms of adaptation has been genomic[10,11], together with the structural or biochemical approaches that had been employed until then. Proteins are the direct translation of all the information contained in the genetic code of any organism, therefore it is instructive to look for regular patterns of substitutions in the genome to identify adaptation strategies. While these studies have given important results for high-temperature or high-salinity adaptation[12,13], they were not as successful in the quest for the understanding of pressure adaptation. The specific mechanisms characterizing adaptation to High Hydrostatic Pressure (HHP) are the topic of this work, as pressure and temperature are equally important thermodynamic variables impacting the folding state of a protein[14].

Unravelling the adaptation to HHP specifically in thermophilic organisms is of interest mainly for the investigations about the origin of life, which has been hypothesized to have taken place near hydrothermal vents[15,16]. The reasons for this are detailed in the next section, while the following sections present a general discussion on the concept of adaptation from the point of view of protein folding, and some adaptation mechanisms are presented, in order to understand the typical approach that is taken in these cases, and to later explain why this has not given the same results concerning pressure adaptation.

### 1.1.1 Origin of life

Uncovering the origin of life on our planet has always been a key topic in science, as a result of the endless human curiosity. Some hypotheses have been put forth as to how the first proto-organisms started to populate the young Earth[17–20], and one of the most accepted is that of hydrothermal vents[15,16].

The radiation coming from the young sun and the completely different atmosphere of that time would have probably impeded any sign of life on the surface, therefore, hydrothermal vents are a promising candidate in this sense. Given their depth, the water above could have provided efficient shielding from the radiation, and the high temperature of those environments could have been an endless supply of thermal energy to perform the first

chemical reactions without the need of enzymes to cross energy barriers.

In the framework of this hypothesis, the first proto-cells would have then resurfaced at one point in history, and gave way to the evolution of life as we know it today. Therefore, the study of modern pressure-adapted organisms could be of interest in this field, as the understanding of their adaptation pathways could shed light on the "de-adaptation" of those early organisms when they came from the sea and started colonizing the Earth.

### 1.1.2 Protein folding

Proteins are the most abundant biomolecules present in cells, and they constitute their machinery. They are essentially bio-polymers, constituted by twenty-one different building block, the *amino acids*. Their common feature is the presence of an amino group and a carboxylic group, bound to a central carbon atom, named *alpha carbon* ($C_\alpha$). This group of atoms is the *backbone* of amino acids. What differentiates the twenty-one amino acids is the *side-chain*, which is also bound to the $C_\alpha$, and confers different properties to them. Amino acids are normally classified based on the electrostatic properties of their side-chain, which can be hydrophobic (Alanine, Valine, Leucine, Isoleucine, Methionine, Phenylalanine, Tyrosine and Tryptophan), polar but uncharged (Serine, Threonine, Asparagine and Glutamine) or polar and charged, be it acidic (Aspartate and Glutamate) or basic (Lysine and Arginine). There are five special cases, Cysteine, which contains a thiol group, and is able to form disulfide bonds, Selenocysteine, which is a very rare analogous of Cysteine but with a selenium atom instead of sulphur, Glycine, in which the side-chain is composed by a single hydrogen atom, Proline, whose side-chain forms a cycle with the nitrogen atom in the amino group, and Histidine, presenting an imidazole group that can change its protonation state, rendering it either basic or polar uncharged depending on the pH. Figure 1.2 presents their molecular structures.

The amino acids are joined together by the so-called *peptide bond*, which is substantially an *amide* bond, formed by the condensation of the carboxylic group of an amino acid with the amino group of the following one. The sequence of residues in a protein is called the *primary* sequence, and this is what is coded in the genome of an organism. However, proteins are not linear chains, they have distinctive 3D structures arising from the *folding*

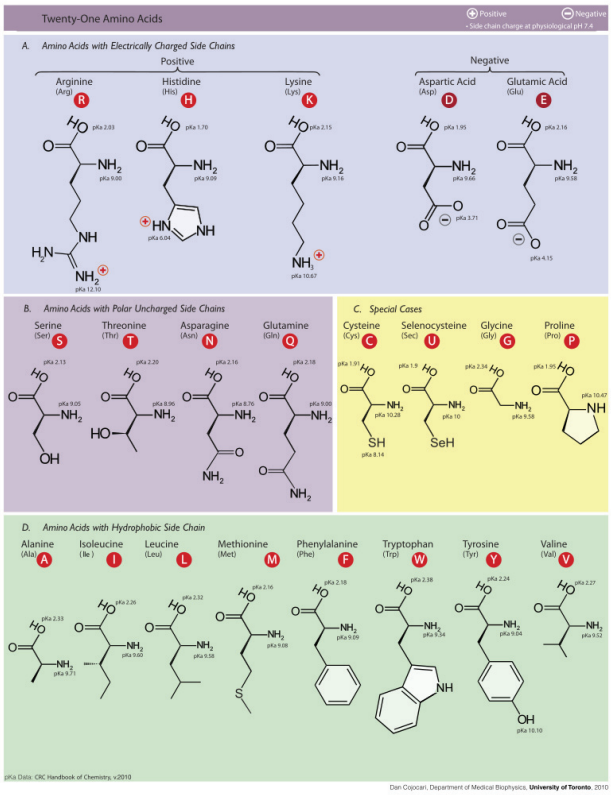


*Figure 1.2 – Molecular structure of the twenty-one proteinogenic amino acids.*

of the chain, and their structures are strictly correlated with their function. Furthermore, it has been demonstrated that the structure of a protein depends entirely on its primary sequence[21]. Therefore, since protein sequences are the direct translation of the genetic code of a given organism, genomic studies are a powerful tool to study adaptation pathways.

The problem of the folding of globular proteins has been puzzling biophysicists for more than 50 years[22] and has been extensively reviewed[23,24]. The main role is played by hydrophobic interactions and electrostatic forces, driven by the presence of water as a solvent. It is known that liquid water forms an extended network of hydrogen bonds, thus when a non-polar molecule is dissolved in it, water has to reorient in order to saturate its hydrogen bonding capability. Therefore, the mere presence of the apolar molecule limits the number of conformations water can assume, decreasing its entropy. Thus, when multiple hydrophobic molecules are dissolved in water, they tend to coalesce as if they were attracted to each other by a pseudo-force. In reality, this is the result of the minimization of the hydrophobic surface in contact with water, so that its conformational entropy can be maximized (fig.

1.3). This drives the *collapse* of the hydrophobic residues of a protein chain in what is usually referred to as the *hydrophobic core*. It happens after the first secondary structure elements are formed by means of hydrogen bonding between the protein backbone atoms.

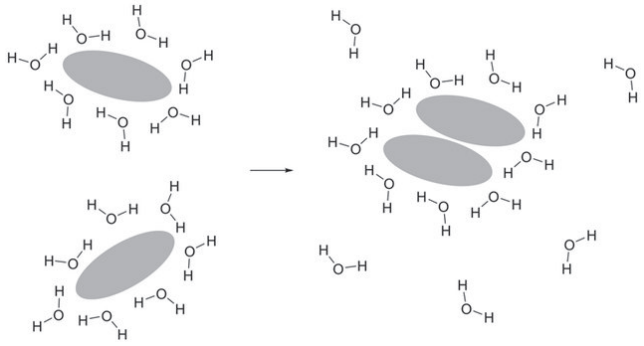

*Figure 1.3 – Depiction of the hydrophobic effect.*

Water also comes into play in the successive phases of protein folding, when electrostatic interactions, mostly charge-dipole (charged residue-water) and dipole-dipole (polar uncharged residue-water), dominate. It is now that the *native* conformation of a protein takes shape, with the stabilization of the 3D structure driven by the favourable electrostatic interactions between water and the polar residues that are on the surface of the protein.

Naturally, all of this is true in the case of what humans think of as *physiological conditions*, that is, 300 K, 1 atm, pH around 7, and moderate levels of salinity (i.e. less than 5%). When those conditions are not met, the driving forces for protein folding are impacted, giving rise to the necessity for *adaptation*.

### 1.1.3 Adaptation of proteins

Understanding the driving forces of protein folding is fundamental to tackle the topic of protein adaptation. Electrostatic and hydrophobic interactions are basic physical phenomena that are greatly impacted by the external conditions. Thermodynamic effects linked to temperature and pressure, together with the presence of electric fields generated by ions in solution (pH and salinity) can greatly perturb the stability of proteins, eventually leading to loss of function or even denaturation.

From all these considerations, it follows that proteins must be adapted to some optimal conditions in order to perform their activity, so that the orga-

nism they belong to can thrive. However, if we ignore protective mechanisms that may be enacted by cells as a whole, the only way a protein could become intrinsically more stable with respect to a given condition is by modifying its sequence. Yet, large modifications of a protein sequence would likely impact its structure, and therefore its function. For this reason, nature has selected organisms that have successfully attained maximum stability in their proteome, while avoiding excessive modifications that would negatively impact their function. This is the core problem of studying adaptation, as an organism could employ the least possible amount of mutations to survive, and the signature of this in its genome might not be entirely evident.

Before considering specific adaptation strategies, let us examine the toolbox that organisms can employ in order to adapt, the amino acids. Among the twenty-one amino acids, about half of them are apolar, and this is not surprising given the pivotal role of the hydrophobic effect in protein folding. However, they all possess different characteristics, notably their molecular weight, that may be advantageous in some adaptations. For instance, if a protein is adapted to a condition that weakens the hydrophobic effect, it may have a larger fraction of apolar residues to counteract it, or, a protein that requires reduced interactions with water may employ them on the surface. Polar residues could also be exploited to tailor protein-water interactions in different ways. A protein that needs to interact less with water may substitute a charged residue with a polar one, or add an oppositely charged residue near the first one, to form a salt bridge and neutralize the charge. Salt bridges may also be used to stabilize the 3D structure of a protein when they involve residues that are far in the primary sequence, but close in space, and the same is true for Cysteines when they form disulfide bonds, with the added bonus of it being a covalent bond, and thus stronger. Glycines can be used to improve local flexibility of the protein chain, and Prolines to decrease it. Undoubtedly, this is not an exhaustive list of the strategies that could be applied in proteins to adapt, but it gives some examples of the mechanisms that could be employed by organisms to adapt their proteome to different extreme conditions.

### 1.1.4 Adaptation to temperature

For many years, researchers have tried to understand adaptation to temperature, mainly because of the important biotechnological implications of thermophilic proteins, properly exemplified by the discovery of the DNA Polymerase from *Thermus aquaticus*[25], which prompted the development of the *Polymerase Chain Reaction* (PCR) and revolutionized molecular biology. Eventually, it will also be important to understand adaptation to climate change in the future.

High-temperature adaptation is, among other kinds of adaptation, one of the most understood to this date. In the context of proteins, it is known that thermal energy is able to affect their stability and leads to denaturation. This happen as the hydrophobic effect, one of the main driving forces in protein folding, depends on temperature[26]. When performing an energetic balance, one has to take into account enthalpic and entropic contributions : at higher temperature, the entropic cost for water molecules to solvate hydrophobic residues is lower, as the thermal energy balances the loss of configurational entropy needed to separate a hydrophobic contact. This ultimately leads to a decrease in the free energy associated with the hydrophobic effect (fig. 1.4) and to the unfolding of a protein.

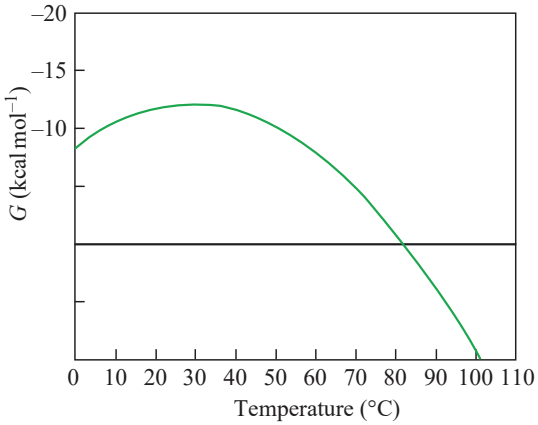


*Figure 1.4 – Typical stability diagram of a protein. At high temperature, the free energy difference between the folded and unfolded state becomes negative, causing the unfolding of the protein.*

To cope with this, thermophilic proteins have larger hydrophobic cores[3], so that they can maintain stability up to higher temperatures. Moreover, they present a higher rigidity at room temperature[27] in order to reach the proper amount of flexibility at their optimum temperature. Other features of temperature-adapted proteins are increased disulfide bonds and

salt bridges[28]. All of this contributes both to the structural and thermodynamic stability, keeping proteins in their most functional state even at very high temperatures. In fact, Tsukamoto et al.[29] found an important increase in hydrophobic residues and in aromatic-aromatic interactions in the Rhodopsin from *Thermus thermophilus* by comparing its structure to bacteriorhodopsin and other homologues. Moreover, they showed that increased hydrogen bonding stabilizes the lumen of the proton pump at high temperature, while the active residues involved in proton transport or in the binding or the chromophore are conserved. To establish the role of disulfide bonding, Cacciapuoti et al.[30] performed thermal stability experiments by circular dichroism on the 5'-deoxy-5'-methylthioadenosine phosphorylase II by the hyperthermophilic archaeon *Sulfolobus solfataricus*. Homologues of this protein from mesophilic organisms do not present intra-subunit disulfide bonding, therefore they also prepared disulfide-lacking mutants and found that their unfolding temperature was greatly reduced, highlighting the considerable stabilizing effect of the disulfide bonds.

All these considerations clearly point towards a well defined adaptation pattern. Indeed, temperature adaptation bears some clear genomic markers. Those have recently been exploited to perform phylogenetic studies to ascertain whether the last common ancestor of Bacteria and Archaea was a hyperthermophilic organism[31] (fig. 1.5). By aligning both protein and ribosomal RNA sequences from a large variety of species, it was possible to construct a phylogenetic tree based on inferred optimal growth temperature.

Adaptation to low temperature, however, is a different matter. In order to function when low thermal energy is available, *psychrophilic* proteins present considerably higher flexibility[3]. To achieve this, they feature less hydrophobic residues in the core and more of them on the surface, to decrease hydration and increase their conformational freedom. Moreover, they tend to employ smaller residues to decrease the surface area available for interaction with the surrounding water. In fact, Gianese and colleagues[32], found a correlation between growth temperature and the fraction of hydrophobic residues present on the surface. They also describe how decreased hydrogen bonding increases the intra-subunit flexibility in some proteins, and that in general charged residues are often substituted by apolar ones, or deleted entirely.

However, multiple adaptation strategies have been found in psychrophiles

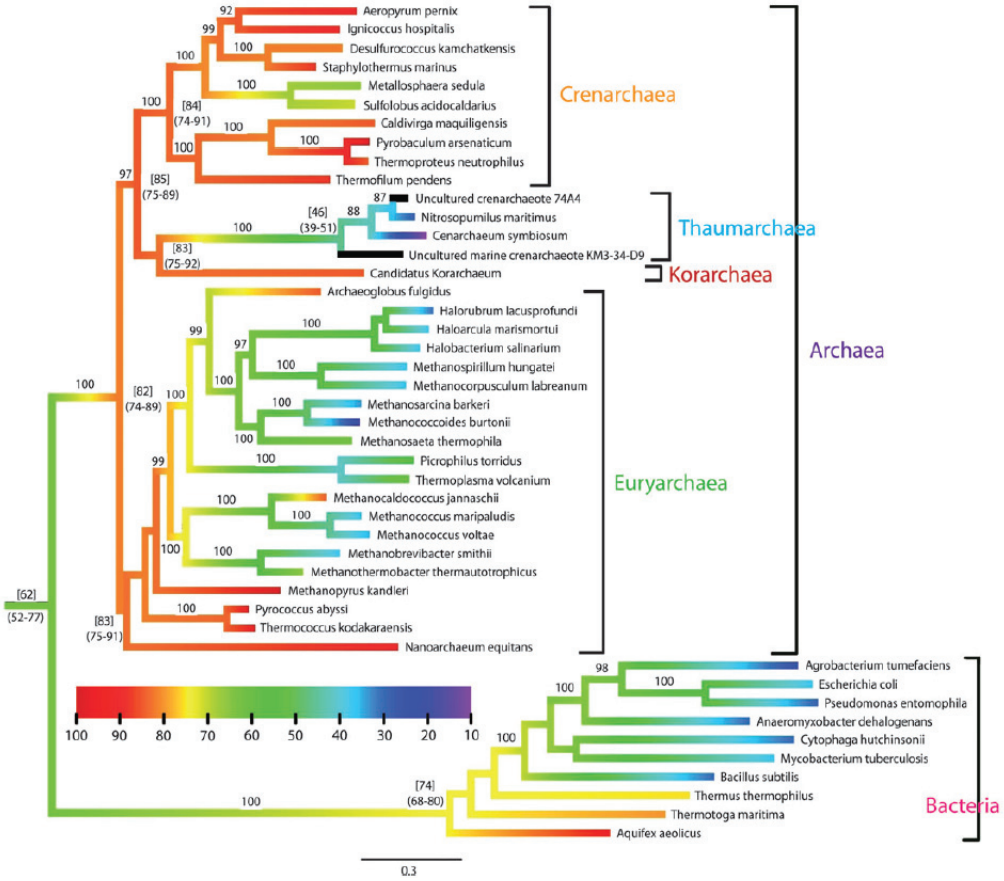


*FIGURE 1.5 – Evolution of optimal growth temperature from a hyperthermophilic ancestral state over the protein archaeal tree. The colour scale is in degree Celsius. Mean estimates of temperature at key nodes are given between square brackets, with the 95% confidence interval between round brackets. Taken from*[31].

that go beyond the substitutions of amino acid in proteins[33]. They have been shown to express some peculiar genes that are characteristic to them and do not exist in mesophiles, and are involved in controlling and preventing ice formation in their cytoplasm. Compatible solutes are also produced in order to act as cryo-protectants for the cell's proteome (fig. 1.6).

### 1.1.5 Adaptation to salinity

A considerable number of organisms, regardless of their temperature optima, live in conditions where salt content is very high in their environment[34]. Therefore, their proteins have to compete for hydration with the ions that are present in solution. To tackle this problem, they employ mostly acidic residues on their surface in order to maintain their hydration[35]. Mo-

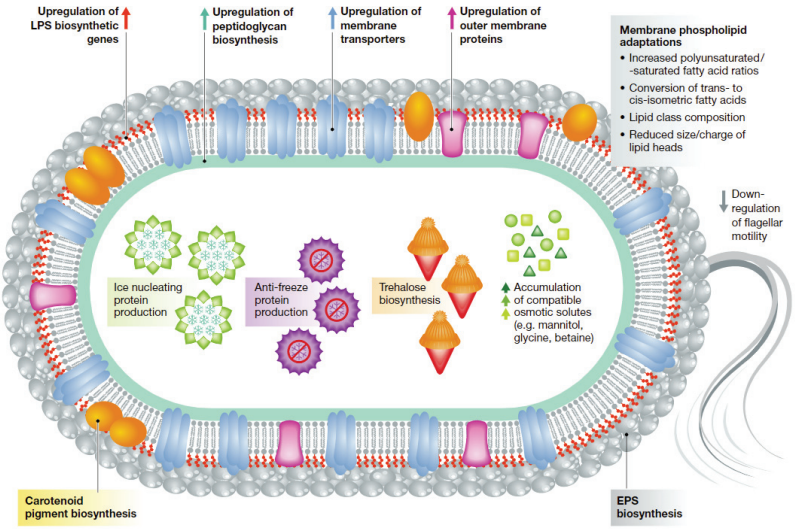


*Figure 1.6 – Principal adaptation mechanisms in prokaryotic psychrophiles. Taken from*[33].

reover, their highly negative charge favours electrostatic repulsion between proteins, to prevent aggregation and maintain optimal solubility even in considerably high ionic strength environments.

Another strategy employed by halophilic organisms is the production of co-solutes to contrast the extreme salt concentrations they face in their habitats. This is also valid for acidophiles and alkaliphiles, organisms that live not only at high salt concentration but also at the extremes of pH[3]. Therefore, halophilic proteins must adapt to the very high salt concentration present in the cytoplasm of these organisms. As a consequence, when salt concentration is decreased, these proteins tend to lose their activity or unfold and completely lose their structure[35].

This concept is indeed one of the keys to halophilic adaptation. Ortega et al.[36] studied a halophilic protein, the $\beta$-lactamase from *Oceanobacillus iheyensis* and compared it to the mesophilic IgG-binding domain from *Peptostreptococcus magnus*. With Nuclear Magnetic Resonance (NMR) spectroscopy, they confirmed that the stabilization of the folded state in the halophilic protein is due to the acidic residues accommodating salt cations on the protein surface, eliminating the competition between the salt and the protein for hydration. This effect concomitantly destabilizes the unfolded state, enhancing the protein's stability. Furthermore, they mutated the mesophilic protein according to the amino acid composition they found for halophiles (fig. 1.7), and they indeed succeeded in creating a halotolerant protein with only five mutations.

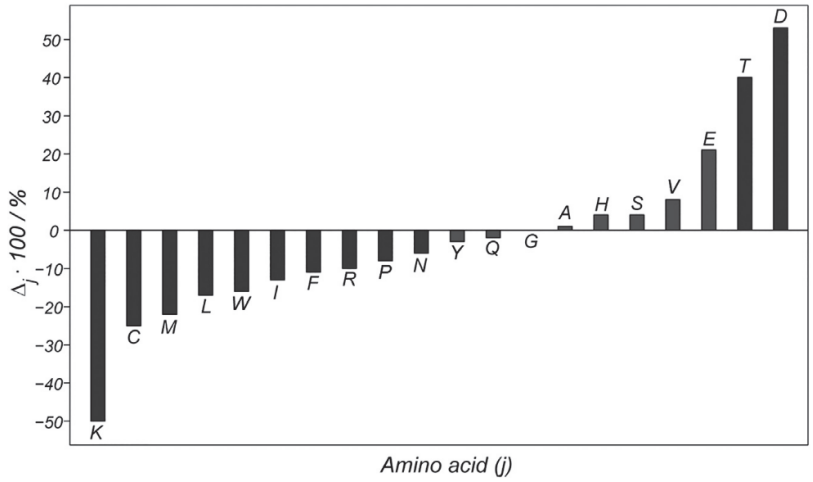


*FIGURE 1.7 – Amino acid occurrence in halophilic proteins, relative to their mesophilic orthologs. Taken from*[36].

The case of halophilic proteins is thus another example where a clear distinction in amino acid composition was evident from a genomic analysis and has been confirmed by experimental results on actual proteins[37,38]. We will see, in the following, that this is not the case for pressure adaptation.

## 1.2 Adaptation to pressure

HHP is one of the most ubiquitous conditions present on our planet. By simple calculations[1], it can be estimated that 75% of the oceans' volume has a depth of at least 1000 m, constituting the so-called *deep biosphere* (fig. 1.8). This thesis work is focused on the study of this particular adaptation mechanism.

Before taking into consideration the matter of adaptation, it is useful to draw a simple, but important, classification of organisms based on their capability, or lack thereof, to withstand pressure. *Piezosensitive* organisms are the common life forms we are used to, that is, species that optimally live at atmospheric pressure, and that are inhibited by higher pressure, leading to their death at relatively small values. *Piezotolerant* organisms are those that still live optimally at atmospheric pressure, but have the capability to survive at higher pressure, albeit at a lower growth rate. *Piezophiles* are organisms that grow optimally at HHP, but are still able to survive at atmospheric pressure. Finally, *strict* or *obligate piezophiles* are the species that need HHP to

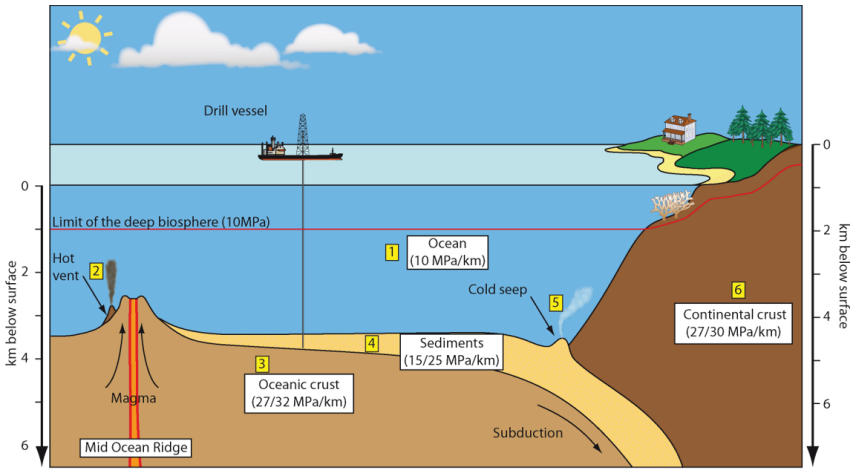


*FIGURE 1.8 – Representation of the deep biosphere. In the ocean, pressure rises by around 10 MPa per km of depth, while in the continental and oceanic crust or in sediments, the rate is different (27-30, 27-32, or 15-25 MPa/km respectively). Posing the limit for the deep biosphere at 10 MPa, the red line represents this limit in both these media. Some particular environments found in the deep sea are also represented. Taken from*[4].

thrive, being incapable of any growth at ambient pressure. Figure 1.9 reports a graph that represents this classification.

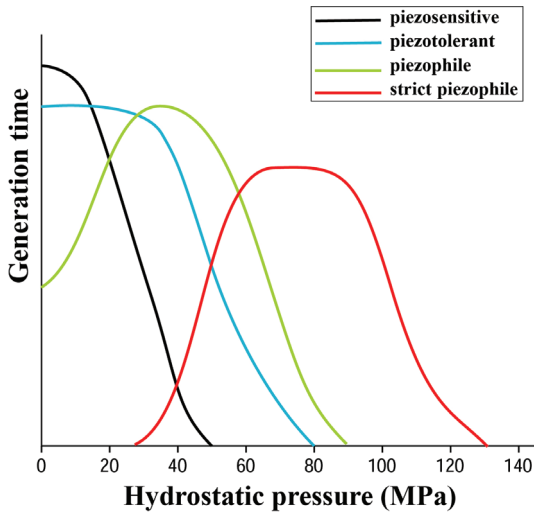


*FIGURE 1.9 – Graphical representation of the classification of organisms based on their pressure tolerance.*

The first challenge that is encountered when studying adaptation to HHP is the isolation of this condition. Nearly all of the piezophilic species known to date are either thermophiles[39] or psychrophiles[40]. This is because the most common habitat affected by HHP is the deep sea, where the temperature is almost constant at 3°C[1]. A feature of the deep sea, however, is the presence of ridges with important volcanic activity where hydrothermal vents, the preferred habitat of thermophilic piezophiles, are located. Only a handful

of species that thrive under pressure and at intermediate temperatures have been discovered[41], and the lack of information on them renders the isolation of the contribution of pressure to the adaptation a difficult topic to tackle.

Concerning proteins, some studies have been performed in which comparisons have been attempted to describe the properties of piezophilic proteins. For instance, Kawano et al.[42] studied the RNA polymerase from *Shewanella violacea*, a psychrophilic and piezophilic bacterium, and compared it to that of *E. coli*. They found a very different pressure response regarding the activity of the two proteins (fig. 1.10), and ascribed that to the different proline and glycine content by comparing the two sequences, which would greatly impact protein flexibility. Moreover, they modelled the structure of the piezophilic protein and commented on the $\beta$-sheet loop of the $\sigma^{70}$ domain, which is responsible for stabilizing the entire enzyme's quaternary structure under HHP.

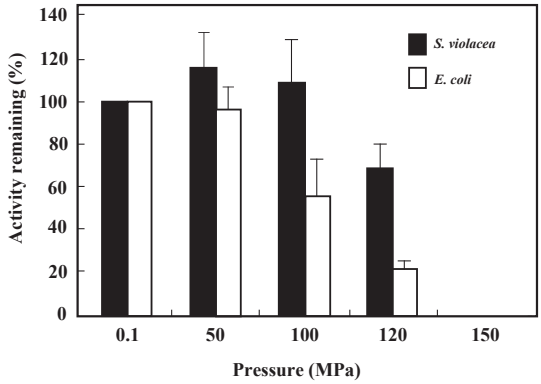


*Figure 1.10 – Effects of high hydrostatic pressure on transcriptional activity of RNA polymerase of* S. violacea *and* E. coli *using a non-specific DNA template. Taken from*[42].

Shrestha et al.[43] performed a similar study by comparing the Inorganic Pyrophosphatase from *Thermococcus thioreducens* with Hen Egg White Lysozyme. They employed neutron scattering in order to focus more on the internal dynamics of the proteins, and highlighted how the piezophilic enzyme is more flexible and is able to retain its peculiar dynamics even at HHP, while lysozyme becomes rigid and likely inactive.

However, the downfall of such studies is that they could not attempt to ascribe the properties they found to differences in the primary sequences of the proteins. This is because they are analysing proteins that are very far apart in the phylogenetic tree, are not homologues and also differ in multiple adaptation traits. Therefore, the sequence similarity is excessively low, if not

completely absent, rendering any sequence comparison questionable.

One way to circumvent this is to study very closely-related organisms that have similar optimum growth temperatures. In a recent study, Peoples et al[10] have used this approach on the psychrophilic genus *Colwellia*, finding some interesting insights on their adaptation. They found a correlation between the pressure optimum and the basicity of the proteome (fig. 1.11a and b). Moreover, they describe how the piezophilic proteomes are, on average, more hydrophobic than the piezosensitive counterparts (fig. 1.11c and d). They point out how this could be of use in avoiding water intrusion into proteins as a result of HHP. They also report an increase in genomic islands and mobile elements in the genomes of the piezophilic *Colwellia*, which could also be part of their adaptation strategy for HHP.

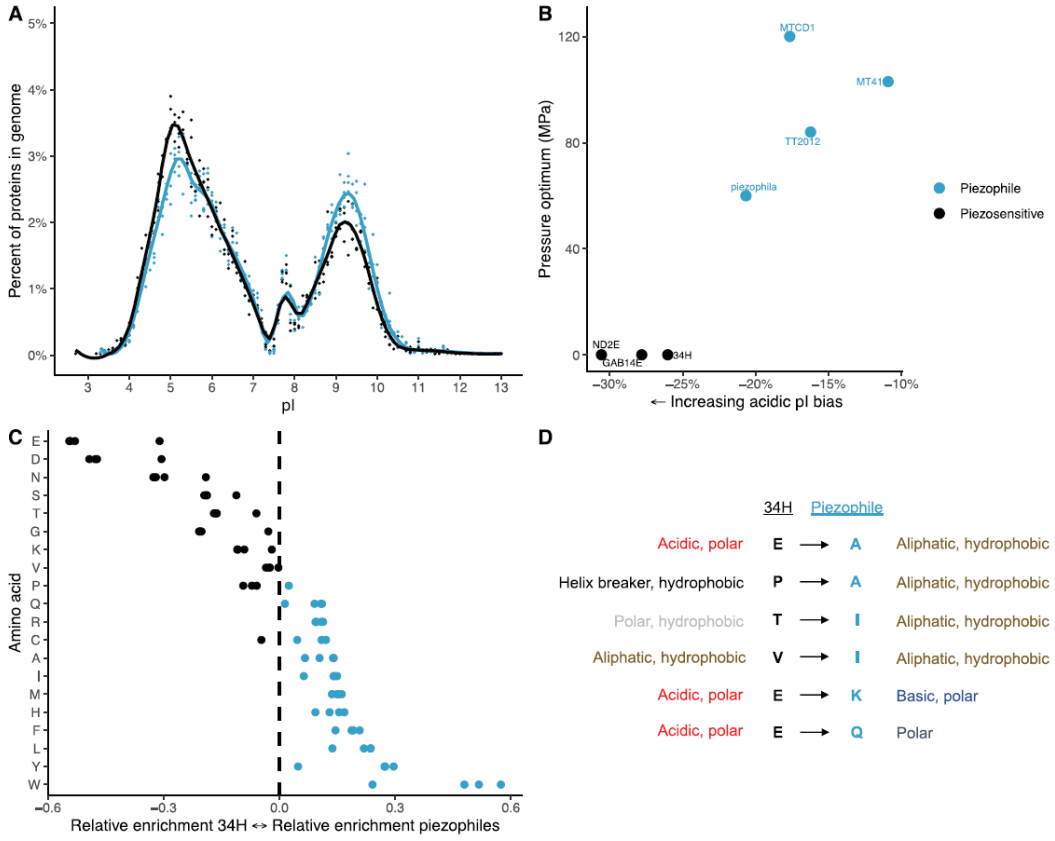


*Figure 1.11 – a, Isoelectric point distribution of proteins within piezophilic (blue points) or piezosensitive (black) strains. b, Isoelectric point protein bias within each strain as a function of their growth pressure. c, Asymmetry index values indicating preference of amino acids in the piezophiles or* C. psychrerythraea *34H within orthologous proteins present in all strains. d, Specific amino acid substitutions from* C. psychrerythraea *34H to the piezophiles within orthologous proteins. Taken from*[10].

A similar approach has been also attempted on two closely-related Archaea,

namely *Pyrococcus furiosus* an *Pyrococcus abyssi*. Di Giulio[44] performed a genomic comparison between these two species and presented the results of a statistical analysis on the amino acid substitutions between the two genomes. Based on these results, a *Pressure Asymmetric Index* (PAI) was calculated and correlated to the polarity and the molecular weight of the residues (fig. 1.12.). The author concludes that smaller residues appear to have a more piezophilic character by virtue of the fact that they would increase protein flexibility and would enhance the packing efficiency of their hydrophobic cores. Moreover, polar residues tend to have a higher PAI on average, and it is concluded that polar interactions with water on the surface of proteins must be important for piezophilic proteins. This study has the merit of comparing very closely-related species, and the comparison is made between orthologous proteins, however, the small sample size is likely responsible for the rather weak correlations that are observed.

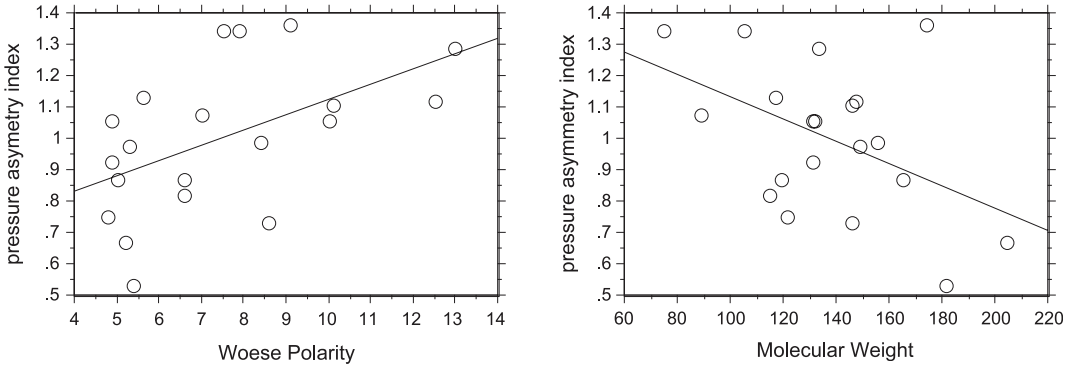


*Figure 1.12 – Relation between the values of the hydrostatic pressure asymmetric index and the polarity values of Woese[45] (left panel), relation between PAI and the molecular weight of amino acids (right panel). Taken from[44].*

In fact, a recent study involving the entire archaeal genus of *Thermococcales*, instead of simply two species, did not give the expected results (unpublished data). The rationale for this study was the same, that is, to take into account closely-related organisms that grow mostly with the same optima, while differing in their optimum pressure. Thermococcales have been intensely studied and serve as model organisms in the search of adaptation pathways regarding the cell membrane[7–9,46–48], yielding a wealth of information about those mechanisms. However, the adaptation strategy of their proteome remains elusive.

In order to gain more insight about this subject, experiments on whole cells have been performed[49–52], showing that the piezophilic species possess a proteome with different characteristics. These results form the basis of this

work, and they are described in detail in section 1.3. However, these studies do not enable the identification of the specific substitutions bringing about these effects.

### 1.2.1 Effects of pressure on life : the biological point of view

Conventional organisms, or *mesophiles*, have been thoroughly studied in their response to HHP, mainly with application to food safety[53,54]. The growth of these organisms is greatly impacted by pressure application, on many different levels.

Nucleic acids are actually stabilized by pressure, as it strengthens the hydrogen bonds between nitrogen bases, and brings about an increase in the melting temperature. From the point of view of pure stability this would be a positive feature, however, organisms need to separate the strands of DNA in order to replicate or repair it, and pressure hinders the double-strand to single-strand transition, hampering growth[55].

Concerning lipids, they form the boundary between cells and the external environment, therefore their stability is also of interest when studying the pressure response of organisms. The main effect pressure has on the bilayer is to decrease its fluidity by promoting increased packing of the hydrophobic tails of the phospholipids. This in turn hinders their motion and their ability to reorient, decreasing membrane permeability and affecting its interaction with proteins that are essential for transport and other functions[56]. This prevents the cells to take advantage of chemical gradients between the inside and the outside that are needed to generate chemical energy, impeding normal cell function. Moreover, the loss of membrane flexibility under pressure prevents cells from adapting their shape throughout their life cycle, and also greatly impacts cell motility, meaning that not only they cannot transport nutrients inside because of the decreased permeability, but they also lose the ability to retrieve said nutrients. However, cells can normally respond to this stress by adding insaturations (Bacteria) or cyclic elements (Archaea) in their lipid tails, to increase their flexibility in a process known as *homeoviscous adaptation*[8].

Proteins are the most pressure-sensitive components of cells. The pressure values found in the deep biosphere are clearly incapable of impacting their

covalent structure (i.e. the primary sequence)[4], and also to unfold them, as even for mesophilic proteins several hundreds of MPa are necessary[2]. Therefore, the biggest impact pressure has on proteins, is on the shifting of the population equilibria between different conformations, impacting their activity. The most sensible are multimeric structures, as they are easily separated by pressure[57].

Following the aforementioned considerations, it is natural to wonder how piezophiles found a solution to these difficulties, especially in the case of strict piezophiles. Their adaptation is so deeply rooted that they completely lost the ability to survive when pressure is absent. However, to understand these mechanisms, the physical aspects about the effects of pressure on macromolecules must be understood first.

### 1.2.2 Effects of pressure on life : the physical point of view

The foundation of every physical study involving pressure is Le Châtelier's principle[58]. It states that in any equilibrium reaction, pressure always favours the state of least volume. The direct consequence of it is that most enzymatic reactions could actually be favoured by pressure, provided that the enzyme itself is stable enough to retain its functional conformation.

However, it has long been known that pressure has deleterious effects on proteins. Since 1914, when Bridgman discovered that egg white coagulated when exposed to HHP[59], pressure denaturation of proteins has been an established fact. Although the biophysics community had lost interest in it for a considerable amount of time, it became a trending topic again when Hawley[60] reported for the first time a thorough characterization of the pressure-temperature phase diagram of chymotrypsinogen.

Nowadays, great progress has been made in the understanding of the macroscopic thermodynamics governing protein unfolding under pressure[61–64], although the microscopic aspects that are involved in it are still elusive.

Different phenomena contribute to the microscopic stability of proteins under pressure, such as the hydrophobic effect[65], the presence of cavities[63,64], electrostriction[66] and crowding[67]. However, the main driving force to pressure unfolding is the fact that in most cases the unfolded state has a lower

volume than the folded one.

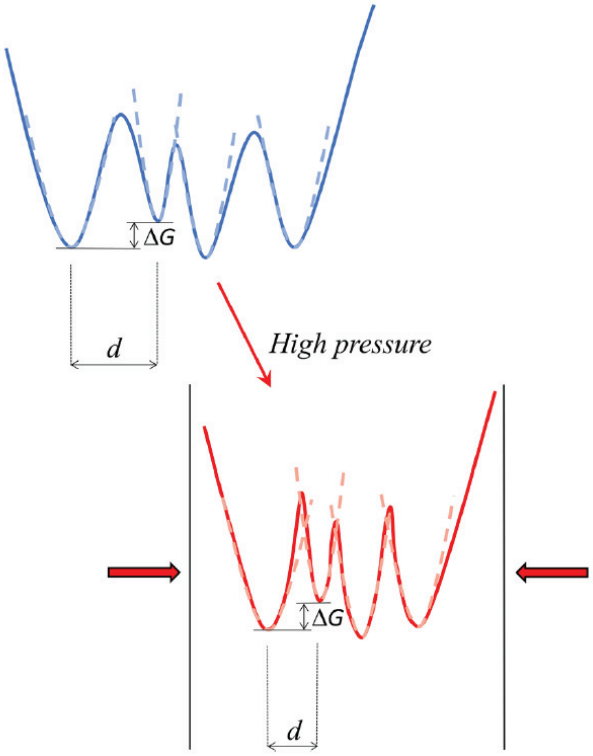


*Figure 1.13 – Schematic representation of the effects of pressure on the protein energy landscape. Taken from [68].*

The most accepted model for protein dynamics is the so-called *Free Energy Landscape* model, put forth by Frauenfelder and colleagues[69]. In the framework of this model, proteins are thought to exist in different Conformational Substates (CS), and at thermodynamic equilibrium a protein can fluctuate between different CS depending on the thermal energy it possesses. The effect of pressure can be viewed as a constraint on the volume difference between these substates, so that the protein can only explore a limited number of them, namely those that do not present a considerable difference in volume[68,70]. Viewing this in a simplified picture, one can imagine that the potential energy landscape of a protein is "shrunk", as the protein cannot perform the large-scale motions that would largely affect their volume (fig. 1.13).

This concept is important when it comes to adaptation, as protein unfolding is not the major concern. The highest pressure value on Earth at which life has been found is around 110 MPa, at the bottom of the Mariana trench[1], and it is not enough to denature even mesophilic proteins. This becomes especially

important when considering thermophilic proteins, as the adaptation to high temperature confers them an extended p-T stability[71].

Therefore, pressure adaptation has to be regarded as the search of a fine balance between stability and flexibility, allowing proteins to maintain their most active conformation even in harsh conditions. To achieve this, our hypothesis is that specific substitutions in key position of the proteome are needed, to precisely tailor the energy landscape of pressure-adapted proteins. This could explain why finding a general pattern on the genomes of these organisms is difficult, and why having structural and dynamical information is of the utmost importance to tackle this topic.

## 1.3 Background of our studies

In the previous sections, some of the difficulties associated with studying pressure adaptation have been analysed. However, the previously mentioned studies on whole-cells gave very important insight on this process in Thermococcales.

First, all of those experiments were performed on two organisms, namely *Thermococcus barophilus*[72] and *Thermococcus kodakarensis*[73], which are very closely related, and almost isogenic. They are both hyperthermophilic, optimally growing at 85°C, and present identical growth characteristics in the same conditions, except for pressure. On one hand, *T. barophilus* is a *piezophile* isolated at a depth of 3550 m at the Snakepit hydrothermal vent on the Atlantic ridge[72]. It is still able to grow at atmospheric pressure, however, its optimum is at 40 MPa. On the other hand, *T. kodakarensis* is a *piezosensitive* organisms, and was isolated in a hot spring on the Kodakara island in Japan[73], therefore its optimum is at atmospheric pressure, and its growth is inhibited already at 15 MPa.

These two organisms make perfect candidates for this kind of study, as they differ only in pressure adaptation, and they have been reasonably well studied as model Archaea. Moreover, their genomes have been also fully sequenced[74,75].

These studies have been possible thanks to the use of neutron scattering, a technique that has also been extensively used in this work and will be

thoroughly described in the next chapter. Briefly, neutrons very strongly interact with hydrogen atoms, and given that proteins make up almost 50% of the cell biomolecules, and they also contain high percentages of hydrogen, their dynamics can be efficiently probed even inside live cells. Moreover, by suspending the cells in $D_2O$, the dynamics of their internal water can be analysed.

Different experiments have been performed, focusing on many aspects : the dynamics of the proteome itself[49,50], crowding[51] and the impact of organic osmolytes[52].

It has been found that the two organisms' proteomes have a markedly different response to HHP. They both present a pressure-dependent jump in the proteome dynamics, however, while this jump is sudden and takes place at very low pressure in *T. kodakarensis*, it is smoother and comes to completion near the optimum pressure for *T. barophilus*. Moreover, remarkable differences have also been found in their hydration water dynamics : the population of hydration water is lower in *T. barophilus* and this seems to confer higher flexibility to its proteome[50].

The effect of crowding was also studied, by lysing the cells into the sample container to release crowding. This experiment revealed that when lysing the piezosensitive cells, the flexibility of their proteome increases, and the hydration water population decreases, rendering the proteome less pressure-sensitive, similar to the piezophilic one[51].

Lastly, the effect of organic osmolytes has also been studied. *T. barophilus* has been shown to produce mannosylglicerate when grown under pressure stress, that is, at atmospheric pressure. As it is adapted optimally to 40 MPa, atmospheric pressure is perceived as a stress by the organism, therefore, it compensates by upregulating stress-response genes and by producing the osmolyte[76]. The results evidence how this molecule acts mainly on the properties of hydration water, rendering the proteome less pressure sensitive, so that it can maintain its characteristics even at low pressure[52].

All of these contributions to proteome dynamics are clearly interconnected and demonstrate that the two proteomes have different characteristics that stem from the genomic differences between the two organisms. However, studying whole cells does not give detailed informations on how substitutions in the genome give rise to these properties. We seek to answer this

question, and the next section presents the strategy we employed to do it.

## 1.4 Objectives

The main question that we sought to answer is about the existence of a specific amino acid substitution pattern that characterizes pressure adaptation. The failure of the genomic approach and the insufficient detail of the whole-cell studies lead us to design a novel bottom-up approach in search for the answer.

Biological systems are notoriously very complex, and answering certain questions often requires a simplification of the problem. Therefore, to gain more insight about pressure adaptation, we set out to study single proteins instead of entire organisms or entire genomes. This gave us the opportunity to study their dynamics and their response to pressure in much greater detail with respect to what it was possible to do before.

The organisms that were chosen are again *T. barophilus* and *T. kodakarensis* for the same reasons described before, and the proteins we selected were shown to be under positive selection for high-pressure. This means that the evolution of their primary sequence is positively correlated with the phenotype of their host (piezophile versus piezosensitive) for the selected evolutive pressure, that is, HHP in this case. This choice is motivated by the fact that such proteins should well represent the average behaviour of the proteome, so that the new results can be directly compared to the previous experiments.

The function of these protein is also known, so that their behaviour can be better put into context, and, eventually, their activity under pressure could be studied in future work. Moreover, the selected proteins are small, so that NMR spectroscopy can be effectively used to probe some of their properties.

We carried out a complete structural and dynamical study on two couples of piezophilic/piezosensitive proteins, by employing X-ray crystallography to unravel their structures at high resolution, and Molecular Dynamics (MD) simulations to relax them to a thermodynamically relevant state. Neutron scattering was then used to obtain information on the dynamics of these proteins under extreme conditions (high temperature and HHP), and High

Pressure Nuclear Magnetic Resonance (HP-NMR) provided useful insights and residue-specific information, to bridge the gap between dynamics and structure. This gives us a complete and detailed picture on the divergence between the selected proteins, allowing us to pinpoint the specific substitutions responsible for HHP adaptation.

From here, a first attempt in building a comprehensive model to explain pressure adaptation on the genome level can be pursued.

# METHODS

## 2.1 Proteins

### 2.1.1 Production of the proteins

In the previous section, the reasons behind the choice of the proteins employed in this work have been discussed. However, some of the experimental techniques we set out to employ, especially neutron scattering, require considerable quantities of sample. Thus, directly extracting them from their host organism would have posed an insurmountable challenge, therefore we opted for recombinant expression. It consists in inserting genetic code into a host organism that will in turn produce the proteins.

Before illustrating our strategy, it is useful to remind some concepts about gene expression (fig. 2.1). This process is divided in two phases, *transcription* and *translation*. During the first phase, the RNA polymerase binds to the DNA where the gene to be expressed is present, and starts building a complementary copy of it, but using *ribonucleotides* instead of *deoxy-ribonucleotides*. This is called *messenger* RNA, or mRNA, and it will be then translated by the *ribosome*, the protein-building machinery present in cells. The ribosome binds to a specific Ribosome Binding Site (RBS) in the mRNA, and starts then building the polypeptide. The amino acids needed for this are brought by the *transport* RNA, or tRNA, which presents *codons*, triads of nucleotides, that are complementary to those on the mRNA, thus *translating* the genetic sequence into an amino acid sequence.

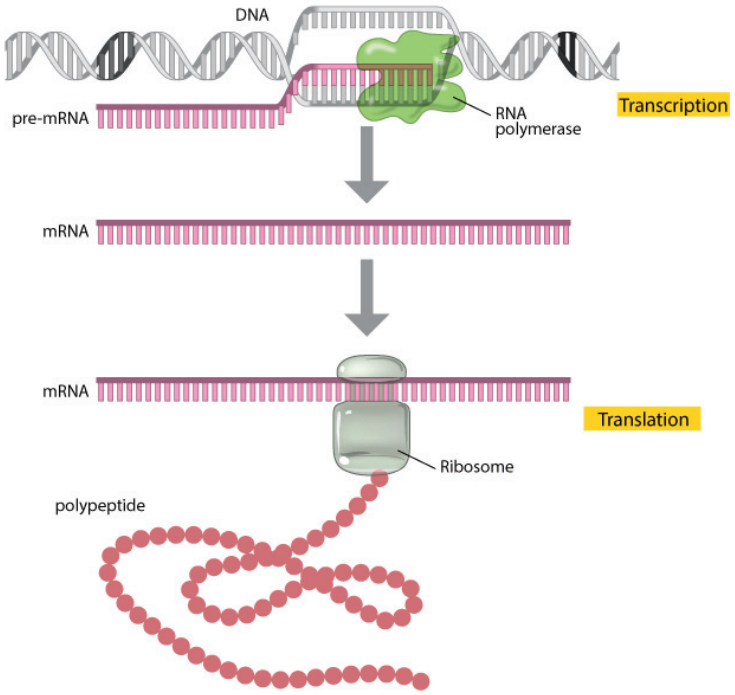


*Figure 2.1 – Schematic illustration of the gene expression process. Taken from*[77].

The most used host in molecular biology is undoubtedly *Escherichia coli*, thanks to its fast growth and ease of manipulation[78]. The technique used is that of *transformation*, where a circular fragment of double-stranded DNA, a *plasmid*, is inserted into *E. coli* cells by heat-shocking them.

To obtain the source DNA, the sequences coding for the target proteins were obtained from the complete genomes of *T. barophilus* and *T. kodakarensis* in the GenBank (accession codes CP002372 and AP006878 respectively), and then *codon-optimized*. This is necessary as the different codons can be associated to the same amino acid, but some organisms have higher concentration of specific tRNA in their cytoplasm[79]. Therefore the codon usage in *E. coli* is different than those of *T. barophilus* and *T. kodakarensis*, and optimizing the sequences can greatly improve the expression efficiency. The codon-optimized genes were then ordered from GENEWIZ Europe, a company that provides synthetic genes on demands.

The genes alone, however, are not sufficient for recombinant expression, they have to be coupled with an *expression system*. A very popular, reliable and vastly utilised choice is that of the T7 expression system[80]. It consists in a particular strain of *E. coli*, namely BL21(DE3), and a specific plasmid. This strain is genetically engineered to be efficient in protein production, and has

also been transfected with the DE3 bacteriophage in order to insert the gene coding for the T7 RNA polymerase in its genome. This RNA polymerase is vastly more efficient than the native one from *E. coli*, and it is also placed under the control of the *lac* promoter, so that the expression of the target protein can be tightly regulated by the experimenter by the addition of Isopropyl-$\beta$-D-1-thiogalactopyranoside (IPTG). This molecule, and analogue of galactose, can initiate the transcription of the T7 polymerase by binding to the *lac* repressor, which normally impedes the transcription of the genes placed downstream of its binding site.

The other component of this system is the plasmid, in our case pET-16b. Its main features are a ribosome-binding site specific for *E. coli*, a gene conferring resistance to ampicillin to the host, and the T7 promoter. The latter is what enables the transcription of the plasmid's content : when the culture is induced by the addition of IPTG, cells will start to produce the T7 RNA polymerase, which in turn will start the transcription of the plasmid's content thanks to the specific promoter. The produced mRNA will be then translated by the ribosomes thanks to the RBS.

The advantage of this system is that it is very efficient. In fact, when the production of the T7 polymerase is triggered, the cells devote a major part of their resources almost exclusively to its production[80]. However, this means that expression has to be tightly regulated, as the target protein could be toxic and lead to cell death in the host. This is achieved by the use of the *lac* promoter. Another advantage of this system is that IPTG cannot be degraded by *E. coli*, therefore even a small amount of it is sufficient to induce expression in a culture, and it will carry on until the culture is actually stopped.

### 2.1.2 Purification on the proteins

After culturing the cells to over-express the target proteins, they have to be extracted and purified. For protein extraction the options are rather limited by the fact that surfactants and detergents cannot be used, as they would unfold them. Another commonly used method is to lyse cells by incubating them with lysozyme, however, it has a similar molecular weight to our target proteins, and this would have greatly hampered the next steps of the purification. Therefore, we employed physical methods to *lyse* the cells. Freeze-thawing the cell suspension is a very effective method. During the

freezing part of the cycle, the water inside cells turns to ice, and therefore it expands, damaging the cell membrane. Thawing is required only to repeat the cycle multiple times. The disadvantage of this method is that it does not damage the chromosomal DNA present in the cells, which is the major responsible of the very high viscosity of the lysate. Therefore, additional treatment by sonication is needed in order to shear the DNA, decrease the viscosity, and homogenize the lysate for further processing. The tip of the sonicator vibrates at ultrasonic frequencies, and it generates cavitation bubbles in the suspension. The explosion of these bubbles in turn produces shock-waves that transfer mechanical energy to the components of the cells, taking them apart. However, proteins are able to resist this treatment and they maintain their structure even in these conditions.

After the elimination of cellular debris (DNA, membrane, cell wall, insoluble proteins and very high molecular weight components) by centrifugation, the result is a solution containing the soluble protein fraction, which can now be purified. A great advantage in the purification of proteins coming from hyperthermophilic organisms is that at this step, a heat-treatment can be performed. At 85°C, the target proteins are close to their organisms' optimum growth temperature, therefore they are able to comfortably retain their structure and solubility. Conversely, the native proteins from the host organism, *E. coli*, are past their denaturation temperature, therefore they unfold and aggregate, falling out of solution, and they can be then removed by centrifugation. The solution recovered after this step has a considerably lower total protein concentration, as most of the contaminants have been removed, and can thus be concentrated to greatly facilitate the successive purifications steps. This step was very efficient in performing the bulk of the purification, with a minimal loss of the target protein, thanks to its thermal stability.

To obtain a completely pure product, we employed a further purification step, that is, Size-Exclusion Chromatography (SEC). This technique makes use of a *column*, packed with a specific *resin*, and permits to separate the components of a protein solution based on their size.

The resin consists of porous beads that are immobilized by cross-linking. The principle of the technique is based on diffusion. The pores have definite sizes, and the time a particle will spend diffusing into the pore will depend on its size, with bigger particles being unable to penetrate them and directly

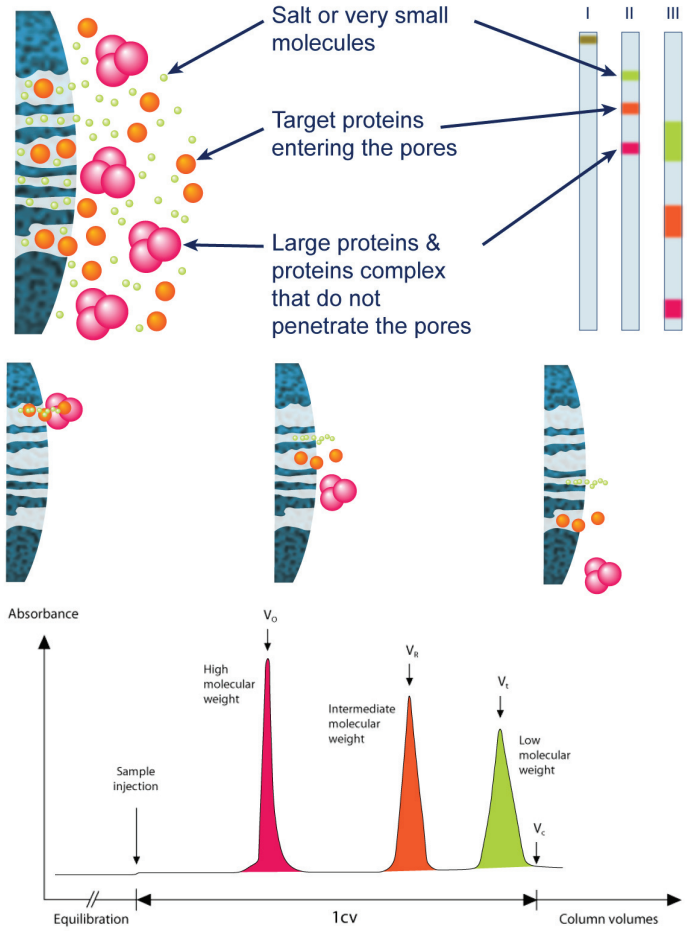


*Figure 2.2 – Scheme of the separation process in SEC. Taken from Interchim.*

travelling through the column, and small particles spending a considerable amount of time diffusing through the beads, being greatly slowed down. Therefore the bigger particles will be eluted first, and the smaller particles will be eluted last (fig. 2.2).

By combining these two methods, we were able to get pure proteins with minimal loss and very high efficiency, as shown in the supplementary information of both articles.

## 2.2 Neutron Scattering

Since many decades, scientists have been using light to investigate the properties of matter by means of different techniques. Among those, X-ray diffraction was the first to succeed in providing atomic resolution, because the wavelength range of this radiation ($10^{-2}$ $nm$ to 10 $nm$) is comparable to interatomic distances. However, X-rays are highly energetic (100 $eV$ to 100 $keV$), hence they are not perfectly suited for the study of biological or soft matter samples, as they can destroy them. This and other problems can be overcome by using neutrons instead of light.

The neutron was discovered in 1932 by James Chadwick[81], in a period during which the exact nature of the atom was still unclear. But, thanks to this discovery and to that of *nuclear fission*, the first nuclear reactors were built. Since then, numerous scientific discoveries, such as that of *antiferromagnetism* by Louis Néel[82] have been made thanks to neutron scattering experiments, made possible by the existence of large scale facilities based either on reactor sources, like the *Intitut Laue-Langevin* in Grenoble and the *High Flux Isotope Reactor* at the *Oak Ridge National Laboratory*, or on spallation sources, like the *ISIS Neutron and Muon Source* at the *Rutherford Appleton Laboratory* in Oxfordshire. The general properties of the neutron are summarized in table 2.1.

| | |
|---|---|
| mass | $1.674927471 \times 10^{-27}$ $kg$ |
| spin | $\frac{1}{2}$ |
| charge | 0 |
| magnetic moment | $-1.91304273$ $\mu_N$ |

*Table 2.1 – Basic properties of the neutron*

In nuclear reactors, neutrons appear as a result of the fission process of an uranium fuel element and are then thermalized inside a *moderator* (heavy water). Therefore, the velocity distribution of the neutrons becomes a Maxwell-Boltzmann distribution, centered around a value that corresponds to the moderator temperature. The outcome of this process is a beam of *thermal* neutrons ($\langle E \rangle \sim 25$ $meV$ for a moderator temperature of 300 $K$) that can be used to perform numerous types of experiments.

Thermal neutrons have various interesting characteristics. First of all, their

de Broglie wavelength is comparable to interatomic distances, enabling them to probe the structure and dynamics of any kind of sample from a microscopic point of view. Moreover, their energy is of the same order as $k_B T$ at room temperature, hence it is similar to that of many excitations in soft and condensed matter.

Neutrons are uncharged particles, thus they penetrate deeply into matter and they interact with matter only through the *strong nuclear force*. This is a great advantage, since the dimensions of nuclei are typically much smaller than the neutron wavelength, so the process can be treated as a point-like interaction. Furthermore, this force depends on the atomic number and the isotope of an element in a non-systematic way, and it is much stronger for hydrogen. On the contrary, in X-ray scattering the light elements are almost invisible (fig. 2.3).

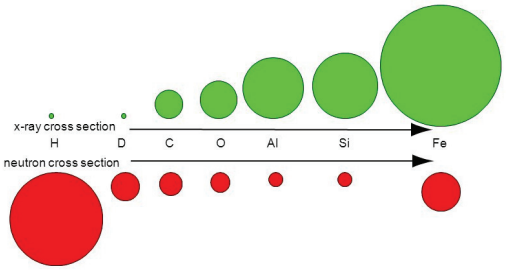


*Figure 2.3 – Illustration of X-ray and neutron scattering cross-sections for some elements and isotopes.*

Neutrons also have a *magnetic moment*, thus they are also able to probe the magnetic properties of matter by means of what is called *magnetic scattering*.

### 2.2.1 General theory

In this section, the basic theoretical aspects involved in neutron scattering will be described, in order to point out the power of this technique. This treatment is not exhaustive by any means, and the proper development of this theory can be found in classic textbooks by Squires[83] or Bée[84].

The first element to take into account to describe this formalism is the geometry of a typical neutron scattering experiment (fig. 2.4), where a monochromatic beam of neutrons, with momentum $\mathbf{k}$ is impinging on the sample along the $z$ axis. The neutrons are then scattered in all directions after inter-

acting with the sample, and we will indicate their final momentum with $\mathbf{k}'$. The direction in which the neutrons are scattered is defined by the angles $\theta$ and $\phi$, which in turn define the solid angle $d\Omega$ inside which the neutrons are detected during an experiment.

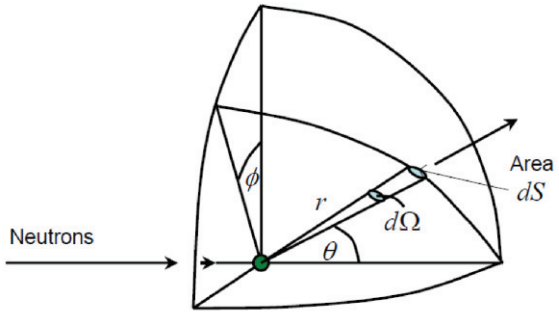


*FIGURE 2.4 – Typical scattering geometry. $d\Omega$ is the solid angle in the direction of the scattered neutrons, and it is defined by the angles $\theta$ and $\phi$. Taken from [83].*

In the case of planar geometry, the angle $\theta$ is directly connected with the difference in the neutron's momentum *before* and *after* the interaction with the sample. Therefore, it is useful to define the difference $\mathbf{q} = \mathbf{k} - \mathbf{k}'$, which is referred to as the *scattering vector* (fig. 2.5), and its modulus depends on the scattering angle $\theta$ according to $q = \frac{4\pi}{\lambda}\sin\left(\frac{\theta}{2}\right)$.

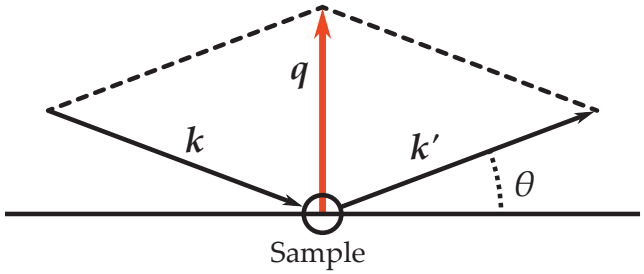


*FIGURE 2.5 – Graphical visualization of the scattering vector* $\mathbf{q}$.

We can now define another important quantity : the *scattering cross-section*. It is designated by $\sigma$, has the dimensions of an area and is commonly measured in *barns* ($10^{-24}$ $cm^2$). The scattering cross-section can be interpreted as the probability for a neutron to be scattered by a nucleus, thus it is defined as the ratio between the number of scattered neutrons and the flux of incident neutrons. However, in a scattering experiment, the scattered neutrons are typically measured in a given direction, and they will also exchange energy with the sample. Hence, it is useful to define the *double differential scattering cross-section* $\frac{d^2\sigma}{d\Omega dE}$, that is, the probability for a neutron to be scattered in a direction into the solid angle $d\Omega$ (fig. 2.4) with an energy gain of $dE$.

The last missing piece in this theory concerns the interaction between the neutrons and the sample. The forces responsible for scattering have a range of about $10^{-14}$ to $10^{-15}$ $m$, while the typical wavelength of thermal neutrons is of the order of $10^{-10}$ $m$. This allows us to suppose spherical symmetry, and a point-like interaction, which is described by the *Fermi pseudopotential*[85] :

$$V(\mathbf{r}) = \frac{2\pi\hbar^2}{m} b\ \delta(\mathbf{r} - \mathbf{R}), \tag{2.1}$$

where $\hbar$ is the reduced Planck constant, $m$ is the mass of the neutron and $\mathbf{R}$ is the position of a nucleus in the sample. A new quantity has been defined here, the *scattering length* $b$, and its value depends on the nucleus the neutron is interacting with. To date, we still do not have a complete theory for the strong nuclear interaction, therefore we are not able to predict the value of $b$ for all the nuclei, and those values have been determined experimentally.

Omitting the full quantum-mechanical treatment needed to derive it, we can now give the general expression for the *double-differential scattering cross-section* :

$$\frac{d^2\sigma}{d\Omega dE} = \frac{1}{N}\frac{k'}{k}\frac{1}{2\pi\hbar}\sum_{i,j}\int_{-\infty}^{\infty}\left\langle b_i b_j e^{-i\mathbf{q}\cdot\mathbf{R}_i(0)} e^{i\mathbf{q}\cdot\mathbf{R}_j(t)}\right\rangle e^{-i\omega t}dt. \tag{2.2}$$

The brackets in eq. 2.2 indicate a thermal average over all the possible states of the system, $N$ is the number of nuclei in the sample, and $\omega$ represents the energy exchanged by the neutrons according to $\hbar\omega = E_{k'} - E_k$. The physical properties of the system are contained in the Heisenberg position operator $\mathbf{R}(t)$ relative to all nuclei in the sample.

### 2.2.2 Coherent and Incoherent scattering

Until now, we neglected the role of spin in the scattering process. Different spin states even of the same isotope result in different scattering lengths. This leads to very interesting consequences when the coupling between the neutron spin and the nuclear spin is taken into account. An usual approximation, that is verified in most of the cases, is to consider the spin state and the positions of the nuclei to be independent or, in other words, that the

probability to find a nucleus with scattering length $b_i$ is not correlated to the probability of finding it in the position $\mathbf{R}_i$. This enables us to decouple the thermal average over the scattering lengths and the positions. The term in brackets of eq. (2.2) thus becomes

$$\left\langle b_i b_j \right\rangle \left\langle e^{-i\mathbf{q}\cdot\mathbf{R}_i(0)} e^{i\mathbf{q}\cdot\mathbf{R}_j(t)} \right\rangle . \tag{2.3}$$

Furthermore, this assumption allows us to decouple the average of the scattering lengths that refer to *different* sites, thus

$$\begin{aligned} \left\langle b_i b_j \right\rangle &= \left\langle b_i \right\rangle \left\langle b_j \right\rangle = \left\langle b \right\rangle^2 \quad \text{if} \quad i \neq j \\ \left\langle b_i b_j \right\rangle &= \left\langle b_i^2 \right\rangle = \left\langle b^2 \right\rangle \qquad \text{if} \quad i = j. \end{aligned} \tag{2.4}$$

In general, $\left\langle b \right\rangle^2$ and $\left\langle b^2 \right\rangle$ are *not* the same. We can thus define two quantities, called the *coherent* and *incoherent* scattering lengths :

$$b_{coh} = \left\langle b \right\rangle \; ; \quad b_{inc} = [\left\langle b^2 \right\rangle - \left\langle b \right\rangle^2]^{1/2}. \tag{2.5}$$

Equation (2.2) can be rewritten in order to separate the contribution to the sum given by the same nucleus, and the definitions in 2.5 can be applied, giving

$$\begin{aligned} \frac{d^2\sigma}{d\Omega dE} =& \frac{1}{N}\frac{k'}{k}\frac{1}{2\pi\hbar} b_{coh}^2 \sum_{i,j} \int_{-\infty}^{\infty} \left\langle e^{-i\mathbf{q}\cdot\mathbf{R}_i(0)} e^{i\mathbf{q}\cdot\mathbf{R}_j(t)} \right\rangle e^{-i\omega t} dt \quad + \\ &+ \frac{1}{N}\frac{k'}{k}\frac{1}{2\pi\hbar} b_{inc}^2 \sum_{i} \int_{-\infty}^{\infty} \left\langle e^{-i\mathbf{q}\cdot\mathbf{R}_i(0)} e^{i\mathbf{q}\cdot\mathbf{R}_i(t)} \right\rangle e^{-i\omega t} dt. \end{aligned} \tag{2.6}$$

By defining $\sigma_{coh} = 4\pi b_{coh}^2$ and $\sigma_{inc} = 4\pi b_{inc}^2$, we finally get

$$\frac{d^2\sigma}{d\Omega dE} = \frac{\sigma_{coh}}{4\pi N}\frac{k'}{k}\frac{1}{2\pi\hbar}\sum_{i,j}\int_{-\infty}^{\infty}\left\langle e^{-i\mathbf{q}\cdot\mathbf{R}_i(0)}e^{i\mathbf{q}\cdot\mathbf{R}_j(t)}\right\rangle e^{-i\omega t}dt \quad + \\ + \frac{\sigma_{inc}}{4\pi N}\frac{k'}{k}\frac{1}{2\pi\hbar}\sum_{i}\int_{-\infty}^{\infty}\left\langle e^{-i\mathbf{q}\cdot\mathbf{R}_i(0)}e^{i\mathbf{q}\cdot\mathbf{R}_i(t)}\right\rangle e^{-i\omega t}dt. \tag{2.7}$$

The first term of (2.7) is called *coherent* scattering cross-section and it gives information about the positions of *different* nuclei at different times. It is therefore the result of *interference effects* arising from the particular structure of the sample and from the collective motions of its atoms. Instead, the second therm is called *incoherent* scattering cross-section and it accounts for the position of *the same* nucleus at different times. It is affected by the dynamics of atoms and of their local environment.

### 2.2.3 The scattering function

We will now focus our attention on the definition and on the physical meaning of the *scattering function* $S(\mathbf{q}, \omega)$. In the previous section we derived a general expression for the double differential scattering cross-section. The probability of interaction between the neutrons and the nuclei present in the sample are included in the coherent and incoherent cross sections, $\sigma_{coh}$ and $\sigma_{inc}$, whereas the integrals contain only the sample's properties. This leads us to the following definition

$$\frac{d^2\sigma}{d\Omega dE} = \frac{1}{4\pi\hbar}\frac{k'}{k}\left[\sigma_{coh}S_{coh}(\mathbf{q},\omega) + \sigma_{inc}S_{inc}(\mathbf{q},\omega)\right], \tag{2.8}$$

where

$$S_{coh}(\mathbf{q},\omega) = \frac{1}{2\pi}\frac{1}{N}\sum_{i,j}\int_{-\infty}^{\infty}\left\langle e^{-i\mathbf{q}\cdot\mathbf{R}_i(0)}e^{i\mathbf{q}\cdot\mathbf{R}_j(t)}\right\rangle e^{-i\omega t}dt \tag{2.9}$$

and

$$S_{inc}(\mathbf{q},\omega) = \frac{1}{2\pi}\frac{1}{N}\sum_{i}\int_{-\infty}^{\infty}\left\langle e^{-i\mathbf{q}\cdot\mathbf{R}_i(0)}e^{i\mathbf{q}\cdot\mathbf{R}_i(t)}\right\rangle e^{-i\omega t}dt \tag{2.10}$$

are the *coherent* and *incoherent scattering functions* and $N$ is the number of scatterers present in the sample.

The scattering function depends on two variables, the exchanges of energy and momentum during the process, namely $\hbar\omega$ and $\hbar\mathbf{q}$. It can be shown that $S(\mathbf{q}, \omega)$ is a time Fourier transform :

$$S(\mathbf{q}, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} I(\mathbf{q}, t) e^{-i\omega t} dt, \tag{2.11}$$

where

$$I(\mathbf{q}, t) = \frac{1}{N} \sum_{i,j} \left\langle e^{-i\mathbf{q}\cdot\mathbf{R}_i(0)} e^{i\mathbf{q}\cdot\mathbf{R}_j(t)} \right\rangle \tag{2.12}$$

is called the *intermediate scattering function*, and it describes the time evolution of the motions occurring in the sample. Similarly, from the incoherent scattering function we can derive the *self* intermediate scattering function :

$$S_{inc}(\mathbf{q}, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} I_s(\mathbf{q}, t) e^{-i\omega t} dt, \tag{2.13}$$

where

$$I_s(\mathbf{q}, t) = \frac{1}{N} \sum_{i} \left\langle e^{-i\mathbf{q}\cdot\mathbf{R}_i(0)} e^{i\mathbf{q}\cdot\mathbf{R}_i(t)} \right\rangle. \tag{2.14}$$

To better explain the information contained in these functions, we are going to real space, by taking now the space Fourier transform of the intermediate function :

$$G(\mathbf{r}, t) = \frac{1}{(2\pi)^3} \int I(\mathbf{q}, t) e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{q}. \tag{2.15}$$

$G(\mathbf{r}, t)$ is the *time dependent pair correlation function*, that is the probability of finding a nucleus in the position $\mathbf{r}$ at time $t$ given that another nucleus is at the origin at time $t = 0$, and it is expressed as :

$$G(\mathbf{r},t) = \frac{1}{(2\pi)^3}\frac{1}{N}\sum_{i,j}\int \left\langle e^{-i\mathbf{q}\cdot\mathbf{R}_i(0)} e^{i\mathbf{q}\cdot\mathbf{R}_j(t)} \right\rangle e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{q}. \tag{2.16}$$

With the same reasoning, starting from the self intermediate scattering function, we find the time dependent *self* correlation function, which expresses the probability of finding a nucleus in a position $\mathbf{r}$ at time $t$ given that the *same* nucleus was at the origin at time $t = 0$ :

$$G_s(\mathbf{r},t) = \frac{1}{(2\pi)^3}\frac{1}{N}\sum_{i}\int \left\langle e^{-i\mathbf{q}\cdot\mathbf{R}_i(0)} e^{i\mathbf{q}\cdot\mathbf{R}_i(t)} \right\rangle e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{q}. \tag{2.17}$$

Equations (2.16) and (2.17) express what we already stated about the scattering function and the correlation function. Now we know that they are related by a space-time Fourier transform

$$S(\mathbf{q},\omega) = \frac{1}{2\pi}\int d\mathbf{r}\int_{-\infty}^{\infty} G(\mathbf{r},t) e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)} dt, \tag{2.18}$$

therefore the behaviour of the scattering function at low frequencies reflects the long-time limit of the correlation function, which describes the dynamics of the slow processes occurring within the sample. Conversely, the high-frequency features of the scattering function are related to the faster internal motions of the system under study. The same relation exists between the properties of the scattering function in reciprocal space and those of the correlation function in real space : the low-$\mathbf{q}$ behaviour of the scattering function depends on the long-range spatial correlations in the structure and dynamics of the sample, whereas its high-$\mathbf{q}$ limit gives information on short-scale correlations.

### 2.2.4 Neutron Scattering Techniques

The scattering function is a powerful tool to investigate both the structure and the dynamics of a sample on a large range of length and time scales. However, it is very difficult, if not impossible, to investigate all the properties of a sample in a single measurement. For this reason, numerous different neutron techniques and instruments have been developed during the years[84]. We can

distinguish them by the energy exchange that usually takes place during an experiment (fig. 2.6).

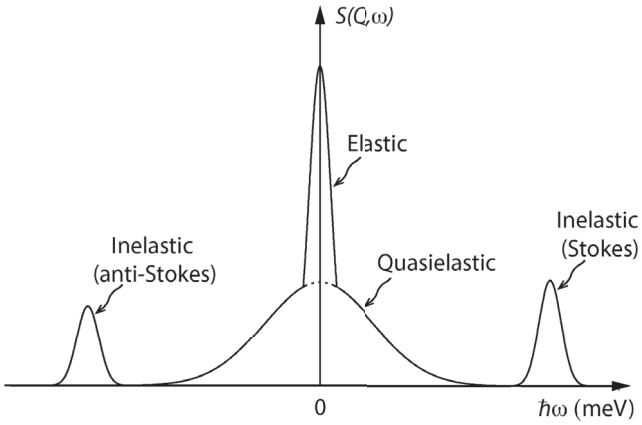


*Figure 2.6 – An example of a typical neutron scattering spectrum as a function of exchanged energy at a given* $\mathbf{q}$[83].

In an Elastic Neutron Scattering (ENS) experiment, the scattered neutrons have the same energy as the incident ones. In both Quasielastic Neutron Scattering (QENS) and Inelastic Neutron Scattering (INS) the energy of the neutron is changing. However, in the case of QENS the exchange is very small ($1\ \mu eV - 1\ meV$) and is often interpreted as a broadening of the elastic peak due to diffusive motions, while in INS the energy exchange is higher ($1-100\ meV$), compatible with vibrations, and the peaks are clearly separated from the elastic peak.

The state of the sample influences the experiment as well, because it defines the importance of the coherent or incoherent parts on the total signal. As a simple example we can think about a biological sample, which is usually in solution or hydrated powder form and contains a large number of hydrogen atoms. The large incoherent cross section and the disordered state of the sample will lead to a signal dominated by incoherent contributions of hydrogen nuclei. On the other hand, the signal coming from a highly ordered crystal will be mostly coherent. With this in mind, we can distinguish some general types of neutron experiments.

In an *Elastic Coherent Neutron Scattering* experiment, or *Diffraction* as it is commonly called, the measured signal will reflect the spatial correlation of the nuclei, therefore giving information about the structure of the sample.

*Inelastic Coherent Neutron Scattering* probes motions that are spatially correlated, like collective vibrations, or phonons, in ordered crystals.

*Elastic Incoherent Neutron Scattering* can be used to study the dynamics of hydrogen containing samples on a rather fast time scale, as we will see in the next section.

*Quasi-Elastic Incoherent Neutron Scattering* gives informations about slower processes and it can be used to describe the geometry of these motions.

*Inelastic Incoherent Neutron Scattering* allows the study of individual vibrations and of the environment around the scattering atoms. It is complementary to other vibrational spectroscopies, like infrared or Raman.

### 2.2.5 Elastic Incoherent Neutron Scattering

Asserting that an elastic scattering technique can be used to study the dynamics of a sample may seem confusing at first. The reason why this is possible resides in the fact that the purely elastic scattered intensity is not constant during an experiment, it is varying depending on external conditions, as temperature for example. This is at the heart of Elastic Incoherent Neutron Scattering (EINS).

Let us take the limit for $t \to \infty$ of the self intermediate function

$$\lim_{t\to\infty} I_s(\mathbf{q}, t) = \frac{1}{N}\sum_i \left\langle e^{-i\mathbf{q}\cdot\mathbf{R}_i(0)} e^{i\mathbf{q}\cdot\mathbf{R}_i(\infty)} \right\rangle = I_s(\mathbf{q}, \infty). \tag{2.19}$$

Note that in this limit, $I(\mathbf{q}, \infty)$ is not a function of time anymore. We can now write the incoherent scattering functions in the limit for $\omega \to 0$ as the Fourier transform of (2.19)

$$\begin{aligned} S_{inc}(\mathbf{q}, 0) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} I_s(\mathbf{q}, \infty) e^{-i\omega t} dt = \\ &= \frac{1}{2\pi}\frac{1}{N}\sum_i \left\langle e^{-i\mathbf{q}\cdot\mathbf{R}_i(0)} e^{i\mathbf{q}\cdot\mathbf{R}_i(\infty)} \right\rangle \int_{-\infty}^{\infty} e^{-i\omega t} dt = \\ &= \frac{1}{N}\sum_i \left\langle e^{-i\mathbf{q}\cdot\mathbf{R}_i(0)} e^{i\mathbf{q}\cdot\mathbf{R}_i(\infty)} \right\rangle \delta(\omega). \end{aligned} \tag{2.20}$$

The time-independent factor multiplying the *delta* function into $S_{inc}(\mathbf{q}, 0)$ is often called Elastic Incoherent Structure Factor (EISF), and it represents the fraction of scatterers that give contributions to the elastic peak, represented

by the delta function. As we can see from (2.20), the $i$-th nucleus will give a non-zero contribution to the EISF only if its position is *not* changed between $t = 0$ and $t = \infty$. In other words, only non-moving nuclei will scatter neutrons elastically. Therefore, one can study the activation of motions by measuring the EISF as a function of an external parameter, e.g. temperature or pressure. This can be done by fitting it with suitable models[86–89].

There is another factor that has to be taken into account : neutron spectrometers are not infinitely precise. The measured EISF is always the result of a *convolution* between the actual $S_{inc}(\mathbf{q}, 0)$ and a *resolution function* $R(\mathbf{q}, E, \Delta E)$, which is typical for the particular spectrometer being used for the measurement.

$$S_{measured}(\mathbf{q}, 0) = S_{inc}(\mathbf{q}, 0) \otimes R(\mathbf{q}, E, \Delta E), \tag{2.21}$$

where $\otimes$ is the convolution operator and $\Delta E$ is the energy resolution of the spectrometer. Hence, the elastic peak has not the form of a delta function any more, it is now a finite peak-shaped function with a width of $\Delta E$. This means that every neutron which has gained or lost an amount of energy that is less than $\Delta E$ will still give a contribution to the EISF.

The effect of this in the time domain is to reduce the limiting time value of the self intermediate function from infinity to a finite value $\tau$. This value can be related to the energy resolution via the time-energy Heisenberg uncertainty relation, thus $\tau \sim \frac{\hbar}{\Delta E}$. We can now understand why particular spectrometers with a definite *energy resolution* are said to have different *time windows*.

A common experiment that is performed with this technique is the *elastic scan*, a measurement of the elastic intensity while continuously changing the temperature of the sample. This allows to probe the activation of fast motions in the sample as the temperature is increased, since more neutrons will exchange energy with the moving scatterers, and the elastic scattering intensity would thus decrease.

In this work, elastic scans have been performed at different pressure points, to probe protein dynamics. The model we used is a *two-state model*, first developed by Stoeckli and colleagues[90] to explain hydrogen dynamics in carboxylic acids. This model proved to be very useful in the description of protein dynamics as well, and it was used by Doster et al[86] in the first pioneering studies about the *Protein Dynamical Transition* in Myoglobin. Librizzi

et al[68] also employed it on a similar system to study the pressure effect on protein dynamics.

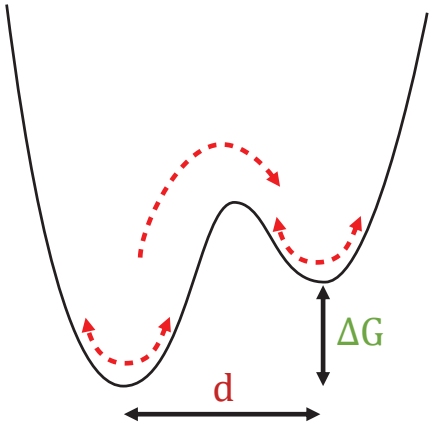


FIGURE 2.7 – *Scheme of the two-state model.*

The main hypothesis of the two-state model is that hydrogen can concomitantly perform two kind of motions, i) harmonic vibrations about their equilibrium position, and ii) jumps between two asymmetric potential wells (fig. 2.7). The $q$-dependence of the elastic intensity according to this model has the form :

$$S(q) = e^{-q^2\Delta x_0^2}\left[1 - 2p_1p_2\left(1 - \frac{sin(qd)}{qd}\right)\right], \tag{2.22}$$

where $\Delta x_0^2$ is the harmonic component of the Mean Square Displacement (MSD), $p_1$ and $p_2$ are the populations of the two wells, and $d$ is the distance between them. The free energy asymmetry can be deducted from the shift in their population with temperature, and by assuming that it follows the Arrhenius law ($p_1/p_2 = exp(-\Delta H/RT + \Delta S/R)$, where $R$ is the gas constant), it can be separated into the enthalpic and entropic components. Hence, the main advantage of this model is that it directly reflects an average representation of the protein energy landscape, as a function of external parameters.

### 2.2.6 Quasielastic Neutron Scattering

From the theoretical point of view, QENS describes the *broadening* of the elastic line. The $S(q, \omega)$ is therefore measured as a function of the neutron energy exchange other than the momentum exchange. However, the energy

exchange in this case is very small, thus the quantized excitation (i.e. vibrations) are ignored, as the peaks arising from them are clearly separated from the broadened elastic line in the energy space. The main contribution to the signal are in general *diffusive* and *sub-diffusive* motions, and in the case of protein samples they arise from the incoherent scattering of hydrogen atoms.

In a strict sense, following statistical physics, a purely diffusive motion should not give any elastic contribution, therefore it is not correct to describe the QENS signal as a broadening of an elastic line. However, in proteins the situation is more complicated.

Hydrogen atoms are mostly evenly distributed in proteins[91] and while a fraction of them may perform diffusive motions, a different fraction may only perform *localized* motions, and other fractions may perform slower motions that cannot be resolved by the spectrometer (fig. 2.8). This is a direct consequence of the hierarchical arrangement of the energy landscape of proteins[69], therefore great care has to be taken in the choice of the spectrometer, and thus the energy resolution, that will be employed to study a protein. A different resolution results therefore in a different *tier* of the protein energy landscape that can be studied.

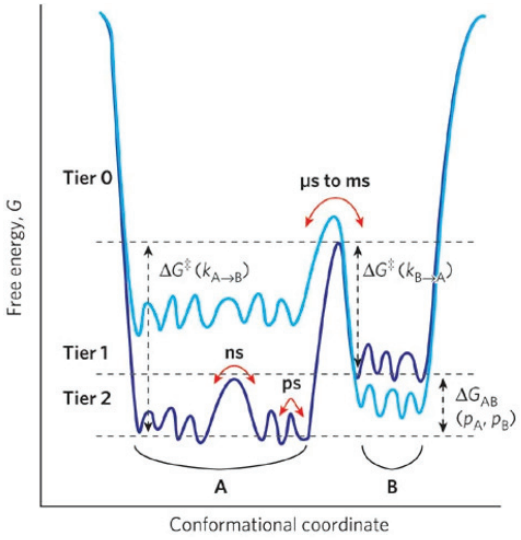


*Figure 2.8 – Representation of the hierarchical energy landscape of proteins. Taken from*[92].

The immediate consequence of the dynamical heterogeneity and the coupling of protein dynamics over several time scales is that the overall shape of the QENS spectrum of a protein is in general not *Lorentzian*, as it would

be for purely diffusive motions. This problem is generally tackled in two different ways, the analysis of the intermediate scattering function (eq. 2.14) in the time domain with models that take into account the non-exponential nature of the relaxation processes[93–95], or a spectral analysis of $S(q, \omega)$ in the frequency domain.

In this work, we have undertaken the second approach, as it allows the separation of different dynamical contributions to the signal and the characterization of their geometry from their q-dependence[84,96,97]. From a careful model-free analysis, we identified the main contribution to the signal and successively built a model to describe them. The latter includes localized motions arising mainly from methyl group rotations and $CH_2$ librations, plus a restricted *jump-diffusion* motion arising from the residue's side-chains.

The thorough mathematical derivation of our model can be found in Article 1 and in the supplementary information, however, the choice of the model by Hall and Ross[98] to characterise the *jump-diffusion* motion deserves further comment. Most works on protein dynamics studied with QENS have been taking advantage of the model by Singwi and Sjölander[99], which is very general and was developed to describe the diffusion of liquid water in bulk. It supposes that water molecules alternatively perform vibrations around an equilibrium position for a certain time, and then diffuse freely for another interval. In fact, the form that is frequently used is the asymptotic form where the second interval tends to zero, effectively describing instantaneous jumps between the different equilibrium positions.

The main difference between the latter form and the Hall-Ross model lies in the assumption made on the jump length distribution, which is uniform in the first case and Gaussian in the second. While a uniform jump distribution is a reasonable assumption for the dynamics of liquid water, which does not undergo any geometrical constraint, we deemed the choice of a Gaussian distribution more appropriate to describe the dynamics of side-chains in a protein. Moreover, the Hall-Ross model also enabled us to take into account the effects of confinement on protein dynamics by the analysis of the EISF.

## 2.3 Nuclear Magnetic Resonance Spectroscopy

### 2.3.1 Basic principles

The discovery dates back to 1938 by Isidor Isaac Rabi[100], who was later awarded the Nobel Prize in 1944. The basic theoretical principles governing NMR are purely quantum-mechanical, and rely on the properties of nuclear *spin*. Its associated angular momentum is *quantized*, though in normal circumstances, different values of spin correspond to the same energy value, meaning that the energy levels associated to spin are *degenerate*.

However, for charged particles, one can always associate a *magnetic moment* $\mu$ to its angular momentum $I$, through the *gyromagnetic ratio* $\gamma$ :

$$\boldsymbol{\mu} = \gamma \mathbf{I}. \tag{2.23}$$

Therefore, when a charged particle with non-zero spin is placed in a magnetic field $\mathbf{B_0}$, the spin will align to it. The potential energy for this interaction is given by :

$$E = -\boldsymbol{\mu} \cdot \mathbf{B_0} \tag{2.24}$$

In the simplest case, for a spin $\frac{1}{2}$ particle, there are only two possible states allowed by the quantization, and the spin can align either in a parallel or antiparallel fashion to the magnetic field (fig. 2.9).

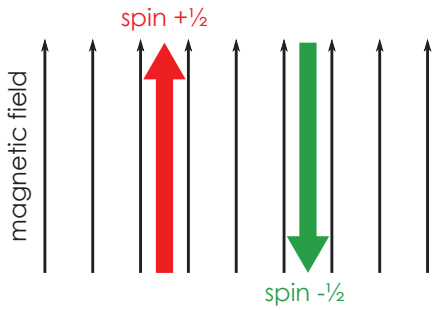


*Figure 2.9 – Alignment of $\frac{1}{2}$ spins in a magnetic field.*

This eliminates the degeneracy for the two energy levels, and one can start to observe *transitions* between them by using electromagnetic fields, inducing the *resonance* phenomenon (fig. 2.10). As the energy difference between the two levels is of the order of $\mu eV$, using the $E = h\nu$ relation, one finds that the frequency of the electromagnetic wave needed to excite the transition falls

into the range of Radio-frequencies (RF).

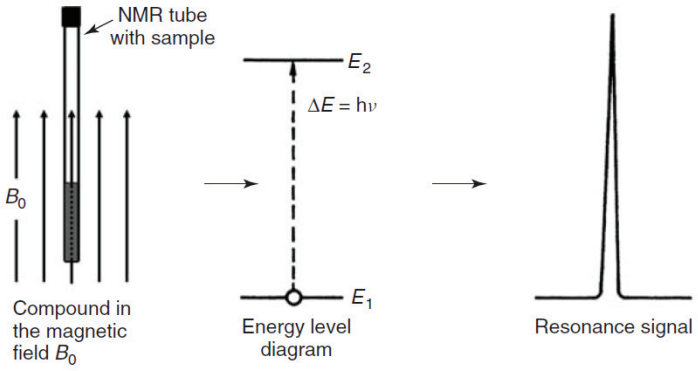


*Figure 2.10 – Basic scheme of an NMR experiment. Taken from*[101].

As an example, with a 21.7 T magnet, the splitting between the energy levels for a proton (hydrogen nucleus) will be of 600 MHz. In fact, an NMR spectrometer would be usually identified by its proton resonant frequency instead of its magnet strength (e.g. one would say *a 600 MHz spectrometer* instead of *a 21.7 T spectrometer*). However, it must be stressed again that this technique is applicable only to samples that contain NMR-active isotopes in it, that is, nuclei that have non-zero spin.

When a sample is placed in a magnetic field, its electrons will also react to it, as they are also charged and possess both spin and *orbital* angular momentum. They will thus generate a *local magnetic field* that will in turn modify the resonance frequency of the nucleus. This effect is known as *chemical shift*. The outstanding sensibility to the electronic environment of each nucleus is what has driven the importance of NMR spectroscopy in chemical sciences.

These shifts are usually very small, and they are measured in parts per million (ppm) (e.g. a 10 ppm shift means that the frequency shift would be 1000 Hz in a 100 MHz spectrometer). This is also useful to compare data between spectrometers that have different resonant frequencies, as ppm are normalized to it. Nevertheless, thanks to the very high sensitivity of this technique, different features can be resolved. A more complete and thorough treatment of both the theoretical and experimental details of NMR can be found in textbooks[101,102].

### 2.3.2 Protein NMR

Recently, with the advent of superconducting magnets with very high fields and cryogenic probes with unmatched sensitivity, the application of NMR to proteins has found widespread use.

The main challenge in the study of proteins by NMR is constituted by their size. First, the sheer number of resonances from a single protein makes their spectra considerably crowded, and a great sensitivity is needed to discern between different peaks. Second, the width of an NMR peak is connected with the spin relaxation time, and the biggest contribution to it comes from the rotational correlation time of the protein molecule[102]. This means that in bigger proteins, which will present longer correlation times because of their hydrodynamic properties, the peaks will be rather broad, defeating the purpose of distinguishing them in a very crowded spectrum.

For these reasons, numerous advanced techniques have been developed to study proteins, taking advantage of *multidimensional* NMR. They consist in the use of *correlations* among different nuclei by transferring the magnetization between them. These techniques can be categorized either based on which nuclei they use, or based on how they achieve the magnetization transfer. Multidimensional techniques that make use of the same isotope are called *homonuclear*, while those that correlate different nuclei are called *heteronuclear*. The magnetization transfer can either happen *through-bond* or *through-space*, exploiting the Nuclear Overhauser Effect (NOE)[103]. The latter is a dipole-dipole coupling between magnetic moments, and its intensity is proportional to $1/r^6$, therefore it gives information on the distances between the coupled spins.

The major drawback of heteronuclear techniques is that they require the *isotopic labelling* of the protein under study, as the proton is the only NMR-active nucleus normally present. Therefore, $^{15}$N or $^{13}$C labelling are necessary for these experiments. However, when a protein can be successfully labelled with both $^{15}$N and $^{13}$C, it is possible to solve its entire structure by means of *triple-resonance* experiments, by transferring the magnetization through the backbone performing the so-called *sequential walking*, and then applying spatial constraints by performing *through-space* experiments.

In this work we exclusively employed the $^{1}$H and $^{15}$N Heteronuclear Single

Quantum Coherence (HSQC) technique, since the structures of the proteins were already known by X-ray crystallography. As the name suggests, this technique is heteronuclear, making use of the proton and $^{15}$N resonances, and the magnetization transfer happens *through-bond*. The main contribution to the signal in this type of experiment is given by the amide H-N bond in the protein backbone (fig. 2.11). Resonances from the side-chains of nitrogen-containing amino acids are present nonetheless, though they are normally shifted with respect to amide resonances, and they do not prevent a proper interpretation of the spectra.

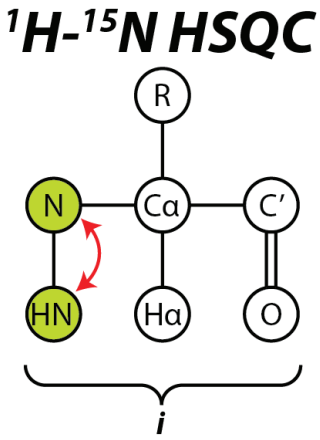


*Figure 2.11 – The magnetization transfer happening on the protein backbone during an HSQC experiment.*

To perform the experiment, the proton resonance is first excited, as it is the most sensitive and yields a higher signal to noise ratio, the magnetization is then transferred to the $^{15}$N and, after an *evolution time*, transferred back to the proton, and the signal is then recorded. This sequence is repeated multiple times by varying the evolution time. This variation is what gives the indirect information about the heteronucleus ($^{15}$N in our case).

The result of this is that the peaks that are normally overlapped in the proton dimension, are "spread out" into the further $^{15}$N dimension (fig. 2.12). Therefore, after performing the *assignment* of the peaks by further 3-dimensional experiments, information can be gathered about every single residue in the protein. Performing series of HSQC experiments while varying an external parameter allows the study of the response of protein conformation to said parameter at a residue level, with an extent of detail that is not possible with other techniques.

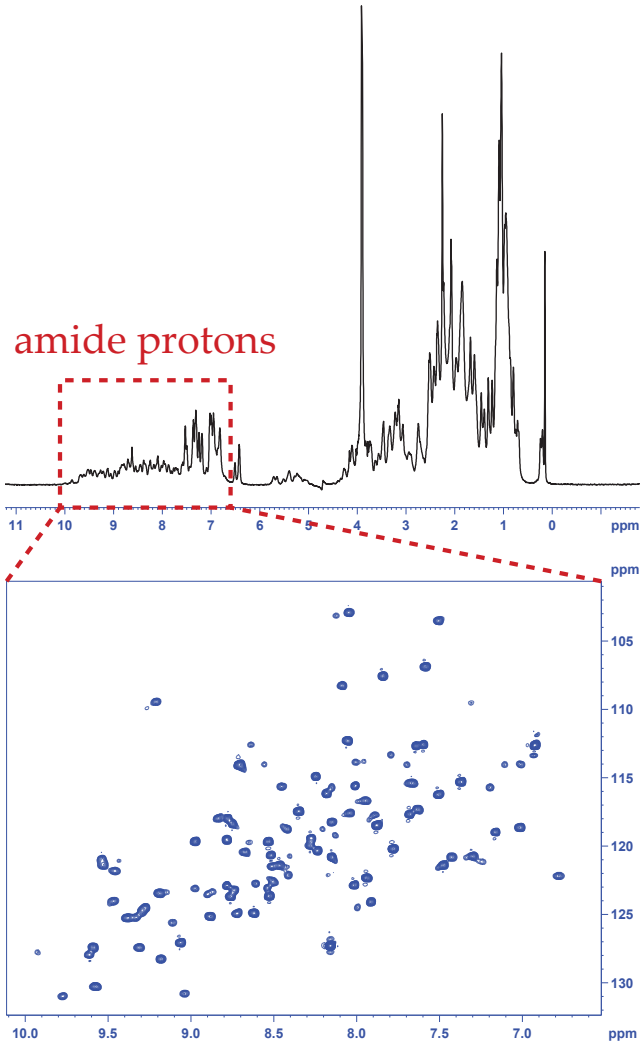


*Figure 2.12 – 1D proton spectrum of the Ribosomal protein S24e from* T. barophilus *(top) and 2D* $^{1}H$-$^{15}N$ *HSQC spectrum in the same conditions.*

### 2.3.3 High Pressure NMR

Multiple perturbing agents have been historically used to probe protein folding, such as temperature, pH or chaotropic agents. Recently, pressure has been employed to the same extent, exploiting its advantages with respect to other methods[63,64,104,105]. Pressure unfolding is almost always a reversible process, and it can be studied while the system is in thermodynamic equilibrium. Moreover, thanks to the technological advances in the field, it is now possible to reliably perform these experiments in a wide pressure range[106]. Another advantage of pressure is that it can shift population equilibria between different protein conformation, uncovering low-lying excited states[104] and giving insight on interesting thermodynamical properties of proteins.

However, there are still limits to the maximum attainable pressure even with state of the art equipment, therefore, the use of chemical denaturants such as guanidinium hydrochloride is still necessary to obtain full unfolding in a narrower pressure range. Nonetheless, the needed quantity of chaotropic agent is still considerably lower than what is needed to fully denature a protein using only the denaturant (1-2 M versus 6-8 M).

In this work, we employed HP-NMR in two different ways. We performed 2D $^{1}$H-$^{15}$N HSQC in both cases, the first set of experiments was carried out with the use of the denaturant, the second without.

Protein unfolding is a slow process (ms to s) compared to the time-scale of the NMR experiment ($\mu$s), therefore it appears as an *all-or-none* transition, where the state of the system is reflected by two sets of spectral features that are in equilibrium. This equilibrium is shifted by the application of pressure, leading to the unfolding of the protein. Thus, the process can be followed by integrating the intensity of the peaks relative to the folded state, which have been assigned to each protein residue. Their intensity will decrease with increasing pressure, and another set of peaks relative to the unfolded state will appear (fig. 2.13).

The first set of experiments thus consisted in a series of HSQC spectra as a function of pressure, as described in Article 2. Thermodynamic quantities, such as the unfolding free energy or the unfolding volume difference, were measured at the residue level to obtain insight about the pressure stability of the two proteins, to probe volume effects, and to pinpoint the amino acid substitutions responsible for pressure adaptation.

The second set of experiments was performed similarly, though without the presence of the chaotropic agent in solution. This implies that the proteins will not undergo unfolding even at the highest pressure probed in our range (1-2500 bar), thus the intensity of the peaks does not vary significantly. What we are interested in are the *changes* in resonance frequencies. Chemical shifts have been thoroughly used as a probe for bond distances[105,107–109], therefore their change with pressure reports on the *compressibility* of those bonds. Acquiring this information at the residue level allowed us to characterize the influence of amino acid substitutions to the flexibility and conformational freedom of the studied proteins.

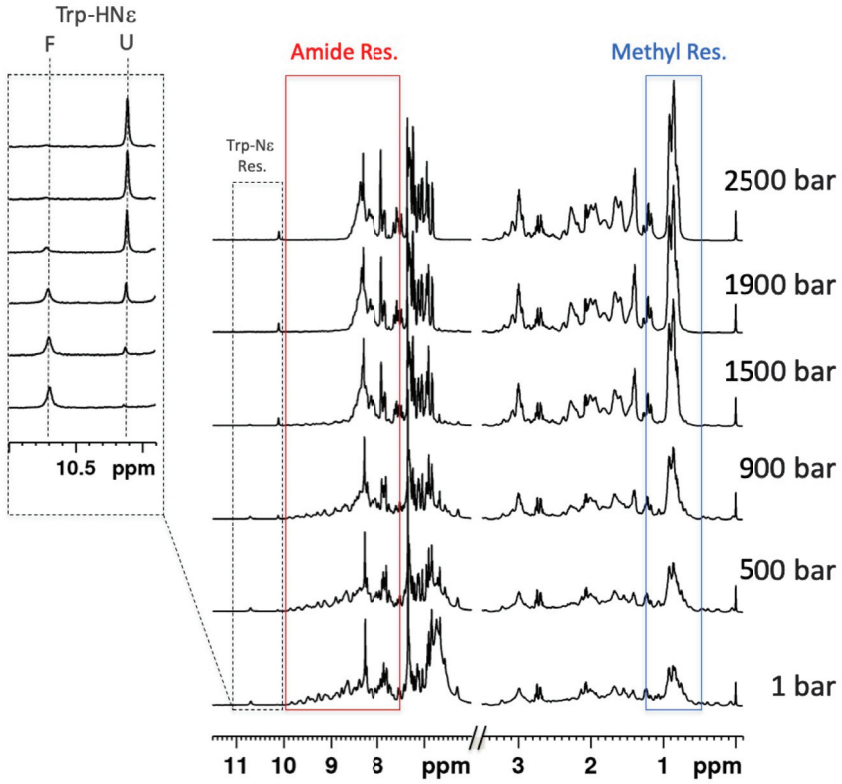


*FIGURE 2.13 – Pressure dependence of the 1D proton spectrum of Titin I27 Ig-like domain in the presence of GuHCl. The all-or-none nature of the transition is clear from the tryptophan resonance (left). Taken from* [106].

## 2.4 X-ray Crystallography

### 2.4.1 Basic principles

Since the first ever protein structure of sperm-whale myoglobin has been solved in 1958 [110], hundreds of thousands have followed. Structural information is of the utmost importance when studying proteins, and its knowledge has led scientists to fundamental discoveries. Nowadays, thanks to the employment of crystallization kits, robots and automated synchrotron beamlines [111], this technique is within reach for a greater number of people.

From the theoretical point of view, the background for X-ray crystallography is similar to that of neutron scattering, described in more detail in section 2.2.1. The main difference between neutron and X-ray scattering is the interaction potential. While neutrons are uncharged and interact only with the nuclei in a sample, X-rays are a form of electromagnetic radiation, and will thus interact with all the charged particles in matter, especially with

electrons. Moreover, looking at X-rays as particles, they are photons, which are bosons with integer spin, therefore their spin will not couple to that of electrons. Thus, there is no incoherent scattering for X-rays in a strict sense.

X-ray crystallography is a diffraction technique, meaning that the scattering is elastic and coherent, and it arises from the ordered structure of the sample. In the case of proteins, the order comes from the periodic arrangement of *many* proteins in the crystal lattice, and not from the molecules themselves. The electron density of the single molecule does not give any diffraction peaks, it *modulates* the intensity of the peaks arising from the crystal lattice of multiple molecules. The result is a diffraction pattern, from which those intensities can be calculated (fig. 2.14). As we have seen in section 2.2.3, the scattering function is the space Fourier transform of the pair correlation function, thus one could take the inverse Fourier transform of the diffraction pattern and obtain the electron density map for the protein.

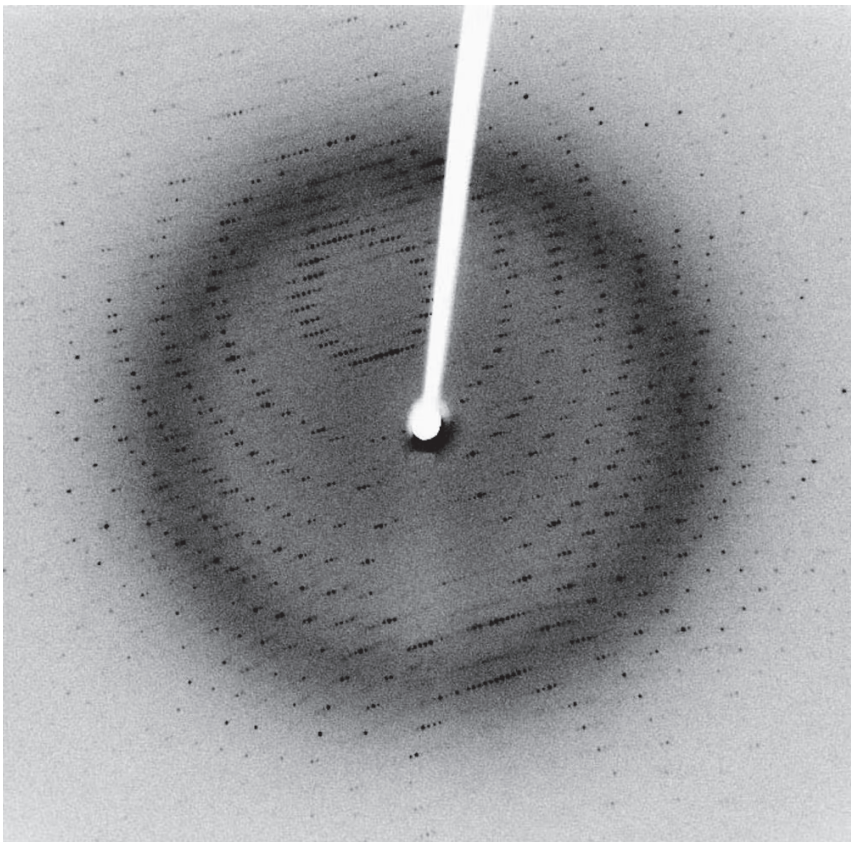

*Figure 2.14 – Typical diffraction pattern from a protein.*

The scattered waves from the proteins will also have their *phase* shifted by the interaction. However, this cannot be measured, as the detectors can only quantify the intensity of the electric field, not its amplitude and phase.

They are both needed in order to calculate the inverse Fourier transform univocally, and while the amplitudes can be computed by taking the square root of the intensities, the information about the phases is lost. This is the origin of the so-called *phase problem*[112].

A common way of illustrating this problem is the well known "Fourier duck" image (fig. 2.15). It presents the images of a duck and a cat, together with their Fourier transforms. Then, an image is constructed by combining the amplitude information of the duck with the phase information from the cat. The inverse Fourier transform of this combination gives back an image that clearly resembles the cat, even though it contains the amplitude information from the duck. This shows the wealth of information that is contained into the phases, and that is irretrievably lost when detecting only the intensities of a diffraction pattern.

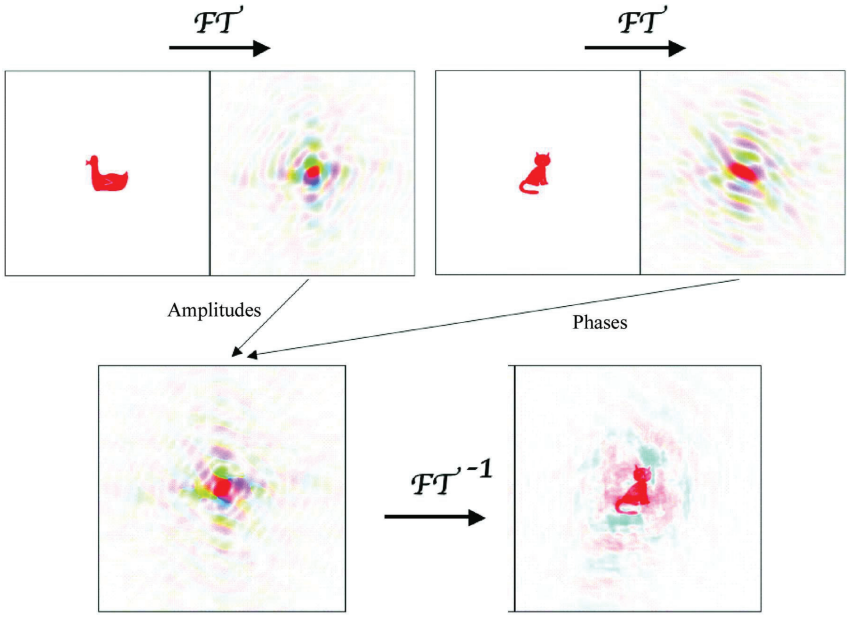


FIGURE 2.15 – *Illustration of the phase problem. Taken from*[112].

Several methods have been developed to retrieve the phases of a crystal structure, and have been thoroughly reviewed throughout the years[112–114]. The one we employed in this work is Molecular Replacement (MR).

### 2.4.2 Molecular Replacement phasing

MR was first described by Rossmann[115,116] and it relies on the prior knowledge of a homologous structure to the unknown one that is being resolved.

The template could be another X-ray structure that has already been solved and shares at least 30% sequence homology with the target, or a *de novo* model constructed by means of homology modelling.

From the starting template, the Patterson map is constructed, which is the Fourier transform of the intensities rather than the amplitudes. It contains information about the interatomic distances, and it can be directly compared to the one constructed from the measured intensities for the unknown structure. This comparison is done by means of a correlation function. The template structure is then rotated around its three degrees of freedom in order to maximize the correlation function. After this, the template is also translated in the three dimensions to further improve the correlation. After each rotation or translation, the Patterson map is recalculated from the template, and correlated again with the target. Figure 2.16 shows a very simplified example of this process.

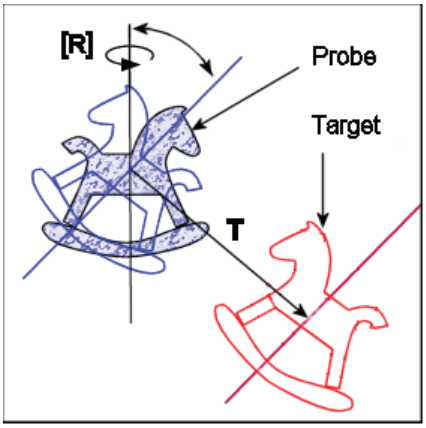


*Figure 2.16 – Principle of the Molecular Replacement method.*

At the end of the rotation-translation process, when the correlation between the two Patterson maps is maximum, the phases are calculated from the correctly orientated template, and they are used to retrieve the electron density map for the unknown structure. From this map, a molecular model for the unknown structure can be built and then refined, giving the final structure for the target protein.

This method has proven to be very effective, and its results are very accurate when a protein with very high sequence homology to the target has already been solved. Moreover, with the recent advances in structure prediction[117], performing MR based on *de novo* models is becoming common place.

## 2.5 Molecular Dynamics Simulations

### 2.5.1 Basic principles

Since the moment computers started to be powerful enough, scientists have been using them to complement their experiments. MD is probably the best example of this. Describing the dynamics of a protein completely by experimental means is still not a reachable objective, and developing an entirely analytical theory to predict it is not possible. The sole alternative is to solve this problem by employing numerical methods[118,119].

MD simulations calculate the coordinates of a system as a function of time, its *trajectory*, by integrating Newton's equation of motion for each atom, giving dynamical information down to very fast time scales, of the order of ps or less. If said system respects the *ergodic* hypothesis, that is, the time average of its evolution is equivalent to an ensemble average over a great number of identical systems, then simulations can also be used to extract average thermodynamical information, provided that the simulation is long enough for the system to explore a meaningful number of representative conformations.

To integrate Newton's equations, the force acting on each atom has to be evaluated. Several energy terms are summed to calculate the force, including *bonded* and *non-bonded* interactions. The first are modelled by harmonic potentials that describe the vibrational motions of 2 or 3-body systems made of covalently bound atoms, while a torsional potential is used for 4-body dihedral angle motions. To reduce the computational cost, bonds involving light atoms such as hydrogen may be fixed to neglect very high frequency vibrations using algorithms such as SHAKE[120] or SETTLE[121].

Non-bonded interactions can be either electrostatic or Van der Waals (VdW). The first are long-range and are evaluated using the Particle Mesh Ewald (PME) method[122], which consist in a splitting of the potential between a short-range and a long-range part. The first converges very quickly and is directly summed, while the second reaches 0 very slowly as $1/r$. Therefore, a Fast Fourier Transform (FFT) is performed, and the sum is evaluated in the *reciprocal space*, saving computational cost. VdW forces are taken into account by using a Lennard-Jones (L-J) potential. Its dependency on the distance is

$\propto r^{-6}$, thus its convergence is more rapid than the electrostatic potential. Therefore, the L-J potential is frequently cut-off with a smooth function at a reasonable distance to save computational cost. All the constants and parameters that define the potentials, both bonded and non-bonded, are contained in the *force field*, which is essentially a table that contains those values, that have been optimized starting from experimental results[123].

To be realistic, a simulation has to contain water, and while some mean-field *implicit solvent* models have been developed, their results are not satisfactory compared to *explicit* water models like TIP3P[124]. This is a 3-site model for water,with partial charges placed on the three atoms, and it also includes L-J parameters to describe dispersive and repulsive forces. Moreover, counterions have to be added to neutralize the system if the studied protein is charged, as long-range electrostatic forces may induce artifacts. Furthermore, to minimize the effect of the water box, *periodic boundary conditions* need to be applied, so that the protein "sees" an image of itself in the neighbouring boxes, effectively simulating a macroscopic protein solution.

Lastly, to make accurate predictions that can be compared with experimental data, simulations have to be performed in the NPT ensemble, meaning that the number of atoms, the temperature, and the pressure have to remain constant throughout the trajectory. Many temperature control algorithms exist, such as velocity rescaling, which reassigns the velocity to each atom at every time step and distributes them according to a Maxwell-Boltzmann distribution at the desired temperature. However, the most commonly used algorithm is the Langevin thermostat[125], which thermally couples the system to an external bath via stochastic equations, so that the temperature is fixed, but it can fluctuate. For pressure control numerous methods exist as well, and one of the most common is the modified Nosé-Hoover with Langevin dynamics[126], in which a piston maintains the pressure on the system, but it is also coupled to the Langevin thermal bath and allows for fluctuations without prior knowledge of the compressibility of the system.

Overall, numerous methods exist to render simulations as accurate as possible, and with the continuous increase in computational power, MD is becoming a very accessible method even to the average user that does not have a supercomputer at hand.

### 2.5.2 Calculation of cavities in protein structures

The importance of cavities for the pressure stability of proteins has been discussed throughout this work. However, crystal structures present two problems that hamper the accurate calculation of cavities. First, they lack hydrogen atoms, which would clearly introduce systematic errors in the calculation, and second, it is not guaranteed that the structure is found in a thermodynamically relevant state.

To tackle this, we performed MD simulations after adding the missing hydrogen atoms, in order to relax the structures in different conditions and then calculate the cavities on them, so that we could compare these results with our experiments and get additional insight.

A multitude of methods and software are available for cavity detection, and a thorough review by Krone and colleagues[127] summarizes all the options based on the underlying method they employ.

In this work, we made use of the CASTp web-server[128]. This method is very robust, as it is based on purely geometric arguments[129,130], and its results are consistent and considerably less dependent from user input, unlike many of the alternative methods. Moreover, this method was the most successful in the identification of solvent-inaccessible internal cavities, which are what we are mainly interested in.

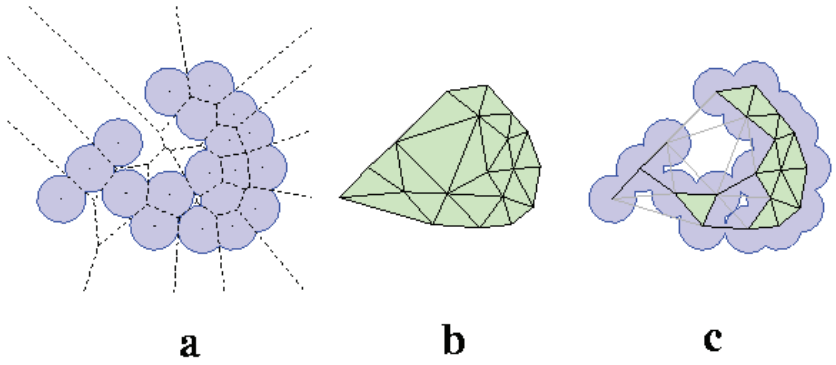


*FIGURE 2.17 – Conceptual illustration of the CASTp method. a) collection of atoms (blue discs) forming a 2D molecule and its Voronoi diagram (dashed lines), b) Delaunay triangulation of the convex hull of the molecule, c) dual complex, or alpha shape, of the initial molecule.*

This method is based on two concepts, namely the *Voronoi diagram* and the *Delaunay triangulation*[129,130]. In the 2D example of fig. 2.17, the disks represent atoms, and the dotted lines in panel a correspond to their Voronoi

diagram. Panel b presents the *convex hull* of the molecule, which is the minimum convex shape that encloses all the atom centers. To visualize it, we can imagine placing pins in the centers of the disks, and then using a rubber band to enclose all of them. By joining the centers of the atoms in each adjacent Voronoi cell, we obtain the Delaunay triangulation, which is enclosed into the convex hull (b). The Delaunay triangulation and the Voronoi diagram are mathematically equivalent and contain the same information. Finally, we remove the Voronoi edges that are not inside the molecule, which is equivalent to removing the edges of the Delaunay triangles that intersect them (c). This is the *dual complex* or *alpha shape* of the original molecule, and it consists of triangles that are not empty, meaning that all their edges are entirely inside the molecule. From here, it is straightforward to identify the cavities and to calculate their surface by summing the areas of the empty triangles and subtracting the area of the disks surrounding them. It is also clear which cavities are accessible to the solvent and which are not.

Generalizing this to the three-dimensional case of an actual protein, this method can provide accurate and reliable calculations about cavities in protein structures, and it is able to distinguish between internal and solvent accessible cavities, allowing one to analyse them differently. Its only major downside is that the source code and the software itself are not freely available, therefore it is not possible to extend it to analyse entire trajectories instead of just static structures.

# RESULTS

## 3.1 Article 1 : Unravelling the mechanisms of pressure adaptation to high pressure in proteins

### 3.1.1 Foreword

The first article, published on the cover of the *International Journal of Molecular Sciences*, concerns the first couple of proteins we studied, the two *Phosphomannose Isomerases* from *T. barophilus* (Tba PMI) and *T. kodakarensis* (Tko PMI). Our goal was to characterise their structural and dynamical properties. In contrast to previous studies[42,43], here we are comparing proteins that are very closely-related, and they differ only in pressure adaptation. We performed two neutron scattering experiments on two different time scales that enabled us to probe he dynamic of hydrogen atoms on the picosecond time scale, and to characterise the geometry of these motions.

Despite the efforts, performing NMR spectroscopy on these samples turned out not to be feasible, owing to the dimeric nature of these proteins. Their relatively large size was reflected in a rather low tumbling time, which is inversely proportional to the width of the NMR peaks, negatively impacting the resolution. Therefore, we employed X-ray crystallography to obtain structural information on the proteins, coupling it to MD simulations in order to relax the structures to a solution-like state before comparing the results to the neutron data.

For this work I prepared the protein samples for neutron scattering with the

assistance of Dr. Stéphane Fontanay, I designed and performed the experiments, and I built the models for data analysis, performing it with custom-made GNU Octave scripts I wrote myself. I prepared the protein samples for X-ray crystallography and assisted the preparation of the crystals performed by Dr. Cécile Dubois at the *Centre de Biochimie Structurale* in Montpellier, with the help of Prof. Christian Roumestand. After the analysis and refinement of the structural data by Dr. François Hoh, I performed the MD simulations and analysed the data and wrote the paper. Dr. Philippe Oger and Prof. Judith Peters conceived and designed the research strategy.

International Journal of
*Molecular Sciences*

MDPI

*Article*

# Unravelling the Adaptation Mechanisms to High Pressure in Proteins

**Antonino Caliò [1], Cécile Dubois [2], Stéphane Fontanay [1], Michael Marek Koza [3], François Hoh [2], Christian Roumestand [2], Philippe Oger [1,*] and Judith Peters [3,4,5,*]**

1 Université de Lyon, UCBL, INSA Lyon, CNRS, MAP UMR 5240, 69621 Villeurbanne, France; antonino.calio@insa-lyon.fr (A.C.); stephane.fontanay@insa-lyon.fr (S.F.)
2 CBS, INSERM U1054, CNRS UMR 5048, Université de Montpellier, 34090 Montpellier, France; cecileduboisl98@gmail.com (C.D.); hoh@cbs.cnrs.fr (F.H.); roume@cbs.cnrs.fr (C.R.)
3 Institut Laue Langevin, 38042 Grenoble, France; koza@ill.fr
4 Université Grenoble Alpes, CNRS, LiPhy, 38400 Grenoble, France
5 Institut Universitaire de France, 75231 Paris, France
* Correspondence: philippe.oger@insa-lyon.fr (P.O.); jpeters@ill.fr (J.P.)

**Abstract:** Life is thought to have appeared in the depth of the sea under high hydrostatic pressure. Nowadays, it is known that the deep biosphere hosts a myriad of life forms thriving under high-pressure conditions. However, the evolutionary mechanisms leading to their adaptation are still not known. Here, we show the molecular bases of these mechanisms through a joint structural and dynamical study of two orthologous proteins. We observed that pressure adaptation involves the decoupling of protein–water dynamics and the elimination of cavities in the protein core. This is achieved by rearranging the charged residues on the protein surface and using bulkier hydrophobic residues in the core. These findings will be the starting point in the search for a complete genomic model explaining high-pressure adaptation.

**Keywords:** high pressure adaptation; protein dynamics; origins of life; neutron scattering



**Citation:** Caliò, A.; Dubois, C.; Fontanay, S.; Koza, M.M.; Hoh, F.; Roumestand, C.; Oger, P.; Peters, J. Unravelling the Adaptation Mechanisms to High Pressure in Proteins. *Int. J. Mol. Sci.* **2022**, *23*, 8469. https://doi.org/10.3390/ijms23158469

Academic Editor: Istvan Simon

Received: 6 July 2022
Accepted: 27 July 2022
Published: 30 July 2022



## 1. Introduction

According to one of the most credited hypotheses on the origins of life, it appeared in the deep sea [1], protected from the deleterious radiation of the young sun, but close to energy sources (e.g., hydrothermal vents) that could sustain relevant chemical reactions [2]. Therefore, it would have appeared under high hydrostatic pressure (HHP) conditions. Modern HHP-adapted organisms (piezophiles) display pressure-dependent physiology; however, deciphering their adaptation to HHP is a very challenging task because it is often concomitant with other environmental adaptations [3]. Indeed, they usually thrive in very cold [4], i.e., the deep ocean, or very hot environments [5], i.e., hydrothermal vents. The first demonstration of proteome structural adaptation in piezophiles originated from comparative whole-cell studies between two nearly isogenic piezophilic and piezosensitive microorganisms, namely *Thermococcus barophilus* and *Thermococcus kodakarensis*, which share identical growth characteristics, except for HHP adaptation. Their proteomes exhibit different dynamical properties, together with a remarkable difference in the response to HHP of the hydration water [6]. In contrast to accepted models, proteins of the piezosensitive microorganism appear less sensitive to increasing HHP, while those of the piezophile are more flexible, and undergo pressure-dependent rearrangements at a pressure value close to the optimum of the organism [6,7]. Above this threshold, the piezophile proteome becomes pressure-insensitive [6–9]. Thus, unexpectedly, the adaptation to HHP in piezophiles seems to imply that the cell's proteome is both more sensitive and more resistant to HHP [6]. These dynamical characteristics can also be preserved by piezophile cells under low-pressure stress through the accumulation of organic osmolytes [10]. Interestingly, a similar insensitivity to HHP has been observed for concentrated protein solutions in the presence of

organic osmolytes [11]. Another peculiarity of piezophiles is the response of their proteome's hydration shell to HHP: its size is reduced and water appears less mobile [6]. It is thus probable that structural adaptation in proteins of piezophiles affects amino acids at the water–protein interface. Therefore, two different processes seem to be responsible for HHP adaptation: (i) a structural (i.e., genomic) adaptation, modifying protein sequences to alter their dynamics, and/or (ii) the modulation of protein–water interaction. To date, all attempts to identify the structural signature of HHP adaptation at the genome level have failed, likely because it only involves amino acids that interact with the hydration water or take part in the formation of internal cavities, which can be greatly destabilized by HHP [12]. If the macroscopic thermodynamics of proteins under pressure is quite well established [12–14], its influence on their microscopical properties and their dynamics is still a debated subject [15–21]. Many different, and often contrasting, contributions govern the structural and dynamical stability of proteins with respect to HHP [14], such as the presence of solvent inaccessible cavities [12], electrostriction [22] and the pressure dependence of the hydrophobic effect [23]. Concerning the fast dynamics, there is evidence that pressure tends to slow it down and inhibit conformational changes that require large amplitude motions [15,21]. To investigate protein HHP adaptation, Elastic Incoherent Neutron Scattering (EINS) and Quasi-Elastic Neutron Scattering (QENS) have been employed to study the dynamics of two orthologous proteins from *T. barophilus* (Tba) and *T. kodakarensis* (Tko), which only differ by their optimal growth pressure. Our data show profound differences both in their dynamics and in their interaction with the surrounding water layer, giving the first hints about the molecular mechanisms involved in HHP adaptation.

## 2. Results

### *2.1. Crystal Structures and Molecular Dynamics Simulations*

Genes TERMP_00744 and TK_0503, coding for a *Phosphomannose isomerase* (PMI), were cloned into the pET-16b over-expression vector [24] for Neutron Scattering experiments, and in the pT7-7 vector [25] for crystallization. The two PMIs have been overexpressed in *E. coli* (BL21 (DE3) pLysS) and purified by heat-treating the cell lysate and by size-exclusion chromatography (Figure S3, see Section 4 for further details). X-ray crystallography was performed to obtain the structures of the proteins (Figures 1 and S5). Both present very similar features: their structure is dominated by $\beta$-sheet contributions with turns and disordered regions connecting them, forming a jelly-roll barrel structure, and both present a dimeric quaternary structure. The active site is located in a pocket inside the barrel (Figure S6) and is conserved between the two proteins (His44, His46, Glu51 and His85). Interestingly, the catalytic metal ion coordinated by these four residues is found to be magnesium in Tba PMI and zinc in Tko PMI; however, the structure's resolution does not allow for making a definitive assignment, as the Van der Waals radius of both ions is smaller than the resolution. The two orthologs differ at 16 positions (Figure S1), and most of the substitutions are located at the protein–water interface and involve mainly polar and charged residues (Figure 1), with the notable exceptions of I35V, located in the hydrophobic core of the protein (Figure 2a,b), and I100V, located at the monomer–monomer interface (Figure 2c,d). The crystal structures have been subsequently relaxed and equilibrated in solution through Molecular Dynamics (MD) simulations, and the internal cavities have been revealed (Figure 2). This already shows the impact that these two seemingly conservative substitutions have on the internal packing of the two proteins. It must be noted that the length of the trajectories may not be sufficient to fully relax the structures, but it is compatible with the time-scale of the neutron scattering experiments and would thus reasonably reproduce the structural and dynamical properties probed by these techniques. Moreover, it must be noted that the cavities are probed on average representative structures; therefore, their fluctuations in time are neglected.

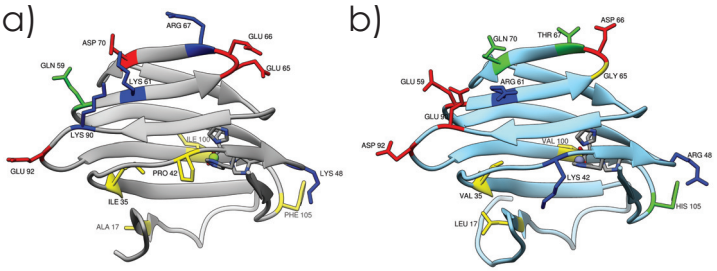


**Figure 1.** Crystal structures of Tba PMI (**a**) and Tko PMI (**b**) with labelled substituted residues for comparison. Residues are color-coded for polarity (red for acidic, blue for basic, green for polar, and yellow for hydrophobic), and the active site is also shown (details in Figure S7).

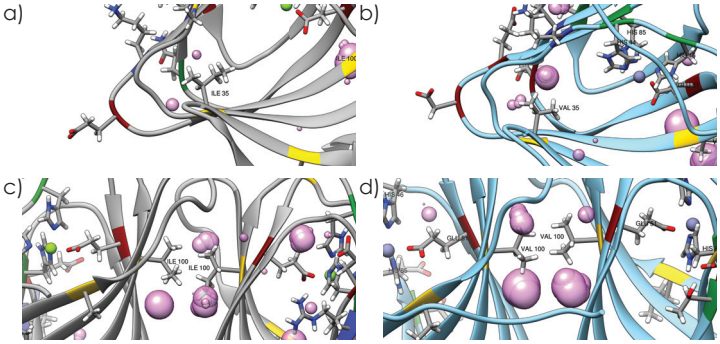


**Figure 2.** Internal cavities in the two proteins located near Ile-Val substitutions in the protein core (**a**,**b**) and at the dimer interface (**c**,**d**) after the relaxation of the crystal structures by MD at 310 K and 1 bar.

### *2.2. Elastic Incoherent Neutron Scattering (EINS)*

EINS was used to access the motions of the two proteins [26,27] in solution, and to probe their response to HHP, while considering their structural differences, to shed light on their adaptation strategies. Incoherent neutron scattering probes the single-particle self-correlation function [28] and, in protein samples, the signal is dominated by hydrogen atoms, thanks to their very large incoherent cross-section [29]. In the case of elastic scattering, there is no energy exchange between the incident neutrons and the sample, meaning that, in the time domain, the correlation function is probed in the long-time limit. Given that the instrument has a finite energy resolution (8 μeV for IN13, see Section 4), this time limit is not at infinity, but it defines the *time window* of the instrument (∼100 ps for IN13). Figure 3 shows the scattering curves for Tba PMI at three representative temperatures. As expected, the scattering intensity shows a general decrease with temperature, consistent with the activation of anharmonic motions [30]. Some contributions from global diffusion of the protein are likely present given the low concentration of the samples. Nonetheless, due to the very close similarity in the primary sequence and the almost identical molecular weight of the two proteins, it is reasonable to assume that these contributions would similarly affect the measured signal for both samples. Hence, the observed differences can be ascribed to the distinct internal dynamics of the two proteins.

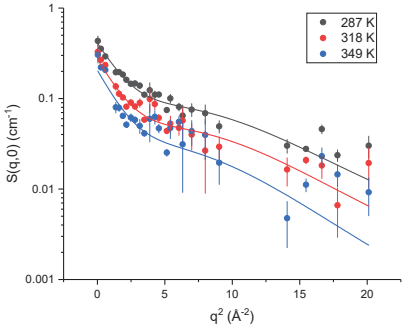


**Figure 3.** Scattering curves for Tba PMI at 150 bar and at some representative temperatures, with the corresponding two-state model fits. The decrease of the elastic intensity with temperature is expected, as more motions enter the time window of the experiment and the involved atoms scatter neutrons inelastically.

Data have been interpreted in the framework of a two-state model [31], which assumes two harmonic potential wells with an associated Mean Square Displacement (MSD) $\Delta x_0^2$, separated by a distance $d$ and a free energy $\Delta G = \Delta H - T\Delta S$ (see Methods). The temperature independence of $d$ (Figure S10), $\Delta H$ and $\Delta S$ (consistently with the assumed Arrhenius behaviour of the two wells' populations) allowed the employment of a global fitting procedure at each pressure point, in which the only temperature-dependent parameter was $\Delta x_0^2$. This granted the minimization of the number of free fitting parameters and greatly improved the quality and stability of the fittings. Figure 4 shows, for both samples, the single-well MSD, $\Delta x_0^2$, and the total MSD, $\Delta x_{tot}^2 = \Delta x_0^2 + \frac{p_1 p_2}{3} d^2$, which takes into account the jump distance between the two wells and their populations. The absolute values of $\Delta x_{tot}^2$, which report on the amplitude of the internal motions, appear very similar for both samples around 350 K, indicating that the proteins display a similar degree of flexibility in proximity of the optimum growth temperature for both organisms. However, the two samples show a clearly different temperature dependence of the MSD, i.e., the slope of the curve which is inversely proportional to the protein's resilience [32,33]: while Tko PMI displays a smooth increase, a change of slope in Tba PMI at around 320 *K* and 1 bar evidences the lower resilience, i.e., higher *softness*, of the protein. This transition is not present at higher pressures, and a linear temperature dependence of the MSD is observed.

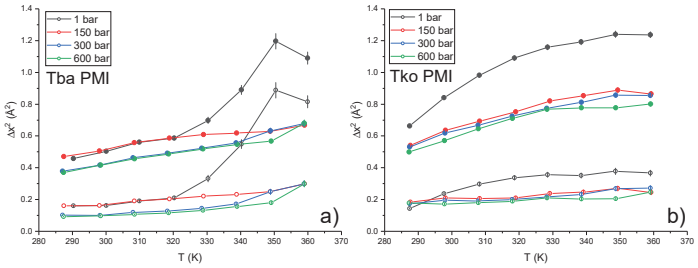


**Figure 4.** Total MSD (full circles) and MSD into the single wells ($\Delta x_0^2$, open circles) for Tba PMI (**a**) and Tko PMI (**b**), lines are a guide to the eye. The absolute value of the MSD reports on the amplitude of hydrogen atoms motion, while the slope of the curve as a function of temperature indicates how much energy is necessary to increase said amplitude, i.e., the *resilience* of the protein [32].

These results are in line with those from whole-cell studies on the same two organisms, which demonstrated the higher flexibility of the proteome of *T. barophilus* [6,7,34] and the existence of pressure-induced structural rearrangements [7,34] in this strain. For *T. barophilus*, the transition occurs smoothly between 1 and 300 bar, close to the optimal growth pressure for the organism, i.e., 400 bar [35]. In *T. kodakarensis*, the transition takes place at a much

lower pressure range as shown by the sharp decrease of the MSD from 1 to 150 bar. Further striking differences between the two proteins can be found by looking at the pressure dependence of the other parameters extracted from the two-state model fitting (Figure 5). In particular, the distance between the wells, $d$ (panel a), is essentially pressure-independent for Tba PMI, while a sizeable decrease is detected for Tko PMI. This behaviour has already been observed in the model protein myoglobin [21] and has been explained in terms of an increased roughness of the protein energy landscape, arising from the difficulty of the protein to explore the conformational substates characterized by bigger volume differences, in agreement with Le Châtelier's principle [36]. In contrast, the energy landscape of Tba PMI appears to be extremely stable with respect to pressure application, resembling what Shrestha et al. [20] found for the *T. thioreducens* IPPase. However, in that case, the comparison with the piezosensitive counterpart was considerably less significant, as hen egg-white lysozyme is not related to the IPPase. The stability of Tba PMI is confirmed by the behaviour of the thermodynamic parameters $\Delta H$ (Figure 5b) and $\Delta S$ (Figure 5c), as they both decrease sharply from 1 to 150 bar and then become pressure-independent, while they both increase for Tko PMI. A decrease in the $\Delta S$ has been connected with decreased hydration [31] and, indeed, such a decrease in the hydration shell size has been detected in the piezophilic proteome [6]. This is consistent with the two proteins having different interactions with water. The exclusion of water from the hydration shell of Tba PMI appears to be the key to the dynamical stability of the protein under high pressure.

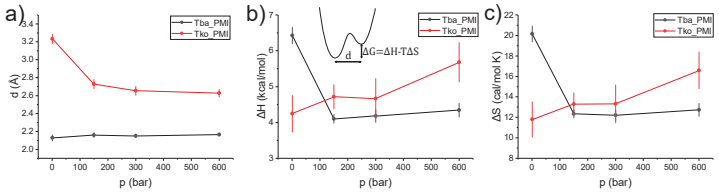


**Figure 5.** Temperature-independent parameters extracted from the two-state model as a function of pressure for Tba PMI (black symbols) and Tko PMI (red symbols): distance between the two wells ($d$, (**a**)), enthalpy ($\Delta H$, (**b**)) and entropy ($\Delta S$, (**c**)) difference, lines are a guide to the eye. (**b**) also contains a pictorial representation of the model showing how the different parameters influence the shape of a proteins' energy landscape.

### *2.3. Quasi-Elastic Neutron Scattering (QENS)*

QENS was used to extract detailed information on the fast dynamics of the protein in solution, probe the effect of HHP, and reveal how specific substitutions in the sequences could affect them, giving insight into the mechanism of pressure adaptation.

As in EINS, the QENS signal is dominated by the incoherent contribution of hydrogen atoms, but in this case, a small energy transfer between the incident neutrons and the sample is allowed: it is thus possible to separate the contributions to the signal arising from different motions, while EINS gives an average picture. The analysis of the spectra gives access to localized and diffusional motions taking place on a specific time scale, which is ∼10 ps for IN5, and to characterize their geometry. Figure 6 shows a fit example, in which the two components we identified are highlighted: a broad and $q$-independent contribution due to fast localized motions, and a narrow contribution arising from confined jump–diffusion of protein residues. The logarithm of the HWHM of the broad component (Figure 7a,b) follows an Arrhenius behaviour ($\Gamma_{loc}(T) = \Gamma_0 \exp(-E_A/RT)$, where $E_A$ is the activation energy, $R$ is the gas constant and $\Gamma_0$ is the pre-exponential constant) at all pressure values for both samples. Tba PMI shows enhanced pressure stability compared to its piezosensitive counterpart concerning fast localized motions. The activation energy values suggest that the rotation of methyl groups is the dominant process from which this contribution arises (values ranging from 1.5 to 3.8 kcal/mol have been reported [37]). Moreover, the value of $E_A$ for methyl rotations has been shown to decrease in efficiently packed hydrophobic environments [37]. Hence, our data indicate that the extent of compression of the protein hydrophobic core is larger in Tko

PMI than in Tba PMI, presumably due to the presence of bigger cavities in the former [12]. Concerning the narrow component, the extracted parameters are the mean jump length $\langle l \rangle$ and the residence time $\tau$, which represents the mean time between two successive jumps.

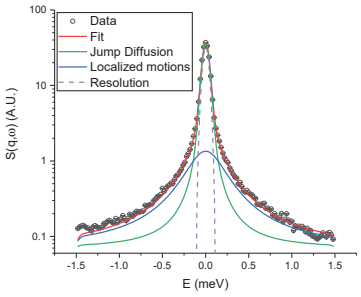


**Figure 6.** Fit example of Tba PMI at 1 bar and 286 K, at a $q$ value of 1.14 Å$^{-1}$. Black circles represent the corrected data, the two Lorentzian contributions owing to localized motions and jump–diffusion are shown respectively as blue and green solid lines, and the total fit is shown as a red solid line. The resolution function is also shown. Error bars are smaller than the symbol size.

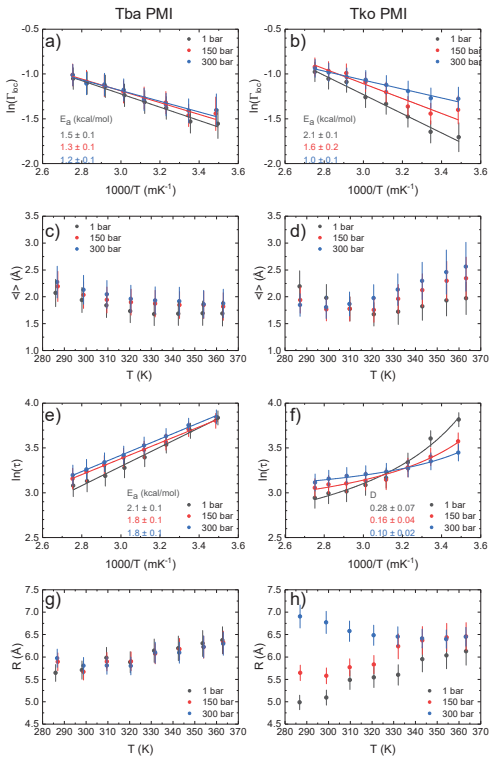


**Figure 7.** Natural logarithm of the broad component HWHM as a function of inverse temperature for Tba PMI (**a**) and Tko PMI (**b**) at all pressure values. Lines are linear fittings assuming an Arrhenius behaviour. Activation energy values are reported on the figure color-coded with the plots. Mean jump length from the Hall–Ross model as a function of temperature and pressure for Tba PMI (**c**) and Tko PMI (**d**). Natural logarithm of the residence time as a function of inverse temperature at all pressure values. Lines are fits to the Arrhenius law ((**e**), Tba PMI) or to the Vogel–Fulcher–Tamman law ((**f**), Tko PMI). Values for the activation energy (Tba PMI) and the fragility index (Tko PMI) are reported on the figure color-coded with the plots. Values of confinement radius $R$ for Tba PMI (**g**) and Tko PMI (**h**).

The temperature dependence of $\langle l \rangle$ (Figure 7c,d) for Tba PMI is rather weak and does not change with pressure, testifying to the structural stability of the protein in the whole temperature and pressure range studied. For Tko PMI, this quantity shows a similar behaviour at 1 bar, where the protein is expected to be functional, while higher pressures seem to have a destabilizing effect. While it would appear straightforward to compare this quantity to the distance between the wells $d$ derived from EINS data (Figure 5a), it must be stressed that the latter results from *all* the internal motions that are activated on the 100 *ps* time scale, and thus gives an average representation of the protein's energy landscape, while $\langle l \rangle$ refers to a particular motion, namely the jump–diffusion of side chains, and it relates to a different time scale. Furthermore, for the sake of comparison with other works, a *pseudo*-diffusion coefficient related to internal dynamics can be calculated according to $D_{pseudo} = \langle l \rangle^2 / 2\tau$ [38] (Figure S15). The difference in the temperature dependence of $\tau$ (Figure 7e,f) for the two proteins is remarkable: it follows the Arrhenius law ($\tau(T) = \tau_0 \exp(E_A/RT)$, note the sign reversal compared to before, as $\tau = \hbar/\Gamma$) for Tba PMI, while it follows the Vogel–Fulcher–Tamman (VFT) law ($\tau(T) = \tau_0 \exp\left(\frac{DT_0}{T-T_0}\right)$, where $1/D$ is the fragility index, not to be confused with the aforementioned pseudo-diffusion coefficient $D_{pseudo}$, and $T_0$ is the Vogel temperature) for Tko PMI. The latter is typical of glass-forming systems [39–41], but it has also been observed in proteins and interpreted as the signature of protein–water coupled dynamics [42–45]. An Arrhenius behaviour arises from activated processes, and it is expected for jump–diffusion, while a VFT behaviour is usually connected with cooperative processes. The appearance of VFT behaviour on such a fast time-scale is intriguing, and it shows that side-chain relaxations are strongly coupled to hydration water dynamics in Tko PMI. This correlates with the higher $\Delta S$ in the EINS data and suggests that the decoupling of protein dynamics from its environment could be the key to pressure adaptation for Tba PMI. To push the analogy further, materials with a low fragility index exhibit a linear behaviour far from the Vogel Temperature $T_0$ which can be reasonably fitted with the Arrhenius law, and are referred to as *strong* glass-forming materials, while *fragile* materials show VFT behaviour in a considerably wider temperature range [46]. Thus, the difference in dynamical properties between Tba PMI and Tko PMI could be assimilated to that between strong and fragile glass-forming materials. It appears that Tko PMI's dynamics are dominated by cooperative motions and that high pressure can destabilize them, as highlighted by the increase in fragility (i.e., decreasing $D$) with increasing pressure, while Tba PMI's dynamics are dominated by pressure-insensitive activated processes. This difference is likely arising from the distinct amino-acidic composition of the two proteins, and it could explain the superior pressure stability of Tba PMI. The *Elastic Incoherent Structure Factor* (EISF) gives information about the geometry of the motions inside the resolution of the instrument, and has been analysed as described in the Methods section (Figures S16 and S17). The behaviour of the confinement radius $R$ extracted from the EISF (Figure 7g,h) highlights another remarkable difference between the two proteins. Tba PMI shows a weak temperature dependence of this parameter, compatible with thermal expansion, and again shows no pressure dependence. On the other hand, the temperature dependence of Tko PMI appears to be stronger, and a sizeable increase in the confinement radius is detected with increasing pressure. This result could appear counter-intuitive, but it can be rationalized by thinking of $R$ as an average measure of the protein's solvent-accessible cavities: higher pressure forces water into them, increasing their volume, while concomitantly decreasing the protein's specific volume, in agreement with Le Châtelier's principle. This volume increase in Tko PMI could also explain the enhanced mean jump length seen at high pressure (Figure 7d) and the stronger coupling with water displayed by the protein side-chains (Figures 5c and 7f). Such assignment is substantiated by the results of MD simulations: an equivalent radius (i.e., the radius of a sphere with the same volume) has been calculated from the volume of the ligand pocket of both proteins simulated in different conditions (Figure S9). At low pressure and temperature, the found value is remarkably close to what is found by QENS, as the ligand pocket is the dominant solvent-accessible cavity in both proteins. However, the values

deviate at high temperature and pressure, as QENS gives an average measure of *all* the cavities in the protein, and more cavities become of comparable size to the ligand pocket under extreme conditions. Nevertheless, MD values follow a very similar trend to the QENS values, and this can also be seen visually in Figure 8.

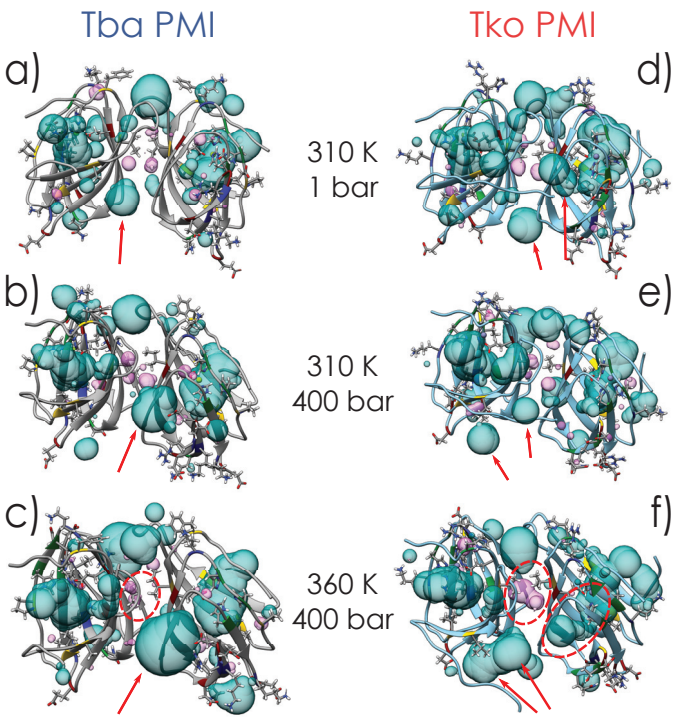


**Figure 8.** Representation of internal (magenta) and solvent-accessible (cyan) cavities for Tba PMI (**a**–**c**) and Tko PMI (**d**–**f**) at different T and P conditions. Structures are represented in cartoons, and substituted residues are evidenced in sticks (with the corresponding part of the ribbon coloured for residue type, blue for basic, red for acidic, green for polar and yellow for hydrophobic). Red arrows follow the evolution of some cavities, and the opposite behaviour of the internal cavities in the dimer interface of the two proteins is highlighted by red circles. Another red circle in (**f**) also shows how the ligand pocket in one monomer of Tko PMI has actually become a channel from side to side under extreme conditions.

## 3. Discussion

This study aims to compare proteins under positive selection originating from nearly isogenic organisms, differing only in their adaptation to pressure. Hence, for the first time, the different dynamical properties exhibited by the two proteins, as revealed by Neutron Scattering, and their structural properties, elucidated by X-ray crystallography and Molecular Dynamics simulations, can be correlated to decipher the adaptation to HHP regardless of other adaptative traits. The results show that the mechanisms by which Tba PMI counteracts the effect of HHP are: (i) the prevention of excessive amounts of water from penetrating the solvent-accessible cavities and limiting their destabilization, (ii) the inhibition of protein–water cooperative relaxations, and (iii) the reduction of internal (i.e., inaccessible to the solvent) cavity volume. This becomes of particular importance when considering the ligand pocket of the protein, i.e., the largest solvent-accessible cavity: its structure must be preserved under HHP to properly carry out the enzymatic reaction (Figure 9c) and to play its role in the metabolism of the organism.

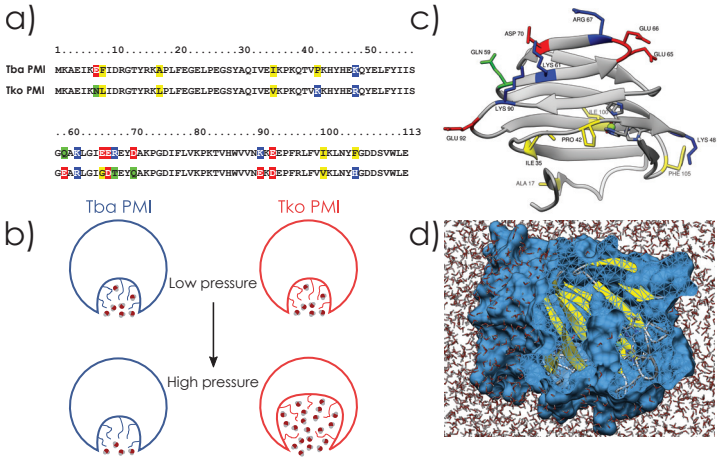


**Figure 9.** (**a**) sequence alignment of the two proteins, substitutions are highlighted (red for acidic, blue for basic, green for polar and yellow for hydrophobic residues). (**b**) schematic representation of the two proteins and effect of high pressure on them. (**c**) cartoon representation of Tba PMI, substitutions are represented in sticks with the same color-code as (**b**). (**d**) vertical cut on the surface representation of Tba PMI, highlighting the ligand pocket.

Because of these results, it is possible to characterize the impact of each substitution in the two protein sequences by comparing their structures and dynamics. The substitutions will be identified with the first letter being the residue present in Tba PMI, and the second letter for that in Tko PMI. The two proteins differ at 16 positions and, as shown in the 3D representation of the protein (Figure 9a,b), all of the substituted residues present their side-chain exposed to the solvent, except for I35V and I100V. It is known that even very small volume changes in the interior of a protein (such as a single point mutation) can greatly affect its pressure stability, especially when this residue is located close to an internal cavity and the substitution is affecting its volume [12]. To this extent, the I35V substitution appears to be key in the stabilization of the hydrophobic core (Figure 2a,b), with the added importance of being close to the ligand pocket, limiting its deformation under extreme conditions (Figure 8). The active site is very conserved except for K48R, which is not affecting charge or volume, and P42K, where the proline residue might soften the rigid $\beta$-strand structure and render it more resilient to pressure changes in Tba PMI. In the dimer-forming region, the I100V substitution optimizes the contact between the two monomers and decreases internal cavity volume (Figure 2c,d), stabilizing Tba PMI's quaternary structure under extreme conditions and seemingly maintaining optimal functionality (Figure 8). The remaining substitutions (E7N, Q59E, K61R, E65G, E66D, R67T, D70Q, K90E, E92D, F105H) are located at the protein–water interface and are likely involved in the modulation of the protein's surface charge distribution, affecting its coupling with hydration water. This is visible when calculating the electrostatic potential on the surface of both proteins (Figure S8): Tba PMI shows a higher positive potential than its piezosensitive counterpart around the entrance of the ligand pocket, a higher negative potential in the middle region, and less uncharged regions in general. This excess charge could in turn increase the orientational constraints on the water molecules in the first hydration shell, effectively shielding the protein. Furthermore, the charge reorganization on Tba PMI's surface promotes the formation of salt bridges (e.g., Lys61 with Asp70), applying different constraints on the protein structure and optimizing its response to high pressure, while also hindering the interaction that those residues could have with water molecules. Two more substitutions are found in F8L and A17L; however, they do not appear to be involved in any particular mechanism, and will thus be considered random conservative mutations.

The overall effect of these substitutions on Tba PMI is to efficiently maintain the protein structure, especially around the ligand pocket. On the contrary, at 360 K and 400 bar, Tko PMI's ligand pocket substantially becomes a channel (Figure 8f) spanning from one side to the other of the monomer, and we speculate that protein function would be at best severely hindered in these conditions. Hence, the amino-acid substitutions between the two PMIs allow for drawing a first adaptation pattern related to HHP. This is characterized by a decrease in polar residues in favour of charged ones, particularly glutamate and lysine, in the piezophilic protein. Moreover, precise substitutions involving isoleucine instead of valine enable the piezophilic protein to tailor its occupation of void volume both in its core and at the dimerization interface, enhancing its pressure stability. However, the relative contributions of the different amino-acid substitutions identified in this work need to be confirmed by further studies using direct methods (multidimensional NMR spectroscopy, site-directed mutagenesis), and eventually other piezophilic proteins.

## 4. Materials and Methods

### *4.1. Protein Expression and Purification*

Recombinant Phosphomannose Isomerases from *T. barophilus* and *T. kodakarensis* have been produced by cloning synthetic codon-optimized genes (purchased from GENEWIZ Europe) into the protein over-expression plasmid pET-16b [24] (Novagen), which was then transformed into *E. coli* BL21(DE3) pLysS strain (Novagen). Ten liter cultures were grown at 37 °C in LB medium supplemented with 100 µg/mL ampicillin until $OD_{600} = 0.5$, induced with a final concentration of 1 mM IPTG and further grown overnight at 25 °C. Cells were harvested by centrifugation at 17.000× *g* for 30 min, washed in isotonic solution (0.9% NaCl) and resuspended in 400 mL of 50 mM $NaH_2PO_4$, 300 mM NaCl, pH 8 buffer. Cells were then lysed by five freeze-thaw cycles in liquid nitrogen (1 min) and at 50 °C (3 min), and homogenized by sonication (maximum power for 15 min at 50% duty cycle). The soluble fraction was recovered by centrifugation at 12,000× *g* and 4 °C for 60 min. It was then heated to 75 °C for 1 h to remove the non-thermostable proteins from the *E. coli* expression host. Protein debris were removed by centrifugation at 12,000× *g* and 4 °C for 60 min. The extraction was repeated a second time for maximum recovery. The supernatant was concentrated to ∼20 mL by ammonium sulfate precipitation and further purified by Size Exclusion Chromatography on an AKTA® FPLC system, using an XK50-60 column packed with 1 L of Superdex® 75 Prep-Grade resin, calibrated with the GE Healthcare® Low Molecular Weight kit (Figure S2). During this step, both proteins evidenced a dimeric quaternary structure, as they eluted at double the expected MW (Figure S3). Fractions containing the protein were then pooled, concentrated by ultrafiltration (Amicon® Ultra-15 centrifugal filter units, Millipore, Burlington, MA, USA) and lyophilised. The purity of the proteins was assessed by SDS-PAGE, and was greater than 99% (Figure S4). To prepare the samples, the lyophilised protein powder was gently dissolved in $D_2O$ (Sigma-Aldrich, Saint Louis, MO, USA) under nitrogen atmosphere, at a concentration of 120 mg/mL. Protein solutions rather than hydrated powders were employed in order to optimally transmit hydrostatic pressure to the sample. Proteins employed in the two different experiments belonged to the same production batch. Proteins for crystallization were obtained by employing the same protocol with two modifications: the expression plasmid was pT7-7 [25], which also exploits the T7 expression system but lacks the His-Tag sequence, and the culture volume was scaled down to 1.5 L.

### *4.2. X-ray Crystallography*

The purified proteins were diluted to 10 mg/mL in 10 mM Tris-HCl pH 8, 300 mM NaCl. Crystallization trials were performed at 20 °C using the hanging-drop vapour-diffusion method in 96-well micro-plates and a Mosquito HTS robot (TPLabtech) with 100 nL of protein mixed with 100 nL of reservoir. Tko PMI crystals were obtained after one week from condition E8 of the Structure screen 1 + 2 kit (Molecular Dimensions), containing 0.2 M ammonium phosphate monobasic, 0.1 M Tris pH 8.5, and 50% *v/v* MPD.

Tba PMI crystals were obtained after two weeks from condition F7 of the Morpheus kit (Molecular Dimensions), containing a 0.12 M monosaccharides mix, 0.1 M buffer system 2 pH 7.5, 30% *v/v* precipitant mix 3. Crystals were flash-frozen in liquid nitrogen. X-ray diffraction datasets were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble) at the ID30A-1 beam line (Massif1 [47,48]) using a pixel detector (PILATUS3 2M) and auto-processed by the XDSAPP package [49]. Tba PMI crystals belong to the $I_{222}$ space group and contain one molecule in the asymmetric unit, while Tko PMI crystals belong to the $P_{31}$ space group and contain eight molecules in the asymmetric unit. The structures of both proteins were determined by molecular replacement with Phaser [50] from the Phenix package [51], using models from the Alphafold2 server [52]. After model building using Coot [53] and Refine [54] from the Phenix package (Tba PMI) or REFMAC5 [55] (Tko PMI), the final structures exhibited an $R(\%)/R(\%)_{free}$ of 0.18/0.20 at 1.7 Å (Tba PMI) and 0.23/0.29 at 2.2 Å (Tko PMI). Final refinement statistics for the structures are listed in Tables S1 and S2. The atomic coordinates and structure factors of Tba PMI and Tko PMI have been deposited in the Protein Data Bank with accession numbers 7ZVM and 7ZVY, respectively.

### *4.3. Elastic Incoherent Neutron Scattering (EINS)*

EINS measurements were performed on the IN13 backscattering spectrometer at the Institut Laue-Langevin (ILL, Grenoble, France). At the elastic position, IN13 has an incident wavelength of 2.23 Å and a nearly $q$-independent resolution of 8 μeV FWHM, which gives a time window of ∼100 ps [56], allowing for probing local motions of hydrogen atoms since their incoherent scattering cross section is an order of magnitude larger than that of other isotopes [29]. Temperature was controlled by means of a closed-cycle dry cryofurnace (Displex+), and continuous up-scans were performed in the 283 K to 363 K range at 0.08 K/min. The scattering intensity was also measured while the temperature was lowered back to 283 K before the next pressure point to check for hysteresis and, once its absence was verified, the downscans were merged with the upscans to improve statistics. HHP was transmitted to the sample by means of the high-pressure stick, cell and controller developed by the SANE team at ILL [57], and four pressure points were investigated (1, 150, 300 and 600 bar). The high-pressure cell is cylindrical and made of a high-tensile aluminium alloy (7026) and has a 6 mm internal diameter [58]. A piston separates the pressure-transmitting liquid (Fluorinert™ FC-770 [59]) from the sample, and a cylindrical aluminium insert (4 mm diameter) was used to decrease sample volume and to minimize multiple scattering. Raw data were corrected for transmission, empty cell and $D_2O$ scattering, normalized to a vanadium standard and then binned in temperature in 10 K intervals using the LAMP [60] software available at ILL. EINS data have been interpreted in the framework of the *two-state model* [31], which models hydrogen atoms' motions as a combination of vibrations in two harmonic potential wells, which give the Debye–Waller contribution with the associated Mean Square Displacement (MSD) $\Delta x_0^2$, and jumps between them. The wells are separated by a distance $d$ and have a free-energy difference $\Delta G$ that can be separated into the enthalpic and enthropic contributions according to $\Delta G = \Delta H - T\Delta S$. The elastic scattering function $S(q, \omega = 0)$ as a function of the scattering vector $q$ (related to the scattering angle $\theta$ and the neutron's wavelength $\lambda$ according to $q = \frac{4\pi}{\lambda} sin(\frac{\theta}{2})$) thus reads:

$$S(q,0) = e^{-\Delta x_0^2 q^2}\left[1 - 2p_1p_2\left(1 - \frac{\sin(qd)}{qd}\right)\right], \quad (1)$$

where $p_1$ and $p_2$ represent the population of each well, assumed in our case to follow the Arrhenius law ($p_1/p_2 = exp(-\Delta H/RT + \Delta S/R)$, where $R$ is the gas constant). It must be stressed, however, that, in the investigated temperature range, large-scale motions could enter the experimental window. It is thus desirable to view the two wells as an average representation of the protein's free-energy landscape that is accessible at each temperature and pressure value.

*4.4. Quasi-Elastic Neutron Scattering (QENS)*

QENS measurements were carried out on the IN5 time-of-flight (TOF) spectrometer [61] at ILL at 5 Å incident wavelength. In this configuration, the energy resolution was ∼70 µeV HWHM, giving a time window of ∼10 ps, suitable for investigating fast localized protein motions. Temperature was controlled with the standard ILL Orange Cryofurnace in the same range as the EINS experiment, and continuous scans at 0.4 K/min were acquired. The same HHP equipment was used for pressure transmission at 1, 150 and 300 bar. The 600 bar point could not be measured because of time constraints. The same corrections as in the EINS data treatment were applied (see Supplementary Materials ) and, after temperature binning, TOF data were further corrected for detector efficiency and detailed balance [28], then converted to $S(q,\omega)$ (where $\hbar\omega$ is the energy that a neutron exchanges with the sample) and re-binned in 20 spectra with evenly spaced (0.02 meV) energy points at $q$ values from 0.07 to 2.57 Å$^{-1}$. Only spectra having a sufficient dynamic range (−1.5 to +1.5 meV) were considered in the analysis, giving a final $q$ range of 0.6–1.8 Å$^{-1}$. The whole treatment was performed with LAMP [60]. First, a model-free analysis of the corrected data was performed. This consists of fitting a sum of Lorentzian functions [62] and leaving their parameters free in order to identify the different dynamical contributions to the measured signal, and then analysing the $q$ dependence of their HWHM to define a suitable model that properly fits the data (see Supplementary Materials). Two main contributions have been identified in our case (adding a third Lorentzian did not improve the quality of the fit): the broad component displayed a substantially $q$-independent width, thus representing fast localized motions (e.g., methyl group rotations [30]), while the narrow component's width exhibited a saturation behaviour at high $q$, characteristic of jump–diffusion processes of protein side-chains [63]. Among the different models that have been tested [38,64,65], the Hall and Ross model [38] gave the most satisfactory results (Figure S14). Therefore, the model function has been built by considering an elastic fraction (represented by the *Elastic Incoherent Structure Factor*, or EISF, $A_0(q)$) plus a q-independent Lorentzian, representing the localized motions [28] ($\Gamma_{loc}$), and then convoluted by another Lorentzian, which represents the jump–diffusion process in the Hall–Ross model [38]. This component is characterized by its $q$-dependent HWHM ($\Gamma_j(q)$), which depends on the time between two successive jumps ($\tau$, also named *residence time*) and the average length by which hydrogen atoms jump ($\langle l \rangle$). The theoretical scattering function $S(q,\omega)$ thus reads:

$$S(q,\omega) = \frac{A_0(q)}{\pi}\frac{\Gamma_j(q)}{\Gamma_j(q)^2+\omega^2} + \frac{1-A_0(q)}{\pi^2}\frac{\Gamma_j(q)+\Gamma_{loc}}{(\Gamma_j(q)+\Gamma_{loc})^2+\omega^2} \quad (2)$$

with

$$\Gamma_j(q) = \frac{\hbar}{\tau}\left(1-\exp\left(-\frac{q^2\langle l\rangle^2}{2}\right)\right). \quad (3)$$

The model function is then convoluted with the resolution function (derived from a measurement of vanadium, as it is a dominant elastic incoherent scatterer, as shown in Supplementary Materials), multiplied by a $q$-dependent scale factor proportional to the Debye–Waller factor [28], and then fitted to the data using a global fitting approach (i.e., by fitting the whole $S(q,\omega)$ at once instead of fitting single spectra at different $q$ values separately), which gives the parameters $\Gamma_{loc}$, $\tau$ and $\langle l \rangle$. In order to minimize the number of free parameters and to avoid ambiguities in the global fitting procedure, $A_0(q)$ has been calculated by integrating the spectra in the elastic region, and dividing this value by the total integral of the spectra, following its definition [28]. The calculated $A_0(q)$ have then been used as fixed parameters in the global fitting, permitting to fit the whole $S(q,\omega)$ with only three free parameters ($\Gamma_{loc}$, $\tau$ and $\langle l \rangle$) and giving solid and consistent results. Global diffusion of the protein was not taken into account as the broadening arising from it would be lower than the resolution of the instrument in this configuration (see Supplementary Materials). To complete the picture, the geometry of these motions has been characterised by analysing the EISF. It has been modelled taking into account methyl rotation ($A_{3-j}$

with $a_M = \sqrt{3}R_M = 1.715$ Å [63]) and restricted jump–diffusion of protein residues ($A_j$, from the Hall–Ross model [38]) according to:

$$A_0(q) = p + (1-p)\left[sA_j(q) + (1-s)A_{3-j}(q)\right]$$
$$\text{with}$$
$$A_j(q) = j_0^2\left(\frac{qR}{2}\right) \quad (4)$$
$$A_{3-j} = \frac{1}{3}[1 + 2j_0(qa_M)]$$

where $p$ represents the fraction of immobile H atoms (i.e., slower than the time-scale of the experiment), $s$ is the fraction of H atoms experiencing confinement during their jump–diffusion motion, $j_0$ is the zeroth-order Bessel function of the first kind and $R$ is the confinement radius.

*4.5. Molecular Dynamics Simulations*

Molecular Dynamics (MD) simulations have been performed with NAMD 2.14 [66] on the P2CHPD computing centre of the Université Claude Bernard Lyon 1. The system was prepared using the CHARMM36m force field on the CHARMM-GUI server [67] by building a cubic water box and leaving a 15 Å padding on each side of the protein. The system was neutralized with 150 mM NaCl, resulting in a total content of 19,407 $H_2O$ molecules, 55 $Na^+$ ions and 59 $Cl^-$ ions for Tba PMI, and 17,923 $H_2O$ molecules, 51 $Na^+$ ions and 53 $Cl^-$ ions for Tko PMI. The system was first minimized for 20,000 steps, then gradually heated to the desired temperature (310 K or 360 K) by reassigning the velocities every step for 31,000 or 36,000 steps (0.01 K/step) in the NVE ensemble, then equilibrated in the NVT ensemble for 300 ps with a Langevin thermostat (2 $ps^{-1}$ damping coefficient) and a 2 fs time step, and finally left to evolve in the NPT ensemble at the desired pressure (1 or 400 bar) for 10 ns (2 fs time step) by employing the modified Nosé–Hoover Langevin barostat implemented in NAMD [68], with a 200 fs period and a decay time of 100 fs. Periodic boundary conditions were applied, with particle-mesh Ewald long range electrostatics, using a grid spacing of 1 Å along with a sixth order B-spline charge interpolation scheme. A 12 Å cut-off was used for non-bonded interactions, with a smooth switching function starting at 10 Å. Bonds were constrained using the SETTLE algorithm, and coordinates were output every 10 ps. Analysis of the trajectories was carried out on VMD [69] and, after assessing the stability of the system (Figure S7), the last 5 ns were used to calculate the average coordinates of each atom. This clearly does not correspond to a physical state of the system; therefore, the whole trajectory was aligned with the average structure and the frame with the lowest RMSD (calculated using the $C_\alpha$ coordinates) was chosen as the most representative physical state of the system. The same process was repeated for every temperature and pressure condition, and the resulting average structures were used for the calculation of cavities using the CastP server [70]. All the images have been generated using UCSF Chimera 1.16 [71].

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms23158469/s1. References [72–80] are cited in the supplementary materials.

**Author Contributions:** A.C., P.O. and J.P. conceived and designed the experiments; A.C. performed the production and purification of recombinant proteins with the assistance of S.F.; C.D. prepared the crystals; F.H. and C.R. analysed the diffraction data and refined the crystal structures; A.C., J.P. and M.M.K. performed the neutron scattering experiments; A.C. analysed the neutron scattering data and performed MD simulations. A.C. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the French National Research Agency (programme ANR 17-CE11-0012-01 to P.O. and J.P.) and by the Mission pour les Initiatives Transverses et Interdisciplinaires

of the CNRS (OriginsUnderPressure and LifeAdapt research projects). A.C. is supported by a PhD grant for international students by the French Ministry of Science and Technology.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are available at http://doi.org/10.5291/ILL-DATA.8-04-876 for the IN13 experiment (8-04-876, from 22 August 2023 or upon reasonable request), and at http://dx.doi.org/10.5291/ILL-DATA.8-05-458 for the IN5 experiment (8-05-458, from 21 September 2023 or upon reasonable request). The X-ray structures of Tba PMI and Tko PMI have been deposited to the Protein Data Bank and are available in the under accession codes 7ZVM and 7ZVY, respectively.

**Acknowledgments:** The authors would like to acknowledge the Institut Laue-Langevin for the allocation of beamtime to perform these experiments, James Maurice and the whole High-Pressure Division of the SANE group for technical support during the experiments, and Miguel Angel Gonzalez for his help during the re-writing of data reduction routines for LAMP. We are grateful to Didier Nurizzo and Matthew Bowler at the European Synchrotron Radiation Facility (ESRF) for providing assistance in using beamline ID30A-1. The authors gratefully acknowledge the ESRF for provision of synchrotron radiation facilities via Block Allocation Group beamtime. We are indebted to Grazia Cottone for her invaluable advice and fruitful discussions on the MD simulations.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Miller, S.L.; Bada, J.L. Submarine hot springs and the origin of life. *Nature* **1988**, *334*, 609–611. [CrossRef] [PubMed]
2. Martin, W.; Baross, J.; Kelley, D.; Russell, M.J. Hydrothermal vents and the origin of life. *Nat. Rev. Microbiol.* **2008**, *6*, 805–814. [CrossRef]
3. Brininger, C.; Spradlin, S.; Cobani, L.; Evilia, C. The more adaptive to change, the more likely you are to survive: Protein adaptation in extremophiles. *Semin. Cell Dev. Biol.* **2018**, *84*, 158–169. [CrossRef] [PubMed]
4. Price, P.B. A habitat for psychrophiles in deep Antarctic ice. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 1247–1251. [CrossRef] [PubMed]
5. Auerbach, G.; Ostendorp, R.; Prade, L.; Korndörfer, I.; Dams, T.; Huber, R.; Jaenicke, R. Lactate dehydrogenase from the hyperthermophilic bacterium Thermotoga maritima: The crystal structure at 2.1 å resolution reveals strategies for intrinsic protein stabilization. *Structure* **1998**, *6*, 769–781. [CrossRef]
6. Martinez, N.; Michoud, G.; Cario, A.; Olliver, J.; Franzetti, B.; Jebbar, M.; Oger, P.; Peters, J. High protein flexibility and reduced hydration water dynamics are key pressure adaptive strategies in prokaryotes. *Sci. Rep.* **2016**, *6*, 32816. [CrossRef] [PubMed]
7. Peters, J.; Martinez, N.; Michoud, G.; Cario, A.; Franzetti, B.; Oger, P.; Jebbar, M. Deep Sea Microbes Probed by Incoherent Neutron Scattering Under High Hydrostatic Pressure. *Z. Phys. Chem.* **2014**, *228*, 1121–1133. [CrossRef]
8. Golub, M.; Martinez, N.; Michoud, G.; Ollivier, J.; Jebbar, M.; Oger, P.; Peters, J. The Effect of Crowding on Protein Stability, Rigidity, and High Pressure Sensitivity in Whole Cells. *Langmuir* **2018**, *34*, 10419–10425. [CrossRef] [PubMed]
9. Salvador-Castell, M.; Golub, M.; Martinez, N.; Ollivier, J.; Peters, J.; Oger, P. The first study on the impact of osmolytes in whole cells of high temperature-adapted microorganisms. *Soft Matter* **2019**, *15*, 8381–8391. [CrossRef] [PubMed]
10. Cario, A.; Jebbar, M.; Thiel, A.; Kervarec, N.; Oger, P.M. Molecular chaperone accumulation as a function of stress evidences adaptation to high hydrostatic pressure in the piezophilic archaeon Thermococcus barophilus. *Sci. Rep.* **2016**, *6*, 29483. [CrossRef] [PubMed]
11. Cinar, H.; Fetahaj, Z.; Cinar, S.; Vernon, R.M.; Chan, H.S.; Winter, R.H.A. Temperature, Hydrostatic Pressure, and Osmolyte Effects on Liquid–Liquid Phase Separation in Protein Condensates: Physical Chemistry and Biological Implications. *Chem. Eur. J.* **2019**, *25*, 13049–13069. [CrossRef]
12. Roche, J.; Caro, J.A.; Norberto, D.R.; Barthe, P.; Roumestand, C.; Schlessman, J.L.; Garcia, A.E.; García-Moreno E., B.; Royer, C.A. Cavities determine the pressure unfolding of proteins. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 6945–6950. [CrossRef] [PubMed]
13. Hawley, S.A. Reversible pressure-temperature denaturation of chymotrypsinogen. *Biochemistry* **1971**, *10*, 2436–2442. [CrossRef] [PubMed]
14. Royer, C.A. Revisiting volume changes in pressure-induced protein unfolding. *Biochim. Biophys. Acta (BBA)—Protein Struct. Mol. Enzymol.* **2002**, *1595*, 201–209. [CrossRef]
15. Frauenfelder, H.; Alberding, N.A.; Ansari, A.; Braunstein, D.; Cowen, B.R.; Hong, M.K.; Iben, I.E.T.; Johnson, J.B.; Luck, S.; Marden, M.C.; et al. Proteins and pressure. *J. Phys. Chem.* **1990**, *94*, 1024–1037. [CrossRef]
16. Ortore, M.G.; Spinozzi, F.; Mariani, P.; Paciaroni, A.; Barbosa, L.R.S.; Amenitsch, H.; Steinhart, M.; Ollivier, J.; Russo, D. Combining structure and dynamics: Non-denaturing high-pressure effect on lysozyme in solution. *J. R. Soc. Interface* **2009**, *6*, S619–S634. [CrossRef]
17. Erlkamp, M.; Marion, J.; Martinez, N.; Czeslik, C.; Peters, J.; Winter, R. Influence of Pressure and Crowding on the Sub-Nanosecond Dynamics of Globular Proteins. *J. Phys. Chem. B* **2015**, *119*, 4842–4848. [CrossRef] [PubMed]

18. Marion, J.; Trovaslet, M.; Martinez, N.; Masson, P.; Schweins, R.; Nachon, F.; Trapp, M.; Peters, J. Pressure-induced molten globule state of human acetylcholinesterase: Structural and dynamical changes monitored by neutron scattering. *Phys. Chem. Chem. Phys.* **2015**, *17*, 3157–3163. [CrossRef] [PubMed]
19. Meinhold, L.; Smith, J.C.; Kitao, A.; Zewail, A.H. Picosecond fluctuating protein energy landscape mapped by pressure–temperature molecular dynamics simulation. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 17261–17265. [CrossRef] [PubMed]
20. Shrestha, U.R.; Bhowmik, D.; Copley, J.R.D.; Tyagi, M.; Leão, J.B.; Chu, X.Q. Effects of pressure on the dynamics of an oligomeric protein from deep-sea hyperthermophile. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 13886–13891. [CrossRef]
21. Librizzi, F.; Carrotta, R.; Peters, J.; Cupane, A. The effects of pressure on the energy landscape of proteins. *Sci. Rep.* **2018**, *8*, 2037. [CrossRef] [PubMed]
22. Kharakoz, D.P. Protein Compressibility, Dynamics, and Pressure. *Biophys. J.* **2000**, *79*, 511–525. [CrossRef]
23. Grigera, J.R.; McCarthy, A.N. The Behavior of the Hydrophobic Effect under Pressure and Protein Denaturation. *Biophys. J.* **2010**, *98*, 1626–1631. [CrossRef] [PubMed]
24. Studier, F.W.; Moffatt, B.A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **1986**, *189*, 113–130. [CrossRef]
25. Tabor, S.; Richardson, C.C. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 1074–1078. [CrossRef] [PubMed]
26. Frauenfelder, H.; Parak, F.; Young, R.D. Conformational Substates in Proteins. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 451–479. [CrossRef]
27. Frauenfelder, H.; Young, R.D.; Fenimore, P.W. Dynamics and the Free-Energy Landscape of Proteins, Explored with the Mössbauer Effect and Quasi-Elastic Neutron Scattering. *J. Phys. Chem. B* **2013**, *117*, 13301–13307. [CrossRef] [PubMed]
28. Bée, M. *Quasielastic Neutron Scattering: Principles and Applications in Solid State, Chemistry, Biology and Materials Science*; Adam Hilger: London, UK, 1988.
29. Sears, V.F. Neutron scattering lengths and cross sections. *Neutron News* **1992**, *3*, 26–37. [CrossRef]
30. Roh, J.H.; Novikov, V.N.; Gregory, R.B.; Curtis, J.E.; Chowdhuri, Z.; Sokolov, A.P. Onsets of Anharmonicity in Protein Dynamics. *Phys. Rev. Lett.* **2005**, *95*, 038101. [CrossRef]
31. Doster, W.; Cusack, S.; Petry, W. Dynamical transition of myoglobin revealed by inelastic neutron scattering. *Nature* **1989**, *337*, 754–756. [CrossRef]
32. Zaccai, G. How Soft Is a Protein? A Protein Dynamics Force Constant Measured by Neutron Scattering. *Science* **2000**, *288*, 1604–1607. [CrossRef] [PubMed]
33. Bicout, D.J.; Zaccai, G. Protein Flexibility from the Dynamical Transition: A Force Constant Analysis. *Biophys. D* **2001**, *80*, 1115–1123.
34. Peters, J. High Hydrostatic Pressure—A Key Element to Investigate Molecular Dynamics in Biosystems. *Front. Phys.* **2022**, *9*, 801539. [CrossRef]
35. Marteinsson, V.; Birrien, J.; Reysenbach, A.; Vernet, M.; Marie, D.; Gambacorta, A.; Messner, P.; Sleytr, U.; Prieur, D. Thermococcus barophilus sp. nov., a new barophilic and hyperthermophilic archaeon isolated under high hydrostatic pressure from a deep-sea hydrothermal vent. *Int. J. Syst. Evol. Microbiol.* **1999**, *49*, 351–359. [CrossRef]
36. Le Châtelier, H. Sur un Énoncé gènèral des lois des Équilibres chimiques. *C. R. Acad. Sci.* **1884**, *99*, 786–789.
37. Xue, Y.; Pavlova, M.S.; Ryabov, Y.E.; Reif, B.; Skrynnikov, N.R. Methyl Rotation Barriers in Proteins from 2H Relaxation Data. Implications for Protein Structure. *J. Am. Chem. Soc.* **2007**, *129*, 6827–6838. [CrossRef] [PubMed]
38. Hall, P.L.; Ross, D. Incoherent neutron scattering functions for random jump diffusion in bounded and infinite media. *Mol. Phys.* **1981**, *42*, 673–682. [CrossRef]
39. García-Colín, L.S.; del Castillo, L.F.; Goldstein, P. Theoretical basis for the Vogel-Fulcher-Tammann equation. *Phys. Rev. B* **1989**, *40*, 7040–7044. [CrossRef]
40. Rault, J. Origin of the Vogel-Fulcher-Tammann law in glass-forming materials : The $\alpha$-$\beta$ bifurcation. *J. Non-Cryst. Solids* **2000**, *271*, 177–217. [CrossRef]
41. Granato, A. A derivation of the Vogel-Fulcher-Tammann relation for supercooled liquids. *J. Non-Cryst. Solids* **2011**, *357*, 334–338. [CrossRef]
42. Fomina, M.; Schirò, G.; Cupane, A. Hydration dependence of myoglobin dynamics studied with elastic neutron scattering, differential scanning calorimetry and broadband dielectric spectroscopy. *Biophys. Chem.* **2014**, *185*, 25–31. [CrossRef]
43. Mallamace, F.; Chen, S.H.; Broccio, M.; Corsaro, C.; Crupi, V.; Majolino, D.; Venuti, V.; Baglioni, P.; Fratini, E.; Vannucci, C.; et al. Role of the solvent in the dynamical transitions of proteins: The case of the lysozyme-water system. *J. Chem. Phys.* **2007**, *127*, 045104. [CrossRef] [PubMed]
44. Chen, S.H.; Liu, L.; Fratini, E.; Baglioni, P.; Faraone, A.; Mamontov, E. Observation of fragile-to-strong dynamic crossover in protein hydration water. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 9012–9016. [CrossRef] [PubMed]
45. Lagi, M.; Chu, X.; Kim, C.; Mallamace, F.; Baglioni, P.; Chen, S.H. The Low-Temperature Dynamic Crossover Phenomenon in Protein Hydration Water: Simulations vs Experiments. *J. Phys. Chem. B* **2008**, *112*, 1571–1575. [CrossRef] [PubMed]
46. Vilgis, T.A. Strong and fragile glasses: A powerful classification and its consequences. *Phys. Rev. B* **1993**, *47*, 2882–2885. [CrossRef]

47. Bowler, M.W.; Nurizzo, D.; Barrett, R.; Beteva, A.; Bodin, M.; Caserotto, H.; Delagenière, S.; Dobias, F.; Flot, D.; Giraud, T.; et al. MASSIF-1: A beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules. *J. Synchrotron Radiat.* **2015**, *22*, 1540–1547. [CrossRef]
48. Svensson, O.; Gilski, M.; Nurizzo, D.; Bowler, M.W. Multi-position data collection and dynamic beam sizing: Recent improvements to the automatic data-collection algorithms on MASSIF-1. *Acta Crystallogr. Sect. D* **2018**, *74*, 433–440. [CrossRef] [PubMed]
49. Kabsch, W. *XDS*. *Acta Crystallogr. Sect. D* **2010**, *66*, 125–132. [CrossRef]
50. McCoy, A.J.; Grosse-Kunstleve, R.W.; Adams, P.D.; Winn, M.D.; Storoni, L.C.; Read, R.J. *Phaser* crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658–674. [CrossRef] [PubMed]
51. Liebschner, D.; Afonine, P.V.; Baker, M.L.; Bunkóczi, G.; Chen, V.B.; Croll, T.I.; Hintze, B.; Hung, L.W.; Jain, S.; McCoy, A.J.; et al. Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in *Phenix*. *Acta Crystallogr. Sect. D* **2019**, *75*, 861–877. [CrossRef]
52. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef]
53. Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and development of *Coot*. *Acta Crystallogr. Sect. D* **2010**, *66*, 486–501. [CrossRef] [PubMed]
54. Afonine, P.V.; Grosse-Kunstleve, R.W.; Echols, N.; Headd, J.J.; Moriarty, N.W.; Mustyakimov, M.; Terwilliger, T.C.; Urzhumtsev, A.; Zwart, P.H.; Adams, P.D. Towards automated crystallographic structure refinement with *phenix.refine*. *Acta Crystallogr. Sect. D* **2012**, *68*, 352–367. [CrossRef]
55. Vagin, A.A.; Steiner, R.A.; Lebedev, A.A.; Potterton, L.; McNicholas, S.; Long, F.; Murshudov, G.N. *REFMAC*5 dictionary: Organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. Sect. D* **2004**, *60*, 2184–2195. [CrossRef] [PubMed]
56. Natali, F.; Peters, J.; Russo, D.; Barbieri, S.; Chiapponi, C.; Cupane, A.; Deriu, A.; Bari, M.T.D.; Farhi, E.; Gerelli, Y.; et al. IN13 Backscattering Spectrometer at ILL: Looking for Motions in Biological Macromolecules and Organisms. *Neutron News* **2008**, *19*, 14–18.
57. Lelièvre-Berna, E.; Demé, B.; Gonthier, J.; Gonzales, J.P.; Maurice, J.; Memphis, Y.; Payre, C.; Oger, P.; Peters, J.; Vial, S. 700 MPa sample stick for studying liquid samples or solid-gas reactions down to 1.8 K and up to 550 K. *J. Neutron Res.* **2017**, *19*, 77–84. [CrossRef]
58. Peters, J.; Trapp, M.; Hughes, D.; Rowe, S.; Demé, B.; Laborier, J.L.; Payre, C.; Gonzales, J.P.; Baudoin, S.; Belkhier, N.; et al. High hydrostatic pressure equipment for neutron scattering studies of samples in solutions. *High Press. Res.* **2012**, *32*, 97–102. [CrossRef]
59. Sidorov, V.A.; Sadykov, R.A. Hydrostatic limits of Fluorinert liquids used for neutron and transport studies at high pressure. *J. Phys. Condens. Matter* **2005**, *17*, S3005–S3008. [CrossRef]
60. Richard, D.; Ferrand, M.; Kearley, G.J. Analysis and visualisation of neutron-scattering data. *J. Neutron Res.* **1996**, *4*, 33–39. [CrossRef]
61. Ollivier, J.; Mutka, H.; Didier, L. The New Cold Neutron Time-of-Flight Spectrometer IN5. *Neutron News* **2010**, *21*, 22–25. [CrossRef]
62. Van Hove, L. Correlations in Space and Time and Born Approximation Scattering in Systems of Interacting Particles. *Phys. Rev.* **1954**, *95*, 249–262. [CrossRef]
63. Grimaldo, M.; Roosen-Runge, F.; Hennig, M.; Zanini, F.; Zhang, F.; Jalarvo, N.; Zamponi, M.; Schreiber, F.; Seydel, T. Hierarchical molecular dynamics of bovine serum albumin in concentrated aqueous solution below and above thermal denaturation. *Phys. Chem. Chem. Phys.* **2015**, *17*, 4645–4655. [CrossRef] [PubMed]
64. Singwi, K.S.; Sjölander, A. Diffusive Motions in Water and Cold Neutron Scattering. *Phys. Rev.* **1960**, *119*, 863–871. [CrossRef]
65. Chudley, C.T.; Elliott, R.J. Neutron Scattering from a Liquid on a Jump Diffusion Model. *Proc. Phys. Soc.* **1961**, *77*, 353–361. [CrossRef]
66. Phillips, J.C.; Hardy, D.J.; Maia, J.D.C.; Stone, J.E.; Ribeiro, J.V.; Bernardi, R.C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, 044130. [CrossRef] [PubMed]
67. Jo, S.; Kim, T.; Iyer, V.G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865. [CrossRef]
68. Feller, S.E.; Zhang, Y.; Pastor, R.W.; Brooks, B.R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613–4621. [CrossRef]
69. Humphrey, W.; Dalke, A.; Schulten, K. VMD—Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [CrossRef]
70. Tian, W.; Chen, C.; Lei, X.; Zhao, J.; Liang, J. CASTp 3.0: Computed atlas of surface topography of proteins. *Nucleic Acids Res.* **2018**, *46*, W363–W367. [CrossRef]
71. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [CrossRef] [PubMed]
72. Vannier, P.; Marteinsson, V.T.; Fridjonsson, O.H.; Oger, P.; Jebbar, M. Complete genome sequence of the hyperthermophilic, piezophilic, heterotrophic, and carboxydotrophic archaeon thermococcus barophilus mp. *J. Bacteriol.* **2011**, *193*, 1481–1482. [CrossRef] [PubMed]

73. Fukui, T.; Atomi, H.; Kanai, T.; Matsumi, R.; Fujiwara, S.; Imanaka, T. Complete genome sequence of the hyperthermophilic archaeon thermococcus kodakaraensis kod1 and comparison with pyrococcus genomes. *Genome Res.***2005**, *15*, 352–363. [CrossRef]
74. Svensson, O.; Malbet-Monaco, S.; Popov, A.; Nurizzo, D.; Bowler, M.W. Fully automatic characterization and data collection from crystals of biological macromolecules. *Acta Crystallogr. Sect. D* **2015**, *71*, 1757–1767. [CrossRef] [PubMed]
75. Svensson, O.; Gilski, M.; Nurizzo, D.; Bowler, M.W. A comparative anatomy of protein crystals: Lessons from the automatic processing of 56000 samples. *IUCrJ* **2019**, *6*, 822–831. [CrossRef] [PubMed]
76. Bowler, M.W.; Svensson, O.; Nurizzo, D. Fully automatic macromolecular crystallography: The impact of massif1 on the optimum acquisition and quality of data. *Crystallogr. Rev.* **2016**, *22*, 233–249. [CrossRef]
77. Nurizzo, D.; Bowler, M.W.; Caserotto, H.; Dobias, F.; Giraud, T.; Surr, J.; Guichard, N.; Papp, G.; Guijarro, M.; Mueller-Dieckmann, C.; et al. RoboDiff: Combining a sample changer and goniometer for highly automated macromolecular crystallography experiments. *Acta Crystallogr. Sect. D* **2016**, *72*, 966–975. [CrossRef]
78. Hutin, S.; Van Laer, B.; Mueller-Dieckmann, C.; Leonard, G.; Nurizzo, D.; Bowler, M.W. Fully autonomous characterization and data collection from crystals of biological macromolecules. *JoVE* **2019**, *145*, e59032. [CrossRef] [PubMed]
79. Ortega, A.; Amorós, D.; García de la Torre, J. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.* **2011**, *101*, 892–898. [CrossRef]
80. Millero, F.J.; Dexter, R.; Hoff, E. Density and viscosity of deuterium oxide solutions from 5–70.deg. *J. Chem. Eng. Data* **1971**, *16*, 85–87. [CrossRef]

# Unravelling the mechanisms of adaptation to high pressure in proteins

**Antonino Caliò[1], Cécile Dubois[2], Stéphane Fontanay[1], Michael Marek Koza[3], François Hoh[2], Christian Roumestand[2], Philippe Oger[1], Judith Peters[3,4,5*]**

*[1]Univ Lyon, UCBL, INSA Lyon, CNRS, MAP UMR5240, 69622 Villeurbanne, France,*

*[2]CBS, INSERM U1054, CNRS UMR 5048, Univ Montpellier, 34090 Montpellier, France,*

*[3]Institut Laue Langevin, Grenoble, France,*

*[4]Univ Grenoble Alpes, CNRS, LiPhy, Grenoble, France,*

*[5]Institut Universitaire de France, Paris, France*

*corresponding author: jpeters@ill.fr

# Supplementary Informations

## Genes and proteins sequences

The complete genomes of *Thermococcus barophilus* MP and *Thermococcus kodakarensis* KOD1 are available on GenBank under the accession codes CP002372[1] and AP006878[2] respectively. The two genes are identified as TERMP_00744 (Tba PMI) and TK_0503 (Tko PMI). The codon-optimized sequences used in this work to express the two proteins are reported here:

Tba PMI (TERMP_00744)

GCTGAACATATGAAAGCCGAAATTAAGGAGTTCATTGACAGGGGAACTTA
TAGGAAAGCCCCATTGTTTGAAGGTGAGCTTCCTGAAGGGAGTTACGCTC
AAATAGTTGAAATTAAACCCAAGCAGACGGTTCCAAAACACTATCATGAA
AAACAATATGAACTATTTTACATAATTAGCGGACAAGCAAAGCTCGGCAT
TGAAGAAAGAGAATATGATGCAAAACCAGGGGACATATTTTTAGTTAAGC
CCAAGACTGTTCATTGGGTTGTCAATAAAAAGGAAGAGCCATTCAGGCTT
TTTGTGATTAAGCTGAACTACTTTGGAGATGATAGCGTTTGGCTTGAGTG
AGGATCCTTCC

Tko PMI (TK_0503)

GGCTGAACATATGAAAGCAGAAATCAAAAATTTAATTGACCGCGGTACAT
ATCGTAAATTACCGTTATTCGAAGGTGAATTGCCTGAAGGTTCATACGCT
CAAATTGTAGAAGTAAAACCTAAGCAGACTGTAAAAAAACACTACCACGA
ACGTCAATACGAACTGTTCTACATTATCTCTGGTGAAGCTCGTCTTGGTA
TTGGTGATACAGAATATCAAGCAAAGCCTGGTGATATTTTTCTAGTTAAA
CCAAAAACCGTACACTGGGTAGTTAATGAAAAAGACGAACCATTCCGTCT
TTTTGTTGTTAAGTTAAATTATCACGGTGATGACTCTGTATGGTTAGAAT
GAGGATCCTTCC

The restriction sites for *NdeI* (blue) and *BamHI* (red) have been indicated. The genes were received from GENEWIZ in pUC57-Kan plasmid, digested with *NdeI* and *BamHI*-HF® (New England Biolabs) and then ligated into the Multi Cloning Site of the pET-16b plasmid (digested with the same enzymes) by means of T4 DNA Ligase (Thermo Fisher). All construct propagation was done using NEB-5$\alpha$® *E. coli* cells (New England Biolabs). Plasmids were then extracted, purified (Macherey-Nagel NucleoSpin® Plasmid kit) and sequenced to verify the constructs, then transformed into the BL21(DE3) pLysS strain for expression. All the listed products have been employed according to the standard protocols issued by the manufacturers.

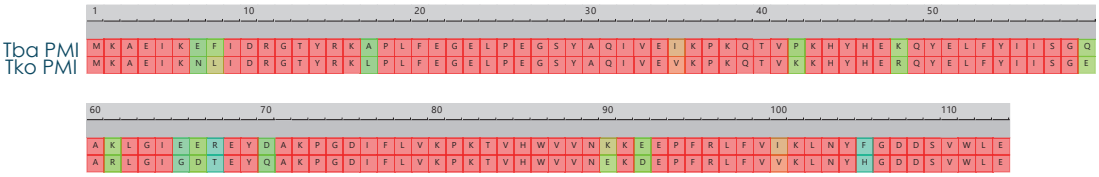


*Figure S 1: Sequence alignment of Tba PMI and Tko PMI, color coded for sequence similarity in a Blue-Green-Red scale (blue for least similar, red for most similar).*

Figure S1 reports the aligned sequences of the two proteins. They both are 113 residues long, and there are 16 substitutions between them, which are color-coded for sequence similarity (BLOSUM30) in a Blue-Green-Red scale (blue being least similar, red most similar).

## Protein production and purification

The proteins have been expressed and purified as reported in the Methods section in the main text. Here we report the calibration run for the XK50-60 column using the GE Healthcare® Low Molecular Weight kit (fig. S2), which consists of five proteins: Conalbumin (75 kDa), Ovalbumin (44 kDa), Carbonic Anhydrase (29 kDa), Ribonuclease A (13.7 kDa) and Aproritin (6.5 kDa). The void volume of the column was determined to be 360 $ml$ by using the Blue Dextran 2000 supplied with the kit. The kit was used according to the manufacturer's protocol.

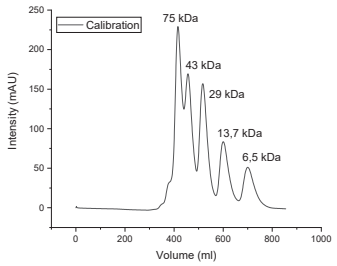


*Figure S 2: FPLC calibration run.*

Figure S3 reports an example purification run for Tko PMI, showing that it elutes at a volume compatible with a MW of 30 kDa (to be compared with fig. S2), but SDS-PAGE analysis of the fractions shows a band at the 15 kDa mark, thus confirming the dimeric nature of the protein. Fractions from 25 to 48 were pooled together and used as the pure fraction, while the rest of the fractions shown

to contain the protein by SDS-PAGE (fractions 12 to 24 and 49 to 70) were processed separately and kept as a backup.

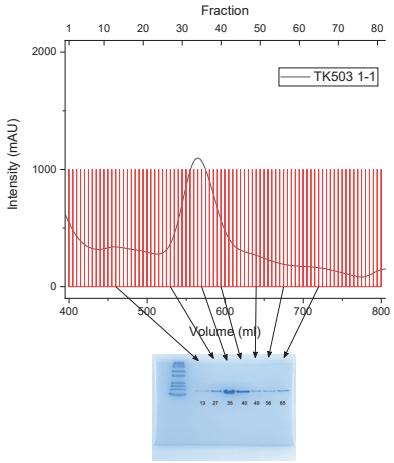


*Figure S 3: Example FPLC run. Fractions are given on the upper axis, and the SDS-PAGE analysis of some representative ones is given.*

Figure S4 reports an SDS-PAGE analysis of Tba PMI after pooling of the FPLC fractions and concentration, attesting the purity of the protein. The only steps performed after this check were lyophilization, and then the subsequent dissolution in $D_2O$ immediately before the neutron experiments.

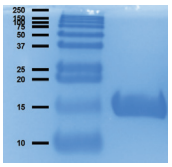


*Figure S 4: SDS-PAGE analysis of the final product (Tba PMI).*

## X-ray Crystallography and Molecular Dynamics

X-ray diffraction datasets were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble) at the ID30A-1 beam line (Massif1[3–9]) using a pixel detector (PILATUS3 2M) and auto-processed by XDSAPP package[10]. The following tables report the final refinement statistics for Tba PMI (Table S1) and for Tko PMI (Table S2).

| **Data collection** | **Tba PMI** |
|---|---|
| Space group | I222 |
| Unit cell parameters (a,b,c) | 48.500 Å 52.458 Å 112.892 Å |
| ($\alpha$, $\beta$, $\gamma$) | 90.0° 90.0° 90.0° |
| Wavelength (Å) | 0.96546 |
| Resolution limit (Å) | 47.58-1.58 (1.62-1.58) |
| Completeness (%) | 99.54 (98.6) |
| Multiplicity | 5.0 (4.4) |
| I/$\sigma_I$ | 10.2 (1.27) |
| $R_{merge}$ | 0.104 (1.015) |
| CC(1/2) | 0.994 (0.18) |
| Total number of observations | 71358 (3145) |
| Total number unique | 19202 (712) |
| **Refinement** | |
| $R_{work}/R_{free}$ (%) | 0.1735/0.212 |
| **Number of atoms** | |
| All atoms | 933 |
| Protein | 871 |
| Water | 60 |
| Ligands | Mg: 1 |
| **B-factors (Å$^2$)** | |
| Protein | 28 |
| Mg | 47 |
| Water | 42 |
| **RMSD** | |
| Bond lengths (Å) | 0.0169 |
| Bond angles (°) | 1.86 |
| Ramachandran favored (%) | 100 |
| Ramachandran outliers (%) | 0 |

*Table S 1: Refinement parameters for Tba PMI.*

| **Data collection** | **Tko PMI** |
|---|---|
| Space group | $P3_1$ |
| Unit cell parameters (a,b,c) | 91.16 Å 91.16 Å 113.38 Å |
| ($\alpha$, $\beta$, $\gamma$) | 90.0° 90.0° 120.0° |
| Wavelength (Å) | 0.96546 |
| Resolution limit (Å) | 64.79-2.16 (2.2-2.16) |
| Completeness (%) | 99.2 (98.4) |
| Multiplicity | 2.2 |
| $I/\sigma_I$ | 7.8 (1.1) |
| $R_{merge}$ | 0.078 (0.661) |
| CC(1/2) | 0.977 (0.297) |
| Total number of observations | 387303 (33137) |
| Total number unique | 40326 (3632) |
| **Refinement** | |
| $R_{work}/R_{free}$ (%) | 0.23/0.29 |
| **Number of atoms** | |
| All atoms | 7172 |
| Protein | 7118 |
| Water | 50 |
| Ligands | Zn: 4 |
| **B-factors (Å$^2$)** | |
| Protein | 68 |
| Mg | 20 |
| Water | 55 |
| **RMSD** | |
| Bond lengths (Å) | 0.027 |
| Bond angles (°) | 3.21 |
| Ramachandran favored (%) | 90 |
| Ramachandran outliers (%) | 2 |

*Table S 2: Refinement parameters for Tko PMI.*

Figure S5 shows the structural alignment of the resulting crystal structures. The substituted residues have also been labelled. Unfortunately, the fast dynamics of the N terminals of both proteins impacted the measured electron densities from the crystals, impeding the characterization of the first two substitutions (E7N and F8L).

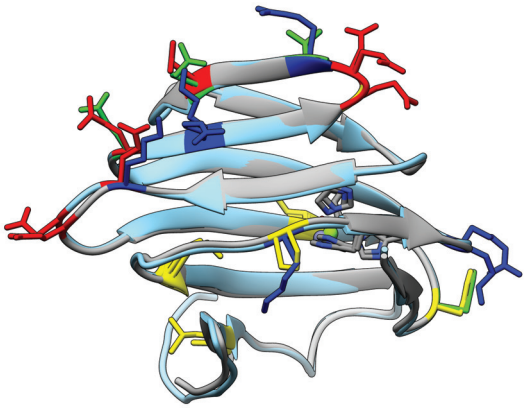

*Figure S 5: Structural alignment of the two crystal structures. Tba PMI is shown as gray ribbon and Tko PMI as light blue. Substitutions are shown in sticks and colored by residue type (red for acidic, blue for basic, green for polar and yellow for hydrophobic).*

Figure S6 shows the active site, located into the $\beta$-barrel and constituted by the conserved residues His44, His46, Glu51 and His85. The ligand is a magnesium ion in Tba PMI and a zinc ion in Tko PMI. However, this assignment is not definitive, as both ions have a Van der Waals radius smaller than the resolution.

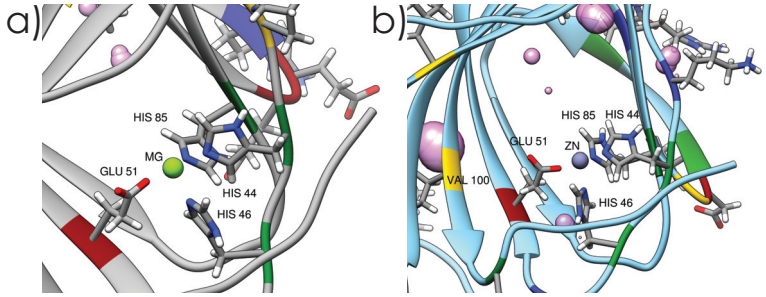


*Figure S 6: Detail of the active side of Tba PMI (panel a) and Tko PMI (panel b). The involved residues (His44, His46, Glu51 and His85) and the metal ions (Mg for Tba PMI and Zn for Tko PMI) are shown.*

Molecular dynamics simulations were performed as described in the methods section of the main text. Figure S7 shows the $C_\alpha$ RMSD values for the six trajectories.

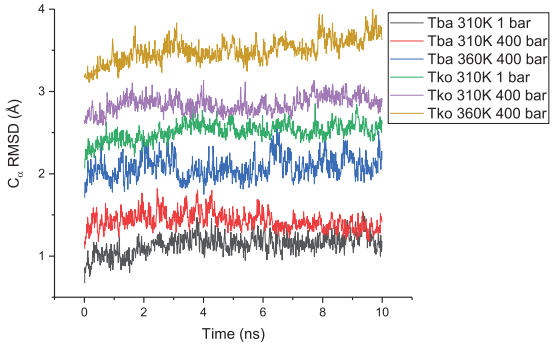


*Figure S 7: $C_\alpha$ RMSD values calculated for the six trajectories. Values have been vertically shifted for clarity.*

The most representative structures were calculated as described in the main text, and the cavities have been calculated with Castp[11].

Figure S8 shows the surface electrostatic potential of the two proteins (panels a and b) and the vertical section of their ligand pocket. Tba PMI presents a higher positive potential around the mouth of the pocket, and also a higher negative potential in the middle region. This could provide increased shielding from water penetration under high pressure conditions. Moreover, the narrower mouth seemingly plays a role as well in this process.

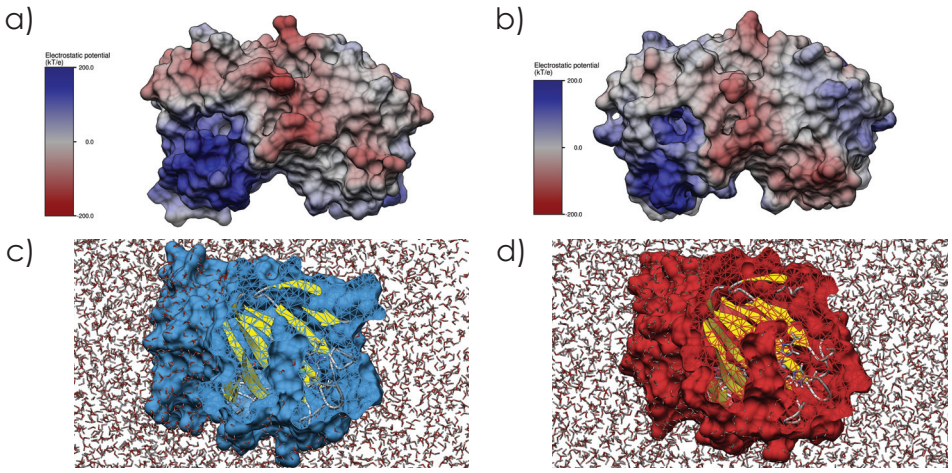


*Figure S 8: Surface representation (at 310 K and 1 bar) of Tba PMI and Tko PMI colored by electrostatic potential (panel a and b). One unit of $K_BT/e$ is equivalent to $\sim 27\ mV$ at 310 K. Panel c (Tba PMI) and d (Tko PMI) show a vertical section of the surfaces to highlight the ligand pocket.*

Figure S9 shows the equivalent radius calculated in different conditions. After submitting the six equilibrated structures (as described before) to Castp, it returns a value of the molecular volume for each cavity. The two proteins being dimers, the two pockets' volumes were averaged, and then the equivalent radius has been calculated as $R = (\frac{3}{4\pi}V)^{1/3}$, i.e. the radius of a sphere with the same volume. The trend is remarkably similar to that found by QENS.

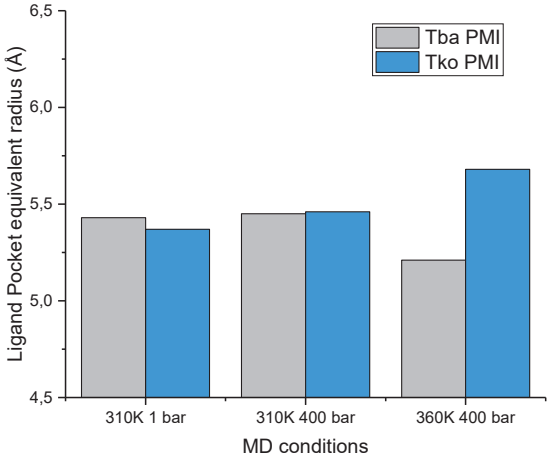


*Figure S 9: Equivalent radius of the ligand pocket calculated from the Castp results at different simulation conditions.*

## Diffusion coefficient calculation

The diffusion coefficient was calculated with HYDROPRO[12] for both proteins at 353 K (at atmospheric pressure). This was the highest temperature (and lowest pressure) reached during experiments, the result would thus give an upper limit for $D$. Moreover, HYDROPRO calculates $D$ in the infinite dilution limit, therefore we expect the real diffusion coefficient in our samples at 120 mg/ml to be considerably lower. The calculation was performed by using the density and viscosity of pure $D_2O$[13]. The results were $22.33$ Å$^2/ns$ and $22.24$ Å$^2/ns$ for Tba PMI and Tko PMI respectively, thus giving a quasi-elastic broadening of $\Gamma = \hbar D q^2 = 42.5 \mu eV$ for Tba PMI, and $42.3 \mu eV$ for Tko PMI at the highest $q$ value investigated here. Comparing this with the resolution of IN5 in our experimental conditions ($70$ $\mu eV$ HWHM), we can conclude that translational diffusion of the proteins does not give a measurable contribution to the signal.

## EINS and QENS Analysis

Both EINS and QENS data were corrected for empty cell, solvent scattering and detector efficiency (measured by means of a vanadium standard) by taking into account their transmission and the solvent volume fraction, according to the following expression:

$$S_{corr}(q,\omega) = \frac{(\frac{1}{t_{sample}} S_{sample} - \frac{1}{t_{cell}} S_{cell}) - \phi(\frac{1}{t_{D2O}} S_{D2O} - \frac{1}{t_{cell}} S_{cell})}{\frac{1}{t_{vana}} S_{vana}} \quad (1)$$

where $t$ denotes the transmission, and the $(q,\omega)$ dependence has been dropped for clarity. $\phi$ denotes the solvent volume fraction in the samples, which has been calculated by dividing the elastic scattering intensity of the solvent by that of the sample at around $q = 1.6$Å$^{-1}$ at each temperature and pressure value, where $D_2O$ displays a broad coherent peak. The transmission can be directly measured on IN13 and the measured values were used for both data sets in virtue of the very weak dependence of the transmission on the neutron wavelength ($2.23$Å for IN13, $5$Å for IN5). The transmission of the samples was 91% for Tba PMI and 93% for Tko PMI, we could thus reasonably assume that multiple scattering effects are negligible. Figure S10 shows the temperature independence of $d$ when fitting the IN13 data with the two-state model[14], thus justifying the global fitting approach with $d$ as a shared parameter.

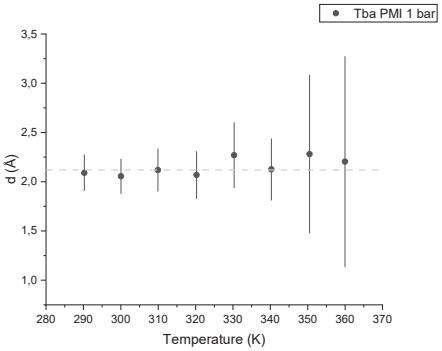


*Figure S 10: Temperature independence of the parameter d in the two state model.*

Figure S11 displays a vanadium spectrum, from which the resolution function was extracted by fitting a Gaussian curve. Its parameters at all $q$ values were then used to model the resolution function and convolute it to $S(q,\omega)$ to get the experimental scattering function $S_{exp}(q,\omega)$, according to:

$$S_{exp}(q,\omega) = B(q) + \mathcal{R}(q,\omega) \circledast [D(q)S(q,\omega)] \tag{2}$$

where $B(q)$ is a flat background accounting for fast vibrational motions, $\mathcal{R}(q,\omega)$ is the resolution function, $\circledast$ is the convolution operator in $\omega$, $D(q)$ is a scale factor proportional to the Debye-Waller factor and $S(q,\omega)$ is the theoretical scattering function.

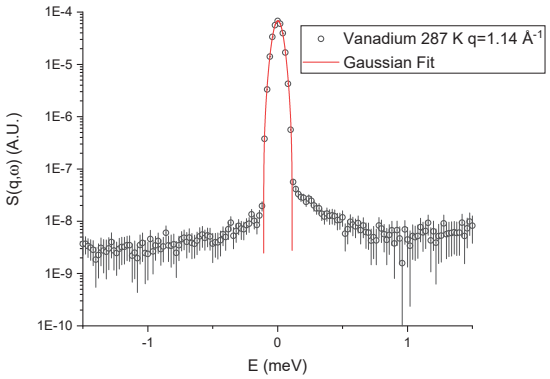


*Figure S 11: Spectrum of vanadium at 287 K and at $q = 1.14\mathring{A}^{-1}$ (black circles) and gaussian fitting (red line).*

We first performed a model-free approach by fitting a sum of Lorentzian functions and leaving their parameters free. As shown in Figure S12, when fitting with three Lorentzians the third always converged to a negligible contribution: its area was two orders of magnitude lower than that of $\mathcal{L}_1$ and $\mathcal{L}_2$, and its width was higher than the instrument's dynamic range in our conditions, therefore two Lorentzians were used. Their HWHM ($\Gamma$) are shown in Figure S13 and S14.

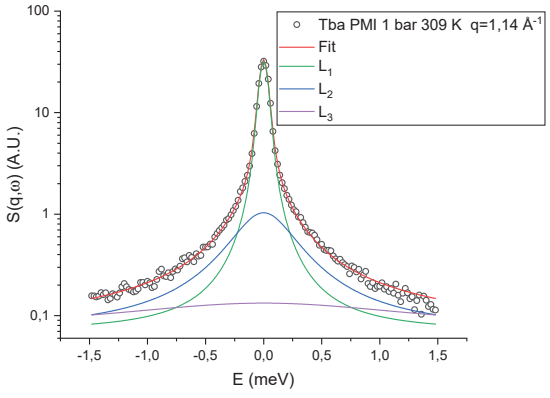


*Figure S 12: Example of a three-Lorentzians fitting.*

The broad component $\mathcal{L}_2$ shows a rather weak $q$-dependence (fig. S13), it has thus been assigned to localized motions, which give rise to a $q$-independent Lorentzian plus an elastic contribution[15],

according to:

$$S_{loc}(q,\omega) = A_0(q)\delta(\omega) + \frac{1 - A_0(q)}{\pi}\frac{\Gamma_{loc}}{\Gamma_{loc}^2 + \omega^2} \tag{3}$$

where $A_0(q)$ is the *Elastic Incoherent Structure Factor*, and $\Gamma_{loc}$ is the $q$-independent HWHM related to localized motions.

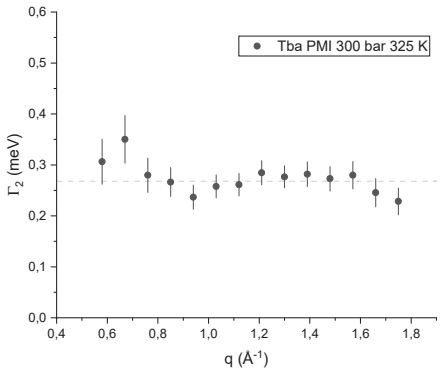


*Figure S 13: Width of the broad component as a function of the scattering vector q.*

The width of the narrow component $\mathcal{L}_1$ exhibited the signature characteristics of jump-diffusion motions, that is, monotonically increasing at low $q$ and reaching a plateau value at high $q$ (fig. S14). Different jump-diffusion models have been tested[16–18], and the one by Hall and Ross[16] was found to better fit the data compared to the one by Singwi and Sjölander, commonly used to model jump-diffusion motions in proteins.

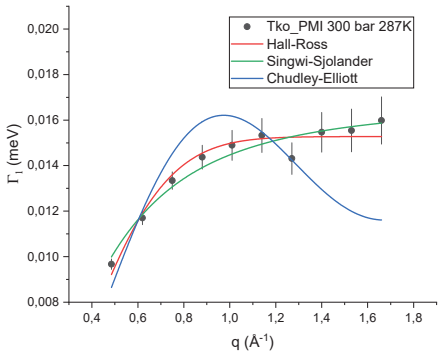


*Figure S 14: Width of the narrow component as a function of the scattering vector q.*

These motions thus bring about another Lorentzian contribution, according to:

$$S_{jump}(q,\omega) = \frac{1}{\pi}\frac{\Gamma_j(q)}{\Gamma_j^2(q) + \omega^2} \tag{4}$$

with

$$\Gamma_j(q) = \frac{\hbar}{\tau}\left(1 - \exp\left(-\frac{q^2\langle l\rangle^2}{2}\right)\right). \tag{5}$$

where $\tau$ is the mean time between two jumps (residence time) and $\langle l\rangle$ is the mean jump length, that

is, the mean of the jump distribution function, which is assumed to be Gaussian according to the Hall-Ross model. The overall theoretical scattering function (eq. 2 in the main text) results from the convolution of these two contributions. Since hydrogen atoms motions dominate the QENS signal, a convolution is necessary instead of a simple sum, as they can concomitantly perform both motions. As an example, a methyl group in an isoleucine can rotate (giving rise to $S_{loc}$) and, at the same time, perform jump-diffusion with the rest of the residue's side-chain (giving rise to $S_{jump}$).

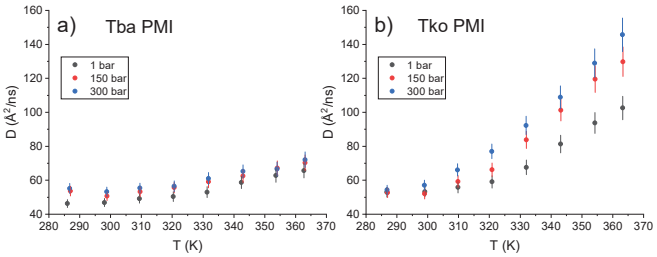


*Figure S 15: Pseudo-diffusion coefficient calculated for Tba PMI and Tko PMI at all temperature and pressure values.*

Figure S15 shows the pseudo-diffusion coefficient $D_{pseudo}$ described in the main text, which can be used to compare these results with other works.

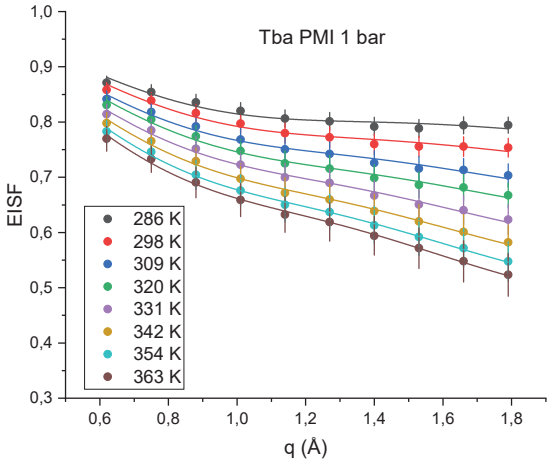


*Figure S 16: Example fit of the EISF at all temperatures for Tba PMI at 1 bar.*

Figure S16 shows a fit of the EISF at all temperatures for Tba PMI at 1 bar, according to the model defined in equation 4 in the main text, and figure S17 displays the fitting parameters not shown in the main text.

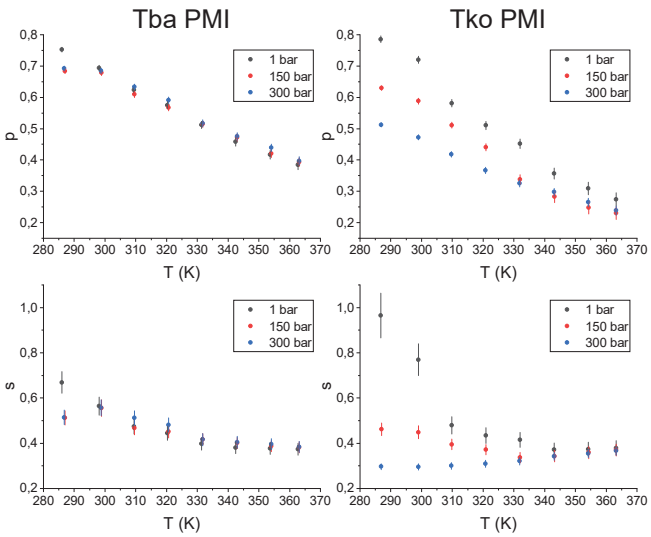


*Figure S 17: Temperature independence of the parameter d in the two state model.*

Both $p$ (immobile fraction) and $s$ (fraction of atoms performing confined jump-diffusion) are essentially pressure-independent, in line with all the other results, while for Tko PMI it is evident that pressure is activating new motions on one hand (decrease in $p$), and inhibiting the jump-diffusion motions on the other hand.

# References

[1] Vannier, P., Marteinsson, V. T., Fridjonsson, O. H., Oger, P. & Jebbar, M. Complete genome sequence of the hyperthermophilic, piezophilic, heterotrophic, and carboxydotrophic archaeon thermococcus barophilus mp. *Journal of Bacteriology* **193**, 1481–1482 (2011).

[2] Fukui, T. *et al.* Complete genome sequence of the hyperthermophilic archaeon thermococcus kodakaraensis kod1 and comparison with pyrococcus genomes. *Genome research* **15**, 352–363 (2005).

[3] Svensson, O., Gilski, M., Nurizzo, D. & Bowler, M. W. Multi-position data collection and dynamic beam sizing: recent improvements to the automatic data-collection algorithms on MASSIF-1. *Acta Crystallographica Section D* **74**, 433–440 (2018).

[4] Svensson, O., Malbet-Monaco, S., Popov, A., Nurizzo, D. & Bowler, M. W. Fully automatic characterization and data collection from crystals of biological macromolecules. *Acta Crystallographica Section D* **71**, 1757–1767 (2015).

[5] Svensson, O., Gilski, M., Nurizzo, D. & Bowler, M. W. A comparative anatomy of protein crystals: lessons from the automatic processing of 56000 samples. *IUCrJ* **6**, 822–831 (2019).

[6] Bowler, M. W. *et al.* MASSIF-1: a beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules. *Journal of Synchrotron Radiation* **22**, 1540–1547 (2015).

[7] Bowler, M. W., Svensson, O. & Nurizzo, D. Fully automatic macromolecular crystallography: the impact of massif-1 on the optimum acquisition and quality of data. *Crystallography Reviews* **22**, 233–249 (2016).

[8] Nurizzo, D. *et al.* RoboDiff: combining a sample changer and goniometer for highly automated macromolecular crystallography experiments. *Acta Crystallographica Section D* **72**, 966–975 (2016).

[9] Hutin, S. *et al.* Fully autonomous characterization and data collection from crystals of biological macromolecules. *JoVE* e59032 (2019).

[10] Kabsch, W. *XDS*. *Acta Crystallographica Section D* **66**, 125–132 (2010).

[11] Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Research* **46**, W363–W367 (2018).

[12] Ortega, A., Amorós, D. & García de la Torre, J. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophysical Journal* **101**, 892–898 (2011).

[13] Millero, F. J., Dexter, R. & Hoff, E. Density and viscosity of deuterium oxide solutions from 5-70.deg. *Journal of Chemical & Engineering Data* **16**, 85–87 (1971).

[14] Doster, W., Cusack, S. & Petry, W. Dynamical transition of myoglobin revealed by inelastic neutron scattering. *Nature* **337**, 754–756 (1989).

[15] Bée, M. *Quasielastic Neutron Scattering: Principles and Applications in Solid State, Chemistry, Biology and Materials Science* (Adam Hilger, 1988).

[16] Hall, P. L. & Ross, D. Incoherent neutron scattering functions for random jump diffusion in bounded and infinite media. *Molecular Physics* **42**, 673–682 (1981).

[17] Singwi, K. S. & Sjölander, A. Diffusive motions in water and cold neutron scattering. *Phys. Rev.* **119**, 863–871 (1960).

[18] Chudley, C. T. & Elliott, R. J. Neutron scattering from a liquid on a jump diffusion model. *Proceedings of the Physical Society* **77**, 353–361 (1961).

## 3.2 Article 2 : How do proteins cope with pressure ?

### 3.2.1 Foreword

Inspired by the results obtained from the first couple of proteins, this time we analysed a different couple, namely the two *Ribosomal Proteins S24e* from the same two organisms. They were chosen as they were both expected to be monomeric, thus enabling their study by NMR spectroscopy. This allowed for a better characterization of the single substitutions, thanks to the residue-specific information gained from these experiments.

*Figure 3.1 – High-Pressure cell after the breakage. Note the crack extending longitudinally from the threading.*

We combined this with a neutron scattering experiment and the structural information obtained again by X-ray crystallography and MD simulations. The neutron experiment was performed in very similar conditions to the one for the PMIs, so that we could eventually confirm our previous findings, and investigate the generality of the properties of piezophilic proteins. Unfortunately, during the last temperature run for the *T. kodakarensis* protein, the high-pressure cell suffered a breakage at 800 bar (fig. 3.1), terminating the data collection in the middle of the ramp. The sample was lost, therefore the ramp could not be resumed.

For this work, I prepared the protein samples for neutron scattering, design and performed the experiments (with the help of Dr. Alessio De Francesco), and I carried out the data analysis. I assisted Dr. Cécile Dubois in the preparation of the $^{15}$N labelled proteins for the NMR experiments, and during the preparation of the crystals. I assisted to the NMR experiments at the CBS with Prof. Christian Roumestand, and I performed the MD simulations with the assistance of Dr. Grazia Cottone, after the resolution of the structures by Dr. François Hoh. Dr. Philippe Oger and Prof. Judith Peters conceived and designed the research strategy. This article has been revised by all the authors and will be shortly submitted to a journal.

# How do proteins cope with pressure?

**Antonino Caliò[1], Cécile Dubois[2], Alessio De Francesco[3,4], François Hoh[2], Philippe Barthe[2], Grazia Cottone[5], Christian Roumestand[2], Judith Peters[4,6,7*], Philippe Oger[1*]**

[1]*Univ Lyon, UCBL, INSA Lyon, CNRS, MAP UMR5240, 69622 Villeurbanne, France,*

[2]*CBS, INSERM U1054, CNRS UMR 5048, Univ Montpellier, 34090 Montpellier, France,*

[3]*CNR-IOM & INSIDE@ILL c/o Operative Group in Grenoble (OGG), 38042 Grenoble, France,*

[4]*Institut Laue Langevin, 38042 Grenoble, France,*

[5]*Department of Physics and Chemistry-Emilio Segrè, University of Palermo, Palermo 90128, Italy,*

[6]*Univ Grenoble Alpes, CNRS, LiPhy, 38400 Grenoble, France,*

[7]*Institut Universitaire de France, 75231 Paris, France*

*corresponding authors: philippe.oger@insa-lyon.fr, jpeters@ill.fr

## Abstract

High hydrostatic pressure is known to destabilize most cellular functions in common life forms, nonetheless, numerous organisms have been discovered to populate habitats characterized by extreme conditions. For instance, hydrothermal vents feature high temperature and high pressure, and are also one of the candidates as an environment in which life on Earth originated. However, the specific pathway leading to the adaptation to high pressure in such organisms is still unknown. In this work, we show the contribution of amino-acid substitutions in two orthologous proteins by means of neutron scattering and NMR spectroscopy. We found direct evidence for the influence of cavities on the stability of the protein core, and how their elimination is achieved in the piezophilic protein owing to bulkier residues. These results constitute the foundation upon which a model explaining high-pressure adaptation at the genomic level can rely on.

Keywords: *High-Pressure Adaptation, Origins of life, Neutron Scattering, NMR Spectroscopy*

## Introduction

More than 70% of life forms on Earth live under High Hydrostatic Pressure (HHP) conditions[1,2], however, the adaptation mechanism to HHP in these organisms has not been unveiled yet. The knowledge of such mechanisms is of interest to investigate different scenarios concerning the origin of life on our planet[3,4]. One of the complicating factors in the search for this adaptation pathway is the fact that piezophilic organisms are often also adapted to high[5] or low[6] temperature, impeding a full disentanglement of those different contributions to genomic adaptation[7]. Although some attempts have been made to tackle this issue with cold-loving bacteria[8], genomic studies carried out on hyperthermophilic Archaea did not lead to significant conclusions, but gave some indications about genes that are under positive selection for HHP growth. Moreover, whole-cell studies[9–12] evidenced profound differences in the proteome of two nearly isogenic species belonging to the genus of *Thermococcales*, namely *Thermococcus barophilus* and *Thermococcus kodakarensis*. The first is a piezophilic Archaeon isolated in the Snakepit hydrothermal vent at a depth of 3550 m, growing best at 400 bar[13],

while the latter is piezosensitive and grows best at atmospheric pressure[14]. The nearly identical optimum growth conditions for these two Archaea, except for their optimum pressure, ensures that any difference in the behaviour of their proteomes arises from the adaptation pathway to HHP that *T. barophilus* has undergone during its evolution. In a previous work[15], we focused our attention on two proteins, the *Phosphomannose Isomerases* from these two species, and found the first hints about the amino-acid substitutions responsible for their different pressure stability. Neutron scattering was used to probe their dynamics under extreme conditions, in parallel with X-ray crystallography and molecular dynamics simulations to elucidate their structural properties. We propose a similar approach here, complemented by Nuclear Magnetic Resonance (NMR) spectroscopy, to study the *Ribosomal Proteins S24e* from the same two organisms. NMR has been thoroughly used in the study of proteins[16,17], and has also been coupled to HHP equipment in order to access different protein conformations by means of pressure application[18–20]. In the framework of this study, 2D Heteronuclear Single Quantum Coherence (HSQC) experiments gave us access to local information at the residue level, gaining direct evidence on the contribution of each residue to volume effects governing the stability of the proteins. This complements and extends the Elastic Incoherent Neutron Scattering (EINS) results and, together with the structural information obtained by X-ray crystallography, gives a complete picture on the effect that each substitution has on the response of these proteins to HHP. Moreover, our results are in excellent agreement with our previous findings, indicating that such response to pressure could be common feature of different pressure-adapted proteins.

# Results

## Crystal Structures and Molecular Dynamics simulations

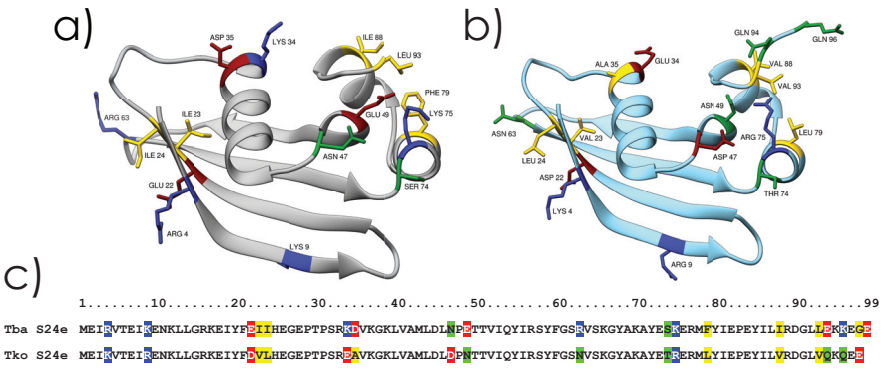


*Figure 1: Crystal structures of Tba S24e (a) and Tko S24e (b) and alignment of their sequences (c) with labelled substituted residues for comparison. Residues are color-coded for polarity (red for acidic, blue for basic, green for polar, and yellow for hydrophobic).*

The two Ribosomal proteins S24e were obtained by recombinant expression of genes TERMP_02015 (*T. barophilus*) and TK_01696 (*T. kodakarensis*), purified by heat treatment and Size-Exclusion Chromatography (SEC), and crystallised as described in Methods. Figure 1 shows the two crystal structures. They both present the same fold, consisting of four $\beta$-strands and an opposing $\alpha$-helix, with smaller helices in the C-terminal region, similarly to other archaeal orthologs from *T. acidophilum*

(2G1D), *P. abyssi* (2V94) and *M. maripaludis* (1YWX). The C-terminal represents the main contact surface with the small subunit of the ribosome, as evidenced in a work by Armanche et al.[21] by cryogenic electron microscopy. Curiously, they also propose a promiscuous behaviour of these proteins, in that they could both bind to the small and large ribosomal subunit indifferently.

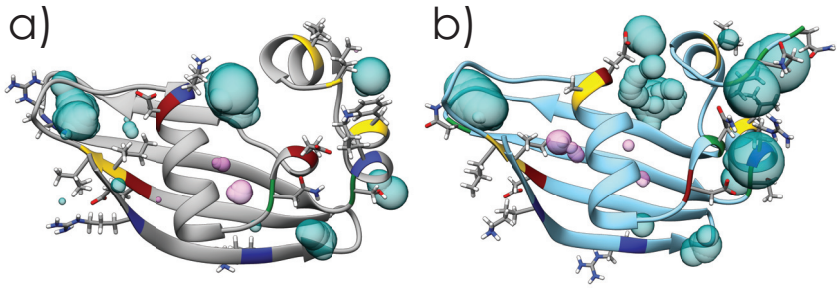


*Figure 2: MD relaxed structures of Tba S24e (a) and Tko S24e (b) with internal (magenta) and solvent-accessible (cyan) cavities highlighted. The colouring of substituted residues is the same as fig. 1.*

Molecular Dynamics (MD) simulations have been used to relax the structures, similarly as in[15], and the cavities have been revealed using the same methods (fig. 2). The two orthologs differ in 18 positions, 19 counting residue 99 which is present in Tba S24e but absent in Tko S24e. The I23V substitution resembles what had been found for the Phosphomannose Isomerases from the same two organisms in our previous work: the pressure-adapted protein employs a bulkier residue to eliminate an internal cavity and improve the pressure stability of the protein. Another interesting substitution is E49N, as the negatively-charged glutamate is able to form a salt bridge with Lys75, enhancing the structural stability of the C-terminal. The latter presents other differences among the two proteins and, although it cannot be proven in the present work, we speculate that the conservative substitutions in it (F79L, I88V and L88V) are seemingly involved in the optimization of the protein-ribosome contact. Indeed, in our previous study[15], we found that a conservative substitution was involved in the optimization of the monomer-monomer contact in the Phosphomannose Isomerase, therefore we suggest a similar role of these conservative substitution located at the protein-ribosome contact surface. The rest of the substitutions are located at the protein-water interface, as also observed for the Phosphomannose Isomerases.

## Elastic incoherent neutron scattering (EINS)

EINS was employed to gather information about the fast dynamics of the two proteins on a 100 ps time-scale. The use of $D_2O$ solutions of proteins assured proper transmission of hydrostatic pressure as well as an optimal signal to noise ratio, by virtue of the difference in scattering cross sections between hydrogen (present in the protein) and deuterium (present in the solvent)[22]. A two-state model, described in[23], has been used to interpret the data (fig. 3) and to extract both dynamical and thermodynamical information: it assumes that hydrogen atoms can explore two distinct potential wells separated by a distance $d$ and a free-energy difference $\Delta G = \Delta H - T\Delta S$, and they can perform harmonic motions in each well with an associated Mean Square Displacement (MSD) $\Delta x_0^2$. It must be stressed that global diffusive motions are probably present in the time window of the

experiment, given the temperature range and the relatively low concentration of the samples in our conditions. However, both proteins are almost identical in molecular weight and overall shape of the molecules, thus it can be reasonably assumed that the observed differences arise mostly from the distinct local dynamical properties of the two proteins. After fitting each curve separately, $d$ was found to be temperature-independent (fig. S1), therefore a global fitting approach has been employed, together with the assumption that the populations of the two states follow an Arrhenius behaviour. This enabled us to fit the whole temperature scans with only the MSD as a temperature-dependent parameter, plus $d$, $\Delta H$ and $\Delta S$ as temperature-independent parameters. The total MSD of the two

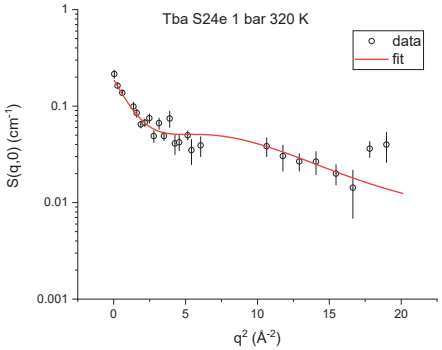


*Figure 3: Example of a two-state model fit on the scattering intensity of Tba S24e at 1 bar and 320 K.*

samples clearly follow very different pressure dependences (fig. 4): in Tba S24e it decreases going from 1 to 200 bar, and then stabilizes on very similar values up to 800 bar. In Tko S24e the same decrease from 1 to 200 bar can be seen, but it is more significant, and it further decreases from 600 to 800 bar. Besides, it can be observed that all the curves for Tba S24e maintain the same slope at all pressure values, indicating that the protein is indeed capable of keeping the same *resilience*[24,25] (i.e. flexibility) throughout the whole studied pressure range, although high pressure induces a slight decrease in the amplitude of the motions. On the other hand, for Tko S24e the slope is noticeably lower at higher pressure, showing how the protein becomes stiffer and less responsive far away from its optimal conditions. The decrease observed in both proteins at 1 bar (and at 800 bar for Tba S24e, although to a lesser extent) going from 350 to 360 K is puzzling. A reduction of the MSD in protein solutions at high temperature has been observed in the past[26], and ascribed to a slow-down of global diffusion during the aggregation of protein molecules after denaturation, a notoriously irreversible phenomenon. The proteins employed here have proven to be thermally stable at least up to 85°C (or 358K) during their purification, and the neutron scans have been checked for reversibility, therefore we exclude that denaturation is taking place in our conditions.

The other parameters extracted from the two-state model are $d$, the distance between the potential wells, which serves as an average measure of the roughness of the energy landscape of a protein[27], and the enthalpy and entropy differences $\Delta H$ and $\Delta S$. The latter reports on the hydration of the excited state in this simplified picture[23]. Figure 5a shows that $d$ is essentially pressure-independent for Tba S24e, while a sizeable decrease is evident for Tko S24e. This evidences the pressure stability of the piezophilic protein's energy landscape and the loss of flexibility of its piezosensitive counterpart due to the increased roughness of its energy landscape. Such behaviour is mirrored by the opposite trend

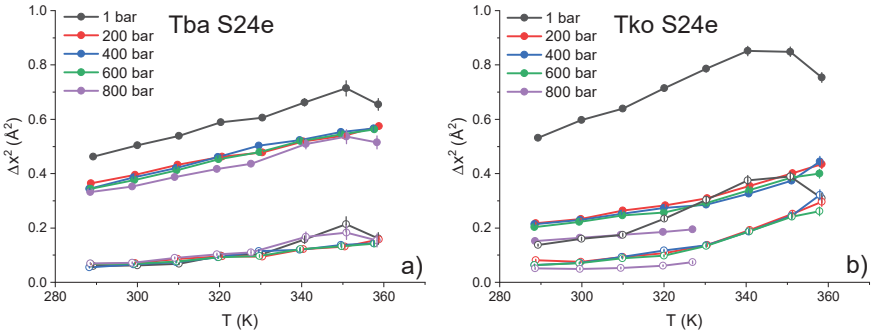


*Figure 4: MSD both inside the potential well ($\Delta x_0^2$, open circles) and total ($\Delta x_{tot}^2 = \Delta x_0^2 + \frac{p_1 p_2 d^2}{3}$, full circles) as a function of temperature and pressure for Tba S24e (a) and Tko S24e (b). $p_1$ and $p_2$ are the populations of the two potential wells, as described in Methods. Lines are a guide to the eye.*

shown by the two proteins for both $\Delta H$ and $\Delta S$ (fig. 5b and c), seemingly arising from the different interactions they have with their hydration water. The decreasing $\Delta S$ for Tba S24e evidences how the protein is able to maintain its dynamical properties at high pressure by interacting less with water, while increased hydration in Tko S24e leads to a rougher energy landscape and results in a loss of flexibility. All these results are in agreement with a previous work from our group on two different proteins belonging to the same organisms. Therefore, it seems compelling to assume that the observed behaviour may be common to pressure adapted proteins.

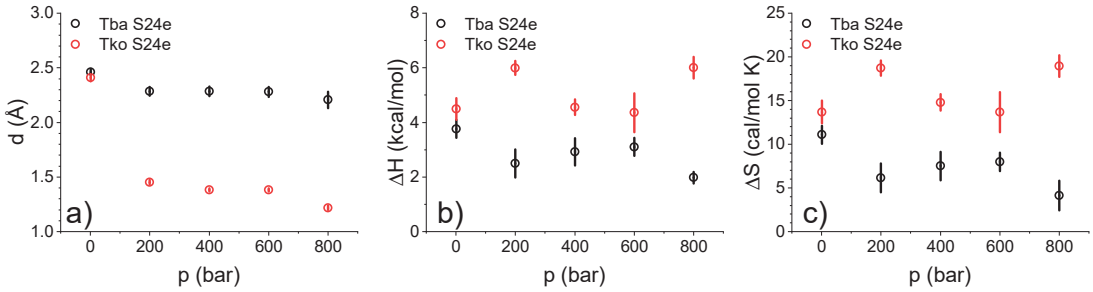


*Figure 5: Distance between the two potential wells d (a), enthalpy difference $\Delta H$ (b) and entropy difference $\Delta S$ (c) between the two states for both samples as a function of pressure. These parameters have been held temperature-independent during the global fitting procedure.*

## Nuclear Magnetic Resonance Spectroscopy

The amide groups nitrogen and proton NMR resonances of Tba S24e and Tko S24e have been assigned using [$^1$H,$^{15}$N] double-resonance 3D NMR spectroscopy (see Methods) with the classical sequential assignment strategy. $^1$H and $^{15}$N resonances have been assigned for 76% and 83% amide groups of non-proline (95 and 94) residues of Tba S24e and Tko S24e, respectively (fig. S2). Most of the amide groups that could not be assigned were not present in the Heteronuclear Single Quantum Coherence (HSQC) spectra due to extreme broadening, probably caused by motions in the intermediate exchange regime. This was the case for residues in the loop joining the strands $\beta$3-$\beta$4 in the $\beta$-sheet, or in the loop between $\beta$4 and the N-terminal of the $\alpha$-helix sitting against the $\beta$-sheet, or for residues located in the turns joining the two C-terminal short $\alpha$-helices. The inspection of the

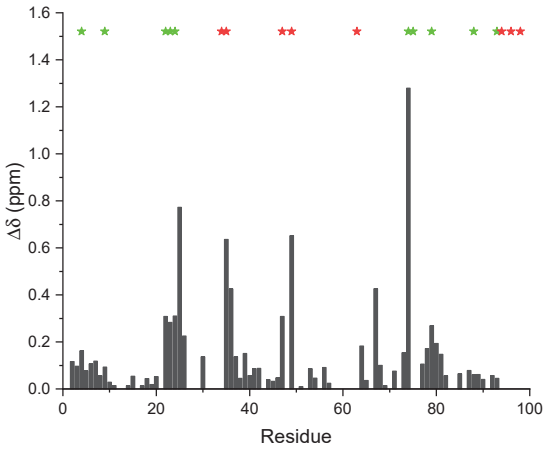


*Figure 6: Weighted $^{1}H/^{15}N$ averaged chemical shift differences measured between Tba S24e and Tko S24e. The location of the mutated side chains is indicated with stars, green for conservatives, red for non-conservative..*

weighted $^{1}H/^{15}N$ chemical shift differences $\Delta\delta$, measured between the two proteins (eq. 2) shows that chemical shift perturbations are almost exclusively restricted to the vicinity of the mutated sites (fig. 6), supporting the great similarity observed between the two X-ray structures.

### *HP-NMR: Compressibility study.*

2D [$^{1}H$,$^{15}N$] HSQC spectra of 15N uniformly labeled Tba S24e and Tko S24e were recorded as a function of pressure (fig. S3). Without denaturant, the proteins cannot be unfolded in the pressure range allowed by our experimental setup. Only pressure-dependent shifts of the native-state amide resonance frequencies are observed, due to compression and possible population of higher energy low-lying states[28,29]. These pressure dependent changes of the $^{1}H$ and $^{15}N$ amide chemical shift can be fitted well by a second order Taylor expansion (eq. 3). Of interest in this study, the origin of the amide $^{1}H$ chemical shift changes induced by pressure has been studied extensively and has been shown to be strongly correlated with the change of the hydrogen-bond lengths[28–31]. Thus, the linear downfield shifts of amide protons observed for Tba S24e and Tko S24e originate from the compression of hydrogen bonds induced by pressure, whereas the few observed upfield shifts indicate a stretching of the corresponding hydrogen bonds, due to local rearrangements of the protein structure under pressurization. Interestingly, we observed a similar profile in the amplitude and sign of the $B$ linear coefficients measured for amide protons in both proteins, even though their average ratio $\left\langle \frac{B(Tba\ S24e)}{B(Tko\ S24e)} \right\rangle = 1.1 \pm 0.5$ suggests a slightly higher compressibility for the H-Bonds in Tba S24e (fig. S4). The dependence of the $B$ linear coefficient for $^{15}N$ chemical shifts is more complex: it involves changes in hydrogen bonding as well as backbone dihedral angles and side chain orientations[32]. Again, a similar profile was observed for the $B$ values for both proteins, with slightly lower values in the case of Tba S24e ($\left\langle \frac{B(Tba\ S24e)}{B(Tko\ S24e)} \right\rangle = 1.0 \pm 0.4$). The calculated residue-specific values of the linear coefficients can also be put into context with the two proteins' sequences and structures (fig. 7a and b). The residues that are in the vicinity of the cavity in Tko S24e appear to be more compressible compared to their counterparts in Tba S24e, likely because of the larger free space they

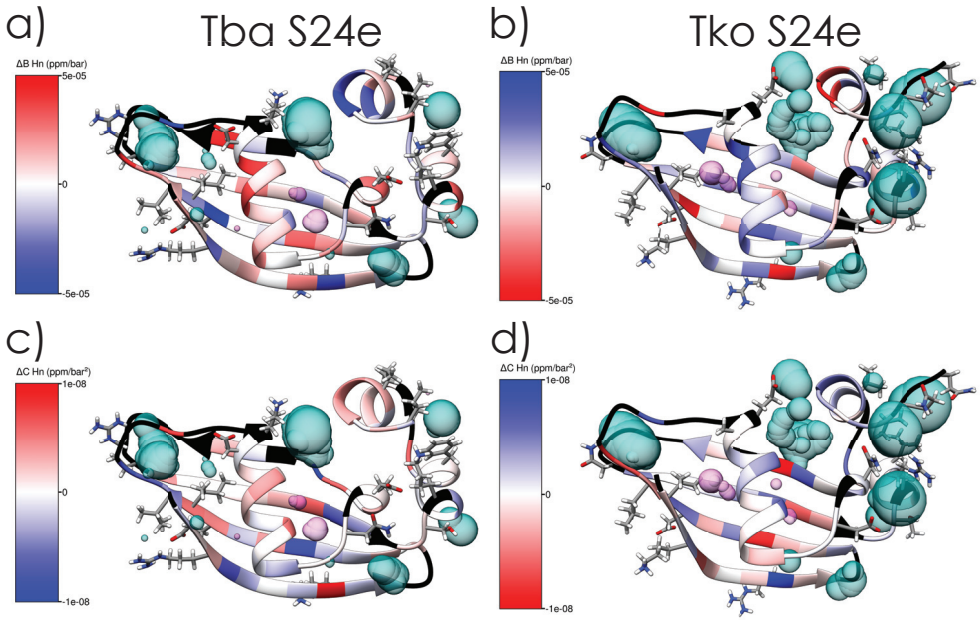


*Figure 7: Difference in $B$ (a,b) and $C$ (c,d) coefficients for amide protons between the two proteins. The calculated values are $\Delta B\ ^1H_n = B(Tba\ S24e) - B(Tko\ S24e)$ and $\Delta C\ ^1H_n = C(Tba\ S24e) - C(Tko\ S24e)$, and the color scale is inverted for Tko S24e. Therefore residues coloured in red are more compressible and residues in blue are less compressible in the respective protein. Mutated residues are shown in licorice. Internal cavities (magenta) and solvent-accessible cavities (cyan) are also shown, and they have been calculated as described in the text.*

can explore. However, the rest of the $\beta$-sheet and the central helix seem more compressible in Tba S24e, in agreement with the higher flexibility observed by neutron scattering at high pressure. In the terminal region, the residues in Tba S24e that are in contact with the rest of the protein are less compressible, indicating that the protein is more compact, while those exposed to the surface display higher compressibility. The better contact between the terminal and the main body of the protein in Tba S24e is also evidenced by the smaller surface-accessible cavity present at their interface. The quadratic coefficients $C$ for $^1$H are also reported (fig. 7c and d), as they are useful indicators of the population of low-lying excited states[28,29]. Here the differences between the two proteins are less striking, but it is nonetheless clear how the residues in the main and C-terminal helices of Tba S24e present increased non-linearity in the shifts of their resonance frequencies compared to Tko S24e. These results seem to suggest that the ability of Tba S24e to access different conformations with similar overall volume and its higher compressibility are the key to its superior pressure stability.

### *HP-NMR: Denaturation study.*

When 2D [$^1$H,$^{15}$N] HSQC spectra were recorded in the presence of 2M guanidinium hydrochloride (GuHCl), we observed a decrease in overall intensity of each native state peak as a function of pressure. Concomitantly we observed an increase in the intensity of the peaks centered around 8.5 ppm in the proton dimension, due to the population of the unfolded state. This behaviour was perfectly reversible: 1D spectra and 2D [$^1$H,$^{15}$N] HSQC spectra recorded for each protein after slow return to atmospheric pressure (500 bar steps, 1 hour relaxation time at each pressure) were perfectly superim-

posable to the spectra recorded before pressurization. These observations indicate that each residue of the proteins is in slow exchange between the chemical environments it experiences in the folded and unfolded states. Hence, the loss of intensity of the native state resonances reflects directly the decrease in population of the folded state as detected locally by each residue. This allowed us to fit the local pressure unfolding curves, obtained from the decrease in intensity of each individual peak for all resolved resonances individually to a two-state pressure-induced unfolding model as described in Methods (eq. 4, fig. S5 and S6), yielding residue specific values for the apparent volume change ($\Delta V_u = -\Delta V_f$) and apparent free energy difference ($\Delta G_u = -\Delta G_f$) of unfolding[19], from which residue specific values for half-denaturation pressure $p_{1/2}$ can be calculated (see Methods, fig. S7). A total of 66 residues, common to both proteins, gave non-overlapping cross-peaks both in the folded and unfolded state (and at all the pressure points in between), with sufficient intensity at atmospheric pressure to be accurately fitted. At 2 M GuHCl, both proteins exhibit moderate stability, with average values $\langle \Delta G_u \rangle = 3.0 \pm 0.1$ kcal/mol for Tba S24e and $3.1 \pm 0.2$ kcal/mol for Tko S24e. Similarly, average $\langle \Delta V_u \rangle$ values of $-72 \pm 3$ ml/mol and $-80 \pm 4$ ml/mol were obtained for the two proteins, respectively. Although the two proteins have similar $\Delta G$, the $p_{1/2}$ values are significantly different ($1700 \pm 30$ bar and $1620 \pm 30$ bar for Tba S24e and Tko S24e, respectively), suggesting better stability for Tba S24e. This can be attributed to lower $\Delta V$ values for Tba S24e, suggesting a better hydrophobic packing in the core of the molecule. Also, significantly lower values of $p_{1/2}$ are observed for both proteins in their C-terminal $\alpha$-helical domain, indicating that this domain may unfold at lower pressures. When looking at the local values of these two quantities (fig. 8a and b for $p_{1/2}$, c and d for $\Delta V$), the first striking result is the perfect correlation between the I23V substitution, the elimination

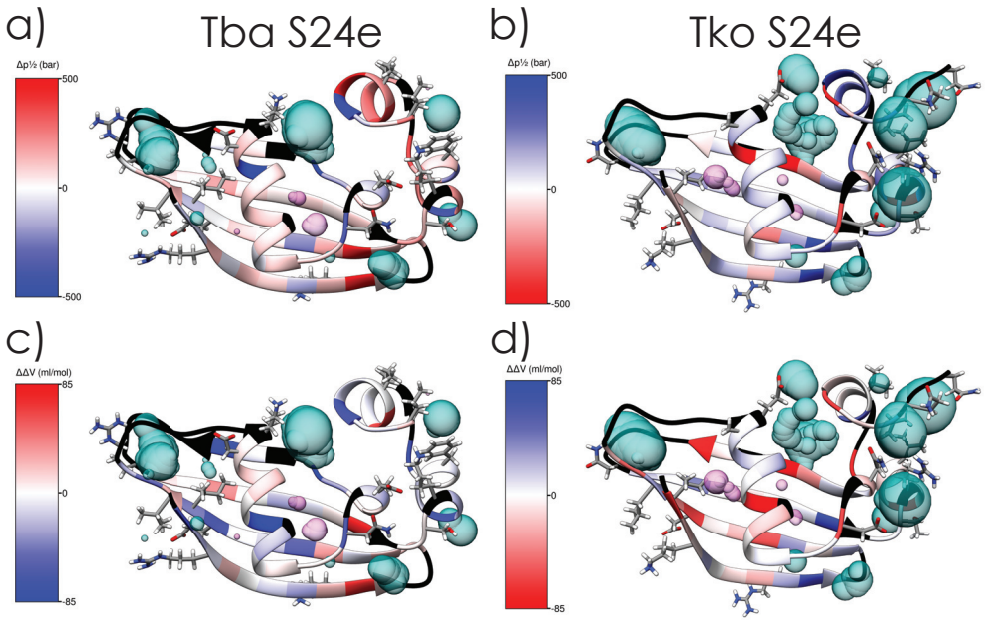


*Figure 8: Difference in $p_{1/2}$ (a, b) and $\Delta V$ (c, d) between the two proteins. The calculated values are $\Delta p_{1/2} = p_{1/2}(Tba\ S24e) - p_{1/2}(Tko\ S24e)$ and $\Delta\Delta V = \Delta V(Tba\ S24e) - \Delta V(Tko\ S24e)$, and the color scale is inverted for Tko S24e. Therefore residues coloured in red have higher $p_{1/2}$ and $\Delta V$ in the respective protein. Mutated residues are shown in licorice. Internal cavities (magenta) and solvent-accessible cavities (cyan) are also shown, and they have been calculated as described in the text.*

of the associated cavity in Tba S24e, and the strong decrease of the unfolding volume for residues 22 and 23. This constitutes the first *direct* evidence of the importance of such conservative substitutions, which were already identified in our previous work[15]. In the terminal region all residues, among those that gave a measurable signal, present a lower unfolding volume in Tba S24e compared to Tko S24e, and their superior pressure stability is also reflected in the half denaturation pressure values. Although this region is the least stable in both proteins, the substitution seems to have a measurable effect, and we speculate that this would be of major importance in the efficiency of the binding to the ribosome *in vivo*.

# Discussion

In this work we studied the structural and dynamical stability of the two Ribosomal proteins S24e from the nearly isogenic species *T. barophilus* ($p_{opt} = 400\ bar$) and *T. kodakarensis* ($p_{opt} = 1\ bar$) by neutron scattering and NMR spectroscopy. The close resemblance of the present EINS results to those previously obtained from another couple of proteins belonging to the same organisms (Phosphomannose Isomerases[15], PMIs) is a very encouraging result, as the proteins studied here bear more of a structural role, strongly interacting with other proteins and nucleic acids[21], and they are not enzymes. The PMIs are instead globular proteins existing in the cytoplasm and carrying out their enzymatic function[33], without having any noticeable interaction with other compounds except for their substrate. The higher resilience of the pressure-adapted protein (fig. 4) allows it to be flexible enough to accommodate the changes in pressure (fig. 7a), while still retaining the ability of exploring a meaningful portion of its energy landscape (fig. 5 and 7c), as shown by the EINS data and further supported by the compressibility study carried out by HP-NMR. On the other hand, the pressure-sensitive protein is less compressible, meaning that more energy is necessary to compress its bonds (fig. 7b), and this increased rigidity translates to a limited exploration of the protein's energy landscape (fig. 5 and 7d). Among the various substitutions found between the two orthologs, I23V stands out in its important contribution to the elimination of a cavity, bringing about a better packing of the protein core and substantially improving its pressure stability. This gives further confirmation that conservative substitutions can have a critical role, although they are often considered irrelevant. N47D, E49N and K75R conserve the overall charge in that region of the protein, but the displacement of the negative charge enables Glu49 to form a salt bridge with Lys75 in Tba S24e, giving stability to the terminal. F79L, I88V and L93V give a very small, but measurable, increase in the flexibility of Tba S24e's terminal, and, based on the evidence from I23V here and I100V in the case of the PMIs[15], we speculate that the bulkier residues employed by the piezophilic protein could be involved in the optimization of the hydrophobic contact surface between the protein and the ribosome. K34E, D35A and R63N impact the surface charge of the proteins, and are partly responsible for the different interaction of the two proteins with their surrounding hydration water, as evidenced by EINS. The rest of the substitutions (R4K, K9R, E22D, I24L, S74T) are conservative, and they do not appear to impact protein stability in an extensive fashion, nor do they confer an advantage to either protein. We will not comment on E94Q, K96Q and G98E, as some of those residues were either not visible in the X-ray

structures or did not give a measurable signal in the NMR experiments. The combination of methods used for this work enabled us to uncover the role of amino-acid substitutions on the pressure stability of proteins with unprecedented detail. It is now clear how the knowledge of the primary sequence does not suffice to identify the adaptation pattern to high pressure, and that structural information is of paramount importance to this end. Thanks to the astonishing progress that has been made in the field of structure prediction[34], a genome-wide protein structural analysis could indeed be carried out in the future, in order to build a robust model to interpret pressure adaptation at the genomic level.

# Acknowledgements

The authors would like to acknowledge the Institut Laue-Langevin for the allocation of beamtime to perform these experiments, and James Maurice and the whole High-Pressure Division of the SANE group for technical support during the experiments. We are grateful to Didier Nurizzo and Matthew Bowler at the European Synchrotron Radiation Facility (ESRF) for providing assistance in using the ID30A-1 and ID30B beamlines. The authors gratefully acknowledge the ESRF for provision of synchrotron radiation facilities via Block Allocation Group beamtime. The NMR platform at CBS is supported by French Infrastructure for Integrated Structural Biology (FRISBI) grant No. ANR-10-INSB-05. This work was supported by the French National Research Agency (programme ANR 17-CE11-0012-01 to P.O. and J.P.) and by the Mission pour les Initiatives Transverses et Interdisciplinaires of the CNRS (OriginsUnderPressure and LifeAdapt research projects). A.C. is supported by a PhD grant for international students by the French Ministry of Science and Technology.

# Data Availability

Data are available at http://doi.ill.fr/10.5291/ILL-DATA.8-04-886 for the IN13 experiment (8-04-886). The X-ray structures of Tba S24e and Tko S24e are available in the Protein Data Bank under accession codes 8AA8 and 8A8S respectively.

# Methods

## Protein expression and purification

Recombinant Ribosomal proteins S24e from *T. barophilus* and *T. kodakarensis* have been produced by cloning synthetic codon-optimized genes (purchased from GENEWIZ Europe) into the protein over-expression plasmid pT7-7[35] (Novagen), which was then transformed into *E. coli* BL21(DE3) pLysS strain (Novagen). 10 L cultures were grown at $37°C$ in LB medium supplemented with 100 µg/ml ampicillin until $OD_{600} = 0.5$, induced with a final concentration of 1 mM IPTG and further grown overnight at $37°C$. Cells were harvested by centrifugation at 17.000 g for 30 min, washed in isotonic solution ($0.9\%$ NaCl) and resuspended in 400 ml of 20 mM TRIS, 150 mM NaCl, pH 7.5 buffer. Cells were then lysed by 5 freeze-thaw cycles in liquid nitrogen (1 min) and at $50°C$ (3 min),

and homogenized by sonication (maximum power for 15 min at 50% duty cycle). The soluble fraction was recovered by centrifugation at 12.000 g and $4°C$ for 60 min. It was then heated to $75°C$ for 1 h to remove the non-thermostable proteins from the *E. coli* expression host. Protein debris were removed by centrifugation at 12.000 g and $4°C$ for 60 min. The extraction was repeated a second time for maximum recovery. The supernatant was concentrated to ∼20 ml by ammonium sulfate precipitation and further purified by Size Exclusion Chromatography on an AKTA® FPLC system, using an XK50-60 column packed with 1 L of Superdex® 75 Prep-Grade resin, calibrated with the GE Healthcare® Low Molecular Weight kit (fig. S8). Fractions containing the protein (fig. S9) were then pooled, concentrated by ultrafiltration (Amicon® Ultra-15 centrifugal filter units, Millipore) and lyophilised. The purity of the proteins was assessed by SDS-PAGE, and was greater than $99\%$ (fig. S10). To prepare the samples for neutron scattering, the lyophilised protein powder was gently dissolved in $D_2O$ (Sigma-Aldrich) under nitrogen atmosphere, at a concentration of 120 mg/ml. Protein solutions rather than hydrated powders were employed in order to optimally transmit hydrostatic pressure to the sample. The same batch has been used for crystallization. $^{15}N$ labelled proteins for 2D-NMR have been similarly produced: the same genes have been cloned in the pET derived plasmid pdbccdb_3C_His (kindly provided by F. Allemand, CBS Montpellier) and then transformed in the same *E. coli* strain. 500 ml cultures in M9 minimal medium, supplied with $^{15}NH_4Cl$ (Eurisotop®), were grown at $37°C$ until $OD_{600} = 0.6$, induced with a final concentration of 0.25 mM IPTG and further grown overnight at $25°C$. Cells were harvested by centrifugation at 18.000 g for 30 min, and resuspended in 140 ml of 50 mM $NaH_2PO_4$, 300 mM NaCl, pH 8 buffer containing 1 µg/ml Lysozyme (Sigma-Aldrich®). After 30 min incubation on ice, the suspension was sonicated four times for 3 min at 40% power and the soluble fraction was recovered by centrifugation at 18.000 g and $4°C$ for 20 min. The supernatant was then bound to a His-Trap® column on an AKTA® FPLC system, and then eluted in the same buffer by concentration gradient up to 500 mM Imidazole. Relevant fractions were pooled and the His-tag was cleaved by means of the HRV 3C protease (kindly provided by A. deVisch, CBS Montpellier) and then dialysed overnight against 50 mM $NaH_2PO_4$, 300 mM NaCl, pH 8 buffer. Further purification was carried out by Size Exclusion Chromatography using an XK26-60 Superdex® 75 Prep-Grade column (GE Healthcare), the relevant fractions were again pooled, dialysed overnight against 1 L of 10 mM TRIS, 300 mM NaCl, 0.1 mM PMSF, 0.1mM EDTA pH 7.4 buffer, and then concentrated up to 8.8 mg/ml (0.8 mM) and stored at $-80°C$ until the NMR experiments.

## X-ray Crystallography

The purified proteins were diluted to 10 mg/ml in 10 mM Tris-HCl pH 8, 300 mM NaCl. Crystallization trials were performed at 20°C using the hanging-drop vapour-diffusion method in 96-well micro-plates and a Mosquito robot (TPLabtech) with 100 nl of protein mixed with 100 nl of reservoir. Tko S24e crystals were obtained after one week from condition E8 of the Structure Screen 1+2 kit (Molecular Dimensions), containing 0.2 M ammonium phosphate monobasic, 0.1 M TRIS pH 8.5, and 50% (v/v) MPD. Tba S24e crystals were obtained after two weeks from condition F7 of the Morpheus kit (Molecular Dimensions), containing 0.12 M monosaccharides mix, 0.1 M buffer system 2 pH 7.5, and 30% (v/v) precipitant mix 3. Crystals were flash-frozen in liquid nitrogen. X-ray diffrac-

tion datasets were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble) at the ID30A-1 beam line (Massif1[36–40]), using a pixel detector (PILATUS3 2M) and auto-processed by XDSAPP package[41]. Tba S24e crystals belong to the $P_1$ space group and contain two molecules in the asymmetric unit, while Tko PMI crystals belong to the $P_{21}$ space group and contain one molecule in the asymmetric unit. The structures of both proteins were determined by molecular replacement with Phaser[42] from the Phenix package[43], using models from the Alphafold 2 server[34]. After model building using Coot[44] and refinement by Refine[45] from the Phenix package, the final structures exhibited an R(%)/R(%)$_{free}$ of 0.18/0.22 at 1.74Å (Tba S24e) and 0.22/0.27 at 1.66 Å (Tko S24e). Final refinement statistics for the structures are listed in Table S1 and S2. The atomic coordinates and structure factors of Tba S24e and Tko S24e, have been deposited in the Protein Data Bank with accession numbers 8AA8 and 8A8S, respectively.

## Elastic incoherent neutron scattering (EINS)

EINS measurements were performed on the IN13 backscattering spectrometer at the Institut Laue-Langevin (ILL, Grenoble, France). At the elastic position, IN13 has an incident wavelength of $2.23$ Å and a nearly $q$-independent resolution of $8\ \mu eV$ FWHM, which gives a time window of $\sim 100\ ps$[46], allowing to probe local motions of hydrogen atoms since their incoherent scattering cross section is an order of magnitude larger than that of other isotopes[22]. Temperature was controlled by means of a closed-cycle dry cryofurnace (Displex+), and continuous up-scans were performed in the $283\ K$ to $363\ K$ range at $0.08\ K/min$. The scattering intensity was also measured while the temperature was lowered back to $283\ K$ before the next pressure point to check for hysteresis and, once its absence was verified, the downscans were merged with the upscans to improve statistics. HHP was transmitted to the sample by means of the high-pressure stick, cell and controller developed by the SANE team at ILL[47], and five pressure points were investigated (1, 200, 400, 600 and 800 $bar$). The high-pressure cell is cylindrical and made of a high-tensile aluminium alloy (7026) and has a $6\ mm$ internal diameter[48]. A piston separates the pressure-transmitting liquid (Fluorinert™ FC-770[49]) from the sample, and a cylindrical aluminium insert ($4\ mm$ diameter) was used to decrease sample volume and to minimize multiple scattering. During the $800\ bar$ measurement for Tko S24e, the cell suffered a breakage and the time left in our beamtime was not enough to repeat the measurement, therefore it was not possible to cover the entire temperature range for this measurement. Raw data were corrected for transmission, empty cell and $D_2O$ scattering, normalized to a vanadium standard and then binned in temperature in $10\ K$ intervals using the LAMP[50] software available at ILL. EINS data have been interpreted in the framework of the *two-state model*,[23] which describes hydrogen atoms motions as a combination of vibrations in two harmonic potential wells, giving rise to the Debye-Waller contribution with the associated Mean Square Displacement (MSD) $\Delta x_0^2$, and jumps between them. The wells are separated by a distance $d$ and have a free-energy difference $\Delta G$ that can be separated into the enthalpic and entropic contributions according to $\Delta G = \Delta H - T\Delta S$. The elastic scattering function $S(q, \omega = 0)$ as a function of the scattering vector $q$ (related to the scattering angle $\theta$ and the

neutron's wavelength $\lambda$ according to $q = \frac{4\pi}{\lambda} sin(\frac{\theta}{2})$) thus reads:

$$S(q,0) = e^{-\Delta x_0^2 q^2} \left[ 1 - 2p_1 p_2 \left( 1 - \frac{\sin(qd)}{qd} \right) \right], \quad (1)$$

where $p_1$ and $p_2$ represent the population of each well, assumed in our case to follow the Arrhenius law ($p_1/p_2 = exp(-\Delta H/RT + \Delta S/R)$, where $R$ is the gas constant). It must be stressed, however, that in the investigated temperature range, large-scale motions could enter the experimental window. It is thus desirable to view the two wells as an average representation of the protein's free-energy landscape that is accessible at each temperature and pressure value.

## Nuclear Magnetic Resonance Spectroscopy

### *NMR assignments of the HSQC spectra*

Protein samples were dissolved in 200 µL of aqueous buffer containing 20 mM Tris-HCl pH 7.5, 300 mM NaCl and 5% D2O for the lock, at a concentration of about 1 mM. Experiments were recorded at 40°C on a Bruker AVANCE III 800 MHz equipped with a 5 mm Z-gradient TCI cryogenic probe head. [$^1$H,$^{15}$N] HSQC spectra were assigned with the help of 3D [$^1$H,$^{15}$N] NOESY-HSQC (mixing time 150 ms) and TOCSY-HSQC (isotropic mixing: 60 ms) NMR double-resonance experiments, using the sequential attribution strategy. Water suppression was achieved with the WATERGATE sequence[51]. $^1$H chemical shifts were directly referenced to the methyl resonance of DSS, while $^{15}$N chemical shifts were referenced indirectly to the $^{15}$N/$^1$H = 0.101329118 absolute frequency ratios. All NMR experiments were processed with Gifa[52]. The weighted $^{15}$N/$^1$H averaged chemical shift difference between Tba S24e and Tko S24e were calculated as:

$$\Delta\delta = \sqrt{\Delta\delta(^1H)^2 + \left(\frac{1}{5}\Delta\delta(^{15}N)\right)^2} \quad (2)$$

### *HP-NMR: Compressibility study.*

In order to measure the pressure linear dependence of amide chemical shifts, series of 2D [$^1$H,$^{15}$N] HSQC spectra were recorded on proteins dissolved in the buffer described above at 40°C and 10 different hydrostatic pressures (1, 100, 400, 700, 1000, 1300, 1600, 1900, 2200, and 2500 bar). High-Pressure (HP) NMR spectra were acquired on a Bruker AVANCE III 600 MHz equipped with a 5 mm Z-gradient BBI probe head, using a 5 mm o.d. ceramic tube from Daedelus Innovations (Aston, PA, USA). Hydrostatic pressure was applied to the sample directly within the magnet using the Xtreme Syringe Pump also from Deadelus Innovations. The measured $^1$H and $^{15}$N chemical shift were then fitted to a second order Taylor expansion:

$$\delta(p) = \delta_0(p_0) + B(p - p_0) + C(p - p_0)^2 \quad (3)$$

where $\delta_0(p_0)$ is the chemical shift at atmospheric pressure $p_0$ of 1 bar, and $B$ and $C$ the first (linear) and second (quadratic) order pressure coefficients[53].

### *HP-NMR: Denaturation study.*

To permit unfolding of the protein within our accessible pressure range (1-2500 bar), 2 M guanidinium chloride was added to the NMR samples. Reliable assignment of the amide cross-peaks in the HSQC spectra at the denaturant concentration used in the high-pressure experiments was ensured by guanidinium chloride titration. Series of 2D [$^1$H,$^{15}$N] HSQC spectra were recorded at 40°C and 15 different hydrostatic pressures (1, 50, 100, 300, 500, 700, 900, 1100, 1300, 1500, 1700, 1900, 2100, 2300 and 2500 bar), with a 2 hours relaxation time after every pressure change, to allow the protein to reach full equilibrium. Relaxation times for the folding/unfolding reaction were previously obtained from series of 1D NMR experiments recorded after 200 bar P-Jump, following the increase of the resonance band corresponding to the methyl groups in the unfolded state of the protein[20]. The cross-peak intensities for the folded species were measured at each pressure, then fitted with a two-state model:

$$I(p) = \frac{I_u + I_f e^{-\frac{\Delta G_f^0 + p\Delta V_f^0}{RT}}}{1 + e^{-\frac{\Delta G_f^0 + p\Delta V_f^0}{RT}}} \tag{4}$$

where $I(p)$ is the cross-peak intensity measured at a given pressure, $I_f$ and $I_u$ corresponds to the cross-peak intensity in the folded state (1 bar) and in the unfolded state (2500 bar), respectively. $\Delta V_f^0$ and $\Delta G_f^0$ stands for the residue specific apparent volume change at equilibrium and free energy at atmospheric pressure, respectively. $I_f$, $I_u$, $\Delta V_f^0$ and $\Delta G_f^0$ values were used as floating parameters in the fit. The residue-specific half-denaturation pressures were calculated from the obtained values of $\Delta V_f^0$ and $\Delta G_f^0$ as $p_{1/2} = \frac{\Delta G_f^0}{\Delta V_f^0}$.

## Molecular Dynamics simulations

Molecular Dynamics (MD) simulations have been performed with NAMD 2.14[54] on the P2CHPD computing centre of the Université Claude Bernard Lyon 1. The system was prepared using the CHARMM36m force field on the CHARMM-GUI server[55] by building a cubic water box and leaving a 15 Å padding on each side of the protein. The system was neutralized with $150\ mM$ NaCl, resulting in a total content of 12642 $H_2O$ molecules, 36 $Na^+$ ions and 37 $Cl^-$ ions for Tba PMI, and 13780 $H_2O$ molecules, 39 $Na^+$ ions and 39 $Cl^-$ ions for Tko PMI. The system was first minimized for 20000 steps, then gradually heated to 310 K by reassigning the velocities every step for 31000 steps (0.01 K/step) in the NVE ensemble, then equilibrated in the NVT ensemble for 300 ps with a Langevin thermostat (2 $ps^{-1}$ damping coefficient)and a 2 fs time step, and finally left to evolve in the NPT ensemble at 1 bar for 10 ns (2 fs time step) by employing the modified Nosé-Hoover Langevin barostat implemented in NAMD[56], with a 200 fs period and a decay time of 100 fs. Periodic boundary conditions were applied, with particle-mesh Ewald long range electrostatics, using a grid spacing of 1 Å along with a sixth order B-spline charge interpolation scheme. A 12 Å cut-off was used for non-bonded interactions, with a smooth switching function starting at 10 Å. Bonds were constrained

using the SETTLE algorithm, and coordinates were output every 10 ps. Analysis of the trajectories was carried out on VMD[57] and, after assessing the stability of the system (fig. S12), the last 5 ns were used to calculate the average coordinates of each atom. This clearly does not correspond to a physical state of the system, therefore, the whole trajectory was aligned with the average structure and the frame with the lowest RMSD (calculated using the $C_\alpha$ coordinates) was chosen as the most representative physical state of the system. The resulting structures were used for the calculation of cavities using the CastP server[58]. All the images have been generated using UCSF Chimera 1.16[59].

# References


[1] Jannasch, H. W. & Taylor, C. D. Deep-sea microbiology. *Annual review of microbiology* **38**, 487–487 (1984).

[2] Oger, P. M. & Jebbar, M. The many ways of coping with pressure. *Research in Microbiology* **161**, 799–809 (2010).

[3] Miller, S. L. & Bada, J. L. Submarine hot springs and the origin of life. *Nature* **334**, 609–611 (1988).

[4] Martin, W., Baross, J., Kelley, D. & Russell, M. J. Hydrothermal vents and the origin of life. *Nature Reviews Microbiology* **6**, 805–814 (2008).

[5] Auerbach, G. *et al.* Lactate dehydrogenase from the hyperthermophilic bacterium thermotoga maritima: the crystal structure at 2.1 å resolution reveals strategies for intrinsic protein stabilization. *Structure* **6**, 769–781 (1998).

[6] Price, P. B. A habitat for psychrophiles in deep antarctic ice. *Proceedings of the National Academy of Sciences* **97**, 1247–1251 (2000).

[7] Brininger, C., Spradlin, S., Cobani, L. & Evilia, C. The more adaptive to change, the more likely you are to survive: Protein adaptation in extremophiles. *Seminars in Cell and Developmental Biology* **84**, 158 – 169 (2018).

[8] Peoples, L. M. *et al.* Distinctive gene and protein characteristics of extremely piezophilic colwellia. *BMC Genomics* **21**, 692 (2020).

[9] Peters, J. *et al.* Deep sea microbes probed by incoherent neutron scattering under high hydrostatic pressure. *Zeitschrift für Physikalische Chemie* **228**, 1121 – 1133 (2014).

[10] Martinez, N. *et al.* High protein flexibility and reduced hydration water dynamics are key pressure adaptive strategies in prokaryotes. *Scientific Reports* **6** (2016).

[11] Golub, M. *et al.* The effect of crowding on protein stability, rigidity, and high pressure sensitivity in whole cells. *Langmuir* **34**, 10419–10425 (2018).

[12] Salvador-Castell, M. *et al.* The first study on the impact of osmolytes in whole cells of high temperature-adapted microorganisms. *Soft Matter* **15**, 8381–8391 (2019).

[13] Marteinsson, V. *et al.* Thermococcus barophilus sp. nov., a new barophilic and hyperthermophilic archaeon isolated under high hydrostatic pressure from a deep-sea hydrothermal vent. *International Journal of Systematic and Evolutionary Microbiology* **49**, 351–359 (1999).

[14] Atomi, H., Fukui, T., Kanai, T., Morikawa, M. & Imanaka, T. Description of thermococcus kodakaraensis sp. nov., a well studied hyperthermophilic archaeon previously reported as pyrococcus sp. kod1. *Archaea* **1**, 204953 (1900).

[15] Caliò, A. *et al.* Unravelling the adaptation mechanisms to high pressure in proteins. *International Journal of Molecular Sciences* **23** (2022).

[16] Ishima, R. & Torchia, D. A. Protein dynamics from nmr. *Nature Structural Biology* **7**, 740–743 (2000).

[17] Bax, A. & Clore, G. M. Protein nmr: Boundless opportunities. *Journal of Magnetic Resonance* **306**, 187–191 (2019).

[18] Royer, C. A. Revisiting volume changes in pressure-induced protein unfolding. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1595**, 201–209 (2002).

[19] Roche, J. *et al.* Cavities determine the pressure unfolding of proteins. *Proceedings of the National Academy of Sciences* **109**, 6945–6950 (2012).

[20] Dubois, C., Herrada, I., Barthe, P. & Roumestand, C. Combining high-pressure perturbation with nmr spectroscopy for a structural and dynamical characterization of protein folding pathways. *Molecules* **25** (2020).

[21] Armache, J.-P. *et al.* Promiscuous behaviour of archaeal ribosomal proteins: Implications for eukaryotic ribosome evolution. *Nucleic Acids Research* **41**, 1284–1293 (2012).

[22] Sears, V. F. Neutron scattering lengths and cross sections. *Neutron News* **3**, 26–37 (1992).

[23] Doster, W., Cusack, S. & Petry, W. Dynamical transition of myoglobin revealed by inelastic neutron scattering. *Nature* **337**, 754–756 (1989).

[24] Zaccai, G. How soft is a protein? A protein dynamics force constant measured by neutron scattering. *Science* **288**, 1604–1607 (2000).

[25] Bicout, D. J. & Zaccai, G. Protein flexibility from the dynamical transition: A force constant analysis. *Biophysical Journal* **80**, 1115–1123 (2001).

[26] Grimaldo, M. *et al.* Hierarchical molecular dynamics of bovine serum albumin in concentrated aqueous solution below and above thermal denaturation. *Phys. Chem. Chem. Phys.* **17**, 4645–4655 (2015).

[27] Librizzi, F., Carrotta, R., Peters, J. & Cupane, A. The effects of pressure on the energy landscape of proteins. *Scientific reports* **8**, 2037 (2018).

[28] Akasaka, K. & Li, H. Low-lying excited states of proteins revealed from nonlinear pressure shifts in 1h and 15n nmr. *Biochemistry* **40**, 8665–8671 (2001).

[29] Kitahara, R., Hata, K., Li, H., Williamson, M. P. & Akasaka, K. Pressure-induced chemical shifts as probes for conformational fluctuations in proteins. *Progress in Nuclear Magnetic Resonance Spectroscopy* **71**, 35–58 (2013).

[30] Wagner, G., Pardi, A. & Wuethrich, K. Hydrogen bond length and proton nmr chemical shifts in proteins. *Journal of the American Chemical Society* **105**, 5948–5949 (1983).

[31] Sitkoff, D. & Case, D. A. Theories of chemical shift anisotropies in proteins and nucleic acids. *Progress in Nuclear Magnetic Resonance Spectroscopy* **32**, 165–190 (1998).

[32] La Penna, G., Mori, Y., Kitahara, R., Akasaka, K. & Okamoto, Y. Modeling 15n nmr chemical shift changes in protein backbone with pressure. *The Journal of Chemical Physics* **145**, 085104 (2016).

[33] Slein, M. W. Phosphomannose isomerase. *Journal of Biological Chemistry* **186**, 753–761 (1950).

[34] Jumper, J. *et al.* Highly accurate protein structure prediction with alphafold. *Nature* **596**, 583–589 (2021).

[35] Tabor, S. & Richardson, C. C. A bacteriophage t7 rna polymerase/promoter system for controlled exclusive expression of specific genes. *Proceedings of the National Academy of Sciences* **82**, 1074–1078 (1985).

[36] Svensson, O., Gilski, M., Nurizzo, D. & Bowler, M. W. Multi-position data collection and dynamic beam sizing: recent improvements to the automatic data-collection algorithms on MASSIF-1. *Acta Crystallographica Section D* **74**, 433–440 (2018).

[37] Svensson, O., Malbet-Monaco, S., Popov, A., Nurizzo, D. & Bowler, M. W. Fully automatic characterization and data collection from crystals of biological macromolecules. *Acta Crystallographica Section D* **71**, 1757–1767 (2015).

[38] Bowler, M. W. *et al.* MASSIF-1: a beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules. *Journal of Synchrotron Radiation* **22**, 1540–1547 (2015).

[39] Bowler, M. W., Svensson, O. & Nurizzo, D. Fully automatic macromolecular crystallography: the impact of massif-1 on the optimum acquisition and quality of data. *Crystallography Reviews* **22**, 233–249 (2016).

[40] Nurizzo, D. *et al.* RoboDiff: combining a sample changer and goniometer for highly automated macromolecular crystallography experiments. *Acta Crystallographica Section D* **72**, 966–975 (2016).

[41] Kabsch, W. *XDS*. *Acta Crystallographica Section D* **66**, 125–132 (2010).

[42] McCoy, A. J. *et al. Phaser* crystallographic software. *Journal of Applied Crystallography* **40**, 658–674 (2007).

[43] Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallographica Section D* **75**, 861–877 (2019).

[44] Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta Crystallographica Section D* **66**, 486–501 (2010).

[45] Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with *phenix.refine*. *Acta Crystallographica Section D* **68**, 352–367 (2012).

[46] Natali, F. *et al.* IN13 backscattering spectrometer at ILL: Looking for motions in biological macromolecules and organisms. *Neutron News* **19**, 14–18 (2008).

[47] Lelièvre-Berna, E. *et al.* 700 MPa sample stick for studying liquid samples or solid-gas reactions down to 1.8 K and up to 550 K. *Journal of Neutron Research* **19**, 77–84 (2017).

[48] Peters, J. *et al.* High hydrostatic pressure equipment for neutron scattering studies of samples in solutions. *High Pressure Research* **32**, 97–102 (2012).

[49] Sidorov, V. A. & Sadykov, R. A. Hydrostatic limits of fluorinert liquids used for neutron and transport studies at high pressure. *Journal of Physics: Condensed Matter* **17**, S3005–S3008 (2005).

[50] Richard, D., Ferrand, M. & Kearley, G. J. Analysis and visualisation of neutron-scattering data. *Journal of Neutron Research* **4**, 33–39 (1996).

[51] Piotto, M., Saudek, V. & Sklenář, V. Gradient-tailored excitation for single-quantum nmr spectroscopy of aqueous solutions. *Journal of Biomolecular NMR* **2**, 661–665 (1992).

[52] Pons, J.-L., Malliavin, T. E. & Delsuc, M. A. Gifa v. 4: A complete package for nmr data set processing. *Journal of Biomolecular NMR* **8**, 445–452 (1996).

[53] Arnold, M. R., Kremer, W., Lüdemann, H.-D. & Kalbitzer, H. R. 1h-nmr parameters of common amino acid residues measured in aqueous solutions of the linear tetrapeptides gly-gly-x-ala at pressures between 0.1 and 200 mpa. *Biophysical Chemistry* **96**, 129–140 (2002).

[54] Phillips, J. C. *et al.* Scalable molecular dynamics on cpu and gpu architectures with namd. *The Journal of Chemical Physics* **153**, 044130 (2020).

[55] Jo, S., Kim, T., Iyer, V. G. & Im, W. Charmm-gui: A web-based graphical user interface for charmm. *Journal of Computational Chemistry* **29**, 1859–1865 (2008).

[56] Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The langevin piston method. *The Journal of Chemical Physics* **103**, 4613–4621 (1995).

[57] Humphrey, W., Dalke, A. & Schulten, K. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics* **14**, 33–38 (1996).

[58] Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Research* **46**, W363–W367 (2018).

[59] Pettersen, E. F. *et al.* Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**, 1605–1612 (2004).

# How do proteins cope with pressure?

**Antonino Caliò[1], Cécile Dubois[2], Alessio De Francesco[3,4], François Hoh[2], Philippe Barthe[2], Grazia Cottone[5], Christian Roumestand[2], Judith Peters[4,6,7*], Philippe Oger[1*]**

[1]*Univ Lyon, UCBL, INSA Lyon, CNRS, MAP UMR5240, 69622 Villeurbanne, France,*

[2]*CBS, INSERM U1054, CNRS UMR 5048, Univ Montpellier, 34090 Montpellier, France,*

[3]*CNR-IOM & INSIDE@ILL c/o Operative Group in Grenoble (OGG), 38042 Grenoble, France,*

[4]*Institut Laue Langevin, 38042 Grenoble, France,*

[5]*Department of Physics and Chemistry-Emilio Segrè, University of Palermo, Palermo 90128, Italy,*

[6]*Univ Grenoble Alpes, CNRS, LiPhy, 38400 Grenoble, France,*

[7]*Institut Universitaire de France, 75231 Paris, France*

*corresponding authors: philippe.oger@insa-lyon.fr, jpeters@ill.fr

# Supplementary Informations

## EINS Analysis

EINS data were corrected for empty cell, solvent scattering and detector efficiency (measured by means of a vanadium standard) by taking into account their transmission and the solvent volume fraction, according to the following expression:

$$S_{corr}(q,\omega) = \frac{(\frac{1}{t_{sample}}S_{sample} - \frac{1}{t_{cell}}S_{cell}) - \phi(\frac{1}{t_{D2O}}S_{D2O} - \frac{1}{t_{cell}}S_{cell})}{\frac{1}{t_{vana}}S_{vana}} \quad (1)$$

where $t$ denotes the transmission, and the $(q,\omega)$ dependence has been dropped for clarity. $\phi$ denotes the solvent volume fraction in the samples. The transmission of the samples, directly measured on IN13, was 89% for Tba S24e and 88% for Tko 24e, we could thus reasonably assume that multiple scattering effects are negligible. Figure S1 shows the temperature independence of $d$ when fitting the IN13 data with the two-state model[1], thus justifying the global fitting approach with $d$ as a shared parameter.

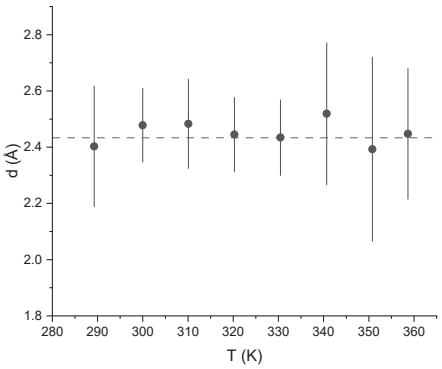


*Figure S 1: Temperature independence of the parameter d in the two state model.*

# NMR Supplementary Figures

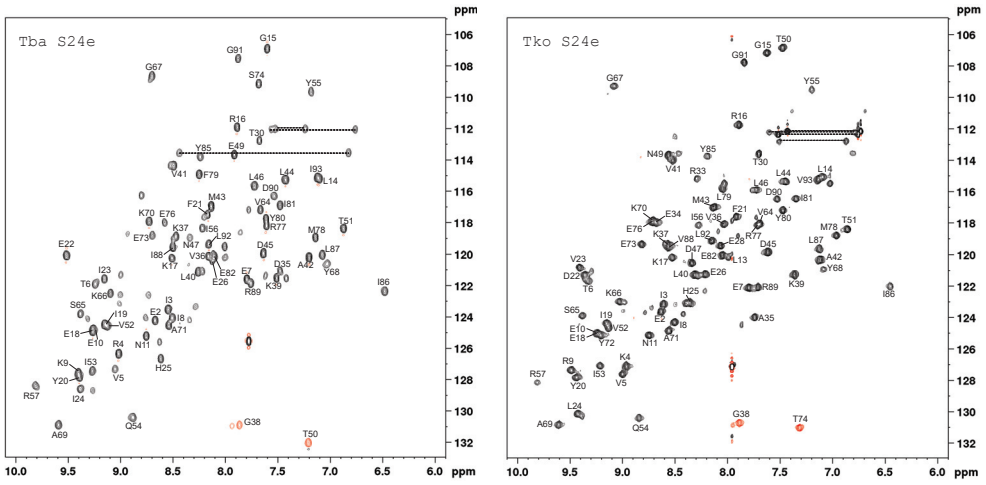


*Figure S 2: 2D [$^{1}$H-$^{15}$N] HSQC spectra of Tba S24e (Left) and Tko S24e (Right) recorded at 800 MHz on $^{15}$N-uniformly labeled samples dissolved in 20 mM Tris-HCl buffer pH 7.5, 300 mM NaCl, at 40°C. Cross peak assignments are indicated using the one-letter amino acid and number code from the PDB. Cross-peaks coloured in red indicate frequency aliasing in the nitrogen dimension.*

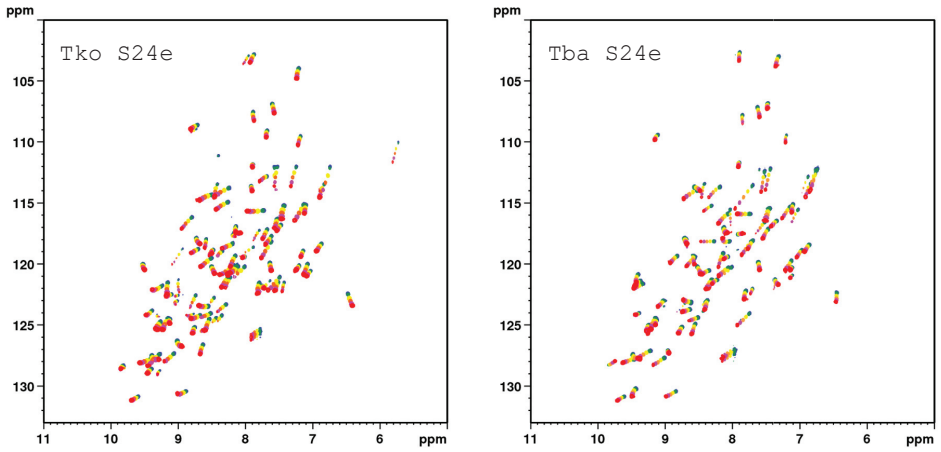


*Figure S 3: Overlay of 2D [$^{1}$H-$^{15}$N] HSQC spectra of Tba S24e (Left) and Tko S24e (Right) recorded at increasing pressure: 1 bar (blue), 100 bar (green), 700 bar (yellow), 1300 bar (orange), 1900 bar (pink) and 2500 bar (red).*

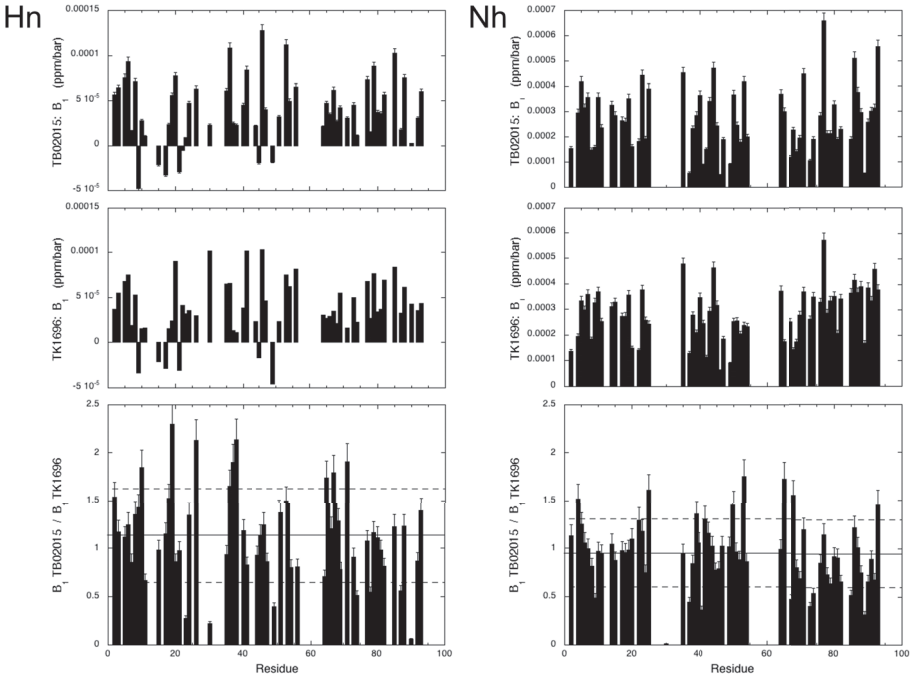


*Figure S 4: Linear coefficient $B_1$ obtained from the fit to a second order polynomial function (eq. 3 in the main text) of the $^{1}H$ (left panels) and $^{15}N$ (right panels) chemical shift evolution as a function of pressure for Tba S24e (TB02015) and Tko S24e (TK1696). The bottom panels report the ratio of the $B_1$ coefficients between the two proteins. The solid line stands for the average values, the dotted lines for the average value ± one standard deviation.*

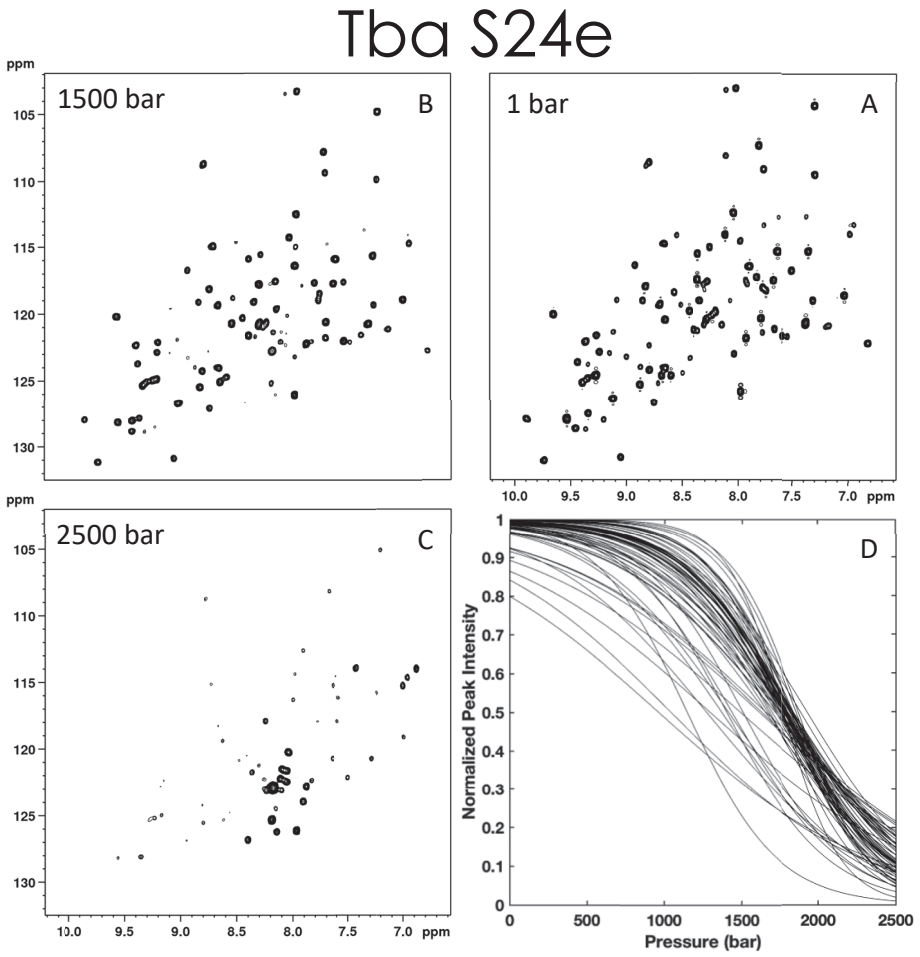


*Figure S 5: NMR detected high pressure unfolding of Tba S24e at 40°C and 2 M guanidinium hydrochloride. (A-C) Examples of [$^{1}$H-$^{15}$N] HSQC NMR spectra at different pressures as indicated; (D) Overlay of the normalized residue-specific denaturation curves obtained from the fits of the pressure-dependent sigmoidal decrease of the native residue cross-peak intensities in the HSQC spectra with eq. 4.*

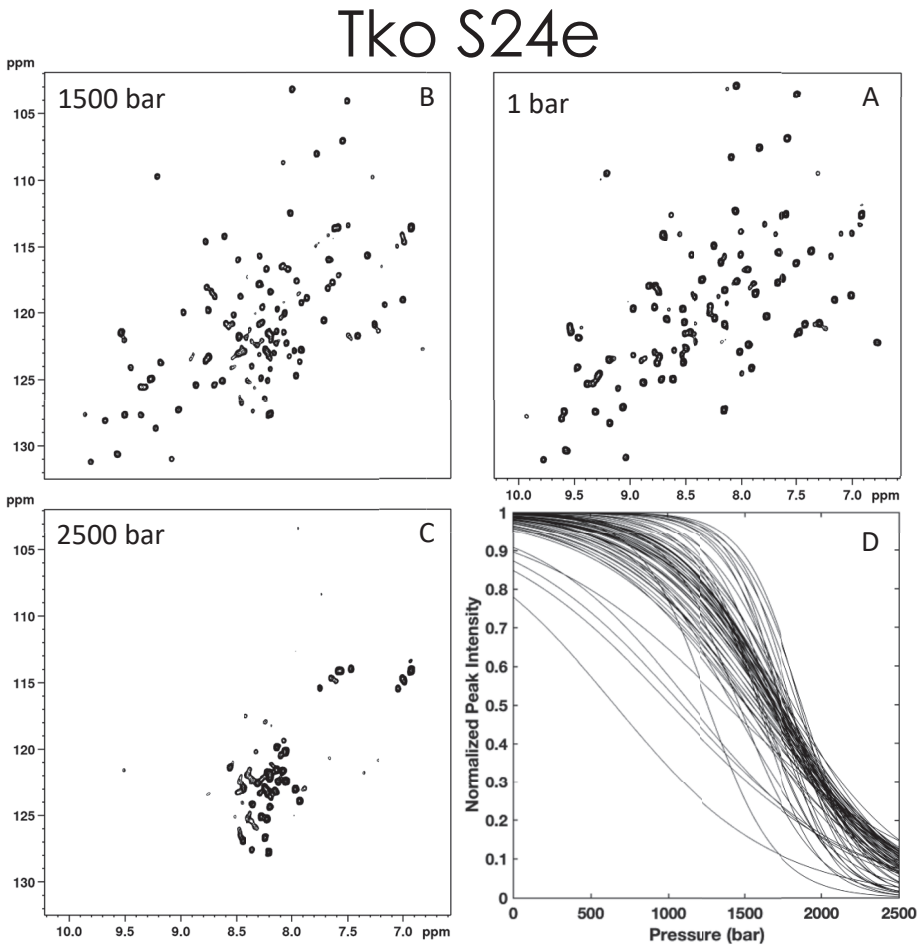


*Figure S 6: NMR detected high pressure unfolding of Tba S24e at 40°C and 2 M guanidinium hydrochloride. (A-C) Examples of [$^{1}$H-$^{15}$N] HSQC NMR spectra at different pressures as indicated; (D) Overlay of the normalized residue-specific denaturation curves obtained from the fits of the pressure-dependent sigmoidal decrease of the native residue cross-peak intensities in the HSQC spectra with eq. 4.*

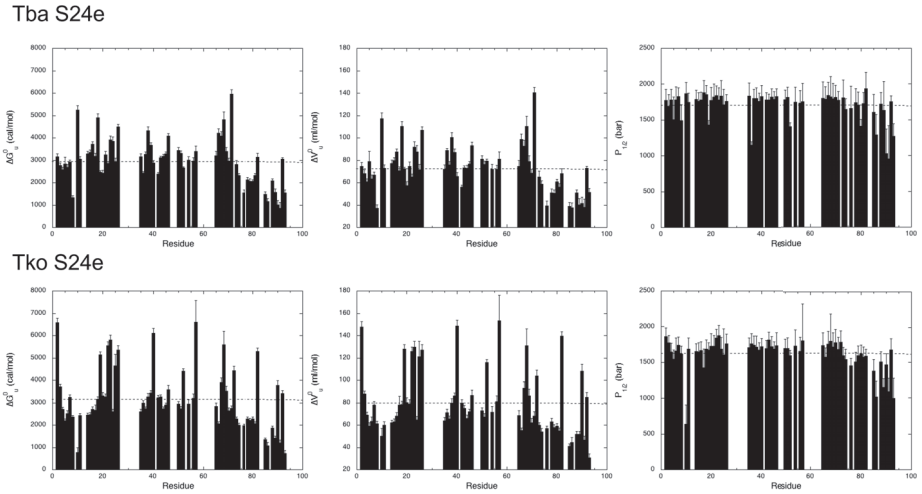


*Figure S 7: Steady-state thermodynamic parameters obtained for (Top) Tba S24e and (Bottom) Tko S24e. Plots versus the protein sequence of apparent residue-specific $\Delta V_f^0$ (central panels), $\Delta G_f^0$ (left panels), obtained through the fit of the intensity decrease of the 2D [$^1$H-$^{15}$N] HSQC native cross-peaks with pressure, and $p_{1/2}$ (right panels) values obtained from the ratio $\Delta G_f^0/\Delta V_f^0$. The solid lines stand for the average values.*

## Genes and proteins sequences

The complete genomes of *Thermococcus barophilus* MP and *Thermococcus kodakarensis* KOD1 are available on GenBank under the accession codes CP002372[2] and AP006878[3] respectively. The two genes are identified as TERMP_02015 (Tba S24e) and TK_01696 (Tko S24e). The codon-optimized sequences used in this work to express the two proteins are reported here:

Tba S24e (TERMP_02015)

GCTGAACATATGGAAATTCGTGTTACCGAAATCAAGGAAAACAAACTGC
TGGGTCGTAAAGAAATTTACTTTGAGATTATCCACGAAGGTGAACCTAC
CCCGAGCCGTAAAGACGTTAAGGGTAAACTGGTAGCGATGTTAGACCTG
AACCCTGAAACCACGGTAATCCAATACATCCGTTCCTACTTCGGTAGCC
GTGTTTCTAAAGGTTATGCGAAGGCTTATGAATCTAAAGAACGTATGTT
CTACATCGAACCGGAATACATTCTGATCCGCGACGGTCTGATTGAAAAG
AAAGAGGGTGAATGAGGATCCTTCC

Tko S24e (TK_01696)

GGCTGAACATATGGAAATCAAGGTTACCGAGATTCGCGAAAACAAACTGC
TGGGTCGTAAAGAGATCTATTTTGATGTCCTTCACGAAGGTGAGCCGAC
CCCGTCTCGCGAGGCGGTTAAAGGCAAACTGGTGGCTATGCTGGATCTC
GACCCGAATACCACCGTTATCCAATATATTCGCAGCTACTTCGGTTCTA
ACGTTTCTAAAGGTTACGCCAAGGCCTACGAAACTCGTGAGCGTATGCT
TTACATCGAACCGGAATACATCCTCGTCCGTGATGGTCTGGTTCAGAAA
CAGGAAGAATGAGGATCCTTCC

The restriction sites for *NdeI* (blue) and *BamHI* (red) have been indicated. The genes were received from GENEWIZ in pUC57-Kan plasmid, digested with *NdeI* and *BamHI*-HF® (New England Bio-

labs) and then ligated into the Multi Cloning Site of the pET-16b plasmid (digested with the same enzymes) by means of T4 DNA Ligase (Thermo Fisher). All construct propagation was done using NEB-5$\alpha$® *E. coli* cells (New England Biolabs). Plasmids were then extracted, purified (Macherey-Nagel NucleoSpin® Plasmid kit) and sequenced to verify the constructs, then transformed into the BL21(DE3) pLysS strain for expression. All the listed products have been employed according to the standard protocols issued by the manufacturers.

## Protein production and purification

The proteins have been expressed and purified as reported in the Methods section in the main text. Here we report the calibration run for the XK50-60 column using the GE Healthcare® Low Molecular Weight kit (fig. S8), which consists of five proteins: Conalbumin (75 kDa), Ovalbumin (44 kDa), Carbonic Anhydrase (29 kDa), Ribonuclease A (13.7 kDa) and Aproritin (6.5 kDa). The void volume of the column was determined to be $360\ ml$ by using the Blue Dextran 2000 supplied with the kit. The kit was used according to the manufacturer's protocol. Figure S9 reports an example purification

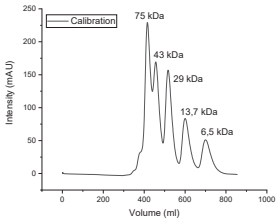


*Figure S 8: FPLC calibration run.*

run for Tko S24e, showing that it elutes at a volume compatible with a MW of 12 kDa as expected (to be compared with fig. S8). Figure S10 reports an SDS-PAGE analysis of Tba S24e after pooling

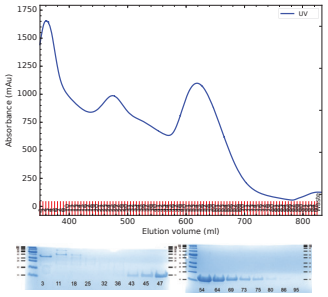


*Figure S 9: Example FPLC run. Fractions are given on the lower axis, and the SDS-PAGE analysis of some representative ones is given.*

of the FPLC fractions and concentration, attesting the purity of the protein. The only steps performed after this check were lyophilization, and then the subsequent dissolution in $D_2O$ immediately before the neutron experiments.

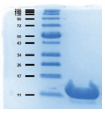

*Figure S 10: SDS-PAGE analysis of the final product (Tba S24e).*

## X-ray Crystallography and Molecular Dynamics

X-ray diffraction datasets were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble) at the ID30A-1 beam line (Massif1[4–10]) using a pixel detector (PILATUS3 2M) and auto-processed by XDSAPP package[11]. The following tables report the final refinement statistics for Tba S24e (Table S1) and for Tko S24e (Table S2).

| **Data collection** | **Tba S24e** | | |
|---|---|---|---|
| Space group | $P_1$ | | |
| Unit cell parameters (a,b,c) | 24.306 Å 35.067 Å 48.897 Å | | |
| ($\alpha$, $\beta$, $\gamma$) | 90.0° 90.0° 90.0° | | |
| Low resolution limit (Å) | 48.88 | 48.88 | 1.78 |
| High resolution limit (Å) | 1.74 | 9.06 | 1.74 |
| $R_{merge}$ (within I+/I-) | 0.094 | 0.070 | 0.599 |
| $R_{merge}$ (all I+ and I-) | 0.072 | 0.057 | 0.874 |
| $R_{meas}$ (within I+/I-) | 0.132 | 0.098 | 0.847 |
| $R_{meas}$ (all I+ and I-) | 0.103 | 0.081 | 1.236 |
| $R_{pim}$ (within I+/I-) | 0.094 | 0.070 | 0.599 |
| $R_{pim}$ (all I+ and I-) | 0.072 | 0.057 | 0.874 |
| $R_{merge}$ in top intensity bin | 0.050 | - | - |
| Total number of observations | 27153 | 194 | 1083 |
| Total number unique | 15670 | 120 | 682 |
| $I/\sigma_I$ | 4.8 | 12.6 | 0.6 |
| Mn(I) half-set correlation CC(1/2) | 0.989 | 0.976 | 0.446 |
| Completeness | 95.4 | 99.2 | 82.1 |
| Multiplicity | 1.7 | 1.6 | 1.6 |
| Mean($\chi^2$) | 0.75 | 0.96 | 0.61 |
| **Refinement** | | | |
| Number of reflections | 14884 (785) | | |
| after outlier rejection | 9956 (919) | | |
| Completeness | 87.81% (81.04%) | | |
| $R_{work}$ | 0.186 (0.281) | | |
| $R_{free}$ | 0.225 (0.350) | | |
| RMS(bonds) | 0.0107 | | |
| RMS(angles) | 1.71 | | |
| MolProbity score | 1.64 | | |
| Ramachandran outliers | 0.00% | | |
| favored | 96,82% | | |

*Table S 1: Refinement parameters for Tba S24e.*

| **Data collection** | **Tko S24e** | | |
|---|---|---|---|
| Space group | $P_{21}$ | | |
| Unit cell parameters (a,b,c) | 23.55 Å 55.27 Å 37.57 Å | | |
| ($\alpha$, $\beta$, $\gamma$) | 90.0° 98.98° 120.0° | | |
| Low resolution limit | 37.118 | 37.118 | 1.687 |
| High resolution limit | 1.659 | 4.499 | 1.659 |
| $R_{merge}$ (all I+ and I-) | 0.033 | 0.027 | 0.32 |
| $R_{meas}$ (all I+ and I-) | 0.043 | 0.034 | 0.4 |
| $R_{pim}$ (all I+ and I-) | 0.027 | 0.021 | 0.324 |
| $R_{pim}$ (within I+/I-) | 0.025 | 0.024 | 0.240 |
| Total number of observations | 15467 | 1445 | 639 |
| Total number unique | 9962 | 581 | 453 |
| $I/\sigma_I$ | 10.5 | 24.4 | 2.2 |
| Completeness | 87.9 | 97.8 | 79.8 |
| Multiplicity | 1.6 | 2.5 | 1.4 |
| CC(1/2) | 0.997 | 0.994 | 0.498 |
| **Refinement** | | | |
| Resolution range | 30.815 - 1.659 (1.718 - 1.659) | | |
| Number of reflections | 9956 (919) | | |
| after outlier rejection | 9956 (919) | | |
| Completeness | 87.81% (81.04%) | | |
| $R_{work}$ | 0.2219 (0.2817) | | |
| $R_{free}$ | 0.2673 (0.3601) | | |
| RMS(bonds) | 0.009 | | |
| RMS(angles) | 1.314 | | |
| Ramachandran outliers | 0.00 % | | |
| favored | 98,94 % | | |
| Rotamer outliers | 1,2 % | | |
| C-beta deviations | 3 | | |
| Clashscore | 10.09 | | |

*Table S 2: Refinement parameters for Tko S24e.*

Figure S11 shows the structural alignment of the resulting crystal structures. Unfortunately, the fast dynamics of the C terminals of both proteins impacted the measured electron densities from the crystals, impeding the characterization of the last three substitutions (E94Q, K96Q and G98E).

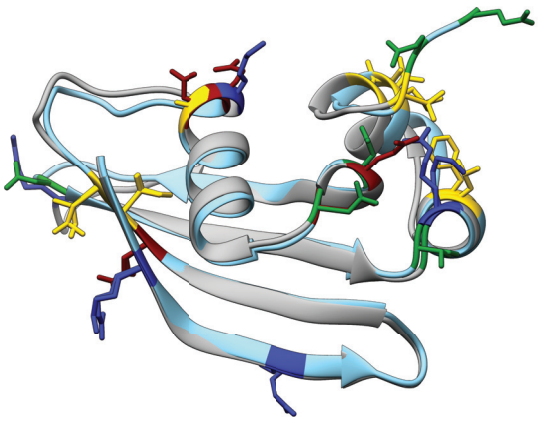

*Figure S 11: Structural alignment of the two crystal structures. Tba S24e is shown as gray ribbon and Tko S24e as light blue. Substitutions are shown in sticks and colored by residue type (red for acidic, blue for basic, green for polar and yellow for hydrophobic).*

Molecular dynamics simulations were performed as described in the methods section of the main text. Figure S12 shows the $C_\alpha$ RMSD values for the two NPT trajectories, with the first frame taken as a reference.

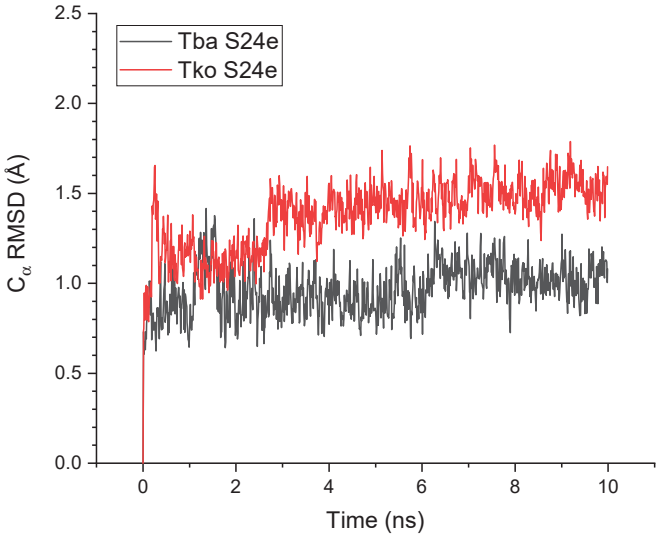


*Figure S 12: $C_\alpha$ RMSD values calculated for the two NPT trajectories with the first frame as a reference (Tba S24e in black, Tko S24e in red).*

# References

[1] Doster, W., Cusack, S. & Petry, W. Dynamical transition of myoglobin revealed by inelastic neutron scattering. *Nature* **337**, 754–756 (1989).

[2] Vannier, P., Marteinsson, V. T., Fridjonsson, O. H., Oger, P. & Jebbar, M. Complete genome sequence of the hyperthermophilic, piezophilic, heterotrophic, and carboxydotrophic archaeon thermococcus barophilus mp. *Journal of Bacteriology* **193**, 1481–1482 (2011).

[3] Fukui, T. *et al.* Complete genome sequence of the hyperthermophilic archaeon thermococcus kodakaraensis kod1 and comparison with pyrococcus genomes. *Genome research* **15**, 352–363 (2005).

[4] Svensson, O., Gilski, M., Nurizzo, D. & Bowler, M. W. Multi-position data collection and dynamic beam sizing: recent improvements to the automatic data-collection algorithms on MASSIF-1. *Acta Crystallographica Section D* **74**, 433–440 (2018).

[5] Svensson, O., Malbet-Monaco, S., Popov, A., Nurizzo, D. & Bowler, M. W. Fully automatic characterization and data collection from crystals of biological macromolecules. *Acta Crystallographica Section D* **71**, 1757–1767 (2015).

[6] Svensson, O., Gilski, M., Nurizzo, D. & Bowler, M. W. A comparative anatomy of protein crystals: lessons from the automatic processing of 56000 samples. *IUCrJ* **6**, 822–831 (2019).

[7] Bowler, M. W. *et al.* MASSIF-1: a beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules. *Journal of Synchrotron Radiation* **22**, 1540–1547 (2015).

[8] Bowler, M. W., Svensson, O. & Nurizzo, D. Fully automatic macromolecular crystallography: the impact of massif-1 on the optimum acquisition and quality of data. *Crystallography Reviews* **22**, 233–249 (2016).

[9] Nurizzo, D. *et al.* RoboDiff: combining a sample changer and goniometer for highly automated macromolecular crystallography experiments. *Acta Crystallographica Section D* **72**, 966–975 (2016).

[10] Hutin, S. *et al.* Fully autonomous characterization and data collection from crystals of biological macromolecules. *JoVE* e59032 (2019).

[11] Kabsch, W. *XDS*. *Acta Crystallographica Section D* **66**, 125–132 (2010).

# CONCLUSIONS

## 4.1 General summary

The results presented in this thesis encompass the dynamical and structural properties of piezophilic proteins compared side by side with piezosensitive orthologs, to understand the adaptation strategy to HHP in piezophilic Archaea. We performed our investigation by using neutron scattering, NMR spectroscopy, X-ray crystallography and MD simulations.

We studied two couples of protein from *Thermococcus barophilus* and *Thermococcus kodakarensis*, namely the Phosphomannose Isomerases and the Ribosomal Proteins S24e. They both gave different and complementary information that all points in the same direction.

By comparing the results from the neutron experiments performed on both proteins, we were able to confirm the findings from the previous whole-cell experiments. This shows that both proteins represent the average proteome really well and are good candidates for this study, as expected.

The two EINS experiments on both couples gave very similar results, hinting to some common properties of pressure-adapted proteins. Both the piezophilic proteins have very stable energy landscapes, meaning that their conformational freedom is not impacted by high pressure. Both experiments point out that this is a result of decreased hydration and decoupling of the protein surface from the surrounding water layer. On the contrary, both piezosensitive proteins display the opposite behaviour, that is, their energy landscape is greatly impacted by pressure, hampering their flexibility as a

result of water penetration.

Another common feature of both couples of proteins is the presence of bigger internal cavities in the piezosensitive ones. This is addressed in the piezophilic proteins by using Isoleucine instead of Valine, in order to occupy more volume in the hydrophobic core and enhance their pressure stability. Moreover the reduction of cavity volume is another explanation for the superior stability of the piezophilic proteins' energy landscape. The use of 2D HP-NMR spectroscopy to probe the two ribosomal proteins allowed us to gain direct evidence on the importance of this substitution, and to confirm that it should not be regarded as a simple conservative mutation.

In the case of the PMIs, we found that the reduced interaction between the piezophilic protein and water stems from the decoupling of cooperative protein-water dynamics, which are evident, on the contrary, for its piezosensitive counterpart. The destabilization of the active site in the piezosensitive protein is also highlighted from both the QENS data and the simulations, as well as the lower stability of its quaternary structure, where another Ile-Val substitution is located.

A smaller hydration shell and a more flexible proteome in the piezophilic species was one of the main results from the whole-cell studies that preceded this work[50]. Our results confirm this, and evidence how both couples of proteins present a behaviour in line with the average of their respective proteomes. Moreover, it should also be considered that the proteins we studied belong to different classes, the PMIs being enzymes, and the ribosomal proteins having a structural role, therefore, the generality of these results is even more compelling.

The first conclusions can be therefore drawn by contextualising all these results together. This discussion is divided in two sections, focusing respectively on the hydrophobic core and the surface of proteins.

## 4.2 The hydrophobic core

Globular proteins always present a hydrophobic core, and its formation is one of the main driving forces for protein folding. The four proteins considered for this study all possess a rather large core, relatively to their

size. This is expected, as all of them are resistant to high-temperature[3], being part of hyperthermophilic organism's proteomes.

Their cores are mostly well conserved among orthologs, except for a limited number of substitutions. These are very conservative, and would thus appear not to be part of any specific adaptation pathway. We have shown that this is not the case, as those substitutions act on a physical quantity that is of the utmost importance when studying pressure response : volume. It has long been known how cavities affect the pressure stability of proteins[63,64], however, this work provides the first direct evidence of their involvement in adaptation to HHP.

As previously discussed, pressure unfolding is not a concern at the relatively low pressure values at which life exists. The main effect is a population shift among conformational substates characterized by different specific volumes[68]. Therefore, reducing void volume into the protein core enables the piezophilic proteins to have a stable energy landscape that is not impacted by pressure application, and permits to retain optimal protein functionality. This hypothesis is supported by all our experimental findings, and explains why it is not necessary to use extremely bulky residues (e.g. Phe or Trp) to eliminate cavities, as they would destabilize the protein structure and probably lead to a different folding pathway altogether. Instead, the additional methyl group in isoleucine compared to valine, which it often substitutes, appears to be big enough to occupy more space and stabilize the core (and the dimer interface as well in the case of PMIs), but also small enough not to perturb the folding pathway of the proteins.

Interestingly, the Ile-Val substitution has been found to be important for the psychrophilic bacteria of the genus *Colwellia* too[10], Isoleucine being preferentially employed in the piezophilic species. However, another study performed on sequences of orthologous proteins from *Pyrococcus furiosus* and *Pyrococcus abyssi*[44], two closely-related Thermococcales, argues the opposite. By defining a Pressure Asymmetry Index (PAI), the author finds a negative correlation between PAI and amino acid molecular weight, meaning that smaller residues are found to be more "barophilic" than bulkier ones, and suggests that smaller residues may increase flexibility and improve packing efficiency. Although this result could seem to contradict our findings, we argue that such study does not take into account the 3D structure of the proteins, and therefore does not distinguish between residues in the core or

on the surface, ignoring which of the two contributions dominates the final result. Indeed, smaller residues on the surface of a protein would decrease its surface area, and thus its interaction with water, which is a desirable property for pressure-adapted proteins. However, it is known that proteins with a high sequence similarity often have very similar folding pathways, and a folded protein has to bear many structural constraints. Thus, it seems unlikely that a small residue, that would also have less dynamical degrees of freedom, could offer more flexibility and better packing in the core, as demonstrated by our results. However, on average, the number of substitutions appears to be higher on the surface than in the core, which has to be substantially conserved to maintain the same fold across orthologous proteins from different species, explaining why surface substitutions would be the dominant contribution in a genomic study. This shows that knowledge of the structure and dynamics is fundamental when studying volume effects on proteins.

Overall, the main conclusion that can be drawn from our results is that the role of the hydrophobic core is crucial for pressure adaptation. Cavities need to be eliminated to increase structural stability under HHP, but this has to be done without perturbing the folding pathway of the protein by means of a few, very targeted substitutions. It is therefore the result of a very fine balance where even a single methyl group makes a substantial difference.

## 4.3 The surface

As opposed to the core, the surface of globular proteins is mostly hydrophilic, to facilitate the interaction with surrounding water and render it soluble. One of the main consequence of this is that the internal motions of proteins are coupled to those of the solvent[93,131,132]. Protein dynamics are also correlated throughout the whole molecule[133,134], therefore surface effect could affect internal motions as well.

The effects of pressure on the surface of proteins are less obvious and intuitive than those in the core. However, the ubiquitous presence of charges makes electrostatic effects dominant. The decoupling of protein-water cooperative dynamics is evident from our data, and it could stem from the fact that some substitutions lead to the formation of salt bridges, i.e. E49N in the

ribosomal proteins. This means that more residues will interact with each other rather than with water, effectively neutralizing parts of the piezophilic protein surface. To give a concrete example, in the case of the aforementioned substitution, the total charge does not change, as next to E49N another substitution takes place (N48D). As a consequence of this, in Tko S24e Asp48 is too far from Arg75 to form a salt bridge, leaving two charges free to interact with water. In Tba S24e, Glu49 easily forms a salt bridge with Lys75, so that water cannot interact with either of them.

This concept could appear counter-intuitive, as it is known that water is essential for function and acts as a "lubricant" for protein dynamics[135–137]. While this is true at low pressure, at higher values the water in the hydration shell of proteins becomes electrostricted[138,139], and its dielectric constant rises[140], favouring electrostatic interactions with the protein surface, and at the same time inhibiting the flexibility of the residues it interacts with. Hence, decoupling their surface residues' dynamics from water is fundamental in piezophilic proteins to retain their flexibility under extreme conditions.

The downside of this picture is that identifying a precise substitution pattern would become challenging, as each protein would need "personalized" substitutions in order to neutralize their surface and decouple them from the surrounding water. However, at the same time, it would explain why sequence-based methods have not succeeded in the identification of this pattern simply by looking for the employment of a specific charge, on average, in the whole piezophilic proteome. Nonetheless, this new information together with the findings on the hydrophobic core represent a considerable advance in the field of pressure adaptation.

## 4.4 Future perspectives

The main focus of this work has been the explanation of high pressure adaptation based on the difference in sequence between the studied proteins. However, other factors have been discussed that are possibly involved in the adaptation strategy of these organisms, namely crowding[51] and the use of organic osmolytes[52]. The field of molecular crowding is expanding very rapidly, and many interesting studies are presently ongoing about its effects on the dynamics of proteins. These investigations could be extended in the

future to pressure-adapted proteins, in order to study the importance of crowding as a pressure adaptation strategy.

Concerning organic osmolytes, they have been shown to have great impact on protein dynamics and its pressure response[141,142]. In fact, we have already performed a neutron scattering experiment on the two PMIs in the presence of trimethylamine-N-oxide (TMAO). These results are not discussed in this thesis, but they will be the subject of an article that is already in preparation.

The work described in this thesis opens many future perspectives for the study of pressure adaptation. First, to further confirm these results, different mutants could be constructed by targeting the amino acid substitutions we found to be relevant. Presently, site-directed mutagenesis has become a routine technique, and it could be employed to further confirm (or deny) our results. Moreover, in the specific case of the PMIs, a pressure-dependent biochemical characterization could be performed, to investigate the activity of these proteins as a function of pressure, and to probe its correlation with our results. However, this would not be a trivial problem, as the currently state-of-the-art activity test for the isomerization of mannose-6-phosphate to fructose-6-phosphate requires the use of two additional proteins[143], therefore a new assay would have to be envisioned to eliminate additional variables in the test.

Further information could also be obtained by studying other pressure-adapted proteins. Two more couples from *T. barophilus* and *T. kodakarensis* were also selected at the beginning of this study, and they could be employed in a similar manner to extend our data set and investigate the generality of the properties we described. Proteins from other organisms could also be adopted to this end, as thorough experimental characterizations of pressure-adapted proteins are lacking in the literature.

Finally, with the great progress done in Artificial Intelligence (AI) and protein structure prediction[117], an entire database containing every protein with a known primary sequence has been made available to the scientific community. This opens the possibility for structural studies on entire proteomes[144], given the availability of computational resources.

For instance, the structural information from the database could be employed to separate residues that are in the core or on the surface of the selected proteins, and this could be then cross-correlated with the results

from a statistical analysis on their sequences, dividing the amino acids into these two subsets. Furthermore, the cavities could be systematically studied on the structures, and the residues involved in cavity formation could form another subset for the analysis. Therefore, a substitutional pattern supported by structural evidence could be built and then tested, i.e. by mutating piezosensitive proteins and then probing their pressure stability.

A similar work, suitably confirmed by experimental evidence, could undoubtedly answer the long-standing question about pressure adaptation. It would give engineers the tools to construct pressure-resistant proteins at will, and it would help evolutionary biologists in the search for the origin of life.

# RÉSUMÉ

## Introduction

Depuis les origines de la vie sur la Terre, les organismes ont dû s'adapter aux variations des conditions environnementales de leurs habitats pour survivre. Aujourd'hui, on connait une grande diversité d'organismes qui ont colonisé des environnements qui sont considérés comme extrêmes, car il ont des conditions environnementales très éloignées de celles de l'Homme (pH neutre, température moyenne de 20°C, pression atmosphérique, présence d'oxygène, etc., fig. 1.1). On estime par exemple que plus de 70% de la vie sur la Terre prospère sous Haute Pression Hydrostatique (HHP, >100 atm)[1], dans la profondeur de la mer ainsi que dans la croûte terrestre et les sédiments[4] (fig. 1.8). Outre l'intérêt principal de comprendre comment des organismes peuvent réaliser l'intégralité de leur cycle de vie à des pression supérieures à 1000 fois la pression atmosphérique, l'intérêt pour ces investigations est motivé par la recherche des origines de la vie sur Terre, puisqu'une des hypothèses les plus acceptées aujourd'hui propose que la vie aurait pu se développer au niveau des sources hydrothermales de l'océan profond, là où il y avait une grande quantité d'énergie thermique à disposition, ainsi qu'une protection efficace des radiations intenses émis par le soleil de la Terre jeune[15,16]. Enfin, la compréhension des mécanismes d'adaptation à l'HHP permettra des innovations biotechnologiques pour l'ingénierie de protéines résistantes à la pression, dont l'utilisation étendra le domaine en P et T des biotechnologies.

Il est bien établi aujourd'hui que la HHP a des effets négatifs sur la crois-

sance des microorganismes non-piézophiles, aussi bien au niveau cellulaire qu'au niveau moléculaire. Notamment, les protéines sont parmi les composants cellulaires les plus impactées par l'HHP, alors que leur stabilité est essentielle pour la survie des organismes. Pour la cellule, il est nécessaire de maintenir son protéome fonctionnel. Pour réaliser cela, les protéines ont dû évoluer au course des générations pour s'adapter et optimiser leurs propriétés physico-chimiques. Cette adaptation est obtenue en modifiant leur composition en acides aminés, qui est codée dans la séquence du génome. Notamment, des études conduites sur des cellules entières[49–52] ont identifié les différences existant entre les protéomes de deux Archées hyperthermophiles quasi-isogéniques, *Thermococcus barophilus* et *Thermococcus kodakarensis*. Ces expériences tiennent compte uniquement des propriétés moyennes du protéome, et par conséquent ne permettent pas d'identifier les motifs de substitution caractéristiques de l'adaptation à l'HHP. Cependant, ils montrent que l'interaction entre le protéome et sa couche d'eau d'hydratation est une des clés adaptatives chez les piézophiles. Plusieurs stratégies ont été envisagées pour identifier les patrons de substitution en acides aminés caractéristiques de l'adaptation à la HHP[10,43,44], mais aucune n'a permis la mise en évidence d'une signature génomique associée à la HHP à ce jour.

Dans ce travail de thèse, on essaye de répondre à cette question en étudiant les propriétés dynamiques et structurales de deux couples de protéines appartenant aux deux organismes quasi-isogéniques mentionnés ci-dessus, i.e. *Thermococcus barophilus* et *Thermococcus kodakarensis*. J'ai mise en place une approche "Bottom-up" pour décrire la différence de dynamique des protéines piézophiles et piézosensibles à l'échelle moléculaire, afin d'identifier le patron de substitution d'acides aminés ainsi que la logique moléculaire à ces substitutions, et ensuite de construire un modèle d'adaptation à la HHP. Pour cela, j'ai utilisé la diffusion neutronique pour explorer leur dynamique, la cristallographie à rayons X pour élucider leur structure, et la résonance magnétique nucléaire (NMR) pour obtenir des informations spécifiques sur le substitutions des acides aminés.

## Méthodes

Les protéines cibles ont été choisies car 1) nous avons pu montrer qu'elles sont sous sélection positive pour l'HHP, 2) leur fonction est connue, et 3)

leur taille est compatible avec la spectroscopie NMR (taille < 150 résidus). Deux couples de protéines, les *Phosphomannose Isomerases* et les *Protéines Ribosomales S24e* de *T. barophilus* et de *T. kodakarensis*, ont été produites par surexpression recombinante dans *Escherichia coli*, et purifiées par traitement thermique et par Chromatographie par Exclusion de Taille (SEC).

La diffusion neutronique est une technique qui a été beaucoup utilisée pour l'étude de la dynamique des protéines. Le neutron n'étant pas chargé, il va interagir avec les noyaux des atomes et non avec les électrons comme les rayons X. De plus, l'interaction des neutrons avec les noyaux d'hydrogène est bien plus importante que celle avec les autres noyaux, et leur énergie est compatible avec les excitations thermiques dans la matière condensée ou molle. Ces caractéristiques font du neutron une sonde idéale pour l'étude de la matière biologique. On a utilisé la diffusion incohérente (EINS et QENS) pour analyser la dynamique des protéines choisies sur une échelle de temps de 10-100 ps.

La spectroscopie de Résonance Magnétique Nucléaire (NMR) donne des informations très précises sur l'environnement chimique des spins nucléaires qui absorbent de la radiation à radio-fréquence (RF) quand ils se trouvent dans un champ magnétique. Les techniques de NMR multidimensionnelles comme l'Heteronuclear Single Quantum Coherence (HSQC) en particulier, permettent de suivre les changements conformationnels des protéines au niveau de chaque acide aminé de la protéine individuellement. Cette technique, utilisée conjointement avec la HHP, rend possible l'étude des changements de volume et de compressibilité des protéines en conditions de HHP et de température contrôlées.

La cristallographie à rayons X permet l'élucidation des structures des protéines à très haute résolution, au niveau atomique. Toutefois, les atomes d'hydrogène ne sont pas visibles. Pour cette raison, on a utilisé les simulations de Dynamique Moléculaire (MD) pour les calculer et relâcher les structures vers une conformation compatible avec celle des protéines en solution. Ceci a permis une étude détaillée des cavités, qui a été mis en corrélation avec les résultats des autres techniques.

## Résultats

Les résultats des expériences de diffusion neutronique confirment ce qui avait été observé pour les cellules entières, c'est à dire que les protéines piézophiles sont plus *souples*, mais ont une résilience supérieure comparées à leur homologues piézosensibles. Ceci est la conséquence de l'interaction différente qu'elles ont avec leur eau d'hydratation. Les protéines piézophiles ont la capacité de résister à la pénétration de l'eau et elles maintiennent leur flexibilité même sous HHP, grâce à la stabilité supérieure de leur paysage énergétique.

Une autre caractéristique commune des deux couples de protéines est la présence de cavités plus grandes dans le cœur hydrophobe des piézosensibles. Les substitutions Isoleucine-Valine permettent aux homologues piézophiles de mieux occuper l'espace vide, réduisant ainsi le volume de telles cavités et optimisant leur résistance à la HHP. Les expérience de NMR multidimensionnelles nous ont donné la preuve directe de ce mécanisme, et ont mis en évidence l'importance des substitutions conservatives dans le processus d'adaptation à la HHP.

## Conclusions

Les résultats obtenus montrent premièrement l'importance des substitutions conservatives dans le cœur hydrophobe des protéines. Aux valeurs de pression rencontrées dans les environnements où la vie existe, le dépliement des protéines n'est pas un risque. Néanmoins, une diminution des variations de volume entre les états conformationnels des protéines augmente la stabilité de leur paysage énergétique sous pression, et ceci revêt une importance capitale pour maintenir leur activité optimale sous conditions extrêmes. Toutefois, des substitutions avec des résidus trop encombrants pourraient perturber la voie de repliement d'une protéine. Il faut donc maintenir le balance entre l'occupation de volume et la stabilité du repliement.

Les substitutions des acides aminés à la surface des protéines sont aussi fondamentales, puisqu'elles règlent l'interaction avec l'eau d'hydratation. Il est bien connu que l'eau est essentielle pour la fonction des protéines,

puisqu'elle sert de *lubrifiant* pour leur dynamique[135–137]. Pourtant, une interaction trop forte avec l'eau pourrait être nuisible sous HHP, étant donné que l'eau pourrait empêcher la dynamique des résidus à la surface en raison de l'électrostriction[138,139]. Pour cette raison, le découplage de la dynamique de ces résidus de celle de l'eau est important. Le point négatif de cette situation est que le substitutions nécessaires pour atteindre ce découplage pourraient être assez différentes pour chaque protéine, ce qui pourrait rendre la recherche d'un motif d'adaptation général assez complexe. Toutefois, cela explique assez bien le raisons de la défaillance des méthodes génomiques qui ont été tentées jusqu'ici pour identifier les motifs de substitutions liés à l'adaptation à la HHP.

En résumé, mes résultats représentent un avancement considérable dans la recherche sur l'adaptation à la HHP. Je démontre l'importance des informations structurales et dynamiques pour étudier cette adaptation et j'ai identifié les premiers indices sur la stratégie d'adaptation des Archées extrêmophiles à la HHP. Dans le futur, grâce aux avances en matière d'Intelligence Artificielle (AI), il devrait être possible de corréler un grand nombre de structures pour confirmer/infirmer ces conclusions, et pour construire un modèle pour l'adaptation à l'HHP au niveau génomique.

# ACKNOWLEDGEMENTS

It could be argued that this very section is the one where every PhD student gets the biggest "writer's block". The amount of people that contribute to such works is so overwhelming that the fear of forgetting anyone overcomes even the strongest of us. It has been a long journey, very difficult at times, but entirely worth it nevertheless. I cannot even start to list the amount of things I have learned during these three years, all thanks to all the people who have been beside me throughout this adventure.

I will start with an anecdote, happened on the first day I set foot in my new lab. Philippe came to me with a vial and said "This is the DNA for your protein, now we'll put it in cells and then we're gonna cut it!". At the time I was just a simple student that had just graduated his Physics master, and had only seen proteins coming out of freezers into Sigma Aldrich boxes. I had no idea if he had some kind of subatomic scissors, or he was going to use black magic to do that, but then, with care and patience (lots of it!), he taught me that and so much more. I owe it to him if my mind today is able to reason in a much more interdisciplinary way, always trying to see things from different perspectives to gain a deeper understanding of the problem. He has always pushed me to be better in this sense and I am very grateful.

My relationship with Judith actually goes back further than the start of this project. I first met her during a summer school at the ILL, when I had my first-hand contact with the magical world of neutron scattering, and then during an internship I did at ILL with another PhD student of hers and Phil's, now Dr. Loreto Misuraca. She has seen and overseen my whole transition from "Oh my god is the reactor gonna explode if I touch that?!" to "Yeah

you can leave me alone on the instrument, I'm good", with a sprinkle of "Judith... The pressure cell just exploded..." in the meantime. She has made me the experimentalist I am today, while always being one of the nicest and kindest people I have ever met in my life. Her always positive attitude towards scientific discussion is outstanding and I can confidently say that she is the prototype that every PhD supervisor should follow if they want to be better towards their students, including her lightning-fast response times whenever there was something to review. I really do not know how to put into words how grateful I am to her, so I sincerely hope that this will give you an idea.

I would also like to acknowledge Professor Christian Roumestand, who welcomed me into his lab in Montpellier to perform the NMR experiments. He taught me a lot about a technique I knew existed, but I ignored the immense power it can have when used by someone as knowledgeable as him. I also owe him more or less a tonne of coffee that I consumed in the meantime. Thanks also to the rest of the team at the CBS for their invaluable help, Cécile, François, Philippe Barthe, Karine and Mounia.

Another very deserved acknowledgment goes to Alessio an Marek, my local contacts for my experiments at ILL, the High-Pressure division of the Sample Environment group, James and Claude, and Miguel Angel Gonzalez for invaluable help with the software.

A very big thanks goes to the members of my team *Microbiologie des Environnements Extrêmes*, particularly Stéphane and Yoann, for having supported and guided me especially during the first time when I was doing things that were completely obscure to me. It is also thanks to them if I eventually managed.

And what would a PhD experience be without office banter ? It is well known that like organisms adapt to high pressure, PhD students always find a way to make their time more enjoyable, at least when they're not tirelessly working. For that I have to thank Marta, Maxime, Alvaro, Josie and Margot. The discussions with them always made everything better, be it about science, football, food or how annoying the French bureaucracy is.

Last, but definitely not least, I would like to thank all the friends I made along the way. After moving to a new city you feel lost and alone, even though you just started an exciting project and everyone at work is nice. Thankfully, this

did not last long for me, as I was lucky enough to meet wonderful people that have never ever made me feel alone anymore. Donato, Andrea, Elisa, Samuele, Joe, Lucio, Francesca, Roberto, Giulia and Dharshini, thank you!

I am also deeply grateful to my family for having always supported and encouraged me throughout the good and bad moments I've had during these three years. THANK YOU!

# REFERENCES


[1] Jannasch, H. W. & Taylor, C. D. Deep-sea microbiology. *Annual review of microbiology* **38**, 487–487 (1984).

[2] Jaenicke, R. Stability and stabilization of globular proteins in solution. *Journal of Biotechnology* **79**, 193–203 (2000).

[3] Brininger, C., Spradlin, S., Cobani, L. & Evilia, C. The more adaptive to change, the more likely you are to survive : Protein adaptation in extremophiles. *Seminars in Cell and Developmental Biology* **84**, 158 – 169 (2018).

[4] Oger, P. M. & Jebbar, M. The many ways of coping with pressure. *Research in Microbiology* **161**, 799–809 (2010).

[5] Woese, C. R. Bacterial evolution. *Microbiological Reviews* **51**, 221–271 (1987).

[6] Bicout, D. J. & Field, M. J. Stochastic dynamics simulations of macromolecular diffusion in a model of the cytoplasm of escherichia coli. *The Journal of Physical Chemistry* **100**, 2489–2497 (1996).

[7] Tourte, M., Schaeffer, P., Grossi, V. & Oger, P. M. Membrane adaptation in the hyperthermophilic archaeon pyrococcus furiosus relies upon a novel strategy involving glycerol monoalkyl glycerol tetraether lipids. *Environmental Microbiology* **24**, 2029–2046 (2022).

[8] Oger, P. M. & Cario, A. Adaptation of the membrane in archaea. *Biophysical Chemistry* **183**, 42–56 (2013). Biomolecular systems under extreme environmental conditions.

[9] Salvador-Castell, M. *et al.* Characterisation of a synthetic archeal membrane reveals a possible new adaptation route to extreme conditions. *Communications Biology* **4**, 653 (2021).

[10] Peoples, L. M. *et al.* Distinctive gene and protein characteristics of extremely piezophilic colwellia. *BMC Genomics* **21**, 692 (2020).

[11] Simonato, F. *et al.* Piezophilic adaptation : a genomic point of view. *Journal of Biotechnology* **126**, 11–25 (2006).

[12] McDonald, J. H., Grasso, A. M. & Rejto, L. K. Patterns of temperature adaptation in proteins from Methanococcus and Bacillus. *Molecular Biology and Evolution* **16**, 1785–1790 (1999).

[13] McDonald, J. H. Patterns of Temperature Adaptation in Proteins from the Bacteria Deinococcus radiodurans and Thermus thermophilus. *Molecular Biology and Evolution* **18**, 741–749 (2001).

[14] Kauzmann, W. Thermodynamics of unfolding. *Nature* **325**, 763–764 (1987).

[15] Miller, S. L. & Bada, J. L. Submarine hot springs and the origin of life. *Nature* **334**, 609–611 (1988).

[16] Martin, W., Baross, J., Kelley, D. & Russell, M. J. Hydrothermal vents and the origin of life. *Nature Reviews Microbiology* **6**, 805–814 (2008).

[17] Miller, S. L. A production of amino acids under possible primitive earth conditions. *Science* **117**, 528–529 (1953).

[18] Kvenvolden, K. *et al.* Evidence for extraterrestrial amino-acids and hydrocarbons in the murchison meteorite. *Nature* **228**, 923–926 (1970).

[19] Mojzsis, S. J. *et al.* Evidence for life on earth before 3,800 million years ago. *Nature* **384**, 55–59 (1996).

[20] Daniel, I., Oger, P. & Winter, R. Origins of life and biochemistry under high-pressure conditions. *Chem. Soc. Rev.* **35**, 858–875 (2006).

[21] Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).

[22] Dill, K. A. & MacCallum, J. L. The protein-folding problem, 50 years on. *science* **338**, 1042–1046 (2012).

[23] Creighton, T. E. Protein folding. *Biochemical journal* **270**, 1 (1990).

[24] Dill, K. A. Dominant forces in protein folding. *Biochemistry* **29**, 7133–7155 (1990).

[25] Chien, A., Edgar, D. B. & Trela, J. M. Deoxyribonucleic acid polymerase from the extreme thermophile thermus aquaticus. *Journal of Bacteriology* **127**, 1550–1557 (1976).

[26] Privalov, P. L. & Gill, S. J. Stability of protein structure and hydrophobic interaction. vol. 39 of *Advances in Protein Chemistry*, 191 – 234 (Academic Press, 1988).

[27] Tehei, M. & Zaccai, G. Adaptation to high temperatures through macromolecular dynamics by neutron scattering. *The FEBS Journal* **274**, 4034–4043 (2007).

[28] Chan, C.-H., Yu, T.-H. & Wong, K.-B. Stabilizing salt-bridge enhances protein thermostability by reducing the heat capacity change of unfolding. *PLOS ONE* **6**, 1–8 (2011).

[29] Tsukamoto, T. *et al.* X-ray crystallographic structure of thermophilic rhodopsin : implications for high thermal stability and optogenetic function. *Journal of Biological Chemistry* **291**, 12223–12232 (2016).

[30] Cacciapuoti, G., Fuccio, F., Petraccone, L., Del Vecchio, P. & Porcelli, M. 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**, 1136–1143 (2012).

[31] Groussin, M. & Gouy, M. Adaptation to environmental temperature is a major determinant of molecular evolutionary rates in archaea. *Molecular biology and evolution* **28**, 2661–2674 (2011).

[32] Gianese, G., Bossa, F. & Pascarella, S. Comparative structural analysis of psychrophilic and meso- and thermophilic enzymes. *Proteins : Structure, Function, and Bioinformatics* **47**, 236–249 (2002).

[33] De Maayer, P., Anderson, D., Cary, C. & Cowan, D. A. Some like it cold : understanding the survival strategies of psychrophiles. *EMBO reports* **15**, 508–517 (2014).

[34] Eisenberg, H., Mevarech, M. & Zaccai, G. Biochemical, structural, and molecular genetic aspects of halophilism. vol. 43 of *Advances in Protein Chemistry*, 1–62 (Academic Press, 1992).

[35] Madern, D., Ebel, C. & Zaccai, G. Halophilic adaptation of enzymes. *Extremophiles* **4**, 91–98 (2000).

[36] Ortega, G., Diercks, T. & Millet, O. Halophilic protein adaptation results from synergistic residue-ion interactions in the folded and unfolded states. *Chemistry & biology* **22**, 1597–1607 (2015).

[37] Müller-Santos, M. *et al.* First evidence for the salt-dependent folding and activity of an esterase from the halophilic archaea haloarcula marismortui. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* **1791**, 719–729 (2009).

[38] Arakawa, T., Yamaguchi, R., Tokunaga, H. & Tokunaga, M. Unique features of halophilic proteins. *Current Protein and Peptide Science* **18**, 65–71 (2017).

[39] Auerbach, G. *et al.* Lactate dehydrogenase from the hyperthermophilic bacterium thermotoga maritima : the crystal structure at 2.1 å resolution reveals strategies for intrinsic protein stabilization. *Structure* **6**, 769–781 (1998).

[40] Price, P. B. A habitat for psychrophiles in deep antarctic ice. *Proceedings of the National Academy of Sciences* **97**, 1247–1251 (2000).

[41] Pradel, N. *et al.* The first genomic and proteomic characterization of a deep-sea sulfate reducer : Insights into the piezophilic lifestyle of desulfovibrio piezophilus. *PLOS ONE* **8**, 1–11 (2013).

[42] Kawano, H. *et al.* Differential pressure resistance in the activity of rna polymerase isolated from shewanella violacea and escherichia coli. *Extremophiles* **8**, 367–375 (2004).

[43] Shrestha, U. R. *et al.* Effects of pressure on the dynamics of an oligomeric protein from deep-sea hyperthermophile. *Proceedings of the National Academy of Sciences* **112**, 13886–13891 (2015).

[44] Di Giulio, M. A comparison of proteins from pyrococcus furiosus and pyrococcus abyssi : barophily in the physicochemical properties of amino acids and in the genetic code. *Gene* **346**, 1–6 (2005).

[45] Woese, C. R., Dugre, D., Dugre, S., Kondo, M. & Saxinger, W. On the fundamental nature and evolution of the genetic code. In *Cold Spring Harbor symposia on quantitative biology*, vol. 31, 723–736 (Cold Spring Harbor Laboratory Press, 1966).

[46] Tourte, M. *et al.* Novel intact polar and core lipid compositions in the pyrococcus model species, p. furiosus and p. yayanosii, reveal the largest lipid diversity amongst thermococcales. *Biomolecules* **10** (2020).

[47] Tourte, M. *et al.* The exploration of the thermococcus barophilus lipidome reveals the widest variety of phosphoglycolipids in thermococcales. *Frontiers in Microbiology* **13** (2022).

[48] Salvador-Castell, M., Tourte, M. & Oger, P. M. In search for the membrane regulators of archaea. *International Journal of Molecular Sciences* **20** (2019).

[49] Peters, J. *et al.* Deep sea microbes probed by incoherent neutron scattering under high hydrostatic pressure. *Zeitschrift für Physikalische Chemie* **228**, 1121 – 1133 (2014).

[50] Martinez, N. *et al.* High protein flexibility and reduced hydration water dynamics are key pressure adaptive strategies in prokaryotes. *Scientific Reports* **6** (2016).

[51] Golub, M. *et al.* The effect of crowding on protein stability, rigidity, and high pressure sensitivity in whole cells. *Langmuir* **34**, 10419–10425 (2018).

[52] Salvador-Castell, M. *et al.* The first study on the impact of osmolytes in whole cells of high temperature-adapted microorganisms. *Soft Matter* **15**, 8381–8391 (2019).

[53] Smelt, J. Recent advances in the microbiology of high pressure processing. *Trends in food science & technology* **9**, 152–158 (1998).

[54] Farkas, D. F. & Hoover, D. G. High pressure processing. *Journal of food science* **65**, 47–64 (2000).

[55] Macgregor Jr, R. B. The interactions of nucleic acids at elevated hydrostatic pressure. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **1595**, 266–276 (2002).

[56] Winter, R. & Jeworrek, C. Effect of pressure on membranes. *Soft Matter* **5**, 3157–3173 (2009).

[57] Balny, C., Masson, P. & Heremans, K. High pressure effects on biological macromolecules : from structural changes to alteration of cellular processes. *Biochimica et biophysica acta* **1595**, 3–10 (2002).

[58] Le Châtelier, H. Sur un Énoncé gènèral des lois des Équilibres chimiques. *Comptes-rendus de l'Académie des sciences* **99**, 786–789 (1884).

[59] Bridgman, P. The coagulation of albumen by pressure. *Journal of Biological Chemistry* **19**, 511–512 (1914).

[60] Hawley, S. A. Reversible pressure-temperature denaturation of chymotrypsinogen. *Biochemistry* **10**, 2436–2442 (1971).

[61] Zipp, A. & Kauzmann, W. Pressure denaturation of metmyoglobin. *Biochemistry* **12**, 4217–4228 (1973).

[62] Frauenfelder, H. *et al.* Proteins and pressure. *Journal of Physical Chemistry* **94**, 1024–1037 (1990).

[63] Royer, C. A. Revisiting volume changes in pressure-induced protein unfolding. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1595**, 201–209 (2002).

[64] Roche, J. *et al.* Cavities determine the pressure unfolding of proteins. *Proceedings of the National Academy of Sciences* **109**, 6945–6950 (2012).

[65] Grigera, J. R. & McCarthy, A. N. The behavior of the hydrophobic effect under pressure and protein denaturation. *Biophysical Journal* **98**, 1626–1631 (2010).

[66] Kharakoz, D. P. Protein compressibility, dynamics, and pressure. *Biophysical Journal* **79**, 511–525 (2000).

[67] Erlkamp, M. *et al.* Influence of pressure and crowding on the subnanosecond dynamics of globular proteins. *The Journal of Physical Chemistry B* **119**, 4842–4848 (2015).

[68] Librizzi, F., Carrotta, R., Peters, J. & Cupane, A. The effects of pressure on the energy landscape of proteins. *Scientific reports* **8**, 2037 (2018).

[69] Frauenfelder, H., Parak, F. & Young, R. D. Conformational substates in proteins. *Annual Review of Biophysics and Biophysical Chemistry* **17**, 451–479 (1988).

[70] Meinhold, L., Smith, J. C., Kitao, A. & Zewail, A. H. Picosecond fluctuating protein energy landscape mapped by pressure–temperature molecular dynamics simulation. *Proceedings of the National Academy of Sciences* **104**, 17261–17265 (2007).

[71] Avagyan, S., Vasilchuk, D. & Makhatadze, G. I. Protein adaptation to high hydrostatic pressure : Computational analysis of the structural proteome. *Proteins : Structure, Function, and Bioinformatics* **88**, 584–592 (2020).

[72] Marteinsson, V. *et al.* Thermococcus barophilus sp. nov., a new barophilic and hyperthermophilic archaeon isolated under high hydrostatic pressure from a deep-sea hydrothermal vent. *International Journal of Systematic and Evolutionary Microbiology* **49**, 351–359 (1999).

[73] Atomi, H., Fukui, T., Kanai, T., Morikawa, M. & Imanaka, T. Description of thermococcus kodakaraensis sp. nov., a well studied hyperthermophilic archaeon previously reported as pyrococcus sp. kod1. *Archaea* **1**, 204953 (1900).

[74] Vannier, P., Marteinsson, V. T., Fridjonsson, O. H., Oger, P. & Jebbar, M. Complete genome sequence of the hyperthermophilic, piezophilic, heterotrophic, and carboxydotrophic archaeon thermococcus barophilus mp. *Journal of Bacteriology* **193**, 1481–1482 (2011).

[75] Fukui, T. *et al.* Complete genome sequence of the hyperthermophilic archaeon thermococcus kodakaraensis kod1 and comparison with pyrococcus genomes. *Genome research* **15**, 352–363 (2005).

[76] Cario, A., Jebbar, M., Thiel, A., Kervarec, N. & Oger, P. M. Molecular chaperone accumulation as a function of stress evidences adaptation to high hydrostatic pressure in the piezophilic archaeon thermococcus barophilus. *Scientific Reports* **6**, 29483 (2016).

[77] Clancy, S. & Brown, W. Translation : Dna to mrna to. *Nature Education* (2008).

[78] Baneyx, F. Recombinant protein expression in escherichia coli. *Current opinion in biotechnology* **10**, 411–421 (1999).

[79] Burgess-Brown, N. A. *et al.* Codon optimization can improve expression of human genes in escherichia coli : A multi-gene study. *Protein expression and purification* **59**, 94–102 (2008).

[80] Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology* **189**, 113–130 (1986).

[81] Chadwick, J. The existence of a neutron. *Proceedings of the Royal Society of London A* **136**, 692–708 (1932).

[82] Néel, L. Propriétés magnétiques des ferrites ; ferrimagnétisme et antiferromagnétisme. In *Annales de physique*, vol. 12, 137–198 (1948).

[83] Squires, G. L. *Introduction to the Theory of Thermal Neutron Scattering* (Cambridge University Press, 2012), 3 edn.

[84] Bée, M. *Quasielastic Neutron Scattering : Principles and Applications in Solid State, Chemistry, Biology and Materials Science* (Adam Hilger, 1988).

[85] Fermi, E. Motion of neutrons in hydrogenous substances. *Ricerca Scientifica* **2**, 13–52 (1936).

[86] Doster, W., Cusack, S. & Petry, W. Dynamical transition of myoglobin revealed by inelastic neutron scattering. *Nature* **337**, 754–756 (1989).

[87] Zaccai, G. How soft is a protein ? a protein dynamics force constant measured by neutron scattering. *Science* **288**, 1604–1607 (2000).

[88] Bicout, D. J. & Zaccai, G. Protein flexibility from the dynamical transition :a force constant analysis. *Biophysical Journal* **80**, 1115–1123 (2001).

[89] Zeller, D., Telling, M. T. F., Zamponi, M., García Sakai, V. & Peters, J. Analysis of elastic incoherent neutron scattering data beyond the gaussian approximation. *The Journal of Chemical Physics* **149**, 234908 (2018).

[90] Stoeckli, A. *et al.* Dynamics of hydrogen bonds in carboxylic acids. *Physica B+ C* **136**, 161–164 (1986).

[91] Engler, N., Ostermann, A., Niimura, N. & Parak, F. G. Hydrogen atoms in proteins : Positions and dynamics. *Proceedings of the National Academy of Sciences* **100**, 10243–10248 (2003).

[92] Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–72 (2008).

[93] Jansson, H., Bergman, R. & Swenson, J. Role of solvent for the dynamics and the glass transition of proteins. *The Journal of Physical Chemistry B* **115**, 4099–4109 (2011).

[94] Jansson, H., Kargl, F., Fernandez-Alonso, F. & Swenson, J. Dynamics of a protein and its surrounding environment : A quasielastic neutron scattering study of myoglobin in water and glycerol mixtures. *The Journal of chemical physics* **130**, 05B613 (2009).

[95] Khodadadi, S., Pawlus, S. & Sokolov, A. Influence of hydration on protein dynamics : combining dielectric and neutron scattering spectroscopy data. *The Journal of Physical Chemistry B* **112**, 14273–14280 (2008).

[96] Grimaldo, M. *et al.* Hierarchical molecular dynamics of bovine serum albumin in concentrated aqueous solution below and above thermal denaturation. *Phys. Chem. Chem. Phys.* **17**, 4645–4655 (2015).

[97] Rusevich, L. *et al.* Protein and solvent dynamics of the water-soluble chlorophyll-binding protein (wscp). In *EPJ Web of Conferences*, vol. 83, 02016 (EDP Sciences, 2015).

[98] Hall, P. L. & Ross, D. Incoherent neutron scattering functions for random jump diffusion in bounded and infinite media. *Molecular Physics* **42**, 673–682 (1981).

[99] Singwi, K. S. & Sjölander, A. Diffusive motions in water and cold neutron scattering. *Phys. Rev.* **119**, 863–871 (1960).

[100] Rabi, I. I., Zacharias, J. R., Millman, S. & Kusch, P. A new method of measuring nuclear magnetic moment. *Physical review* **53**, 318 (1938).

[101] Günther, H. *NMR spectroscopy : basic principles, concepts and applications in chemistry* (John Wiley & Sons, 2013).

[102] Cavanagh, J., Fairbrother, W. J., Palmer III, A. G. & Skelton, N. J. *Protein NMR spectroscopy : principles and practice* (Academic press, 1996).

[103] Overhauser, A. W. Polarization of nuclei in metals. *Physical Review* **92**, 411 (1953).

[104] Akasaka, K. & Li, H. Low-lying excited states of proteins revealed from nonlinear pressure shifts in 1h and 15n nmr. *Biochemistry* **40**, 8665–8671 (2001).

[105] Kitahara, R., Hata, K., Li, H., Williamson, M. P. & Akasaka, K. Pressure-induced chemical shifts as probes for conformational fluctuations in proteins. *Progress in Nuclear Magnetic Resonance Spectroscopy* **71**, 35–58 (2013).

[106] Dubois, C., Herrada, I., Barthe, P. & Roumestand, C. Combining high-pressure perturbation with nmr spectroscopy for a structural and dynamical characterization of protein folding pathways. *Molecules* **25** (2020).

[107] Wagner, G., Pardi, A. & Wuethrich, K. Hydrogen bond length and proton nmr chemical shifts in proteins. *Journal of the American Chemical Society* **105**, 5948–5949 (1983).

[108] Sitkoff, D. & Case, D. A. Theories of chemical shift anisotropies in proteins and nucleic acids. *Progress in Nuclear Magnetic Resonance Spectroscopy* **32**, 165–190 (1998).

[109] La Penna, G., Mori, Y., Kitahara, R., Akasaka, K. & Okamoto, Y. Modeling 15n nmr chemical shift changes in protein backbone with pressure. *The Journal of Chemical Physics* **145**, 085104 (2016).

[110] Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662–666 (1958).

[111] Bowler, M. W. *et al.* MASSIF-1 : a beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules. *Journal of Synchrotron Radiation* **22**, 1540–1547 (2015).

[112] Taylor, G. The phase problem. *Acta Crystallographica Section D : Biological Crystallography* **59**, 1881–1890 (2003).

[113] Hauptman, H. Phasing methods for protein crystallography. *Current opinion in structural biology* **7**, 672–680 (1997).

[114] Giacovazzo, C. Phasing in crystallography : a modern perspective. *Rendiconti Lincei* **24**, 71–76 (2013).

[115] Rossmann, M. & Blow, D. M. The detection of sub-units within the crystallographic asymmetric unit. *Acta crystallographica* **15**, 24–31 (1962).

[116] Rossmann, M. G. The molecular replacement method. *Acta Crystallographica Section A : Foundations of Crystallography* **46**, 73–82 (1990).

[117] Jumper, J. *et al.* Highly accurate protein structure prediction with alphafold. *Nature* **596**, 583–589 (2021).

[118] Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nature structural biology* **9**, 646–652 (2002).

[119] Hollingsworth, S. A. & Dror, R. O. Molecular dynamics simulation for all. *Neuron* **99**, 1129–1143 (2018).

[120] Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints : molecular dynamics of n-alkanes. *Journal of computational physics* **23**, 327–341 (1977).

[121] Miyamoto, S. & Kollman, P. A. Settle : An analytical version of the shake and rattle algorithm for rigid water models. *Journal of computational chemistry* **13**, 952–962 (1992).

[122] Darden, T., York, D. & Pedersen, L. Particle mesh ewald : An n log (n) method for ewald sums in large systems. *The Journal of chemical physics* **98**, 10089–10092 (1993).

[123] Jo, S., Kim, T., Iyer, V. G. & Im, W. Charmm-gui : A web-based graphical user interface for charmm. *Journal of Computational Chemistry* **29**, 1859–1865 (2008).

[124] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **79**, 926–935 (1983).

[125] Berendsen, H. J., Postma, J. v., Van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics* **81**, 3684–3690 (1984).

[126] Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation : The langevin piston method. *The Journal of Chemical Physics* **103**, 4613–4621 (1995).

[127] Krone, M. *et al.* Visual analysis of biomolecular cavities : State of the art. In *Computer Graphics Forum*, vol. 35, 527–551 (Wiley Online Library, 2016).

[128] Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. Castp 3.0 : computed atlas of surface topography of proteins. *Nucleic acids research* **46**, W363–W367 (2018).

[129] Edelsbrunner, H., Facello, M., Fu, P. & Liang, J. Measuring proteins and voids in proteins. In *Proceedings of the Twenty-Eighth Annual Hawaii International Conference on System Sciences*, vol. 5, 256–264 (IEEE, 1995).

[130] Liang, J., Edelsbrunner, H., Fu, P., Sudhakar, P. V. & Subramaniam, S. Analytical shape computation of macromolecules : Ii. inaccessible cavities in proteins. *Proteins : Structure, Function, and Bioinformatics* **33**, 18–29 (1998).

[131] Paciaroni, A., Cinelli, S. & Onori, G. Effect of the environment on the protein dynamical transition : a neutron scattering study. *Biophysical journal* **83**, 1157–1164 (2002).

[132] Fomina, M., Schirò, G. & Cupane, A. Hydration dependence of myoglobin dynamics studied with elastic neutron scattering, differential scanning calorimetry and broadband dielectric spectroscopy. *Biophysical Chemistry* **185**, 25–31 (2014).

[133] Levantino, M. *et al.* Ultrafast myoglobin structural dynamics observed with an x-ray free-electron laser. *Nature communications* **6**, 1–6 (2015).

[134] Librizzi, F., Caliò, A. & Cupane, A. Dynamical properties of myoglobin in an ultraviscous water-glycerol solvent investigated with elastic neutron scattering and ftir spectroscopy. *Journal of Molecular Liquids* **268**, 242–248 (2018).

[135] Fenimore, P. W., Frauenfelder, H., McMahon, B. & Young, R. Bulk-solvent and hydration-shell fluctuations, similar to $\alpha$-and $\beta$-fluctuations in glasses, control protein motions and functions. *Proceedings of the National Academy of Sciences* **101**, 14408–14413 (2004).

[136] Caliskan, G. *et al.* Protein and solvent dynamics : how strongly are they coupled ? *The Journal of chemical physics* **121**, 1978–1983 (2004).

[137] Schiró, G., Caronna, C., Natali, F. & Cupane, A. Molecular origin and hydration dependence of protein anharmonicity : an elastic neutron scattering study. *Physical Chemistry Chemical Physics* **12**, 10215–10220 (2010).

[138] Danielewicz-Ferchmin, I., Banachowicz, E. & Ferchmin, A. Protein hydration and the huge electrostriction. *Biophysical chemistry* **106**, 147–153 (2003).

[139] Merzel, F. & Smith, J. C. Is the first hydration shell of lysozyme of higher density than bulk water ? *Proceedings of the National Academy of Sciences* **99**, 5378–5383 (2002).

[140] Uematsu, M. & Frank, E. Static dielectric constant of water and steam. *Journal of Physical and Chemical Reference Data* **9**, 1291–1306 (1980).

[141] Cinar, H. *et al.* Temperature, hydrostatic pressure, and osmolyte effects on liquid–liquid phase separation in protein condensates : Physical chemistry and biological implications. *Chemistry – A European Journal* **25**, 13049–13069 (2019).

[142] Al-Ayoubi, S., Schummel, P., Golub, M., Peters, J. & Winter, R. Influence of cosolvents, self-crowding, temperature and pressure on the sub-nanosecond dynamics and folding stability of lysozyme. *Physical Chemistry Chemical Physics* **19**, 14230–14237 (2017).

[143] Gracy, R. W. & Noltmann, E. A. Studies on phosphomannose isomerase : I. isolation, homogeneity measurements, and determination of some physical properties. *Journal of Biological Chemistry* **243**, 3161–3168 (1968).

[144] Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).